①

# AN AUTOMATED TASK LOAD INDEXING SYSTEM

Robert S. Kennedy, James G. May
William P. Dunlap, Marshall B. Jones,
and Martin G. Smith
Essex Corporation
1040 Woodcock Road, Suite 227
Orlando, FL 32803

9 March 1991

Final Report for Period August 1990-March 1991
Contract Number N62269-90-C-0279

DOD DISTRIBUTION STATEMENT

Approved for public release;
distribution unlimited

Prepared for
Naval Air Development Center
Warminster, PA 18974-5000

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

# REPORT DOCUMENTATION PAGE

| | |
|---|---|
| 1a. REPORT SECURITY CLASSIFICATION | 1b. RESTRICTIVE MARKINGS |
| 2a. SECURITY CLASSIFICATION AUTHORITY | 3. DISTRIBUTION/AVAILABILITY OF REPORT |
| 2b. DECLASSIFICATION/DOWNGRADING SCHEDULE | |
| 4. PERFORMING ORGANIZATION REPORT NUMBER(S) | 5. MONITORING ORGANIZATION REPORT NUMBER(S) |

| | | |
|---|---|---|
| 6a. NAME OF PERFORMING ORGANIZATION<br>Essex Corporation | 6b. OFFICE SYMBOL *(If applicable)* | 7a. NAME OF MONITORING ORGANIZATION<br>DCMAD Baltimore |
| 6c. ADDRESS *(City, State, and ZIP Code)*<br>1040 Woodcock Road, Suite 227<br>Orlando, FL 32803 | | 7b. ADDRESS *(City, State, and ZIP Code)*<br>200 Towsontown Blvd., West<br>Towson, MD 21204-5299 |
| 8a. NAME OF FUNDING/SPONSORING ORGANIZATION<br>Naval Air Development Center | 8b. OFFICE SYMBOL *(If applicable)*<br>N62269 | 9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER<br>N62269-90-C-0279 |

8c. ADDRESS *(City, State, and ZIP Code)*
Warminster, PA 18974-5000

10. SOURCE OF FUNDING NUMBERS

| PROGRAM ELEMENT NO. | PROJECT NO. | TASK NO. | WORK UNIT ACCESSION NO. |
|---|---|---|---|
| | | | |

11. TITLE *(Include Security Classification)*
An Automated Task Load Indexing System (Unclassified)

12. PERSONAL AUTHOR(S)
R. S. Kennedy, J. G. May, W. P. Dunlap, M. B. Jones, and M. G. Smith

| 13a. TYPE OF REPORT | 13b. TIME COVERED | 14. DATE OF REPORT *(Year, Month, Day)* | 15. PAGE COUNT |
|---|---|---|---|
| Final | FROM 8/90 TO 3/91 | 8 March 1991 | 44 |

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

| FIELD | GROUP | SUB-GROUP |
|---|---|---|
| | | |
| | | |

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
Work load, intelligent systems, eye movements, cockpit resource allocation, expert systems, adaptive function allocation

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

The accelerating tempo of military operations increases the task demands and work-related stresses imposed on human operators. As new technological designs are incorporated into military systems, a metric is required to ensure that human cognitive work load limits are not exceeded. Subjective techniques are available, but generally depend on the self-report of users conducted after the work is finished. A need exists for objective methods which can follow time-course changes as work is performed and to develop biocybernetic approaches to task allocation. Some progress has been achieved in using a variety of electrophysiological techniques, including EEG and neural-evoked potentials. However, these objective techniques tend to be intrusive, relatively artificial, and nonportable. We believe there are more readily obtainable bioelectric measures that can serve as simple external indicants of work load, and which can eventually be bundled in

| | |
|---|---|
| 20. DISTRIBUTION/AVAILABILITY OF ABSTRACT<br>☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS | 21. ABSTRACT SECURITY CLASSIFICATION<br>UNCLASSIFIED |

| | | |
|---|---|---|
| 22a. NAME OF RESPONSIBLE INDIVIDUAL<br>C. P. Skriver | 22b. TELEPHONE *(Include Area Code)*<br>(215) 441-2457 | 22c. OFFICE SYMBOL<br>Code 6021 |

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

19. Abstract (continued)

portable, "vest-pocket" systems to be employed during studies of cockpit display function allocation. In Phase I, we conducted electrophysiological recordings of the action of the eye while subjects attended and performed on tasks with different visual demands and task difficulty (complexities). The bioelectric actions we recorded included eye movements (frequency, amplitude, velocity, acceleration, range, etc.) and eye blinks (frequency, duration, etc.) and provided for 25 direct measures. Two developments were undertaken: (1) Two different objective scales of task load were developed based on visual and mental task demands. These served as independent variables. Different dependence on visual system measures (e.g., frequency of eye movements, blink duration) on tasks with differing visual and mental requirements were differentially predictive of the two objective scales. (2) Customized computer software for automated presentation of the tasks and scoring of the electrophysiological responses were developed for desk-top personal computers. This software system was mechanized and implemented and is now fully up and running. Because within-subject changes correlated at a statistically meaningful level with the visual task demands and with the mental work load, this procedure holds promise as a method for calibrating individuals against known visual and mental task loading so that laboratory-based systems like NADC's reconfigurable cockpit can be used to study adaptive function allocation. In Phase II this system would be further developed to: (1) run on line and in real time and be validated in an aircraft or simulator system to determine quality assurance boundaries; (2) be made compatible with standard data analytic packages (BMD, SAS, SPSS); (3) be made fully portable for field usage; (4) create algorithms which will permit partition between mental task loading versus visual task demands in cockpit workplace design and development; (5) create field manuals for use by systems developers to objectively assess visual work load parameters of various aspects of aviation activity; (6) be field tested at a Navy development laboratory as part of their work load R&D programs.

The Phase I effort was the first step in the design of an automated task load analysis system for biocybernetic modification and function allocation of aircraft cockpit display systems for rapid and portable on-site measurement of aviation cockpits and workspaces. The availability of such a package could provide aircraft manufacturers and others with common metrics for conducting human factors engineering design, test and evaluation of workstations of all kinds. Nonintrusive visually based measures of an operator's interest or attention could have far-reaching commercial applications.

Accession For
NTIS CRA&I ✓
DTIC TAB
Unannounced
Justification

By
Distribution/

Availability Codes

Dist
A-1

## ACKNOWLEDGMENTS

The authors are indebted to Elizabeth Voss and Barbara Holmes, research assistants. The success of this project was due much to their dedication and willingness to appear at the laboratory in order to accommodate the irregular schedules of subjects. They did this willingly and with good spirit. Additionally, thanks are given to Pat Patrizzi for her usual excellent word processing support and to Mary Kay Osteen for her editorial assistance.

THIS PAGE INTENTIONALLY LEFT BLANK

## TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# INTRODUCTION

In aviation's early days, the pilot's task was only to fly the aircraft and what few cockpit instruments there were provided information about the aircraft. As time went on, the tactical potential of the aircraft was recognized and cockpit "real estate" reflected this new application with geometric increases in the number of instruments installed. More recently not only has workload increased, but the pilot's task has changed significantly and our exploding technologies permit the presentation of information via multiple media and some of these are new approaches. Therefore, now, not only is it necessary to control one's position in space, but the pilot uses all sensory channels sometimes in new ways as the manager of a weapons system. All three uniformed services have become increasingly aware of these changes in workload, task content and display media. Major development programs are under way to seek solutions to unburden the pilot such as improved function allocation, automation, intelligent and adaptive systems, appropriate media selection, etc.

One such program is the Reconfigurable Cockpit (RC) currently on-going at the Naval Air Development Center, Warminster, PA. In that effort, multiple visual displays can be configured and changed so that the effects on human workload and performance can be empirically determined in laboratory studies. A recent report (Morrison, Gluckman, & Deaton, 1990) describes their first cockpit automation study where task difficulty and other behavioral parameters were explored. Future plans call for evaluation of the effects of changing the characteristic of one or more displays, possibly within an on-going tactical mission segment.

We concur with Gopher and Braune (1984) that the three most logical possibilities are subjective, behavioral or electrophysiological techniques. Performance _per se_ can be measured or indirect measures of performance like physiological assays or subjective reports may be obtained. All these have merit and disadvantages and experimentation such as this is not without difficulties. The value of direct measures of performance is self-evident, but such measures require considerable development and usually still have metric problems (lack of stability, low reliability, different tasks have different indicants, etc). Indirect measures have the obvious disadvantage that they are not the performance itself, but can sometimes helpfully augment the behavioral measures, particularly if sufficient linkages to performance or aspects of the task can be demonstrated beforehand. The major advantages of indirect measures is that they may not be as likely to be task specific and so differing combinations of display configurations and workload can be indexed against a common metric. Also many of the metric problems can be solved by high data acquisition and analysis rates obtained over full mission segments and in real time. We discuss these three methods more fully below.

## SUBJECTIVE MEASURES

A vast majority of the scientific literature on the effects of work load has involved subjective measures where a performer makes a conscious judgment regarding how well he or she has performed. Several subjective measurement scales have been developed. These scales include the Cooper-Harper rating scale (Cooper & Harper, 1969), a modification of the Cooper-Harper rating scale (Sheridan & Stassen, 1979), Likert-type scales of fatigue (Gray, 1980), motion sickness symptomatology (Lane & Kennedy, 1988), bipolar rating techniques (Hart, Childress, & Bortolussi, 1981) and the derivative and more current NASA TLX system (Hart, 1990; Hart & Staveland, 1988), mood scales (cf., e.g., Storm, 1980), alertness scales (Peacock, Glube, Miller, & Clune, 1983), the Subjective Workload Assessment Technique (SWAT) (Reid, Shingledecker, & Eggemeier, 1981; Reid, Shingledecker, Nygren & Eggemeier, 1981), and Gopher and Braune's (1984) application of magnitude estimation originally developed by S. S. Stevens (1951). Generally, these scales are employed by the individual operators and after the work is performed, but peer evaluations are also popular (e.g., Gal, 1975) and very useful, and sometimes ratings or protocols (Ericsson & Simon, 1984; Berbaum, Kennedy & Hettinger, 1991) can be used while the work is ongoing and scored afterwards. Another method uses behaviorally-anchored rating scales (Campbell, Dunnette, Arvey, & Hellervik, 1973) which obtain operators' ratings of effort and of inclination or disinclination to continue with the task.

A comparison of studies utilizing these subjective measures is complicated by the lack of standardization, the use of different rating dimensions, and inconsistency of results between tasks. Additionally, and perhaps more importantly, these scales often show low correlations with objective measures of task performance (Wickens & Yeh, 1983) so that their usefulness in predicting work load demands may be questioned. We believe that one of the reasons for these low correlations is that often comparisons are being made between metrics which are subject dependent (EVEN THOUGH THEY MAY BE OBJECTIVE) with others which are subject independent (WHETHER THEY BE OBJECTIVE OR SUBJECTIVE). For example, the Subjective Workload Assessment Technique (SWAT) (Reid et al., 1981) is generally used to evaluate a system's work load characteristics and, when mean scores are employed, are subject-independent. However, eye blink (Goldstein, Stern, & Bauer, 1985), while an objective measure, is largely subject dependent, being different in different subjects. It may come as no surprise that the two metrics may not be correlated. We plan to attend to this logical distinction in our work on this project as it can affect measurement precision. We discuss this issue more completely elsewhere (Kennedy, May, Jones, & Fowlkes, 1989), but it is a recurring theme in this report and we believe should be developed further. Inattention to the implication of this model can invalidate an entire experiment or systems analysis.

In summary, the advantages of using subjective scales lie in their ease of administration and the lack of need for extensive instrumentation that may interfere with the performance of the primary task. Subjective measures have been used to assess the relationship between performance and work load in physical tasks (Borg, 1978), cognitive tasks (Borg, 1978), and manual control tasks (Cooper & Harper, 1969). Although significant correlations were obtained in all of these studies, the correlations were among subjective judgments of work load and not with objective measures of performance. Thus, subjective methods are subject to criticism of "method variance" and results may be limited in generality in that they yield information available from only one component of a task, that is, that which enters the performer's consciousness, and therefore may neglect aspects of information processing that are automatic, but which nevertheless consume processing capacity. A major drawback is that often it is of interest to evaluate time-course changes within a technical mission or mission segment as task difficulty changes or as different display options are studied. Subjective methods do not adapt well to these real-time, on-line requirements.

## BEHAVIORAL MEASURES

A second approach to the measurement of work load involves obtaining direct behavioral (performance) measures. Here, an evaluation is made of an operator's overt task behavior (e.g., speed or accuracy of performance). This method draws heavily on "resource theory" (Wickens, 1984). One variation on such an approach involves administering a primary task simultaneously with an additional, secondary task (Shingledecker, 1982). As the difficulty level of the primary task is increased, a point will be reached when the operator's processing capacity is exceeded, and the performance decrement on the secondary task will be inversely proportional to the primary load. If the primary task consumes all processing capacity, then there will be no functional reserve when a secondary task is added and performance will immediately degrade. With this method it is essential, of course, that the primary task remain primary, a problem not always handled satisfactorily (Damos, Bittner, Kennedy, & Harbeson, 1981; Kantowitz & Weldon, 1985). Although the behavioral approach appears to offer much promise with respect to the measurement of work load, a major drawback lies in the possibility that operators will develop a bias toward one task or another or effect criterion shifts during performance. For this reason it is important that the operator's performance be stabilized on the primary task to some predetermined level and monitored thereafter.

Another approach is to take a task which can be varied in difficulty level and show correlations between task loading and performance (Kennedy, 1971, Kennedy et al., 1989). A variation on this method would be to take differing tasks and use response per minute as an inverse index of work load. A third technique would be to determine the visual task demands by measuring the incident visual angles of the material placed before the subject and determining the amount of ocular motility necessary to perform the task. We plan to incorporate all three of these variations as measures of cognitive load against which we expect the eye movement parameters to vary. Here again, attention must be paid to attempts at correlating a performance measures (e.g., hits vs. percent correct on different work load tasks) with physiological or subjective metrics. Whether the experiment is planned on a within-subject vs. between-subject design is an additional consideration.

## ELECTROPHYSIOLOGICAL MEASURES

An alternative to using subjective and behavioral measures to study work load is to take direct physiological measures (e.g., heart rate and its derivatives, respiration, GSR, ERP, neuroendocrine changes) during sustained task performance. This method eliminates the possibility of distortion which may occur from subjective reports and generally does not interfere with the work. The drawback to this approach is that some measures of autonomic nervous system function may be more likely to reflect stress induced by the task (Shingledecker, 1982) rather than its cognitive load, and often these measures may lack stability and have insufficient reliability for statistical power (Cohen, 1977). Some of them (like ERP's) may intrude on the work to be performed (Krebs, Wingert, & Cunningham, 1977; O'Donnell, 1981) and nearly all of them require averaging (Goldstein et al., 1985; Donchin & Kramer, 1986) over many events, epochs, subjects, and exposures. We also repeat the caution described above that having objective measures like heart rate or aerobic capacity does not assure the measure is subject-independent. Indeed, quite the opposite is likely.

Research has been conducted to study physiological indicants and much of the recent work has employed blink (Stern 1990) and heart rate (Moray et al., 1986; Wierwille, Rahimi, & Casali, 1985; Hart & Hauser, 1987.

The research in this field is broad and we have reviewed it elsewhere (Kennedy, May, Jones, & Fowlkes, 1989). Several years ago, we also reviewed (Kennedy, 1972) several studies showing that aspects of eye movement activity were correlated with the mental state of the subject. However, most of the studies reviewed in that report were not directly addressed to the issue of work load but proceeded from arousal theory and habituation of the orienting response (Lynn, 1966). The thesis was not that eye movements could index arousal or other attentive states of the subject. Although heart rate measures are promising, we believe the studies on blink by Stern (1990) appear to be the furthest along in studying these measures. Moreover, while we know that blink frequency and blink duration can be influenced by emotion, motivation, and fatigue, as well as work load and visual demands, we think that these are sufficiently orderly relations that any bioelectrical recording of the visual system should include blink metrics.

Based on these relations, we began a basic research program for the U.S. Air Force Office of Scientific Research (Kennedy, May, Jones, & Fowlkes, 1989). In the Air Force experiment, while searching for a velocity indicant of arousal, we found that other aspects of eye movements (viz. aggregate eye movement extent) (May, Kennedy, Williams, Dunlap, & Brannan, 1999) could serve as an index of the objective information load imposed on the operator, particularly during auditory monitoring in the dark. Then, in a study for the USAF School of Aerospace Medicine at Brooks Air Force Base (Kennedy, Fowlkes, & Smith, 1989), we systematically varied task loading over a range of difficulties and sessions while eye movement elements and performance served as dependent variables. The objective was to determine whether task demands which are inherent in the stimulus (e.g., number of channels monitored, time on task) covaried with characteristics of the dependent variables. Although the approach was empirical, the elements selected for study followed from the theory we developed previously (Kennedy, 1972, 1978) and findings are reported in the scientific literature.

In that study, it was demonstrated that group performance on an auditory tone counting task varied with task difficulty ($p < .01$), suggesting that the work load had been successfully manipulated. Of the eye movement measures, acceleration of eye movements (or the slope of the regression line relating velocity of saccades to amplitude) bore a strong relationship to task difficulty, becoming steeper (faster eye movements) for 80% ($N = 15$) of the subjects ($p < .01$) when the high-task loading condition was compared to the lower. Eye movement frequency and eye blinks did not appear to be different in the two task loadings ($p = .40$). For the other eye movement measures explored: (a) amplitude of saccades tended to increase in the high task load condition for the majority of the subjects (61%), but this difference was not significant ($p < .15$); and (b) aggregated eye movement velocities under the high work load were generally greater in the high work load condition for the majority of the subjects (73%; $p < .02$). We also commented on an observed methodological problem in studies of this type. We have described this difficulty above in connection with attempting to correlate subject-dependent electrophysiological measures (e.g., heart rate) during work load with outcomes which might be subject independent (such as number of channels monitored). We elaborate on this paradigmatic approach to work load below.

It should not be surprising that eye movements, particularly those involving binocular foveal fixation and scanning, can represent very sensitive measures of alertness and of cognitive and motor performance. More than other sensory systems (Snider & Lowy, 1968), the eye has embryological connections to the cortex (Gregory, 1973; Weale, 1960). Eye movements are intimately related to the functional integrity of the Central Nervous System (CNS) centers thought to be responsible for arousal and alertness, particularly the reticular nuclei (Cohen, Feldman, & Diamond, 1969; Yules, Krebs, & Gault, 1966). Embryologically, the retina of the eye develops from the same substrate as the brain (Snider & Lowy, 1968; Gregory, 1973; Weale, 1960), and so when global integrating characteristics of the nervous system are sought (e.g., a rapid assessment of the operator's information processing and decision making capability status), scientists have focused on visual functions. Characteristics of spontaneous eye movements have been related to hemispheric specialization of cognitive, affective, and physiological variables (Bakan & Strayer, 1973). Eye movements have also been used to detect drug effects in terms of changes in fixation, gazing, and scanning of various visual stimuli involved in object and word recognition tasks (Monty, Hall, & Rosenberger, 1975). Quantitative relationships have been established among various components of eye movements, variations in instrument scanning strategies, task difficulty, and pilot work load (Krebs et al., 1977). Changes in amplitude of pursuit eye movements have also been used as objective measures of visual fatigue in a variety of visual tracking tasks (Malmstrom et al., 1981).

The relation between eye movements and mental work is not simply dependent upon the visual aspects of a task. This is best seen in nonvisual tasks. For example, Loren and Darrow (1962) compared electro-oculograms (EOG) to EEG recording during a mental multiplication task in a dark room with eyes closed. Increase in eye movements following onset of this nonvisual task was a much more reliable and consistent index of mental work than any of the EEG measures, although small reductions in occipital lobe alpha waves were noted. The reciprocal relationship between eye movements and incidence of alpha waves, the EEG index of lowered arousal, was shown clearly by Gardner (1967) who reported increased rates of eye movements during the absence of alpha waves in response to auditory verbal material. Increase in the velocity of saccadic eye movements as a function of heightened alertness induced by amphetamines in cats was reported by Crommelinck and Roucoux (1976).

In summary, the present report sets out to determine the feasibility of a family of visually based bioelectric measures, because we believe they hold the most promise for being automated, nonintrusive, and sensitive. The purpose is (a) to develop such measures, (b) demonstrate correspondence with other measures, (c) automate the scoring to the extent that it can be performed in real time. The goal is to provide a psychobiological index which can be bundled into a nonintrusive hardware/software system to measure cognitive task load in the human operator. Availability of such a system would permit indexing changes when new display concepts are introduced and signalling opportunities for adapting the display to the operator's capacities. We selected several eye movement and blink parameters to serve as indicants of task load characteristics. Visual tasks which differ in demand characteristics (field size, mental activity, complexity) serve as the behavioral controls. Our previous work (Kennedy, Fowlkes, & Smith, 1989; Kennedy, May, Jones, & Fowlkes, 1989) emphasized auditory tasks and suggested that measurement of some eye movement parameters (e.g., range, velocity, and acceleration) could provide a viable and sensitive indicant of cognitive work load. We hypothesized that further development of these indicants was warranted where eye movement measures were related to graded levels of a set of visually-based tasks which have factorial diversity. Workload of the tasks will be objectively (psychophysically) indexed by task characteristics (e.g., number of channels monitored) and by performance scores (e.g., number of correct responses per minute) and by visual demands (amount of eye movement activity per degree of retinal incidence). Phase I sets out to demonstrate the feasibility of these objectives. Successful development of such metrics would aid in development of a mathematical model to direct biocybernetic allocation of displayed information customized for the state (or trait) of the operator's capabilities.

In order to conduct a coherent program related to the neuroscience of work load, several technical problems must be addressed and solved. These technical problems become the five key tasks in this effort which are described below.

# METHOD

## TECHNICAL PLAN

The project had five main tasks:

(a) Administer cognitive tasks on portable microcomputers which differ on several dimensions (complexity, content, difficulty, visual demands, etc.). Administer computerized cognitive marker tests which tap a broad spectrum of task factors related to the pilot's task and work load metrics which relate to differing demand characteristics of these cognitive tasks.

(b) Analyze and relate eye movement indices (dependent variables) to different task features (independent variables). Distinguish between subject-dependent versus subject-independent metrics.

(c) Determine the feasibility of Mechanizing and implementing customized software for automated scoring of various eye movement parameters. Create a customized data acquisition and analysis system.

(d) Determine the feasibility of an on-line real-time analysis system to permit biocybernetic (adaptive) function allocation of cockpit displays.

(e) Begin formulation of a mathematical model for the neuroscience of work load. Develop metrics for task variables and dependent variables.

## GENERAL PROCEDURE

The above five tasks were divided into: (1) the Experimental Effort -- eye movement metrics which might respond to scaled work load were examined. (2) the System Development Effort -- focus was on development of eye movement algorithms and design of a portable system that can be fielded.

### Experimental Effort

In the first part of the Experimental Effort, eye movement metrics were to be scaled to work load on a tone counting task and the Automated Performance Test System (APTS) subtests. A within-subject design was employed and 30 subjects participated in five sessions conducted on separate days. Subjects were paid $5.00 for each hour that they participated. In Sessions 1 and 2 the subjects were familiarized with the experimental procedures and practiced the computerized tests. In Sessions 3 and 4 two levels of task difficulty of the tone counting test were administered (Kennedy, Fowlkes, & Smith, 1989) while the subjects were instrumented with surface electrodes and eye movements were measured. The plan was to analyze these data in order to determine which of the physiologic measures (amplitude, velocity, and acceleration of saccades, eye blink frequency and duration) bore a monotonic relationship with task load and time on task and the effects of practice on the measures. The final session entailed an administration of the full APTS battery while eye movements were recorded. Table 1 lists these tasks.

## Subjects

A pool of 30 college students were recruited to serve as subjects. Their ages ranged from 18 to 30 and they were paid $5.00 per hour for their participation. The subjects received informed consent forms and were otherwise used in accordance with established policies of human use according to nationally published guidelines. Each subject participated in five separate experimental sessions with electrodes being used in only three of these sessions. On days 1 and 2, subjects were given a computerized performance test five times each day with a short break in between each test. This test comprised selected portions of the APTS battery, described below, and is designed to assess human performance on various cognitive tasks.

## Work Load Tasks

The tasks used include the Counting Test (Jerison, 1956; Kennedy, 1972), and tests from the APTS (Kennedy, Baltzley, & Osteen, 1988; Kennedy, Baltzley, Wilkes, & Kuntz, 1989). The Counting Test, with which we have had success previously, has been modified for visual and auditory presentation. Use of this task is an extension of our previous work (Kennedy, Fowlkes, & Smith, 1989; Kennedy, May, Jones, & Fowlkes, 1989). The APTS is a battery of mental acuity tests which incorporates tests of verbal, spatial, and motor ability. The subtests selected for this study have been studied repeatedly by us in a series of experiments and we already have a considerable amount of information about their metric properties, effects of practice, etc. In particular, the factor structure of the test battery is rich and the task content varied so that a broader applicability of the eye movement metrics which surface can be tested. Performance on these tasks has also been shown to be related to military tasks (Turnage, Kennedy, Gilson, & Nolan, 1989; Turnage & Bliss, 1990) and to military test performance (Kennedy, Baltzley, & Osteen, 1988).

TABLE 1. Experimental Design Specifications*

| | Sessions | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| | 5 Trials each of APTS** | 5 Trials each of APTS | 2 Trials of Counting | 2 Trials of Counting | 1 Trial of APTS |
| THT | 20 | 20 | -- | -- | 20 |
| MK | 60 | 60 | -- | -- | 60 |
| STM | 60 | 60 | -- | -- | 60 |
| RT4 | 60 | 60 | -- | -- | 60 |
| MP | 60 | 60 | -- | -- | 60 |
| PC | 60 | 60 | -- | -- | 60 |
| GR | 60 | 60 | -- | -- | 60 |
| NPT | 20 | 20 | -- | -- | 20 |
| CS | 60 | 60 | -- | -- | 60 |
| Ch1 | -- | -- | 15 min | 15 min | -- |
| Ch2 | -- | -- | 15 min | 15 min | -- |

*Time in seconds unless otherwise noted.
**Subjects received practice equivalent to 1/2 the test time on the first administration only.

Legend:

THT = Two-Hand Tapping
MK = Manikin
STM = Sternberg's Short-Term Memory
RT4 = Reaction Time
MP = Math Processing
PC = Pattern Comparison
GR = Grammatical Reasoning
NPT = Nonpreferred Tapping
CS = Code Substitution
Ch1 = One Channel Counting
Ch2 = Two Channel Counting

## The Performance Tests

a. Tone Counting Task. The counting task was originally developed by Jerison (1956) and was later modified by Kennedy (1971) for auditory and visual presentation. The task presents at irregular intervals tones of three differing pitches (high, medium, and low) or lights (20 degrees left, middle, and 20 degrees right) 8, 6, and 5 times per minute respectively, for an extended period of time. For the simple (one channel) version, the subject is to count the low tone's/left light's occurence and when it has been presented four times press the low tone's button and begin counting again. Score is percent correct obtained by the formula hits/omits + commits + hits x 100. For the two and three channel versions of the test, the middle and high tones/right lights are also monitored and kept separately in the subject's working memory. In our experience, everyone can do the simple test almost without error for short periods, but errors occur with longer term monitoring periods. For the complex (three channel) test, performance is approximately 65% accurate, on the average, and almost no one can obtain 100% for any 5-minute epoch of performance.

b. APTS Tests. The mental acuity tests selected for inclusion in this study were from the microcompter-based APTS and have been researched and developed by us (e.g., Kennedy, Baltzley, Wilkes, & Kuntz, 1989). Performances have been shown to be related to military tasks (Turnage, Kennedy, Gilson, & Nolan, 1989) and to military test performance (Kennedy, Baltzley, & Osteen, 1988). The battery consists of a menu of fully automated human performance measures. Previous subtest evaluation research has demonstrated retest reliabilities $\geq$ 0.707, with mean, standard deviation, and differential stability achievable in 8 to 12 minutes of practice. The battery of subtests requires approximately 18 minutes of real-time testing. Candidate individual subtests for use in the proposed research are discussed below.

o Tapping (two tests: THT and NPT). Tapping tests are motor skills/performance tasks that may be placed throughout the test battery serving as a check against interfering factors during battery administration (e.g., boredom). The participant is required to press the indicated keys as fast as he or she can with two fingers from each hand (THT) or two-fingers from their nonpreferred (NPT) hand. There are two 10-second trials of each per session. Performance is based on the number of alternate key presses made in the allotted time.

o Grammatical Reasoning (GR). The Grammatical Reasoning test requires the participant to read and comprehend a simple statement about the order of two letters, A and B. Five grammatical transformations on statements about the relationship between the letters or symbols are made. The five transformations are: (1) active versus passive construction, (2) true versus false statements, (3) affirmative versus negative phrasing, (4) use of the verb "precedes" versus the verb "follows," and (5) A versus B mentioned first. There are 32 possible items arranged in random order. The subject's task is to respond "true" or false," depending on the verity of each statement, with performance scored according to the number of transformations correctly identified. Grammatical Reasoning is presented as one, 60-second trial of testing. The task is described as measuring "higher mental processes" with reasoning, logic, and verbal ability, important factors in test performance.

- Mathematical Processing (MP). This test includes arithmetical operations as well as value comparison of numeric stimuli. The participant performs one addition or subtraction operation in a single presentation. Then a response is made indicating whether the obtained total is greater or less than a prespecified value of five. The problems are randomly generated using only numbers 1 through 9. There are response deadlines for the problems corresponding to the demand characteristic of the test. Mathematical Processing is presented as one 60-second trial of testing.

- Code Substitution (CS). The Code Substitution test is a mixed associative memory and perceptual speed test with visual search, encoding, decoding, and rote recall, important performance factors. The computer displays nine alpha characters across the top of the screen and beneath them the digits 1 through 9 within parentheses. The subject's task is to associate the digit with the alpha character and to repeat the assigned digit code when presented with alpha characters. CS is presented as one, 60-second trial of testing. Previous studies of CS have indicated that the task is acceptable for use in repeated-measures research.

- Pattern Comparison (PC). The Pattern Comparison task requires the participant to determine if two simultaneously displayed patterns of asterisks are the same or different. Patterns are randomly generated with similar and different pairs presented in random order. Pattern Comparison is presented as one, 60-second trial of testing.

- Manikin (MK). This performance test involves the presentation of a simulated human figure in either a full-front or full-back facing position. The figure is shown to have two easily differentiated hand-held patterns. One of the two patterns hand-held matches a pattern appearing below the figure. The subject's task is to determine which hand of the figure holds the matching pattern and respond by pressing the appropriate microprocessor key. Pattern type, hand associated with the matching pattern and front-to-back figure orientation, are randomly determined. Manikin is presented as one, 60-second trial of testing. The MK test is a perceptual measure of spatial transformation of mental images and involves spatial ability.

- Short-Term Memory (STM). The Short-Term Memory Task presents a set of four letters for one second (positive set) followed by a series of single letters presented for two seconds (probe letters). The subject's task is to determine if the probe letters accurately represent the positive set and respond with the appropriate key press. Performance is based on the number of probes correctly identified. Short-Term Memory is described as a cognitive-type task which reflects short-term memory scanning rate.

- Reaction Time-Choice (RT-4). The Four-Choice Visual RT test involves the presentation of a visual stimulus and measurement of a response latency to the stimulus. The subject's task is to respond as quickly as possible with a keypress to a simple visual stimulus. On this test, four boxes are displayed and a short tone signals a "change" in the status of one of the boxes. One of the boxes visually changes and the subject responds as rapidly as possible with a keypress beneath the box. Reaction Time is presented as one, 60-second trial of testing. Simple RT has been described as a perceptual task responsive to environmental effects.

Scoring: "Hits" were used as the chief score for all tests, if appropriate. Other possible directly obtained metrics (viz. latency) are essentially redundant and we find derived metrics like percent correct can permit comparison across tasks, have the disadvantage they minimize what are likely to be reliable within subject differences. This factor thereby reduces statistical power (Cronbach, 1990) and we have shown percent correct to suffer from the same measurement defects a difference scores, slopes, and ratios (Dunlap, Kennedy, Fowlkes, & Harbeson, 1989; Seales, Kennedy, & Bittner, 1978). Therefore, we avoid this use unless no good alternative is available. Reaction Time is scored as an average latency of all trials; Tapping is the number of alternations and the counting tasks used percent correct.

## System Development

General. Considerable effort has been expended to create an automated system. To this end, a number of subjects were run in preliminary attempts at identifying which eye movements would be measured. From previous research (Kennedy et al., 1989; May et al., 1990) we knew that, in the dark, eye movement extent would covary with task load but, to determine the feasibility of using real tasks with different mental loads and visual task demands, we sought a metric which, on the one hand, could be independent of visual demands and, on the other hand, independent of different mental load or mental content. With such a metric we envisioned that it could be employed during evaluation of systems which require these disparate characteristics and demands. Therefore, in addition to preliminary work toward development of an algorithm for assessing the visual activity indirectly from task elements (field of view, response per minute), we also set out to create an analytic system that would be addressable to measure all possible characteristics of the ocular activity (blinks and eye movements, accelerations, frequencies, etc.) and to perform the analysis automatically.

The primary goal was to develop eye movement based algorithms to predict work load. Data from the experimental effort would be used to assess the relationship between eye movement data and task loading (i.e. response per minute). These data would allow us to create the work load algorithms. Because real-time analysis of data was not required for the experiment, we used the Essex mid-speed microprocessor (12Mhz i286) for these experimental sessions and the same automatic interactive eye movement scoring techniques developed in Phase I and in our previous work for the U.S. Air Force Office of Scientific Research. The data were transferred to 1.44Mb floppy and analyzed on the Essex Northgate Elegance 20 Mhz i386 with 165 Mb 14 ms hard disk, 4 Mb memory, 80387-20 math co-processor, and 800 x 600 VGA graphics with Princeton Graphics Ultra 14 analog monitor.

### Eye Movement Recording

Eye movements (i.e., frequency, amplitude, velocity, acceleration, fixation duration, frequency and duration of eye blinks) were recorded using electro-oculography (EOG). Six 4mm silver/silver chloride electrodes were used for all EOG recordings. Electrode leads were fed to amplifiers featuring characteristics suitable for EOG recording. A MetroByte Corporation DAS-16F 8 channel (bipolar) analog-to-digital converter capable of sampling up to 100,000 samples per second served as the interface to the microprocessor. Three channels were sampled, two for vertical eye movements and one for horizontal eye movements. Each of the three channels were sampled at 256 samples per second.

The calibration board contained 9 red LEDs imbedded in a 4-foot square panel which was painted black. The LEDs were controlled via software using the 8 digital I/O channels provided by the DAS-16F. With the subject seated 5 feet from the center of the board, and at eye level, 40 degrees of vertical and 40 degrees of horizontal distance separate the top/bottom, left/right LEDs, with 10 degrees of separation between each LED. The software calibration routine successively illuminated the calibration LEDs in the horizontal plane in the following order: -20 degrees, 0 degrees, +20 degrees, 0 degrees, -10 degrees, 0 degrees, +10 degrees, 0 degrees, -20 degrees, +20 degrees, -20 degrees, and 0 degrees.

### Eye Movement Scoring

Figure 1 shows schematically an eye movement and the various elements which may be scored. The software program was constructed so that a saccade was defined as a rapid change in direction which persists for more than 31.25 msec, and involved high opposite direction of movement in each channel from the baseline. This was used as a criterion to be above the noise in our system on the one hand, and below the point where more than one eye movement would be involved. Then amplitudes were saved as well as average velocity at the midpoint of a saccade (i.e., peak velocity); the amplitude of each saccade was saved; the duration; and frequency. This analysis provided for automated recording each second of 25 variables. Figure 2 provides a schematic record of an eye blink. The criterion for an eye blink was: displacement in both eyes involving depolarization in one eye with hyperpolarization in the other eye. The amplitude, duration and acceleration were resolved as shown. These data were also assessed each second and cumulated over the period of the exposure. These variables appear in Table 2. In all cases the data were averaged over the particular experimental session and, where appropriate, normalized so that they may be interpreted as response per minute.

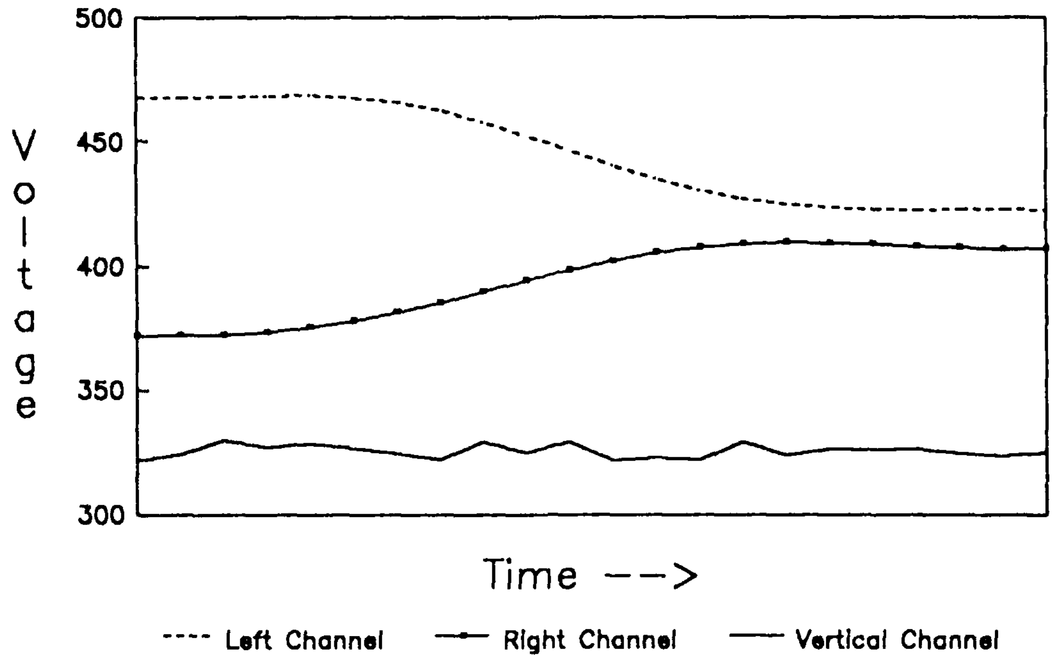

Figure 1a. Eye movement detection.

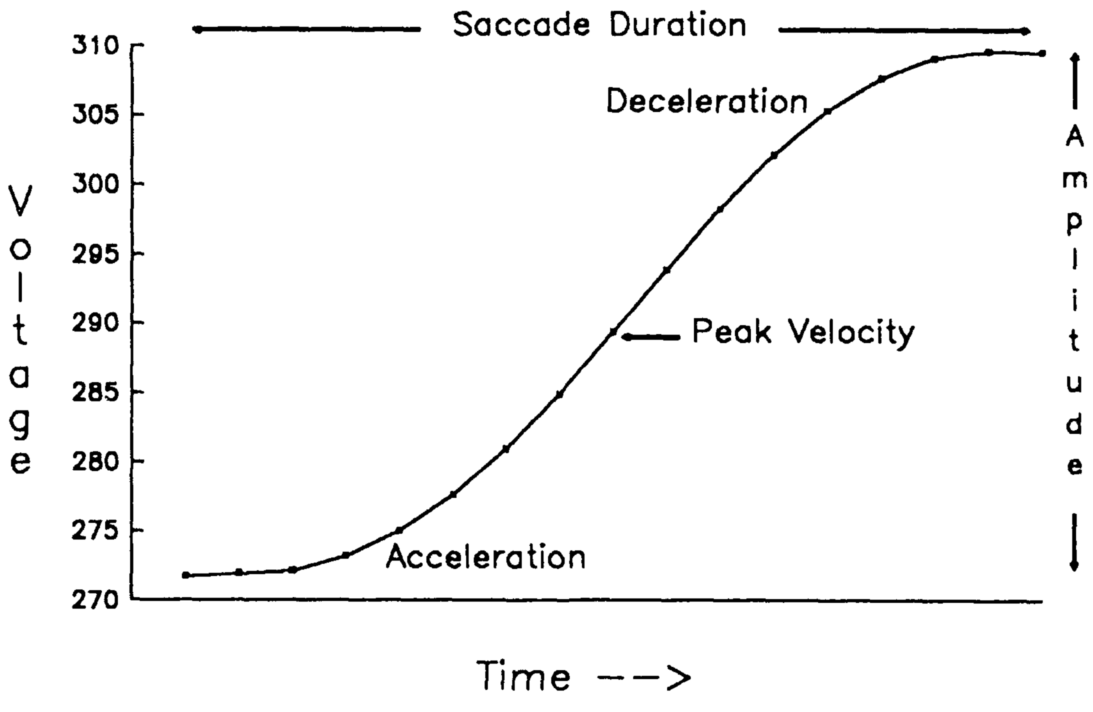

Figure 1b. Eye movement measurements.

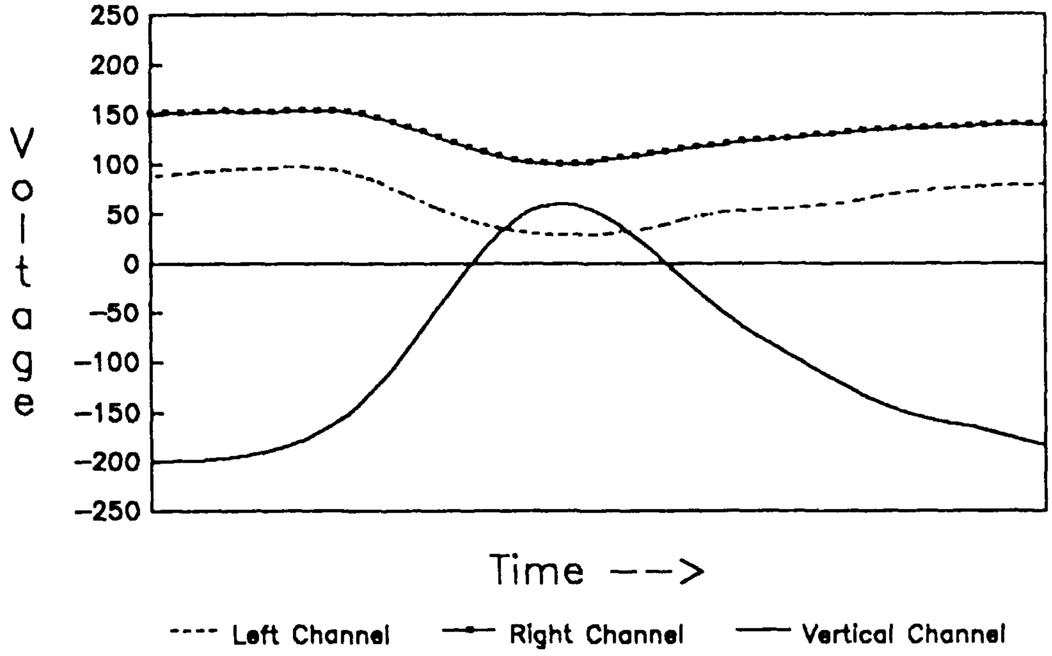

Figure 2a. Blink detection.

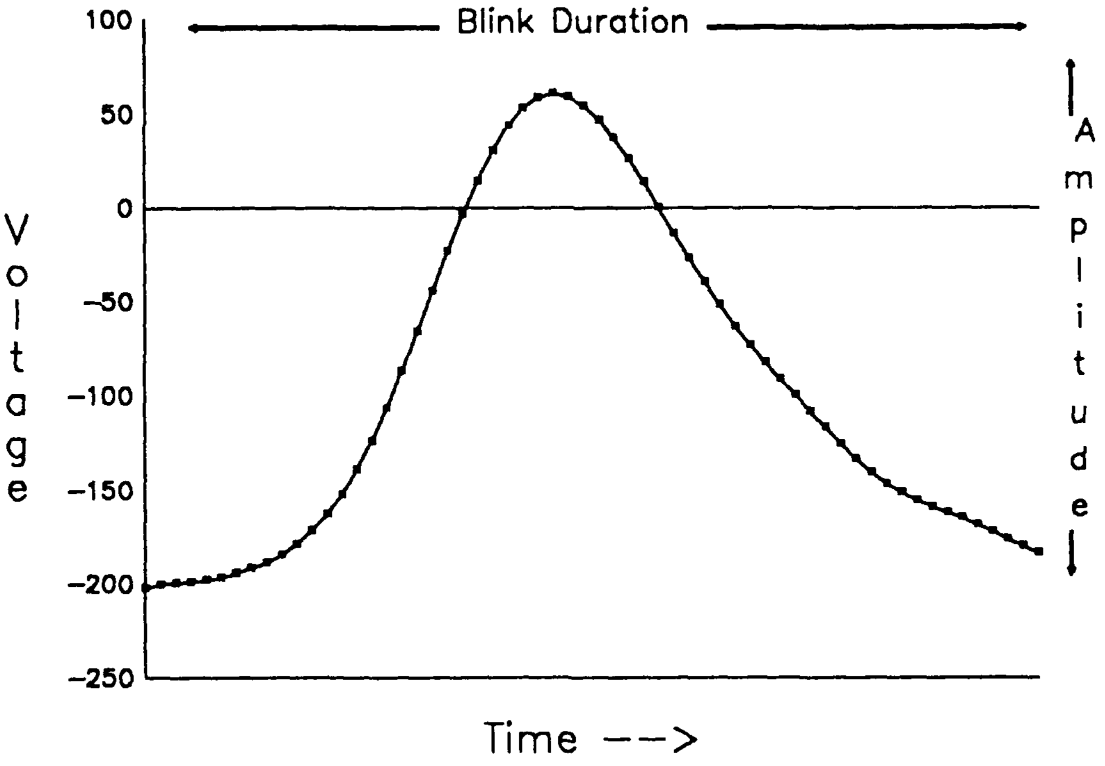

Figure 2b. Blink measurements.

TABLE 2. Eye Movement and Blink Dependent Variables
Automatically Recorded and Saved for Subsequent Analysis

| | | |
|---|---|---|
| 1) | BLINKS | Blink Frequency |
| 2) | BLKDMIN | Blink Duration (Minimum) |
| 3) | BLKDMAX | Blink Duration (Maximum) |
| 4) | BLKDMEAN | Blink Duration (Mean) |
| 5) | BLNKAMIN | Blink Amplitude (Minimum) |
| 6) | BLNKAMAX | Blink Amplitude (Maximum) |
| 7) | BLNKAMN | Blink Amplitude Mean |
| 8) | EYEMOVE | Eye Movement Frequency |
| 9) | SACDMIN | Saccade Duration (Minimum) |
| 10) | SACDMAX | Saccade Duration (Maximum) |
| 11) | SACDMEAN | Saccade Duration Mean |
| 12) | LAMPMIN | Left Amplitude (Minimum) |
| 13) | LAMPMAX | Left Amplitude (Maximum) |
| 14) | LAMPMEAN | Left Amplitude Mean |
| 15) | LVELMEAN | Left Velocity Mean |
| 16) | LACCMEAN | Left Acceleration Mean |
| 17) | LDECMEAM | Left Deceleration Mean |
| 18) | RAMPMIN | Right Amplitude (Minimum) |
| 19) | RAMPMAX | Right Amplitude (Maximum) |
| 20) | RAMPMEAN | Right Amplitude Mean |
| 21) | RVELMEAN | Right Velocity Mean |
| 22) | RACCMEAN | Right Acceleration Mean |
| 23) | RDECMEAN | Right Decleration Mean |
| 24) | LSLIT | Left Eye Amplitude Range |
| 25) | RSLIT | Right Eye Amplitude Range |

## Automated Task Load Analysis System

Figure 3 shows the result of an actual eye movement recording. This panel depicts one second, the epoch employed as the unit of analysis in this study. However, these data can easily be aggregated over epochs of any length (minutes, hours, etc.). In the final version of the computer program, epoch length will be fully addressable. This is the first step in our analysis from an average subject.

It may be seen in this record that the left eye and right eye are recorded separately and this record can be used to differentiate blinks from large excursion eye movements as well as to measure wave forms, duration, and other characteristics of the eye movement. The recording system is updated 256 times per second, and the bioelectric electrodes employed permit a 1-degree resolution of eye movement; eye movements in excess of 400 degrees per second can be resolved.

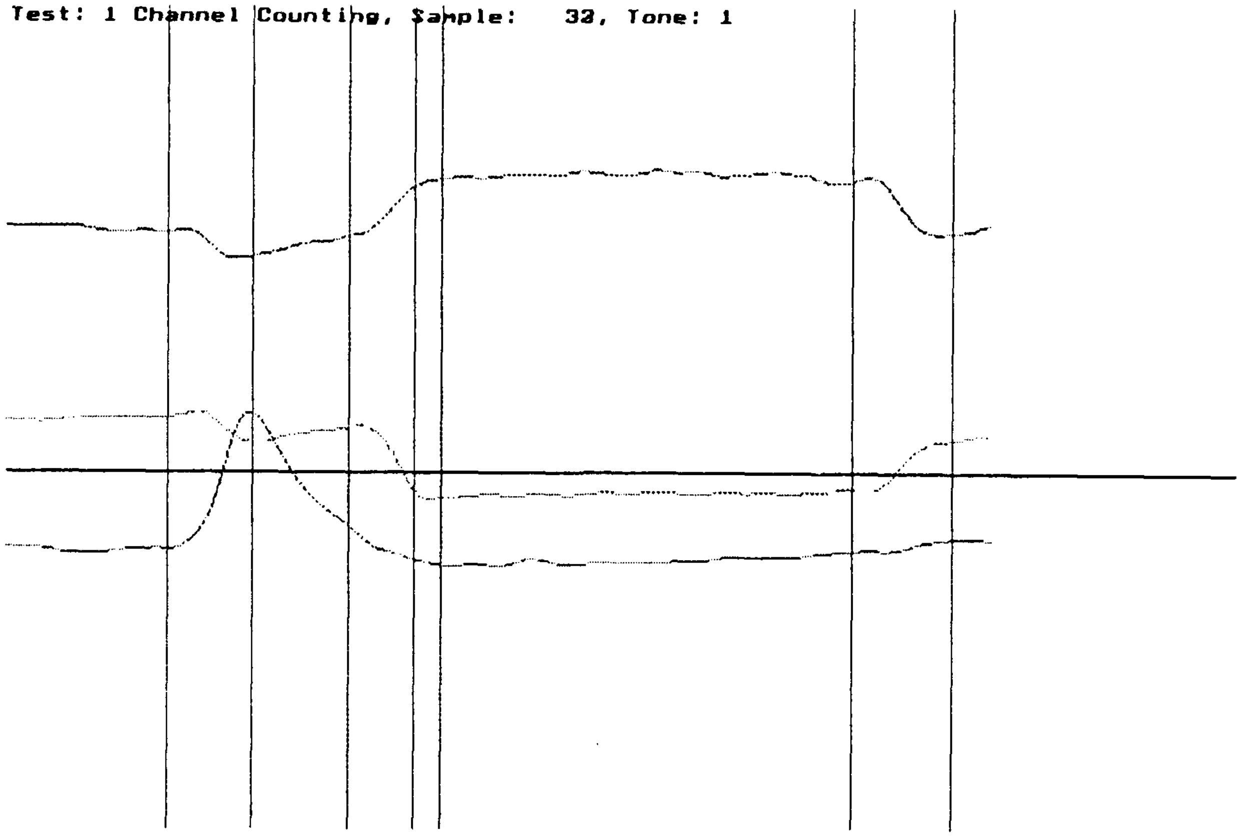

Figure 3. Results of an actual eye movement recording.

## DEVELOPMENT OF TASK LOAD METRICS

It was the thesis of this investigation that the independent variables would be the visual and mental demand characteristics. These characteristics could be considered differences in content and (more likely) differences in loading. We sought to measure these elements by performance (hits/minute) and by bioelectric actions of the visual system (eye movements and blink). It was our plan to relate these using between subject (i.e., subject independent) metrics and within subject (subject dependent) metrics. As mentioned above, it is our view that the work load literature is not always clear with this distinction. Attempts to correlate these in a single experiment have found less than perfect relationships (Gopher & Braune, 1984; Wierwille, Rahimi, & Casali, 1985). We believe the logical inconsistency ("between" being correlated with "within") may be partly the reason for low correlations. However there may be ways that such difficulties may be circumvented.

We therefore followed a series of steps. We developed a series of psychophysiologically-based metrics (e.g., 25 eye movement and blink parameters), measured performance (hits per minute) and then developed subject-dependent (estimated difficulty) and subject-independent (visual activity required for performance) metrics.

# RESULTS

## INDEPENDENT VARIABLES

Three related metrics were devised in order to quantify the computerized cognitive tasks as independent variables of work load.

1. At the conclusion of the APTS battery performance, the subjects were instructed to rank the various tasks for their difficulty level on a 10-point scale. The consistency of these ratings across subjects was high, yielding a Cronbach's alpha of 0.90. The group average of this difficulty ranking for each task is given in the first column of Table 3. Note that whereas each individual subject's estimate may be a "subject-dependent" metric of work load, the average value may be considered subject independent. We believe this easily obtained metric would compare well with the NASA-TLX.

2. In previous studies (Kennedy, Baltzley, Wilkes, & Kuntz, 1989; Kennedy, Baltzley, & Osteen, 1988), after comparable periods of practice, each of the APTS tests was shown to become stable according to strict psychometric criteria. Using data from several experiments as a basis, the average number of correct responses per task was normalized for session length in order to obtain a response per minute (RPM) index of each task. For example, in this analysis, average Tapping scores were 111 times per minute and Grammatical Reasoning, a more mentally complex task, achieved scores of 18 responses per minute. These values also appear in Table 3.

3. Lastly, the visual demands of the different tasks were estimated by using the amount of ocular motility necessary to "see" and read the various characters of the APTS test. Since the screen was 11 inches wide and was viewed from 22 inches, the horizontal retinal angle of the screen text was 28.8 degrees, but each APTS task had different dimensions and visual task requirements. For example, with Grammatical Reasoning, the length of the line of text was 8.25 inches (20 degrees of retinal angle). From experience and observation, we know that each Grammatical Reasoning problem is generally re-read once. Therefore, based on an average frequency of response of 18 per minute and two scans per problem and 20 degrees of retinal angle, the estimated visual demand characteristics of this task can be calculated.

The results of this analysis appear in summary form in Table 3. It may be seen that Grammatical Reasoning and Code Substitution had the highest estimated difficulty rating and Reaction Time and Tapping the lowest. Response rate revealed a similar relation, but these values did not appear linear and so rankings were employed (column 3) in subsequent analyses. Visual demands (ocular motility) calculations (column 4) showed Manikin to be the most visually demanding and Tapping the least, but these numbers too were nonlinear. Therefore, rankings for these values were used in column 5. Columns 1, 3, and 5 were employed in subsequent analyses.

TABLE 3. Results of Workload Metrics: Group Difficulty Rating, Average Response Rate Actual, Response Rate Ranking, Ocular Motility Demand Characteristics, Ocular Motility Ranking

| Subtest | Subjective Estimate of Difficulty | Response Rate Actual | Response Rate Ranking | Ocular Motility Demand Characteristics | Ocular Motility Ranking |
|---|---|---|---|---|---|
| Reaction Time | 2.0 | * | * | * | * |
| Tapping Task | 3.2 | 111.0 | 7 | 78 | 1 |
| Pattern Comp. | 3.7 | 45.1 | 6 | 457 | 4 |
| Manikin Test | 4.8 | 44.6 | 5 | 2233 | 7 |
| Math Processing | 6.2 | 30.2 | 4 | 119 | 3 |
| Short-Term Memory | 6.3 | 33.0 | 3 | 108 | 2 |
| Code Substit. | 7.3 | 29.4 | 2 | 688 | 5 |
| Grammatical Reasoning | 8.8 | 17.8 | 1 | 769 | 6 |

Note: Subtests that are underlined are the four judged to have the least visual requirements in terms of degrees of visual angle.
*NA

## DEPENDENT VARIABLES

### Counting Test

Performances on the Counting Test for sessions 1 and 2 are combined and are shown in Figure 4. Performance scores on dual counting (the harder task) were poorer than on the single counting task. It should be remembered that an even higher task load condition is sometimes employed with the counting procedure where subjects are required to keep track of every fourth instance of three stimuli. Because the present procedure only employed single and dual stimulus counts, we can consider these to represent low and moderate work load conditions; therefore, the condition labeled high in the figures is only relatively high. Given this consideration, the consistency of the findings relating performance to bioelectric events which follow are even more noteworthy.

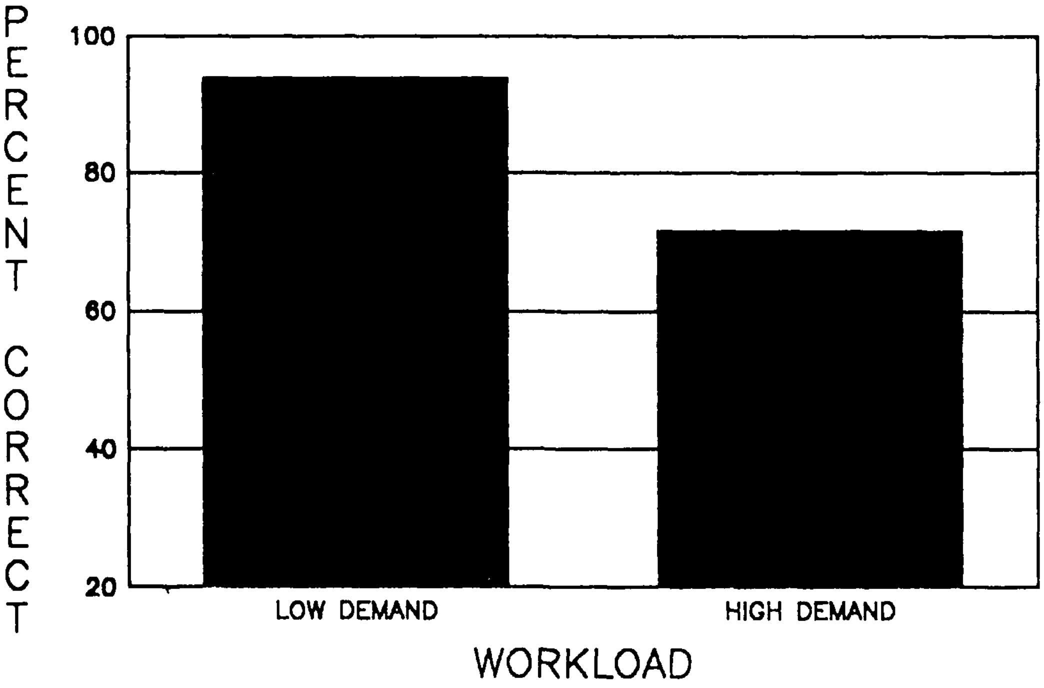

Figure 4. Workload results for counting task.

APTS

Scores for the nine tests for the three sessions (two practice sessions with five trials for each APTS test, and on the third session single trial for each APTS test during electrophysiological readings) may be found in Figures 5-13. It may be seen that each test improves markedly over the 10 practice sessions and by session 3 (trial 11) most of the learning has been accomplished. This implies that performances were stable after session 2, which is concordant with data from other experiments (e.g., Kennedy, Baltzley, Dunlap & Kuntz, 1989).

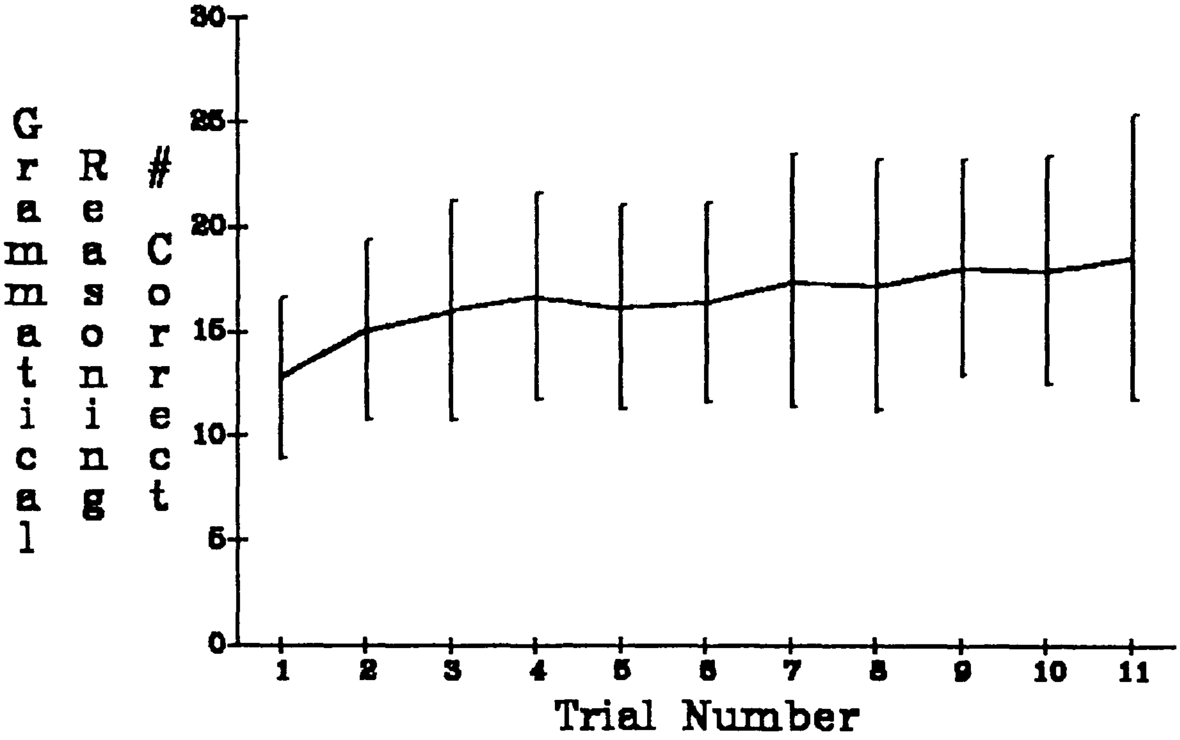

Figure 5. Number correct for Grammatical Reasoning by trial.

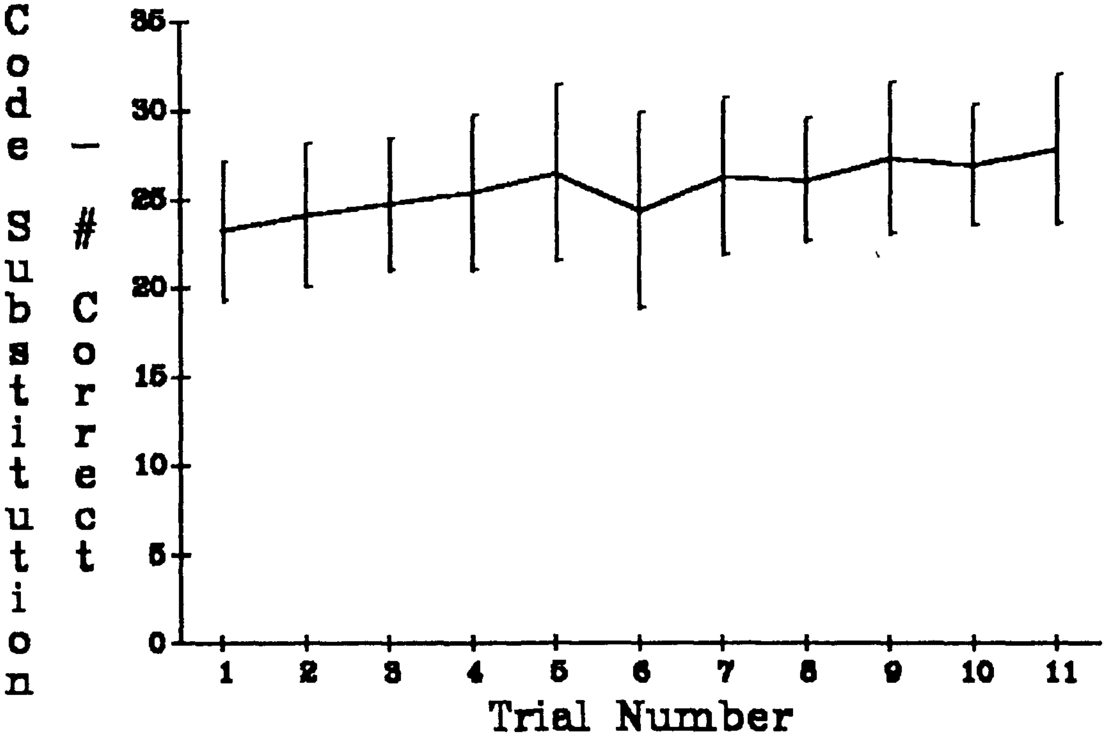

Figure 6. Number correct for Code Substitution by trial.

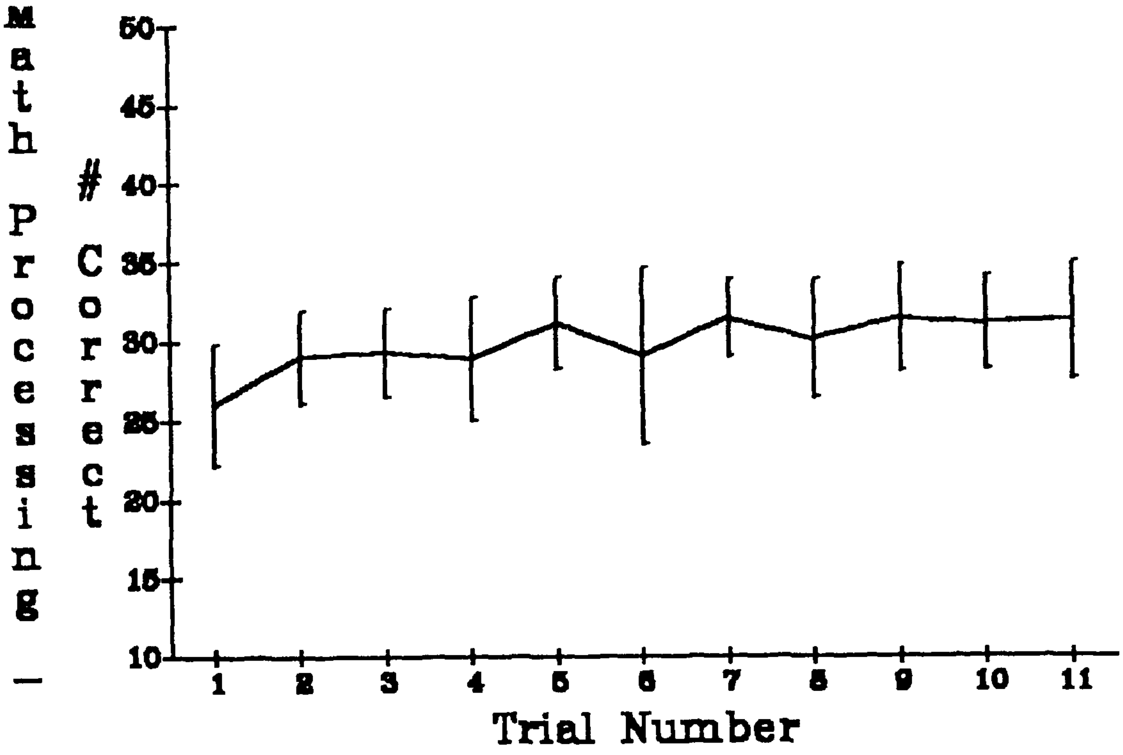

Figure 7. Number correct for Math Processing by trial.

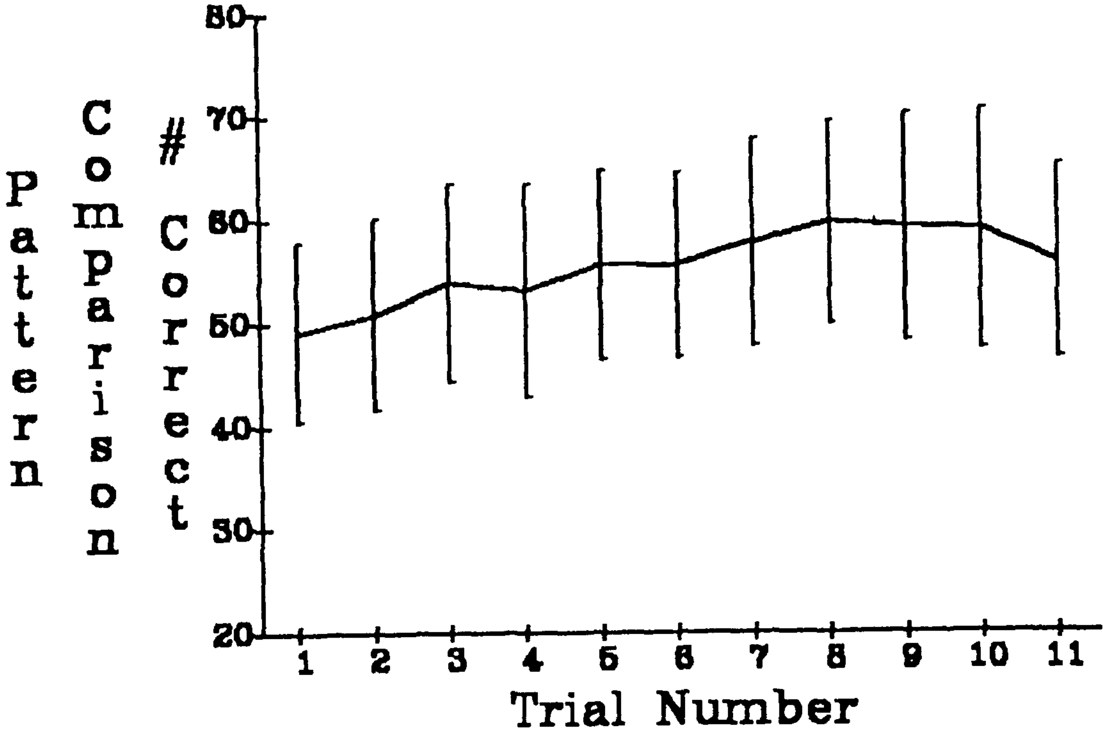

Figure 8. Number correct for Pattern Comparison by trial.

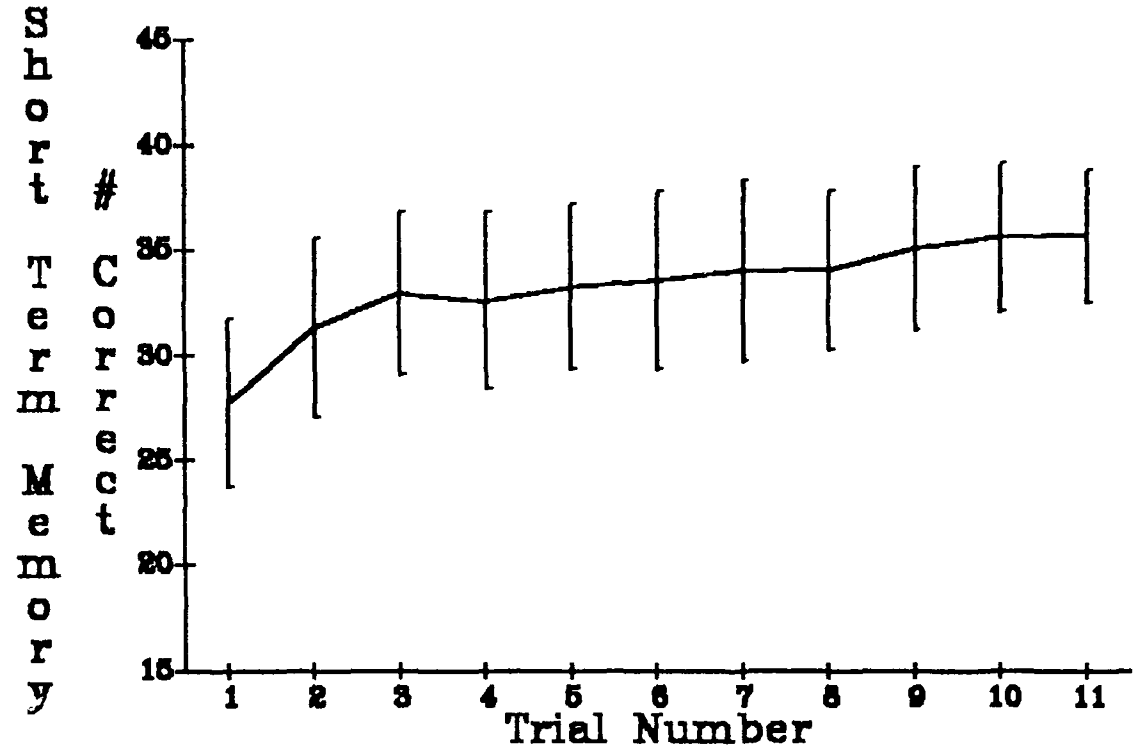

Figure 9. Number correct for Short-Term Memory by trial.

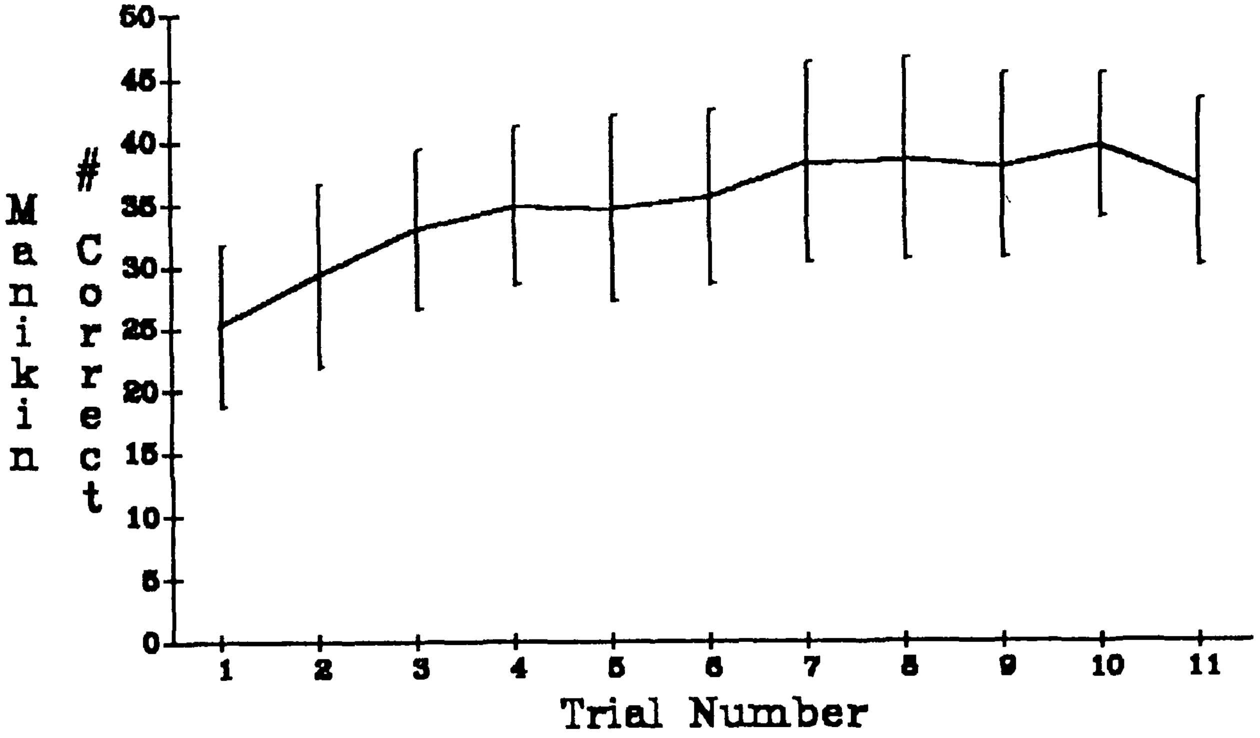

Figure 10. Number correct for Manikin by trial.

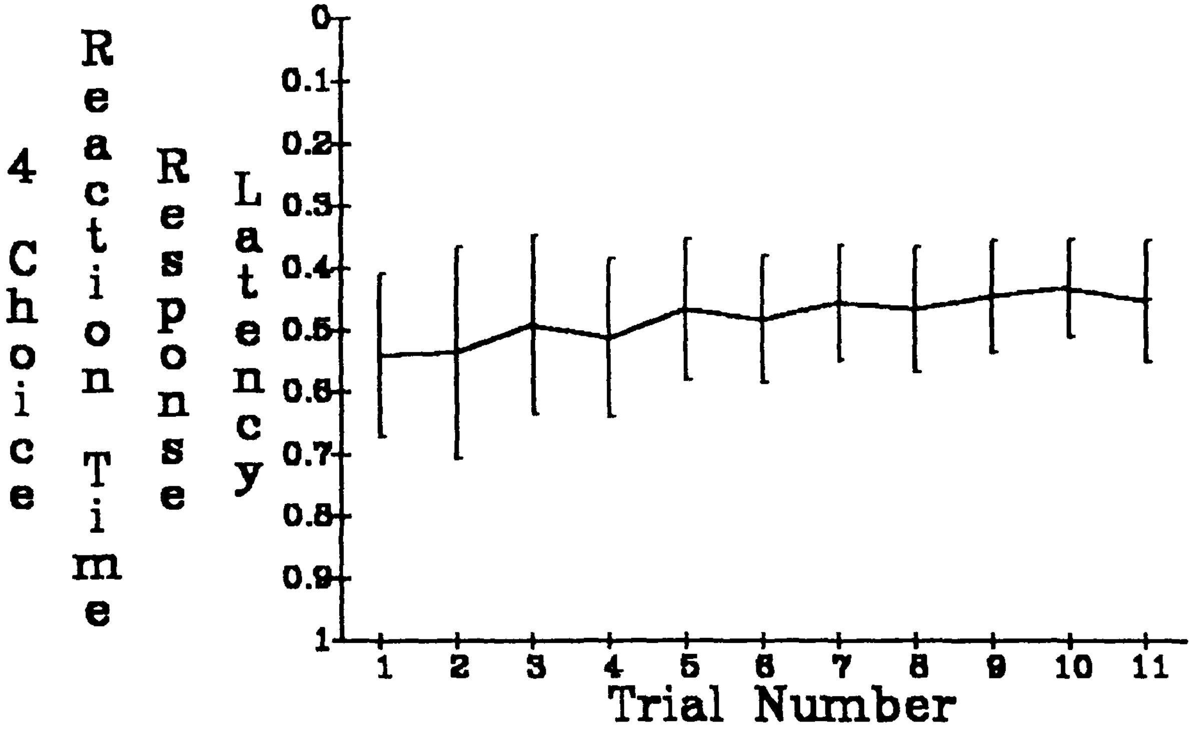

Figure 11. Average response latency for 4-choice Rection Time.

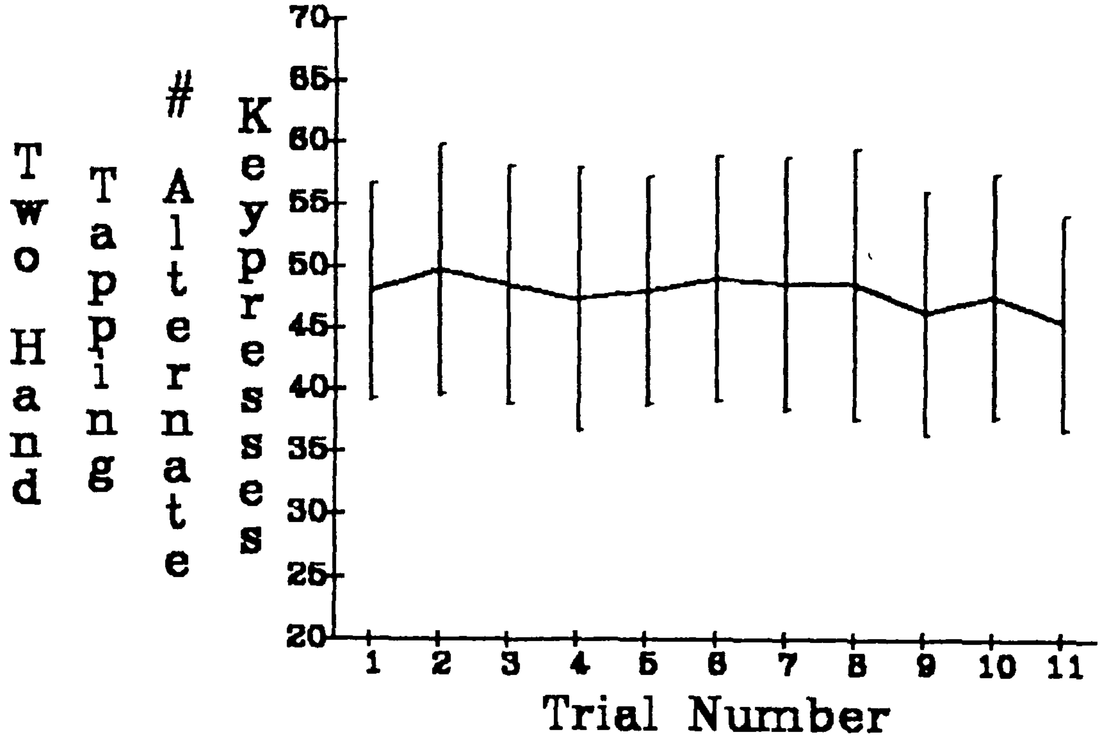

Figure 12. Number of alternate keypress Two-Hand Tapping.

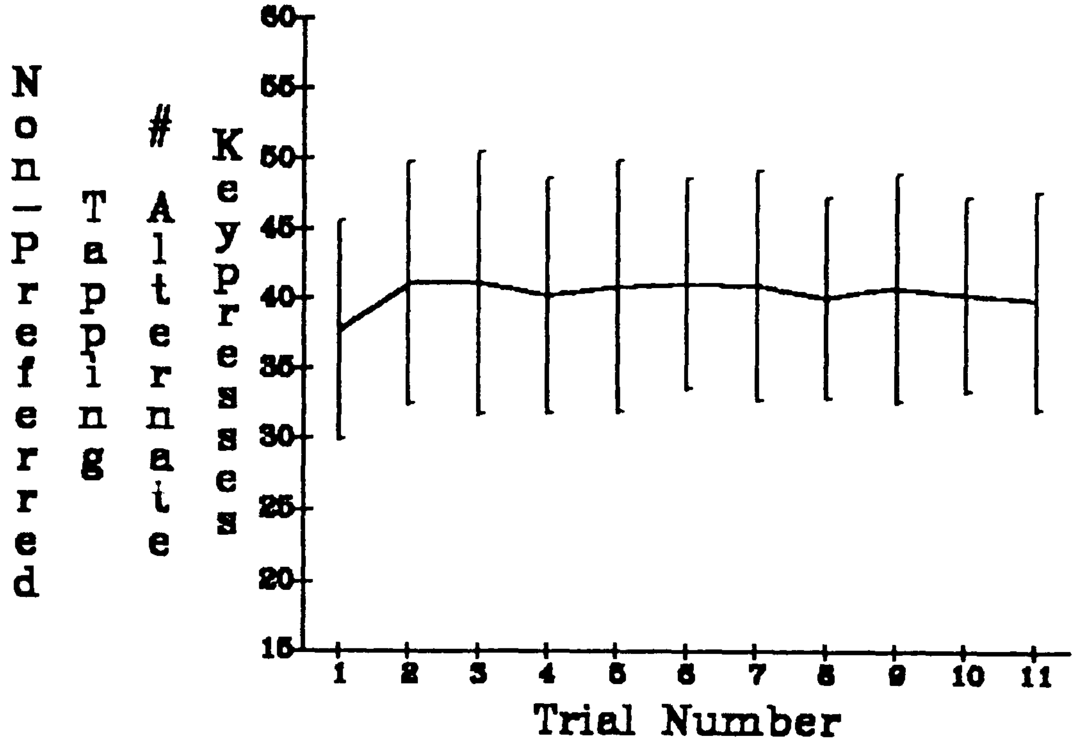

Figure 13. Number of alternate keypress Nonpreferred Tapping.

## Electrophysiological Measures

### Counting Test

Effects of Task Load on Eye Measures. Results of task load (counting test) on eye movement indices are depicted in Figure 14-18. These data were analyzed by the t-test for correlated measures, and the results of these tests are presented at the top of each figure. Both the blink duration, number of blinks and the number of blinks per minute were significant. Blink durations were longer and there were fewer at the higher work load condition. Individual subject data revealed that 100% of the subjects changed in the direction indicated by the group means in Figures 14-16 for each of these three measures as task load increased. Neither eye movement range nor numbers of eye movements were significant, and the subjects' responses were highly variable and inconsistent over these measures, implying different monitoring strategies were used. The task could be performed visually or auditorially, which may have invited these different strategies.

Eye Movement Correlations Across Task Loads. Inspection of the Counting Test data revealed fairly strong individual differences that were maintained across low and high load conditions on the Counting test. These trends resulted in strong correlations between the low and high task load data of 0.93, and 0.97 for number of blinks and blink duration. These correlations are all significant and strongly suggest that these indices will work most effectively as within-subject as opposed to between-subject measures. This means that establishing an accurate and stable baseline for each subject under a low task load condition is quite important for the sensitive use of these measures as within-subject indicators of increased task load.

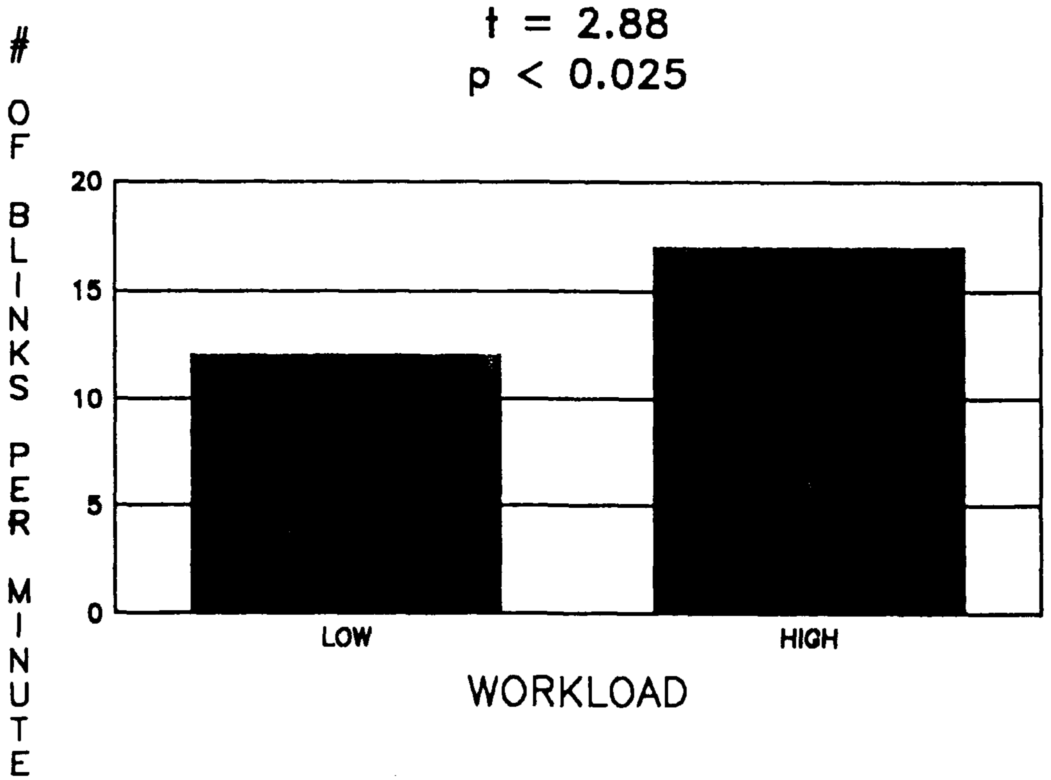

Figure 14. Number of blinks per minute by work load.

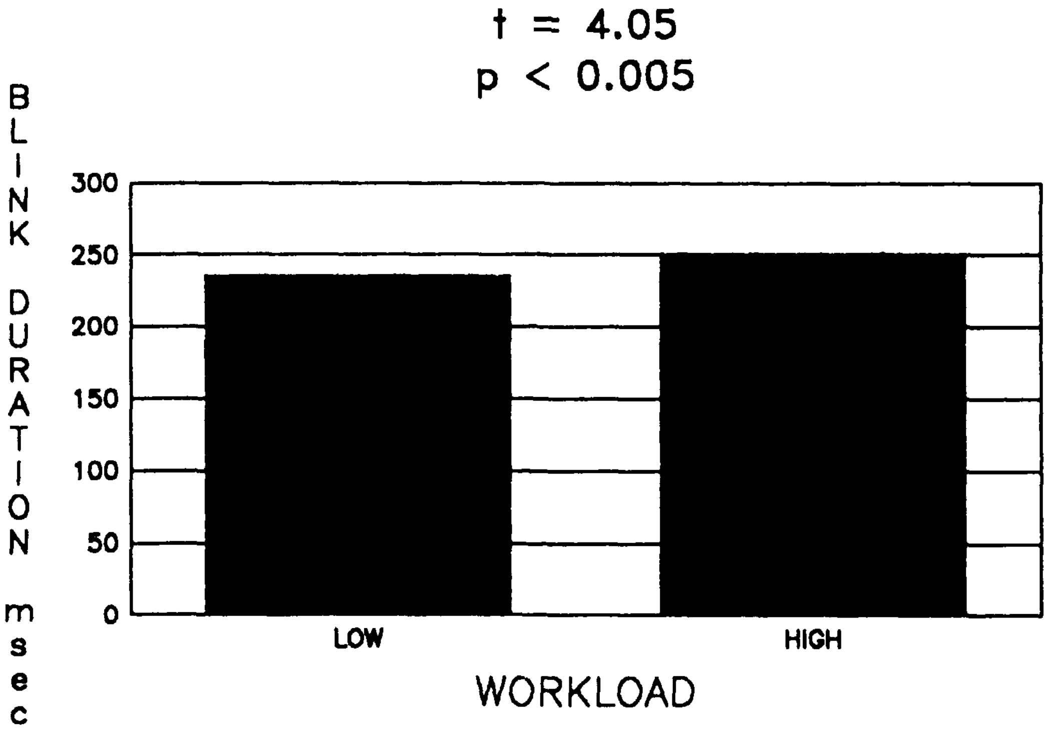

Figure 15. Blink duration by work load during counting test.

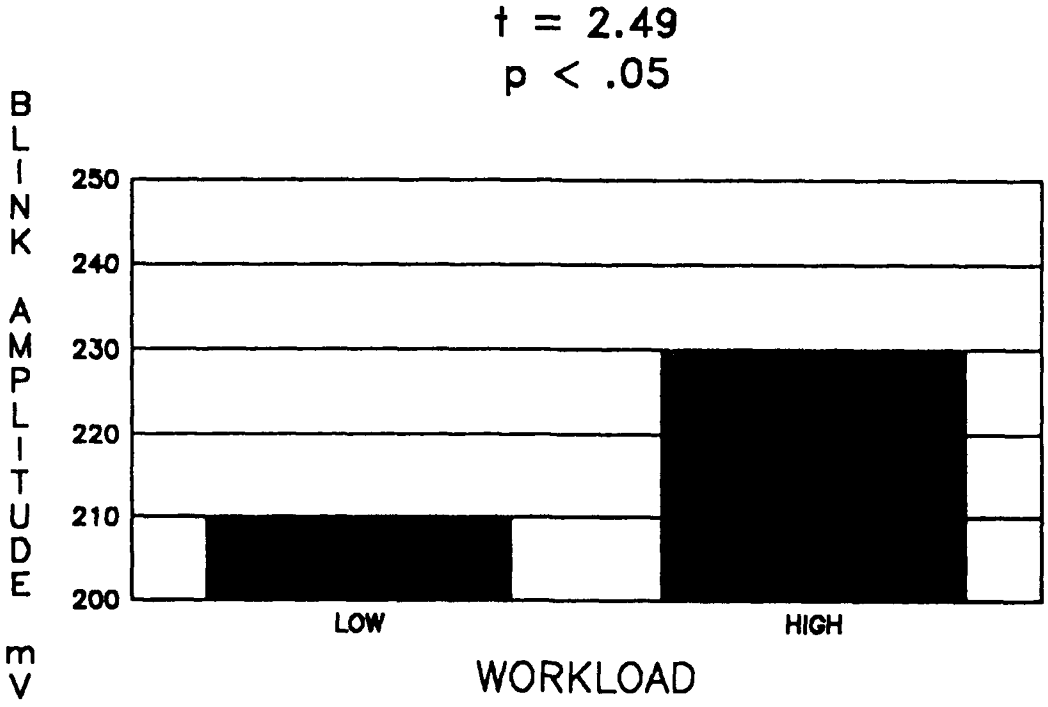

Figure 16. Blink amplitude by work load during counting test.

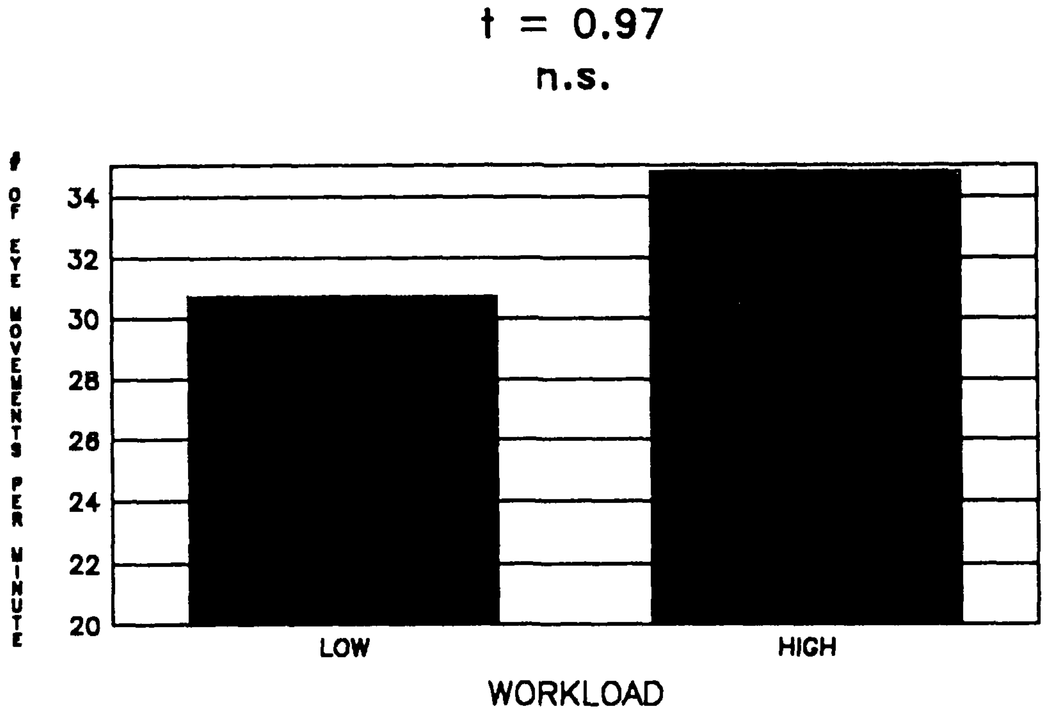

Figure 17. Number of eye movements per minute by work load.

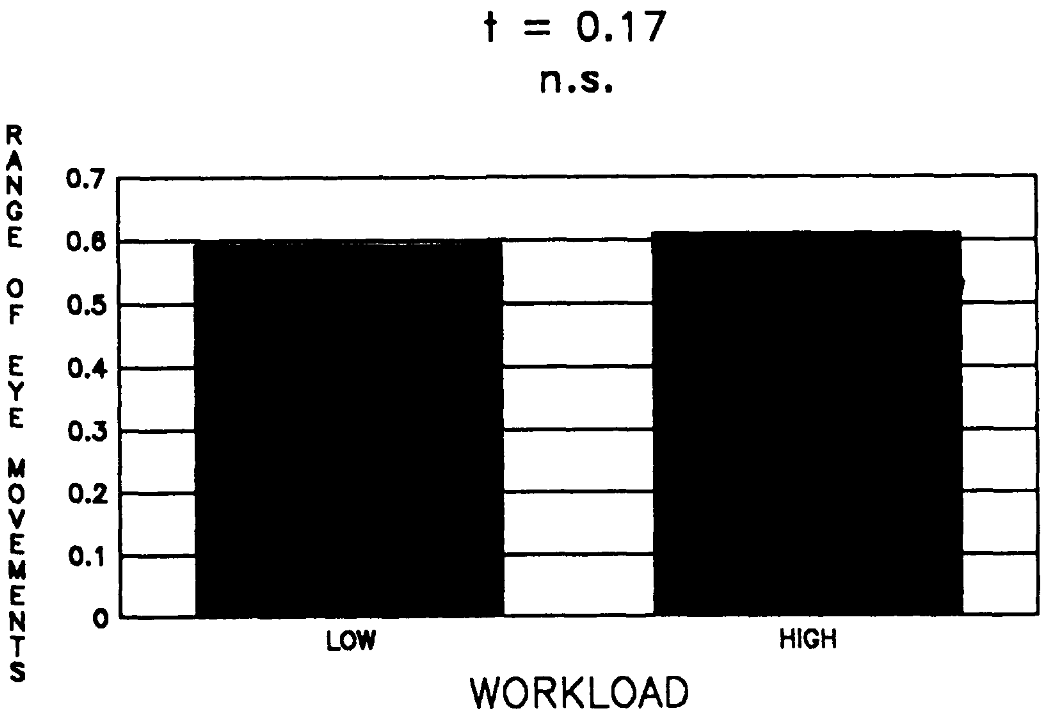

Figure 18. Range of eye movements by work load.

APTS Tests

Five variables emerged as prospectively useful because they appeared nonredundant. These included: (1) number of blinks, (2) blink duration, (3) number of saccades, (4) average velocity, and (5) overall visual extent. After 10 subjects had completed their APTS testing, the 25 dependent measures from electrophysiological readings were cast into an inter measure correlation matrix. Several measures appeared to overlap considerably and so the data set was reduced to eight dependent measures based on low redundancy (velocity) and/or the measure was theoretically meaningful (e.g., number of blinks and blink duration).

Eye Movement Indices and APTS Tests. Table 4 lists the three independent variables of task loading variables (difficulty, response per minute and visual demand) as well as mean scores for the five eye movement metrics.

TABLE 4. Difficulty Ratings, Response Rates, Blink Data, and Eye Movement Data for the Various Subtests of the APTS Battery ( N=15 )

| | Independent Variables | | | Dependent Variables | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| SUBTEST | Difficulty Rating Subjective | Response Rate Ranking | Ocular Motility Ranking | Number of Blinks | Blink Duration (msec) | Blink Amplitude (volts) | Number of Saccades | Saccade Velocity | SLIT |
| Reaction Time | 2.0 | -- | -- | 11 | 221 | .25 | 70 | 2.53 | .70 |
| Tapping Tests | 3.2 | 7 | 1 | 18 | 182 | .21 | 47 | 5.24 | .61 |
| Pattern Comparison | 3.7 | 6 | 4 | 7 | 212 | .22 | 136 | 2.02 | .31 |
| Manikin | 6.2 | 5 | 7 | 20 | 196 | .19 | 107 | 2.18 | .70 |
| Mathematical Processing | 6.2 | 4 | 2 | 11 | 247 | .26 | 9 | 6.35 | .30 |
| Short-Term Memory | 6.3 | 3 | 3 | 18 | 227 | .25 | 25 | 6.65 | .35 |
| Code Substitution | 7.3 | 2 | 5 | 24 | 204 | .25 | 142 | 2.88 | .83 |
| Grammatical Reasoning | 8.8 | 1 | 6 | 16 | 222 | .25 | 120 | 2.43 | .66 |

Blink frequency ranged from 5-25/minute, a figure which compares favorably with reports in the literature (Stern, 1990). The other dependent variables are also consistent with electrophysiological studies of eye movement.

Table 5 shows a correlation matrix between the three task load (i.e., independent) variables and several eye movement and blink dependent variables for one subject. It is obvious that two metrics (difficulty and RPM) are virtually identical, but the ocular motility metric adds new information. It should be noted that the remaining correlations are for one subject and so the comparison of the variables is illustrative. Since these are totally within-subject comparisons, conventional, statistical examination is not warranted for evaluation of relationships between task load and dependent measure. Therefore, we produced correlation matrices similar to the above for all 12 of the subjects who completed the five experiment sessions and the "sign" of the relationships was tallied. Several relationships were statistically significant based on positive (or negative) correlations in all 12 of the subjects ($\underline{p} < .001$). We have indicated with asterisks those which are significant by the sign test of Walker and Lev, 1953 (p. 458).

Most notably, this included the correlation between blink duration and either task difficulty or response per minute matrix. This relation again is in a direction opposite to what others (Stern, 1990) find (viz. longer duration for the more difficult tasks). Next, number of eye movements appeared positively related (in all 12 subjects) to the objectively derived visual task demand matrix. Several other suggestive relationships (viz correlation between velocity and all three task load variables) were present ($P<.07$) but were weaker. It may also be seen in Table 4 that most of the five dependent measures in this subject are essentially independent of each other. Similar relations appeared for the other subjects.

It is also noteworthy that, if attention is focused on those subtests where visual scan requirements were less than or equal to approximately 5 degrees of visual angle (the four subtests underlined in Table 4), the correlations between response rate and difficulty rating become -0.38, 0.91, and -0.70, for blink number, blink duration, and eye movement range, respectively. The latter two correlations are congruent in size and direction to the earlier findings with varied task load within the counting task. It should further be pointed out that all of the performance tests used were self-paced, therefore, the individual subject could easily vary task load for even the subtest rated most difficult, by responding at a slower rate. This, of course, is a general problem for interpreting difficulty ratings as monotonic with task load.

## TABLE 5. Intercorrelations Between Task Load Variables and Eye Movement Variables

| | RPM RANK | MOT RANK | BLINKS | BLKDMEAN | BLKAMN | SACCADES | VEL | SLIT |
|---|---|---|---|---|---|---|---|---|
| DIFFRANK | -.98 | .50 | .07** | .26* | .15** | .27*** | -.05 | .09 |
| RPMRANK | | -.56 | -.06** | -.27* | -.13** | .32*** | .08 | -.09 |
| MOTRANK | | | -.04 | .09 | .02 | .61**** | -.30* | .11 |
| BLINKS | | | | -.23 | -.16 | .02 | .05* | .11* |
| BLKDMEAN | | | | | .58** | -.03 | .08 | -.14 |
| BLKAMN | | | | | | .06 | .01 | -.15 |
| SACCADES | | | | | | | -.31 | .25* |
| VEL | | | | | | | | .04 |

Signif: **** p < .001; *** p < .01; ** p < .02; * p < .07

## DISCUSSION

A human operator in an airborne weapons system is bombarded with multiple visual and other sensory inputs. These inputs frequently exceed, or nearly exceed, the capacities of the operator. Since so much of the information is typically brought in visually, a need exists for an objective method to index display characteristics and configurations to assist in the development of an expert system in order to evaluate alternative display features (i.e., adaptive function allocation). It is logical to expect that measurement of eye position and frequency could serve as an appropriate technique. However, considerable methodological and technical difficulties attend the successful prosecution of any development programs in this regard, but there are two promising factors in any such study. First, there have been a succession of studies with positive findings in the past half dozen years (Kennedy et al., 1989; 1989; May et al., 1990; Stern, 1990; Stern et al., 1984) where characteristics of eye movement and blink were related to aspects of task loading. Second, the heavy computation demands that ordinarily attend eye movement recording and analysis are becoming increasingly less expensive. This means that it is now possible to conduct with desktop systems what only major laboratories were able to accomplsih several years ago. Therefore, to anticipate future capabilities of portable computer systems, we believe the timing is appropriate to undertake a combined software mechanization and implementation system at the same time as a behavioral electrophysiological program reveals what aspects of eye movement should be analyzed.

In this Phase I experiment workload was characterized in two ways, both derived from the tasks under study but not from the use of data derived from the subjects of this study. The first characterization (or metric) was the average number of correct responses per minute by a large (N > 50) sample, and the second was the estimated visual demand based on the ocular motility necessary to "see" the problem presented. Both measures were subject-independent. We found that: (1) Different dependence on visual system measures (e.g., frequency of eye movements, blink duration) on tasks with differing visual and mental requirements were differentially predictive of the two objective scales. (2) Customized computer software for automated presentation of the tasks and scoring of the electrophysiological responses were developed for desk-top personal computers. This software system was mechanized and implemented and is now fully up and running.

It is our opinion that because within-subject changes correlated at a statistically meaningful level with the visual task demands and with the mental work load, this procedure holds promise as a method for calibrating individuals against known visual and mental task loading so that laboratory-based systems like NADC's reconfigurable cockpit can be used to study adaptive function allocation. Future studies should develop the system to: (1) run on line and in real time and be validated in an aircraft or simulator system to determine quality assurance boundaries; (2) be made compatible with standard data analytic packages (BMD, SAS, SPSS); (3) be made fully portable for field usage; (4) create algorithms which will permit partition between mental task loading versus visual task demands in cockpit workplace design and development; (5) create field manuals for use by systems developers to objectively assess visual work load parameters of various aspects of aviation activity; (6) be field tested at a Navy development laboratory as part of their work load R&D programs.

The Phase I effort was the first step in the design of an automated task load analysis system for biocybernetic modification and function allocation of aircraft cockpit display systems for rapid and portable on-site measurement of aviation cockpits and workspaces. The availability of such a package could provide aircraft manufacturers and others with common metrics for conducting human factors engineering design, test and evaluation of workstations of all kinds. Nonintrusive visually based measures of an operator's interest or attention could have far-reaching commercial applications.

It should be mentioned that the two measures selected for this work were not chosen because we considered them the best available measures of workload but, rather, because they were readily available and familiar. The latter point was decisive. We had had extensive experience with both the APTS battery and with visual demand characterstics and felt comfortable on that basis in using the two measures as rough-and-ready indices of workload. If a longer period of performance (i.e., > six months) had been available to carry out experiments we would have choosen as measures of workload what seems to us to be the best available rather than the most feasible measure for use with limited resources. Therefore we would propose that if this effort is moved into a second phase, a new "gold standard" for workload should be introducted into this electrophysiological approach.

That standard will be the NASA-TLX scale (Hart, 1990; Hart & Staveland, 1988). We agree with the developers of this scale that subjective assessment is still the most valid indicator of workload. Hart (1990) and Hart and Staveland (1988), however, have carried this proposition much further by identifying specific sources of workload, distinguishing sources of variance (some of them unwanted) in workload assessment, and devising ways of minimizing between-subject differences. In Phase II we will use the NASA-TLX as the best available metric. This decision means that a candidate eye-movement metric must correlate substantially with the NASA-TLX to warrant further consideration. Perfect correlation with the NASA-TLX is not, of course, desirable because in that case the one or the other measure(s) would be superfluous. On the other hand, the eye-movement measure(s) must identify substantially the same tasks as imposing heavy (or light) workload as does the NASA-TLX. If this association is strong, the possibility emerges that the eye-movement measure(s) can extend the NASA-TLX in important ways.

The NASA-TLX is a subjective measure. This fact is not so much a weakness as an omission. If valid objective measures can be developed, then their addition to our measurement armamentarium would give us a more complete or all-sided assessment procedure. In this connection, it should be pointed out that eye-movement measures, while objective, are not subject-independent. The same measure in different subjects will take different values for the same task. The use of an objective measure does not, therefore, "get around" all of the difficulties with subjective measurement. Between-subject differences, for example, will still be a problem. We discuss this further below.

Eye-movment measures do, however, offer some clear-cut advantages. First, an eye-movement measure can be taken while the task is being perfomed (in real time). One does not have to wait until the task is over to ask the subject about it. Second, one can take an eye-movement measurement over very small spans of time. This second advantage allows us to track the course of workload over time in ways that could be very significant, especially in a military context, or when adding/subtracting displayed task elements or changing them in any way.

Consider a pilot who is asked to perform an additional task in the cockpit. At first, this task imposes a substantial additional workload, because the pilot is unfamiliar with it and has not yet learned to perform it with minimal attention. As the pilot becomes more experienced with the additional task, he learns to perform it with a minimal commitment of attentive resources. As he does, the workload imposed by the task diminishes--not because the task has changed but because the pilot has learned how to perform it with greater ease and less attention. This kind of change could be tracked over time by eye-movement measures. As a result, we could assess workload effects and possibly very important ones that would not be detected by a NASA-TLX administerer at the end of the flight. The gains to be achieved by eye-movement measures are not a matter of substituting one measure or kind of measure for another but, rather, of extending a proven measure in new directions, of allowing the measurement of workload phenomena that could not otherwise be addressed. It is such an advantage that we see in eye-movement measures, and the tracking in time of workload processes, especially *in situ*, is *the kind of new phenomenon* that we foresee as being brought within the range of study.

We would envision that it would require perhaps two or three experiments "off-line" and in our laboratory to gain acceptable experience with the NASA TLX workload technique using our psychophysically scaled task (the Counting Test) as well as several of the APTS cognitive and information processing tasks. Then, at the end of the first year or the beginning of the second we would propose the introduction of the automated eye movement data acquisition and analysis system into an NADC Reconfigurable Cockpit study as a piggy back to on on-going experiment. In order to assure successful accomplishment of such an enterprise, a pair of subobjectives must first be realized:

1. A standard methodology needs to be developed where work load or input can be psychophysically scaled by task selection of "calibrated" tasks. In the present study, we believe the Counting Test series, plus tests from APTS, fit this requirement. Through software subroutine, both types of performances can now be automatically presented and we have successfully demonstrated performance differences and eye movement differences in various combinations of these tasks. However, the present experiment was largely a demonstration of feasibility of this approach and much longer data samples of eye movenents will be required. It will also be

necessary to conduct test-retest reliability sessions of the eye-movement and blink metrics.

2. Scoring of ocular activity automatically, by computer, so that analyses can be conducted shortly after exposure and then, at sometime later, has been accomplished. Figures 1-3 show how some of these are performed. It is our opinion that the customized automated system that we develop should feed naturally into and make use of standard computer-based analyses packages used in the behavioral, physiological, and engineering sciences. It should also be menu driven, user friendly, and flexible for use by DOD cockpit automation and workload experts. This has not yet been done. As we completed the present effort, the transformation of the data was a significant obstacle. However, much of this will be easier the second time.

Intersecting these two efforts, and overarching them, is an analytic, really a meta-theoretic concern. Table 5* above serves as a point of departure and is repeated here below for convenience.

TABLE 5. Intercorrelations Between Task Load Variables and Eye Movement Variables

| | RPM RANK | MOT RANK | BLINKS | BLKDMEAN | BLKAMN | SACCADES | VEL | SLIT |
|---|---|---|---|---|---|---|---|---|
| DIFFRANK | -.98 | .50 | .07** | .26* | .15** | .27*** | -.05 | .09 |
| RPMRANK | | -.56 | -.06** | -.27* | -.13** | .32*** | .08 | -.09 |
| MOTRANK | | | -.04 | .09 | .02 | .61**** | -.30* | .11 |
| BLINKS | | | | -.23 | -.16 | .02 | .05* | .11* |
| BLKDMEAN | | | | | .58** | -.03 | .08 | -.14 |
| BLKAMN | | | | | | .06 | .01 | -.15 |
| SACCADES | | | | | | | -.31 | .25* |
| VEL | | | | | | | | .04 |

Signif: **** p < .001; *** p < .01; ** p < .02; * p < .07

Table 5 concerns a single subject. The correlations reported are calculated among the eight measures over tasks. The sample size, if you like, is the number of tasks, in this instance, 10. This kind of correlation has been called the "P-technique" by R. B. Cattell (1949). It contrasts sharply with the usual R-technique, in which correlations are calculated among measures over subjects or the less familiar Q-technique, in which correlations are calculated among subjects over measures (the inverse of R-technique).

---

*Source: see page 31.

In the present case, there is still another methodological approach that might have been taken. The application of R-technique to the eye movement data results in as many matrices or separate analyses as there are subjects, that is, 12. The analysis of results individual by individual has only recently developed in a fully formal way. The more conventional approach would have been to calculate correlations among the eight measures "within subjects." In this approach the 10 values that eye blink (for example) takes for the same subject on the 10 tasks are averaged and the deviation of each eye-blink measurement from this average determined. The same is done for the other five subject-dependent measures. Correlations are then run among measures over all (10X12=) 120 data points, where each data point is a deviation from the average for one subject. This way of proceeding is the one followed in analysis of covariance to obtain a "within-group correlation" freed from the effects of between-group differences. In the same way, a "within-subject correlation" is freed from the effects of between-subject differences. It could be argued, therefore, that a "within-subject" correlation matrix should have been calculated. It would have given us a single correlation matrix that could reasonably have been taken as a comprehensive or general representation of the association among these eight measures within subjects or, put somewhat differently, freed from the between-subject differences.

The decision to calculate 12 correlation matrices (one for each subject) and not to merge them into a single within-subject matrix is a mundane data-analytic maneuver with large theoretical implications. This decision is essentially the same as the one to work with individual animals rather than means of groups of animals (B. F. Skinner) or to study the efficacy of behavioral modification in individual patients by means of withdrawal, ABA, or multiple baseline designs rather than in groups of patients by t-tests or analysis of variance. And the reasons for studying individuals are the same in both cases. The mean of a group does not necessarily characterize any individual member of that group. There may be no member at or near the group mean. In the same way a within-subject correlation of .5 (say) may not characterize any one of our 12 subjects. Six of them, for example, might have correlations of .75 and the other six of .25. The question at issue is the level of description at which one wishes to work.

We agree with Hart and Staveland (1988) that workload should be human centered rather than task centered, but we propose to go a step further. Our belief is that workload measurement, to be successful, must be carried out at the individual level. It is not enough to know that, on the average, Task A imposes a heavier workload than Task B and Task B than Task C. If this ordering really exists, then for each subject it should be the case that A is heavier than B is heavier than C. Occasionally, there may be exceptions. One subject may have a demonstrable hearing loss that makes C more work for him than B. Another may have a reading deficit that makes B more work than A. In each instance, however, there should be independent, external evidence to justify the inversion.

An insistence on working at the individual level involves more than a demand for strong evidence. It also involves a distinct brand of methodology. Cattell's (1949) P-technique has already been mentioned, as have baseline and withdrawal designs. The study of individual subjects has a long and distinguished history, stretching back from B. F. Skinner to the beginnings of physiology with Claude Bernard. A formal apparatus, however, for the analysis of experiments with an N of 1 is a more modern innovation (Hersen & Barlow, 1976). Workload measurement, in our opinion, needs to incorporate these newer designs of data-analytic methods if it is to establish results of real usefulness. It especially needs to develop statistical approaches for handling individual correlation matrices (P-technique) like the one in Table 5. The development of such approaches and their application to workload measurement will be a major thrust of any work we do in the future in this area.

There are several other items which emerged from these studies which require discussion.

It was found that the number of eyeblinks tended to increase with task load, which is congruent with indications of previous research (Stern, 1990). Surprisingly, eyeblink duration also increased as task load increased, which is counterintuitive in the sense that it represents longer interruption of visual input during the higher task load. Also, this highly significant finding is in a direction opposite to that described by Stern (1990) as a function of increased work load. While Stern shows that fatigue can cause duration to increase, at present, we have no explanation for this disparity. One result of longer blink duration with load is present in our 1-minute APTS tasks and our 15-minute counting tasks. Also, our tasks are not markedly different from those of Sterns, Walrath, and Goldstein (1984), who used a discrimination task presented via different modalities. Alternatively, as was seen (in Table 2), the amount of ocular motility that may be expected to occur in the different mental tasks of the APTS varies by several orders of magnitude and so these elements too need to be taken into account in studying the Counting Test results. We were not able to quantify ocular motility in our counting test to the extent we did with our APTS tests and so visual demands could also interact with this metric. Resolution of these issues awaits future study.

We hypothesize that changing visual requirements across the APTS battery subtests served to mask some of the between task differences in task demands as indexed by the three eye movement measures surfaced in the earlier phase of the study. For example, Grammatical Reasoning entails the reading of sentences,for which necessary eye movements are likely to differ in fundamental aspects from the requirements of math processing, for which digits appear at a central screen location. On the other hand, we feel that if we had varied task difficulty within the particular subtest, which we could have done by requiring faster response speed or providing only the most difficult test items, we may well have seen work load differences in terms of increased blink duration and eye movement range within the particular subtest.

It also should be pointed out that eye movement data collection for each of the APTS subtests took place over considerably shorter intervals (1 min.) than the similar data collection for the counting task (15 min.). For this reason, one would expect considerably greater reliability and precision for data collected over the longer interval via the Spearman-Brown relationship. Determination of optimal intervals over which to average eye movement work load indices, in terms of relaibility, accuracy, and precision, should be an important aspect of subsequent research. In particular, we believe that as a pilot has new displays added, as may be envisioned in studying automation issues in the NADC Reconfigurable Cockpit (Morrison, Gluckman, & Deaton, 1990), it will be necessary to have a within-session estimate of work load which can follow a time course of change.

A nonintrustive indicant of attention can index changing task difficulty in tactical mission performance as new displays are added to cockpit real estate. Applications are as intelligent or adaptive (biocybernetic) system by which function allocation of cockpit displays can be effected and evaluated. Further, a bioelectric index during simulated or experimentally based studies of cockpit display can also provide objective assessment of task difficulty and work load which permit test and evaluation of systems for human use as well as improve the basis on which they are designed. Finally, such metrics may also reflect individual differences in monitoring capability and thereby aid in job classification as well as job selection.

In Phase II additional experiments will be conducted and the prototype automated data analysis system will be miniaturized and made more transportable. It should be noted that we will simultaneously record and score eye movements while the subject is performing the task. Therefore, a very high speed microprocessor is necessary. It is expected that by the time this work is conducted, the availability and cost of the 35 Mhz i486 machines will be similar to the Essex high speed microcomputer, but with greater disk and storage memory.

If the physiological indicants studied in Phase I bear a relationship to time on task and task load, we propose in Phase II to assess the generality of their relationship to work load by testing them against other tasks which are known to vary in work load. To the extent that we could be accommodated in on-going work in the Reconfigurable Cockpit at the NADC, we would want to try out out electrophysiological system as soon as practicable. In Phase III, we propose to design and develop a go-everywhere "bio-pack", a strap on device, which will allow for rapid and portable on-site measurement of work load and thus improve human quality control. This effort will take cognizance of the experimental results, and where practical, incorporate the features found beneficial (e.g., two-dimensional eye movement recording, eye blink rejection, etc.), and perhaps technical variables not yet studied (e.g., bruxation, buccal muscle tension, etc.) which may limit its usage in flight studies.

# CONCLUSIONS

The accelerating tempo of military operations increases the task demands and work-related stresses imposed on human operators. As emerging cockpit display systems are incorporated into military tactical aircraft, a metric is required to ensure that human cognitive work load limits are not exceeded. Subjective techniques are available, but generally depend on the self-report of users (Hart & Staveland, 1988), and a need exists for an objective method. Some progress in this endeavor has been achieved in using as objective work load indicators a variety of electrophysiological techniques, including EEG and neural-evoked potentials. However, these objective techniques tend to be intrusive, relatively artificial, and nonportable. We believe there are more readily obtainable measures that can serve as simple external indicants of work load, and which can eventually be bundled in portable, "vest-pocket" systems to be employed in applied work load assessment. In Phase I, we conducted electrophysiological recordings of the action of the eye while subjects attended and responded to work with different visual task demands. The bioelectric actions we recorded included eye movements (frequency, position, velocity, acceleration, range, etc.) and eye blinks (duration, frequency, acceleration, etc.). Three developments were undertaken: 1) an experimental paradigm was created whereby visual and mental tasks with disparate demand characteristics (memory, reaction time, search, spatial perception; information processing) were presented under computerized control; 2) a software package was developed to run on desktop computers which reduce, score, and analyze these data; and then 3) another software package was developed to be used with desktop personal computers to compare bioelectric output variables to characteristics of the visual work load demands. The general findings are that elements of eye activity (range of movement, number of blinks and blink duration) correlated at a statistically meaningful level with the visual task demands and with the work load. Other bioelectric variables are available with this automated scoring package. These were also examined and some show promise, but either demonstrated similar relationships to those above, but were statistically weaker, or they were related to different aspects of the visual task loads. In Phase II this system would be further developed along with field manuals for use by systems developers to objectively assess visual work load parameters of various aspects of aviation activity.

The Phase I effort was the first step in the design of an automated task load analysis system for improving function allocation and adapting task loading to the operators' capacities as indexed by subjective, objective, and behavioral indicants. This will permit rapid and portable on-site measurement of tactical aviation cockpits and workspaces. The commercial availability of such a package would provide a desktop capability, with a common metric, for aircraft manufacturers and others in the private sector for conducting human factors engineering design, test and evaluation of workstations of all kinds.

# REFERENCES

Bakan, P., & Strayer, F. F. (1973). On reliability of conjugate lateral eye movements. Perceptual and Motor Skills, 36, 429-430.

Berbaum, K. S., Kennedy, R. S., & Hettinger, L. J. (1991). Visual tasks in helicopter shipboard landing. Submitted for publication to Journal of Applied Ergonomics.

Borg, G. (1978). Subjective aspects of physical and mental load. Ergonomics, 21, 215-220.

Campbell, J. P., Dunnette, M. D., Arvey, R. D., & Hellervik, L. V. (1973). The development and evaluation of behaviorally based rating scales. Journal of Applied Psychology, 57, 15-22.

Cattell, R. B. (1949). $r_p$ and other coefficients of pattern similarity. Psychology, 14, 179-198.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences. New York: Academic Press.

Cohen, A. S.(1977). Is the duration of an eye fixation a sufficient criterion referring to information input? Perceptual and Motor Skills, 45, 766.

Cohen, B., Feldman, M., & Diamond, S. P. (1969). Effects of eye movement, brain-stem stimulation, and alertness on transmission through lateral geniculate body of monkey. Journal of Neurophysiology, 32, 583-594.

Cooper, G. E., & Harper, R. P. (1969). The use of pilot rating in the evaluation of aircraft landing qualities. National Aeronautics and Space Administration, Ames Research Center, TN-D-5153, Moffett Field, CA.

Crommelinck, M., & Roucoux, A. (1976). Characteristics of cat's eye saccades in different states of alertness. Brain Research, 103, 574-578.

Damos, D. L., Bittner, A. C., Jr., Kennedy, R. S., & Harbeson, M. M. (1981). Effects of extended practice on dual-task tracking performance. Human Factors, 23(5), 627-631.

Donchin, E., & Kramer, A. (1986). The event-related brain potential as an index of information processing: A program of basic research (Technical Report AFOSR-CPL-86-1). Air Force Office of Scientific Research, Bolling AFB, DC.

Gal, R. (1975). Assessment of seasickness and its consequences by a method of peer evaluation. Aviation, Space, and Environmental Medicine, 46, 836.

Gardner, K. (1967). Some patterns of fixation saccadic eye movements. Psychonomic Science, 7, 146-146.

Goldstein, R., Stern, J., & Bauer, L. (1985). A psychophysiological mapping of cognitive processes (Contract No. F49620-83-C-0059). Bolling Air Force Base, DC: Air Force Office of Scientific Research.

Gopher, D., & Braune, R. (1984). On the psychophysics of work load: Why bother with subjective measures? Human Factors, 26(5), 519-532.

Gray, S. F. (1980). Forty-eight versus twenty-four hour duty for USAF missile crews: A feasibility study using subjective measures of fatigue (SAM-TR-80-39). Brooks Air Force Base, TX: USAF School of Aerospace Medicine.

Gregory, R. L. (1973). Eye and brain. New York: McGraw Hill.

Hart, S. G. (1990). Crew workload strategies in advanced cockpits. Proceedings of the Aviation Safety/Automation Program Conference (pp. 105-125). Moffett Field, CA: National Aeronautics and Space Administration, Ames Research Center.

Hart, S. G., Childress, M. E., & Bortolussi, M. (1981). Defining the subjective experience of work load. Proceedings of the Human Factors Society 25th Annual Meeting, pp. 527-531. Human Factors Society, Santa Monica, CA.

Hart, S. G. & Hauser, J. R. (1987). Inflight application of three pilot work load measurement techniques. Aviation, Space, and Environmental Medicine, 58, 402-410.

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

Hersen, M., & Barlow, D. H. (1976). Single case experimental designs: Strategies for studying behavior change. New York: Pergamon Press.

Jerison, H. J. (1956). Effect of a combination of noise and fatigue on a complex counting task (WADC TR-55-360). Dayton, OH: Wright Air Development Center, Wright-Patterson AFB.

Kantowitz, B. H., & Weldon, M. (1985). On scaling performance operating characteristics: Caveat Emptor. Human Factors, 27(5), 531-547.

Kennedy, R. S. (1971). A comparison of performance on visual and auditory monitoring tasks. Human Factors, 13(2), 93-97.

Kennedy, R. S. (1972). The relationship between habituation to vestibular stimulation and vigilence: Individual differences and subsidiary problems. (NAMRL Monograph 20) Doctoral dissertation, University of Rochester. (Also in Dissertation Abstracts International, 1972, No. 72-78, 764).

Kennedy, R. S. (1978). Bioelectric indicants of diver's ability to perform useful work. Proceedings of the Undersea Medical Society Workshop. Bethesda, MD: Undersea Medical Society. (NTIS No. AD A060675)

Kennedy, R. S., Baltzley, D. R., Dunlap, W. P., Wilkes, R. L., & Kuntz, L. A. (1989, May). Microcomputer-based tests for repeated-measures: Metric properties and predictive validities (EOTR-89-02). Orlando, FL: Essex Corporation.

Kennedy, R. S., Baltzley, D. R., & Osteen, M. K. (1988). A microcomputer test battery: Normative data and sensitivity to military stressors. Paper presented at the 30th Annual Military Testing Association Conference, Arlington, VA.

Kennedy, R. S., Baltzley, D. R., Wilkes, R. L., & Kuntz, L. A. (1989). Psychology of computer use: IX. A menu of self-administered microcomputer-based neurotoxicology tests. Perceptual and Motor Skills, 68, 1255-1272.

Kennedy, R. S., Fowlkes, J. E., & Smith, M. G. (1989). Enhancing cognitive performance through biocybernetic feedback: An individual differences approach (Contract No. F33615-88-C-0640). Brooks Air Force Base, OH: USAF School of Aerospace Medicine.

Kennedy, R. S., May, J. G., Jones, M.B., & Fowlkes, J.E. (1989). Development of saccade length index of taskload for biocybernetic application (Final Report AFOSR Contract F49620-87-C-002). Orlando, FL: Essex Corporation.

Krebs, M. J., Wingert, J. W., & Cunningham, T. (1977). Exploration of an oculometer-based model of pilot workload (NASA-CR-145153, 76-SRC/39). Minneapolis, MN: Honeywell, Inc. Systems & Research Center.

Lane, N. E., & Kennedy, R. S. (1988). A new method for quanitfying simulator sickness: Development and application of the simulator sickness question- naire (SSQ) (EOTR 88-7). Orlando, FL: Essex Corporation.

Loren, S. A., & Darrow, C. W. (1962). Eye movements, EEG, GSR, and EKG during mental multiplication. Electroencephalography and Clinical Neurophysiology, 14, 739-746.

Lynn, R. (1966). Attention, arousal and the orientation reaction. Oxford: Pergamon Press.

Malmstrom, F. V., Randle, R. J., Murphy, M. R., Reed, L. E., & Weber, R. J. (1981). Visual fatigue: The need for an integrated model. Bulletin of the Psychonomic Society, 17(4), 183-186.

May, J. G., Kennedy, R. S., Williams, M. C., Dunlap, W. P., & Brannan, J. R. (1990). Eye movement indices of mental work load. Acta Psychologica, 75, 75-89.

Monty, R. A., Hall, R. J., & Rosenberger, M. A. (1975). Eye movement responses of heroin addicts and controls during word and object recognition. Neuropharmacology, 14, 693-702.

Moray, N., Turksen, B., Aidie, P., Drascic, D., Eisen, P., Kruschelnicky, E., Money, L., Schonert, H., & Thornton, C. (1986). Progress in mental work load measurement. Proceedings of the 30th Annual Meeting of the Human Factors Society, 1121-1123.

Morrison, J. G., Gluckman, J. P., & Deaton, J. E. (1990). Adaptive function allocation for intelligent cockpits. Cockpit automation study 1: Baseline study. Warminster, PA: Naval Air Development Center.

O'Donnell, R. D. (1981). Development of a neurophysiological test battery for work load assessment in the U.S. Air Force. Proceedings of the International Conference on Cybernetics and Society (pp. 298-402). Atlanta, GA: IEEE, Systems, Man & Cybernetics Society.

Peacock, B., Glube, R., Miller, M., & Clune, P. (1983). Police officers' reponses to 8 and 12 hour shift schedules. Ergonomics, 26(5), 479-493.

Reid, G. B., Shingledecker, C. A., & Eggemeier, F. T. (1981). Applications of conjoint measurement to work load scale development. Proceedings of the Human Factors Society 25th Annual Meeting. Human Factors Society, Santa Monica, CA.

Reid, G. B., Shingledecker, C. A., Nygren, T. E., & Eggemeier, F. T. (1981). Development of multi-dimensional subjective measures on work load. Proceedings of the International Conference on Cybernetics and Society, IEEE Systems Man and Cybernetics Society, Atlanta, GA.

Sheridan, T. B., & Stassen, H. G. (1979). In N. Moray (Ed.), Mental workload: Its theory and measurement (pp. 219-233). New York: Plenum.

Shingledecker, C. A. (1982). Performance evaluation of the embedded secondary task technique. Preprints of 1982 Annual Scientific Meeting, Aerospace Medical Association. Washington, DC: Aerospace Medical Association, 151-152.

Snider, R. S., & Lowy, K. (1968). Evoked potential and microelectrical analysis of sensory activity within the cerebellum. In Fourth Symposium on the Role of the Vestibular Organs in Space Exploration (pp. 145-258), Pensacola, FL.

Stern, J. A. (1990). The eyeblink: Effective and cognitive influences. Proceedings of the International Conference on Anxiety, pp. 1-3.

Stern, J. A., Walrath, L.C., & Goldstein, R. (1984). The endogenous eyeblink. Psychophysiology, 21, 22-23.

Stevens, S. S. (Ed.). (1951). Handbook of experimental psychology. New York: Wiley.

Storm, W. F. (1980). E-4B crew fatigue associated with 30-hour IOT&E mission (SAM-TR-80-40). Brooks Air Force Base, TX: USAF School of Aerospace Medicine.

Turnage, J.J., & Bliss, J. P. (1990, October). An analysis of skill transfer for tank gunnery performance using TOPGUN, VIGS, and ICOFT trainers (ARI Tech. Rep. No. 916). Orlando, FL: Army Research Institute, PM Trade.

Turnage, J. J., Kennedy, R. S., Gilson, R. D., & Nolan, M. D. (1989). A microcomputer test battery of vision tests for the prediction of flight training performance. Paper presented at the 34th Annual Meeting of the Human Factors Society.

Walker, H., & Lev, J. (1953). Statistical inference. New York: Holt.

Weale, R. A. (1960). The eye and its function. London: Hatton Press.

Wickens, C. D., & Yeh, Y. Y. (1983). The disassociation between subjective work load and performance: A multiple resource approach. Proceedings of the Human Factors Society 27th Annual Meeting (pp. 244-248). Santa Monica, CA: Human Factors Society.

Wierwille, W. W., Rahimi, M. & Casali, J. G. (1985). Evaluation of 16 measures of mental work load using a simulated flight task emphasizing mediational activity. Human Factors, 27(5), 489-502.

Yules, R. B., Krebs, C. Q., & Gault, F. P. (1966). Recticular formation control of vestibular system. Experimental Neurology, 16, 349-358.