AD-A169 941

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

DTIC ELECTE JUL 2 4 1986 S D D

1a. REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 86-0416
6a. NAME OF PERFORMING ORGANIZATION Essex Corporation / 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research/NL
6c. ADDRESS (City, State, and ZIP Code) 1040 Woodcock Road, Suite 227 Orlando FL 32803	7b ADDRESS (City, State, and ZIP Code) Building 410 Bolling AFB, DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR / 8b. OFFICE SYMBOL (If applicable) NL	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620-85-C-0121
8c. ADDRESS (City, State, and ZIP Code) Building 410 Bolling AFB, DC 20332-6448	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61102F; PROJECT NO. 3005; TASK NO A1; WORK UNIT ACCESSION NO

11. TITLE (Include Security Classification)

Eye Movements as an Index of Mental Workload

12. PERSONAL AUTHOR(S)

J. G. May, R. S. Kennedy, M. C. Williams, W. P. Dunlap, & J. R. Brannan

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Final	FROM 15Jul85 TO 14Mar86	17 March 86	19

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Eye movement velocity, saccadic latency, saccadic extent, mental workload, human performance
15	08		

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

Two investigations were carried out to assess the feasibility of using eye movement measures as nonintrusive indicants of mental workload. In the first experiment, measures of saccadic latency and eye movement velocity were obtained during alternating eye movement scans while subjects were differentially task loaded by simple, moderate, and complex auditory tone counting. The latency and eye movement velocity measures changed but did not differ reliably as tone counting complexity (workload) was increased. In the second experiment, the spatial extent of spontaneous saccades was measured under three levels of tone counting complexity. The results indicated that the extent of such eye movements varied inversely (p less than .0151) as tone counting complexity increased. This index appears to hold promise for the development of an objective indicator of mental workload.

DTIC FILE COPY

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Alfred R Fregly	22b. TELEPHONE (Include Area Code) (202) 767-5021	22c. OFFICE SYMBOL NL

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

146

AFOSR-TR- 86-0416

EYE MOVEMENTS AS AN INDEX OF MENTAL WORKLOAD

James G. May*, Robert S. Kennedy#,
Mary C. Williams*, William P. Dunlap+,
and Julie R. Brannan*

Submitted to:
Dr. Alfred Fregly, AFOSR
Directorate of Life Sciences
Building 410
Bolling AFB, DC 20332-6448

Submitted by:
Essex Corporation
1040 Woodcock Road, Suite 227
Orlando, FL 32803
(305) 894-5090

FINAL REPORT
Contract F49620-85-C-0121
Phase I, Topic #174

*Department of Psychology, University of New Orleans, Lakefront, New Orleans, LA 70148
#Essex Corporation, Orlando Office, 1040 Woodcock Road, Orlando, FL 32803
+Department of Psychology, Tulane University, New Orleans, LA 70118

INTRODUCTION

Workload refers to demands imposed on a human operator by a given task, and workload measurement involves an attempt to characterize conditions under which task demands can or cannot be met by the performer (Gopher & Braune, 1984). Because it has been suggested that there may be little or no deterioration in performance until the point of failure is closely approached (Schmidt, 1978), sensitive measures of workload are of vital importance. Workload assessment can be used not only to evaluate pilot performance requirements, but also to predict workload changes with system modification. Additionally, by assessing workload impact on individuals for tasks of constant difficulty, workload indices, if reliable, can be used to determine individual differences in capability and thereby aid in personnel selection.

It is generally accepted that humans are limited capacity information processors, and that human performance is a function of both individual processing capabilities and of task demands (Kahneman, 1973; Moray, 1967; Wickens, 1984). It is inevitable, therefore, that in certain situations, human performers reach the upper limits of their ability to cope with task demands, and performance is jeopardized when these limits are approached or exceeded. For this reason, a need has arisen for the development of reliable, accurate, and nonintrusive measures of mental workload.

Various methods have been devised for the measurement of workload, but great disagreement remains concerning which method provides the most reliable and valid measure. A number of comprehensive reviews of the workload literature are available (Chiles & Alluisi, 1979; Moray, 1979, 1982; Wierwille, 1979; Wierwille & Williges, 1978, 1980; Williges & Wierwille, 1979; Wickens, 1984) and three symposia (AGARD, 1977, 1978; Frazier & Crombie, 1982) describe the state-of-the-art.

The three general approaches to the measurement of workload employed are subjective, behavioral, and physiological metrics. Subjective and physiological measures provide scalar indices of workload, but tend to be insensitive to demands on cognitive resources, while behavioral measures offer greater diagnostic capability of performance capacity on multiple dimensions (Wickens, 1984). The vast majority of workload research has involved subjective measures, where a performer makes a conscious judgment regarding the difficulty of the task at hand. Several subjective measurement scales have been developed (see Gopher & Braune, 1984, for review), all of which require the operator to rate the subjective workload associated with the performance of a particular task. These scales include the Cooper-Harper rating scale (Cooper & Harper, 1969), a modification of the Cooper-Harper rating scale (Sheridan & Simpson, 1979), bipolar rating techniques (Bird, 1981; Hart, Childress, & Bortolussi, 1981), the Subjective Workload Assessment Technique (SWAT) (Reid, Shingledecker, & Eggemeier, 1981; Reid, Shingledecker, Nygren & Eggemeier, 1981), and Gopher & Braune's (1984) application of magnitude estimation originally developed by S. S. Stevens (1957).

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, these scales often show low correlations with objective measures of task performance (Wickens, Sandry,

& Hightower, 1982) so that their usefulness in predicting performance is compromised. The advantages of using such scales lies 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 of workload have been used to assess the relationship between performance and workload in physical tasks (Borg, 1978; Johannsen et al., 1979; Tulga, 1978; Verplank, 1977), cognitive tasks (Borg, 1978; Borg, Bratfisch, & Dornic, 1971, 1972; Bratfisch, Borg, & Dornic, 1972), and manual control tasks (Cooper, 1957; Cooper & Harper, 1969). Although significant correlations were obtained in all of these studies, the correlations were among subjective judgments of workload and not with objective measures of performance. Thus, subjective methods are limited to the information available to 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.

An alternative to subjective measures of workload is to take direct physiological measures (e.g., heart rate, respiration, GSR, ERP) during task performance. Such an approach eliminates the possibility of subjective distortion and generally does not interfere with task performance. The drawback to this approach is that measures of autonomic nervous system function may be more likely to reflect stress induced by the task rather than information processing load (Shingledecker, 1982), and often these measures may lack stability and have insufficient reliability for statistical power (Cohen, 1977). Some of them also may intrude on the work to be performed (Krebs, Wingert & Cunningham, 1977; O'Donnell, 1979) and several of them require averaging (Goldstein, Stern & Bauer, 1985; Donchin & Kramer, 1986; Kaufman & Williamson, 1983).

A final approach to the measurement of workload involves obtaining direct behavioral (performance) measures. Here, an evaluation of an operator's overt task behavior (e.g., speed or accuracy of performance) is made. One such approach involves administering a primary task simultaneously with an additional, secondary task (Shingledecker, 1982). As the difficulty level of the secondary task is increased, a point will be reached when the operator's processing capacity is exceeded, the performance decrement on the primary task will be inversely proportional to the secondary 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. Workload, then, can be indexed by the difference between single and dual task performance. 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 workload, 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. Perhaps the most efficacious approach to the assessment of workload would be a combination of objective and overt performance measures. Simultaneous application of physiological and performance measures may be a step toward linking human performance to underlying mechanisms.

We are aware of the large investigative effort underway studying event related brain potentials (e.g., Donchin & Kramer, 1986; Gevins et al., 1984; Goldstein, Stern & Bauer, 1985; Hoffman, Houck, MacMillan, Simons & Oatman, 1985; Lewis 1982; 1983a,b,c) as indicators of cognitive activity and workload. But these potentials are triggered responses and require some averaging. Because visually guided behavior is one of the most prominent characteristics of all diurnal primates we considered that eye movement behavior might also hold promise for biocybernetic applications. More specifically, intricate eye-hand and eye-head coordination represents a most sophisticated cybernetic mechanism involved in human spatial orientation, attention, and complex information processing. Consequently, eye movements, particularly those involving binocular foveal fixation and scanning, could represent very sensitive measures of alertness, cognitive, and motor performance. More than other sensory systems (Snider & Lowy, 1968) the eye has embryological connections to the cortex (Gregory, 1973; Weale, 1960; Patten, 1951). In view of the central role of eye movements in visual, cognitive, and refined motor functions, it is not surprising that numerous studies have begun to relate various quantitative aspects of eye movements to attention, cognitive capacity, mental effort, fatigue, drug state, and the integrity of the underlying neural mechanisms. A large literature has always existed which reported on the relevance of where a person was looking for performance. The present approach examines eye movement activity as an implied index of the overall mental alertness of the individual.

Several years ago we (Kennedy, 1972) reviewed the literature at that time which correlated aspects of eye movement activity to the mental state of the subject. The reported studies were not programmatic nor even thematic, and since that time several texts have appeared (e.g., Carpenter, 1977; Ditchburn, 1973; Senders, Fisher, & Monty, 1979) but we are aware of no consistent trends. Impetus for the present effort began with our work in vestibular nystagmus where we showed that keeping track performance covaried with changes in fast phase activity (Kennedy, 1972). Subsequently, in a pilot study (Kennedy, 1978), there appeared to be evidence for eye movement velocity being related to performance but these findings were not persued. Relatedly, 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).

More obliquely relevant is Guedry's (1965) review where he referenced about 20 papers where the subject's mental state modified recorded vestibular nystagmus, and it was later shown that nystagmus fast phase was absent in patients who lacked a pontine reticular formation (Daroff & Hoyt, 1971). Cohen, Feldman, and Diamond (1969) and Yules, Krebs, and Gault, (1966) note that 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. Characteristic and spontaneous eye movements have been related to hemispheric specialization of cognitive, affective, and physiological variables (Bakan & Strayer, 1973). Lastly, Wierwille, Rahimi & Casali (1985) have had some success with eye blinks and fixation duration in a simulator but less successful were Wilson, O'Donnell, and Wilson (1982), who explored eye movement activity in an A-10 ground-based flight simulator.

It is interesting that one of the technologically more difficult problems in evoked potential recording is rejecting the parts of electrical brain potential changes that are related to eye movements, which are viewed as artifacts (Gevins et al., 1984). These so-called artifacts were the proposed topic of study in the current research plan. While our hypothesis was that the velocity of eye movements would be greater during high versus low workload we would also look at different aspects of eye movements (viz., latency, extent, dwell times, etc.). Our purpose was to surface an eye movement indicator which would bear a monotonic relationship to workload, states of preparedness, alertness, and attention.

METHODS

Experiment 1

Subjects - Five subjects (one male and four females) ranging in age from 24 to 34 years participated in this experiment. All had normal vision and hearing, and were well practiced on the tone counting task (cf. Kennedy, 1972 for a review).

Apparatus - Eye movements were recorded from the left eye via an infrared tracking method and electro-oculographic techniques. Signals from infrared tracking apparatus (Eye Trac, Model 160) and HgCl electrodes positioned at the inner and outer canthi were amplified (X1000) and fed into an FM tape recorder together with trigger pulses associated with fixation light alternation. Offline data reduction involved recording individual saccades from each channel into two channels of a signal processor (Nicolet Model 1072), measuring the latencies of each, and deriving the peak velocity of each through differentiation. The complex counting test of Jerison (1956) was modified to be presented auditorily (Kennedy, 1972) because it has been shown to be sufficiently stable and the amount of pretraining required was minimal (Kennedy & Bittner, 1980). Tone counting tasks of various complexities (workload) were administered with a microcomputer (NEC PC 8201A) which was programmed to present a series of high, medium, and low frequency tones (duration = 500 msec) in a pseudo-random series at an average rate of 2.0 Hz. Performance data were computed and stored on the microprocessor.

Procedure - Each subject practiced the tone counting tasks until performance exceeded 70% correct. The first task required that they count only the low tones (low task load) and press a key after each fourth low tone. Thirty-six low tones were presented together with 28 medium tones and 24 high tones. The second task required that they count the middle tones (medium task load) and depress a different key after each fourth middle tone. The third task required that they combine the two previous tasks and depress a different key after the occurrence of each fourth low and fourth middle tone (high task load). The program recorded number correct, number missed, and number incorrect (false positives). An error caused the scoring routine to reset. The subjects first performed each of the counting tasks without alternative fixation. Within each subsequent session the subject alternated fixation from left and right, fixating either of the two red LEDs spaced 20 degrees apart horizontally. Next, they performed the low, middle, and combined counting tasks (in order) while alternatively fixating. The rate at which the fixation lights were alternatively illuminated was aperiodic and averaged 0.2 Hz. Ten to 12 saccades were required throughout the duration of the tone-counting tasks. As a check on practice effects, the low task was performed again at the end of each session. All subjects used a bite-bar to maintain stable head position.

RESULTS

Performance Data

The tone counting accuracy scores (see Figure 1) obtained during the pretest (low, medium, and high task loads) and during alternating fixation were submitted to an analysis of variance which revealed a significant main effect for conditions ($F = 5.02$, df 6, 24, $p = .0018$). Subsequent Newman-Kuels Range tests revealed that pretest scores for the low task load differed significantly from those for the high task load ($p = .0080$), and scores for the medium load task differed significantly from those for the high task load ($p = .0124$). This finding indicated that pretest tone counting was indeed more difficult when two tone types were counted. During alternate fixation, however, these differences were not obtained, suggesting that alternate fixation may have interfered with task performance such that performance difference due to workload were no longer significant.

Eye Movement Data

The saccades were digitized and displayed with a signal processor at an epoch of 204.8 msec. Each trace began at the time the fixation lights were alternately illuminated and the latency of an eye movement could be measured with 2 msec resolution. Typical saccades for left and right fixation together with the method of latency measurement are depicted in Figure 2 for infrared (ET) and EOG recording. Traces containing eye blink artifact were excluded. Each trace was then differentiated and measures of peak velocity were obtained (see Figure 2). At least eight such measures were obtained for each condition. Average latencies and peak velocities were derived for each condition for each subject. The group averages for these measures, together with the standard error of the mean, are presented as a function of experimental condition (task) in Figure 3. The horizontal dotted line indicates the group mean for that measure without the counting task.

Measures of saccadic latency and velocity for both EOG and ET recordings, under different conditions of workload, were submitted to four separate analyses of variance. For both the EOG and ET data, measures of eye movement velocity did not differ significantly from pretest levels under any of the workload conditions. A main effect for workload conditions (none, low, medium, high, and a second low workload) was significant for both latency measures ($F = 17.49$, df 4, 16, $p = .0000$, and $F = 10.43$; df 4, 16; $p = .0002$) for EOG and ET respectively. For the EOG measures, subsequent Newman-Kruels tests revealed significant increases in latency between the pretest measures and each of the counting conditions (first low - p. = .0000; medium - $p = .0011$; high - $p = .0009$; second low - $p = .0066$). In addition, significant differences in latency were obtained between medium and high conditions ($p = .0244$) and between first low and high conditions ($p = .0014$). Unfortunately, a significant decrease was obtained between the first and the second low task conditions ($p = .0014$). For the ET measures, the pattern of results was almost identical: Newman-Kruels tests revealed that all latencies under the workload task were significantly increased relative to the pretest measures

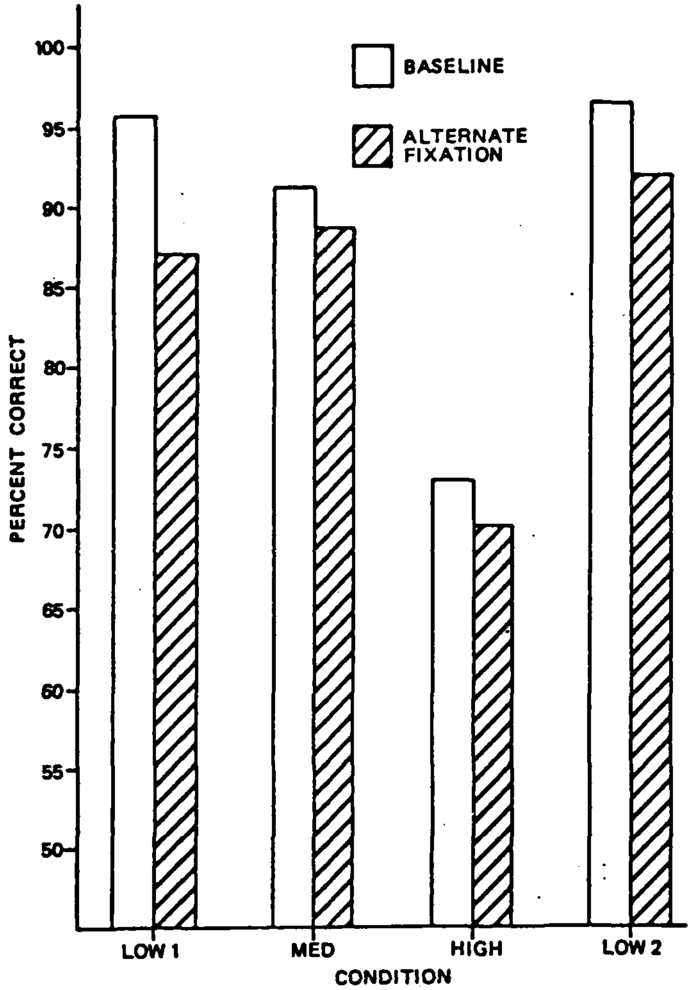

Figure 1. Tone counting accuracy scores before and during alternating fixation.

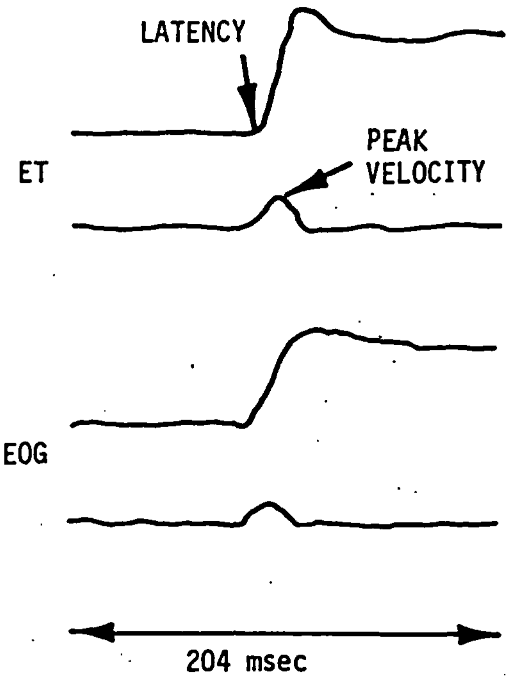

Figure 2. Typical saccades for left and right fixation for infrared and EOG recording.

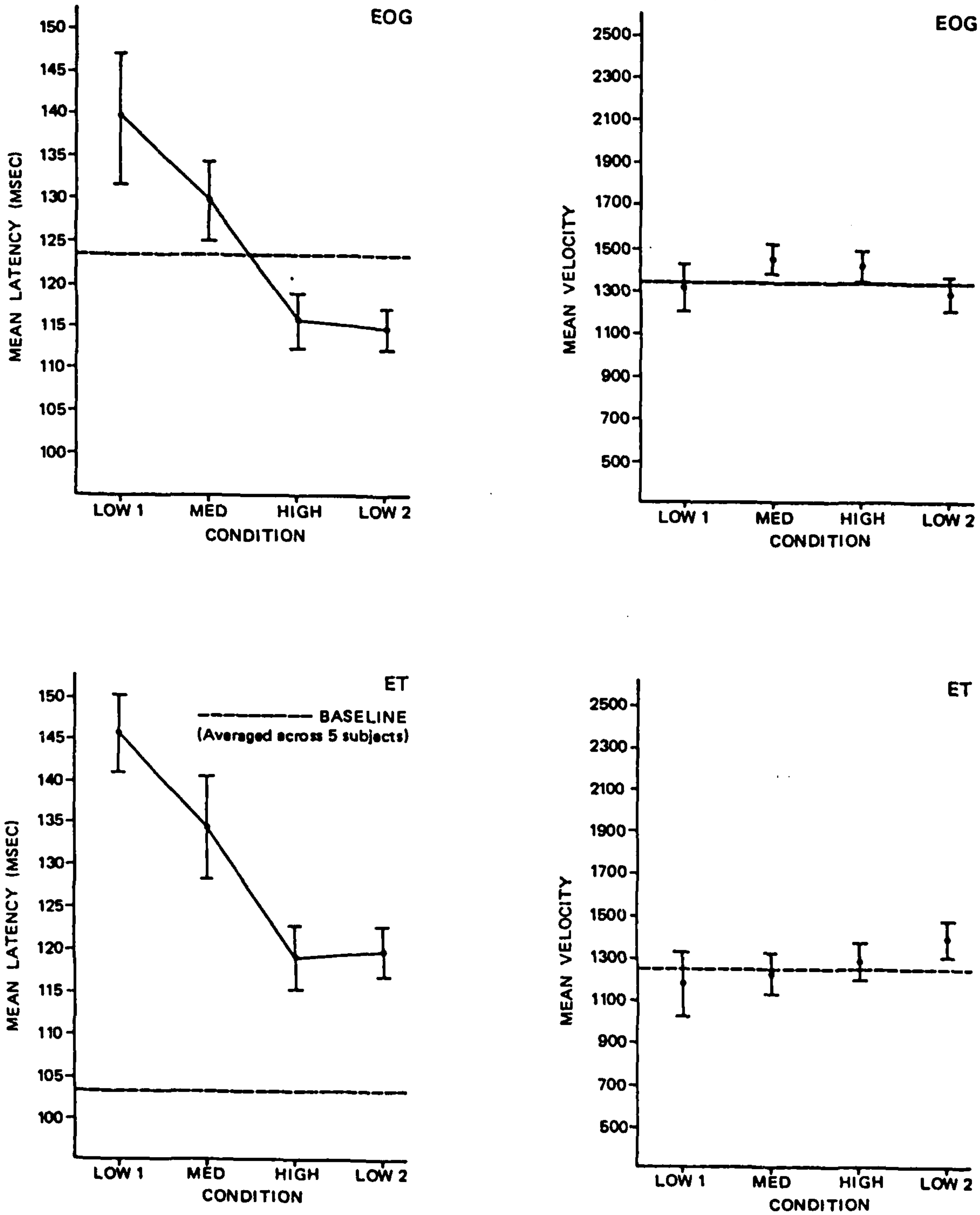

Figure 3. Group averages for latencies and peak velocities.

(first low - p = .0002; medium - p = .0016; high - p = .0290). Also, significant differences in latency occurred between the first low and the high condition (p = .0116), and again the latency under the first low condition was significantly higher than that for the second low condition (p = .0118).

Discussion

First, the interpretation of the results of Experiment I is complicated by the fact that differences in task performance due to increased workload, present during pretesting, diminished to the point of nonsignificance once subjects were required to alternatively fixate. This result might be due either to the possibility that workload was not sufficiently varied so that with practice workload differences diminished, or to the possibility that alternate fixation interfered with workload task performance to a point where differences were no longer detectable.

Performance results notwithstanding, it seems clear that although trends for higher velocity under workload were present, this index is too variable a measure to show significant impacts of workload manipulation. It should be remembered that the distance across which the eye moved was constant across conditions and not free viewing in the dark as was employed previously (Kennedy, 1978).

Eye movement latency appeared at first to be a more promising variable, but the data would suggest that low workloads have the greatest impact on latency. A more parsimonious explanation is that, initially, latency was longer when the workload task was introduced, but with practice, latency decreased decidedly. The significant decrease of eye movement latency from first to second test under low work load supports such a learning or practice interpretation and may related to the finding of Malmstrom et al. (1983). It was encouraging to note that measures with ET and EOG were quite parallel and of approximately equal sensitivity.

Experiment II

In light of the results of Experiment I, it would seem that more extensive manipulation of workload, together with less intrusive measures, might be required to reveal useful eye movement indicants of mental workload. For this reason we carried out Experiment II, which entailed measuring the spatial extent of spontaneous saccades during free viewing in a dimly lighted room of low mesopic levels and more demanding tone counting conditions.

Subjects - Five subjects (four also participated in the first experiment) were used. There were two males and three females and their ages ranged from 24 to 45 years.

Apparatus - The infrared eyetracking instrument was used to record eye movements from the left eye. These signals were applied to the modulation input of a voltage-controlled frequency generator (Wavetek, Model 148), the output of which was fed into the signal processor (Nicolet, Model 1072), which was programmed to accumulate a time-interval histogram. In this fashion eye-movement extent was coded in terms of frequency modulation and depicted as two adjacent frequency histograms -- one for leftward eye movements and one for rightward eye movements. The resultant histograms were plotted on an X-Y plotter (Hewlett-Packard, Model 7044A).

Tone-counting tasks were again administered with the microprocessor (NEC, Model 8201A) which was programmed to present a random series of 36 low tones, 28 medium tones, and 24 high tones. Tone durations were .5 seconds and the same temporal distribution was repeated every 60 seconds, but the subjects did not identify a pattern. Responses were entered and cataloged on the microprocessor and scoring included the number correct, incorrect, and missed. Three tone-counting tasks were used. Task one required a response after each fourth low tone (low task load), task two required a response after each fourth low tone and each fourth middle tone (medium task load), counted separately and kept track of separately. Task three required a response after each fourth low, medium, and high tone (high task load). Three separate keys were used to indicate the three different tone counts. Scoring was always reset in the event of a miss or an incorrect response.

Procedure - Each subject was allowed one practice run on each tone-counting task prior to data collection. Each session began with a fixation condition in which the subject fixated a small cross (subtending 10 min. of visual angle) for five minutes during which eye-movements were recorded. Next, they performed an alternating fixation task which required 20 degree saccades at an aperiodic rate (0.2 Hz) for five minutes while eye movements were recorded. Following this they were allowed to move their eyes freely for five minutes during which eye movements were recorded. After these baseline conditions, they were asked to perform the one-, two-, and three-channel counting tasks under free viewing for five minutes each while eye movements were being recorded.

RESULTS

Performance Data

Presented in Table 1 are the percent correct performance scores on the counting task as a function of workload (number of channels monitored). As can be seen, under the low workload condition (1 channel monitored) performance was nearly perfect (96%) whereas under high workload conditions a substantial percentage of errors were made ($F(1,4) = 9.10$, $p = .0393$, for the linear component).

TABLE 1. MEANS AND STANDARD DEVIATIONS OF SPONTANEOUS SACCADIC LENGTH AND PERFORMANCE SCORES (PERCENT CORRECT) AS FUNCTIONS OF LOW (1 CHANNEL), MEDIUM (2 CHANNELS), AND HIGH (3 CHANNELS) LEVELS OF WORKLOAD

		Low	Medium	High
Saccade Length	Mean	3.25	3.01	2.44
	SD	(2.30)	(3.08)	(2.32)
Performance (% correct)	Mean	0.96	0.82	0.64
	SD	(0.09)	(0.19)	(0.23)

Eye Movement Data

The histograms obtained under all six experimental conditions for a single subject are presented in Figure 4. The results for the other four subjects were similar and are omitted. It may be seen that under conditions of steady fixation (Panel A) the distribution of frequency modulation was quite narrow, indicating that the extent of leftward or rightward movement was quite small. Under conditions of 20 degree alternate fixation (Panel B), the distribution of frequency modulation is bimodal, indicating that the extent of leftward and rightward eye movements was quite extensive. These data were used to calibrate the abscissa (saccade length) in degrees of visual angle. Under conditions of free viewing (Panel C), the distribution of frequency modulation was intermediate between fixation and alternating fixation, indicating that the range of eye movements during this condition fell somewhere between steady fixation and saccades of 20 degrees. The effects of tone counting are depicted in Panels D through F for one-, two-, and three-channel counting. It is evident that as task load increased, the extent of frequency modulation decreased. Thus, the index of interest is a measure of the extent of eye movements under these different conditions of workload. For this purpose, the range of the histogram was computed and transformed to degrees of saccade and was further normalized by dividing by the range of saccades under fixation. This was done because there were substantial overall differences in both fixated and spontaneous eye movements. Average normalized spontaneous saccadic length as a function of workload is presented in Table 1 where it is clear that saccadic length decreased as a function of workload ($F\ (1,4) = 16.65$, $p = .0151$, for the linear component). To further substantiate the relation of saccadic length to workload, correlation coefficients between saccadic length and performance were computed for each subject, which averaged $r = .64$, and ranged from .37 to .99.

Discussion

It is clear from the performance data of both experiments that the modified Jerison counting task offers considerable control of task workload and provides an excellent behavioral index of that parameter. Although the performance scores varied on average from 64% to 96% in the present study, this technique can be made more difficult by the inclusion of more tone categories and increased rate of tone presentation to expand the range of workloads investigated. Such a manipulation might well improve the correlation obtained between saccadic and behavioral measures of workload. It would even be possible to empirically adjust the difficulty level of the task based on previous performances. This could then be employed to create a task of empirically determined isodifficulties for all subjects which in equivalently motivated subjects would imply equal workload and performance.

The results of Experiment I indicated that eye movement velocity during alternate fixation did not vary significantly when task difficulty was increased. Although the latency measures did increase with increased workload, this effect was confounded with large practice effects and is, therefore, an equivocal candidate for an objective index of workload. Additionally, measures derived from such a paradigm, which requires controlled eye movements (alternate fixation) cannot easily be obtained in most real-world activities.

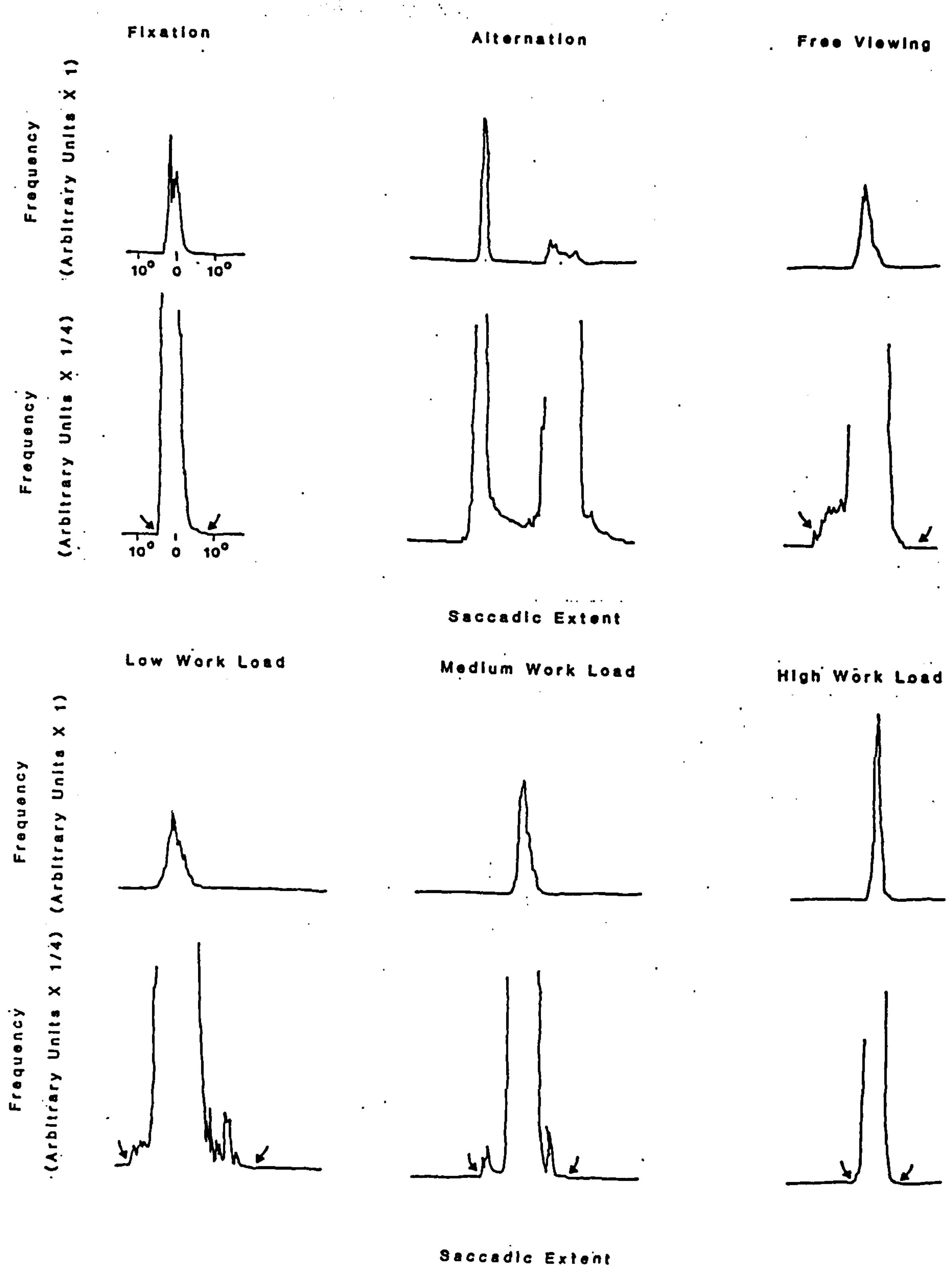

Figure 4. Histogram for six conditions for one subject.

The results of Experiment II are much more encouraging in that the extent of spontaneous saccades was significantly restricted as task difficulty increased. These measures could be obtained easily in many situations that require dynamic information processing, but a number of potential problems must be addressed. Because only a small number of subjects were used in the present experiment, and only a rather primitive index (the range) of saccadic extent was employed, future studies should address this relationship with a larger number of subjects, and more sophisticated measures of saccade distance. The extent to which these procedures might be used in situations where visual information is to be processed is another important consideration. It may be that a decrease in saccadic extent is also observed with increased workload in situations where visual monitoring of events is necessary as the data of Hall, (1976) and Malmstrom, Randle, Murphy, Reed, & Weber (1983) imply. The degree to which this relationship exists may depend on whether the primary visual task requires precise fixation or tracking performance, but many visual activities do not. Clearly we should reexamine this relationship with a VISUAL monitoring (counting) task that is analogous to the previously used auditory counting task.

If the extent of spontaneous saccades is a sensitive index of workload, then the decrease in mental effort which derives from repeated practice should be associated with an increase in the extent of spontaneous saccades. If this were the case, then this measure may provide an indirect index of the degree to which a task has become automatic (Ackerman & Schneider, 1984) and might provide a sensitive measure of individual differences with potential application to personnel selection and training. One of the reasons that the counting test was selected was that we knew it would not improve much with extended practice (Kennedy & Bittner, 1980), but that is not the case with most other performance measures (cf., Newell & Rosenbloom, 1984, for a review). In many cases the workload "rating" that a task possesses can be expected to change as the task is practiced. These relations should be studied.

The techniques employed in the present investigation included two serious limitations. First, the eye tracking apparatus was insensitive to eye movements in non-horizontal meridians. It may be possible to use similar techniques with instruments that track vertical as well as horizontal eye movement signals. Such procedures may provide a more sensitive index of workload. Although we do recognize that the separate innervation of the extraocular eye muscles can result in vertical horizontal differences in eye movement behavior (Guedry & Benson, 1971), it is not anticpated that there would be interactions between eye movement direction and workload, but this should be examined. Second, the apparatus employed in the second experiment did not allow the exclusion of eye blinks. Future work which parcels out these events might also provide improvement in the sensitivity of this method.

REFERENCES

Ackerman, P. L., & Schneider, W. (1984). Individual differences in automatic and controlled information processing (Report No. HARL-ONR-8401). Champaign-Urbana, IL: University of Illinois, Department of Psychology.

AGARD (1977). Methods to assess workload. Proceedings of the Aerospace Medical Panel Symposium (AGARD-CP-216). Neuilly-sur-Seine (France): Advisory Group for Aerospace Research and Development.

AGARD. (1978). Methods to assess workload AGARD Conference Proceedings No. 216. Neuilly-sur-Seine, France: NATO Advisory Group for Aerospace Research & Development.

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

Bird, K. L. (1981). Subjective rating scales as a workload assessment technique. In Proceedings of the 17th Annual Conference on Manual Control, Los Angeles, CA.

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

Borg, G., Bratfisch, O., & Dornic, S. (1971). Perceived difficulty of a visual search task (Report No. 16). Stockholm: Institute of Applied Psychology.

Borg, G., Bratfisch, O., & Dornic, S. (1972). On the problem of perceived difficulty. Scandinavian Journal of Psychology 12, 249-260.

Bratfisch, O., Borg, G., & Dornic, S. (1972). Perceived item difficulty in three tests of intellectual performance capacity (Report No. 29). Stockholm: Institute of Applied Psychology.

Carpenter, J. H. S. (1977). Movements of the eyes. London: Pion.

Chiles, W. D., & Alluisi, E. A. (1979). On the specification of operator or occupational workload with performance-measurement methods. Human Factors, 21, 515-528.

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.

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

Cooper, G. E., & Harper, R. P. (1969). The use of pilot ratings in the evaluation of aircraft handling qualities (NASA Ames Technical Report NASA TN-D-5153). Moffett Field, CA: NASA Ames Research Center.

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). The effects of extended practice on dual-task training. Human Factors, 23, 627-631.

Daniel, R. S. (1966). Electroencephalographic pattern quantification and the arousal continuum. Psychophysiology, 2(2), 146-160.

Daroff, R. B., & Hoyt, W. F. (1971). In Bach-y-Rita and Collins (Eds.). Symposium on the Control of Eye Movements. New York: Academic Press.

Ditchburn, R. W. (1973). Eye-movements and visual perception. New York: Oxford University Press.

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.

Frazier, M. L., & Crombie, R. B. (Eds.). (1982). Proceedings of the Workshop on Flight Testing to Identify Pilot Workload and Pilot Dynamics, 19-21 January 1982 (AFFTC-TR-82-5). Edwards AFB, CA: Air Force Flight Test Center.

Gevins, A. S., Bressler, S. L., Cutillo, B. A., Doyle, J. C., Morgan, N. H., & Zeitlin, G. M. (1984). Neurocognitive pattern analysis of an auditory and visual numeric motor control task (Contract No. F4962082-K-0006). San Francisco: EEG Systems Laboratory.

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

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

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

Guedry, F. E. (1965). Psychophysiological studies of vestibular function. In Neff (Ed.). Contributions to sensory physiology (Vol. 1). New York: Academic Press, 63-135.

Guedry, F. E., & Benson, A. J. (1971). Nystagmus and visual performance during sinusoidal stimulation of the vertical semicircular canals (NAMRL-1131, USAARL 71-16). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Hall, R. J. (1976). Eye movement fixations and gating processes. In R. A. Monty and J. W. Hall (Eds.), Eye movements and psychological processes (Pp. 491-498). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Hoffman, J. E., Houck, M. R., MacMillan III, F. W., Simons, R. F., & Oatman, L. C. (1985). Event-related potentials elicited by automatic targets: A dual-task analysis. Journal of Experimental Psychology: Human Perception and Performance, 11(1), 50-61.

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.

Johannsen, G., Moray, N., Pew, R., Rasmussen, J., Sanders, A., & Wickens, C. (1979). First steps towards a designer's mental workload checklist. In N. Moray (Ed.), Mental workload: Theory and measurement. New York: Plenum.

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

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

Kaufman, L., & Williamson, S. J. (1983). Neuromagnetic investigation of workload and attention (AFOSR-TR-33-0901). Bolling AFB, DC: Air Force Office of Scientific Research. (AD A136172)

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., & Bittner, A.C., Jr. (1980). Development of Performance Evaluation Tests for Environmental Research (PETER): Complex counting. Aviation, Space, and Environmental Medicine, 51, 142-144.

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

Lewis, G. W. (1982). Event related brain electrical and magnetic activity: Toward predicting on-job performance. International Journal of Neuroscience, 18, 159-182.

Lewis, G. W. (1983a). Bioelectric predictors of personnel performance: A review of relevant research at the Navy Personnel Research and Development Center (NPRDC TR-84-3). San Diego, CA: Navy Personnel Research and Development Center.

Lewis, G. W. (1983b). Bioelectric and biomagnetic predictors of military performance: Relevance for aviation personnel. Paper presented at the 1983 Fall Technical Meeting of the Tri-Service Aeromedical Research Panel, Wright-Patterson AFB, OH.

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

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.

Moray, N. (1967). Where is attention limited? A survey and a model. Acta Psychologica, 27, 84-92.

Moray, N. (Ed.). (1979). Mental workload: Its theory and measurement. New York: Plenum Press.

Moray, N. (1982). Subjective mental workload. Human Factors, 24, 25-40.

Newell, A., & Rosenbloom, P. S. (1980). Mechanisms of skill acquisition and the law of practice (CMU-CS-80-145). Pittsburgh, PA: Carnegie-Mellon University.

O'Donnell, R. D. (1979). Contributions of psychophysiological techniques to aircraft design and other operational problems (AGARDograph AG-244). NATO Advisory Group for Aerospace Research and Development.

O'Hanlon, J. F. (1971). Heart rate variability: A new index of driver alertness/fatigue (Technical Report 1712-1). Goleta, CA: Human Factors Research Inc.

Patten, B. M. (1951). Early embryology of the chick. New York: McGraw-Hill Book Company, Inc.

Reid, G. B. (1982). Subjective workload assessment: A conjoint scaling approach. Preprints of 1982 Annual Scientific Meeting, Aerospace Medical Association, Washington, DC: Aerospace Medical Association, 153-154.

Reid, G. B., Shingledecker, C., & Eggemeier, T. (1981). Application of conjoint measurement to workload scale development. In R. Sugarman (Ed.), Proceedings of the 25th Annual Meeting of the Human Factors Society, Santa Monica, CA.

Reid, G. B., Shingledecker, C. A., Nygren, T. E., & Eggemeier, F. T. (1981). Development of multidimensional subjective measures of workload. Atlanta, GA: IEEE Systems, Man & Cybernetics Society.

Schmidt, D. K. (1978). A queing analysis of the air traffic controllers' workload. IEEE Transactions on Systems, Man, and Cybernetics, SMC-8, 492-498.

Senders, J. W., Fisher, D. F., Monty, R. A. (Eds.) (1979). Eye movement in the higher psychological function. Hillsdale, NJ: Erlbaum Associates.

Sheridan, T. B., & Simpson, R. W. (1979). Toward the definition and measurement of the mental workload of transport pilots (FTL Report R 79-4). Cambridge, MA: Massachusetts Institute of Technology, Flight Transportation Laboratory.

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.

Stevens, S. S. (1957). On the psychophysical law. Psychological Review, 64, 153-181.

Tulga, M. K. (1978). Dynamic decision making in multitask supervisory control: Comparison of an optimal algorithm to human behavior (Man-Machine Systems Report). Cambridge, MA: M.I.T. Man-Machine Systems Laboratory.

Verplank, W. L. (1977). Is there an optimal workload in manual control? Unpublished doctoral dissertation, M.I.T., Cambridge, MA.

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

Wickens, C. D. (1984). Engineering psychology and human performance. Columbus, OH: Charles E. Merrill Publishing Co.

Wickens, C. D., Sandry, D., & Hightower, R. (1982). Display location of verbal and spatial material: The joint effects of task hemispheric integrity and processing strategy (Technical Report EPL-82-2/ONR-82-2). Champaign, IL: University of Illinois, Engineering Psychology Research Laboratory.

Wierwille, W. W. (1979). Physiological measures of aircrew mental workload. Human Factors, 21, 575-594.

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

Wierwille, W. W., & Williges, R. C. (1978). Survey and analysis of operator workload assessment techniques (Report No. S-78-101). Blacksburg, VA: Systemetrics, Inc.

Wierwille, W. W., & Williges, B. H. (1980). An annotated bibliography on operator mental workload assessment (Report SY-27R-80). Patuxent River, MD: Naval Air Test Center.

Williges, R. C., & Wierwille, W. W. (1979). Behavioral measures of aircrew mental workload. Human Factors, 21, 549-574.

Wilson, G. F., O'Donnell, R. D., & Wilson, L. (1982). Neurophysiological measures of A-10 workload during simulated low altitude missions (Report No. AFAMRL TR-83-0003). Wright-Patterson AFB, OH: AFAMRL, Human Engineering Division.

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