AD-A167 729

CAORF-16-8122-02

TECHNICAL REPORT
SIMULATION EXPERIMENT

A SIMULATION STUDY OF THE EFFECTS OF SLEEP DEPRIVATION, TIME OF WATCH, AND LENGTH OF TIME ON WATCH ON WATCHSTANDING EFFECTIVENESS

DEPARTMENT OF TRANSPORTATION ★ UNITED STATES OF AMERICA ★

U.S. DEPARTMENT OF TRANSPORTATION

MARITIME ADMINISTRATION
OFFICE OF SHIPBUILDING, OPERATIONS, AND RESEARCH

NATIONAL MARITIME RESEARCH CENTER
KINGS POINT, NEW YORK 11024-1699

JANUARY 1986

LEGAL NOTICE

AD-A167729

BIBLIOGRAPHIC DATA SHEET	1. Report No.	2.	3. Recipient's Accession No.

4. Title and Subtitle
A Simulation Study of the Effects of Sleep Deprivation, Time of Watch, and Length of Time on Watch on Watchstanding Effectiveness

5. Report Date January 1986

6.

7. Author(s)
Anita D. D'Amico, Ph.D., Edmund Kaufman, and Christine Saxe

8. Performing Organization Report No. CAORF 16-8122-02

9. Performing Organization Name and Address
Computer Aided Operations Research Facility (CAORF)
National Maritime Research Center, Kings Point, New York 11024

10. Project/Task/Work Unit No.

11. Contract/Grant No.

12. Sponsoring Organization Name and Address
Office of Shipbuilding, Operations, and Research
Maritime Administration
U.S. Dept. of Transportation
Washington, D.C. 20590

13. Type of Report & Period Covered CAORF Simulation Experiment

14.

15. Supplementary Notes

16. Abstracts

The primary purpose of this research was to measure the effects of sleep deprivation, time of watch, and length of time on watch on a watchstander's work performance and on physiological measures in a simulated open-sea watch. A second purpose was to evaluate marine simulation as a method for studying the effects of such variables on mariner performance.

Sleep deprivation is a common problem among mariners and was the primary variable of interest. Time of watch and length of time on watch were examined because of their interrelationship with the fatiguing effects of sleep deprivation. The standard work schedule of watch officers — four hours on and eight hours off work — is discordant with the typical work/rest schedules and accompanying circadian rhythm of most people. Thus, this watch schedule was expected to mediate the effects of sleep deprivation on watch standing. The length of time which a watch officer has stood a relatively uneventful watch was also expected to interact with the effects of sleep deprivation, such that the deleterous effects of sleep deprivation would be exacerbated as time on watch increased.

(Continued on following page)

17. Key Words and Document Analysis. 17a. Descriptors

Boredom
CAORF
Circadium Rhythm
Fatigue
Monotony
Simulation
Sleep Deprivation
Vigilance
Watch Schedule
Watchstanding

17b Identifiers/Open-Ended Items

17c. COSATI Field/Group

18. Availability Statement	19. Security Classification (This Report)	21. No. of Pages
Approved for Release NTIS Springfield, Virginia	UNCLASSIFIED	
	20. Security Classification (This Page) UNCLASSIFIED	**22. Price** 142

Cont'd

16. Abstracts (Cont'd)

Twenty-five watchstanding mates were assigned to one of four experimental groups: Group 1 slept for 7.5 hours, then stood a morning watch; Group 2 slept for 7.5 hours, then stood an afternoon watch; Group 3 received no sleep before standing a morning watch; Group 4 received no sleep before standing an afternoon watch. All watches were stood on the bridge of a full-scale full-mission shiphandling simulator. Dependent measures were collected before, during and after a four-hour watch. Dependent measures included speed of traffic ship detection, radar observation, and measures of effect, cognition, and physiological state.

Sleep deprivation appeared to have a deleterious effect on watchstanding vigilance, especially for those mates standing the afternoon watch. Conclusions were based on many weak indicators, however, rather than a few very strong effects.

CAORF 16-8122-02

TECHNICAL REPORT
SIMULATION EXPERIMENT

A SIMULATION STUDY OF THE EFFECTS OF SLEEP DEPRIVATION, TIME OF WATCH, AND LENGTH OF TIME ON WATCH ON WATCHSTANDING EFFECTIVENESS

By

Anita D. D'Amico, Ph.D.
Edmund Kaufman
Christine Saxe
CAORF Research Staff

U.S. DEPARTMENT OF TRANSPORTATION

MARITIME ADMINISTRATION
OFFICE OF SHIPBUILDING, OPERATIONS, AND RESEARCH

NATIONAL MARITIME RESEARCH CENTER
KINGS POINT, NEW YORK 11024-1699

JANUARY 1986

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

INTRODUCTION

The purpose of this research was to measure the effects of sleep deprivation, time of watch, and number of hours worked on a watchstander's work performance and on cognitive and physiological measures in a simulated open-sea watch.

Sleep Deprivation — Despite the fact that human error, in whole or in part, accounts for some eighty percent of all marine casualties, it has never been studied systematically. Specifically, fatigue has been targeted as a primary cause of errors of judgement that lead to marine accidents.

Fatigue, as a result of insufficient rest, is a common problem among mariners. As such, it has been the recommendation of both the Panel on Human Error and the National Transportation Safety Board that the effects of fatigue and work schedules be investigated.

Time of Watch — The standard work schedules of the watch officer, four hours on and eight hours off, is discordant with the typical work/rest schedules that most people follow. For most people, work begins in the morning, is continuous for approximately eight hours, and is followed by a long period of rest and sleep. This type of schedule more closely matches the cyclical fluctuations in body temperature known as circadian rhythm. Low points in the circadian cycle are usually correlated with degraded attention and work performance. Because watchstanding occurs around the clock, time of watch was studied in conjunction with fatigue as a contributory factor to human error.

Number of Hours Worked — In addition to problems associated with fatigue and ebbing circadian rhythm, the effects of boredom on watchstanding performance were also examined in this experiment. Boredom is usually a problem that is inherent in any long, monotous task. In the present study, boredom was considered as a function of the number of hours on watch.

OBJECTIVES

The primary purpose of this research was to measure the effects of sleep deprivation, time of watch, and number of hours worked on a watchstander's work performance, mood, cognitive capacities, and physiological states under the conditions of a mate standing an open-sea watch.

A second purpose was to evaluate marine simulation as a method for studying the effects of sleep deprivation, watch schedule, and number of hours worked on mariner performance. This evaluation incorporated the use of performance measures previously used in non-maritime industry research as well as performance measures specific to the maritime industry.

METHODOLOGY

Subjects were 25 experienced second mates. They were randomly assigned to four groups:

	Sleep Condition	Time of Watch
Group 1	7½ hours	0800-1200
Group 2	7½ hours	1200-1600
Group 3	0 hours	0800-1200
Group 4	0 hours	1200-1600

Three independent variables were tested:

Between Subjects
- Amount of Sleep (7½ vs. 0 hours)
- Time of Watch (0800-1200 vs. 1200 to 1600)

Within Subjects
- Number of Hours on Watch (1 vs. 2 vs. 3 vs. 4 hours)

Several dependent variables were assessed as a function of the independent variables:

Work Performance
- Speed with which three traffic ships were sighted
- Speed with which one instrument failure was sighted
- Frequency and duration of radar observations

Measures of Affect
- NPRU Mood Scales
- Stanford Subjective Sleepiness Scale (SSS)
- Profile of Mood States (POMS)

Cognitive Measures
- Visual Search and Scan Test (VSS)
- Digit Span Test
- Baddeley Grammatical Transformation Test

Physiological Measures
- Cardiac Activity
- Muscle Tension
- Physical Activity
- Oral Temperature

Subjects arrived one day prior to the experimental runs and received three hours of familiarization on the CAORF bridge the first day. Pertinent biographical data was collected and personality tests were administered. During the first evening, subjects were monitored closely. No alcoholic beverages were allowed, and physical activity was kept to a minimum. Half the subjects retired to bed at 11 PM while the other half were kept awake all night. The next day, all subjects stood a four-hour, open-sea watch during which three ships appeared and one instrument failure occurred. Dependent measures were collected before, during, and after the watch period. Subjective responses to the experiment were ascertained in a post-experimental interview.

RESULTS

Post-Experimental Interview – The post-experimental debriefing revealed that the simulation appeared realistic in most ways. Only the lack of distractions such as auxiliary paper work or a helmsman to talk to were missing. The experimental manipulations were effective; that is, the fatigued group reported less alertness and, in fact, fell asleep far more frequently than the well-rested group. Subject felt as though their performance degenerated over the course of the watch. Daydreaming was frequently used as a means of alleviating boredom; mindblanking was also used, but with less frequency. Overall the simulated watch appeared to be sufficiently realistic to make the results generalizable to the real world.

Biographical Data and Personality Profiles – With respect to biographical information and personal characteristics, the experimental groups were highly similar, except for age and competitiveness. There was a significant interaction effect on age, and a significant effect for each condition on the Jenkin H. Scale, a measure of one's tendency to be hard-driving and competitive; subjects in the afternoon watch group were more hard-driving and competitive than those in the morning watch condition.

Detection of Ships and Instrument Failure – Four incidents occurred during the watch; three ships appeared on the horizon and the RPM indicator failed. For the first ship sighting, although the differences were nonsignificant, the means for the experimental groups are ordered such that the fatigued groups took longer to detect the first ship than the well-rested groups; the no sleep/afternoon group, in particular, performed the worst. For the second ship sighting, a weak effect for watch condition was found indicating that afternoon watchkeepers were slower to detect the destroyer than the morning watchkeepers. While there were no significant differences found between groups for time to detect the RPM indicator failure, again the no sleep/afternoon group took the longest. Finally, no main effects were observed for detection time of the last ship.

Detection performance over the entire watch was analyzed for the three traffic ships. The results indicate that the full sleep groups took significantly less time overall to detect the traffic ships than did the no-sleep groups. There was also a trend for the morning watch groups to take less time than the afternoon groups.

Frequency and Duration of Radar Observations – Perhaps the single most important function a watch officer performs is to maintain vigilance with respect to the possibility of collision with other ships. Since the utilization of radar is vital in the performance of this function, the effects of fatigue on radar usage are considered a significant measure of mariner performance. To assess this, the CAORF radar was configured so that the mate had to hold down a button in order to use it. In this way, the frequency and duration of radar use could be precisely determined.

Radar activity not specifically associated with the detection of a particular incident increased as the watch progressed, irrespective of sleep or watch condition. Radar activity associated with the detection of a ship also increased during the watch with the exception of the fatigued afternoon group, who appeared to consult the radar less by the end of their watch, although the duration of radar consultations did not change.

Psychological Measures – Measures of mood, subjective fatigue and cognition were administered to each watchstander before and after his watch. Some measures were also administered at mid-watch.

The NPRU Positive and Negative Affect Scale showed significant main effects for sleep condition. The well-rested groups tended to report more positive and less negative feelings than the no-sleep groups. The Negative Affect Scale also revealed significant main effects for watch condition indicating that the afternoon watchstanders experienced more negative affect than the morning watchstanders.

Main effects for sleep condition were also found in the Stanford Subjective Sleepiness Scale; subjects in the no sleep condition were significantly more fatigued than those who had slept.

Two significant results were found in the Profile of Mood States Checklist; the fatigue and vigor scales showed strong main effects for sleep condition. The anger and confusion scales yielded interesting results.

The visual Search and Scan test yielded significant results for sleep condition and watch condition, with the no-sleep and afternoon groups performing more poorly when measured at the beginning of the watch. In some instances, time of testing also influenced results.

Results of the Digit Span test were influenced only by time of testing, and no significant results were found with the Baddeley Grammatical Transformation Test.

Physiological Measures — Subjects' overall muscle tension (EMG) and gross body movements were recorded. Both the 16.5 minutes and the 5.5 minute periods prior to the four attention-demanding tasks were examined. The data suggests that the afternoon watchkeepers were more tense than the morning watchkeepers throughout the watch when measured for 16.5 minutes prior to each task. When measured 5.5 minutes prior to each event, the fatigued group showed more tension than the well-rested group, although these differences were nonsignificant. The well-rested group showed more tension during the last hour of the watch in comparison to the first hour; the fatigued/afternoon group showed the opposite effect.

During the time of actual detection, the fatigued groups appeared to exhibit more muscle tension than the well rested groups. Overall, muscle tension increased as the watch progressed.

Comparisons of EMG level during detection to immediately before detection indicated that during the second hour, the morning watchstanders decreased their tension once a sighting was made while the afternoon watchstanders remained as tense during the sighting as before. This was also observed during the third hour when the no sleep/morning group showed a remarkable decline in muscle tension once a sighting was made.

The afternoon watchstanders tended to have a higher level of EMG arousal after detection rather than before, which may indicate a slow recovery from arousing events. The morning watchstander, on the other hand, recovered quickly, showing a decrease in muscle tension following a sighting.

CONCLUSIONS

The authors concluded that fatigue had a deleterious effect on the vigilance of watchstanders, especially those standing afternoon (as opposed to morning) watches. These conclusions were based on many weak indicators, rather than a few very strong effects.

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Virtually all surveys of ship casualties conclude that approximately 80% have resulted, at least in part, from human error. Yet the factors which predispose ship officers to commit errors have not been the subject of much systematic research. The research described in this paper was an initial effort to study the role of personnel fatigue and work/rest cycles as contributing factors to human error aboard ships. It is a commonly held belief among seamen that fatigue is a significant factor contributing to at-sea casualties (Panel on Human Error in Merchant Marine Safety, 1976); yet human factors such as fatigue have not received the research attention that machine-related factors have heretofore received. Few experiments have been conducted in the maritime industry to analyze which tasks are affected when performed by fatigued personnel and how much fatigue is necessary to produce deleterious effects on performance at sea. The Panel on Human Error in Merchant Marine Safety recommended in their 1976 report that the following action be taken: "The Maritime Administration should conduct research into fatigue, to include effects of duty cycles for specific tasks, physiological day/night cycles, and chronic or long term fatigue."

A review of casualty investigations further substantiates the need for research into the areas outlined by the Panel. For example, on September 21, 1972, at 0140 hours, the TUG CAROLYN and WEEKS BARGE 254 collided with a trestle of the Chesapeake Bay Bridge and Tunnel leading to an economic loss conservatively estimated at $3,397,000. The United States Coast Guard/National Transportation Safety Board (NTSB) Marine Casualty Report (1974) determined one of the probable causes to be incorrect decision-making by the master of the CAROLYN. The report states that ". . . the master's inability to respond adequately to the changing situation was heightened by the fact that the master had had no rest since before 0600 hours on September 20. These conditions may have led the master to take certain actions not compatible with those which he might have reasonably been expected to take." As a result of this collision, the NTSB recommended to the U.S. Coast Guard that it "determine the effects of fatigue on personnel error as a cause of marine casualties. . . ."

More recently, the collision in the Neches River on February 25, 1979, of the MARINE DUVAL, carrying 23,000 long tons of molten sulphur, with the MOBIL VIGILANT, carrying 41,000 tons of crude oil, resulted in damages estimated at over $6,000,000. In light of the lack of vigilance by the pilots involved in this accident, the NTSB recommended that they "review pilot rotation policies relative to vessel movements and avoid assigning pilots to two consecutive long trips without adequate rest between such assignments."

The study reported herein was designed as a first step in response to the recommendations of the Panel on Human Error and the NTSB regarding research on the effects of fatigue on mariner performance. This initial study focused on the effects of fatigue and related factors on the behavior of one particular group of mariners – watchstanding officers. Since this research is the first fatigue investigation conducted on a marine simulator, it also served the purpose of testing various measures of performance for sensitivity to fatigue under simulated work conditions.

1.2 WATCHSTANDING

The safety of a ship's passage is determined in part by the vigilance of its watch officers. The watch officer is the captain's representative and "his primary responsibility at all times is the safe navigation of the vessel" (Bridge Procedures Guide, 1977). The role of the watch officer on modern ships does not differ substantially from that of his predecessors of 1,000 years ago; that is, he must look out for other ships.

According to the Bridge Procedures Guide, written by the International Chambers of Shipping (1977), some of the duties of the officer of the watch are as follows:

> "The officer of the watch should keep his watch on the bridge which he should in no circumstances leave until

properly relieved. A prime responsibility of the officer of the watch is to ensure the effectiveness of the navigating watch. It is of especial importance that at all times the officer of the watch ensures that an efficient lookout is maintained. In a vessel with a separate chartroom, the officer of the watch may visit this, when essential, for a short period for the necessary performance of his navigational duties, but he should previously satisfy himself that it is safe to do so and ensure that a good lookout is kept. . . .

The officer of the watch is responsible for the maintenance of a continuous and alert lookout. This is the most important consideration in the avoidance of casualties. The keeping of an efficient lookout requires (that it) be interpreted in its fullest sense which includes the following: (a) an alert all-round visual and aural lookout to ensure a full grasp of the current situation including the presence of ships and landmarks in the vicinity; (b) close observation of the movements and compass bearing of approaching vessels; (c) identification of ship and shore lights; (d) the need to ensure that the course is steered accurately and that wheel orders are correctly executed; (e) observation of the radar and echo sounder displays; (f) observation of changes in the weather, especially the visibility.

The officer of the watch should bear in mind that the engines are at his disposal. He should not hesitate to use them in case of need. However, timely notice of engine movements should be given when possible. He should also keep in mind the maneuvering capabilities of his ship, including its stopping distance."

While the above description is not an exhaustive list of the responsibilities of a watch officer, it does convey the necessity for vigilance in the performance of his duties.

The term "watch" itself is derived from the Anglo-Saxon "wacion" meaning to be vigilant or to be awake. The earliest uses of the word "watch" connoted continuing without sleep or keeping vigil (McEwen and Lewis, 1953). Remnants of this earlier meaning pervade the modern maritime industry. Watch officers frequently perform their duties with little or no sleep in the previous 24 hours. Watchkeeping without sleep may be self-imposed, as in the case of the seaman returning from a wild night in port, or it may be imposed by the nature of the work, as in the case of discharging or loading cargo.

A watch period is traditionally defined as four hours. A commerce law, enacted on March 4, 1915, requires all U.S. merchant vessels of greater than 100 gross tons (excepting those vessels which exclusively navigate rivers, harbors, bays, sounds, bayous, canals, and lakes other than the Great Lakes) to use a "three-watch system" while at sea. In this system, the watch officer stands watch for four hours, rests for eight hours, and stands watch for another four hours.

The traditional names assigned to the watch periods are:

Afternoon watch:	Noon to 4 PM (1200 to 1600)
First dog watch:	4 PM to 6 PM (1600 to 1800)
Second dog watch:	6 PM to 8 PM (1800 to 2000)
First watch:	8 PM to midnight (2000 to 2400)
Middle, or graveyard, watch:	Midnight to 4 AM (0000 to 0400)
Morning watch:	4 AM to 8 AM (0400 to 0800)
Forenoon watch:	8 AM to noon (0800 to 1200)

In the three-watch system, the first and second dog watches are combined to form one watch. This 4 to 8 watch is virtually always interrupted by suppertime in the evening and by the sunrise in the morning. Because of this interruption in the boredom of the watch, as well as the afternoon leisure time which is available, this watch is highly desirable. The 0800 to 1200 is the next most desirable watch, since the midnight to 8 AM rest period closely approximates on-shore sleep time. The 0000 to 0400 watch is considered the least desirable because of the lack of activity and companionship associated with it, the total darkness of the graveyard watch, and the physical adjustment the officer must make in working from midnight to 4 AM.

The work/rest schedules associated with the three-watch system require watchstanders to begin their second four-hour watch cycle exactly 12 hours after their first watch, a schedule which is unlike most shoreside jobs. The eight hours of time between watch cycles do not provide sufficient time for the mate to both get a full eight hours of sleep and attend to his personal needs. Thus, watchstanders rarely follow the typical work/rest cycles engaged in by most people.

1.3 PROBLEMS INHERENT IN WATCHSTANDING

The most profitable shipping company is one which keeps its ships at sea as much as possible. To accomplish this, modern ships spend as little time as possible in port, keeping "turnaround" periods to a bare minimum. While this schedule serves the needs of the ship owners and operators, it also intensifies some of the problems of watch officers; namely: fatigue, boredom and ebbing circadian rhythms. Each of these is discussed below.

With little time in port to load and unload cargo, the ship's officers may work long, continuous hours in cargo handling and then go ashore for several hours of entertainment. The turnaround time in port is so short, however, that watch officers frequently assume their sea watches without recuperative rest after port time. It is not uncommon for a tired watch officer to keep a cigarette continuously burning between his fingers so that, should he fall asleep while standing watch, the scorching of his fingers will awaken him. The casualty reports reviewed earlier clearly indicate the potential disasters which may result from human error related to such fatigue.

The short turnaround time also maximizes the amount of time spent in the open sea. Weeks of performing a task which requires staring at the horizon leads to one of the most pervasive problems among seamen – boredom. Their verbal reports indicate that boredom increases with the amount of time spent at sea, and with the amount of time which has passed during a four-four watch. Despite the prevalence of boredom on watch, few efforts have been made to evaluate its effects on watchkeeping on the bridge of a ship.

Boredom is a difficult construct to quantify. Much of the research on boredom infers its existence from the length of time during which external stimulation has been lacking. Some of this research suggests that boredom results in slower reaction times to detecting signals on a radar (Thackray, Bailey and Touchstone, 1977). Since this research was conducted using air traffic controllers on a simulated radar task, the generalizability of these findings to watchkeeping on a ship's bridge is questionable. Air traffic controllers must respond rapidly to the appearance of a plane on their screens, whereas watch officers may note the appearance of a ship on radar and wait several minutes or hours before taking action, if indeed any action other than logging the incident is necessary. However, if watchstanders at sea were to show decrements in their performance as a result of boredom similar to those of the air traffic controllers, then the lack of stimulation at sea would be of special operational significance.

Interviews conducted by this author and others on board ships as well as articles on the maritime trade journals (e.g., Safety at Sea, 1981) indicate that the monotony of watchkeeping is considered a negative aspect of the job. Some watchkeepers use daydreaming, star gazing, conversing with anyone who happens to walk through the bridge, playing mental games – virtually anything to divert their attention away from the watchkeeping task – to alleviate the boredom inherent in this task. One can reasonably assume that boredom must have some deleterious effect on vigilance, and that boredom in all likelihood increases with the number of hours one must stand watch alone and, perhaps, with the degree of fatigue.

The "four on-eight off" schedule of watchkeeping also presents problems. The graveyard watch (0000 to 0400) is considered least desirable, while the 0400 to 0800 period is associated with the greatest frequency of collisions. Many factors may contribute to the seaman's reaction to a particular watch cycle. The graveyard watch is always dark. It is frequently the loneliest and most boring. It is also the time when people traditionally sleep. Under the work/rest cycle followed by most people, work is engaged in for a continuous period of time usually beginning in the morning, followed by a long period of rest and sleep. However, the "four on-eight off" schedule makes this typical cycle virtually impossible for the graveyard watch and morning watch officers. The atypical work/rest cycles of those watchstanders in especially problematic when first boarding a vessel after several weeks ashore. While ashore the mates have usually acclimatized to the social and physiological schedules of shoreside people. Once at sea, the watch officer must readjust to the three-watch cycle. It is not known whether watch officers ever physiologically adjust to this atypical work/rest cycle. Physiological adaptation to the "four on-eight off" cycle may be quite difficult, given the natural rhythms of the body's functions.

The body functions of human beings typically follow a daily rhythm which is easily discernible through changes in body temperature, as illustrated in Figure 1. If the temperatures of watch officers follow the same daily rhythm (or circadian rhythm) as that of most people, then the watchstanders reach their biological low point from approximately 2 AM to 6 AM. This period overlaps with the undesirable graveyard watch, and is also associated with the greatest number of collisions at sea (see Figure 2). Since previous research on circadian rhythm (e.g., Mackie and Miller, 1978) has indicated a consistent relationship between fluctuations in body functions and changes in work performance, it would be logical to see similar effects among watchstanders.

In summary, three phenomena which are inherently part of the watchkeeping system, and may well have deleterious

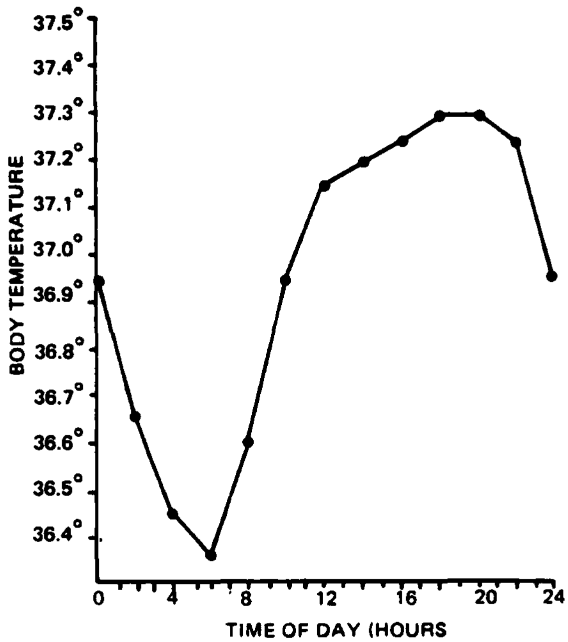

Figure 1. Typical Circadian Cycle of Body Temperature During Normal Activity Day

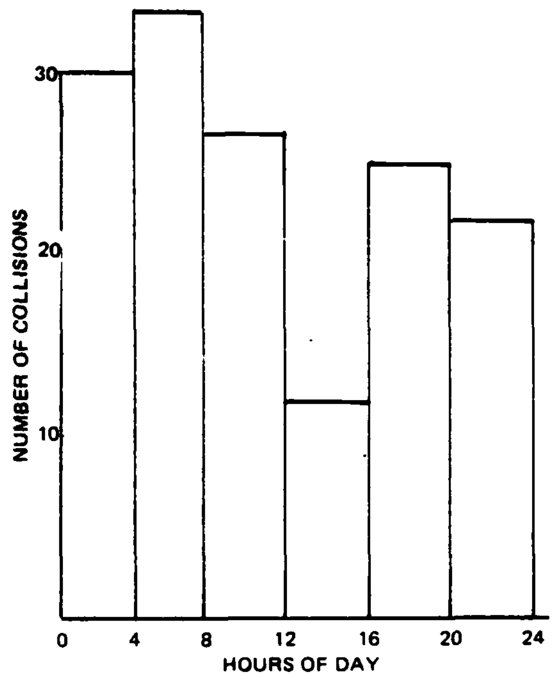

Figure 2. Number of Collisions as a Function of Time of Watch (from D. Rother, 1980, p. 50)

effects on the watchkeeper's performance, are fatigue, boredom and changes in body functioning associated with circadian rhythm. Each of these phenomena was considered when designing this research. In the current study of watchkeeping performance, watch officers with differing amounts of sleep on the previous night, were required to stand a four-hour simulated watch during either the morning or afternoon. The differential amounts of sleep were expected to induce varying degrees of fatigue. In addition, fatigue was expected to build up with the passing of each of the four hours on watch. The assignment of watchstanders to either a morning or afternoon watch was done in an attempt to separate watch officers who were experiencing rises in biological functioning from those who were experiencing a plateau of biological functioning. While such a gross separation does not allow for clear conclusions regarding the relationship of circadian rhythm to watchstanding performance, it does assist the researcher in the interpretation of the effects of sleep loss. Although the boredom is a difficult construct to quantify, inferences about it can be made by structuring the experimental task so that it is lengthy and unstimulating. Thus, the task used in this study was structured so that only four critical events occurred during the simulated four-hour watch. The resultant experimental design was constructed to examine watchstanding performance as a function of sleep deprivation, time of day, and length of time on watch. Each of these concepts is discussed in more detail below.

1.4 SLEEP DEPRIVATION

Fatigue can result from many different sources such as sleep loss, long hours of work, excessive exercise, or illness. While all of these may play a part in the vigilance of a watchstander, only the first, sleep loss, was examined in this research.

Sleep loss is generally considered to be a reduction in sleep from a person's normal amount (approximately 7.5 hours). Previous research has attended to the effects on human performance of virtually every type of sleep reduction — total reduction, partial reduction, gradual reduction over time — as well as redistribution of sleep over a 24-hour period. Much of this work has been conducted for the aviation industry, the military and the automotive transportation industry. The concentration of sleep loss research in these fields of application is an indication of their concern for preventing the devastating human and financial casualties which may result from fatigue-mediated human error.

Studies of sleep loss are frequently embedded in other studies of fatigue resulting from continuous work, since it is impossible to assess the effects of work lasting longer than 16 hours without also assessing the effects of sleep loss. The assessment of fatigue, resulting from both sleep loss and continuous work, is also confounded with the effects of circadian rhythm. Since most people sleep from roughly 2300 to 0800 hours, during their circadian low points, studies which extend the waking state into the early morning hours are simultaneously examining the impact of both sleep loss and circadian low points. Any continuous work lasting longer than 12 hours will also be influenced by fluctuations in circadian rhythm. The potential confounding of the effects of sleep loss, continuous work, and time of day were given serious consideration in the design of this research.

Systematic research on the effects of sleep loss has been ongoing since psychology emerged as a specialized area of knowledge. It is noteworthy that many of the measurement techniques used then and the phenomena observed are still constituents of today's research designs. One of the earliest studies of sleep deprivation in humans was reported in 1896 by Patrick and Gilbert. They studied the effects of approximately 90 hours of wakefulness on the reaction time, discrimination time, motor ability, memory, attention, and subsequent sleep patterns of three subjects. Mixed results, some counterintuitive, were found. For example, reaction time increased but discrimination time decreased, as the number of hours awake increased. Overall mental alertness, physical strength and voluntary motor ability also decreased with wakefulness, but visual acuity increased. In addition, Patrick and Gilbert found that all of their sleep-deprived subjects engaged in visual imaging somewhat like hallucinations or instantaneous dreams while apparently awake. They postulated that these instances of staring at something with eyes open while engaging in a dream-like process were actually like small naps and had a restorative function. In fact, the function of such daydreaming activities are still studied (e.g., Singer and Antrobus, 1963). While no firm conclusions could be drawn from Patrick and Gilbert's study about the nature of sleep and sleep loss, it did set the stage for research which would be conducted 80 years later.

Many different types of tasks, ranging from psychomotor to cognitive tasks, are affected by sleep deprivation. Aircraft carrier pilots subjected to variable sleep schedules simulating combat conditions made more landing errors as the variability of their sleep schedules increased (Brietson, McHugh and Naitoh, 1974). The ability of a pilot to manually track the localizer needle on the aircraft's instrument is also deleteriously affected by sleep deprivation over a period of 34 to 55 hours (Collins, 1976). While the accuracy of solving addition problems is not significantly affected by sleep deprivation for periods lasting up to 68 hours, the speed of solving addition problems is affected (Wilkinson, 1958). Vigilance tasks, such as an auditory signal detection task, may be impaired even when sleep is reduced by only three hours (Wilkinson, Edwards, and Hains, 1966). Sleep-deprived subjects also have difficulty in recalling word lists immediately after presentation (Williams, Gleseking, and Lubin, 1966). The effect of sleep loss on short-term memory has been further substantiated by Hamilton and Wilkinson (1970) and Hartley (1974). Other cognitive tasks which have also shown sensitivity to sleep deprivation are visual search and scan (Folkhard, et al., 1976) and signal detection (Deaton, et al., 1971).

The aspects of a task which make it susceptible to sleep deprivation are best summarized by Johnson and Naitoh (1974). The longer the duration of a task, the greater its susceptibility to sleep loss. Tasks which do not provide the worker with immediate feedback about his performance, which are difficult to perform, highly complex, or are relatively new to the worker are also highly sensitive to the effects of sleep deprivation. In addition, tasks which are paced by someone or something other than the worker, as well as tasks which rely on short-term memory for their successful completion are especially sensitive to sleep loss. It should also be noted that tasks which consist of an automatic response sequence are not significantly affected by sleep loss (Williams, Lubin, and Goodnow, 1959).

Some of the hypothesized mechanisms by which sleep deprivation interferes with task performance include the hypothesis that sleep-deprived people experience brief intermittent lapses of attention which increase in frequency and duration as sleep loss increases (Williams, Lubin and Goodnow, 1959), and the hypothesis that sleep deprivation interferes with the encoding of information in memory (Williams, Gleseking and Lubin, 1966).

Characteristics inherent in the person who is deprived of sleep also mediate the effects of sleep loss. Johnson and Naitoh (1974) summarized these as follows. If the person finds the task at hand interesting, he is less likely to show performance decrement after sleep loss. Likewise, if the person is motivated to perform, even if the task itself is uninteresting, then the effect of sleep loss will be mitigated. The point in the person's circadian rhythm will also affect the impact of sleep loss. Afternoon performance under fatigue conditions is generally better than early morning performance.

In light of these findings, the watchstanding officer who, during his circadian low point, works at a four-hour long, uninteresting task, while receiving no performance feedback, appears to be a prime candidate for being deleteriously affected by sleep deprivation.

In addition to decrements in task performance, sleep loss affects the mood of an individual. Measurement of an individual's mood (e.g., irritability) has been facilitated through previous research efforts which resulted in the development of a number of "mood scales". Changes in the direction of mood scale items, such as irritability and listlessness, are apparent even if no measurable performance decrements exist. These findings, as they relate mood to performance, are consistent regardless of the mood scale used (e.g., Taub and Berger, 1973; Lisper and Kjellberg, 1972; Moses, and Lubin and Naitoh, 1974; Bugge, Opstand and Magnus, 1979). The use of mood as a dependent measure in sleep loss studies is interesting from two perspectives: (1) mood frequently relates to task performance and is therefore a useful indicator of expected performance, and (2) mood is logically related to morale which may have a bearing on task performance.

1.5 RELATIONSHIP OF CIRCADIAN RHYTHMS TO TIME OF WATCH

Most living organisms exhibit biological fluctuations which occur with regularity. These rhythms are manifested in changes in body temperature, blood and urine chemistry, and cardiac rate. The rhythm appears to be inherent in the organism with periodic factors in the environment pacing the biological rhythm, but not causing it.

The typical circadian cycle is at its lower point (measured by body temperature) at about 6 AM. It rises steeply throughout the morning until about noon, after which it levels off. At about 6 PM, a gradual descent occurs until midnight, after which the rate of descent increases to the lowest point at 6 AM (see Figure 1). This typical cycle has been found repeatedly under various conditions. Aschoff (1980) has found similarly shaped functions under conditions of regular sleep with normal daytime activity, regular sleep with daytime bed rest, and sleep deprivation. Cook and Shipley (1980) conducted a field study of ship pilots working irregular hours and found the same type of rhythm. Body temperature is generally used as the standard measure of circadian rhythm; however, cardiac rate, urine catecholomine excretion, and urinary flow are also good indicators.

Previous research indicates that the cyclical fluctuations in body temperature correlate with work performance and mood across several industries. In general, lower temperatures, such as those occurring from 0200 to 0600 are associated with poorer performance, while higher temperatures, such as those in the afternoon, are associated with better performance (Aschoff, 1980). Bugge, Opstand, and Magnus (1979) report that scores on the Profile of Mood States correlate highly with circadian changes and performance in a ranger training course. Mackie and Miller (1978) found that truck driving and bus driving performance, measured by tests of steering and trackkeeping, degrades during circadian low points. These low points are also associated with subjective feelings of fatigue and are consistent across work schedules. Overall, the relationship of circadian rhythm to mood and work performance appears to be a stable phenomenon reported consistently throughout the literature (e.g., Colquhoun, 1969; Frobert, et al., 1972; Hockey and Colquhoun, 1972). There is some evidence (Black, 1967; Baddeley, et al., 1970), however, which suggests that memory tasks show an inverse relationship with circadian rhythm. Such "inverse rhythm" is exceptional, but is apparently a consistently found phenomenon with respect to short-term memory.

Whether standing watch during a period of ebbing circadian rhythms influences a watchstander's performance is not known. If the typical biological fluctuations were to affect watchstanding effectiveness, one would expect more errors to be committed by the watchstander who is working when his biological activities are at their least active stage, such as between 0200 and 0600. The casualty statistics required to shed light on this are not readily available. Although many organizations of international repute (e.g., Lloyds of London, United States Coast Guard, Det Norske Veritas) collect casualty data and issue casualty analyses, almost none of these analyses provide data suitable for studying the impact of circadian rhythm or fatigue on risk of casualty. The existent analyses break down merchant vessel collisions, rammings, groundings, and swampings by a plethora of ship and environmental categories (e.g., weather conditions, angle of impact, location, gross tonnage), yet almost universally fail to categorize casualties by time of day or watch cycle.

One of the few casualty analyses to categorize casualties by time of day was performed by Rother (1980) and is illustrated in Figure 2. It shows the likelihood of a collision is greatest from 0400 to 0800 and least from 1200 to 1600 hours. There are certainly many possible explanations for this: for instance, the 0400 to 0800 period which shows the greatest percentage of casualties may be assigned to the most inexperienced mate (on ships where the chief mate does not stand a watch); it is also generally the foggiest time; however, it is also the time when a person's body temperature is at its lowest point in the circadian cycle — a point traditionally correlated with poorer attention and performance on non-maritime tasks.

Whether ship casualties are directly or indirectly caused by the inattentiveness associated with circadian low points or its relationship to fatigue is not clear and has never been previously addressed. However, the fairly consistent relationship between circadian rhythm and work performance in other industries, leads one to consider ebbing circadian rhythm as a contributor to poor watchkeeping performance. It is also important to consider circadian rhythm when assessing the affects of sleep deprivation on watchkeeping, since any prolonged assessment of watchkeeping will be associated with changes in biological functioning.

1.6 LENGTH OF TIME ON WATCH

There is a substantial body of research in an area broadly known as vigilance in which most of the studies focus on the performance of a monitoring task over a fairly prolonged length of time. Because of the similarity of watchkeeping on a ship to certain vigilance tasks cited in the literature (e.g., radar monitoring), the literature on vigilance was consulted. Vigilance has been defined by Mackworth (1957) as a state of readiness to detect small changes which are intermittent in the external environment. Thus, the detection of intermittent signals or faint perceptual stimuli is frequently used as an operational measure of vigilance. In particular, the number of correct or incorrect detections and the quality of the signal necessary for detection are usually studied as dependent variables. Vigilance is virtually always studied over some relatively long duration of time, since it is believed that the attention and effort necessary to maintain vigilance wanes over time, especially when the signals to be detected are random and infrequent. Research on vigilance bears out this belief. For example, in a study of target detection using a visual display simulating the radar used by air traffic controllers, signal detection latencies decreased significantly over a two-hour period (Thackray, Touchstone and Bailey, 1978).

One of the most relevant studies to the research reported herein was conducted by Colquhoun, Hamilton and Edwards (1974) to assess the effects of prior sleep and circadian rhythm on a four-hour sonar-monitoring task. They assessed the efficiency of signal detection from 0400 to 0800 and again from 2000 to 2400. Regardless of the recency of the subjects' sleep, monitoring was better in the evening at peak points in the circadian rhythm cycle than the morning. However, all groups showed a significant performance decrement over the four-hour time span.

The impact of number of hours worked on vigilance may be mediated by an individual's method of coping with the boredom associated with prolonged vigilance or with his personality traits. Previous research by Bakan, Belton and Toth (1963) indicates that introverts show smaller decrements in vigilance over time than extroverts, and their ability to perform vigilance tasks better than extroverts appears to be related to the level of stimulation in a task. Bakan found in an earlier study (1959) that when signals to be detected were relatively infrequent, introverts performed better; but when signals increased in frequency, introverts and extroverts performed equally well.

One hypothesis used to explain such data is that the introvert is less reliant than the extrovert on external sources of arousal and therefore, is less negatively affected by an unstimulating situation. This hypothesis was corroborated in reports by Davies, Hockey and Taylor (1969) and Corcoran (1965). Corcoran also found that such results carry over into sleep deprived conditions. If this hypothesis were true, then one could further postulate that the individual with a predisposition for an active mental life, such as that associated with daydreaming, may also be better prepared to cope with boredom. In fact, Antrobus and Singer (1964) conducted research which suggests that subjects with varied inner responses, analogous to daydreaming, may be capable of sustaining alertness during "understimulating" monitoring tasks. Their subsequent work (Antrobus, Coleman and Singer, 1967), however, showed a tendency for daydreaming to compete with signal detection when the monitoring task was more highly arousing. Specifically, they found that very active daydreamers had a tendency to commit more errors in a stimulating signal detection task than less active daydreamers.

In summary, daydreaming may serve to alleviate boredom and increase alertness in a vigilance task of low stimulation and long duration, however, it may interfere with a highly stimulating vigilance task. This view is affirmed by Broadbent, (1958).

1.7 METHODS OF STUDYING WATCHSTANDING

The choice of an appropriate method for assessing watchkeeping must be governed by two issues: the internal and the external validities of the results of the research. Internal validity refers to the degree to which changes in the dependent measures (e.g., watchstanding performance) can be attributed to manipulation of the independent variables (e.g., number of hours sleep). External validity refers to the degree to which the results of the research can be generalized to the real world.

To date, the only data relating sleep loss, watch period, or hours worked to mariner performance has been derived from case studies of individual casualties. Within these casualty reports, the determination that fatigue resulting from loss of sleep has contributed to human error has been arrived at through inference or conjecture. These case studies are, by definition, retrospective; they cannot manipulate fatigue and predict the results. Thus, they are seriously deficient in internal validity. In addition, since case studies are conducted only after a casualty has occurred, they afford no systematic way of determining how fatigue interacts with other variables.

One of the most realistic methods of investigating the effects of sleep loss, watch period and hours worked on watchkeeping performance would be to subject watch officers on merchant vessels to various levels of these variables, and then require them to perform their work. While this method offers both internal and external validity, its practicality and ethics precluded its implementation on this particular research project.

To maximize the internal and external validity of this research, while working within the boundaries of practicality, the Computer Aided Operations Research Facility (CAORF) ship simulator operated by the National Maritime Research Center at Kings Point, New York, was used to carry out the research and experienced watch officers were employed as subjects.

1.8 DESCRIPTION OF CAORF SIMULATOR

CAORF consists of a replication of the bridge of a ship and the view from the bridge. The CAORF bridge looks like any bridge on a modern ship. It contains all the instrumentation required to maneuver "ownship"[1] including gyrocompass, throttle, RPM indicator, helm, communication devices, and radar. Looking through the wheelhouse windows, one sees a full-color projected simulation of the outside world. As the mariner takes action to control the ship (such as ordering a course change or increasing speed), a central computer interprets his actions and alters the visual scene accordingly. The central computer also adjusts the instrument readings to correspond to the changing visual scene. Simulated radar displays are also coordinated with the outside visual scenes. Thus, the radar depicts buoys, shorelines, and other ships in their proper locations with respect to ownship. In the case of simulated fog, the radar depicts objects which may not be visible in the visual scene.

The visual scene itself consists of computer-generated images projected onto a cylindrical screen. The mariner has a 240-degree horizontal view, 120 degrees on either side. Illumination can be full daylight, twilight, or moonless night, and may be combined with haze or fog to create any desired range of visibility. For example, to simulate an early morning voyage, the CAORF computers can start with a moonless night, gradually increase illumination to hazy daylight, and then gradually lift the haze so that visibility is unlimited (to the horizon).

Up to six moving ships and numerous stationary ships can be included in the visual scene. These moving traffic ships can be individually controlled in response to ownship. For example, if the master of ownship wishes to pass "port to port" with one of the traffic vessels, he can pick up his radio telephone and communicate with the master of the other ship (whose response is simulated by a member of the experimental team), arrange to pass each other on a one-whistle signal, and then proceed to do so. A "control station" operator then "steers" the other ship to correspond to the agreed upon meeting situation. Even the one-whistle signal is simulated. Communication with the mariner/subject, manipulation of vessels within the visual scene, and initialization of fog, wind, or current conditions are directed from the "control station."

Ownship's movements and positions are also automatically recorded by computer so that ship performance can be assessed. For example, the specific angle of the ship's rudder, the distance of the ship from land masses, and the speed of the ship can be sampled and recorded at various rates (usually every 30 seconds).

[1] *"Ownship" is a marine team used to refer to the ship one is on. In this context it refers to the ship operated from the* CAORF bridge.

A "human factors monitoring station" is also available for monitoring the behavior of the subjects. Closed circuit television cameras and four microphones are located on the bridge to record all activities, comments, and commands.

The use of a shiphandling simulator allows the researcher to manipulate experimental variables while keeping the physical environment constant across all subjects, thereby maximizing internal validity. However, the generalizability of the experimental results to the "real world" can be questioned. One can never be sure that behavior on a simulated bridge conforms to that on a real bridge. But one must have confidence that behavior on a simulated bridge **approximates** that on a real bridge. Early validation work conducted at CAORF provided this confidence.

The CAORF simulator has been validated with respect to its ship response characteristics, perception of visual scenes, and the behavior of mariners. The entire validation process has been reported in detail in a document entitled **Validity of the Computer Aided Operations Research Facility (CAORF**; published in January, 1979.

A brief review of the validation effort follows.

1.8.1 SHIP RESPONSE CHARACTERISTICS

To produce valid research results, it is vital that the response characteristics of ownship be realistic. For example, the mathematical models used in simulation must accurately predict the response of a 165,000 dead weight tons (DWT) tanker to the application of 25-degree right rudder while traveling at 12 knots with a 2-knot following current. The mathematical models used at CAORF have been verified by comparing ship responses they generate to the ship responses generated by independently derived equations.

Of most interest is the comparison of the ship response characteristics found at CAORF to those found in the real world. Three different ship maneuvers were used for comparison: turning circles, zig-zags, and spirals. The validation procedure required these three maneuvers to be executed at CAORF with a simulated 80,000 DWT tanker and then to be executed on a real 80,000 DWT tanker at sea. All three CAORF ship response characteristics compared favorably with the at-sea responses.

1.8.2 PERCEPTION OF VISUAL SCENE

The sizes of traffic ships displaced at various distances on the CAORF projector screen appear to be the same in the real world. This was determined by asking master mariners to estimate the ship type, ship dimensions, and range of traffic vessels at sea and at CAORF. Comparisons of these estimations indicated that master mariners tended to overestimate the distance of ships at ranges less than 7 nautical miles (nm) and to underestimate the distances of ships at ranges greater than 7 nm. This result was attributed to the fuzziness of images on the screen. The master mariners were then trained in range estimation of the CAORF simulator and the subsequent estimations yielded results quite similar to those made in the real world.

The observed overestimation of distances at CAORF does not present a problem for this study, since the visual estimation of traffic ship distances was not a critical task in this research. Any traffic ship sightings made by subjects were to be verified on radar and the position of the traffic ship had to be determined using proper instruments and procedures rather than naked eye estimations. In addition, subjects in this study, as in all CAORF research, were familiarized with the bridge and scene in order to mitigate the effects of a novel situation on the experimental procedure.

1.8.3 MARINER BEHAVIOR

To validate the behavior of mates on the CAORF bridge, behavioral data were collected at sea under various visibility and geographical conditions CAORF scenarios were then designed which represented. the conditions under which data were collected at sea. These scenarios were then run on the simulator with mates on watch and behavioral data were collected using procedures like those employed at sea. The comparison focused on four areas of interest: (1) the mate's overall activity level; (2) the mate's use of radar under different conditions; (3) the proportion of time spent on specific tasks during an entire watch period; and (4) the variability of behavior among mates.

Results indicate that mates are more active on the CAORF bridge than at sea. This is probably due to high motivation and lack of familiarity with the CAORF bridge. During the validation studies, mates at CAORF walked more, tested more instruments, and, in general, showed more interest in their surroundings than in the real world. A more appropriate comparison would have been between mates standing watch on the CAORF bridge and mates standing watch on the bridge of an unfamiliar ship at sea.

Mates also behave more cautiously at CAORF, taking fewer risks than in the real world. (This finding is noteworthy in that it contradicts the assumption, made by many, that CAORF subjects would behave more impulsively or riskily since no danger exists.) The greater activity and cautiousness of mates at CAORF necessitates complete familiarization with the CAORF bridge prior to experimental runs. Knowing that CAORF mates are more cautious than usual also strengthens the confidence one has in those CAORF research results which clearly show mariner errors. That mates are generally more active at CAORF may mute any effects of fatigue. However, should effects of fatigue on the mariner be demonstrated, the results may also be viewed with confidence.

Mates at CAORF use the radar under circumstances similar to the real world. The transfer of radar observation techniques from the real world to the simulator is quite importtant to the validity of a study on watchkeeping, since an important component of keeping a good lookout is timely monitoring of the radar. Mates at CAORF also distribute the time spent on tasks during a watch period in a manner proportional to the real world. Of even further value is the observation that stylistic differences between mates carry over from real life to CAORF. That is, the CAORF bridge does not elicit one style of watchstanding more than others. The "pacers," "dreamers," and "brass and glass shiners" who are found in the real world can also be found at CAORF.

Under any conditions of simulation, behavior will differ somewhat from the real world. However, such differences, if minor, do not discount the usefulness of simulation in testing hypotheses about why accidents occur or in training personnel to avoid tragedies. The methodology outlined in Chapter 2 could not have been conducted in the real world because of the requirement for tight control over the environmental conditions in order to maintain internal validity. Of course, by sailing on many vessels under many different conditions, one could gather data concerning the relationship of fatigue and watchstanding, but the internal validity of these conclusions could then be questioned.

The CAORF simulator is the premier shiphandling simulator in the world. It affords the researcher the opportunity to study human performance on the ship's bridge in a credible environment where control over experimental variables can be maintained. Its use in this research project permitted hypotheses to be tested under conditions closely approximating those at sea without endangering lives of property. Appendix A contains a detailed description or CAORF's characteristics and operating procedures. It is a more extensive description than is necessary to understand this report; however, the interested reader may wish to consult it.

1.9 SELECTION OF EXPERIMENTAL ENVIRONMENT, INDEPENDENT AND DEPENDENT VARIABLES USED IN PRESENT RESEARCH

In this particular study, the CAORF simulator was used as the laboratory in which watchstanding performance was assessed as a function of the effects of sleep deprivation, time of watch and length of time on watch. The effects of sleep deprivation on watchstanding were of primary interest because of the greater attention they have received in casualty reports and marine transportation reports, and because of the expectation that sleep deprivation would have greater effects on watchstanding than the other two variables. Nevertheless, time of watch and length of time on watch were also studied because of their expected interaction with sleep deprivation. The rationale for these expected effects has been drawn from the literature reviewed and presented in this chapter. The test subjects were experienced watchstanding mates divided into four groups on the basis of amount of sleep (none vs. 7.5 hours) and time of watch (morning vs. afternoon). The experimental task required each subject to stand a four-hour watch on the CAORF simulator performing his duties as he would on an actual ship. The dependent variables were assessed before, during, and after the watch.

Certain dependent variables were selected because of their direct relationship to the responsibilities of watchstanders. Watchstanding mates are primarily responsible for keeping a good lookout so that early detection of potential threats to the ship's safety can occur. Such threats include the presence of another ship which may collide with ownship, failure of equipment, or involvement with high seas and winds. The vigilant watchstander should detect the presence of such events very soon after they become discernible, and should conduct his watch in a manner which maximizes the probability of early detection. In this research, the work performance of a mate was measured by the rapidity with which he detected ships and an instrument failure. Because the probability of an early detection should increase with the degree to which the watchstander consults the radar and moves around the bridge, radar observation and physical activity were studied as work-related measures.

Several measures not directly related to watchstanding performance were also examined in this study. The Profile of Mood States Checklist, the Naval Personnel Research Unit Affect Scales, and the Stanford Subjective Sleepiness

Scale were administered in order to examine the effects of the independent variables on certain subjective feelings such as fatigue, vigor, anxiety, depression and confusion. The particular scales used to measure these feelings were selected because previous research has indicated that they are sensitive to sleep loss and time of day.

Measures of cognition were also administered in order to assess the effects of the independent variables on levels of cognitive functioning which may be necessary for watchkeeping. For example, the ability to recall items from short-term memory, visually search and scan the environment, and make logical transformations are necessary for sighting, fixing the position of, and reporting traffic ships at sea. Should sleep loss negatively affect these cognitive functions, then the performance of the watchstander may also be affected. Thus, measures of these cognitive functions were included as dependent variables in this study.

Physiological measures were also examined throughout the course of the watch. Measures of muscle tension (EMG) taken from the frontalis are indicators of alertness, which is a crucial prerequisite to effective watchstanding. Measures of heart rate (EKG) and heart rate variability are generally sensitive to changes in mental workload and alertness, as well as length of time on watch.

Finally, a subjective, retrospective report by the subject regarding his evaluation of the watch was included to gain insight into the watchstander's perspective on his performance.

When one considers these dependent variables together, one can infer a great deal about the effectiveness of a watch officer. Thus, this study attempted to systematically evaluate the effects of sleep deprivation and related operational factors on watchstanding effectiveness.

1.10 OBJECTIVES OF PRESENT RESEARCH

The primary purpose of the present research was to measure the effects of sleep deprivation, time of watch, and length of time on watch on a watchstander's work performance, work-related performance, mood, cognitive measures and physiological measures under a simulation of the conditions of mates standing open-sea watches.

A second purpose was to evaluate marine simulation as a method for studying the effects of sleep deprivation, time of watch and length of time on watch on mariner performance. This evaluation incorporated the use of performance measures previously used in non-maritime industry research as well as performance measures specific to the maritime industry.

1.11 RESEARCH HYPOTHESES

1. A main effect of sleep deprivation was hypothesized:

   Mates with no sleep within a 24-hour period prior to assuming their watches would be less effective in their watchstanding performance than mates with a full night's sleep.

2. An interaction of time of watch and number of hours on watch was hypothesized:

   a. Mates standing a morning watch would be less effective at the beginning of their watch than mates standing an afternoon watch.

   b. No differences in watchstanding effectiveness would be found between morning and afternoon watchstanders towards the end of their respective watches.

3. A main effect of length of time on watch was hypothesized:

   Watchstanding effectiveness would deteriorate from the beginning to the end of the watch.

CHAPTER 2

METHODOLOGY

2.1 OVERVIEW OF PRESENT STUDY'S METHODOLOGY

Three independent variables were tested:

Between Subjects:

Amount of Sleep in Past 24 Hours: 7½ hours versus 0 hours.
Time of Watch: 0800 to 1200 versus 1200 to 1600.

Within Subjects:

Length of Time on Watch: 1 through 4 hours

Several dependent variables were assessed as a function of these independent variables:

Work Performance: Speed with which traffic ships were sighted; speed with which an instrument failure was recognized.

Work-Related Measures: Frequency and duration of radar observations; physical activity.

Measures of Affect: Subjective feeling of fatigue; mood state.

Cognitive Measures: Digit Span test of short-term memory; Visual Search and Scan test; Baddeley Grammatical Transformation Test.

Physiological Measures: Muscle tension; cardiac rate; physical activity.

Subjects were experienced second mates. They arrived one day prior to the experimental runs and received three hours of familiarization on the CAORF bridge during the first day. During the first evening, subjects were monitored closely to minimize the introduction of spurious events. Half of the subjects retired to bed at 11 PM; the other half stayed awake all night and were engaged in a number of nonstrenuous activities. The next day, they each stood a four-hour, open-sea watch during which three ships appeared and one instrument failure occurred (a total of four "events" distributed over the four hours of each watch). Dependent measures were collected before, during, and after the watch period.

This procedure was the refined outcome of a procedure tested during preliminary experimental trials.

2.2 PRELIMINARY STUDY

Because of the planned investment of significant time and money into the conduct of this research, an initial study was conducted in June, 1981, in order to streamline and refine the experimental method. At that time data were collected from six mates, each assigned to one of four experimental conditions: full sleep, morning watch (n=2); full sleep, afternoon watch (n=1); no sleep, morning watch (n=2), no sleep, afternoon watch (n=1). Each subject stood a 3½ hour familiarization watch on the CAORF simulator, followed by a four-hour experimental watch the next day. During the watch, measures of cardiac rate (EKG), muscle tension (EMG), and physical activity were recorded using an ambulatory monitoring system. Oral temperature was taken and the Profile of Mood States Checklist (POMS), Naval Personnel Research Unit (NPRU) Mood Scale, Stanford Subjective Sleepiness Scale, Baddeley Grammatical Transformation Test, Digit Span Test, and Visual Search and Scan Test were administered before and after the watch period. During the middle of the watch period, an experimenter went out onto the bridge and administered all of these except the POMS, Digit Span, and Visual Search and Scan Test to the subject. Three traffic ships appeared during the four-hour watch, and one instrument failure occurred. The time it took each subject to notice each of these incidents was recorded.

The changes in physiological measures over time were examined. Measures of EKG, EMG and physical activity, when recorded during the 0800 to 1200 watch, appeared to change in comparable ways over time. During the first and second half hours of the watch, the subject was fairly physiologically active. This activity declined through the fourth half hour. Midway through the watch the experimenter intervened to administer tests; this mid-watch break appeared to have an arousing affect during which time physiological activity rose dramatically. Thereafter, this activity again declined steadily, reaching a low point at the end of the watch. Because of the apparent stimulating effect of the experimenter's intervention at mid-watch, the procedure was amended for the present study to eliminate this intervention.

NPRU Mood Scale, Stanford Subjective Sleepiness Scale and POMS Checklist indicated that, irrespective of sleep, afternoon watch subjects felt less vigorous or active than morning watch subjects. The afternoon non-sleeper was considerably more fatigued and sleepy than the other subject. The morning watch mates reported more vigorous feelings compared to the afternoon watch mates; afternoon subjects also reported greater confusion on the POMS. Based on these preliminary findings, it was decided to proceed with the use of these self-report measures.

In contrast to these data were the response times to recognizing ships. Sleepers appeared to take longer to recognize the first and second target ships than non-sleepers. Except for the afternoon non-sleeper, all mates became quicker at recognizing target vessels as the watch progressed. One important aspect of ship sightings which was considered in the preliminary study was the modality used for sightings. Ship sightings by radar may be quicker than those made visually. Perhaps by the end of the watch period, mates huddle over the radar. In the present study, both reaction time and modality used for traffic sightings were recorded.

The results of this preliminary research and discussions with the subjects were responsible for two changes to the procedure which were incorporated in the method of the current study. The first modification eliminated the experimenter's walking out onto the bridge to administer tests. Only checklists which could be self-administered were used at mid-watch. The second change required the introduction of a helmsman to the bridge for the first half hour of the watch. The helmsman steered about 25 minutes, was called away, came back to the bridge a few minutes later and told the mate he must do some deck maintenance work. This addition was made for the sake of realism and to acclimate the subject to the simulation.

2.3 SUBJECTS

Twenty-five second mates participated in this experiment. Second mates were selected because they have experience standing both the 0800 to 1200 watch and the 1200 to 1600 watch. Each subject must have sailed within the past year and must have been on shore leave for at least two weeks prior to the beginning of the experiment. They were required to log their sleep schedules and oral temperatures (at least every two hours while awake) for five days prior to the experiment so that circadian cycles and sleep cycles could be ascertained.

Volunteers with a history of heart disease or high blood pressure or those who were under medication for anxiety-related problems were exempted from the pool of subjects. Females were also exempted for two reasons: the population to which the results would be generalized is overwhelmingly male and the potential differences between male and female responses to the independent variables could dilute experimental effects.

Each potential subject was sent an introductory letter and package of sleep logs and thermometers one week prior to the onset of the experiment. Prior to participation in the project, each subject was requested to sign a consent form. A copy of these documents appears in Appendix B.

During the 24 hours prior to experimentation, subjects were not permitted to drink alcohol, use drugs other than for medical purposes (e.g., antibiotics were permitted) or engage in strenuous physical activity.

The 25 mates were randomly assigned to one of four groups. These were:

	Time of Watch	Amount of Sleep	Number of Subjects
Group 1	0800-1200	7½ hours sleep	n=6
Group 2	1200-1600	7½ hours sleep	n=6
Group 3	0800-1200	no sleep	n=6
Group 4	1200-1600	no sleep	n=7

The unequal sample sizes arose from the plan to run three more subjects than were actually needed, in case some subjects were lost. In fact, one of these "insurance" subjects failed to return after the familiarization watch, and another was discarded by the experimenter when he inadvertently read the plans for traffic ship locations and instrument failure. This left one extra subject who for logistical reasons was placed in Group 4. While six subjects per group is quite minimal, the cost associated with reimbursing subjects and using the simulator was so high (roughly $6,400 per subject) that the addition of more subjects would have increased expenditures to a prohibitive level. In addition, collapsing across the two levels of time of watch would provide an n=12 or 13 for each of the sleep conditions. Previous CAORF research has been successful at demonstrating significant differences between treatments using fewer subjects. While these previous projects produced significant effects with few subjects, the measures used to evaluate performance were primarily related to trackkeeping (e.g., Williams and D'Amico, 1980). Whether comparable sensitivity and magnitude of experimental effects could be expected from measures of watchkeeping was not known. However, a preliminary study using a total N=6 showed what appeared to be rather large effects of time of watch and sleep deprivation on mood, and a discernible effect of number of hours of work on physiological measures.

2.4 INDEPENDENT VARIABLES

2.4.1 HOURS OF SLEEP: 7½ HOURS VERSUS 0 HOURS

Based on the literature reviewed in Chapter 1, it is clear that loss of sleep has a deleterious effect on work performance, cognitive efficiency, and mood. Such effects can be seen even by limiting sleep to half the normal amount. It is the responsibility of the watch officer to ensure that he is well-rested before assuming his watch. However, reviews of casualty reports and interviews indicate that this responsibility may be fulfilled by catching a nap for about two hours in a 24-hour period. Whether two hours of sleep is sufficient to prepare the watch officer for his vigil is not known. This study compared 7½ hours of sleep to none at all on the evening prior to the experimental watch. If this study were to find that no sleep has a negative impact on watchstanding, then follow-up research would focus on determining the minimum acceptable amount of sleep. If this study and others were to indicate that 24-hour sleep loss does not interfere with watchstanding, then there may be some merit in discovering the threshold of sleep deprivation beyond which vigilance is impaired.

2.4.2 TIME OF WATCH: 0800 TO 1200 VERSUS 1200 TO 1600

In order to examine changes in watchkeeping as a function of changes in circadian cycles without confounding those effects with length of time on watch, it was necessary to examine watchstanding during at least two distinct periods of the day. The two watch periods had to differ with respect to the circadian phases associated with them and had to be similar with respect to ambient light (i.e., period of daylight). Two sets of watch periods were considered for possible inclusion in this study: the 2000 to 2400 (8 PM to midnight) and 0000 to 0400 set (midnight to 4 AM), or the 0800 to 1200 (8 AM to noon) and 1200 to 1600 set (noon to 4 PM).

The 2000 to 2400 and the 0000 to 0400 set was highly desirable since the first watch is associated with circadian plateau (consistent core body temperature over several hours) and the latter with circadian decline (a steady drop in core body temperature over several hours). In addition, the nighttime darkness would probably provide few sources of stimulation and enhance any effects of fatigue or boredom. However, running the simulator at night would have required employing several people to work a graveyard shift, thereby significantly increasing the cost of the study.

Therefore, the time from 0800 to 1600 was used. The 0800 to 1200 period is associated with a rise in body temperature and the 1200 to 1600 period with a plateau in body temperature. Both watches occur in daylight, without influences of dawn or dusk, and both are preceded by a regular meal. Because of the changing nature of the biological functions during the morning period, and the steady nature of them during the afternoon period, the relationship of these two periods with amount of sleep is fairly complex.

2.4.3 LENGTH OF TIME ON WATCH: 1 THROUGH 4 HOURS

The length of time which an officer has been standing watch may influence his watchkeeping effectiveness. A decline in effectiveness could be due to progressive boredom or fatigue or the result of a declining circadian phase. Likewise, a decline in effectiveness resulting from something such as boredom could be negated by a rise in circadian phase. In this experiment, the length of time on watch was a within-subjects variable.

While length of time is a continuous variable, in this study it is frequently treated as having discrete levels such as "before and after" or "first, second, third and fourth hours." For example, most psychological measures used in this study were assessed before and after the watch, while detection times of traffic ships and an instrument failure were assessed at a rate of one per hour. Because psychological tests had to be administered by an experimenter, they could only be assessed at discrete times. Few measures were actually taken at mid-watch, since the preliminary study showed an excessively long interruption would have been potentially arousing.

Since radar observation data and physiological measures were automatically recorded, it was relatively easy to assess their changes over time, although even these data were averaged across approximately five-minute segments throughout the four-hour watch. A summary of the intervals in which the dependent measures were collected appears in Table 1.

TABLE 1. TIME AT WHICH EACH DEPENDENT MEASURE WAS COLLECTED

Time of Measurement	Dependent Measures
Monitored Continuously During Watch	EKG, EMG, Physical Activity Radar Observation
Pre-Watch	NPRU[1] Positive/Negative Affect Scale Stanford Subjective Sleepiness Scale Digit Span Test Forward/Backward Visual Search and Scan Test Baddeley Grammatical Transformation Test POMS[2] Checklist Oral Temperature
Mid-Watch	NPRU Positive/Negative Affect Scale Stanford Subjective Sleepiness Scale POMS Checklist Oral Temperature
Post-Watch	NPRU Positive/Negative Affect Scale Stanford Subjective Sleepiness Scale Digit Span Test Forward/Backward Visual Search and Scan Test Baddeley Grammatical Transformation Test POMS Checklist Oral Temperature Post-Experimental Debriefing Questionnaire

[1] NPRU, Naval Personnel Research Unit
[2] POMS, Profile of Mood States

2.5 DEPENDENT VARIABLES

The dependent variables employed in this experiment fall into four categories: measures of work performance, affect, cognition, and the physiology. Each is discussed below.

2.5.1 WORK PERFORMANCE

To assess the effects of amount of sleep, time of watch and number of hours on watch on work performance, the subjects were placed in a simulated work environment. This environment consisted of the CAORF ship simulator bridge. The rationale for utilizing a simulated workplace rather than a real one was discussed in Chapter 1.

As part of the simulated working conditions, the view from the CAORF bridge was designed to simulate that of a 80,000 DWT tanker. This particular ship type is a common vessel. Four incidents occurred during the watch cycle: three ships passing at sufficient range from ownship that no course alteration would be required, and one minor instrument failure (see Table 2). Each is detailed below.

Traffic Ship Traveling at 20.4 Knots. In this incident, a containership traveling at 20.4 knots appeared on the horizon (16.8 nautical miles away) at 61 degrees to either the left (port) or 61 degrees to the right (starboard) of ownship's bow. Whether this ship was on the port or starboard beam was not expected to have a significant impact on detection time; however, port and starboard were counterbalanced across subjects. The time it took for the mate to recognize and record each sighting, and the mode he used for detection (radar or visual sighting), served as performance measures. Subjects were instructed to record sightings as soon as they were made, regardless of mode. Because the presence of a traffic ship could be detected earlier using the radar, it was expected that the more effective watch officers would report more radar rather than visual sightings.

Traffic Ship Traveling at 12 Knots. In this incident, a very large crude carrier (VLCC) traveling at 12 knots appeared on the horizon 105 degrees to either the port or starboard of ownship. Port and starboard were counterbalanced. The time it took for the mate to recognize and record each sighting, and the mode he used for detection, served as performance measures.

Naval Destroyer Astern. In this incident, a naval destroyer, traveling at 29.7 knots, overtook ownship from astern on

TABLE 2. INCIDENT ORDER MATRIX

Sleep Condition	Watch Condition	Subject No.	Incident Order: First Hour	Second Hour	Third Hour	Fourth Hour
7.5 HOURS SLEEP	0800 to 1200	10	1L	3	4	2R
		14	2L	4	3	1R
		16	2L	3	4	1R
		24	1L	4	3	2R
		26	1L	3	4	2R
		32	2L	4	3	1R
	1200 to 1600	11	1R	4	3	2L
		15	2R	3	4	1L
		17	2R	4	3	1L
		25	1R	3	4	2L
		27	1R	3	4	2L
		33	1L	4	3	2R
NO SLEEP	0800 to 1200	12	1L	4	3	2R
		19	1R	4	3	2L
		22	2R	3	4	1L
		28	2L	3	4	1R
		30	2L	4	3	1R
		34	1R	3	4	2L
	1200 to 1600	13	1R	3	4	2L
		20	1L	4	3	2R
		21	1R	4	3	2L
		23	2L	3	4	1R
		29	2R	4	3	1L
		31	2R	3	4	1L
		35	2L	3	4	1R

Key: 1 – Vessel at 20.4 kt. Left 299°, Right 61°
2 – Vessel at 12.0 kt. Left 255°, Right 105°
3 – Destroyer at 29.7 kt. from Astern 54°
4 – Instrument Failure

the starboard side. It passed ownship 1 nautical mile astern. An astern approach may go unnoticed for quite some time by a watchstander, who neglects to maintain a visual lookout astern. Therefore, sighting is often accomplished initially by radar. The time it took for the mate to recognize and record this sighting was used as a performance measure.

Instrument Failure. This incident did not involve traffic but concerned an instrument failure on the bridge. The RPM indicator was selected. This particular instrument is not one which regularly demands attention during the course of the watch; rather attention would be directed towards the RPM indicator as part of a systematic review of instrument status. Thus, notice of its failure would require more than a minimal level of vigilance. At the appropriate time, the main engine tachometer appeared to fail, registering zero RPM. The propulsion failure alarm did not light. If the subject proceeded with an alarm test, the result showed a burned out bulb. When the mate called the engine room to inquire or inform the engineer, a control station operator responded that "the engine is turning at X RPM, I'll have to check your bridge tach circuits." Approximately five minutes after this call, the tachometers were restored and the control station personnel, in the guise of an engineer, informed the mate that "a fuse blew and you should have the tachs back now." If the mate asked about the burned out bulb, he was told that it would be replaced at the end of the watch. Time to notice instrument failure was recorded as a performance measure.

The order in which each of these incidents occurred was counterbalanced across subjects. The incident order actually used appears in Table 2.

2.5.2 MEASURES OF AFFECT

NPRU Mood Scales. The NPRU Positive and Negative Affect Scales are currently used in sleep research at the Naval Health Research Center in San Diego and were used in this study to assess the effects of the independent variables on mood. The scales were designed specifically to evaluate the effects of sleep loss on mood. Previous research (e.g., Moses, Lubin and Naitoh, 1974) has demonstrated that these scales are, indeed, sensitive to the effects of fatigue. They are easy to administer and can be used as repeated measures to detect changes over time.

Stanford Subjective Sleepiness Scale. The Stanford Subjective Sleepiness Scale was also used to evaluate the subjects' feelings of fatigue. This scale has been extensively employed in sleep research conducted at the Naval Health Research Center and at the Stanford Sleep Research Center.

Profile of Mood States Checklist (POMS). The POMS has also been used in fatigue research and was employed in this study. Research by Bugge, Opstad and Magnus, (1979) indicates that men who are deprived of sleep and required to work long hours show changes in mood which correlate highly with the adequacy of their performance in a ranger training program. Fluctuations in mood as measured by the POMS have also been correlated with oscillations in circadian rhythm.

The POMS Checklist itself consists of 65 five-point adjective rating scales which produce six mood scores: anxiety, depression, anger, confusion, vigor and fatigue. Of these scales the last two were the most relevant to this study. In addition, the confusion scale was of interest because it was considered to be a subjective measure of clarity of thought.

According to Eichman (in Buros, 1978, p. 1015-1019) "reliability appears to be acceptably high. K-R 20 values range from 0.84 to 0.95 in two samples of 350 and 650 psychiatric patients. Test-retest correlations range from 0.65 to 0.74, with median 0.69. This is a considerable difference, but it is concordant with the purpose of measuring 'transient fluctuating affective states'".

2.5.3 COGNITIVE MEASURES

Visual Search and Scan Test. This test requires the subject to search for specific target letters embedded within a computer-generated array of other letters (Folkhard, et al., 1976). The target letters can be varied in length and do not necessarily have to appear in their original order or adjacent to one another. Each subject was instructed to check the lines in which the target letters appeared. He was allowed two minutes to complete as many lines as possible. By changing the number of target letters, (i.e., two, four, or six) mental workload was varied. Increased fatigue and inattention were expected to result in an increase in errors and a decrease in number of lines scanned.

Digit Span Test. This test of short-term memory requires the subject to immediately recall and repeat a series of numbers presented to him verbally. Each subject was asked to repeat the numbers in the order in which they were presented, or in reverse order. Sleep deprivation research conducted by Williams, et al., (1966) indicates that fatigue negatively affects short-term memory. This finding is consistent with an earlier theory (Williams, et al., 1959) that sleep-deprived people experience brief intermittent lapses of attention. Circadian rhythm, on the other hand, appears to be inversely related to short-term memory. Blake, (1967), found that short-term memory changed inversely to body temperature. The Digit Span Test of short-term memory was included here to test this inverse relationship.

Baddeley Grammatical Transformation Test. This is a simple reasoning test involving the understanding of sentences of various levels of syntactic complexity (Baddeley, 1968). Each item of the test consists of a short sentence such as "A precedes B" or "B follows A" followed by a pair of letters ("AB" or "BA"). The short sentence claims to describe the order of the letters which follow. Each subject was required to read the sentence and the pair of letters which followed, then decide whether the sentence accurately described the letter pair. He was required to check the "True" category if the sentence accurately described the letter pair, and "False" if the sentence did not. For example:

1) "A follows B" – "BA" **True**
2) "B precedes A" – "AB" **False**

Performance on this test has been shown to be sensitive to the intoxicating effects of breathing air at high pressure and to be a measure of workload during driving, (Baddeley, 1968). It was used for this study at the recommendation of Dr. Paul Naitoh of the Naval Health Research Center, who has used this test in other fatigue research.

2.5.4 PHYSIOLOGICAL MEASURES

Each subject was fitted with an unobtrusive multi-channel battery-powered instrumentation tape recorded. The tape recorder was worn on the waist and could run for 24 hours unattended. Each channel, once calibrated, could record different physiological and activity measures. The equipment did not intrude on the subject's movement and the associated electrode connections did not cause discomfort or hinder movement. A more complete description of the instrumentation can be found in Appendix F. The following physiological parameters were monitored.

Muscle Tenseness. Continuous recordings of muscle potential (EMG) were made. The inclusion of this measure is based on the results of studies (Travis and Kennedy, 1948; Kennedy and Travis, 1948) which utilized tasks simulating lookout performance and continuous tracking.

In their early studies, Kennedy and Travis found that when college students were required to react to an intermittent warning light, while simultaneously operating a hand control in a tracking task, the reaction times of the subjects were positively correlated with muscle tension measured from the frontalis. Specifically, they found that average reaction time slowed as tension level decreased; variability in reaction time increased as muscle tension decreased; and the number of failures to detect warning lights increased as muscle tension decreased. Thus, muscle tension appeared to be an indicator of alertness.

EMG was monitored during the entire watch duty. Data were summarized in 5.5 minute segments for close scrutiny and were collapsed into larger time segments for examining grosser effects.

Heart Rate and Heart Rate Variability. The use of heart rate (EKG) and heart rate variability (sinus arrhythmia) has been widely investigated as a measure of physical and mental workload. One of the more relevant studies (O'Hanlon, 1971) investigated the relationship between heart rate variability and driver performance. Data collected on a round-the-clock basis for five days showed that heart rate variability was related to driver alertness/fatigue and fluctuated substantially after the occurrence of momentary events which re-alerted the driver. The more fatigued the individual, the higher the variability; the more alert or diverted the attention of the individual, the lower the variability.

A series of studies conducted at the Naval Postgraduate School (Coons, 1977) further substantiates earlier findings that heart rate is related to signal detection performance and time on watch. Coons found that as the number of minutes engaged in a vigilance task increased to a maximum of 40 minutes, the heart rate and percentage of correct detections decreased.

Heart rate and heart rate variability were collected during the entire watch and summarized in the same manner as EMG.

Physical Movement. This measure gave an estimate of the amount of movement around the bridge as the subject performed various duties. It was expected that fatigued individuals would move less than those who were alert.

Physical movement data were collected and summarized in the same manner as EMG and EKG. Idiosyncratic behavior such as excessive pacing, foot or arm tapping, etc., were noted.

Body Temperature. Oral temperature was recorded every hour (when feasible) prior to the start of the experiment. It was expected that variation in temperature would reflect the circadian rhythm of the individual. The use of body temperature in this manner has been widely substantiated in the literature (see Colquhoun, 1971; Aschoff, 1976; Folkhard, et al., 1976). Body temperature was also recorded before the watch, mid-way through the watch, and immediately after the watch.

2.6 PROCEDURE

Each subject reported at 0700 on Day 1. Day 1 was devoted to a three-hour familiarization watch.

The familiarization watch was very similar to the experimental watch except for the traffic and instrument failure. During the familiarization watch, only two ships appeared on the horizon. Physiological monitoring and pre-, mid-, and post-watch testing occured during the familiarization watch just as they did during the experimental watch.

After the familiarization watch the subjects were taken to dinner and a movie. No alcoholic beverages were permitted. After the movie (approximately 10:30 PM), the sleep group subjects were brought to their bedrooms and put to bed at 11 PM, and awakened at 6:30 AM. The non-sleep group subjects were brought to a game room where they were permitted to play pool, ping pong, video games, pinball, board games, and bowl (limited) until 6:30 AM of Day 2. At 6:30 AM of Day 2, subjects were fed breakfast and brought to the CAORF building. Day 2 was devoted to experimental watches. The four-hour experimental watches (0800 to 1200 and 1200 to 1600) included the four incidents previously described. Dependent measures were collected pre-, mid-, and post-watch. (See Table 1.)

CHAPTER 3

RESULTS

The information presented in this chapter includes the test subjects' subjective responses to the experiment, pertinent biographical data, personality profiles, and a detailed analysis of the dependent measures of work and work-related performance, cognition, affect, and physiological activity. Because of the voluminous amount of information collected, this section reports only the most salient results from each of the performance measure categories.[1] It is organized such that information obtained during the post-experimental debriefing interview is presented first, so that knowledge of the subjects' subjective experiences will be available to the reader when the subsequent evaluations of performance measures are presented. Following the debriefing data, a personality profile of the subjects is presented. The final three sections report the evaluation of the four major categories of performance: work performance; work-related performance, psychological state, including measures of cognition and affect; and physiological measures.

3.1 POST-EXPERIMENTAL INTERVIEW

This interview was guided by three purposes: (1) to evaluate the realism of the simulated watch; (2) to determine mates' personal evaluation of the effects of the independent variables; and (3) to obtain information regarding how the mates coped with fatigue and boredom in actual watches at sea. A copy of the interview appears in Appendix C. A summary of its results appears in Table 3.

3.1.1 EXPERIMENTAL REALISM

In general, greater experimental realism leads to greater external validity, giving one more confidence to apply the results of a study to real life. It may be inferred from responses to the first few questions of the interview that the conditions of this experiment were indeed realistic. Twenty-four of the twenty-five subjects responded positively to the first question, indicating that the experimental task resembled something that they would encounter in real life situations. Similarly, all but two respondents reported that they would sometimes be required to stand a watch without having slept the night before. About half the mates, however, indicated that some of their activities during the simulated run differed from real life. The most common explanations for this were that they would ordinarily have had other work to do during the watch (paperwork) and that there would have been a helmsman there to talk to. (The helmsman was purposefully absent during the simulation to induce a more monotonous watch. While removal of the helmsman is not typical in real life, it does occur on ships during periods of heavy deck work.) A few of the subjects reported that they did not feel the same sense of responsibility they normally would at sea. This, unfortunately, is an unavoidable byproduct of simulation.

For all subjects combined *30% reported that they found themselves dropping off to sleep during the simulated watch, and 17% said that they actually slept for some length of time*. Apparently this is not uncommon in real life; more than half of the subjects (57%) said that they either tend to catch themselves dropping off or actually falling asleep while standing watch aboard their own ships.

When analyzed by group, these items reveal that it was *only the no-sleep group that dropped off to sleep or fell asleep* during the experiment; none of the mates in the full-sleep group did. However, when standing on actual watch on a ship, about half of both groups reported falling asleep.

Responses to Questions 22 and 23 regarding daydreaming were quite similar. Subjects reported daydreaming about 29% of the time during the experimental watch, as compared to 21% in reality. Three mates indicated that they daydreamed as a means of alleviating boredom (Question 11).

[1] In all presentations of results, an alpha level of 0.10 was used to identify statistically significant results. In traditional psychological research the 0.05 alpha level is used; however, in the maritime field, 0.10 is sometimes used.

TABLE 3. SUMMARY OF RESULTS OF POST-EXPERIMENTAL DEBRIEFING QUESTIONNAIRE

Quest. No.	Abbreviated Question	Response Category	Experimental Group Responses			
			Full Sleep Morning	Full Sleep Afternoon	No Sleep Morning	No Sleep Afternoon
1	Task resemble real life	Yes	6	6	5	7
		No	0	0	1	0
2	Stand watch without sleep real life	Yes	5	5	5	5
		No	0	0	0	1
		Missing	1	1	1	1
3	Behave differently than normal	Yes	3	5	1	4
		No	3	1	5	3
4	Rate realism of CAORF experience	1 (poor)	0	0	0	0
		2	0	0	0	0
		3	1	2	3	3
		4	5	3	2	2
		5 (excellent)	0	1	1	2
5	Was task-challenging	Yes	1	0	1	3
		No	5	6	5	4
6	Rate the difficulty of task	Easy	3	5	3	2
		Neutral	2	1	1	4
		Difficulty	1	0	2	1
7	Performance change over time	Yes	4	3	4	5
		No	2	3	2	2
8	Bored during experiment	Yes	6	6	6	5
		No	0	0	0	2
9	How long did it take to become bored	Not bored	0	0	0	2
		15 min or less	2	3	1	0
		20-30 min.	0	3	2	1
		1 hour	1	0	0	1
		1½ hours	2	0	1	1
		2 hours	1	0	1	1
		2½ hours	0	0	1	0
		Missing	0	0	0	1
10	When were you most bored during experiment	Not bored	0	0	0	2
		1st hour	0	2	0	0
		½ way	1	0	0	1
		3rd hour	0	1	1	1
		4th hour	1	0	1	0
		Periods of no activity	2	1	3	3
		Missing	2	2	1	0

TABLE 3. SUMMARY OF RESULTS OF POST-EXPERIMENTAL DEBRIEFING QUESTIONNAIRE (cont'd)

Quest. No.	Abbreviated Question	Response Category	Experimental Group Responses			
			Full Sleep Morning	Full Sleep Afternoon	No Sleep Morning	No Sleep Afternoon
11	Did you do anything to alleviate boredom	Read	2	3	1	1
		Checked equipment	2	3	1	1
		Paced	2	1	2	1
		Consulted charts	2	0	1	0
		Daydreamed	1	0	1	1
		Drank coffee	1	1	1	0
		Smoked	0	1	2	0
		Listened to weather	1	0	0	1
		Doodled	0	1	1	0
		Nothing	1	1	1	0
		Not bored	0	0	0	2
		Missing	0	0	0	1
12	Fatigued during experiment	Yes	1	–	4	5
		No	4	4	1	1
		Missing	1	2	1	1
13	How long before fatigued	Not fatigued	4	4	1	1
		0 hours	–	–	2	3
		1 hour	1	–	–	–
		4 hours	–	–	–	1
		Missing	1	2	3	2
14	When most fatigued during experiment	Not fatigued	4	4	1	1
		Before watch	–	–	–	1
		2nd hour	–	–	2	2
		½ way	1	–	1	1
		3rd hour	–	–	–	1
		4th hour	–	–	1	–
		Missing	1	2	1	1
15	Do anything to alleviate fatigue	Not fatigued	4	4	1	1
		No	–	–	2	2
		Paced	–	–	1	2
		Drank coffee	1	–	1	1
		Smoked	–	–	1	-
		Concentrated harder	–	–	1	–
		Missing	1	2	1	1
16	Rate difficulty of staying awake	Easy	5	4	3	–
		Neutral	–	–	–	5
		Difficult	–	–	2	1
		Missing	1	2	1	1

TABLE 3. SUMMARY OF RESULTS OF POST-EXPERIMENTAL DEBRIEFING QUESTIONNAIRE (cont'd)

Quest. No.	Abbreviated Question	Response Category	Experimental Group Responses			
			Full Sleep Morning	Full Sleep Afternoon	No Sleep Morning	No Sleep Afternoon
17	Which of the following influenced your performance	Fatigue	1	–	4	5
		Boredom	6	4	4	3
		Ill health	–	–	–	1
		Hunger	1	1	–	–
		Poor spirits	–	–	–	–
		Physical discomfort	–	–	–	2
		Lack of interest	–	–	1	–
		None of the above	–	2	2	2
18	How much sleep did you get last night	0 hours	–	–	6	7
		4½ hours	–	1	–	–
		5 hours	1	–	–	–
		6 hours	1	–	–	–
		6½ hours	2	–	–	–
		7 hours	2	1	-	–
		7½ hours	–	1	–	–
		8 hours	–	2	–	–
		9 hours	–	1	–	–
19	Drop off to sleep during experiment	Yes	–	–	3	4
		No	5	5	2	3
		Missing	1	1	1	–
20	Did you actually sleep during experiment	Yes	–	–	3	–
		No	5	5	3	7
		Missing	1	1	–	–
21	Drop off in real life	Yes	3	3	4	4
		No	2	2	1	3
		Missing	1	1	1	-
22	Daydream during experiment	Yes	6	4	3	5
		No	–	1	2	2
		Missing	–	1	1	–
23	What percent of time in real life spent daydreaming	0	–	–	–	2
		5-10%	3	5	3	–
		20-25%	2	–	–	3
		30%	–	1	–	–
		40%	1	–	1	-
		50%	–	–	2	1
24	Mind blank during experiment	Yes	2	1	4	1
		No	4	4	2	6
		Missing	–	1	·	–
25	Mind blank real life	Yes	5	4	5	3
		No	1	1	1	4
		Missing	–	1	–	–

Mates were asked if they experienced "mindblanking," a state in which a person is awake, with his eyes open, but in which his mind is totally blank. About 57% reported that they do experience mindblanking at sea, while half as many (30%) experienced this state during the experiment. For most subjects, this state occurred between one and five times during the course of watchstanding; for one subject in particular, mindblanking occurred quite frequently, reportedly once each five-minute period.

3.1.2 SUBJECTIVE EFFECTS OF INDEPENDENT VARIABLES

A second purpose of this questionnaire was to gather subjective information about the effects of the independent variables: hours of sleep, time of watch, and number of hours worked. Of the entire sample, about half the subjects (47%) indicated that they felt fatigued. Not surprisingly, these mates all belonged to the no-sleep group. Mates were asked in Item 17 if any of a list of factors influenced their work performance during the simulation. Only one mate in the full-sleep group (8.3%) versus nine mates in the no-sleep group (69%) felt that their performance was impaired due to fatigue. As discussed earlier, none of the mates dropped off to sleep at any time during the watch; only mates from the no-sleep group did.

The mates were also asked if they perceived their performance changing over time. Sixty percent responded positively, and forty percent negatively. More mates in the no-sleep group (69%) than in the full-sleep group (50%) perceived a change in their performance over the course of the watch, although this is not a statistically significant difference. In most cases, mates explained that their alertness or concentration either fluctuated or decreased; a few reported that they became fatigued or bored. Only two subjects became more alert over the course of the watch. Both stood the afternoon watch; one had slept the night before and the other had not. (These two mates represent only eight percent of the entire sample.)

Boredom was experienced almost universally among the participants; only 2 mates (8%) of the 25 indicated that they were **not** bored. Boredom set in anywhere from five minutes to two and one-half hours into the watch; the mean time was 54 minutes. Responses to Item 17 indicated that boredom influenced the work performance of 68% of the subjects.

As may have been expected, subjects were most bored during the periods of no traffic or other activity. This finding is not surprising due to the monotony of the visual scene and the simplicity of the experimental task at these times. Many subjects indicated they felt that the absence of their usual paperwork and a helmsman to talk to increased the monotony.

3.1.3 METHODS OF COPING WITH FATIGUE AND BOREDOM

The final purpose of this interview was to determine how mates cope with the problems of fatigue and boredom. It was expected that daydreaming and mindblanking might be two techniques used to alleviate problems associated with fatigue and boredom. Daydreaming was, in fact, utilized by 78% of the sample. Mindblanking was less prevalent, but was still used by 30% of the group.

Those who tried to fight the effects of fatigue tended to drink coffee or move about the bridge. To alleviate boredom, most subjects amused themselves by checking the instruments, investigating the other equipment and familiarizing themselves with the bridge. Others read whatever was available (magazines, Rules of the Road), consulted the charts, smoked, paced, or drank coffee.

3.1.4 SUMMARY OF EXPERIMENTAL DEBRIEFING

In summary, the simulation appeared realistic in most ways. Only the lack of distractions, such as auxillary paperwork or a talkative helmsman, were missing. The experimental manipulations were effective; that is, the fatigued group reported less alertness and, in fact, fell asleep far more frequently than the well-rested group. Furthermore, subjects felt as though their performance degenerated over the course of the watch. Daydreaming was used frequently as a means of alleviating boredom; mindblanking was also used, but with less frequency. Overall, the simulated watch was reported to be sufficiently realistic to make results generalizable to the real world.

3.2 BIOGRAPHICAL DATA AND PERSONALITY PROFILES

During the familiarization day, all subjects were asked to provide pertinent biographical data and to take several personality tests. One reason for the collection of this information was to insure that the experimental groups did not differ from each other with respect to characteristics, other than the independent variables, which might influence their performance. The biographical data, such as number of years experience as a watchstander, were

TABLE 4. GROUP MEANS AND STANDARD DEVIATIONS FOR BIOGRAPHICAL AND PERSONALITY MEASURES

	Full Sleep Morning		Full Sleep Afternoon		No Sleep Morning		No Sleep Afternoon	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age[a]	32.2	7.12	28.667	3.933	26.333	1.633	34.00	9.00
No. of Years Experience as Second Mate	2.933	1.846	1.833	.289	1.5	1.0	3.15	3.971
No. of Years Experience as Deck Officer	5.833	4.107	6.067	3.479	3.5	1.549	6.486	4.143
Jenkins A Scale	213.333	45.311	116.667	46.792	239.5	88.897	216.286	81.414
Jenkins S Scale	188.167	47.368	138.50	72.957	184.167	71.315	181.571	68.417
Jenkins J Scale	264.833	45.923	257.167	52.105	229.5	33.697	262.571	59.966
Jenkins H Scale[b]	127.833	21.367	162.667	13.064	112.5	22.950	103.143	33.746
Life Experiences Survey	4.5	6.686	3.178	6.178	2.5	3.507	1.857	5.275
Life Experiences Survey Locus of Control Scale	9.333	5.715	9.00	2.966	11.00	3.688	9.143	5.429

[a]Significant Interaction, $F(1,20) = 4.76$, $p = 0.04$
$F = 4.760$, $df = 1/20$, $p = 0.04$
[b]Significant Main Effect of Watch, $F = 2.99$, $df = 1/21$, $p = 0.10$
$F(1,21) = 2.99$, $p = 0.10$

selected based upon common sense. The personality instruments were selected because previous research indicated that certain personal characteristics may mediate response to fatigue. A second reason for collecting this information was, in the absence of experimental effects, to further evaluate the influence of personal characteristics on the performance measures.

Two-by-two analyses of variance were performed on all biographical and personality data. These analyses revealed no significant differences between groups for number of years experience as a second mate, or total number of years as a deck officer. However, a significant interaction for age was found, $F(1,20) = 4.76$, $p=0.04$. Group means and standard deviations for these and the personality measures appear in Table 4.

Three personality tests were selected for use in this study. These are: The Jenkins Activity Survey, the Life Experiences Survey, and the Rotter Internal-External Locus of Control Scale.

The Jenkins Activity Survey is a measure of "Type A" behavior, or coronary-prone behavior. This type of behavior pattern is characterized by "extreme competitiveness, striving for achievement, aggressiveness, impatience, haste, restlessness, and feelings of being challenged by responsibility and under the pressure of time," (Jenkins, et al., 1979). In addition to a general measure of the degree to which an individual exhibits the Type A behavior pattern (Scale A), the Jenkins test further yields scores on three scales that measure specific behaviors typical of the Type A personality: the Speed and Impatience Scale (Scale S), which refers to one's sense of time urgency, temper, and irritability; the Job Involvement Scale (Scale J), which measures one's degree of dedication to occupational activity; and the Hard-Driving and Competitive Scale (Scale H), which looks at the extent to which one is hard-driven, conscientious, responsible, serious, and competitive.

It was only the latter (Scale H) which showed significant differences between experimental groups. Those assigned to the afternoon-watch groups were more hard-driving and competitive than those in the morning-watch groups, $F(1,21) = 2.99$, $p=0.10$. Scores on the remaining scales yielded non-significant results.

The Life Experiences Survey was given in order to determine if groups differed significantly in terms of recent occurrences in their personal lives which may produce stress and thereby affect performance on the experimental task. It was included here because previous research has shown a positive relationship between the recency of significant life experiences (such as recent betrothel, recent death of family member) and the probability of being involved in an aircraft accident (Alkov and Borowsky, 1980). A similar relationship was found between the recency of significant life experiences and probability of becoming ill (Rahe, 1974). This research utilized the Life Experiences Survey (LES) developed by Sarason, Johnson and Siegel (1978). The LES elicits information from each subject regarding the recency and impact of 47 different life experiences, and produces results in the form of a positive change score and a negative change score. The LES does not appear to be influenced by mood state (Siegel, Johnson and Sarason, 1979) or social desirability (Sarason, Johnson and Siegel, 1978). When administered to 58 undergraduates at six week time intervals, the test-retest reliability scores using Pearson correlations were 0.53 for the positive change score, .88 for negative change, and 0.64 for total change (Sarason, Johnson and Siegel, 1978).

There were no significant effects found on these scores; that is, the experimental groups were no different with respect to recent life changes.

The third personality test administered was the Rotter Internal-External Locus of Control Scale, which is used to distinguish between "internal personality types" (those who believe they have control over their destinies) and "external personality types" (those who believe their outcomes are determined by factors extrinsic to themselves). This scale was selected because previous research has demonstrated that "internals" perform better on vigilance tasks than "externals" (e.g., Bakan, Belton, and Toth, 1963). The Rotter I-E scale is based on a 29-item forced-choice test including six filler items. The internal consistency measured by Kuder-Richardson coefficients varies from approximately 0.70 to 0.76. The test-retest reliability with a one-month interval (administered in a group setting) has been reported at 0.60 for males and 0.83 for females. The Rotter I-E Scale does correlate with social desirability in a range from -0.07 to -0.35 (Rotter, 1954).

Once again, no significant differences were found between groups with respect to this concept of locus of control.

In sum, with the exception of the Jenkins Scale H and age, there were no real differences found between groups in terms of pertinent biographical information and personality types.

3.3 WORK PERFORMANCE

3.3.1 DETECTION OF SHIPS AND INSTRUMENT FAILURES

Four incidents occurred during the course of the four-hour watch. During the first hour and the fourth hour, either a containership or a very large crude carrier (VLCC) appeared on the horizon passing either to the left or right of ownship. The type of ship and direction of passage were counterbalanced across the four experimental groups. During the second hour and the third hour, a destroyer approached ownship from astern and the RPM indicator on the bridge failed. The position of these latter two incidents was systematically reversed between the second and third hours to achieve counterbalancing across groups. The amount of time taken to detect each of these incidents was used as a dependent measure of work performance.

Given the fact that each type of traffic ship (containership, VLCC, destroyer) was traveling at a different speed, a transformation of detection times was made so that comparisons could be made irrespective of ship speed. The resultant data reflect the time between the first possible detection (on radar) and the time of actual detection. For example, a detection time of two minutes indicates that the mate saw the ship (either visually or on radar) two minutes after it first appeared on radar. Since visibility was only to seven miles with the naked eye and the radar was set at the 12 mile range, the earliest sightings could only occur by radar. In fact, all but one of the sightings did occur by radar first.

3.3.2 DIRECTION OF SHIP PASSAGE

The direction of ship passage was not of primary interest in this study; however it was counterbalanced to control for any spurious effects resulting from it. To examine the

TABLE 5. DETECTION TIMES (IN MINUTES) OF FIRST SHIP AND LAST SHIP BY EXPERIMENTAL GROUP

	First Ship			Last Ship			Difference Between First and Last[1]		
	Mean	SD	n	Mean	SD	n	Mean	SD	n
Full Sleep Morning Watch	1.00	1.67	6	0.20	0.45	5	0.20	0.45	5
Full Sleep Afternoon Watch	4.00	5.40	6	2.17	4.40	6	1.83	7.68	6
No Sleep Morning Watch	4.67	4.50	6	3.50	3.99	6	1.17	4.58	6
No Sleep Afternoon Watch	5.57	5.26	7	3.83	6.88	6	2.33	6.09	6
All Groups Combined	3.95	4.75		2.52	4.56		1.43	5.22	23

[1] The difference scores reported in this column may not reconcile with those in preceding columns due to the exclusion of missing data when computing difference scores. This is evident where the n's for first ship and last ship do not match.

effect of direction of ship passage on time to detect traffic ships, a t-test was computed using direction as the independent variable and detection time as the dependent variable. Separate t-tests were computed for the first sighting and the last sighting. Since detection times had been measured in such a way as to make comparisons across ship types possible, the first and last sightings included data from both the VLCC and the containership. For the first sighting, $t = 0.2883$, $p = 0.76$ for 25 observations. For the last sighting, $t = 0.6544$, $p = 0.52$ for 23 observations. The reduction in the number of observations was due to the exclusion of unreliable data collected in the last hour. The substantially low values of these t statistics indicate that direction of ship passage does not affect detection time in any systematic way. Therefore, direction of ship passage was eliminated as an experimental factor in subsequent statistical tests.

3.3.3 CHANGE IN DETECTION TIMES FROM BEGINNING TO END OF WATCH

To assess the main effect of hours worked on detection time and its interaction with fatigue and time of watch, two levels of analysis were undertaken. The first analysis simply compared the detection time of the first incident to the detection time of the last incident, collapsing across all sleep and watch conditions. Data from the second and third hours were not used in this comparison since the incidents occurring in those hours (namely, an instrument failure and a destroyer approaching astern) were not directly comparable to the incidents occurring in the first and fourth hours. In addition, one data point from the fourth hour was missing due to an experimental error. A comparison of the first and fourth hours using a correlated t-test yielded statistically insignificant results, $t(22) = 1.32$, $p = 0.20$. The means and standard deviations used in this comparison can be found in Table 5 (and in Appendix D, Table D1). Based on this analysis, it appeared that, in general, the time to detect traffic vessels did not change significantly as the watch progressed. However, the direction of the means indicated that response times were more rapid in the fourth hour.

A second level of analysis was undertaken to see if detection times changed differently through the watch for the various sleep X watch conditions; this interaction was tested using a 2 X 2 analysis of variance. The independent variables were sleep condition and watch condition and the dependent measure was the difference score obtained by subtracting the last incident's detection time from the first incident's time. (See Table 5 and Figure 3). The analysis of variance showed no significant main effects or interactions (see Appendix D, Table D2). The lack of statistical significance is due, in part, to the large variability associated in particular with the no-sleep afternoon group.

3.3.4 INDIVIDUAL INCIDENT ANALYSIS

The objective of this set of analyses was to determine the influence of fatigue and time of watch on the detection of specific incidents. Four separate 2 X 2 analyses of variance were computed using sleep condition and watch condition as the independent variables, and time to detect each of four incidents as the dependent measure. The four incidents were as follows: (1) The first incident referred to as "first ship sighting" was the sighting during the first hour of either a containership or VLCC off either the port or starboard bow. Data were collapsed across ship type and

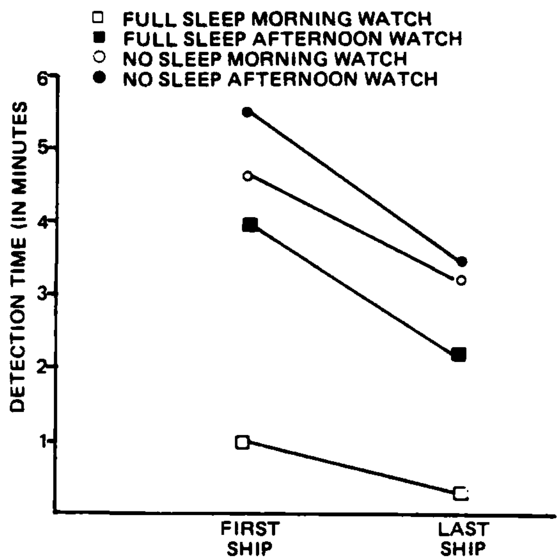

Figure 3. Detection Times (in Minutes) of First Ship and Last Ship by Experimental Group

direction of passage. (2) The "destroyer sighting" incident refers to the detection of a destroyer approaching from astern. Sometimes this occurred during the second hour and other times during the third hour. To break down the analysis of the destroyer sighting by hour of appearance would have yielded too small a sample size to permit statistical analysis. Therefore, the analysis of the destroyer sighting data, for the most part, disregards its position in the time sequence. (3) The "instrument failure" incident tested the mate's response to an RPM indicator failure. This incident occurred either during the second or third hour, complementing the order of the destroyer sighting. (4) The fourth incident referred to as the "last ship sighting" was a repetition of the first; however, it occurred during the last hour of the watch. Results for each of these incidents are reported below.

1. With respect to the **first ship sighting** the fatigued group showed a tendency to take longer to detect the ship in comparison to the well-rested group. While this main effect for sleep condition was not statistically significant ($F(1,21) = 2.08$, $p = 0.16$), the means for this particular incident (as well as for the other ship sightings) indicate a tendency for fatigue to have a deleterious effect on the ability to detect ships (see Figure 3). For example, the full-sleep mornirn group took an average of 1.00 minute to detect the first ship in comparison to the no-sleep afternoon group which took an average of 5.57 minutes. If the traffic ship were traveling at 20 knots, this difference of 4.57 minutes would translate into a distance of approximately 1.5 nautical miles further before detection.

An interesting and somewhat unexpected finding was the relationship of ship type and detection time during the first hour. When detection time associated with the VLCC was compared to detection time associated with the containership, the $t = 1.46$, $df = 23$, $p = 0.18$. This indicates that the detection times were somewhat longer for the VLCC traveling at 12.0 knots as compared to the containership traveling at 20.4 knots. This relationship was not demonstrated in the fourth hour where $t=0.0183$, $df = 21$, $p = 0.98$. Since ship type was responsible for some of the variability in detection time, the 2 X 2 analysis of variance was recomputed extracting ship type as a covariate. The results of this analysis show an increase in the size of the F ratio to 2.737 ($p = 0.114$), thus strengthening the argument that sleep loss deleteriously affects detection time. A comparison of Tables D3 and D4 in Appendix D provides more detailed information.

To summarize the first ship sighting, while the 2 X 2 analyses of variance were not statistically significant at an alpha level of 0.10, the means for the experimental groups are ordered such that the fatigued groups took longer to detect the first ship than the well-rested group. In particular, the no-sleep afternoon group performed the worst. The lack of statistical significance may be due to the rather large variability associated with most of the groups' performances.

2. **Detection of the destroyer** took longest for the no-sleep afternoon group (5.57 minutes), followed by the full-sleep afternoon group (2.83 minutes), no-sleep morning (1.33 minutes), and full-sleep morning (1.00 minute) groups. As in the previous incident, the high within-group variability associated with the no-sleep afternoon group precluded finding statistical significance. Nevertheless, a trend towards a main effect on watch condition was found indicating that the afternoon watchkeepers were slower in detecting the destroyer than the morning watchkeepers, $F(1,21) = 2.51$, $p = 0.13$. Figure 4 depicts these relationships.

In fact, the difference between the average detection times of the fatigued afternoon group and the well-rested morning group was 4.57 minutes. This difference in average group detection times was consistent with the first ship sighting. Since the destroyer was traveling at 29.7 knots,

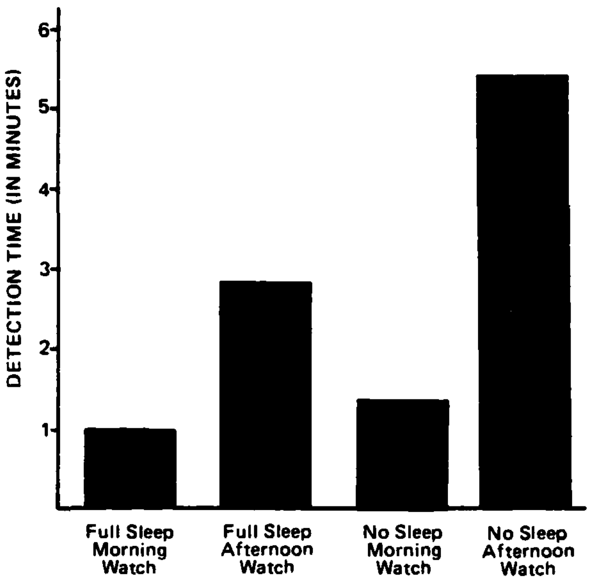

Figure 4. Detection Time (in Minutes) of Destroyer By Experimental Groups

this detection differential means that the destroyer traveled an average distance of 2.25 nautical miles further before the fatigued afternoon group detected it.

To insure that the hour in which the destroyer appeared did not covary with the relationships of detection time to either sleep or watch conditions, the 2 X 2 analysis of variance used to test the effects of sleep and watch on detection of the destroyer was recomputed partialling out the effects of the order of presentation. The results of this analysis were almost identical to those of the previous analysis. The summary tables from these analyses appear in Appendix D, Tables D5 and D6.

To summarize the results of the destroyer sighting, a weak effect of watch condition was found indicating that the afternoon watchkeepers were slower to detect the destroyer than the morning watchkeepers. In particular, the fatigued afternoon group was slowest.

3. Table 6 summarizes the response of each experimental group to the **RPM indicator failure.** While the proportions of watchstanders who noticed the failure did not vary appreciably between groups, of those who did notice the failure, the no-sleep afternoon group took the longest time to detect the failure. Specifically, the no-sleep afternoon group allowed an average of 24.80 minutes to pass before noticing the instrument failure compared to a range of 9.33 to 10.67 minutes for the other three groups.

TABLE 6. RESPONSES TO RPM INDICATOR FAILURE BY EXPERIMENTAL GROUPS

	Proportion of Group that Noticed RPM Indicator Failure					
	Sleep			No Sleep		
AM Watch	0.67 4/6			0.50 3/6		
PM Watch	0.50 3/6			0.71 5/7		
	Group Detection Times (In Minutes) for Subjects Who Noticed RPM Failure					
	Sleep			No Sleep		
	Mean	SD	n	Mean	SD	n
AM Watch	9.75	10.37	4	10.67	10.50	3
PM Watch	9.33	12.74	3	24.80	26.61	5

The cell sizes were too small to assess interactions reliably. Main effects for sleep conditions and watch condition were not statistically significant when tested using a 2 X 2 analysis of variance.

Since the distribution of detection times was substantially skewed, a nonparametric analysis of these data was also performed. Results from a median test yielded no main effects for either sleep or watch conditions.

4. **Detection time for the last ship** did not appear to be affected by either sleep, watch, or the interaction of these factors. Table D7 in Appendix D summarizes these results. The most noteworthy outcome of this incident is that the no-sleep afternoon group demonstrated the most within-group variance, while the full-sleep morning group demonstrated the least. This pattern of within-group variance was repeated across all four incidents and may, in and of itself, be the result of the experimental manipulation.

3.3.5 ANALYSIS OF DETECTION TIMES ACROSS ALL SHIP SIGHTINGS

Finally, detection performance over the entire watch was analyzed by a two-way analysis of variance (ANOVA) using sleep and watch conditions as the independent variables,

TABLE 7. ANALYSIS OF VARIANCE SUMMARY TABLE FOR DETECTION TIME (IN MINUTES) AVERAGED ACROSS ALL THREE SHIPS

Source	Sum of Squares	df	Mean Squares	F Ratio	P
Sleep Condition	39.52	1	39.52	2.94	0.10
Watch Condition	34.67	1	34.67	2.58	0.12
S x W	0.10	1	0.10	0.01	0.93
Error	255.31	19	13.44		
TOTAL	326.39	22	14.84		

	Full Sleep Morning Watch	Full Sleep Afternoon Watch	No Sleep Morning Watch	No Sleep Afternoon Watch
$\bar{X}$	0.40	3.00	3.17	5.50
SD	0.89	3.02	2.39	5.97
N	5	6	6	6

and the average detection time for each subject across all three ship sightings as the dependent measure. (The detection time of the RPM failure was not included in this average, since the nature of that task was somewhat different from the ship sightings.) The analysis (summarized in Table 7) indicates that the full-sleep groups took less time overall to detect the traffic ships than did the no-sleep groups, $F(1,19) = 2.94$, $p = 0.10$. The morning watch groups also took less time than the afternoon groups, although this difference did not reach statistical significance, $F(1,19) = 2.58$, $p = 0.12$.

Thus, it appears that, while it may be difficult to predict a fatigued mariner's detection efficiency for any particular incident he may encounter, his overall performance during a four-hour watch is likely to be considerably worse than that of a mariner who has had adequate rest. Furthermore, the watchkeeper who stands the afternoon watch appears less effective than his morning counterpart, although this difference between watch periods did not reach statistical significance in this study.

3.4 WORK-RELATED PERFORMANCE

The vigilant watchstander should do everything reasonably possible to maximize the probability of detecting a critical event. Towards this end he should consult the radar and move around the bridge frequently. The measures discussed below address these two work-related behaviors.

3.4.1 RADAR OBSERVATION

Since it was necessary for subjects to depress and hold down a button on the radar unit in order to view the display, radar observation data will be discussed in terms of number of radar presses and duration of presses. The combination of hardware and software which automatically recorded these data failed on a few occasions, resulting in the total loss of four subjects' data and the intermittent loss of other data. A review of the videotapes in an attempt to retrieve these data points led to the conclusion that any frequency and duration data compiled by viewing the videotapes would have been potentially unreliable and suspect. Therefore, the analyses described below are based on reduced sample sizes.

3.4.2 OVERVIEW OF RADAR OBSERVATION

For each half-hour of the four-hour watch, the number and duration of presses was calculated. Figures 5 and 6 depict the frequency and duration of radar presses per half-hour period.[1] Tables 8 and 9 present the means and standard deviations for these measures for each experimental group. An examination of the frequencies of radar

[1] Although the four-hour watch was comprised of eight half-hour segments, only the data from the first seven segments was used because the watches of a few subjects were terminated five to ten minutes early.

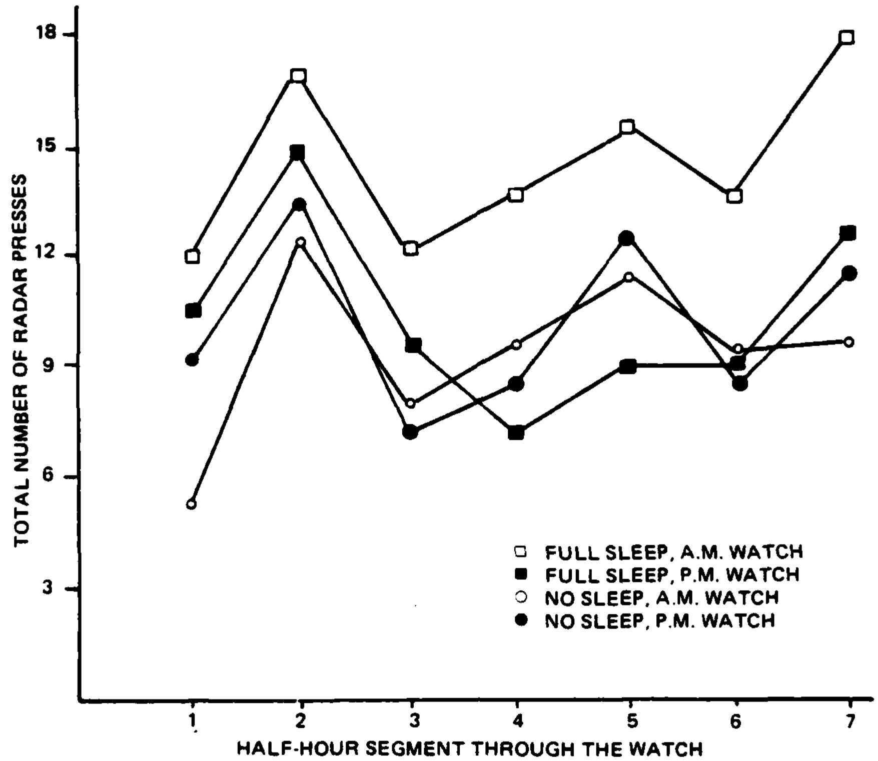

Figure 5. Frequency of Radar Presses in Each Half-Hour Time Segment of the Watch for Each Group

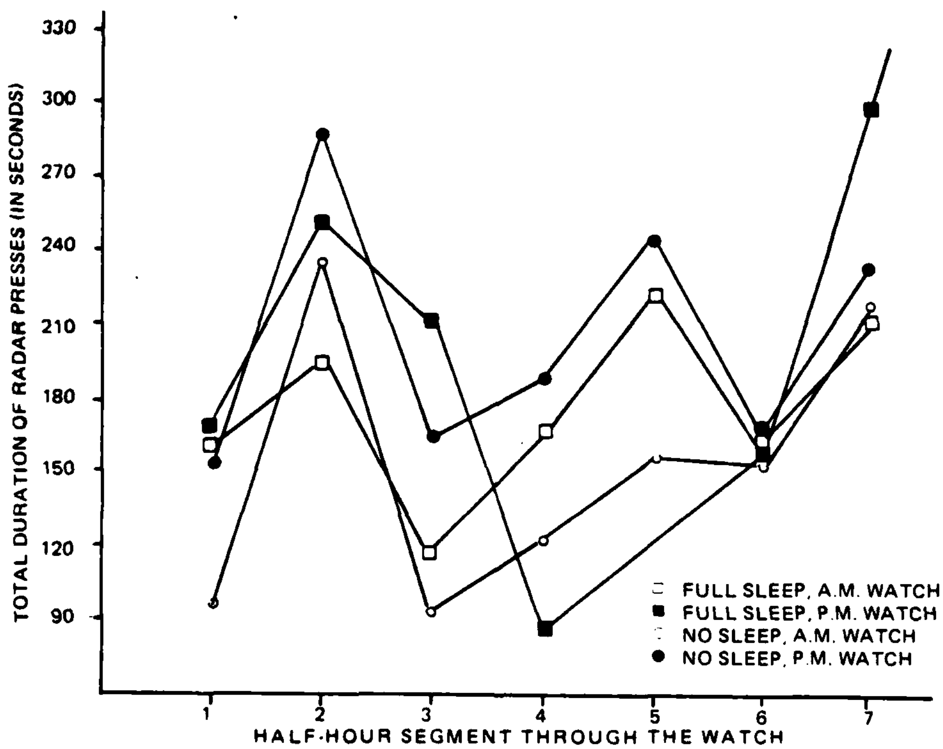

Figure 6. Total Duration of Radar Presses in Each Half-Hour Time Segment of the Watch for Each Group

TABLE 8. FREQUENCY OF RADAR PRESSES FOR EACH HALF-HOUR OF WATCH BY EXPERIMENTAL GROUP

	Full Sleep Morning		Full Sleep Afternoon		No Sleep Morning		No Sleep Afternoon	
Half Hour	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	12.00	5.66	10.20	6.91	5.33	3.78	8.71	9.91
2	17.40	7.70	14.80	4.55	12.50	5.47	13.71	8.12
3	12.00	4.64	9.80	5.36	7.83	4.22	7.43	6.11
4	14.00	3.81	7.00	5.43	9.67	5.28	8.14	6.77
5	15.80	8.58	8.80	8.84	11.67	7.06	12.29	7.76
6	13.60	6.88	8.80	5.89	9.17	4.45	8.57	6.55
7	17.80	7.85	12.60	6.54	9.33	3.61	11.57	7.18

presses reveals that, with one exception, the relationship of the means of all four groups to each other remains fairly constant over the course of the four-hour watch. For most of the watch, the full-sleep morning watch group pressed most frequently, followed by the full-sleep afternoon group, the no-sleep afternoon group, and the no-sleep morning group, in that order. The one exception occurred during the third and fourth half-hour segments, or the second hour, of the watch. At this time, the full-sleep afternoon watch group exhibited a decline in the number of radar presses, while the remaining three groups either increased or remained level. By the end of the watch, all groups were once again following a similar pattern. To test whether sleep or watch conditions influenced the frequency of radar presses at each half-hour of the watch, a series of 2 X 2 analyses of variance was performed. No significant differences between groups were found at the alpha level of 0.10.

Not surprisingly, the duration of radar presses showed a similar pattern; that is, the relationship of the four groups to each other remained relatively constant over the course of the watch, with the exception of the second hour. At this point, the full-sleep afternoon watch group again dropped below the other three, which increased in value. The group means at each half-hour were not significantly different, however. One interesting finding is the fact that, while the full-sleep morning watch had the greatest number of radar presses, this group did not show any extreme scores for duration of radar presses. For most of the watch, the no-sleep afternoon watch group exhibited the greatest duration of radar presses. A further examination of Figures 5 and 6 indicates that radar consultation peaks during the second half-hour of the watch period. One can speculate that the increase in activity was the result of two factors: (1) the introduction of the first traffic ship and (2) some increased level of arousal upon assuming the watch. With

TABLE 9. DURATION OF RADAR PRESSES FOR EACH HALF-HOUR OF WATCH BY EXPERIMENTAL GROUP

	Full Sleep Morning		Full Sleep Afternoon		No Sleep Morning		No Sleep Afternoon	
Half Hour	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	160.40	81.29	169.00	95.35	96.83	88.15	153.43	149.54
2	116.00	42.00	256.40	117.55	236.00	99.07	291.43	219.35
3	116.00	61.51	203.00	181.77	93.33	61.73	163.71	172.95
4	166.00	56.12	83.80	56.46	120.33	56.67	189.00	184.57
5	221.40	120.20	100.76	157.50	97.46	259.29	259.29	219.67
6	170.00	100.16	160.60	104.47	157.33	98.63	177.29	157.19
7	215.60	142.51	326.60	237.68	217.83	164.57	238.29	179.82

respect to the second factor, one might ordinarily expect the increased arousal associated with assuming a watch to occur during the first half-hour; however, in this particular study mates frequently chatted with the helmsman on the bridge during the first 25 minutes of the watch. Once the helmsman was called away to do deck maintenance work (always after 25 minutes) the mate was left alone on the bridge with the expectation that he could be alone for the next 3½ hours of the watch. It is, therefore, speculated that he channeled his arousal into consulting the radar. This increased amount of radar consultation was, of course, concomitant with the introduction of the first traffic ship.

3.4.3 APPROACH TO SPECIFIC ANALYSES OF RADAR OBSERVATION

A more refined analysis was conducted by focusing on smaller samples of radar observation and examining these samples with respect to events in the environment, such as ship sighting. To accomplish this, each subject's four-hour watch was subdivided into five-minute segments, yielding 48 segments of radar data per subject. The total number of radar presses and the total number of seconds, during which the radar was consulted were calculated for each five-minute segment of the watch.

In order to examine radar observation with respect to specific ship sightings or instrument failure, the particular segments of each hour during which the subject reported a ship sighting or instrument failure was noted; this segment will henceforth be referred to as the "detection segment." For instance, if the subject reported the sighting of a destroyer at 0932 of the 0800-1200 watch, the detection segment of the second hour would be Segment 19.

While the detection segment of each hour was the object of close scrutiny, three segments immediately before and three segments immediately after the detection segment were also studied. These "before and after segments" were useful in several ways. For example, a simple difference score between the detection segment and one segment before detection would yield a quantifiable measure of the increment or decrement of radar consultation associated with the time of ship detection. Likewise, a difference score between the detection segment and one segment after detection could be used to indicate how radar observation changes immediately after the detection of a ship. Furthermore, an examination of radar observations two or three segments before detection would show the amount of radar consultation conducted without the stimulation of a ship sighting. This last measure, referred to as the "resting level" of radar observation, seems to be a purer measure of watchstanding vigilance, since it quantifies the degree to which the mate puts himself into a favorable condition for detecting a ship.

A series of statistical tests (mostly analyses of variance) were performed using sleep and watch conditions as the independent variables, and various combinations of detection segments, before detection segments, and after detection segments as the dependent measures. The presentation of results which follows is based on these analyses. Figure 7 and 8 show the experimental group means for the three segments before detection, the segment at the time of detection, and the three segments after detection of each hour of the watch. Figure 7 shows the frequency of radar presses, while Figure 8 shows the duration of radar presses.

3.4.3 RESTING LEVELS OF RADAR CONSULTATION PRIOR TO DETECTION OF AN INCIDENT

As previously stated, radar observation during periods of the watch when there is no unusual event to observe can be considered, at least conceptually, as a good measure of watchstanding vigilance. Figure 9 depicts the group means of radar observation data collected during a five-minute period three segments before detection for each hour of the watch. During the first hour the well-rested morning group appeared to consult the radar more during this segment. However, the differences between groups for measures of frequency and duration were not significant.

In order to compute a measure of average resting level, the three segments before the detection of an incident were then averaged. Figure 10 depicts the average resting level of each experimental group during the watch.

When a comparison was made of average resting level between the beginning and end of the watch, the duration of time spent consulting the radar rose significantly from 14.1 seconds during the first hour to 37.9 seconds during the fourth hour. A paired t-test on these data yielded a $t(20) = -3.36$, $p = 0.003$. An analysis of the frequency of radar presses showed a rise from an average of 1.1 presses during the first hour to 1.8 during the fourth hour, $t(20) = -3.03$, $p = 0.007$. These results indicate a significant increase in radar activity as the watch progresses, irrespective of sleep or watch condition. Such increased activity may be caused by several factors, including increased motivation to perform well, and a desire to counteract what could have been a corresponding increase in boredom. If the latter

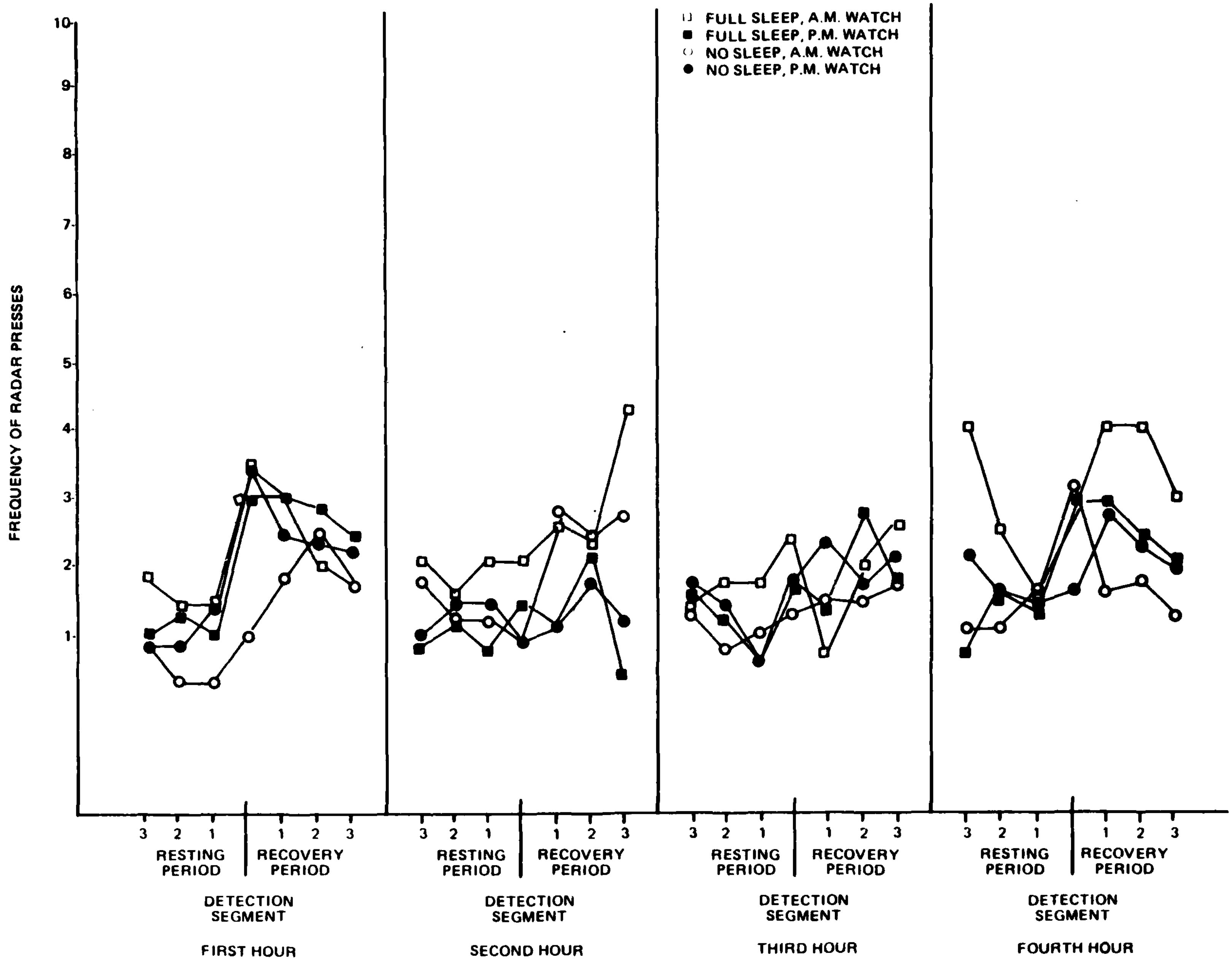

Figure 7. Experimental Group Averages for Frequency of Radar Presses Before, During, and After Detection of Each Hour of the Watch

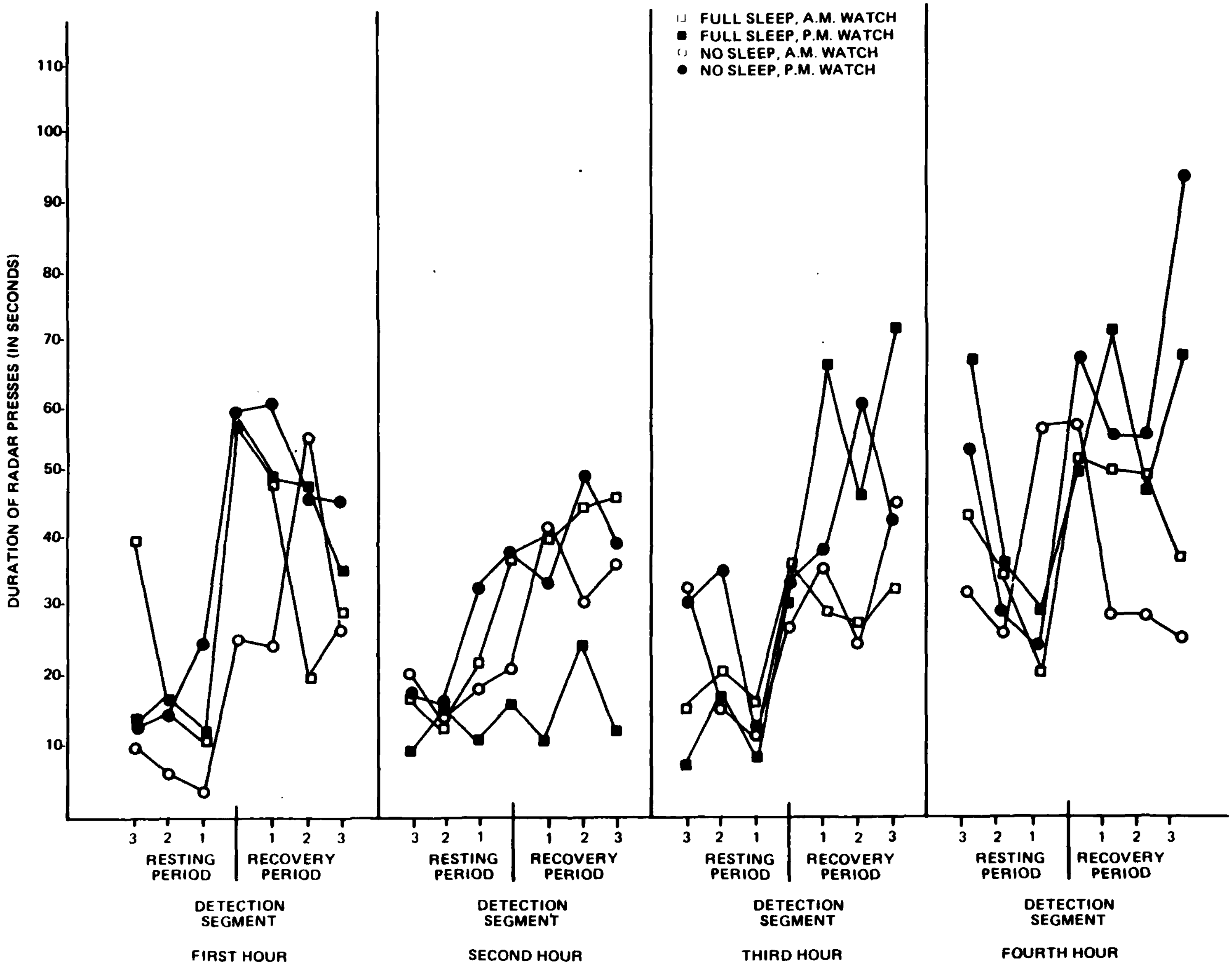

Figure 8. Experimental Group Averages for Duration of Radar Presses

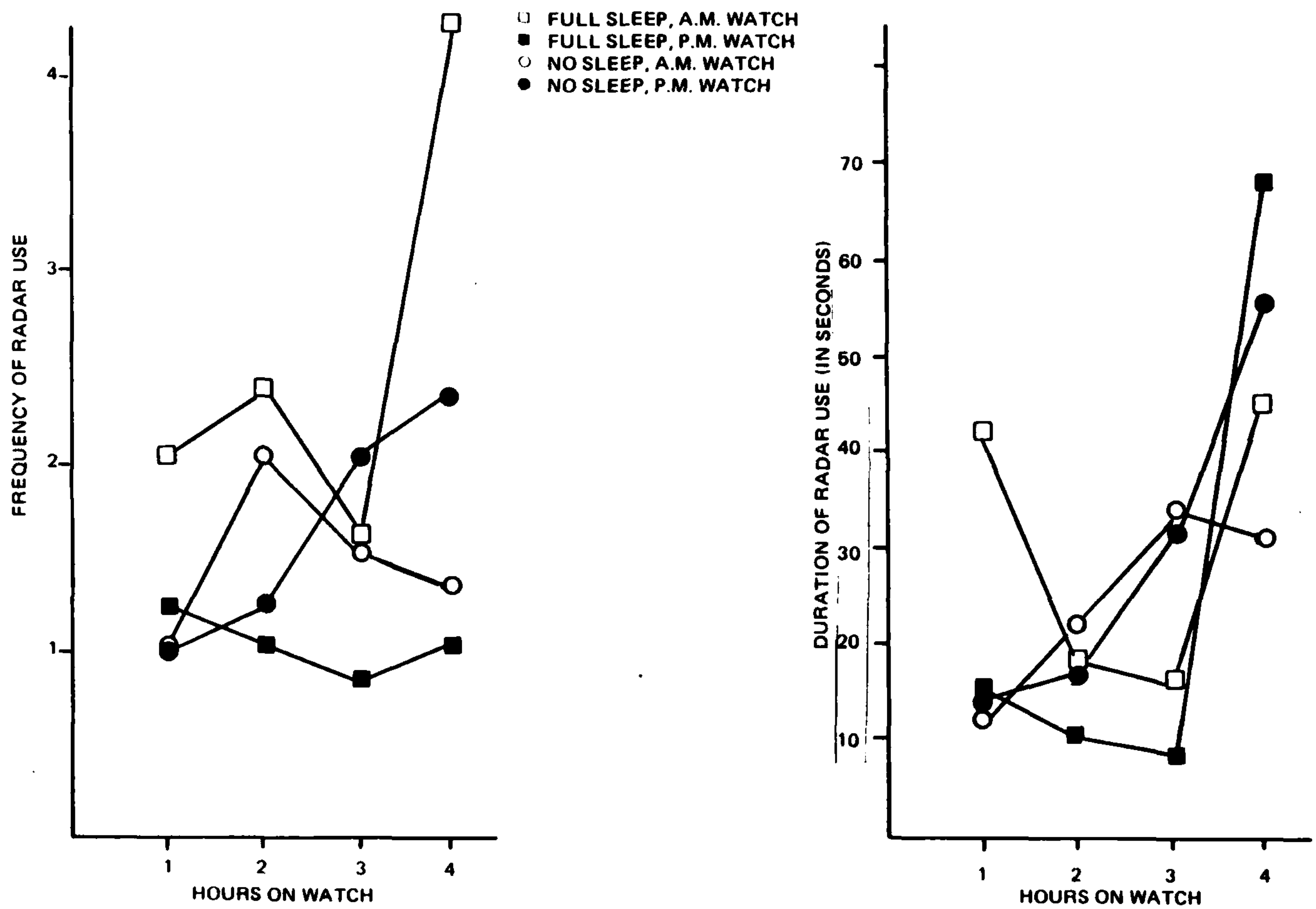

Figure 9. Mean Frequency and Duration of "Resting" Levels of Radar Use Measured During a Five-Minute Period Occurring Three Segments Before Detection of a Critical Incident

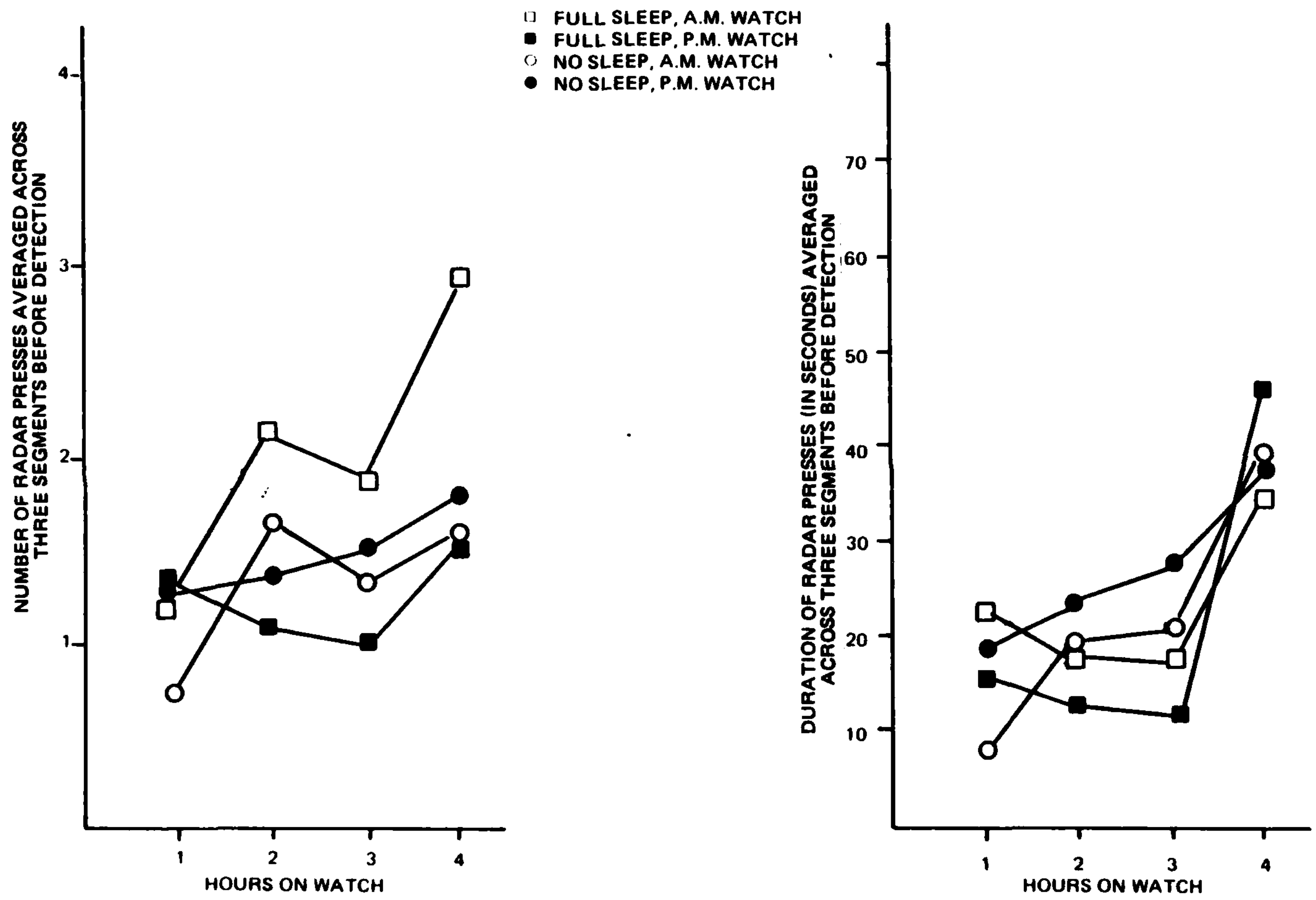

Figure 10. Mean Frequency and Duration of "Resting" Levels of Radar Use Averaged Across Three Five-Minute Segments Before Detection of a Critical Incident

were true, one could argue that a certain amount of boredom may actually increase vigilance by motivating the watchstander to keep busy. Another possible explanation for the increase in radar observation at the end of the watch is the tendency for watchstanders to update the ship status towards the end of the watch in order to apprise the next watchstander of all pertinent information. If this were the case, one could expect that the increase in radar activity during the fourth hour would appear as a rather sudden jump as opposed to a gradual, progressive rise from the first through the fourth hour. In fact, the data bears this out. When one compares the average resting levels between the first, second, and third hours, no significant differences are found. However, a comparison of the fourth hour radar activity to each of the preceding hours yields significant results for all comparisons. For example, the difference between the duration of radar consultation measured during the average resting level of the first and fourth hours is -23.73 [$t(20) = -03.36$, $p = 0.003$], and the difference between the third and fourth hours is -20.02 [$t(18) = 2.97$, $p = 0.008$]. The same relationship holds true for frequency of radar presses. A clear depiction of this phenomenon can be seen in Figure 10. It should also be noted that the tendency to increase radar activity from the first to the fourth hour is not differentially affected by experimental group membership.

3.4.4 RADAR CONSULTATION AT THE TIME OF DETECTION OF AN INCIDENT

Two ways of measuring radar observation in response to an incident were utilized. The first looked at frequency and duration of radar usage during the five-minute segment during which the subject reported a ship sighting. The second method measured the incremental difference in radar usage between one five-minute segment before detection and the five minute period associated with detection. The results of each method are described below.

1. Figure 11 depicts the **use of radar to detect ships during the first and fourth hours.** Data from the second and third hours are not considered, because the instrument failure and ship sighting were counterbalanced in appearance during the second and third hours, resulting in a small n for ship sighting data in each of these hours. The fatigued morning watchstanders appear to use the radar less during the first hour than the other three groups, although the sleep X watch interaction effect is not statistically significant, for frequency of radar use, $F(1,19) = 2.68$, $p = 0.12$. However, the frequency with which each group uses the radar during target detection does change from the beginning to the end of the watch. As can be seen in Figure 11, the fatigued morning group increases its use of radar substantially during the course of the watch from an average of 1.17 presses in the first hour to 3.33 presses in the fourth hour. Of course, given their low level of responsiveness during the first hour, it is not very surprising they would increase their responsiveness somewhat later. The fatigued afternoon group appears to respond qualitatively different by decreasing radar presses from the beginning to the end of the watch. This interaction effect is statistically significant for frequency of radar use (see Table 10) but not for duration.
2. A second way of studying radar use during detection is to measure the **incremental change in radar consultation from one segment before detection to the detection segment.** To assess this incremental change, the amount of radar consultation observed in the five-minute period during which a ship was being sighted was compared to the amount of radar consultation observed in the five-minute period prior to a ship being sighted. Scores based on this difference were subjected to analyses of variance. During the first hour, the full-sleep groups' duration of radar consultation increased by about 45.5 seconds, whereas the no-sleep groups' duration increased by about 28.5 seconds. The differences in responsiveness between these groups was statistically significant, $F(1,11) = 8.433$, $p = 0.014$ (see Table 11). Significant results were also found in the fourth hour where all but the fatigued morning group increased their duration of radar use during the detection segment; this interaction can be seen in Table 12 ($F(1,11) = 3.59$, $p = 0.085$). Statistical significance was not found for frequency of radar presses.

3.4.5 RADAR CONSULTATION DURING RECOVERY FROM DETECTION OF INCIDENT

An inspection of Figures 7 and 8 reveal what appears to be increased use of the radar for several minutes following the sighting of a ship as compared to radar use prior to the sighting. Such an increase in radar observation is not surprising, since the vigilant watchkeeper should track nearby ships, even non-threatening ones, to ensure that the courses of the ships do not bring them onto a collision course. The amount that the radar was used during the five-minute segment immediately following a sighting did not differ between experimental groups, however.

The three segments following detection were average to yield an average level which was subjected to analyses of

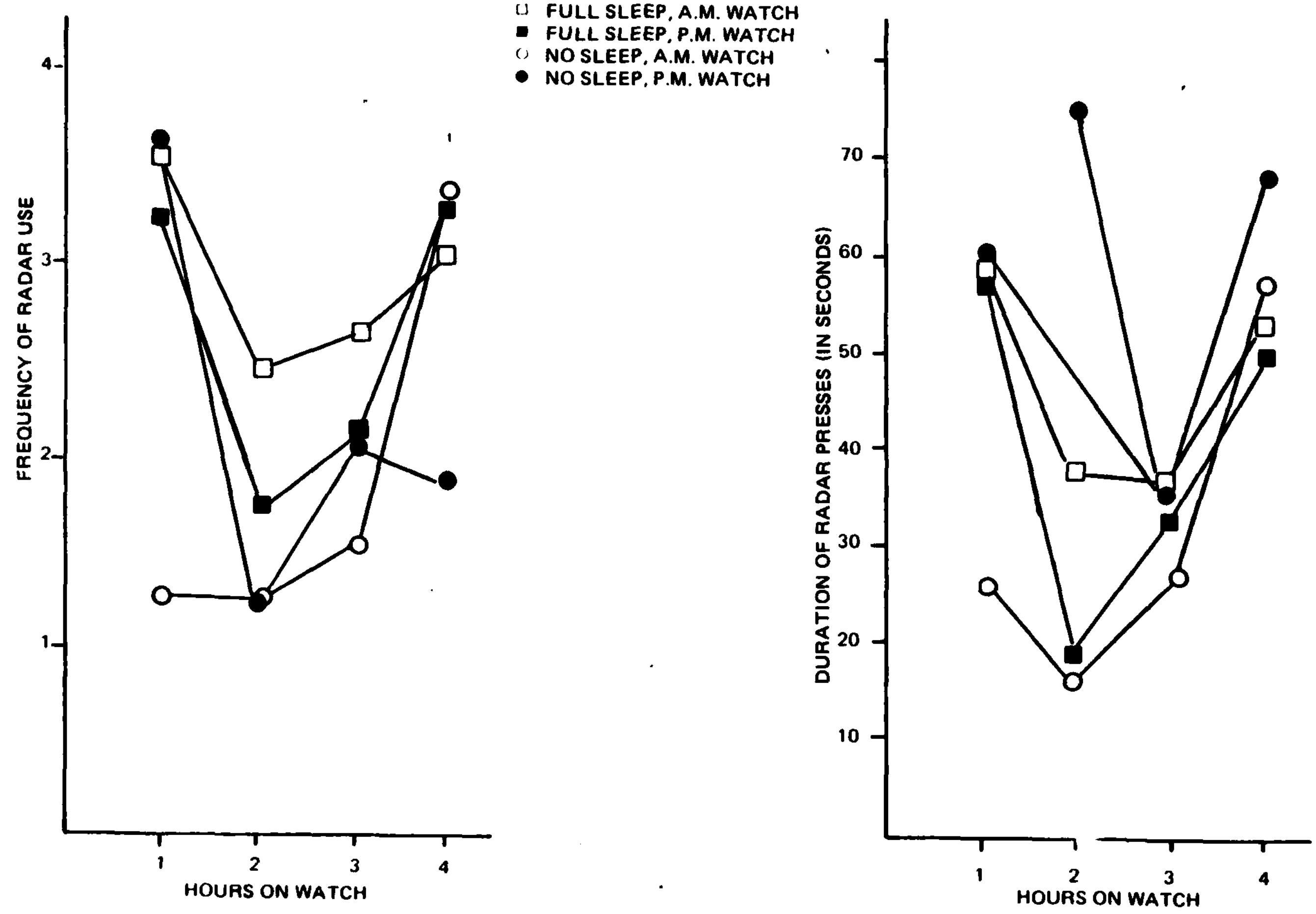

Figure 11. Experimental Group Means Across Watch for Radar Use During Actual Time of Detection

TABLE 10. SUMMARY TABLE FOR DIFFERENCE IN BETWEEN FIRST AND LAST HOUR FREQUENCY OF RADAR PRESSES DURING DETECTION OF SHIPS

Source	Sum of Squares	df	Mean Square	F Ratio	P
Sleep	2.843	1	2.843	0.559	0.465
Watch	13.298	1	13.298	2.616	0.124
S x W	21.285	1	21.285	4.187	0.057
Error	86.417	17	5.083		
Total	124.571	20	6.229		
	Full Sleep Morning	Full Sleep Afternoon	No Sleep Morning	No Sleep Afternoon	
$\overline{X}$	0.75	0.00	-2.17	1.17	
SD	2.36	1.73	2.32	2.48	
n	4	5	6	6	

variance. Figure 12 depicts the mean scores of average recovery for each group during the first and fourth hours of the watch. One can see what appears to be a main effect for time of watch such that the afternoon watchstanders used the radar for longer durations of time than the morning groups. At the fourth hour of the watch, this apparent main effect of time of watch is significant, $F(1,15) = 4.467$, $p = 0.052$, as shown in Table 13. With respect to frequency of radar presses, the well-rested groups appeared to consult the radar more than the fatigued groups; in particular the well-rested, morning group was most active. The main effect of sleep approached statistical significance, $F(1,15) = 2.987$, $p = 0.104$. The interaction effect also approached significance, $F(1,15) = 2.937$, $p = 0.107$.

When experimental groups were compared on changes in radar use between the first and fourth hours, no significant effects were found.

TABLE 11. SUMMARY TABLE FOR DIFFERENCE IN DURATION OF RADAR USE DURING THE SEGMENT IMMEDIATELY PRIOR TO DETECTION AND THE SEGMENT DURING DETECTION IN THE FIRST HOUR[1]

Source	Sum of Squares	df	Mean Square	F Ratio	P
Sleep	5712.90	1	5712.90	8.433	0.014
Watch	0.04	1	0.04	0.000	0.999
S x W	824.89	1	824.89	1.218	0.293
Error	7451.94	11	677.45		
Total	14212.92	14	1015.21		
	Full Sleep Morning	Full Sleep Afternoon	No Sleep Morning	No Sleep Afternoon	
$\overline{X}$ of difference	-46.6	-44.4	-22.00	-34.14	
SD	41.36	16.01	29.93	60.42	
n	4	5	6	6	

[1] Because the difference scores were computed by subtracting the detection segment from one segment before detection, greater negative scores indicate an increase in radar observation as detection takes place.

TABLE 12. SUMMARY TABLE FOR DIFFERENCE IN DURATION OF RADAR USE DURING THE SEGMENT IMMEDIATELY PRIOR TO DETECTION AND THE SEGMENT DURING DETECTION IN THE FOURTH HOUR

Source	Sum of Squares	df	Mean Square	F Ratio	P
Sleep	1554.50	1	1554.50	0.914	0.353
Watch	5418.99	1	5418.99	3.279	0.098
S x W	5933.69	1	5933.69	3.59	0.085
Error	18180.52	11	1652.78		
Total	30181.59	14	2155.83		
	Full Sleep Morning	Full Sleep Afternoon	No Sleep Morning	No Sleep Afternoon	
$\bar{X}$ of difference	-32.00	-20.00	-0.50	-43.17	
SD	29.82	22.41	66.56	49.89	
n	4	5	6	6	

3.4.6 SUMMARY OF RESULTS OF RADAR OBSERVATION

Radar activity not associated with the detection of a particular incident increased as the watch progressed, irrespective of sleep or watch condition. There is reason to believe that this increase in activity is preparatory to turning over the watch to another mate at the end of four hours.

Radar activity associated with the detection of a ship also increases during the watch, with the exception of the fatigued afternoon group, who appear to consult the radar less by the end of their watch, although the actual duration of radar consultation does not change.

3.4.7 PHYSICAL ACTIVITY

No differences with respect to gross physical movements were found between groups. One can conclude that the differences in detection times found between groups can be explained by differences in radar consultation, rather than physical movement. This explanation is further supported

TABLE 13. SUMMARY TABLE FOR DURATION OF RADAR USE AVERAGED OVER THREE SEGMENTS FOLLOWING DETECTION OF THE FOURTH INCIDENT

Source	Sum of Squares	df	Mean Square	F Ratio	P
Sleep	125.773	1	125.773	0.091	0.767
Watch	6180.109	1	6180.109	4.467	0.052
S x W	895.667	1	895.667	0.647	0.434
Error	20752.824	15	1383.521		
Total	28046.555	18	1558.142		
	Full Sleep Morning	Full Sleep Afternoon	No Sleep Morning	No Sleep Afternoon	
$\bar{X}$ of difference	184.00	264.33	164.33	376.67	
SD	46.00	66.08	27.39	75.33	
n	4	4	6	5	

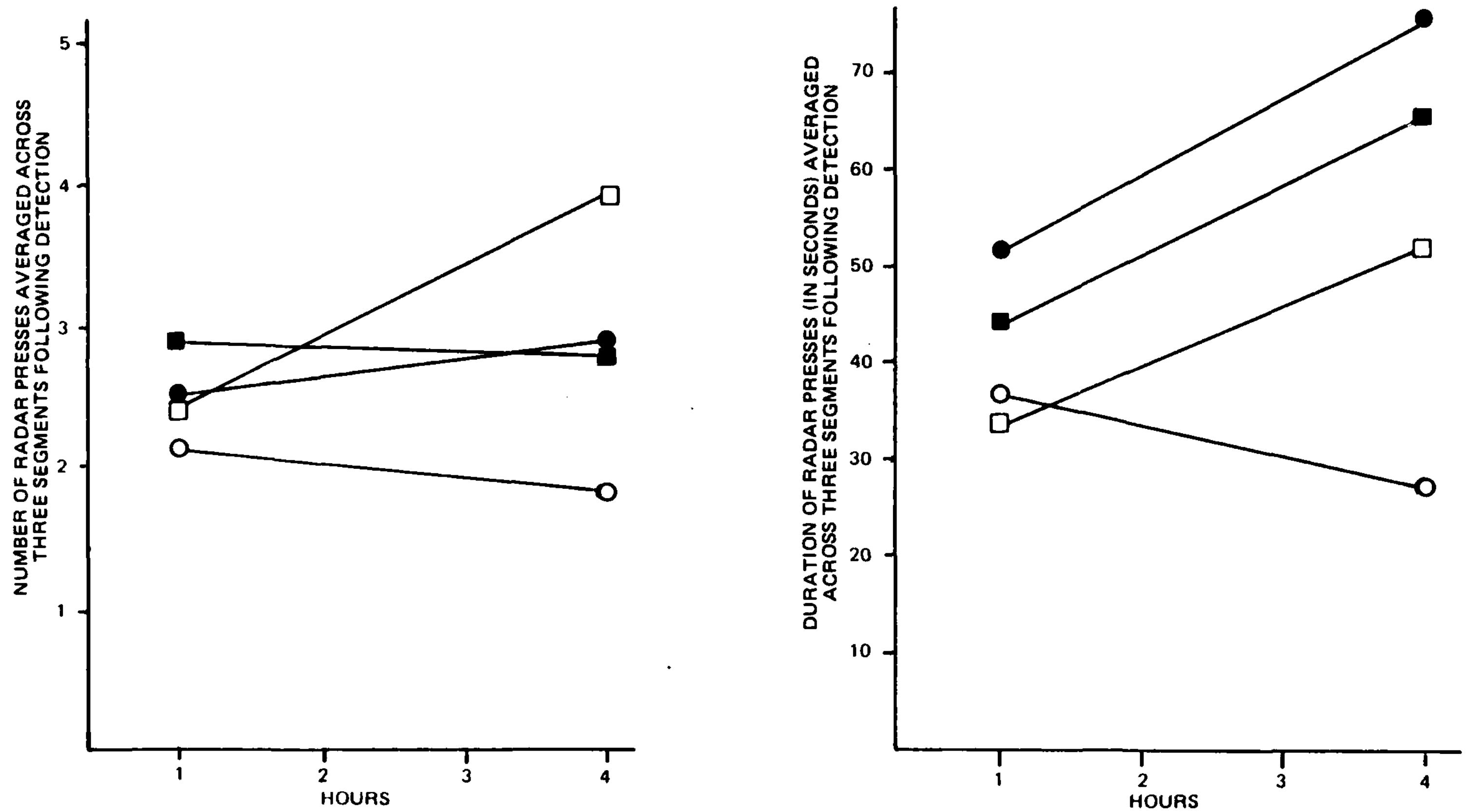

Figure 12. Group Means of Radar Observation Averaged Across Three Segments Following Detection

by the observation that all but one of the ship detections occurred using radar rather than visual sighting. The means and standard deviations of physical activity are presented in Tables 16 through 24 as part of the discussion of physiological measures.

3.5 MEASURES OF AFFECT AND COGNITION

Measures of mood, subjective fatigue, and cognition were administered to each watchstander before and after his watch. Some measures were also administered at mid-watch. An evaluation of the effects of amount of sleep, time of watch, and number of hours on watch on each of these measures is presented in this section.

In order to assess these effects, a series of split-plot analyses of covariance (ANCOVA) (Kirk, 1968) were computed for each psychological measure. The between-subjects variables were amount of sleep (7.5 hours versus 0 hours) and time of watch (morning versus afternoon). The within-subjects variable was the time at which the particular dependent measure was administered on the experimental day (pre- versus mid-versus post-watch). For those measures which were not administered at mid-watch, only two testing times were used in the analysis.

The covariate was the corresponding psychological measure collected post-watch of the familiarization day. For example, when a subject's score on the Visual Search and Scan Two Letter Target Detection test collected on the experimental day was used as the dependent measure, the subject's score on the same test administered post-watch of the familiarization day was used as the covariate.

By using a covariate analysis, it was possible to statistically control for differences between the experimental groups which existed prior to the experimental day. In the example of the Visual Search and Scan Two Letter Test, the full-sleep afternoon watch group performed better than all other groups **on the familiarization day**. This particular group appeared to perform better on several of the tasks and to exhibit moods different from the other groups even before they were subjected to the experimental intervention of sleep deprivation. Lest these **a priori** differences affect the interpretation of data collected on the experimental day, the effects of the familiarization day's testing were partialled out of the experimental day's scores.

The post-watch familiarization day's scores were chosen as the covariate over the pre- and mid-watch scores for several reasons:

1. Some subjects seemed to misunderstand some of the tests upon first administration, while others grasped the objectives immediately. Since some simply failed to perform the task properly at the pre-watch test on familiarization day, these pre-watch data were eliminated as a choice for covariate.

2. Not all measures were collected at mid-watch; therefore, mid-watch measures were eliminated so that the choice of a variable to be used as a covariate could be consistent across all dependent measures with respect to testing time.

3. This left the post-watch familiarization measures as the logical choice. The significant positive correlations of these measures to subsequent measures further substantiates this choice.

Results from each of these analyses of psychological measures are presented below. Means of each psychological measure by experimental group can be found in Tables 14 and 15.

3.5.1 NAVAL PERSONNEL RESEARCH UNIT (NRPU) SCALES

The Naval Personnel Research Unit (NPRU) Positive Affect Scale showed significant main effects for sleep condition, $F(1,19) = 9.61$, $p = 0.006$. The well-rested groups reported more positive feelings than the fatigued groups at all points of measurement – pre-, mid-, and post-watch.

No significant effects of watch schedule or significant changes in positive affect over time were found (see Table D8).

Significant effects for sleep condition were also found in the NPRU Negative Affect Scale, with subjects in the no-sleep groups reporting more negative affect than those in the full-sleep groups, $F(1,19) = 65.63$, $p = 0.001$. A main effect for watch condition was also found, with the afternoon groups reporting more negative affect, $F(1,19) = 6.67$, $p = 0.02$. In all cases, however, the degree of negative affect decreased during the course of the watch $F(2,40) = 4.17$, $p = 0.02$. It was somewhat surprising to find that test subjects, who were wired up and required to stand yet another boring watch in less than 24 hours, would report being in better moods as the watch progressed, rather than in worse moods; but that is what the data revealed. Perhaps this is an example of test subject compliance or a sign of the subjects increasing anticipation that their ordeal would soon be over. Table D9 summarizes the analyses.

TABLE 14. GROUP MEANS[a] ON AFFECT SCALES AT THREE POINTS IN WATCH

Test	Time of Test	Experimental Group			
		Full Sleep Morning	Full Sleep Afternoon	No Sleep Morning	No Sleep Afternoon
NPRU Positive Affect	Pre	1.71	1.51	1.26	1.23
	Mid	1.79	1.59	1.31	1.39
	Post	1.81	1.60	1.33	1.41
NPRU Negative Affect	Pre	0.21	0.35	0.80	1.24
	Mid	0.16	0.30	0.68	0.89
	Post	0.12	0.20	0.59	0.72
Stanford Subjective Sleepiness	Pre	2.11	2.39	3.67	3.17
	Mid	2.11	2.22	4.33	3.50
	Post	2.11	2.06	4.00	3.00
POMS Fatigue Scale	Pre	2.12	2.84	9.97	9.77
	Mid	1.79	1.34	11.30	7.63
	Post	2.12	1.51	12.80	9.77
POMS Vigor Scale	Pre	16.76	13.42	10.58	12.92
	Mid	18.09	15.25	10.75	10.35
	Post	18.09	15.96	8.25	12.63
POMS Confusion Scale	Pre	3.43	4.76	3.80	5.15
	Mid	2.59	3.76	4.80	2.87
	Post	3.09	3.92	5.47	3.30
POMS Anger Scale	Pre	1.26	2.11	0.50	1.61
	Mid	1.09	3.44	0.84	2.97
	Post	0.92	1.44	0.50	0.83

[a]Group means have been adjusted for covariates.

While it is not surprising that the subjects with no sleep should exhibit negative affect. It is noteworthy. Previous fatigue research in other industries has shown the NPRU scales to be sensitive to fatigue. Until the completion of the present research, it had not been ascertained whether the scales would retain their validity in a marine simulation environment. However, these scales' sensitivity to fatigue, in this study, indicates their usefulness in other maritime simulation research. That is, future experiments conducted on marine simulators can use the NPRU affect scales with the confidence that they are sensitive to fatigue and related factors.

3.5.2 STANFORD SUBJECTIVE SLEEPINESS SCALE (SSS)

As expected, main effects for sleep condition were also found in the Stanford Subjective Sleepiness Scale (SSS). The subjects in the no-sleep condition were significantly more fatigued throughout the watch than those who had slept, $F(1,19) = 17.09$, $p = 0.0006$. Unlike the NPRU Negative Affect Scale, scores on the SSS did not substantially change as time on watch progressed. (See Table D10 for ANOVA summary table.)

TABLE 15. GROUP MEANS[a] ON COGNITIVE MEASURES AT TWO POINTS IN WATCH

Test	Time of Test	Full Sleep Morning	Full Sleep Afternoon	No Sleep Morning	No Sleep Afternoon
VSS 2 Letter	Pre	52.44	53.96	48.85	50.32
Lines Scanned	Post	47.84	54.80	47.52	46.32
VSS 4 Letter	Pre	48.26	56.34	43.01	42.78
Lines Scanned	Post	46.46	54.17	47.84	41.45
VSS 6 Letter	Pre	51.01	58.92	45.28	44.97
Lines Scanned	Post	45.01	50.92	48.94	47.47
VSS 2 Letter	Pre	3.07	3.12	3.12	3.53
No. Correct	Post	3.47	4.29	3.79	3.69
VSS 4 Letter	Pre	2.03	3.98	3.50	2.40
No. Correct	Post	2.03	2.98	3.25	3.40
VSS 6 Letter	Pre	4.14	3.73	3.62	3.04
No. Correct	Post	3.74	2.73	3.62	3.04
Digit Span	Pre	7.29	7.10	6.84	7.66
Forward	Post	7.95	8.10	7.68	7.95
Digit Span	Pre	6.16	6.22	5.40	6.23
Backwards	Post	6.24	5.81	5.95	6.62
Baddeley	Pre	25.97	27.84	25.53	27.14
Lines Scanned	Post	25.97	28.00	26.36	27.71
Baddeley	Pre	24.98	26.55	24.42	23.61
No. Correct	Post	23.82	26.22	25.92	24.61

[a]Group means have been adjusted for covariate.

These results served to confirm the anticipated effects of the experimental manipulation; that is, the deprivation of sleep induced the expected state of fatigue. In future research in the marine industry, the SSS may be used with confidence as a measure of fatigue.

3.5.3 PROFILE OF MOOD STATES CHECKLIST (POMS)

The Profile of Mood States (POMS) Checklist, which measures feelings of anger, anxiety, confusion, depression, fatigue, and vigor, yielded significant results for fatigue and its opposite vigor, and interesting results for confusion and anger.

Figures 13 and 14 depict the effects of the experimental variables on feelings of fatigue and vigor as reported on the POMS Checklist. Both fatigue and vigor showed strong main effects for sleep condition. The no-sleep groups felt more fatigued [$F(1,20) = 42.81$, $p = 0.0001$] and less vigorous [$F(1,20) = 13.57$, $p = 0.0015$] than the well-rested groups. The no-sleep morning watchstanders, in particular, reported less vigor. (See Tables D11 and D12 for details.)

The deterioration of the no-sleep morning group's affective state as the watch progressed was also demonstrated with respect to reported confusion. This group reported an increase in confusion as the watch progressed, while the other groups can best be seen graphically in Figure 15 [$F(2,42) =$

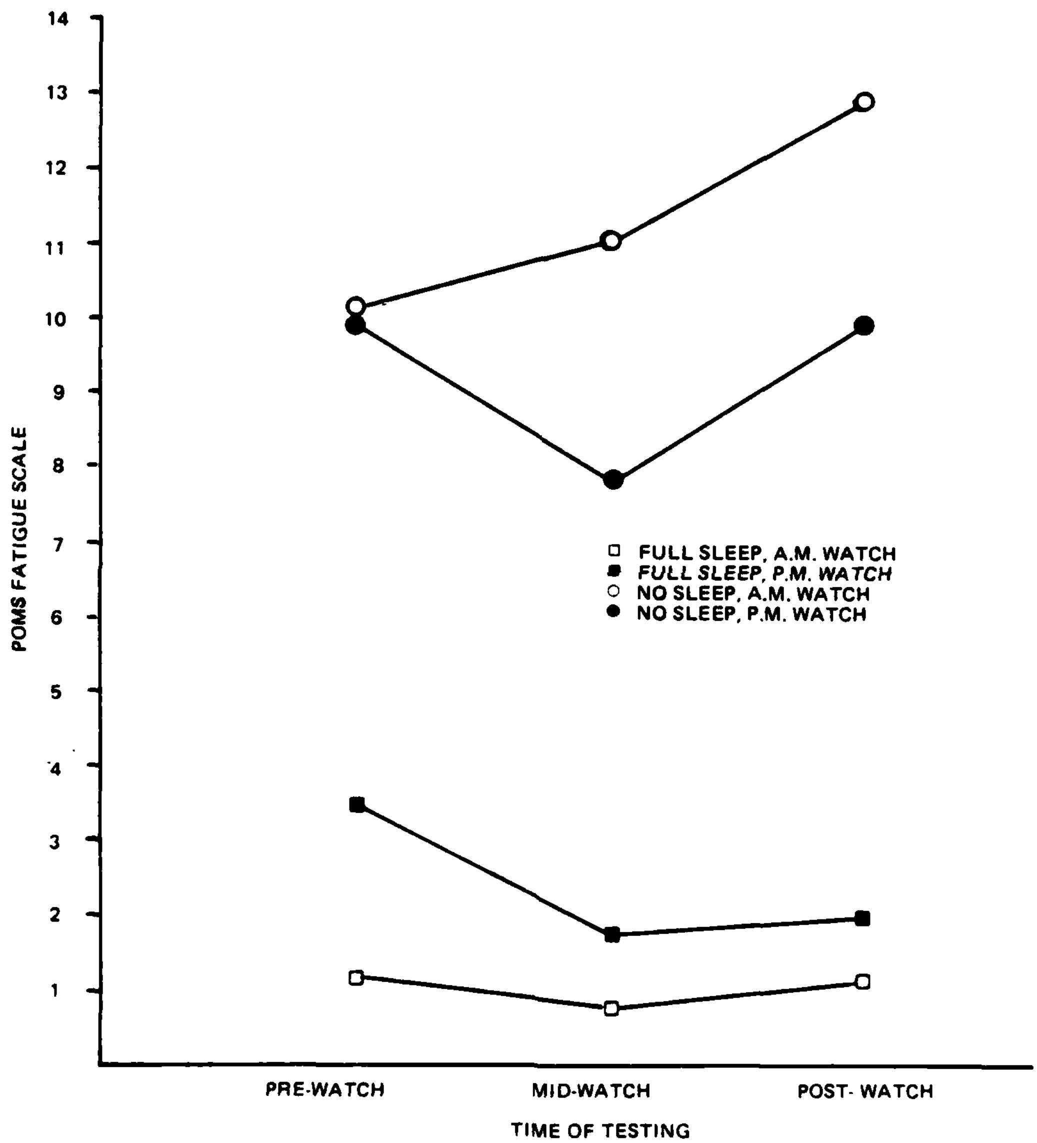

Figure 13. Group Means on POMS Fatigue Scale at Three Points in Watch

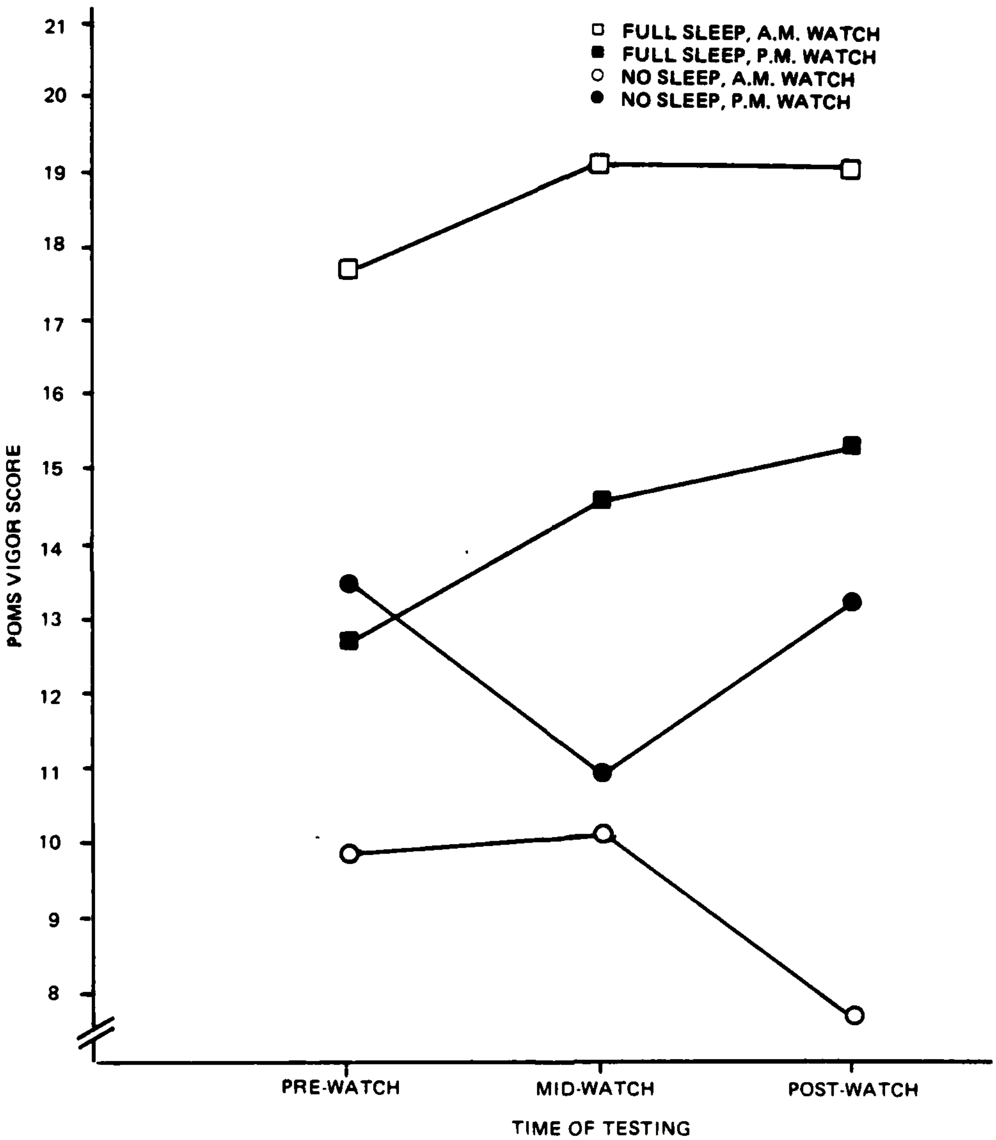

Figure 14. Group Means on POMS Vigor Scale at Three Points in Watch

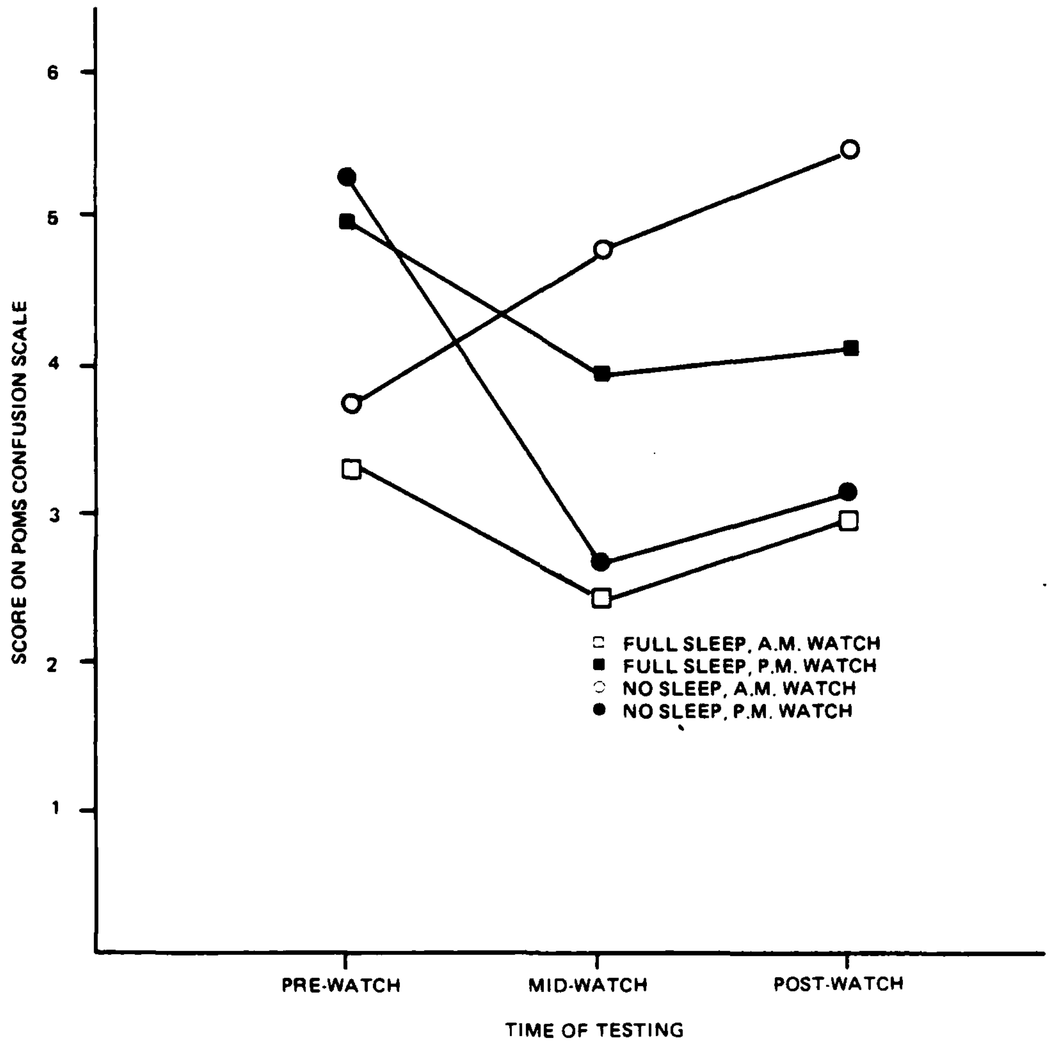

Figure 15. Interaction of Watch x Test Time on POMS Confusion Scale

2.36, p = 0.11]. The analysis of covariance summary table is presented in Table D13.

The afternoon watchstanders, irrespective of sleep condition, reported that they felt more angry than their morning counterparts, F(2,42) = 2.94, p = 0.10). Anger was also expressed more at the beginning and middle of the watch. Table D14 summarizes this analysis. Since anger is not a dependent measure of any particular interest to the experimental hypothesis, this finding is not considered especially meaninful but it is interesting, nevertheless. (A score on anger is a product of the Profile of Mood States Checklist which assessed other moods, such as fatigue and confusion, which were of more interest.) The increased expression of anger by the afternoon groups is probably a result of these subjects having to wait around for their watches to begin and having to stay later at the facility than their morning associates. In the interest of maintaining good rapport with test subjects in the future, it may be advisable to keep waiting time to a minimum.

In summary, the POMS Checklist contains several scales which are sensitive to fatigue, watch schedule, and time of testing. The fatigue and vigor scales were particularly sensitive to the independent variables, while the confusion and anger scales also showed a slightly lesser degree of sensitivity. The POMS could be used in future marine simulation studies which examine fatigue and related issues.

3.5.4 VISUAL SEARCH AND SCAN TEST (VSS)

The Visual Search and Scan Test (VSS) required the subject to scan a list of randomly generated letters for specific combinations of target letters. These targets consisted of combinations of either two letters, four letters, or six letters. Scores were assigned for number of lines scanned and number of targets correctly identified.

The two-letter target was considered the easiest task of this series. The analyses of covariance indicated a slight but statistically insignificant influence of sleep on number of lines scanned such that the no-sleep groups scanned fewer lines than the full-sleep group, F(1,18) = 2.59, p = 0.13. This effect was strengthened slightly when the task was increased in difficulty to a four-letter target, F(1,18) = 2.83, p = 0.11. When the task was increased to a six-letter target, an interesting interaction between sleep condition and time of testing was noted. During the pre-watch administration, both of the no-sleep groups scanned fewer lines than the rested groups. However, at the post-watch administration, the fatigued groups improved upon their previous performance, while the rested groups dropped in performance, F(1,18) = 2.96, p = 0.10. This finding was not predicted and was rather surprising. Figure 16 illustrates these findings. Of further interest is the strong effect of the covariate across all three levels of difficulty of the VSS. The p levels of less than 0.01 associated with the effect of the covariate (performance on familiarization day) may be interpreted as an indication that the number of lines scanned is a function of individual differences between subjects over and above any experimental manipulation. The summary tables associated with these statistical analyses can be found in Tables D15, D16, and D17.

The number of targets correctly identified appears to be influenced by a different set of factors in comparison to the number of lines scanned. For example, in the two-letter target task, sleep influences number of lines scanned; however, only the time of testing appears to influence the number of correct targets. The ANCOVA revealed that significant improvements in target detection occur from pre- to post-watch irrespective of group membership, F(1,19) = 4.86, p = 0.04 (see Table D18). An even more complex interaction occurs in the four-letter target task. A significant time of test X sleep condition X watch schedule interaction indicates that the full-sleep afternoon group detects fewer targets accurately as time progresses through the watch, while the no-sleep afternoon group appears to improve their performance, F(1,15) = 4.22, p = 0.06. When the individual subjects' scores were examined in an effort to elucidate this finding it was noticed that **none** of the full-sleep morning group had perceived any of the targets accurately when tested post-watch on the familiarization day. Since that particular testing was used as a covariate in the analysis, a second analysis of variance was computed without partialling out the effects of the post-watch familiarization day's scores. Results from this second analysis substantiate the previous analyses. Tables D19 and D20 in Appendix D show the results of both analyses. Figure 17 illustrates the interaction effect.

No significant effects were found when the number of correct six-letter targets were used as the dependent measure.

In summary, the Visual Search and Scan task was consistently sensitive to sleep deprivation. The effects of watch schedule and length of time on watch on the VSS were much less consistent. Given its sensitivity to sleep loss and the added advantage of allowing the experimenter to manipulate the variability in difficulty, its use in future research of this type is recommended. However,

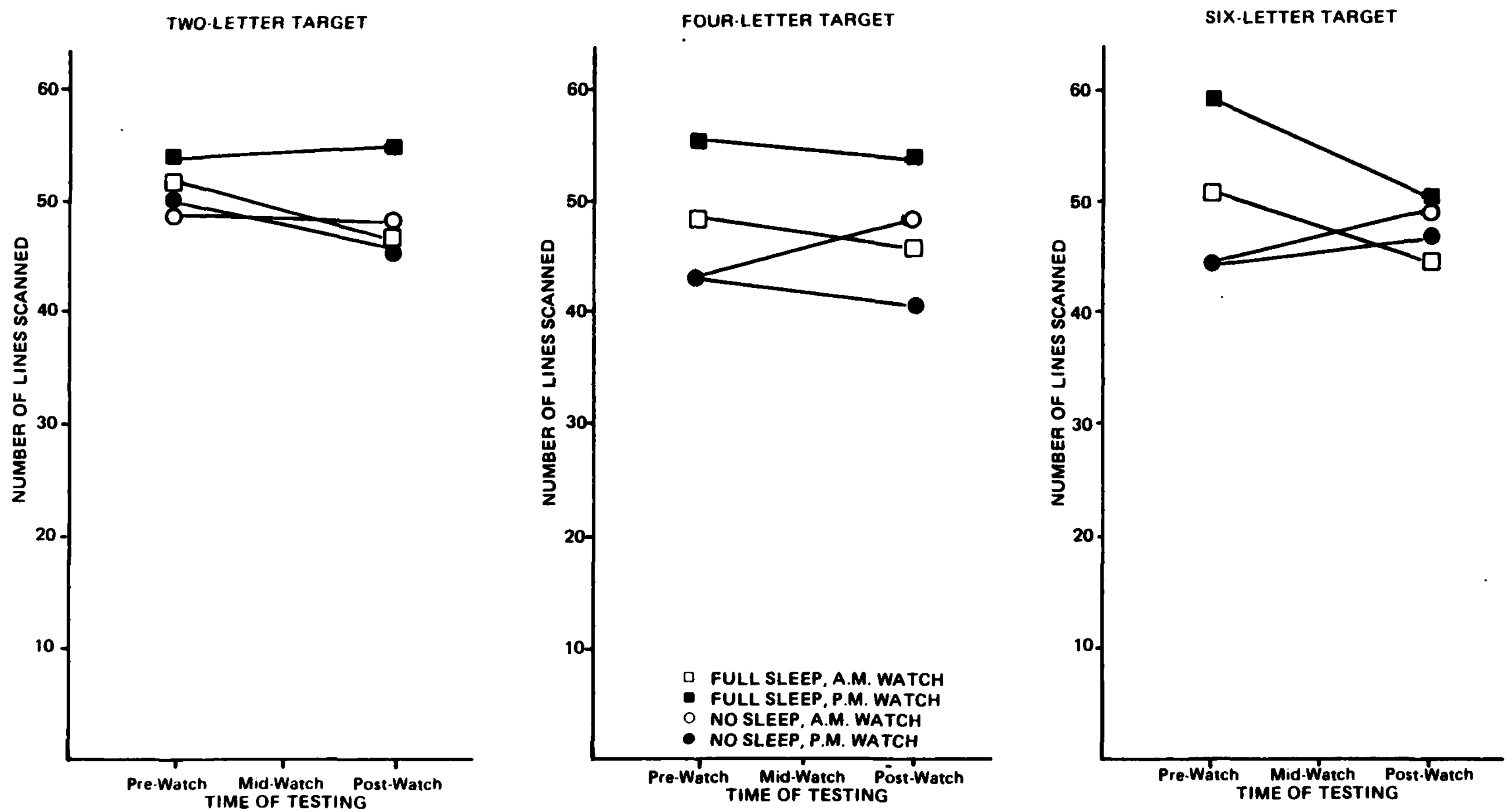

Figure 16. Number of Lines Scanned During Visual Search and Scan Tests of Varying Difficulty

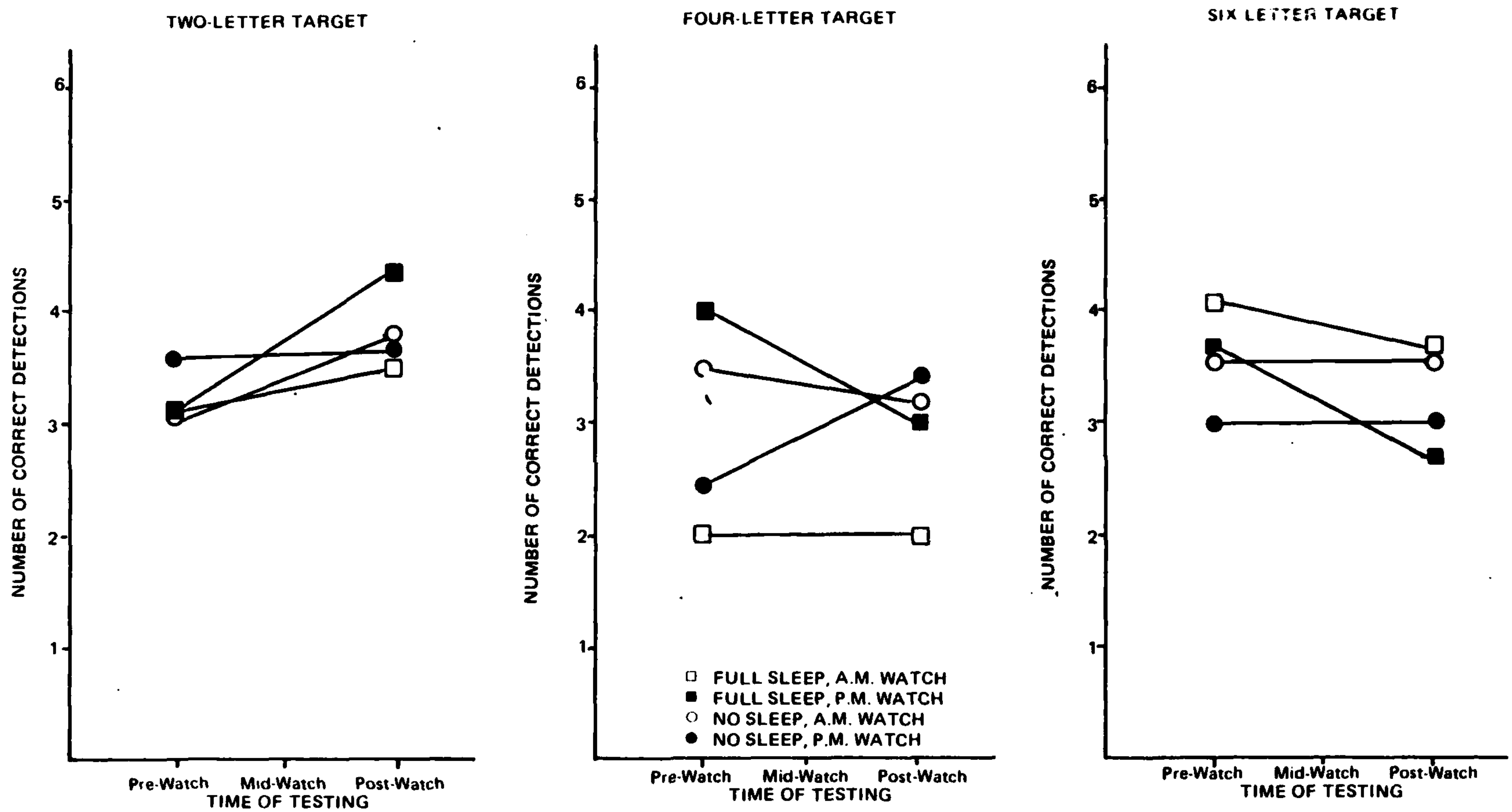

Figure 17. Number of Targets Correctly Detected During Visual Search and Scan Tests of Varying Difficulty

its sensitivity to individual differences must always be borne in mind when utilizing this instrument.

3.5.5 DIGIT SPAN TEST

The Digit Span Test of short-term memory, which required the subject to recall a series of digits in the direction they were announced (i.e., forward), yielded a significant effect of time of testing such that subjects recalled more digits after the watch than before, $F(1,21) = 7.55$, $p = 0.01$ (see Table D21). Recalling digits in the direction opposite from that announced (i.e., backwards) did not yield significant results.

The lack of significant effects of sleep or watch condition on Digit Span may indicate that short-term memory of the watchstanders was not measurably affected by the independent variables, or it may indicate that the Digit Span test is not a sufficiently sensitive measure of an existing effect; therefore, its utility in future marine simulation work is questionable.

3.5.6 BADDELEY GRAMMATICAL TRANSFORMATION TEST

There were no reportable results revealed by the Baddeley Grammatical Transformation Test. This test was originally incorporated into the design at the recommendation of Dr. Paul Naitoh of the Naval Personnel Research Unit in San Diego. Dr. Naitoh has used it successfully in fatigue research there. The lack of statistical significance in the present study may be due to an incompatibility of the Baddeley and the watchstanding tasks, or it simply may not have been affected by the independent variables used in this study. Whatever the reason, it may not be advisable to use it in future simulator research.

3.6 PHYSIOLOGICAL MEASURES

The raw data for the physiological measures consisted of readings of the subjects' muscle activity, physical activity, taken in millivolts, as well as heart rate and oral temperature. A rating of overall muscle tension (here referred to as EMG) was made, based on the activity of the frontalis (forehead) muscle. Measures were also taken of the subjects' gross body movements by means of a device strapped to the wrist. Although heart rate and heart rate variability were also recorded, an error in the recording process made automatic data reduction impossible. Therefore, manual data reduction (counting individual heart beats) was necessary. This method yielded heart rate data which was incomplete and is probably not as reliable as data reduced automatically. In some cases, only two of the six subjects in an experimental group contributed data. As a result, statistical analyses of EKG data were limited. Furthermore, the manual reduction method made heart rate variability very difficult to obtain; therefore, cardiac variability is not reported. The measurement of oral temperature was also of questionable reliability and was unavailable for several subjects. No results are reported for oral temperature.

3.6.1 METHOD OF ANALYSIS

Each subject's run was divided into 5.5 minute segments and an average measure of EMG and physical activity was computed for each segment. A total of 43 segments comprised the entire four-hour watch. An overview of the entire watch, broken down into 5.5 minute segments and further separated by experimental groupings, can be seen in Figures 18 and 19.

The analysis of this information focused on the particular segment in each hour during which the subject detected either a ship or an instrument failure. The analysis also included an examination of three segments before and three segments after each detection period.

The method used to extract pertinent physiological data from the entire four-hour watch is described below. Comprehension of this explanation and the terms defined within it are necessary for an understanding of the data analysis.

3.6.2 DETECTION SEGMENT

The specific 5.5 minute time segment during which each subject detected either a ship or an instrument failure was determined by consulting the experimental log. This was referred to as the "detection segment" and was determined for each subject for each hour of the watch.[1] For example, the detection segment associated with the detection of the destroyer or the instrument failure during the second hour had to have occurred sometime during Segments 11 through 22.

[1] One detection should have occurred per hour; however, in a few cases, no detection of the instrument failure ever occurred.

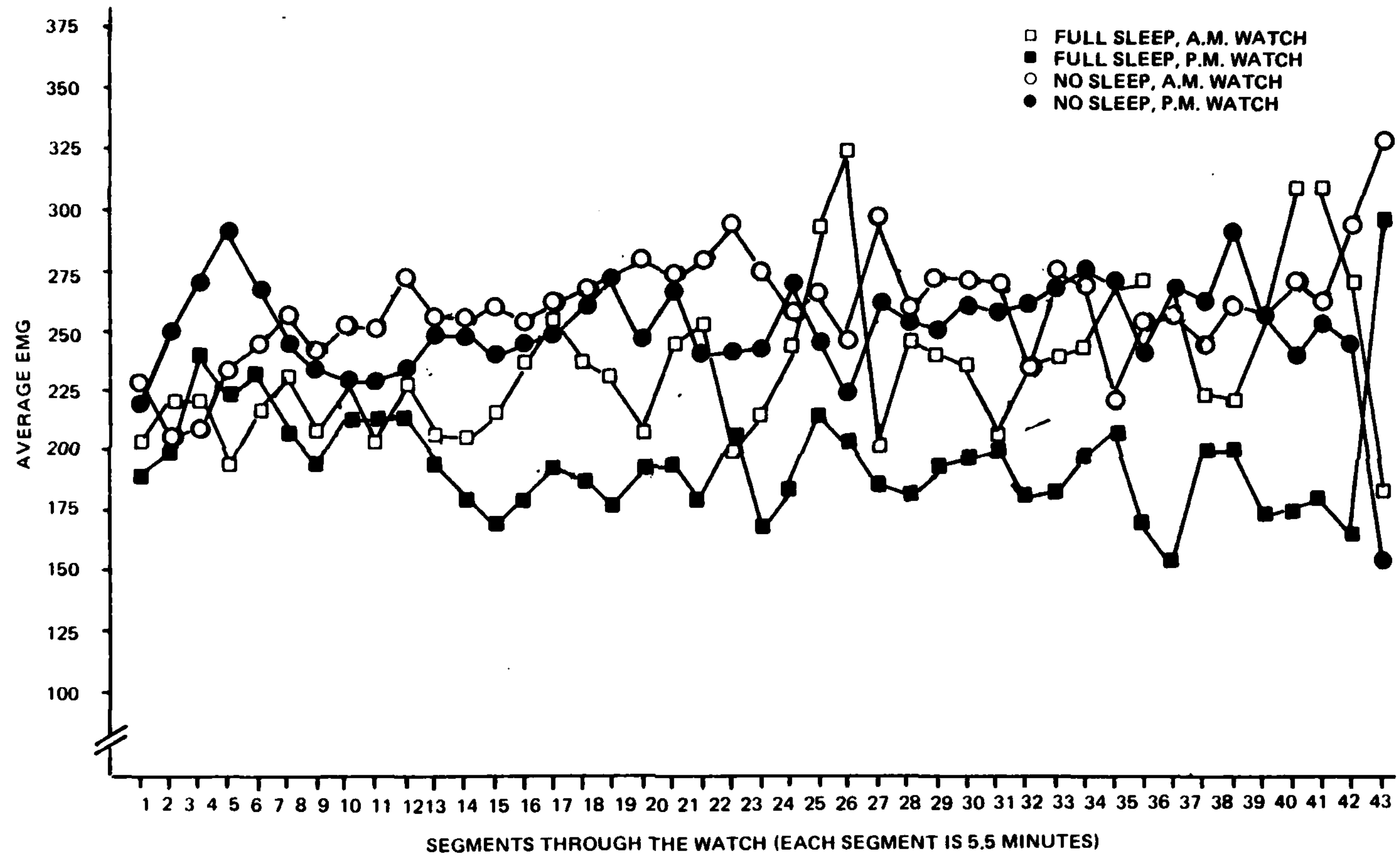

Figure 18. EMG Activity Averaged Every 5.5 Minutes Throughout Four-Hour Watch for Each Experimental Group

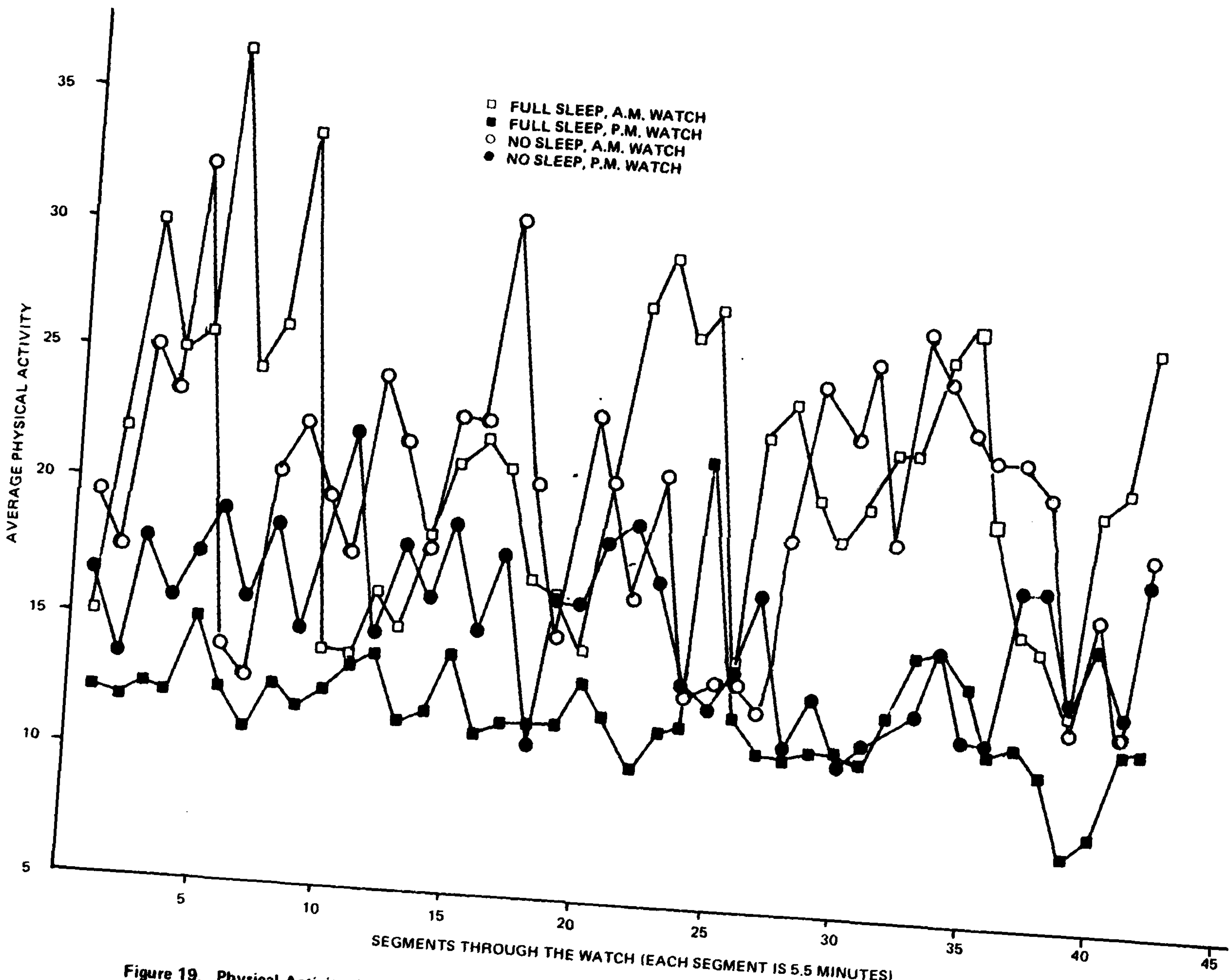

Figure 19. Physical Activity Averaged Every 5.5 Minutes Throughout Four-Hour Watch for Each Experimental Group

3.6.3 BEFORE DETECTION AND AFTER DETECTION SEGMENTS

Three specific segments immediately before and three segments immediately after each detection segment were also noted. This permitted an examination of physiological activity 5.5 minutes, 11 minutes and 16.5 minutes before the detection segment, and for the same periods of time after the detection segment. These "before and after segments" were useful in several ways. For example, a simple difference score between the detection segment and one segment before detection indicated the increment or decrement in physiological activity associated with detection. Likewise, a difference score between the detection segment and one segment after detection indicated immediate physiological recovery after detection. Furthermore, an examination of before detection segments alone yielded information about a subject's resting level, prior to any heightened physiological responses to detection. This measure of resting level was particularly useful in other analyses which looked at the change in resting levels as the watch progressed.

A series of statistical tests (mostly analyses of variance) was performed using sleep and watch conditions as the independent variables, and various combinations of detection segments, before detection segments, and after detection segments as the dependent measures.

The presentation of results which follows is based entirely on the before, during, and after detection segments associated with each hour of the watch. Figures 20, 21 and 22 graphically depict each experimental group's average EMG, physical activity, and EKG scores for these segments. In these figures the three segments before detection are referred to as the "resting period" and the three segments after detection are referred to as the "recovery period." These concepts are used as guides to the discussion of the physiological results.

3.6.4 RESTING LEVEL PRIOR TO DETECTION OF AN INCIDENT

3.6.4.1 Average Resting Level

The physiological resting level measured prior to detection was calculated by averaging the three segments before detection. This yielded data about the general state of physiological arousal which was present prior to the onset of an attention-demanding task. Two by two analyses of variance were conducted for each hour. The results showed that the no-sleep groups generally had a higher level of EMG prior to each incident than did the full-sleep groups (see Table D22). This tendency was significant during the first hour, $F(1,20) = 3.18$, $p = 0.09$, but became less marked as the watch went on. There were no consistent effects of watch condition on resting levels of either EMG, physical movement, or EKG. Table 16 shows the group means, standard deviations, and number of valid scores for the average across three segments prior to detection of each incident.

The effect of number of hours on watch on resting state was also examined. The resting level before the first incident was compared to the resting level before the fourth incident across all subjects using a correlated t-test. The resting state on the average appeared slightly higher at the fourth hour for EMG, physical movement, and EKG, although these differences were not enough to produce statistical significance.

The effect of sleep and watch condition on the change in resting state over time was analyzed by computing the difference in resting state between the first and fourth incident for each subject. These difference scores were then used as the dependent measure for a two-way analysis of variance. The results of this analysis indicated that the full-sleep groups and the no-sleep morning group had a higher level of arousal, both EMG and physical movement, before the fourth incident than before the first. The no-sleep afternoon group, on the other hand, had higher levels of arousal before the first incident. This led to a significant effect of watch condition on the EMG measure, $F(1,19) = 3.83$, $p = 0.07$, although the effects on physical movement and EKG were not significant. Table 17 and Table D23 presents the means and standard deviations for this measure and the analysis of variance for EMG, respectively. The reader may note that the difference score between the first and fourth incident for the no-sleep morning group, which appears in Table 17, does not correspond to the individual incident scores in Table 16. This is due to the reduction in that group's n for the computation of values in Table 17.

3.6.4.2 Resting Level of One Segment Before Detection

A finer examination of the pre-detection data focused on only one segment before detection. This analysis looked only at physiological activity immediately before detection. Table 18 presents the means and standard deviations for each experimental group. At each of the four hours of the watch, the fatigued group appeared to exhibit more muscle tension. This was especially noticeable during the first

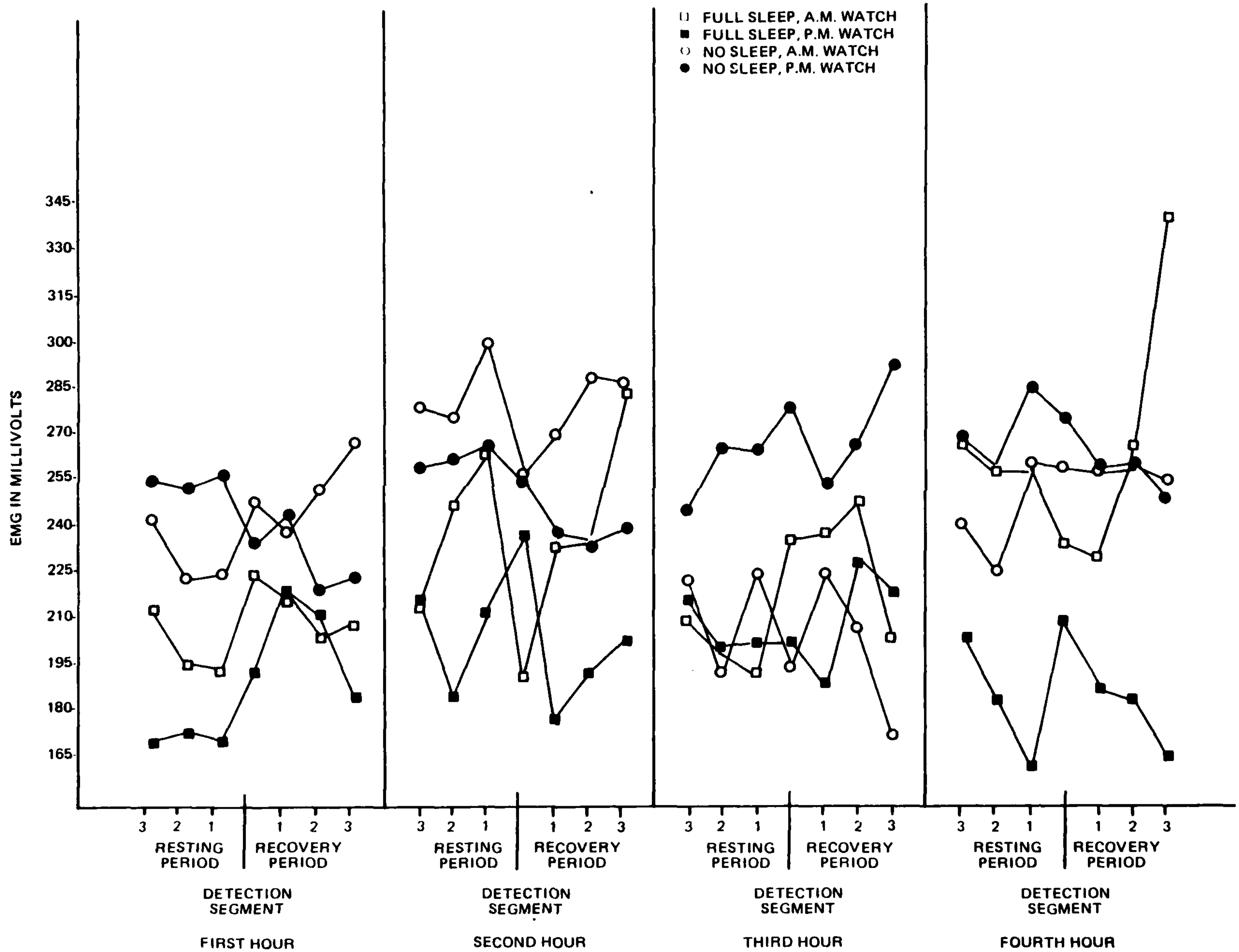

Figure 20. Experiment Group Averages for EMG Before, During, and After Detection at Each Hour of the Watch

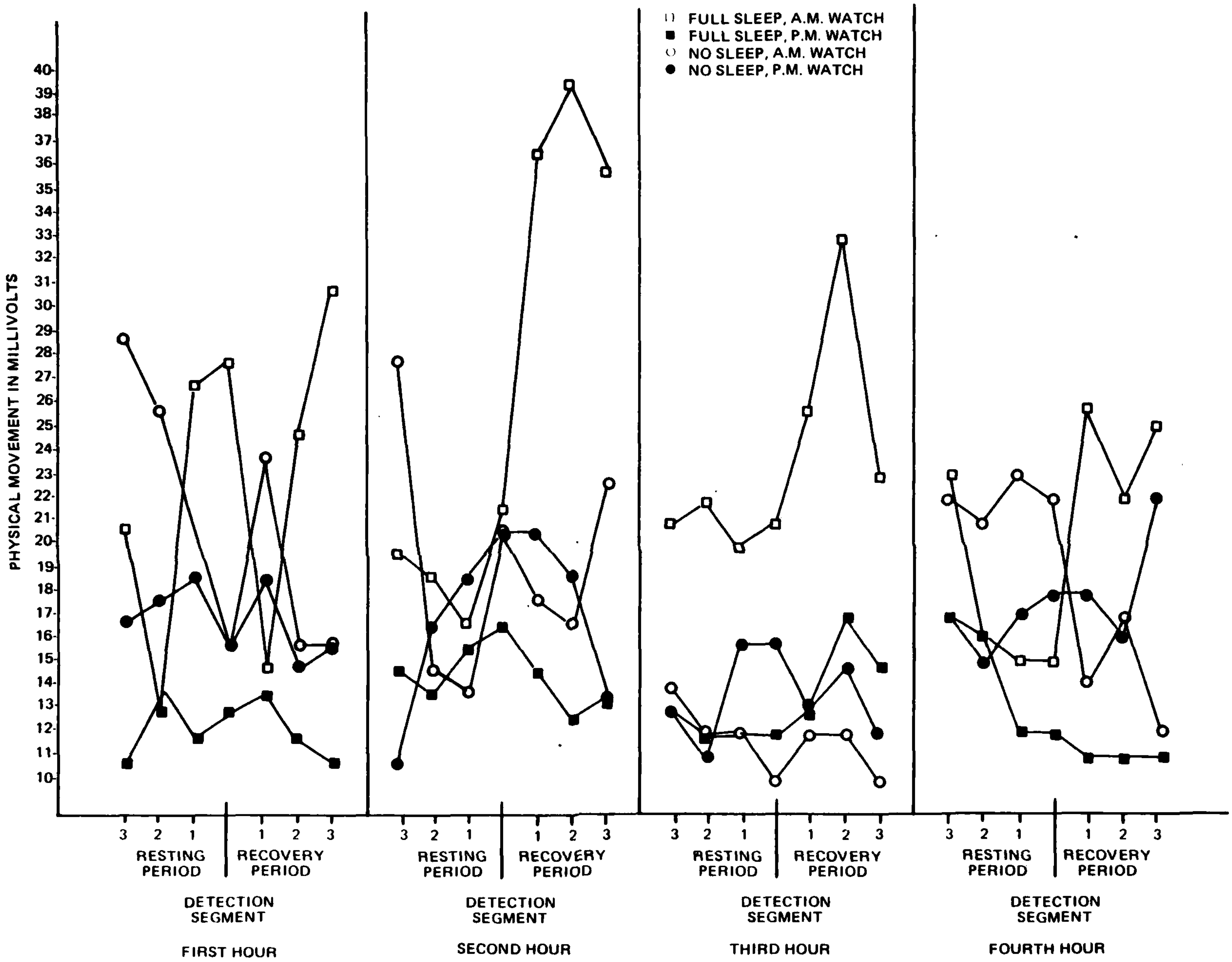

Figure 21. Experimental Group Averages for Physical Movement Before, During, and After Detection at Each Hour of the Watch

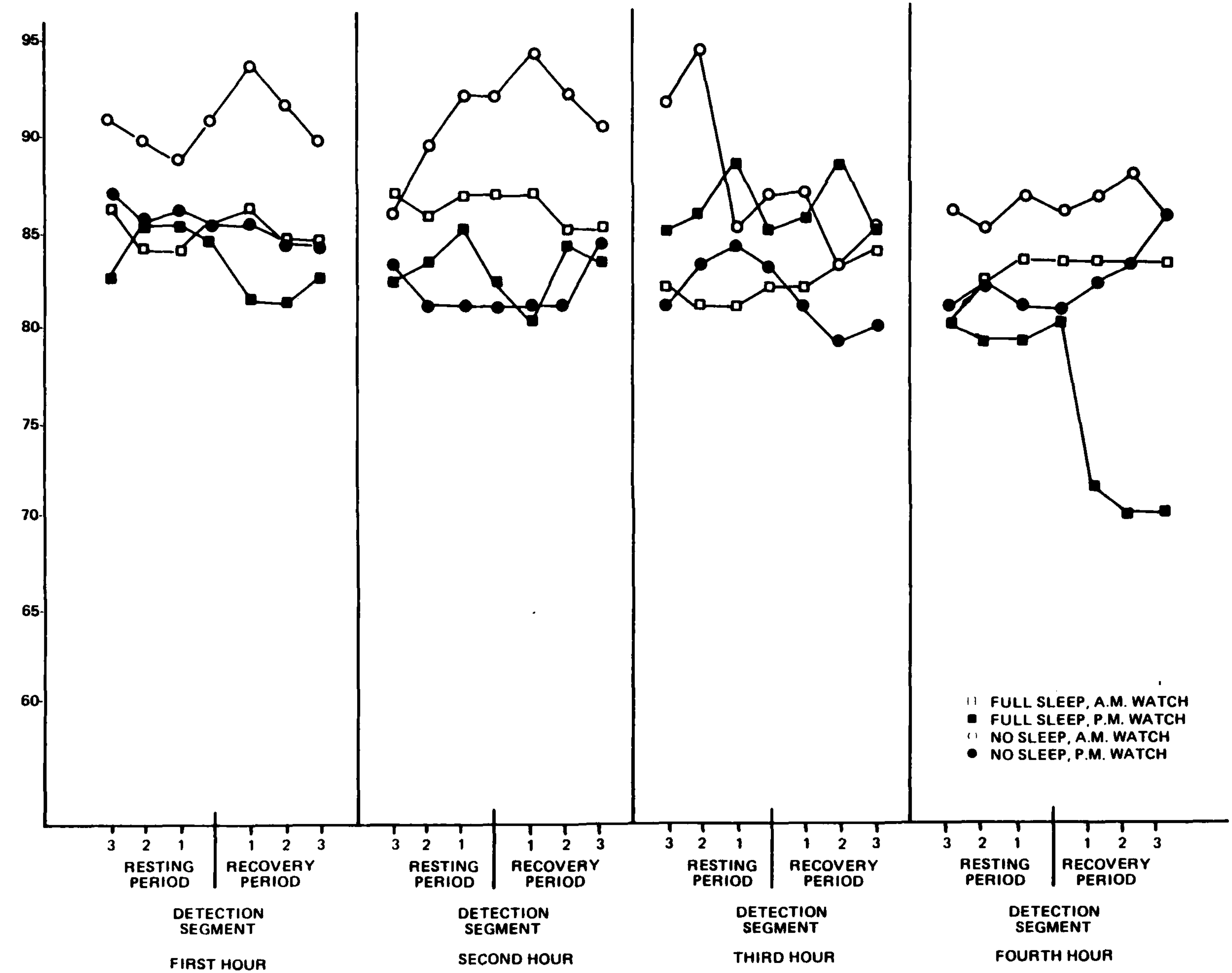

Figure 22. Experimental Group Averages for EKG Before, During and After Detection at Each Hour of the Watch

TABLE 16. PHYSIOLOGICAL ACTIVITY AVERAGED ACROSS THE THREE SEGMENTS BEFORE DETECTION (RESTING LEVEL)

		Full Sleep Morning	Full Sleep Afternoon	No Sleep Morning	No Sleep Afternoon
First Hour					
EMG	Mean	210.5	176.6	252.8	257.7
	SD	39.7	74.5	116.0	91.9
	n	5	6	4	6
Physical Activity	Mean	14.1	12.4	25.2	18.0
	SD	5.9	8.1	21.1	20.5
	n	4	6	4	6
EKG	Mean	87.3	84.3	93.0	86.9
	SD	20.5	3.8	19.3	10.2
	n	5	2	2	6
Second Hour					
EMG	Mean	246.5	208.2	288.4	266.2
	SD	96.8	135.7	153.0	90.9
	n	4	4	5	6
Physical Activity	Mean	19.0	15.1	18.9	15.9
	SD	12.4	13.6	11.1	11.3
	n	4	4	5	6
EKG	Mean	87.6	84.2	91.2	82.6
	SD	16.9	6.8	8.9	6.9
	n	4	2	4	6
Third Hour					
EMG	Mean	205.9	209.8	218.3	262.5
	SD	57.4	99.6	76.7	82.6
	n	5	5	4	6
Physical Activity	Mean	20.9	12.2	13.0	13.3
	SD	20.4	8.2	3.3	6.8
	n	6	5	4	6
EKG	Mean	82.3	87.3	95.0	83.3
	SD	17.8	0	0	8.2
	n	6	1	1	5
Fourth Hour					
EMG	Mean	263.6	187.3	246.4	273.9
	SD	78.8	88.3	133.0	100.7
	n	5	6	6	6
Physical Activity	Mean	18.2	15.1	22.1	16.5
	SD	10.0	13.9	28.8	9.7
	n	5	6	6	6
EKG	Mean	82.7	80.2	87.1	82.3
	SD	19.0	8.7	9.9	6.2
	n	5	2	4	5

TABLE 17. DIFFERENCE BETWEEN RESTING STATE BEFORE FIRST INCIDENT AND RESTING STATE BEFORE FOURTH INCIDENT

		Full Sleep Morning	Full Sleep Afternoon	No Sleep Morning	No Sleep Afternoon
First Hour					
EMG	Mean	80.6	10.7	18.7	-21.6
	SD	91.2	46.5	84.3	53.5
	n	4	6	4	5
Physical Activity	Mean	8.2	2.7	2.4	-6.9
	SD	8.1	6.0	21.0	18.2
	n	3	6	4	5
EKG	Mean	3.4	-4.2	-5.0	-3.0
	SD	6.3	4.9	14.1	6.3
	n	4	2	2	4

NOTE: Negative number indicates lower reading at fourth incident.

and fourth hours, although the difference did not reach significance.

3.6.4.3 Summary of Resting Level of Physiological Measures

The data suggests that the afternoon watchkeepers showed more muscle tension than the morning watchkeepers throughout the watch when measured for 16.5 minutes prior to any attention-demanding task (i.e., while "at rest"). When measured 5.5 minutes prior to the detection of an event, the fatigued group showed more tension than the well-rested group, although these differences were not found to be statistically significant. The well-rested morning watchstanders showed more tension during the last hour of the watch in comparison to the first hour; the fatigued afternoon group showed the opposite effect. Whether the cause of muscle tension related to the stress of fatigue or to alertness was not directly measurable. Nevertheless, it seems sensible that the well-rested group, which showed more tension, would have been more alert rather than stressed; likewise, the fatigued group which showed less tension, would have been alert and more stressed. Thus, it appears that the muscle tension measured in this study was a sign of alertness rather than fatigue. This conclusion is also consistent with the findings of Kennedy and Travis (1948) which indicate that increases in muscle tension are associated with alertness and efficient signal detection.

3.6.5 PHYSIOLOGICAL RESPONSE TO DETECTION OF AN INCIDENT

An evaluation of the effects of sleep and watch conditions on the physiological arousal associated with the detection of a ship or instrument failure was conducted by focusing on the "detection segment" of each hour. During the first and last hours of the watch, either a VLCC or containership was detected. During the second and third hours of the watch, either a destroyer or instrument failure was noticed. Since several subjects failed completely to notice the instrument failure, the sample sizes in the second and third hours were fairly small.

Table 19 presents the means for EMG, physical activity and EKG collected during the detection segment for each experimental group. During all four hours, the fatigued groups exhibited more muscle tension at the time of detection when compared to the well-rested groups. However, statistical testing did not demonstrate the reliability of this observation. The standard deviations associated with these means were very high for some of the experimental groups, virtually eliminating the possibility of finding statistically significant differences with such small sample sizes. Unfortunately, the sizes of the standard deviations did not follow any particular pattern, which precluded many insights into the reasons for such disparate data. One possible source of the high variability may have been inadequate securing of the electrodes. If an electrode measuring EMG loosened during the course of the watch, it may not have been noticeable but could nevertheless have affected the readings. On the other hand, previous literature (e.g., Kennedy and Travis, 1948) indicates that some individuals show remarkably high variability in muscle tension. Such variability appears to be highly individualistic. In addition, high variability in EMG is associated with lower levels of alertness. Appendix E illustrates each individual subject's EMG and physical activity scores over

TABLE 18. LEVEL OF PHYSIOLOGICAL ACTIVITY ONE SEGMENT BEFORE DETECTION

		Full Sleep Morning	Full Sleep Afternoon	No Sleep Morning	No Sleep Afternoon
First Hour					
EMG	Mean	199.4	175.9	228.9	259.7
	SD	46.5	89.8	119.9	103.1
	n	6	6	6	7
Physical	Mean	26.9	11.9	12.1	18.9
Activity	SD	33.0	10.2	4.9	22.4
	n	6	6	6	7
EKG	Mean	83.5	84.5	89.3	86.6
	SD	19.8	2.1	14.4	8.4
	n	6	2	4	7
Second Hour					
EMG	Mean	270.3	215.7	306.5	270.9
	SD	134.5	145.8	155.8	100.8
	n	4	4	5	6
Physical	Mean	17.5	16.3	13.8	18.7
Activity	SD	11.4	16.0	4.7	17.0
	n	4	4	5	6
EKG	Mean	87.8	86.0	92.5	81.8
	SD	16.4	4.2	11.1	6.5
	n	4	2	4	6
Third Hour					
EMG	Mean	197.2	207.3	230.2	269.2
	SD	55.9	100.9	102.4	92.7
	n	5	5	4	6
Physical	Mean	19.6	11.7	12.0	16.3
Activity	SD	16.5	7.8	4.4	13.7
	n	6	5	4	6
EKG	Mean	82.3	89.0	85.5	85.4
	SD	18.1	0	16.3	10.4
	n	6	1	2	5
Fourth Hour					
EMG	Mean	260.3	165.6	264.3	289.1
	SD	56.7	65.9	148.4	94.2
	n	5	6	6	6
Physical	Mean	15.3	12.3	22.8	16.9
Activity	SD	5.8	7.7	27.5	15.6
	n	5	6	6	6
EKG	Mean	83.6	80.0	88.0	81.8
	SD	20.8	8.5	9.8	6.6
	n	5	2	4	5

TABLE 19. LEVEL OF PHYSIOLOGICAL ACTIVITY DURING DETECTION SEGMENT

		Full Sleep Morning	Full Sleep Afternoon	No Sleep Morning	No Sleep Afternoon
First Hour					
EMG	Mean	231.5	198.3	251.5	239.2
	SD	40.5	111.8	70.9	70.2
	n	6	6	6	7
Physical	Mean	27.6	12.7	16.0	15.8
Activity	SD	30.6	9.9	8.8	13.7
	n	6	6	6	7
EKG	Mean	85.5	84.5	91.0	86.3
	SD	20.0	2.1	15.3	8.1
	n	6	2	4	7
Second Hour					
EMG	Mean	196.3	243.2	261.5	259.8
	SD	45.1	129.1	156.3	88.9
	n	4	4	5	6
Physical	Mean	21.8	17.2	21.0	21.3
Activity	SD	15.8	7.5	14.2	20.2
	n	4	4	5	6
EKG	Mean	88.0	83.0	93.3	82.3
	SD	16.2	7.1	10.3	6.7
	n	4	2	4	6
Third Hour					
EMG	Mean	240.5	207.2	197.9	283.2
	SD	92.6	80.4	80.1	112.3
	n	5	5	4	6
Physical	Mean	21.3	12.1	10.5	16.2
Activity	SD	16.3	8.4	3.9	10.0
	n	6	5	4	6
EKG	Mean	83.2	86.0	87.5	84.4
	SD	18.4	0	20.5	8.0
	n	6	1	2	5
Fourth Hour					
EMG	Mean	237.0	214.5	263.1	279.2
	SD	43.0	93.9	132.6	65.5
	n	5	6	6	6
Physical	Mean	14.9	12.3	21.9	18.4
Activity	SD	5.4	6.4	25.9	13.6
	n	5	6	6	6
EKG	Mean	84.2	80.5	87.3	81.6
	SD	17.3	9.2	10.3	6.0
	n	5	2	4	5

time. Those figures graphically depict the dispersion of data just discussed.

The effect of number of hours on watch on arousal during detection was evaluated by comparing EMG and physical movement levels during the first hour detection segment with levels during the fourth hour detection segment using a paired-comparison t-test. Although the means were in a direction such that EMG levels were somewhat higher during the fourth hour than during the first, this trend did not achieve significance (see Table D24). No trend was found for physical movements.

To investigate the extent to which sleep and watch condition influenced the change in arousal over time, a difference score was computed by subtracting the level at the fourth detection segment from the level at the first detection segment. This difference score was then used as the dependent variable in a 2 X 2 ANOVA. No significance effects of sleep or watch condition were found.

In addition to examining physiological activity recorded during the detection segment, data recorded during the detection segment was also compared to data collected during the previous segment. By computing a difference score between the detection segment and one segment before detection, it was possible to measure the increment or decrement in physiological activity associated with detection. Table 20 provides the means and standard deviations for this analysis. Positive numbers indicate an increment. Analyses of variance on the data from each hour investigated the effects of sleep and watch condition on this change. No trend emerged across all four hours; however, during the second hour, the morning watch groups showed a much stronger tendency to decrease in EMG level at detection than did the afternoon watch groups. This tendency was significant, $F(1,18) = 4.40$, $p = 0.05$ (see Table D25). During the third hour, it was the no-sleep morning watch group that showed the strongest tendency to decrease in EMG level at detection. This led to a significant interaction of sleep and watch conditions, $F(1,19) = 5.16$, $p = 0.04$ (see Table D26). With respect to EKG, the cardiac rate of morning watchstanders increased with detection during the first hour to a significant degree more than the afternoon watchstanders, $F(1,15) = 3.99$, $p = 0.06$. During the second hour the fatigued mates appear to increase their EKG rate at detection, although the reliability of this finding is questionable, $F(1,12) = 3.27$, $p = 0.10$.

To assess the effect of number of hours worked on responses to an incident, a correlated t-test was performed. This compared the increment at the first hour with the increment at the fourth hour for each subject. No noteworthy trends were apparent across all subjects.

3.6.5.1 Summary of Physiological Response to Detection of An Incident

The fatigued groups appeared to exhibit more muscle tension during the time of detection than the well-rested groups. In general, muscle tension increased as the watch progressed.

Comparisons of EMG level during detection to immediately before detection indicated that during the second hour, the morning watchstanders decreased their tension once a sighting was made, while the afternoon watchstanders remained as tense during the sighting as before. This was also observed during the third hour when the no-sleep morning group showed a remarkable decline in muscle tension once a sighting was made.

3.6.5.2 Recovery Period

The physiological recovery period was measured by averaging scores from the three segments immediately after the detection segment. This yielded data about the general state of physiological arousal following an attention-getting event. Table 21 shows the average recovery periods for each group over time for EMG and physical activity. There were no apparent effects of sleep or watch condition on this average at any hour during the watch. When the change in average from first to fourth incident was compared for all subjects using a t-test, and for effects of sleep and watch condition using an ANOVA, no statistical significance was found either.

As with the resting level data, recovery was also evaluated as an incremental change in physiological arousal following a period of detection. This measurement simply required the computation of a difference score between the detection segment and one segment after detection. This difference score was a measure of immediate recovery and proved sensitive to several factors. For example, the afternoon groups tended to have a higher level of arousal after detection than during. The morning groups, on the other hand, tended to exhibit higher levels during the detection segment. This effect of watch condition was consistent for EMG during the first three incidents, and achieved significance during the second hour, $F(1,15) = 8.24$, $p = 0.01$.

TABLE 20. DIFFERENCE BETWEEN PHYSIOLOGICAL ACTIVITY OF DETECTION SEGMENT AND ACTIVITY ONE SEGMENT BEFORE DETECTION

		Full Sleep Morning	Full Sleep Afternoon	No Sleep Morning	No Sleep Afternoon
First Hour					
EMG	Mean	32.1	22.4	22.5	-20.6
	SD	33.3	32.7	63.4	66.5
	n	6	6	6	7
Physical Activity	Mean	0.7	0.7	3.9	-3.1
	SD	4.0	0.8	5.8	9.0
	n	6	6	6	7
EKG	Mean	2.0	0.0	1.8	-0.3
	SD	1.5	0.0	2.6	2.2
	n	6	2	4	7
Second Hour					
EMG	Mean	-74.0	27.5	-45.0	-11.1
	SD	99.2	50.8	38.0	63.0
	n	4	4	5	6
Physical Activity	Mean	4.3	0.9	7.2	2.5
	SD	5.6	10.1	15.6	3.6
	n	4	4	5	7
EKG	Mean	0.3	-3.0	0.8	0.5
	SD	2.1	2.8	1.3	1.8
	n	4	2	4	6
Third Hour					
EMG	Mean	43.3	-0.1	-32.3	14.0
	SD	51.7	33.8	57.2	33.0
	n	5	5	4	6
Physical Activity	Mean	1.7	0.4	-1.5	-0.1
	SD	1.7	1.3	1.3	5.0
	n	6	5	4	6
EKG	Mean	0.8	-3.0	2.0	1.3
	SD	2.2	0.0	4.2	2.2
	n	6	1	2	4
Fourth Hour					
EMG	Mean	-23.2	48.9	-1.2	-9.9
	SD	87.5	116.9	46.5	51.5
	n	5	6	6	6
Physical Activity	Mean	-0.4	0.0	-0.8	1.4
	SD	2.7	1.9	4.0	19.0
	n	5	6	6	6
EKG	Mean	0.6	0.5	-0.8	-0.2
	SD	4.8	0.7	2.1	1.9
	n	5	2	4	5

NOTE: Positive numbers indicate an increment, negative numbers indicate a decrement from first to fourth hour.

TABLE 21. PHYSIOLOGICAL ACTIVITY AVERAGED ACROSS THREE SEGMENTS FOLLOWING DETECTION

		Full Sleep Morning	Full Sleep Afternoon	No Sleep Morning	No Sleep Afternoon
First Hour					
EMG	Mean	214.7	209.5	257.7	233.5
	SD	70.9	110.1	97.3	70.4
	n	6	6	6	7
Physical	Mean	16.5	12.4	19.1	16.9
Activity	SD	9.2	8.7	12.2	17.8
	n	5	6	6	7
Second Hour					
EMG	Mean	255.2	196.5	286.5	242.4
	SD	99.1	119.3	145.3	70.0
	n	4	4	5	6
Physical	Mean	37.7	13.8	19.6	18.0
Activity	SD	40.4	10.2	16.6	13.3
	n	4	4	5	6
Third Hour					
EMG	Mean	233.5	215.6	205.4	280.3
	SD	85.8	95.0	77.1	107.0
	n	5	5	4	5
Physical	Mean	27.6	15.3	11.1	13.6
Activity	SD	26.2	16.2	2.3	5.7
	n	6	5	4	6
Fourth Hour					
EMG	Mean	281.7	186.8	260.4	258.9
	SD	129.3	82.0	136.7	74.1
	n	5	5	6	5
Physical	Mean	24.3	11.3	14.3	18.5
Activity	SD	26.9	5.9	9.8	13.4
	n	5	5	6	5

This is summarized in Tables 22 and D27. During the fourth hour, however, all groups showed less activity after the incident than before. No trends were apparent for the physical movement. The effect of time of watch on this recovery was investigated, again by comparing the recovery score of the first hour with the recovery score at the fourth hour using a paired comparison t-test. A tendency for subjects to decrease in EMG after the fourth hour's incident, but to increase in EMG after the incident at the first hour was shown, although the difference did not achieve significance.

The first segment following detection was also examined as a measure of immediate recovery. Table 23 shows the means and standard deviations for this measure. Although the full-sleep groups appeared to have a somewhat lower level of EMG than did not the no-sleep groups, this difference did not achieve significance.

Finally, recovery was quantified over a longer period of time by taking the difference between the scores averaged across the three segments prior to detection and the scores averaged across the three segments following detection.

TABLE 22. DIFFERENCE BETWEEN PHYSIOLOGICAL ACTIVITY AT DETECTION SEGMENT AND ONE SEGMENT AFTER DETECTION

		Full Sleep Morning	Full Sleep Afternoon	No Sleep Morning	No Sleep Afternoon
First Hour					
EMG	Mean	-11.9	25.0	-9.1	9.0
	SD	62.3	36.0	29.8	36.6
	n	6	6	6	7
Physical	Mean	-0.4	1.2	8.5	3.7
Activity	SD	7.5	4.3	18.2	9.3
	n	5	6	6	7
Second Hour					
EMG	Mean	41.9	-60.35	13.8	-16.6
	SD	63.0	54.6	17.6	44.2
	n	4	4	5	6
Physical	Mean	15.5	-2.5	-2.5	0.1
Activity	SD	34.4	7.7	1.2	4.2
	n	4	4	5	6
Third Hour					
EMG	Mean	0.5	-15.5	32.5	-25.7
	SD	42.9	19.9	97.1	48.4
	n	5	5	4	6
Physical	Mean	5.2	1.4	1.6	-3.2
Activity	SD	13.5	2.9	6.1	2.6
	n	6	5	4	6
Fourth Hour					
EMG	Mean	-5.1	-25.6	32.5	-25.7
	SD	56.4	28.5	25.3	45.4
	n	5	6	6	5
Physical	Mean	11.0	-1.6	-8.3	0.3
Activity	SD	21.9	1.2	18.9	6.8
	n	5	6	6	5

Table 24 shows these difference scores for all hours, and indicates that no clear trend emerged using this measure. Evaluating the effects of time on this measure by using a t-test to compare the difference score at the first hour with the difference score at the last hour also yielded nothing of significance.

3.6.5.3 Summary of Recovery Period of Physiological Measures

The afternoon watchstanders tended to have a higher level of EMG arousal after detection rather than before, which may indicate a slow recovery from arousing events. The morning watchstanders, on the other hand, recovered quickly, showing a decrease in muscle tension following a sighting. In neither case were the results statistically significant. The meaning of such results is not clear. The gradual decrease in muscle tension following an incident may indicate a continual level of alertness in the afternoon groups. The association of muscle tension with alertness is consistent with the earlier work in this area.

TABLE 23. LEVEL OF PHYSIOLOGICAL ACTIVITY ONE SEGMENT AFTER DETECTION

		Full Sleep Morning	Full Sleep Afternoon	No Sleep Morning	No Sleep Afternoon
First Hour					
EMG	Mean	219.6	223.3	242.4	248.2
	SD	82.3	99.7	91.0	74.1
	n	6	6	6	7
Physical Activity	Mean	15.1	13.8	24.5	19.6
	SD	6.1	8.2	21.4	22.6
	n	5	6	6	7
Second Hour					
EMG	Mean	238.2	182.8	275.4	243.1
	SD	97.5	99.2	149.4	71.3
	n	4	4	5	6
Physical Activity	Mean	37.3	14.6	18.6	21.4
	SD	38.7	9.2	14.3	23.3
	n	4	4	5	6
Third Hour					
EMG	Mean	241.0	191.7	230.4	257.5
	SD	97.2	79.9	115.8	91.3
	n	5	5	4	6
Physical Activity	Mean	26.5	13.5	12.1	13.0
	SD	26.7	11.1	4.3	8.1
	n	6	5	4	6
Fourth Hour					
EMG	Mean	231.9	188.9	261.1	260.6
	SD	38.4	85.8	143.6	77.6
	n	5	6	6	5
Physical Activity	Mean	25.8	10.7	13.7	17.8
	SD	24.5	6.7	8.3	13.3
	n	5	6	6	5

TABLE 24. DIFFERENCE BETWEEN AVERAGE RESTING LEVEL AND AVERAGE RECOVERY PERIOD[1]

		Full Sleep Morning	Full Sleep Afternoon	No Sleep Morning	No Sleep Afternoon
First Hour					
EMG	Mean	11.1	32.9	32.4	-31.6
	SD	44.6	39.3	31.1	56.4
	n	5	6	4	6
Physical	Mean	4.6	0.0	-3.0	-1.0
Activity	SD	3.3	1.7	8.5	1.5
	n	4	6	4	6
Second Hour					
EMG	Mean	8.8	-11.7	-2.0	-23.8
	SD	36.1	16.9	20.1	41.4
	n	4	4	5	6
Physical	Mean	18.6	-1.4	0.6	2.2
Activity	SD	37.2	3.4	7.0	3.1
	n	4	4	5	6
Third Hour					
EMG	Mean	27.6	5.8	-12.9	8.5
	SD	38.6	19.0	41.1	37.2
	n	5	5	4	5
Physical	Mean	6.6	3.1	-1.9	0.3
Activity	SD	16.7	8.1	2.7	2.8
	n	6	5	4	6
Fourth Hour					
EMG	Mean	18.1	-10.9	14.0	-22.2
	SD	76.5	124.8	23.9	46.2
	n	5	5	6	5
Physical	Mean	6.2	-5.5	-7.8	1.8
Activity	SD	20.6	9.1	19.5	5.1
	n	5	5	6	5

CHAPTER 4

DISCUSSION OF RESULTS AND CONCLUSIONS

Note

Table 25 tabulates the interesting findings of this experiment. The reader may find consulting this table to be a useful prelude to this chapter. Although an alpha level of 0.10 was utilized for statistical significance in this study, several p values greater than 0.10 are reported in Table 25. These p values are included because the direction of the means were of interest in interpreting the results. They were not reported as statistically significant in Chapter 3, but are included here for the purposes of discussion.

4.1 REVIEW OF EXPERIMENTAL GOALS

The primary purpose of this investigation was to evaluate the effect of sleep deprivation, time of watch, and length of time on watch on the effectiveness of a watchstanding officer.

Sleep deprivation was chosen as a variable because it is a common source of fatigue, which was the actual concept of interest. In addition to sleep loss, fatigue may result from several conditions including: strenuous physical exericse, long work hours, mental stress, and medication. Of these possible causes, long work hours and sleep deprivation are two of the more common sources of fatigue among deck officers. These two sources are also sometimes found together, such as when work on a critical task is sustained through regular sleep periods.

Because it is a common occurrence, is easily quantified and controlled, and can be manipulated without the use of instrumentation, sleep deprivation was selected as the experimental variable.

Circadian rhythm was also a variable of interest because it has been related to work performance. Two different watch schedules were selected for study because of their relationship to circadian rhythm and their congruity with real world watch schedules. Perhaps a better method for studying the effects of circadian rhythm would have been to select subjects to perform as watchstanders during their individual high and low points. This method was not selected because it did not permit advance scheduling of simulator time and because it would have required a fair degree of dedication to the experiment by the subject. (In fact, each subject was requested to chart his temperature every two hours for one week prior to the experiment; however, many subjects supplied incomplete or unreliable data making even **post hoc** analysis difficult.) The morning and afternoon periods of the day were chosen as an indirect means of studying the effect of circadian rhythm because they have been associated with different cycles of the circadian rhythm of most people.

The length of time on watch was chosen as an experimental variable because it was thought that the boring effects of watchstanding would accumulate as the watch progressed, the effects of fatigue would also accumulate, and an individual's point on this circadian cycle would shift during the course of four hours.

All of these factors – fatigue, circadian rhythm, and boredom – were of interest because of their potential impact on the attentiveness of watch officers. Inattention can result in failure to notice a critical event and may result in a severe casualty. It was the perceived relationship of these factors to casualty that led to the second purpose of this investigation: to evaluate the efficiency of using simulation to study such potentially hazardous relationships.

The evaluation of simulation methods focused on the perceived realism of the simulation, the validation of its realism using quantifiable measures, and the sensitivity of various performance measures to simulated watchstanding performance. The demonstration of the reliability and validity of performance measures is essential to the correct interpretation of experimental results. In this particular experiment certain measures were used, such as mood scales, whose reliability and validity have already been demonstrated in other applied settings. Nevertheless, it was considered important to test the transferability of these measures to the marine simulation environment. Other performance measures, such as ship detection tasks, were

TABLE 25. SUMMARY OF INTERESTING RESULTS[1]

Variable Name	Main Effects			Interactions			
	Sleep	Watch	Test Time	S x W	T x S	T x W	T x S x W
Detection of First Ship	0.16 0.11						
Detection of Destroyer		0.13 0.14					
Average Detection of All Ships	0.11	0.13					
Change in Resting Level of Radar from First to Last Hour			0.007				
Change in Resting Level of Radar from First to Last Hour			0.003				
Frequency of Radar Use During First Hour				0.12			
Difference in Resting Level of Radar Frequency from First to Last Hour		0.12		0.06			
Incremental Change in Radar Duration at Detection During First Hour	0.01						
Incremental Change in Radar Duration at Detection During Fourth Hour		0.10					
Recovery of Radar Duration Following Detection During Fourth Hour		0.05					
NPRU Positive	0.006						
NPRU Negative	0.0001	0.02	0.02				
Stanford Sleepiness	0.0006						
POMS Fatigue	0.0001						
POMS Vigor	0.002			0.12			
POMS Confusion				0.17		0.11	
POMS Anger		0.11	0.07			0.11	
VSS 2 Letters Lines Scanned	0.13						
VSS 4 Letters Lines Scanned	0.11						
VSS 6 Letters Lines Scanned					0.10		

[1] Values in boxes are p values. Only p values less than 0.20 are reported. Where 2 values appear in a box, the value in the upper left is taken from an analysis of variance summary table, and the lower right is taken from an analysis of covariance table.

TABLE 25. SUMMARY OF INTERESTING RESULTS[1] (Cont'd)

Variable Name	Main Effects			Interactions			
	Sleep	Watch	Test Time	S x W	T x S	T x W	T x S x W
VSS 2 Letters Correct Detections			0.04				
VSS 4 Letters Correct Detections				0.05 0.10			0.08 0.06
Digit Span Forward			0.01				
EMG Resting First Hour	0.09						
Change in EMG Resting from First to Last Hour	0.13	0.07					
Change in EMG at Detection from First to Last Hour			0.15				
Incremental Change in EMG at Detection Second Hour		0.05					
Incremental Change in EMG at Detection Third Hour		0.04					
EMG Recovery Second Hour		0.01					

[1] Values in boxes are p values. Only p values less than 0.20 are reported. Where 2 values appear in a box, the value in the upper left is taken from an analysis of variance summary table, and the lower right is taken from an analysis of covariance table.

TABLE 25. SUMMARY OF INTERESTING RESULTS (Cont'd)

Variable Name	Synopsis
Detection of First Ship	Fatigued groups took longer to detect ship
Detection of Destroyer	Afternoon groups took longer
Average Detection of All Ships	Fatigued and afternoon groups took longer
Change in Resting Level of Radar from First to Last Hour	Radar use increases from first to last hour
Change in Resting Level of Radar from First to Last Hour	Radar use increases from first to last hour
Frequency of Radar Use During First Hour	Fatigued morning group uses radar less
Difference in Resting Level of Radar Frequency from First to Last Hour	Fatigued morning group increased radar use from first to last hour; fatigued afternoon group decreased
Incremental Change in Radar Duration at Detection During First Hour	Rested groups increased radar use at detection more than fatigued groups
Incremental Change in Radar Duration at Detection During Fourth Hour	All but fatigued morning group increased radar use at time of detection
Recovery of Radar Duration Following Detection During Fourth Hour	Afternoon groups used radar more following detection
NPRU Positive	Rested groups were more positive
NPRU Negative	Fatigued groups were more negative, especially fatigued group at beginning of watch
Stanford Sleepiness	Fatigued groups were more sleepy
POMS Fatigue	Fatigued groups reported more fatigue
POMS Vigor	Fatigued groups reported less vigor, especially fatigued morning group
POMS Confusion	Fatigued morning group showed increase in confusion as watch progressed
POMS Anger	Afternoon groups reported more anger, especially at beginning of watch
VSS 2 Letters Lines Scanned	Fatigued groups scanned fewer lines
VSS 4 Letters Lines Scanned	Fatigued groups scanned fewer lines
VSS 6 Letters Lines Scanned	Fatigued groups improved as watch progressed, while rested groups got worse
VSS 2 Letters Correct Detections	Correct detections increased from beginning to end of watch
VSS 4 Letters Correct Detections	Fatigued afternoon group improved, while resting morning got worse as watch progressed
Digit Span Forward	Short term memory improved after watch
EMG Resting First Hour	Fatigued groups had higher EMG
Change in EMG Resting from First to Last Hour	Fatigued afternoon groups had higher EMG at first hour rather than fourth
Change in EMG at Detection from First to Last Hour	Overall, EMG increased over time
Incremental Change in EMG at Detection Second Hour	Morning groups decreased EMG at detection
Incremental Change in EMG at Detection Third Hour	Fatigued morning group decreased EMG at detection
EMG Recovery Second Hour	Afternoon groups decreased EMG after detection more than morning groups

specific to the marine setting and were being evaluated as indicators of watch performance for the first time in marine simulation.

From the subject's perspective, performance measures were basically of two types: directly related or not related to work. For example, radar observation and ship detection carried substantial face validity because they were directly related to the taks of a real watch. On the other hand, tests of short-term memory and mood appeared superfluous to the task of watchstanding, and were. However, both types were needed in order to evaluate the simulation setting and draw conclusions for future research. For instance, while ship detection may be sensitive to sleep deprivation, time of day, or length of time on watch, the specific bearing, range and speed of the vessel may contribute substantially to the liklihood of its detection, regardless of how well rested the watchstander is. If such specific characteristics are important for detection, then future research in fatigue would be ill-conceived if it were to rely solely on such work-related measures. Thus, it would be advisable to also incorporate measures external to the specific task into the research design. Such external measures could be applied in many different situations, thereby eliminating the potential problem of work-related measures with situations-specific sensitivity. Of course, while the use of external measures guarantees flexibility in their utilization, they do not have much face validity. It, therefore, becomes important to "tie" task-related measures to unrelated ones. The study presented in this report attempted to do just that.

The assessment of the sensitivity of various performance measures to fatigue is not only important to research — it is an essential step to the creation of an index of readiness for assuming a watch. Should the Maritime Administration, or NTSB, or USCG wish to measure fitness prior to assuming a watch, it would be necessary to have data concerning the measurement of alertness. A scientific evaluation of such performance measures would meet this requirement.

4.2 FULFILLMENT OF EXPERIMENTAL GOALS

When one reflects upon the success and shortcomings of this experiment, two thoughts come to mind. First, after reviewing all the data and the subjective reports, one feels convinced that 24 hours of sleep deprivation has a deleterious effect on watchstanding. The data also indicate, although not as clearly, that afternoon watchstanders are more negatively affected by fatigue than their morning counterparts. Second, this sense of conviction is derived more from many weak indicators, rather than several strong, statistically significant results. After analyzing voluminous data, one is left with dozens of trends almost all in the predicted direction, but with only a few strong indicators of the effects of the experimental variables.

The lack of strong experimental results may be due to several reasons. First, the sample size of 25 used in this particular study did not provide sufficient power to the experimental design to permit one to reliably see strong experimental effects. Because of the extraordinary expense associated with simulation research (e.g., each simulator hour costs $800), sample sizes used in these studies are generally smaller than in traditional psychological research. Although these smaller sample sizes typically reduce the probability of observing significant effects, their impact on the power of simulation studies is often mitigated somewhat by the greater effect sizes frequently found in marine simulation research. For example, given a study as this one, where tests of the main effects utilize 12 subjects per group, a two-tailed alpha level = 0.10, and power = 0.80, it would be necessary to observe differences of at least one standard deviation between sample means in order to yield statistically significant results. Such differences are relatively large, but nevertheless were observed in this study, particularly in measures relating to subjective feelings.

A second possible reason for not observing strong effects of sleep deprivation and related factors on watchkeeping effectiveness was the inherent motivation and level of skill of the mates used in the study. The mates standing watch on the CAORF simulator bridge knew they were being observed and could very well have responded with increased motivation, which could have then counteracted the effects of the experimental factors. Their subjective reports certainly indicated that they felt some impairment of their performance due to sleep deprivation and length of time on watch, and these subjective reports were corroborated by the strong results seen with subjective measures of mood. Nevertheless, the statistical tests conducted on work-related measures yielded weaker effects, with several p values in the 0.10 to 0.15 range. Thus, it seems clear that the mates felt tired, thought their performance was affected by their feelings of fatigue, yet were nevertheless able to perform their jobs adequately. Both an increased level of motivation associated with participation in the study, an intrinsic sense of professionalism, and the experimental requirement that the mates perform a task in which they were well-skilled and practiced could have served to increase the likelihood of good watchkeeping performance. This explanation is consistent with literature cited in Chapter 1 which reported that mates on the CAORF bridge are more active than their counterparts in the real world, and that fatigue is less likely to affect a well-practiced skill as opposed to a newly acquired skill.

A third possible reason for the failure to observe strong experimental effects may be that the true effect of 24 hours sleep deprivation on watchstanding is, in fact, weak. It may be that 24 hours of sleep loss is simply not adequate to induce potent deleterious efforts on watchstanding. Perhaps despite the fact that mates on actual ships feel tired, they do not let it infringe on their effectiveness in standing a typical uneventful watch.

In sum, the experiment was quite successful in providing the first scientifically sound simulation study of fatigue in the marine setting, but fell short of making incontrovertible statements about the effects of fatigue. Nevertheless, the results provide data which indicate that the work/rest schedules of mates have an effect on their watchkeeping and bear further investigation. The extent of sleep loss and the particular characteristics of a watch required to make a strong, unequivocal impact on watchkeeping effectiveness can be examined in future research. The results of the present study provide insight into research techniques which can be used in such subsequent marine simulation research.

4.2.1 REALISM

Based on the responses of the subjects during post-experimental debriefing, we know that the CAORF simulator provides sufficient realism to make results from its research credible and generalizable. The most common suggestion made by subjects to improve realism was to add more personnel to the bridge, either in the form of a helmsman or other crew members coming and going, and to add more paperwork to their tasks.

The intention in designing this research was to design either a very boring or a very taxing scenario, since either condition can be stressful. The middle ground, while perhaps most probable in real life, may not have provided a sufficiently stressful condition to amplify the effects of fatigue. Many people equate realism with "highly probable in real life"; however, it is possible to generate a realistic situation which is, nevertheless, highly improbable. In fact, this is the **modus operandi** of much safety research. The boring situation was selected over the taxing one mostly for its simplicity. Since this was the first project of its kind on the CAORF simulator, it was desirable to keep it relatively simple (an intention which was probably violated by the mere scope of the project!). We are not convinced that this was the best decision, nor are we convinced otherwise. Rather, we have come away from this project with interest piqued about how fatigue, circadian rhythm, and hours on watch affect the watchstander who is mentally taxed. Does one become more alert out of necessity? Or does the system break down, leading to poor judgments and reaction time? Now that we know that the simulator is a viable tool for answering such questions, energy can be directed toward answering them.

4.2.2 EFFECT OF EXPERIMENTAL VARIABLES ON PERFORMANCE MEASURES

The subjective reports of the subjects show a clear influence of the experimental variables on their vigilance. Approximately one-third of the fatigued group were so affected that they began dropping off to sleep and more than two-thirds felt their performance was impaired by fatigue. More than half of all subjects felt that their vigilance declined as the watch progressed, and this feeling was exacerbated by fatigue. The objective data bears out some of their subjective feelings.

The fatigued watchkeeper was likely to detect ships more slowly than the well-rested watchkeeper. The deleterious effects of fatigue were compounded by standing the afternoon (as opposed to the morning) watch. One example of the differences between well-rested morning and fatigued afternoon mates was particularly striking. In detecting a destroyer approaching from astern, the fatigued afternoon group took 4.5 minutes longer to detect the ship than the well-rested morning group. This detection differential means that the destroyer traveled 2.25 miles further before the fatigued afternoon mates sighted it.

Not only did the fatigued afternoon watchstanders detect ships more slowly, they also decreased their monitoring of the radar during ship detection as the watch progressed, while other groups increased such monitoring over time. One would expect radar use to increase towards the end of a watch, as the mate is preparing for watch relief; yet the tired afternoon mates did not, on the average, do this. If only circadian rhythm was affecting performance, one would expect to see an improvement in performance in the afternoon rather than a decrement. However, the observed decrements in the fatigued afternoon watchstanders were probably the result of an accumulating and overpowering effect of fatigue during the course of the day. The fatigued morning watchstanders' use of the radar at the beginning of the watch was also affected such that radar monitoring in response to a ship sighting did not increase to the same degree as other groups. This finding is consistent with the hypothesis that the effects of fatigue and circadian low points may combine to impair watchstanding effectiveness.

Furthermore, the performance of fatigued deck officers was much more variable than that of rested officers. Across most work-related measures, the well-rested groups performed consistently, while the fatigued groups were fairly disparate in their performance. This may be due to the ability of some people to mitigate the effects of fatigue better than others. A statistical analysis was conducted to ascertain whether the experimental groups differed with respect to certain personality and biographical characteristics. There were no significant differences between groups on measures of recent life experiences, locus of control, speed and impatience, or job involvement. Only the scale measuring hard-driving and competitive characteristics on the Jenkins Activity Survey showed that those assigned to the afternoon watch groups were more hard-driving and competitive than those in the morning groups. Given that the fatigued afternoon group showed greater performance decrements than the other groups on several work-related tasks, it is unlikely that the additional degree of hard-driving, competitive style which they possessed was adequate to ameliorate the effects of fatigue. A detailed analysis of personality as a mediating factor was not undertaken as part of the present research; however, the data are available for future analyses. Likewise, the use of daydreaming or mental game playing may be a mediating factor and would be an interesting aspect to study.

There was also some evidence to suggest the fatigued morning watchstanders are affected more at the beginning of their watches, while the fatigued afternoon watchstanders are affected more towards the end of their watches. While the effect of length of time on watch was not generally strong with respect to most work-related measures, it was quite pronounced with respect to radar observation. The amount of radar monitoring, not associated with an incident detection, increased as the watch progressed, irrespective of sleep or watch condition. Such increased activity may possibly have been the result of an increased motivation to perform well, or a desire to counteract the accumulating effects of boredom. A more plausible and substantiated reason for the increase in radar monitoring towards the end of the watch is the tendency for watch officers to update the ship's status at that time, in preparation for turning over the watch to another officer.

Overall, it appears that fatigue had a significant effect on work-related measures. Regarding the effectiveness of the work-related measures in a simulated environment, their use is recommended in future research of this kind.

The measures of affect and cognition were sensitive to sleep deprivation, time of day, and number of hours on watch. Fatigued watchstanders reported more negative and fatiguing feelings on the NPRU, Stanford Subjective Sleepiness Scales, and the POMS Checklist than those who were well-rested. While this is not surprising, it is noteworthy. Until the completion of the present research, it had not been ascertained whether affect scales used in studies outside of the maritime industry would retain their validity in a marine simulation environment. These scales' sensitivity to fatigue in this study indicates their usefulness in other maritime simulation research.

With respect to time of watch, afternoon groups reported more negative affect than the morning groups. Such poorer mood corresponds to the poorer performance of the fatigued afternoon watchstanders on ship detection and radar observation tasks. Such correspondence between mood, fatigue, and work-related tasks provides further evidence for the importance of subjective feelings in work situations and is consistent with research conducted in laboratories (e.g., Moses, Lubin and Naitoh, 1974).

One particularly curious finding was that across all experimental groups, negative affect on the NPRU Scale decreased during the course of the watch (in contrast to measures of anger on the POMS Checklist which increased for the afternoon groups). It was somewhat surprising to find that test subjects, who were wired up with physiological monitoring equipment and required to stand a second boring watch, would report being in better rather than worse moods as the watch progressed. Perhaps the improvement in affect was indicative of test subject compliance or of anticipation that their participation was drawing near its end. Of even greater interest is that their improvement in mood over the course of the watch corresponds to increased radar monitoring (when no ship was in sight) during the watch. Thus, their morale appears to correspond to their watchstanding effectiveness.

Because of their sensitivity to sleep deprivation, time of day, and length of time on watch, the NPRU Positive and Negative Scale, the Stanford Subjective Sleepiness Scale, and the POMS Checklist are recommended for future research of this type in the marine industry.

The Visual Search and Scan Test also yielded some surprising findings: the fatigued groups performed worse than the rested groups on the easier two-letter target, but did better than the rested groups on the more difficult scanning tasks. Perhaps the level of effort required to adequately perform difficult scanning tasks arouses the fatigued watchkeepers. It is possible that they are so aware of their flagging attention, that they try harder than the others. It is also possible that staying up all night somehow improves one's ability to scan material and detect targets. This explanation is considered unlikely, since it is both counterintuitive and inconsistent with other performance measures, such as traffic ship detection. Given its sensitivity to sleep loss and the advantage of permitting the experimenter to manipulate its level of difficulty, the Visual Search and Scan Test is recommended for use in future research of this type.

The effectiveness of the Digit Span Test in this experimental setting is still undetermined. Performance on it (especially Digits Forward) improved over time, which may be a result of practice or of some other factor, such as circadian rhythm. In any case, such improvement is inconsistent with the hypothesis that accumulating fatigue impairs short-term memory. Some research (for example, Baddeley, et al., 1970) indicates that short-term memory is inversely related to circadian rhythm, such that short-term memory improves when body cycles are at their lower points. The analysis conducted on available data was not sufficiently refined to test this. In general, we would be inclined to not use the Digit Span Test in future work of this type.

The Baddeley Grammatical Transformation Test did not yield significant effects for any of the experimental variables. Furthermore, the scores were very close to perfect, resulting in somewhat of a ceiling effect. Perhaps the cognitive skill required for making logical transformations is not meaningfully affected by sleep deprivation or related factors. Its use in future work of this type is not recommended.

Of the three physiological measures assessed, EMG was clearly superior to physical activity as a measure of fatigue and circadian rhythm. Results of the heart rate analysis, which was conducted manually due to an instrumentation error, yielded data which was incomplete and possibly unreliable. Thus, conclusions cannot be drawn. Results indicated that, with the exception of the fatigued afternoon watchstanders, all groups showed higher levels of arousal on the EMG and physical activity prior to the fourth incident as compared to the first incident. This increase in arousal during the watch corresponds with the increased radar activity and improved affect noted earlier.

When measured during rest periods, the afternoon groups showed more muscle tension than the morning groups. Also, the afternoon groups showed a slight decline in muscle tension as the watch progressed, while the morning groups showed a rise in EMG as the watch progressed. This finding is completely consistent with the general circadian cycle which rises through the morning, flattens out in early afternoon, and begins a slight decline in late afternoon.

When EMG was measured during the actual time of detection, this pattern was not found. After detection occurred (that is, during the recovery period), the afternoon watchstanders tended to maintain a higher level of EMG longer than the morning groups which may indicate a slower recovery from arousing events. Whether this high level of EMG is indicative of something positive (such as alertness) or negative (such as stress) is debatable. However, because of its synchronicity with circadian rhythm, it may be considered an indicator of alertness. This interpretation is also consistent with the work of Kennedy and Travis (1948) which indicated that increases in muscle tension are associated with alertness and efficient signal detection.

When all of the dependent measures are examined together, one notes that the measures of work performance, affect, and EMG correspond to one another. For example, mates in this study were generally more active and alert in the fourth hour than in the first. This phenomenon was observed with respect to radar monitoring, mood, and EMG. Such results are encouraging in that they provide more confidence that the effects of the experimental factors were genuine and systemic. It is also noteworthy that the watchstanding mates, who are a much different group of subjects than those typically found in psychological research, responded to the effects of the independent variables in ways consistent with theory and findings derived from much different subject groups.

4.3 APPLICATION TO THE REAL WORLD

The four-on, eight-off watch schedule has been part of nautical tradition for hundreds of years. It appears to have served the industry adequately and is unlikely to be changed by most operators. Another nautical tradition, albeit an informal one, is the tendency to work long hours, sometimes with little sleep, under certain conditions. While this is occassionally necessary, it is also self-defeating when fatigue impairs performance. In recognition of this, the

Masters, Mates and Pilots (MMP) established union work rules requiring a deck officer to have at least six hours of rest in the past 12 hours prior to assuming a bridge watch after leaving port.[1] According to the mates who participated in this project, this rule is frequently violated. They reported that it is implicit on many ships that the watch will be kept even if the mate is excessively fatigued. This policy, they say, is frequently endorsed by the master. Many of the subjects were enthusiastic about this research because they felt, if scientific evidence demonstrating the deleterious effects of fatigue on vigilance were available, the ship operators and their masters would be more inclined to endorse the MMP work rules and let a tired man rest.

Because of the potential application of the data from this study by unions or shop operators, it is important that the interpretation of the results be conducted judiciously. When interpreting these results, one should bear several factors in mind. First, the data was collected from a relatively small sample size, thus limiting the chance of finding strong effects due to fatigue and related factors. With a larger sample size, it is possible that stronger effects of fatigue may have been found. Second, an alpha level of 0.10 was used to test for statistical significance, rather than the traditional level of 0.05. This more liberal alpha level served to increase the chance of finding statistically significant results, but with the greater attendant risk of falsely identifying significant effects. Third, although several significant effects were found utilizing an alpha level of 0.01, the number of significant effects were not overwhelming. Given an alpha of 0.10, one could expect 10 percent of the statistical tests to have yielded significant effects by chance. In fact, approximately 13 percent of the tests were statistically significant with about another 9 percent showing trends. Therefore, one must question the confidence one has in the results. Nevertheless, it is noteworthy that, with the exception of the Visual Search and Scan task, all the observed statistically significant effects were in the direction consistent with theory or common sense. If these significant effects were merely random "noise," one would expect the observed directions of means to be less systematic and more random. Thus, there is a plausible argument for noting and discussing these results but policies should not be based on them. Fourth, the study was conducted on a simulator, which introduces the possibility that the mates did not behave in a manner generalizable to the real world. Several precautions were taken in the experimental procedure to ensure as much similarity as possible to real world performance. The responses of the mates in conjunction with the results of work on the utilization of the CAORF simulator (cited in Chapter 1) leads one to believe that the mates were actually more attentive on the CAORF bridge than in the real world. Should this be the case, then one could expect (all things being equal) that the significant experimental effects observed on the simulator may actually be stronger in the real world. However, it is also possible that the mates' knowledge that he was standing watch in a simulated environment could have made him feel less responsible and therefore perform less effectively than in the real world. These factors should be kept in mind in interpreting and applying the results.

4.4 RECOMMENDATIONS

One must recognize and accept that with reduced manpower on modern ships, the failure of one man to assume his watch impacts the schedule of several others. It must also be recognized that the failure of one man to keep his watch **adequately** may have a devastating impact on the entire ship. Severe fatigue may significantly impair the adequacy of watchkeeping. The industry would be served well by research which defines the amount of sleep loss critical to performance and the amount and quality of sleep necessary for recovery of critical abilities. Given the results of this study, future research on watchstanding should be conducted in at least two other areas: at sea and in the offices of reporting agencies. The measurement devices tested in this research could be brought to sea as part of a field study of fatigue and watchstanding and the results could then be compared to those of simulation studies. In addition, reporting agencies such as the U.S. Coast Guard, the National Transportation Safety Board, and maritime casualty insurance companies should collect better information regarding sleep deprivation, time of day, and number of hours worked prior to an incident as a regular part of their casualty report procedure.

Research on the interrupted sleep of ship masters and on the work rotation cycles of pilots should also be given research priority. Simulation techniques may be particularly useful in controlling the extraneous factors which make it so difficult to conduct such research while on an actual ship. In general, simulation represents a valuable tool for research in these areas.

[1] "No officer shall be required or permitted to take charge of a watch upon leaving or immediately after leaving port unless he has at least six (6) hours off duty within the twelve (12) hours immediately preceding the vessel's actual sailing time." (Excerpted from: "1978-1981 Master Collective Bargaining Agreement Covering Offshore Vessels Under Contract with the International Organization of Masters, Mates, and Pilots")

CHAPTER 5

SUMMARY

5.1 BACKGROUND

Virtually all surveys of ship casualties conclude that approximately 80 percent have resulted, at least in part, from human error. Yet the factors which predispose ship officers to commit errors have not been the subject of much systematic research. The research described in this report was an initial effort to study the role of personnel fatigue and work/rest cycles as contributing factors to human error aboard ships.

Several problems associated with watchstanding which may have a deleterious effect on watchstanding effectiveness were considered in designing this research, namely: fatigue, boredom, and changes in body functioning associated with circadian rhythm. Since this research was the first fatigue investigation conducted on a marine simulator, several performance measures were also tested for their sensitivity to fatigue under simulated work conditions.

5.2 OBJECTIVES

The primary purpose of the research was to measure the effects of sleep deprivation, time of watch, and length of time on watch on a watchstander's work performance, work-related performance, mood, cognitive measures and physiological measures under a simulation of the conditions of mates standing open-sea watches.

A second purpose was to evaluate marine simulation as a method for studying the effects of sleep deprivation, time of watch, and length of time on watch on mariner performance. This evaluation incorporated the use of performance measures previously used in non-maritime industry research as well as performance measures specific to the maritime industry.

5.3 METHOD

The CAORF simulator was used as the laboratory in which watchstanding performance was assessed as a function of the effects of sleep deprivation, time of watch, and length of time on watch. Sleep deprivation was of primary interest because of the greater attention it has received in casualty reports and marine transportation reports, and because of the expectation that sleep deprivation would have greater effects on watchstanding than the other two variables. Nevertheless, time of watch and length of time on watch were also studied because of their expected interaction with sleep deprivation. The rationale for these expected effects was drawn from the literature reviewed and presented in Chapter 1.

The test subjects were experienced watchstanding mates divided into four groups on the basis of amount of sleep (none vs. 7.5 hours) and time of watch (morning vs. afternoon). They arrived one day prior to the experimental runs and received three hours of familiarization on the CAORF bridge during the first day. During the first evening, subjects were monitored closely to minimize the introduction of spurious events. Half of the subjects retired to bed at 2300 hours; the other half stayed awake all night and engaged in a number of strenuous activities. The next day, they each stood a four-hour, open-sea watch during which three ships appeared and one instrument failure occurred (a total of four "events" distributed over the four hours of each watch). Dependent measures were collected before, during, and after the watch period. This procedure was the refined outcome of a procedure tested during preliminary experimental trials.

Certain dependent variables were selected because of their direct relationship to the responsibilities of watchstanders. Watchstanding mates are primarily responsible for keeping a good lookout so that early detection of potential threats to the ship's safety can occur. Such threats include the presence of another ship which may collide with ownship, failure of equipment, or involvement with high seas and winds. The vigilant watchstander should detect the presence of such events very soon after they become discernible, and should conduct his watch in a manner which maximizes the probability of early detection. In this research, the work performance of a mate was measured by the rapidity with which he detected ships and an instrument failure.

Because the probability of an early detection should increase with the degree to which the watchstander consults the radar and moves around the bridge, radar observation and physical activity were studied as work-related measures.

Several measures not directly related to watchstanding performance were also examined in this study. The Profile of Mood States Checklist, the Naval Personnel Research Unit Affect Scales, and the Stanford Subjective Sleepiness Scale were administered in order to examine the effects of the independent variables on certain subjective feelings such as fatigue, vigor, anxiety, depression and confusion. The particular scales used to measure these feelings were selected because previous research has indicated that they are sensitive to sleep loss and time of day.

Measures of cognition were also administered in order to assess the effects of the independent variables on levels of cognitive functioning which may be necessary for watchkeeping. For example, the ability to recall items from short-term memory, visually search and scan the environment, and make logical transformations are necessary for sighting, fixing the position of, and reporting traffic ships at sea. Should sleep loss negatively affect these cognitive functions, then the performance of the watchstander may also be affected. Thus, measures of these cognitive functions were included as dependent variables in this study.

Physiological measures were also examined throughout the course of the watch. Measures of muscle tension (EMG) taken from the frontalis are indicators of alertness, which is a crucial prerequisite to effective watchstanding. Measures of heart rate (EKG) and heart rate variability are generally sensitive to changes in mental workload and alertness, as well as length of time on watch.

Finally, a subjective, retrospective report by the subject regarding his evaluation of the watch was included to gain insight into the watchstander's perspective on his performance.

When one considers these dependent variables together, one can infer a great deal about the effectiveness of a watch officer. Thus, this study attempted to systematically evaluate the effects of sleep deprivation and related operational factors on watchs ing effectiveness.

5.4 RESULTS

Table 25 summarizes the interesting results of this study. The subjective reports of the subjects showed a clear influence of the experimental variables on their vigilance. Approximately one-third of the fatigued group were so affected that they began dropping off to sleep and more than two-thirds felt their performance was impaired by fatigue. More than half of all subjects felt that their vigilance declined as the watch progressed, and this feeling was exacerbated by fatigue. The objective data bore out their subjective feelings.

The fatigued watchkeeper was likely to detect ships more slowly than the well-rested watchkeeper. The deleterious effects of fatigue were compounded by standing the afternoon (as opposed to the morning) watch.

Not only did the fatigued afternoon watchstanders detect ships more slowly, they also decreased their monitoring of the radar as the watch progressed, while other groups increased such monitoring over time. One would expect radar use to increase towards the end of a watch, as the mate is preparing for watch relief; yet the tired afternoon mates did not, on the average, do this. In addition, the fatigued morning watchstanders' use of the radar at the beginning of the watch was also affected such that radar monitoring in response to a ship sighting did not increase to the same degree as other groups. Thus, it appears that fatigue affected the performance of work-related tasks.

There was also some evidence to suggest the morning watchstanders are affected more at the beginning of their watches, while afternoon watchstanders are affected more towards the end of their watches.

Furthermore, the performance of fatigued deck officers was much more variable than that of rested officers. Across most work-related measures, the well-rested groups performed consistently, while the fatigued groups were fairly disparate in their performance. This may be due to the ability of some people to mitigate the effects of fatigue better than others. An analysis of personality as a mediating factor was not undertaken as part of the research; however, the data are available for future analyses. Likewise, the use of daydreaming or mental game playing may be a mediating factor.

Regarding the effectiveness of the work-related measures in a simulated environment, their use is recommended in future research of this kind.

The psychological measures were sensitive to sleep deprivation, time of day, and number of hours on watch. The use of the NPRU Positive and Negative Affect Scales, the

Stanford Sleepiness Scale, the Profile of Mood States Checklist, and the Visual Search and Scan Test are recommended for future use. This latter test yielded some surprising findings: the fatigued groups performed worse than the rested groups on the easier two-letter target test, but did better than the rested groups on the more difficult scanning tasks. Perhaps the level of effort required to adequately perform difficult scanning tasks alerts the fatigued watchkeepers. It is possible that they are so aware of their flagging attention that they try harder than the others. It is also possible that staying up all night somehow improves one's ability to scan material and detect targets. This explanation was considered unlikely, since it is both counterintuitive and inconsistent with other performance measures, such as traffic ship detection.

The effectiveness of the Digit Span Test in this experimental setting is still undetermined. Performance on it improved over time, which may be a result of practice or of some other factor, such as circadian rhythm. Furthermore, the Digits Forward Test was found to be more sensitive than the Digits Backward. Some previous research (reviewed in Chapter 1) indicated that short-term memory is inversely related to circadian rhythm, such that short-term memory improves when body cycles are at their lower points. The analysis conducted on available data was not sufficiently refined to test this. In general, the use of Digit Span in future work was not recommended.

The Baddeley Grammatical Transformation Test did not transfer well to the marine simulation setting. Its use in future work of this type is not recommended.

Of the three physiological measures assessed, EMG was clearly superior to physical activity as a measure of fatigue and circadian rhythm. Results of the heart rate analysis, which was conducted manually due to an instrumentation error, yielded data which was incomplete and possibly unreliable. Thus, conclusions cannot be drawn. When measured during rest periods, the afternoon groups showed more muscle tension than the morning groups. Also, the afternoon groups showed a slight decline in muscle tension as the watch progressed, while the morning groups showed a rise in EMG as the watch progressed. This finding is completely consistent with the general circadian cycle which rises through the morning, flattens out in early afternoon, and begins a slight decline in late afternoon. When EMG was measured during the actual time of detection, this pattern was not found. After detection occurred (that is, during the recovery period), the afternoon watchstanders maintained a higher level of EMG longer than the morning groups. Whether this high level of EMG was indicative of something positive (such as alertness) or negative (such as stress) is debatable. However, because of its synchronicity with circadian rhythm, it may be considered an indicator of alertness. This interpretation is also consistent with literature reviewed in Chapter 1.

5.5 CONCLUSIONS

It was concluded that fatigue does appear to have a deleterious effect on watchkeeping. However, the results of the study should be interpreted with caution because the results were based on a small sample size and a liberal alpha level. Future research on watchkeeping should be conducted with more emphasis on field studies and better tabulation of casualty data by regulatory and underwriting agencies. Other research should also be conducted on the effect of interrupted sleep on masters and the effect of extended work hours and rotation cycles on pilots.

REFERENCES

Alkov, R. A., & Borowsky, M. S. A questionnaire study of psychological background factors in U. S. Navy aircraft accidents. **Aviation, Space, and Environmental Medicine**, 1980, 51, 860–863.

Antrobus, J. S., Coleman, R., & Singer, J. L. Signal-detection performance by subjects differing in predisposition to day-dreaming. **Journal of Consulting Psychology**, 1967, 31, 487–491.

Aschoff, J. **Circadian rhythms; interference with and dependence on work-rest schedules.** Paper presented at the Symposium on Variations in Work-Sleep Schedules, San Diego, CA, September 1980.

Aschoff, J. Circadian systems in man and their implications. **Hospital Practice**, 1976, 14, 51–57.

Baddeley, A. D. A 3-minute reasoning test based on grammatical transformation. **Psychonomic Science**, 1968, 10, 341-342.

Baddeley, A. D., Hatter, J., Scott, D., & Snashell, A. Memory and time of day. **British Journal of Psychology**, 1970, 22, 605–609.

Bakan, P. Extraversion-introversion and improvement in an auditory vigilance task. **British Journal of Psychology**, 1959, 50, 325–332.

Bakan, P., Belton, J. A., & Toth, J. C. In D. N. Buckner & J. J. McGrath (eds.), **Vigilance: A Symposium**. New York: McGraw-Hill, 1963.

Blake, M. J. F. Time of day effects on performance in a range of tasks. **Psychonomic Science**, 1967, 9, 349–350.

Brietson, C. A., McHugh, W., & Naitoh, P. **Prediction of pilot performance: Biochemical and sleep mood correlates under high workload conditions.** Presented at AGARD Aerospace Medical Panel Specialist Meeting, 29 April–3 May 1974, Oslo, Norway.

Broadbent, D. E. **Perception and Communication**. New York: Pergamon Press, 1958.

Bugge, J. F., Opstad, P. K., & Magnus, P. M. Changes in the circadian rhythm of performance and mood in healthy young men exposed to prolonged, heavy physical work, sleep deprivation, and caloric deficit. **Aviation, Space, and Environmental Medicine**, 1979, 50, 663–668.

CAORF Research Staff, **Validation of Computer Aided Operations Research Facility** (Report No. CAORF 90-7801-01), Kings Point, NY: National Maritime Research Center, January 1979.

Collins, W. E. **Some effects of sleep deprivation on tracking performance in static and dynamic environments** (FAA-AM-76-12), Oklahoma City, OK: FAA Civil Aeromedical Institute, October 1976.

Colquhoun, W. P. Circadian variations in mental efficiency. In W. P. Colquhoun (ed.) **Biological rhythms and human performance**. New York: Academic Press, 1971.

Cook, T. C., & Shipley, P. Human factors studies of the working hours of U.S. ship's pilots: 1. A field study of fatigue. **Applied Ergonomics**, 1980, 11.2, 85–92.

Coons, W. E. **Physiological Correlates of Vigilance Performance.** Naval Postgraduate School, CA, March 1977. (NTIS No. AD-A-039-15).

Corcoran, D. J. Personality and the inverted -u relation. **British Journal of Psychology**, 1965, **56**, 267-273.

Costantini, A. F., Braun, J. R., Davis, J. E., & Iervolino, A. The life change inventory: a device for quantifying psychological magnitude of changes experienced by college students. **Psychological Reports**, 1974, **34**, 991-1000.

Davies, D. R., Hockey, G. J., & Taylor, A. Varied auditory stimulation, temperament differences and vigilance performance. **British Journal of Psychology**, 1969, **60**, 453-457.

Deaton, M., Tobias, J. S., & Wilkinson, R. T. The effect of sleep deprivation on signal detection parameters. **Quarterly Journal of Experimental Psychology**, 1971, **23**, 449-452.

In: Dohrenwend & Dohrenwend (eds.), **Stressful life events: their nature and effects**. New York: Wiley. 1974, 73-86.

Eichman, W. J., Profile of mood states. In Buros, O. K. (ed.), **The Eighth Mental Measurements Yearbook** (Vol. 1). New Jersey: Gryphon Press, 1978.

Folkhard, S., Knauth, P., Monk, T. H., & Rutenfranz, J. The effect of memory load on the circadian variation in performance efficiency under a rapidly rotating shift system. **Ergonomics**, 1976, **19**, 479-488.

Frankmann, J. P., & Adams, J. A. Theories of vigilance. **Psychological Bulletin**, 1962, **59**, 257-272.

Froberg, J., Karlsson, C. G., Levi, L., & Lidberg, L. Circadian variations in performance, psychological ratings, catecholamine excretion, and diuresis during sleep deprivation. **International Journal of Psychology**, 1972, **2**, 23-36.

Hartley, L. R. A comparison of continuous and distributed reduced sleep schedules. **Quarterly Journal of Experimental Psychology**, 1974, **26**, 8-14.

Hockey, G. R. J., & Colquhoun, W. P. Diurnal variation in human performance: a review. In W. P. Colquhoun (ed.), **Aspects of Human Efficiency: Diurnal Rhythm and Loss of Sleep**, Cambridge, UK: The English Universities Press Limited.

Jenkins, C. D., Ph.D., Zyzanski, S. F., Ph.D., & Rosenman, R. H., M.D. **Jenkins Activity Survey Manual**, New York: Psychological Corporation, 1979.

Kennedy, J. L., & Travis, R. C. Prediction and Control of Alertness. I. Control of lookout alertness. **Journal of Comparative and Physiological Psychology**, 1948, **41**, 457-461.

Kennedy, J. L., & Travis, R. C. Prediction and Control of Alertness: II. Continuous Tracking. **Journal of Comparative and Physiological Psychology**, 1948, **41**, 203-210.

Kennedy, J. L., & Travis, R. C. Prediction of speed of performance by muscle action potentials. **Science**, 1947, **105**, 410-411.

Kirk, R. E. **Experimental Design: Procedures for the Behavioral Sciences**. Belmont, CA: Wadsworth, 1968.

Lisper, H. O., & Kjellberg, A. Effects of 24-hour sleep deprivation on rate of decrement in a 10-minute auditory reaction time task. **Journal of Experimental Psychology**, 1972, **96**, 287-290.

Mackie, R. R., & Miller, J. C. "Effects of Hours of Service, Regularity of Schedule, and Cargo Loading on Truck and Bus Driver Fatigue" (Final Report on DOT Contract HS-5-01142), Human Factors Research, Inc., Goleta, CA, 1978.

Mackworth, N. H. Some factors affecting vigilance. **The Advancement of Science**, 1957, **53**, 389-393.

McEwen, W. A., & Lewis, A. H. **Encyclopedia of Nautical Knowledge**. Cambridge, Maryland: Cornell Maritime Press, 1953.

O'Hanlon, J. **Heart Rate Variability: A New Index of Driver Alertness/Fatigue**. Goleta, CA: Human Factors Research Incorporated, July 1971.

Panel on Human Error in Merchant Marine Safety, Maritime Transportation Research Board, Commission on Sociotechnical Systems. **Human Error in Merchant Marine Safety**, Washington, DC: National Academy of Sciences, June 1976.

Patrick, G. W., & Gilbert, J. A. On the effects of loss of sleep. **The Psychological Review**, 1895, **3**, 471–483.

Roscoe, A. H. Use of pilot heart rate measurement in flight evaluation. **Aviation, Space, and Environmental Medicine**, 1976, 86–90.

Rother, D. Ship casualties — An analysis of causes and circumstances. **Lectures and Contributions**, **28**, Bremen, West Germany: Institute of Shipping Economics, 1980.

Rotter, J. B. **Social learning and clinical psychology**. Englewood Cliffs, NJ: Prentice-Hall, 1954.

Safety spotlight: This month—boredom at sea. **Safety at Sea**, January 1981, pp. 41–46.

Sarason, I. G., Sarason, B. R., & Johnson, J. H. **Stressful life events: measurement, moderators, and adaptation**. Office of Naval Research, VA: Organizational Effectiveness Research Program, October 1980, (NTIS No. AD-AO-91-754).

Sarason, I. G., Johnson, J. H., & Siegel, J. M. Assessing the impact of life changes: development of the life experiences survey. **Journal of Consulting and Clinical Psychology**, 1978, **46**, 932-946.

Siegel, J. M., Johnson, J. H., & Sarason, I. G. Mood states and the reporting of life changes. **Journal of Psychosomatic Research**, 1979b, **23**, 103-108.

Singer, J. L. **Daydreaming**. New York: Random House, 1966.

Singer, J. L., & Antrobus, J. S. A factor analytic study of daydreaming and conceptually-related cognitive and personality variables. **Perceptual and Motor Skills**, 1963. Monogr. Suppl. No. 3-V17.

Taub, J. M., & Berger, R. J. Performance and mood following variations in the length and timing of sleep. **Psychophysiology**, 1973, **10(6)**, 559-570.

Teichner, W. H. The detection of a simple visual signal as a function of time of watch. **Human Factors**, 1974, **16**, 339-353.

Thackray, R. I., Bailey, J. P., & Touchstone, R. M. **The effect of increased monitoring load on vigilance performance using a simulated radar display** (FAR-AM-77-18). Oklahoma City, OK: Civil Aeromedical Institute, Federal Aviation Administration, July 1977, (NTIS No. AD-AO-44558).

Thackray, R. I., Touchstone, R. M., & Bailey, J. P. A comparison of the vigilance performance of men and women using a simulated radar task. **FAA Office of Aviation Medicine Reports**, 1978 (Mar), No. 78-11.

United States Coast Guard. **Marine Casualty Report No. 254 — Tug CAROLYN and Barge WEEKS, Chesapeake Bay Bridge and Tunnel, September 20-21, 1972**, (USCG/NTSB-MAR-74-2). Washington, D.C.: National Transportation Safety Board, Bureau of Surface Transportation Safety, January, 1974.

Weckowicz, T. E. Profile of mood states. In Buros, O. K. (ed.), **The Eighth Mental Measurements Yearbook** (Vol. 1). New Jersey: Gryphon Press, 1978.

Wilkinson, R. T. The effects of sleep loss on performance. **Medical Research Council, Applied Psychology Report No. 323/58,** 1958.

Wilkinson, R. T., Edwards, R. S., & Haines, E. Performance following a night of reduced sleep. **Psychonomic Science**, 1966, **5**(12), 471-472.

Williams, H. L., Gieseking, C. F., & Lubin, A. Some effects of sleep loss on memory. **Perceptual Motor Skills**, 1966, **23**, 1287-1293.

Williams, H. L., Lubin, A., & Goodnow, J. J. Impaired performance with acute sleep loss. **Psychological Monographs**, 1959, **73** (14, Whole No. 484).

APPENDIX A

THE COMPUTER AIDED OPERATIONS RESEARCH FACILITY (CAORF)

A.1 DESCRIPTION OF CAORF

CAORF is the sophisticated ship-maneuvering simulator operated by the U.S. Maritime Administration for controlled research into man-ship-environment problems. Controlled experiments, which might require several vessels, cannot be performed readily in the real world and would certainly be ruled out for testing situations that involve potential danger. Such experiments can be performed safely and easily at CAORF. A simplified cutaway of the simulator building is shown in Figure A-1 and the relationships among the major subsystems are illustrated in Figure A-2.

All actions called for by the watch officer on the bridge are fed through a central computer that alters the visual scene and all bridge displays and repeaters in accordance with the calculated dynamic response of ownship and the environmental situation being simulated. CAORF has the capability of simulating any ship, port, or area in the world. The major subsystems are:

- **Wheelhouse**, which contains all equipment and controls needed by the test subject watch officer to maneuver ownship through a scenario, including propulsion and steering controls, navigational equipment and communications gear.
- **Central Data Processor**, which computes the motion of ownship in accordance with its known characteristics, models the behavior of all other traffic ships, and drive the appropriate bridge indicators.
- **Image Generator**, which constructs the computer-generated visual image of the surrounding environment and traffic ships that is projected onto a cylindrical screen for visual realism.
- **Radar Signal Generator**, which synthesizes video signals to stimulate the bridge radars and collision avoidance system for the display of traffic ships and surrounding environment.
- **Control Station**, from which the experiment is started and stopped, traffic ships and environment can be controlled, mechanical failures can be introduced, and external communications with ownship's bridge can be simulated.
- **Human Factors Monitoring Station**, from which unobtrusive observation and video recording of test subject behavior can be carried out by experimental psychologists.

A.2 SIMULATED BRIDGE

The simulated bridge consists of a wheelhouse 20 feet (6.1 m) wide and 14 feet (4.3 m) deep. The equipment on the CAORF bridge is similar to that normally available in the merchant fleet and responds with realistically duplicated time delays and accuracy. The arrangement is based on contemporary bridge design and includes the following equipment:

- **Steering Controls and Displays** – a gyropilot helm unit with standard steering modes, rate of turn indicator, rudder angle/rudder order indications, and gyro repeaters.
- **Propulsion Controls and Displays** – an engine control panel (capable of simulating control from either bridge or engine room) containing a combined engine order telegraph/throttle, an rpm indicator and a switch for selecting the operating mode, such as finished with engine, warm up, maneuvering and sea speed.
- **Thruster Controls and Displays** – bow and stern thrusters and their respective indicators and status lights.
- **Navigation Systems** – two radars capable of both relative and true motion presentations, plus a collision avoidance system. Capability exists for future additions such as a digital fathometer, Radio Direction Finder, and Loran C and Omega Systems.

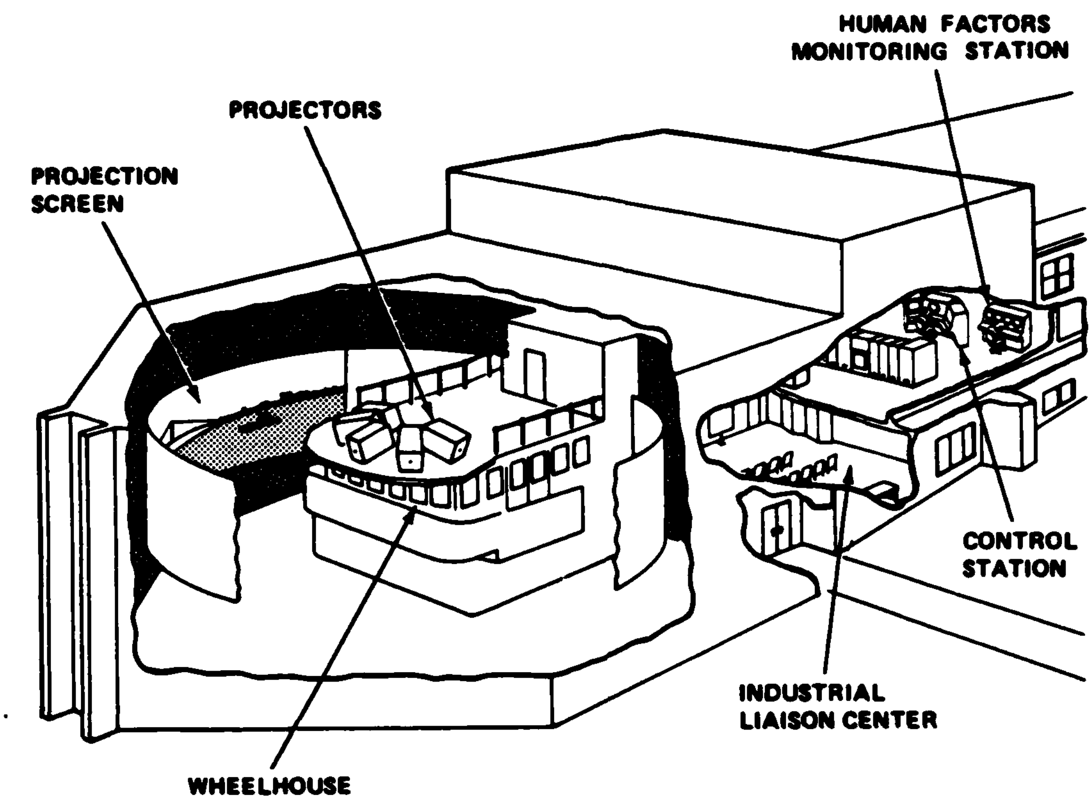

Figure A-1. Cutaway of CAORF Building

- **Communications** – simulated VHF/SSB radio, docking loudspeaker (talkback) system, sound powered phones and ship's whistle.

- **Wind Indicators** – indicate to the bridge crew the true speed and direction of simulated wind.

A.3 OWNSHIP SIMULATION

Any ship can be simulated at CAORF. The computerized equations of motion are adapted to the ship by changing specific coefficients, among which are hydrodynamic, inertial, propulsion, thruster, rudder, aerodynamic, etc. Wind and currents realistically affect ship motion according to draft (loaded or ballasted) and relative speed and direction. Ownship's computer model was validated by comparing various simulated maneuvers (e.g., zig-zag, turning circle, spiral, crash stop, and acceleration tests) with sea trial data.

A.4 IMAGE GENERATION

The visual scene is generated at CAORF to a degree of realism sufficient for valid simulation. The scene (Figure A-3) includes all the man-made structures and natural components of the surrounding scene that mariners familiar with the geographical area deem necessary as cues for navigation.

Thus, bridges, buoys, lighthouses, tall buildings, mountains, glaciers, piers, coastlines, and islands would be depicted in the scene. In addition, the closest traffic ships and the forebody of ownship appear. All elements in the scene appear to move in response to ownship's maneuvers. The sky is depicted without clouds and the water without waves.

For enhanced realism the scene is projected in full color. The perspective is set for the actual bridge height above

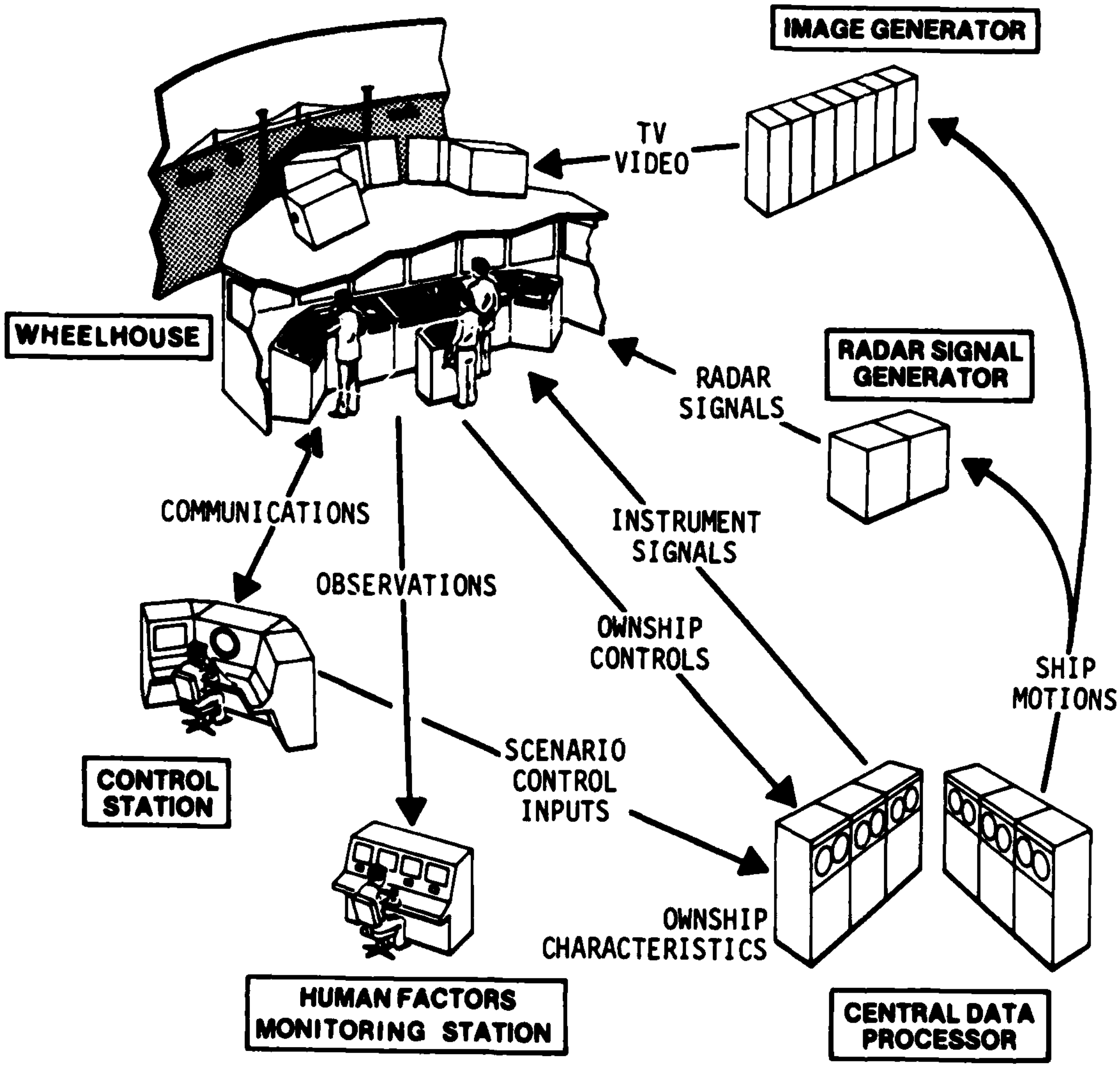

Figure A-2. Major CAORF Subsystems

waterline for the simulated ship. Shadowing can be varied according to the position of the sun at different times of day.

Environmental conditions also affect the scene. The lighting can be varied continuously from full sun to moonless night. At night, lights can be seen on traffic vessels, buoys, piers, and other points ashore. Visibility in the day or night can be reduced to simulate any degree of fog or haze.

A.5 RADAR SIGNAL GENERATION

The Radar Signal Generator produces real-time video signals for driving two radar PPIs. The items displayed are synchronized with the visual scene and include navigation aids, ships, shorelines and other topographical features with appropriate target shadowing, clutter, range attenuation, and receiver noise. The radar gaming area, which covers an area of 150 by 200 miles, extends beyond the visual gaming

Figure A-3. Typical Simulated Visual Scene at CAORF

area, which is 50 by100 miles. Within the radar gaming area, as many as 40 moving traffic ships can be displayed. The radar signal generator also drives the collision avoidance system, which can be slaved to either of the master PPIs.

A.6 CONTROL STATION

The Control Station (Figure A 4) is the central location from which the simulator experiment is controlled and monitored. An experiment can be initiated anywhere within the visual gaming area with any ship traffic configuration. The Control Station enables the researchers to interface with the watchstanding crew on the bridge, to simulate malfunctions, and to control the operating mode of the simulator. The Control Station is also capable of controlling motions of traffic ships and tugs in the gaming area and simulating telephone, intercom, radio (VHF, SSB) and whistle contact with the CAORF bridge crew.

A.7 HUMAN FACTORS MONITORING STATION

The Human Factors Monitoring Station (Figure A-5) is designed to allow collection of data on crew behavior. Monitoring data is provided by five closed-circuit TV cameras and four microphones strategically located throughout the wheelhouse to record all activities, comments and commands.

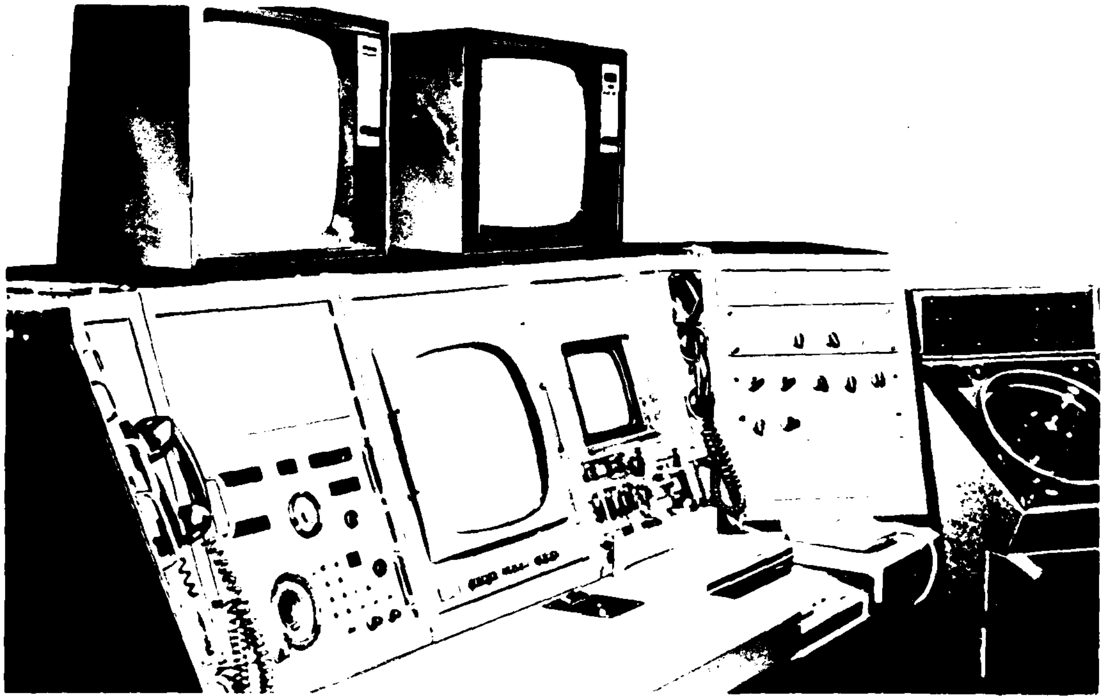

Figure A-4. Control Station

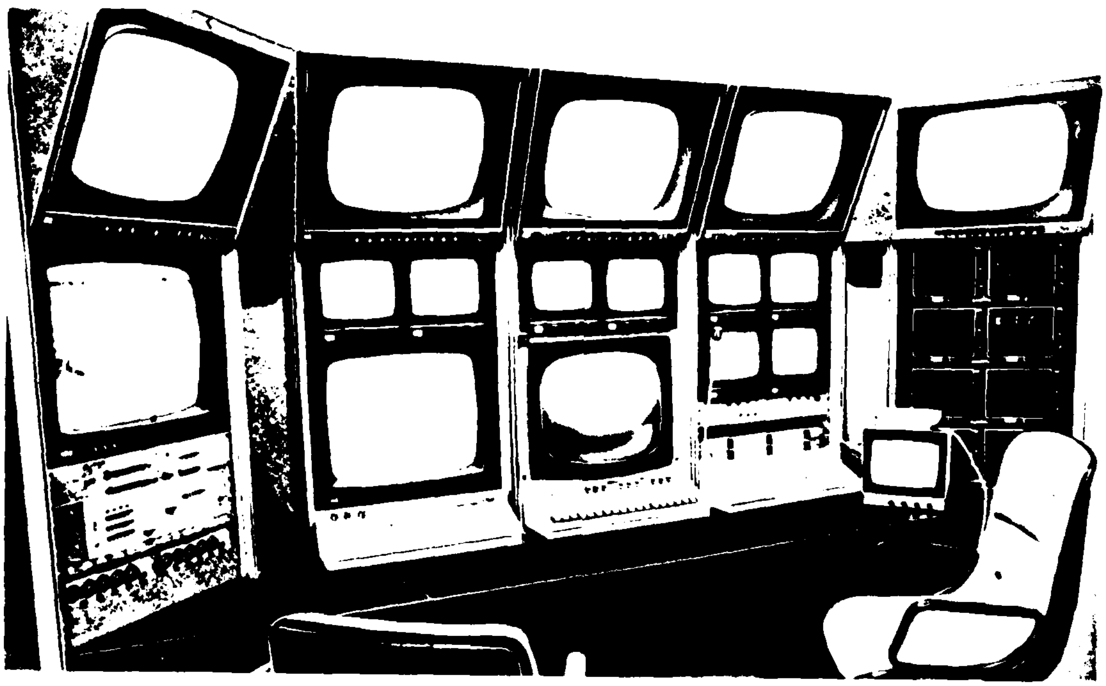

Figure A-5. Human Factors Monitoring Station

APPENDIX B

DOCUMENTATION GIVEN TO SUBJECTS PRIOR TO EXPERIMENT

Dear Participant:

You will be taking part in a government sponsored study designed to investigate some factors which may interfere with the ability of the watchstanding ships's officer to do his job. This research will examine the influences of visibility, fatigue, traffic conditions, basic personality differences, scheduling, etc., on watchstanding behavior. The influences will be measured in the participants by means of a variety of paper and pencil tests and physiological measurements given at various times during a simulated watch, as well as observing watchstanding behavior. This testing process will involve the following procedures:

- a battery of paper and pencil tests designed to assess personality traits (taking about an hour) will be given once to each participant.
- a series of paper and pencil tests designed to measure cognitive functioning (taking about ten minutes) will be given at the beginning, middle, and end of each of the two simulated watch periods.
- a continuous measurement of heart rate, eye movement, muscle tension and body movement will be made by means of seven wires taped to the forehead and chest and feeding in turn into a small tape recorder worn on the belt.

One of the most important factors this study will be investigating will be the effects of sleep loss and scheduling on watchstanding performance. In order to examine this factor scientifically it is essential that we control the sleep schedules and activities of the participants during the study. We will, therefore, make the following requests of all participants:

- All participants are requested to stay overnight at the Merchant Marine Academy, please be sure to bring appropriate toilet articles.
- Participants will be asked to remain on academy grounds during the experiment, and to retire at the time specified by the research staff.
- Some participants (to be randomly selected) will be asked to remain awake on one night. Activities will be provided to help accomplish this; you may also wish to bring a book, board game, etc., to pass the time.
- Participants will be asked not to indulge in strenuous physical activity or drink any alcoholic beverages during the study.

In order to apply the results of this study to the real world it is necessary that we learn something about the normal sleep habits of the participants. Also, we would like to investigate the biological rhythms of the participants to see if these interact with sleep conditions. These rhythms will be measured by means of periodic oral temperature readings. We will ask you to provide us with this information by filling out the enclosed series of forms for one week prior to the beginning of the experiment.

- Sleep logs to be filled out one each morning upon awakening
- Disposable oral thermometers for taking a temperature reading each hour during the day
- Temperature log on which to record temperature readings
- Detailed instructions for using the above forms.

All participants will be reimbursed for traveling expenses and meals, and will receive a fee for services. The payment schedule is as follows:

- Reimbursement for transportation expenses upon presentation of receipts, or 18½¢ per mile for driving plus tolls.

- $23.00 per day for food

- $130.00 per day for services rendered.

Thank you very much for your participation. If any questions or problems should arise, please do not hestitate to call either Anita D'Amico or Ned Kaufman at (516) 482-8200, ext. 560 or 608. Otherwise, we will look forward to seeing you on ____________________ at _________ hours.

Anita D'Amico
Program Manager

ADA:imr
Enclosures

Thank you for volunteering for this CAORF research project. The object of this study is to investigate the effects of fatigue and time of watch on watchstanding.

We are measuring the effects of fatigue by asking some of the participants to stay awake for the duration of the experiment. Those participants who are required to stay awake will be asked to stand a three hour watch today and a four hour watch tomorrow without obtaining any sleep between now and tomorrow. Participants in the "no sleep" group should expect to feel fatigued on the second day and will be given facilities in which to rest after the experiment is over. Selection of participants for this "no sleep" condition is determined by experimental schedule and has nothing to do with personal characteristics of the volunteers. Those in the group which are permitted to sleep will be asked to stand a three hour watch today and a four hour watch tomorrow and will be required to obtain 7½ hours of sleep sometime between 2300 hours this evening and 0900 hours tomorrow.

The effects of time of watch will be investigated by comparing two different watch periods – the 0800 to 1200 watch and the 1200 to 1600 watch. Participants are randomly assigned to these watch periods.

You will be assigned to one of the following four conditions:

I. no sleep, 0800 to 1200 watch

II. no sleep, 1200 to 1600 watch

III. 7½ hours sleep, 0800 to 1200 watch

IV. 7½ hours sleep, 1200 to 1600 watch.

We are also interested in examining how bodily functions change during a watch period. In order to monitor your physiological responses during the watch period, we will be attaching seven wires to your forehead, chest and back. These wires will monitor your eye movement, muscle tension, heart rate and physical activity. Attachment of these monitors poses no danger to you and you will be able to move freely. You may find it slightly difficult to scratch your forehead. Because we are monitoring physiological responses, it is very important that we know whether you are using any medication or have had a recent illness. It is also important that you not consume any alcoholic beverages during your 48 hour stay.

Before, during, and after the watch period, we will ask you to fill out some forms. These paper and pencil tests measure things such as mood and ability to concentrate. At one point we will also ask for some biographical information (such as age, education and sea experience) and for information about your personality. This data will be combined with information obtained from all other participants in CAORF research to draw a profile of the mariners who volunteer for simulator research.

An experimenter will be available to answer any questions you may have about this research. You are free to withdraw from participating in this research project at any time. Thank you for your time and cooperation.

I have read the above statement and consent to participation. I have no history of heart disease or uncontrolled blood pressure. I have informed the experimenter about any medication I am currently taking or have taken in the past 48 hours. I understand that I have been requested to refrain from using any alcohol or drugs during this experiment. I also understand that I am free to withdraw from participating in this research project at any time.

____________________ Name

____________________ Date

APPENDIX C

POST EXPERIMENTAL DEBRIEFING QUESTIONNAIRE

1. Did the experimental task resemble something you would encounter in real life situations?

   yes____

   no ____

   What could be better?

2. In real life would you ever be required to stand watch without having slept the night before?

   yes ____

   no ____

3. During this simulated run did you behave any differently than you normally would at sea?

   yes ____

   no ____

   If yes, please explain.

4. On a scale from 1 to 5 how would you rate the realism of the CAORF experience? (circle one)

Poor				Excellent
1	2	3	4	5
Very Unrealistic				Very Realistic

5. Was the experimental task challenging?

   yes ____

   no ____

   Please explain why or why not.

6. Would you rate your task as: (check one)

   Easy ____

   Neutral ____

   Difficult ____

7. Did you perceive your performance changing one time?

   yes ____

   no ____

   If yes, please describe how it changes.

8. Were you bored during this experiment?

   yes ____

   no ____ (skip to question 12)

9. How long did it take for you to become bored?

   ____ hours ____ minutes

10. When were you most bored during this experiment?

11. Did you do anything to alleviate boredom during this experiment?

    yes ____

    no ____

    If yes, please describe what you did.

12. Were you fatigued during this experiment?

    yes ____

    no ____ (skip to question 16)

13. How long did it take for you to become fatigued?

____ hours ____ minutes

14. When were you most fatigued during the experiment?

15. Did you do anything throughout the experiment to alleviate this fatigue?

16. How would you rate the difficulty of staying awake for the simulated four hour watch?

Easy ____

Neutral ____

Difficult ____

What conditions made it either easy or difficult?

17. Do you feel any of the following factors influence your performance in this experiment?

fatigue ____

boredom ____

ill health ____

hunger ____

poor spirits ____

physical discomfort ____

lack of interest ____

none of the above ____

18. How much sleep did you get last night?

19. Did you find yourself beginning to drop off to sleep at any time during the experiment?

yes ____

no ____ (skip to question 21)

If yes, how often did you find this happening? When during the experiment did it happen the most?

20. Did you sleep for any period of time during the experiment?

yes ____

no ____

If yes, how often during the experiment did you fall asleep? How long did each period of sleep last? Approximately when during the experiment did you sleep most?

21. In real life have you ever found yourself either dropping off or sleeping during a watch?

yes ____

no ____

22. Did you ever daydream during the experiment?

yes ____

no ____

If yes, what percentage of the time would you estimate that you daydreamed? On the average how long did each period of daydreaming last? When during the experiment did you daydream the most?

23. In real life what would you estimate as the percentage of time on watch that you daydream?

24. Sometimes people experience a state where they are wide awake with their eyes open and looking straight ahead, but where their mind is totally blank – a state called mind blanking. Did this ever happen to you at any time during the experiment?

yes ____

no ____

If yes, how often did it happen? What was the duration of each occurrence? Can you describe a specific point in time during the watch when you might have been mind blanking.

25. Do you ever experience such a state during your normal watches?

yes ____

no ____

If yes, how often do they occur? In what situations, conditions, and on which watches are they most likely to occur?

APPENDIX D

SUMMARY TABLES FOR STATISTICAL ANALYSES REPORTED IN CHAPTER 3

TABLE D1. SUMMARY TABLE FOR T-TEST COMPARING DETECTION OF FIRST AND LAST SHIPS

Dependent Variable: Detection Time (in minutes)
Group: All Subjects

Independent Variable	No. of Cases	Mean	Standard Deviation	T-Value	df	2 Tail Probability
First Ship	23	3.95	4.75			
Last Ship	23	2.52	4.56			
Difference		1.43	5.22	1.42	22	0.20

TABLE D2. SUMMARY TABLE FOR ANALYSIS OF VARIANCE COMPARING DETECTION OF FIRST AND LAST SHIPS

Dependent Variable: Difference Between Time to Detect First Incident and Time to Detect Fourth Incident
Independent Variables: Sleep Condition (1 = Full Sleep, 2 = No Sleep)
Watch Condition (1 = Morning, 2 = Afternoon)

Source	Sum of Squares	df	Mean Squares	F Ratio	p
Sleep Condition	2.987	1	2.987	0.097	0.759
Watch Condition	11.048	1	11.048	0.358	0.557
S x W	0.311	1	0.311	0.010	0.921
Error	585.799	19	30.832		
TOTAL	599.652	22	27.257		

	Full Sleep Morning Watch	Full Sleep Afternoon Watch	No Sleep Morning Watch	No Sleep Afternoon Watch
$\bar{X}$ of Difference	0.20	1.83	1.17	2.33
SD of Difference	0.45	7.68	4.58	6.09
N	5	6	6	6

TABLE D3. SUMMARY TABLE FOR ANALYSIS OF VARIANCE FOR DETECTION OF FIRST SHIP

Dependent Variable: Detection Time for First Ship
Independent Variables: Sleep Condition (1 = Full Sleep, 2 = No Sleep)
Watch Condition (1 = Morning, 2 = Afternoon)

Source	Sum of Squares	df	Mean Squares	F Ratio	p
Sleep Condition	62.72	1	62.72	2.08	0.16
Watch Condition	39.13	1	39.13	1.30	0.26
S x W	1.41	1	1.41	0.50	0.53
Error	632.76	21	30.13		
TOTAL	739.99	24	133.39		

	Full Sleep Morning Watch	Full Sleep Afternoon Watch	No Sleep Morning Watch	No Sleep Afternoon Watch
$\bar{X}$	1.00	4.00	4.67	5.57
SD	1.67	5.40	4.50	5.26
N	6	6	6	7

TABLE D4. SUMMARY TABLE FOR ANALYSIS OF COVARIANCE FOR DETECTION OF FIRST SHIP

Dependent Variables: Detection Time for First Ship Sighting
Independent Variables: Sleep Condition (1 = Full Sleep; 2 = No Sleep)
Watch Condition (1 = Morning; 2 = Afternoon)
Covariate: Ship Type (1 = Containership; 2 = VLCC)

Source	Sum of Squares	df	Mean Squares	F Ratio	p
Sleep Condition	52.443	1	52.443	2.737	0.114
Watch Condition	20.003	1	20.003	1.044	0.319
Sleep x Watch	3.242	1	3.242	0.169	0.685
Covariate	38.800	1	38.800	2.025	0.170
Error	383.191	20	19.160		
TOTAL	500.640	24	20.860		

	Full Sleep Morning	Full Sleep Afternoon	No Sleep Morning	No Sleep Afternoon
Mean	1.00	4.00	4.67	5.57
SD	1.67	5.40	4.50	5.26
N	6	6	6	7

TABLE D5. SUMMARY TABLE FOR ANALYSIS OF VARIANCE FOR DETECTION OF DESTROYER

Dependent Variable: Detection Time for Destroyer
Independent Variables: Sleep Condition (1 = Full Sleep, 2 = No Sleep)
Watch Condition (1 = Morning, 2 = Afternoon)

Source	Sum of Squares	df	Mean Squares	F Ratio	p
Sleep Condition	15.56	1	15.56	0.66	0.42
Watch Condition	59.12	1	59.12	2.51	0.13
S x W	9.00	1	9.00	0.38	0.54
Error	493.88	21	23.52		
TOTAL	580.00	24	24.17		

	Full Sleep Morning Watch	Full Sleep Afternoon Watch	No Sleep Morning Watch	No Sleep Afternoon Watch
$\bar{X}$	1.00	2.83	1.33	5.57
SD	1.55	3.19	2.16	8.24
N	6	6	6	7

TABLE D6. SUMMARY TABLE FOR ANALYSIS OF COVARIANCE FOR DETECTION OF DESTROYER

Dependent Variables: Detection Time for Destroyer
Independent Variables: Sleep Condition (1 = Full Sleep; 2 = No Sleep)
Watch Condition (1 = Morning; 2 = Afternoon)
Covariate: Order of Presentation (2 = During Second Hour; 3 = During Third Hour)

Source of Variance	Sum of Squares	df	Mean Square	F Ratio	p
Sleep Condition	16.137	1	16.137	0.655	0.43
Watch Condition	58.450	1	58.450	2.374	0.14
Sleep x Watch	9.744	1	9.744	0.396	0.54
Covariate	0.410	1	0.410	0.017	0.90
Error	492.439	20	24.622		
TOTAL	580.000	24	24.167		

	Full Sleep Morning	Full Sleep Afternoon	No Sleep Morning	No Sleep Afternoon
Mean	1.00	2.83	1.33	5.57
SD	1.55	3.19	2.16	8.24
N	6	6	6	7

TABLE D7. SUMMARY TABLE FOR ANALYSIS OF VARIANCE FOR DETECTION OF LAST SHIP

	Dependent Variable: Detection Time for Last Ship Independent Variables: Sleep Condition (1 = Full Sleep, 2 = No Sleep) Watch Condition (1 = Morning, 2 = Afternoon)				
Source	**Sum of Squares**	**df**	**Mean Squares**	**F Ratio**	**p**
Sleep Condition	21.13	1	21.13	0.96	0.34
Watch Condition	1.97	1	1.97	0.09	0.77
S x W	9.91	1	9.91	0.45	0.51
Error	417.97	19	21.99		
TOTAL	450.43	22	20.47		

	Full Sleep Morning Watch	Full Sleep Afternoon Watch	No Sleep Morning Watch	No Sleep Afternoon Watch
$\bar{X}$	0.80	2.17	3.50	2.83
SD	0.45	4.40	3.99	6.94
N	5	6	6	6

TABLE D8. SUMMARY TABLE FOR SPLIT PLOT ANALYSIS OF COVARIANCE FOR NPRU POSITIVE AFFECT SCALE

Dependent Variable: NPRU Positive Affect Scale

Independent Variables:
- Between Subjects: Sleep Condition (1 = Full Sleep; 2 = No Sleep); Watch Condition (1 = Morning; 2 = Afternoon)
- Within Subjects: Test Time (1 = Pre; 2 = Mid; 3 = Post-Watch on Experimental Day)

Covariate: NPRU Positive Affect Scale – Tested Post-Watch on Familiarization Day

Source of Variance	Sums of Squares	df	Mean Square	F Ratio	p
Between					
Sleep Condition	2.18631	1	2.18631	9.61	0.0059
Watch Condition	0.11620	1	0.11620	0.51	0.4836
S x W	0.26396	1	0.26396	1.16	0.2950
Covariate	5.92999	1	5.92999	26.05	0.0001
Error	4.32458	19	0.22761		
Within					
Test Time	0.17041	2	0.08521	1.29	0.2878
T x S	0.00263	2	0.00132	0.02	0.9803
T x W	0.01254	2	0.00627	0.09	0.9099
T x S x W	0.01138	2	0.00569	0.09	0.9179
Error	2.65183	40	0.06630	0.09	0.9179
TOTAL					

	Full Sleep A.M.	Full Sleep P.M.	No Sleep A.M.	No Sleep P.M.
Mean				
Pre-	1.7433	1.4300	1.2167	1.3200
Mid-	1.8233	1.5150	1.2650	1.4800
Post-	1.8400	1.5267	1.2883	1.4983
Standard Deviation				
Pre-	0.19552	0.3613	0.5767	0.5393
Mid-	0.08664	0.2409	0.6957	0.5606
Post-	0.25846	0.5173	0.6279	0.4434
Mean Adjusted for Covariate				
Pre-	1.7136	1.5078	1.2590	1.2296
Mid-	1.7936	1.5928	1.3073	1.3896
Post-	1.8102	1.6045	1.3307	1.4079
n	6	6	6	6

TABLE D9. SUMMARY TABLE FOR SPLIT PLOT ANALYSIS OF COVARIANCE FOR NPRU NEGATIVE AFFECT SCALE

Dependent Variable: NPRU Negative Affect Scale

Independent Variables:
- **Between Subjects** — **Sleep Condition** (1 = Full Sleep; 2 = No Sleep); **Watch Condition** (1 = Morning; 2 = Afternoon)
- **Within Subjects** — **Test Time** (1 = Pre; 2 = Mid; 3 = Post-Watch on Experimental Day)

Covariate: NPRU Negative Affect Scale — Tested Post-Watch on Familiarization Day

Source of Variance	Sums of Squares	df	Mean Square	F Ratio	p
Between					
Sleep Condition	6.37705	1	6.37705	65.63	0.0001
Watch Condition	0.64850	1	0.64850	6.67	0.0182
S x W	0.08732	1	0.08732	0.90	0.3550
Covariate	0.06108	1	0.06108	0.63	0.4376
Error	1.84619	19	0.09717		
Within					
Test Time	0.69534	2	0.34767	4.17	0.0226
T x S	0.18872	2	0.09436	1.13	0.3323
T x W	0.10789	2	0.05394	0.65	0.5287
T x S x W	0.05860	2	0.02930	0.35	0.7056
Error	3.33205	40	0.08330	0.35	
TOTAL					

	Full Sleep A.M.	Full Sleep P.M.	No Sleep A.M.	No Sleep P.M.
Mean				
Pre-	0.2117	0.3500	0.8000	1.2333
Mid-	0.1667	0.3000	0.6833	0.8850
Post-	0.1217	0.2000	0.5983	0.7167
Standard Deviation				
Pre-	0.2173	0.3728	0.2530	0.4412
Mid-	0.2066	0.2000	0.2787	0.4225
Post-	0.1281	0.1789	0.3924	0.2483
Mean Adjusted for Covariate				
Pre-	0.2088	0.3521	0.7954	1.2387
Mid-	0.1638	0.3021	0.6788	0.8904
Post-	0.1188	0.2021	0.5938	0.7221
n	6	6	6	6

TABLE D10. SUMMARY TABLE FOR SPLIT PLOT ANALYSIS OF COVARIANCE FOR STANFORD SLEEPINESS SCALE

Dependent Variable: Stanford Sleepiness Scale

Independent Variables: Between Subjects — Sleep Condition (1 = *Full Sleep*; 2 = *No Sleep*); Watch Condition (1 = Morning; 2 = Afternoon). Within Subjects — Test Time (1 = Pre; 2 = Mid; 3 = Post-Watch on Experimental Day)

Covariate: Stanford Sleepiness Scale — Tested Post-Watch on Familiarization Day

Source of Variance	Sums of Squares	df	Mean Square	F Ratio	p
Between					
Sleep Condition	37.55556	1	37.55556	17.09	0.0006
Watch Condition	1.78256	1	1.78256	0.81	0.3790
S x W	3.25625	1	3.25625	1.48	0.2384
Covariate	13.02584	1	13.02584	5.93	0.0249
Error	41.75194	19	2.19747		
Within					
Test Time	0.86111	2	0.43056	0.37	0.6949
T x S	1.02778	2	0.51389	0.44	0.6481
T x W	0.52778	2	0.26389	0.23	0.7994
T x S x W	0.02778	2	0.01389	0.01	0.9882
Error	46.88889	40	1.17222		
TOTAL					

	Full Sleep A.M.	Full Sleep P.M.	No Sleep A.M.	No Sleep P.M.
Mean				
Pre-	1.8333	2.6667	3.6667	3.1667
Mid-	1.8333	2.5000	4.3333	3.5000
Post-	1.8333	2.3333	4.0000	3.0000
Standard Deviation				
Pre-	0.4082	1.0328	1.5055	0.9832
Mid-	0.7528	0.5477	2.2509	1.6432
Post-	0.9832	1.2111	2.0976	0.6325
Mean Adjusted for Covariate				
Pre-	2.1085	2.3915	3.6667	3.1667
Mid-	2.1085	2.2248	4.3333	3.5000
Post-	2.1085	2.0581	4.0000	3.0000
n	6	6	6	6

TABLE D11. SUMMARY TABLE FOR SPLIT PLOT ANALYSIS OF COVARIANCE FOR POMS FATIGUE SCALE

Dependent Variable: Poms Fatigue Scale				
Independent Variables:	Between Subjects	Sleep Condition	(1 = Full Sleep; 2 = No Sleep)	
		Watch Condition	(1 = Morning; 2 = Afternoon)	
	Within Subjects	Test Time	(1 = Pre; 2 = Mid; 3 = Post-Watch on Experimental Day)	
Covariate: Poms Fatigue Scale — Tested Post-Watch on Familiarization Day				

Source of Variance	Sums of Squares	df	Mean Square	F Ratio	p
Between					
Sleep Condition	1250.42282	1	1250.42282	42.81	0.0001
Watch Condition	25.90656	1	25.90656	0.89	0.3575
S x W	21.43102	1	21.43102	0.73	0.4018
Covariate	227.11621	1	227.11621	7.78	0.0113
Error	584.16950	20	29.20847		
Within					
Test Time	13.68783	2	6.84392	0.47	0.6306
T x S	14.66725	2	7.33363	0.50	0.6104
T x W	20.29688	2	10.14844	0.69	0.5065
T x S x W	4.27219	2	2.13610	0.15	0.8650
Error	616.57143	42	14.68027		
TOTAL					

	Full Sleep A.M.	Full Sleep P.M.	No Sleep A.M.	No Sleep P.M.
Mean				
Pre-	1.1667	3.3333	10.1667	10.0000
Mid-	0.8333	1.8333	11.5000	7.8571
Post-	1.1667	2.0000	13.0000	10.0000
Standard Deviation				
Pre-	1.9408	2.8048	8.4004	5.6273
Mid-	1.3292	1.7224	7.2319	3.5790
Post-	1.3292	2.3664	7.4297	5.0332
Mean Adjusted for Covariate				
Pre-	2.1216	2.8443	9.9664	9.7723
Mid-	1.7883	1.3443	11.2998	7.6294
Post-	2.1216	1.5110	12.7998	9.7723
n	6	6	6	7

TABLE D12. SUMMARY TABLE FOR SPLIT PLOT ANALYSIS OF COVARIANCE FOR POMS VIGOR SCALE

Dependent Variable: Poms Vigor Scale

Independent Variables: Between Subjects — Sleep Condition (1 = Full Sleep; 2 = No Sleep); Watch Condition (1 = Morning; 2 = Afternoon)

Within Subjects — Test Time (1 = Pre; 2 = Mid; 3 = Post-Watch on Experimental Day)

Covariate: Poms Vigor Scale — Tested Post-Watch on Familiarization Day

Source of Variance	Sums of Squares	df	Mean Square	F Ratio	p
Between					
Sleep Condition	531.34340	1	531.34340	13.57	0.0015
Watch Condition	2.16137	1	2.16137	0.06	0.8166
S x W	100.67240	1	100.67240	2.57	0.1245
Covariate	366.94144	1	366.94144	9.37	0.0062
Error	782.96332	20	39.14817		
Within					
Test Time	1.17166	2	0.58583	0.03	0.9678
T x S	38.08524	2	19.04262	1.07	0.3538
T x W	23.36919	2	11.68460	0.65	0.5254
T x S x W	14.63668	2	7.31834	0.41	0.6667
Error	750.85714	42	17.87755		
TOTAL					

	Full Sleep A.M.	Full Sleep P.M.	No Sleep A.M.	No Sleep P.M.
Mean				
Pre-	17.6667	12.6667	9.8333	13.4286
Mid-	19.0000	14.5000	10.0000	10.8571
Post-	19.0000	15.1667	7.5000	13.1429
Standard Deviation				
Pre-	3.9833	4.7609	6.4627	5.4729
Mid-	4.0497	4.1833	7.2938	5.7280
Post-	3.7947	7.9603	5.6833	4.7759
Mean Adjusted for Covariate				
Pre-	16.7859	13.4178	10.5845	12.9190
Mid-	18.0923	15.2511	10.7511	10.3475
Post-	18.0923	15.9578	8.2511	12.6332
n	6	6	6	7

TABLE D13. SUMMARY TABLE FOR SPLIT PLOT ANALYSIS OF COVARIANCE FOR POMS CONFUSION SCALE

Dependent Variable: Poms Confusion Scale

Independent Variables:	Between Subjects	Sleep Condition	(1 = Full Sleep; 2 = No Sleep)
		Watch Condition	(1 = Morning; 2 = Afternoon)
	Within Subjects	Test Time	(1 = Pre; 2 = Mid; 3 = Post-Watch on Experimental Day)

Covariate: Poms Confusion Scale — Tested Post-Watch on Familiarization Day

Source of Variance	Sums of Squares	df	Mean Square	F Ratio	p
Between					
Sleep Condition	7.43720	1	7.43720	0.86	0.3611
Watch Condition	0.16889	1	0.16889	0.02	0.8901
S x W	17.46171	1	17.46171	2.03	0.1701
Covariate	13.36289	1	13.36289	1.55	0.2275
Error	172.42203	20	8.62114		
Within					
Test Time	7.60905	2	3.80453	1.22	0.3064
T x S	0.74486	2	0.37243	0.12	0.9998
T x W	14.74486	2	7.37243	2.36	0.1070
T x S x W	9.79012	2	4.98506	1.57	0.2209
Error	131.3333	42	3.12690		
TOTAL					

	Full Sleep A.M.	Full Sleep P.M.	No Sleep A.M.	No Sleep P.M.
Mean				
Pre-	3.3333	5.0000	3.8333	5.0000
Mid-	2.5000	4.0000	4.8333	2.7143
Post-	3.0000	4.1667	5.5000	3.1429
Standard Deviation				
Pre-	1.3663	1.7888	2.5625	4.1633
Mid-	1.7607	1.6733	3.1885	1.3801
Post-	1.4142	0.9832	2.6646	1.4638
Mean Adjusted for Covariate				
Pre-	3.4276	4.7576	3.8013	5.1544
Mid-	2.5943	3.7576	4.8013	2.8686
Post-	3.0943	3.9243	5.4680	3.2972
n	6	6	6	7

TABLE D14. SUMMARY TABLE FOR SPLIT PLOT ANALYSIS OF COVARIANCE FOR POMS ANGER SCALE

Dependent Variable: Poms Anger Scale

Independent Variables: **Between Subjects** — **Sleep Condition** (1 = Full Sleep; 2 = No Sleep); **Watch Condition** (1 = Morning; 2 = Afternoon)

Within Subjects — **Test Time** (1 = Pre; 2 = Mid; 3 = Post-Watch on Experimental Day)

Covariate: Poms Anger Scale — Tested Post-Watch on Familiarization Day

Source of Variance	Sums of Squares	df	Mean Square	F Ratio	p
Between					
Sleep Condition	4.41283	1	4.41203	0.49	0.4924
Watch Condition	26.56467	1	26.56467	2.94	0.1017
S x W	0.00219	1	0.00219	0.00	0.9877
Covariate	94.19676	1	94.19676	10.44	0.0042
Error	180.47784	20	9.02309		
Within					
Test Time	16.98942	2	8.49471	2.70	0.0737
T x S	0.16637	2	0.08319	0.03	0.9732
T x W	10.73427	2	5.36714	1.75	0.1055
T x S x W	0.29806	2	0.14903	0.05	0.9525
Error	128.49206	42	3.05933		
TOTAL					

	Full Sleep A.M.	Full Sleep P.M.	No Sleep A.M.	No Sleep P.M.
Mean				
Pre-	1.0000	2.1667	0.1667	2.1421
Mid-	0.8333	3.5000	0.5000	3.4286
Post-	0.6667	1.5000	0.1667	1.2857
Standard Deviation				
Pre-	1.2649	1.9408	0.4082	3.1848
Mid-	1.1690	5.2820	0.8366	4.6853
Post-	1.2111	0.8366	0.4082	1.8898
Mean Adjusted for Covariate				
Pre-	1.2560	2.1066	0.5017	1.6076
Mid-	1.0894	3.4399	0.8351	2.9734
Post-	0.9227	1.4399	0.5017	0.8305
n	6	6	6	7

TABLE D15. SUMMARY TABLE FOR SPLIT PLOT ANALYSIS OF COVARIANCE FOR VISUAL SEARCH AND SCAN, 2 LETTERS, LINES SCANNED

Dependent Variable: Visual Search & Scan 2 Letter Target, Lines Scanned

Independent Variables: Between Subjects — Sleep Condition (1 = Full Sleep; 2 = No Sleep); Watch Condition (1 = Morning; 2 = Afternoon)
Within Subjects — Test Time (1 = Pre; 2 = Mid; 3 = Post-Watch on Experimental Day)

Covariate: Visual Search & Scan 2 Letter target, lines scanned — Tested Post-Watch on Familiarization Day

Source of Variance	Sums of Squares	df	Mean Square	F Ratio	p
Between					
Sleep Condition	182.43271	1	182.43271	2.59	0.1253
Watch Condition	48.06185	1	48.06185	0.68	0.4200
S x W	39.12210	1	39.12210	0.55	0.4662
Covariate	3959.53503	1	3959.53503	56.11	0.0001
Error	1270.28163	18	70.57120		
Within					
Test Time	59.15000	1	59.15000	0.85	0.3675
T x S	1.75317	1	1.75317	0.03	0.8754
T x W	5.47646	1	5.46746	0.08	0.7820
T x S x W	46.86429	19	46.86429	0.68	0.4214
Error	1318.68333		69.40439		
TOTAL					

	Full Sleep A.M.	Full Sleep P.M.	No Sleep A.M.	No Sleep P.M.
Mean				
Pre-	43.4000	60.8333	50.1667	49.6667
Post-	38.8000	61.6667	48.8333	45.6667
Standard Deviation				
Pre-	10.5499	13.6736	10.7036	14.1091
Post-	11.2339	20.9634	8.6352	10.2502
Mean Adjusted for Covariate				
Pre-	52.4381	53.9636	48.8504	50.3209
Post-	47.8381	54.7970	47.5170	46.3209
n	5	6	6	6

TABLE D16. SUMMARY TABLE FOR SPLIT PLOT ANALYSIS OF COVARIANCE FOR VISUAL SEARCH AND SCAN, 4 LETTERS, LINES SCANNED

Dependent Variable: Visual Search & Scan, 4 Letter Target, Lines Scanned

Independent Variables: Between Subjects — Sleep Condition (1 = Full Sleep; 2 = No Sleep); Watch Condition (1 = Morning; 2 = Afternoon)
Within Subjects — Test Time (1 = Pre; 2 = Mid; 3 = Post-Watch on Experimental Day)

Covariate: Visual Search & Scan, 4 Letter target, lines scanned — Tested Post-Watch on Familiarization Day

Source of Variance	Sums of Squares	df	Mean Square	F Ratio	p
Between					
Sleep Condition	646.91068	1	646.91068	2.83	0.1099
Watch Condition	42.98335	1	42.98335	0.19	0.6698
S x W	315.27290	1	315.27290	1.38	0.2557
Covariate	2166.66992	1	2166.66992	9.47	0.0065
Error	4117.09674	18	228.72760		
Within					
Test Time	0.15556	1	0.15556	0.00	0.9611
T x S	39.82222	1	39.82222	0.63	0.3855
T x W	30.48889	1	30.48889	0.48	0.4970
T x S x W	24.02857	1	24.02857	0.38	0.5460
Error	1207.90000	19	63.57368		
TOTAL					

	Full Sleep A.M.	Full Sleep P.M.	No Sleep A.M.	No Sleep P.M.
Mean				
Pre-	40.6000	62.5000	41.3333	44.6667
Post-	38.0000	60.3333	46.1667	43.3333
Standard Deviation				
Pre-	10.9224	19.5729	10.9666	11.2190
Post-	9.3648	20.3142	15.1316	7.8655
Mean Adjusted for Covariate				
Pre-	48.2557	56.3371	43.0057	42.7775
Post-	46.4557	54.1704	47.8390	41.4450
n	5	6	6	6

TABLE D17. SUMMARY TABLE FOR SPLIT PLOT ANALYSIS OF COVARIANCE FOR VISUAL SEARCH AND SCAN, 6 LETTERS, LINES SCANNED

Dependent Variable: Visual Search & Scan, 6 Letter Target, Lines Scanned

Independent Variables: Between Subjects — Sleep Condition (1 = Full Sleep; 2 = No Sleep); Watch Condition (1 = Morning; 2 = Afternoon)

Within Subjects — Test Time (1 = Pre; 2 = Mid; 3 = Post-Watch on Experimental Day)

Covariate: Visual Search & Scan, 6 Letter target, lines scanned — Tested Post-Watch on Familiarization Day

Source of Variance	Sums of Squares	df	Mean Square	F Ratio	p
Between					
Sleep Condition	257.46638	1	257.46638	1.58	0.2252
Watch Condition	75.16486	1	75.16486	0.46	0.5060
S x W	151.93297	1	151.93297	0.93	0.3474
Covariate	4763.44553	1	4763.44553	29.19	0.0001
Error	2937.63781	18	163.20210		
Within					
Test Time	43.82937	1	43.82937	0.45	0.5199
T x S	290.49603	1	290.49603	2.96	0.1015
T x W	7.16270	1	7.16270	0.07	0.7899
T x S x W	0.49603	1	0.49603	0.01	0.9440
Error	1863.41667	19	98.07456		
TOTAL					

	Full Sleep A.M.	Full Sleep P.M.	No Sleep A.M.	No Sleep P.M.
Mean				
Pre-	42.0000	70.1667	41.0000	45.5000
Post-	36.0000	62.1667	44.6667	48.0000
Standard Deviation				
Pre-	18.6145	19.8435	10.2762	18.2181
Post-	6.5574	20.4980	14.7467	11.5065
Mean Adjusted for Covariate				
Pre-	51.0076	58.9153	45.2755	44.9696
Post-	45.0076	50.9153	48.9422	47.4696
n	5	6	6	6

TABLE D18. SUMMARY TABLE FOR SPLIT PLOT ANALYSIS OF COVARIANCE FOR VISUAL SEARCH AND SCAN, 2 LETTERS, CORRECT DETECTIONS

Dependent Variable: Visual Search & Scan, 2 Letter Target, Correct Detections

Independent Variables:			
	Between Subjects	Sleep Condition	(1 = Full Sleep; 2 = No Sleep)
		Watch Condition	(1 = Morning; 2 = Afternoon)
	Within Subjects	Test Time	(1 = Pre; 2 = Mid; 3 = Post-Watch on Experimental Day)

Covariate: Visual Search & Scan, 2 Letter target, Correct Detections — Tested Post-Watch on Familiarization Day

Source of Variance	Sums of Squares	df	Mean Square	F Ratio	p
Between					
Sleep Condition	0.02346	1	0.02346	0.02	0.8993
Watch Condition	0.99556	1	0.99556	0.70	0.4140
S x W	0.22161	1	0.21161	0.16	0.6978
Covariate	4.13886	1	4.13886	2.91	0.1054
Error	25.62781	18	1.42377		
Within					
Test Time	4.11429	1	4.11429	4.86	0.0401
T x S	0.38413	1	0.38413	0.45	0.5089
T x W	0.05079	1	0.05079	0.06	0.8092
T x S x W	1.14603	1	1.14603	1.35	0.2592
Error	16.1000	19	0.84737		
TOTAL					

	Full Sleep A.M.	Full Sleep P.M.	No Sleep A.M.	No Sleep P.M.
Mean				
Pre-	3.0000	3.1667	3.1667	3.5000
Post-	3.4000	4.3333	3.9333	3.6667
Standard Deviation				
Pre-	0.7071	1.3292	0.7528	1.2247
Post-	0.8944	1.8619	0.7528	0.5164
Mean Adjusted for Covariate				
Pre-	3.0705	3.1234	3.1234	3.5278
Post-	3.4705	4.2900	3.7900	3.6945
n	5	6	6	6

TABLE D19. SUMMARY TABLE FOR SPLIT PLOT ANALYSIS OF COVARIANCE FOR VISUAL SEARCH AND SCAN, 4 LETTERS, CORRECT DETECTIONS

Dependent Variable: Visual Search & Scan, 4 Letter Target, Correct Detections

Independent Variables: Between Subjects — Sleep Condition (1 = Full Sleep; 2 = No Sleep); Watch Condition (1 = Morning; 2 = Afternoon)

Within Subjects — Test Time (1 = Pre; 2 = Mid; 3 = Post-Watch on Experimental Day)

Covariate: Visual Search & Scan, 4 Letter target, Correct Detections – Tested Post-Watch on Familiarization Day

Source of Variance	Sums of Squares	df	Mean Square	F Ratio	p
Between					
Sleep Condition	0.49974	1	0.49974	0.29	0.6003
Watch Condition	0.92176	1	0.92176	0.53	0.4786
S x W	5.31012	1	5.31012	3.05	0.1024
Covariate	0.02907	1	0.02907	0.02	0.8990
Error	24.34593	14	1.73900		
Within					
Test Time	0.07206	1	0.07206	0.06	0.8042
T x S	0.77794	1	0.77794	0.69	0.4200
T x W	0.07206	1	0.07206	0.06	0.8042
T x S x W	4.77794	1	4.77794	4.22	0.0578
Error	16.97500	15	1.13167		
TOTAL					

	Full Sleep A.M.	Full Sleep P.M.	No Sleep A.M.	No Sleep P.M.
Mean				
Pre-	2.0000	4.0000	3.5000	2.4000
Post-	2.6000	3.0000	3.2500	3.4000
Standard Deviation				
Pre-	0.7071	0.7071	1.7320	0.8944
Post-	1.9494	0.7071	0.5000	1.3416
Mean Adjusted for Covariate				
Pre-	2.0260	3.9795	3.4969	2.3969
Post-	2.0260	2.9795	3.2469	3.3969
n	5	5	4	5

TABLE D20. SUMMARY TABLE FOR SPLIT PLOT ANALYSIS OF VARIANCE FOR VISUAL SEARCH AND SCAN, 4 LETTERS, CORRECT DETECTIONS

Dependent Variable: Visual Search & Scan, 4 Letter Target, Correct Detections

Independent Variables:			
	Between Subjects	Sleep Condition	(1 = Full Sleep; 2 = No Sleep)
		Watch Condition	(1 = Morning; 2 = Afternoon)
	Within Subjects	Test Time	(1 = Pre; 2 = Mid; 3 = Post-Watch on Experimental Day)

Source of Variance	Sums of Squares	df	Mean Square	F Ratio	p
Between					
Sleep Condition	0.4777	1	0.4777	0.31	0.584
Watch Condition	1.1716	1	1.1716	0.77	0.394
S x W	6.6002	1	6.6002	4.32	0.054
Error	24.44167	16	1.5276		
Within					
Test Time	0.0042	1	0.0042	0.00	0.9518
T x S	1.2287	1	1.2287	1.09	0.312
T x W	0.0042	1	0.0042	0.00	0.952
T x S x W	4.0859	1	4.0859	3.62	0.075
Error	18.0417	16	1.1276		
TOTAL					

	Full Sleep A.M.	Full Sleep P.M.	No Sleep A.M.	No Sleep P.M.
Mean				
Pre-	2.1667	4.0000	3.5000	2.4000
Post-	2.5000	3.0000	3.2500	3.4000
Standard Deviation				
Pre-	0.7528	0.7071	1.7320	0.8944
Post-	1.7607	0.7071	0.5000	1.3416
n	6	5	4	5

TABLE D21. SUMMARY TABLE FOR SPLIT PLOT ANALYSIS OF COVARIANCE FOR DIGIT SPAN FORWARD

Dependent Variable: Digit Span Forward

Independent Variables:
- Between Subjects: Sleep Condition (1 = Full Sleep; 2 = No Sleep); Watch Condition (1 = Morning; 2 = Afternoon)
- Within Subjects: Test Time (1 = Pre; 2 = Mid; 3 = Post-Watch on Experimental Day)

Covariate: Digit Span Forward — Tested Post-Watch on Familiarization Day

Source of Variance	Sums of Squares	df	Mean Square	F Ratio	p
Between					
Sleep Condition	0.07736	1	0.07736	0.05	0.8285
Watch Condition	0.85410	1	0.85410	0.53	0.4743
S x W	0.93085	1	0.93085	0.58	0.4554
Covariate	14.96333	1	14.96333	9.32	0.0063
Error	32.12000	20	1.60600		
Within					
Test Time	6.03571	1	6.03571	7.55	0.0121
T x S	0.23325	1	0.23325	0.29	0.5949
T x W	0.03571	1	0.03571	0.04	0.8347
T x S x W	0.60362	1	0.60362	0.75	0.3948
Error	16.79762	21	0.79989		
TOTAL					

	Full Sleep A.M.	Full Sleep P.M.	No Sleep A.M.	No Sleep P.M.
Mean				
Pre-	7.3333	7.0000	6.6667	7.8571
Post-	8.0000	8.0000	7.5000	8.1429
Standard Deviation				
Pre-	1.6330	1.0955	1.3665	1.3452
Post-	1.0955	0.8944	1.2247	1.0690
Mean Adjusted for Covariate				
Pre-	7.2857	7.1012	6.8424	7.6606
Post-	7.9524	8.1012	7.6757	7.9463
n	6	6	6	7

TABLE D22. SUMMARY TABLE FOR ANALYSIS OF VARIANCE
FOR EMG RESTING LEVEL DURING FIRST HOUR

Dependent Variable: EMG averaged across three segments prior to first incident
Independent Variables: Sleep Condition (1 = Full Sleep, 2 = No Sleep)
Watch Condition (1 = Morning, 2 = Afternoon)

Source	Sum of Squares	df	Mean Squares	F Ratio	p
Sleep Condition	2175.9	1	21795.9	3.18	0.09
Watch Condition	1263.9	1	1263.9	0.18	0.67
S x W	1918.4	1	1918.4	0.28	0.60
Error	1166671.8	17	6830.0		
TOTAL	141140.6	10	7057.0		

	Full Sleep Morning Watch	Full Sleep Afternoon Watch	No Sleep Morning Watch	No Sleep Afternoon Watch
$\bar{X}$	210.5	176.6	252.8	257.7
SD	39.7	74.5	116.0	91.9
N	5	6	4	6

TABLE D23. SUMMARY TABLE FOR ANALYSIS OF VARIANCE
FOR CHANGE IN EMG RESTING LEVEL FROM FIRST TO LAST HOUR

Dependent Variable: Difference in EMG between resting state before first incident and resting state before last incident
Independent Variables: Sleep Condition (1 = Full Sleep, 2 = No Sleep)
Watch Condition (1 = Morning, 2 = Afternoon)

Source	Sum of Squares	df	Mean Squares	F Ratio	p
Sleep Condition	11947.1	1	11947.1	2.64	0.13
Watch Condition	17351.7	1	17351.6	3.83	0.07
S x W	2720.1	1	2720.1	0.60	0.45
Error	63357.5	14	4525.5		
TOTAL	92437.0	17	5437.5		

	Full Sleep Morning Watch	Full Sleep Afternoon Watch	No Sleep Morning Watch	No Sleep Afternoon Watch
$\bar{X}$ of Difference	80.6	10.7	18.5	-21.6
SD of Difference	91.2	46.5	84.3	53.5
N	4	6	4	5

Note: Negative number indicates lower scores at fourth incident.

TABLE D24. SUMMARY TABLE FOR T-TEST COMPARING EMG AT DETECTION AT FIRST AND LAST HOURS

Dependent Variable: EMG Activity During Detection Segment Group: All Subjects						
Independent Variable	**No. of Cases**	**Mean**	**Standard Deviation**	**T-Value**	**df**	**2 Tail Probability**
First Hour	23	227.9	77.2			
Last Hour	23	248.9	89.3			
Difference		-21.1	67.2	-1.51	22	0.15

TABLE D25 SUMMARY TABLE FOR ANALYSIS OF VARIANCE FOR EMG RESPONSE TO INCIDENT IN SECOND HOUR[1]

Dependent Variable: Difference between EMG level at detection and level at one segment before detection, second hour Independent Variables: Sleep Condition (1 = Full Sleep, 2 = No Sleep) Watch Condition (1 = Morning, 2 = Afternoon)					
Source	**Sum of Squares**	**df**	**Mean Squares**	**F Ratio**	**p**
Sleep Condition	173.3	1	173.3	0.04	0.84
Watch Condition	18461.9	1	18461.9	4.40	0.05
S x W	5281.4	1	5281.4	1.26	0.28
Error	62908.2	15	4193.9		
TOTAL	86701.5	18	4816.7		

	Full Sleep Morning Watch	**Full Sleep Afternoon Watch**	**No Sleep Morning Watch**	**No Sleep Afternoon Watch**
$\bar{X}$ of Difference	074.0	27.5	-45.0	-11.1
SD of Difference	99.2	50.8	38.0	63.0
N	4	4	5	6

[1] Positive numbers indicate an increment in EMG activity; negative numbers indicate a decrement.

TABLE D26. SUMMARY TABLE FOR ANALYSIS OF VARIANCE FOR EMG RESPONSE TO INCIDENT IN THIRD HOUR

Dependent Variable: Difference between EMG level at detection segment and one segment before detection, third hour
Independent Variables: Sleep Condition (1 = Full Sleep, 2 = No Sleep)
Watch Condition (1 = Morning, 2 = Afternoon)

Source	Sum of Squares	df	Mean Squares	F Ratio	p
Sleep Condition	3395.8	1	3395.8	1.78	0.20
Watch Condition	1.4	1	1.4	0.001	0.98
S x W	9862.2	1	9862.2	5.16	0.04
Error	30575.2	16	1910.9		
TOTAL	42855.4	19	2308.2		

	Full Sleep Morning Watch	Full Sleep Afternoon Watch	No Sleep Morning Watch	No Sleep Afternoon Watch
$\bar{X}$ of Difference	43.3	−0.1	−32.3	14.0
SD of Difference	51.7	33.8	57.2	33.0
N	5	5	4	6

TABLE D27. SUMMARY TABLE FOR ANALYSIS OF VARIANCE FOR EMG RECOVERY DURING SECOND HOUR

Dependent Variable: Difference in EMG between detection segment and one segment after detection, second hour
Independent Variables: Sleep Condition (1 = Full Sleep, 2 = No Sleep)
Watch Condition (1 = Morning, 2 = Afternoon)

Source	Sum of Squares	df	Mean Squares	F Ratio	p
Sleep Condition	391.7	1	391.7	0.18	0.7
Watch Condition	17498.8	1	17498.8	8.24	0.01
S x W	5951.4	1	5951.4	2.80	0.11
Error	31842.8	15			
TOTAL	55485.1	18	3082.5		

	Full Sleep Morning Watch	Full Sleep Afternoon Watch	No Sleep Morning Watch	No Sleep Afternoon Watch
$\bar{X}$ of Difference	41.9	−60.35	13.8	−16.6
SD of Difference	60.3	54.6	17.6	44.2
N	4	4	5	6

APPENDIX E

INDIVIDUAL SUBJECTS' PHYSIOLOGICAL ACTIVITY MEASURED BEFORE, DURING AND AFTER DETECTION OF AN INCIDENT

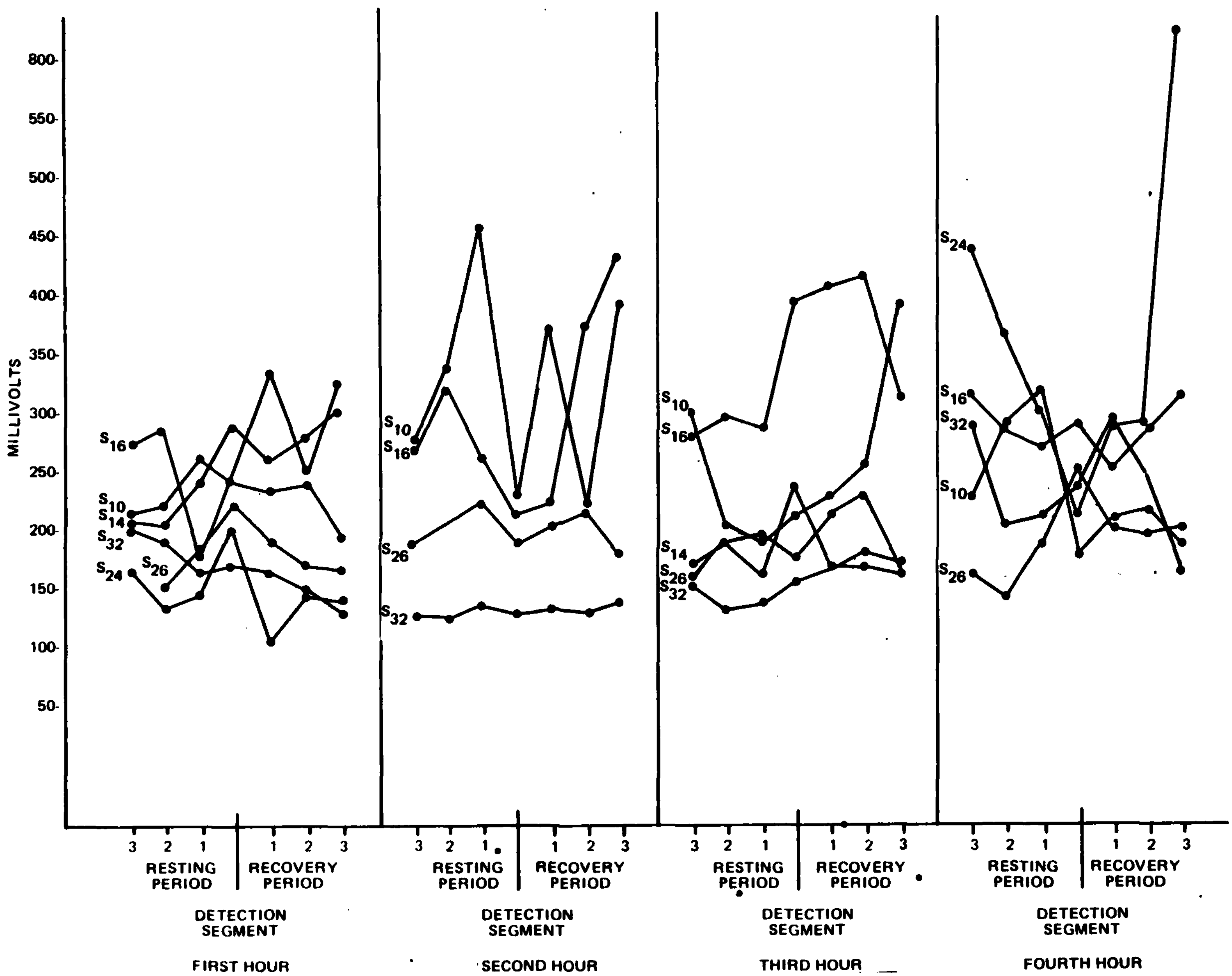

Figure E-1. Measurement of EMG in millivolts for each Subject in the Full Sleep/Morning Group, Before, During and After Detection of an Incident

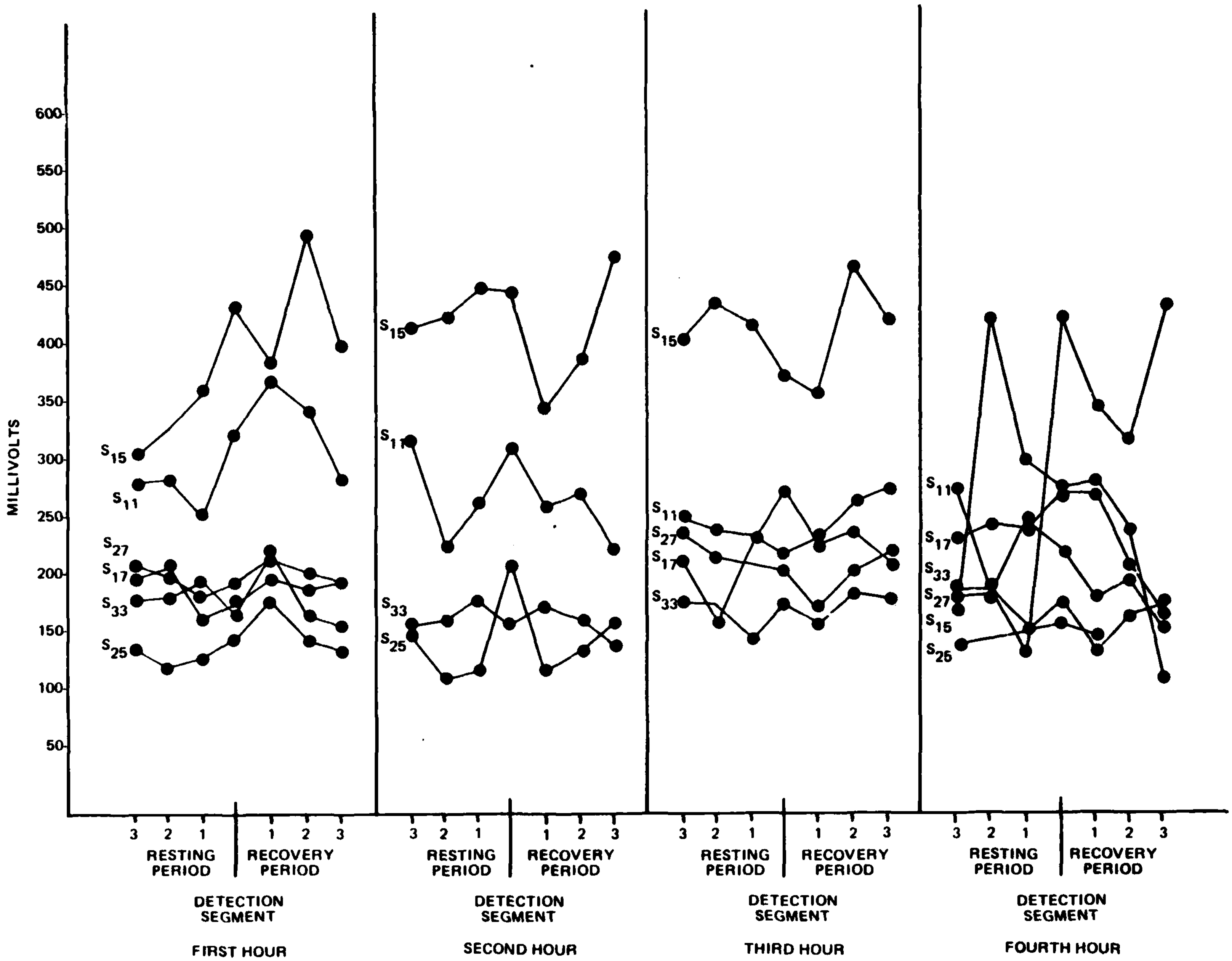

Figure E-2. Measurement of EMG in millivolts for each Subject in the Full Sleep/Afternoon Group, Before, During, and After Detection of an Incident

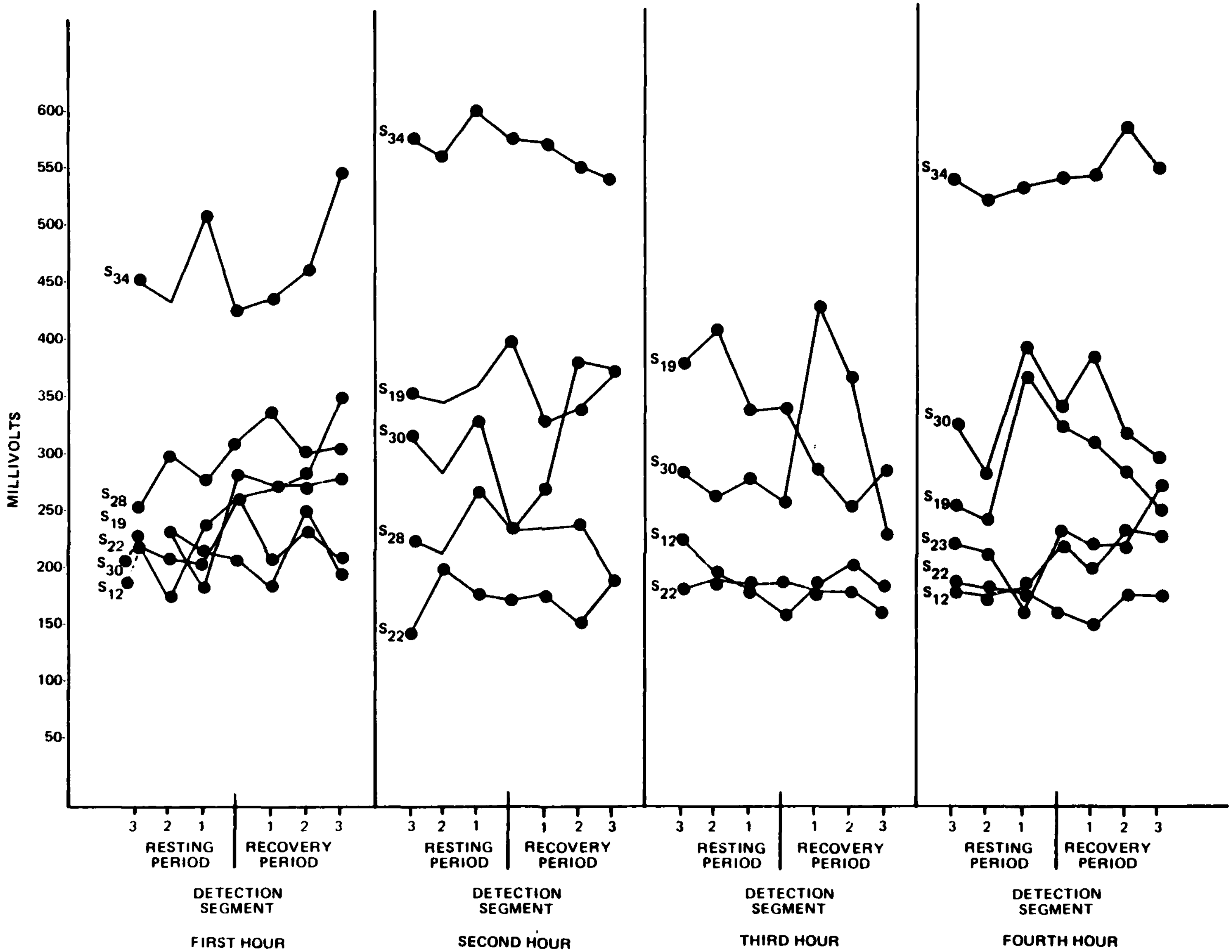

Figure E-3. Measurement of EMG in millivolts for each Subject in the No Sleep/Morning Group, Before, During and After Detection of an Incident

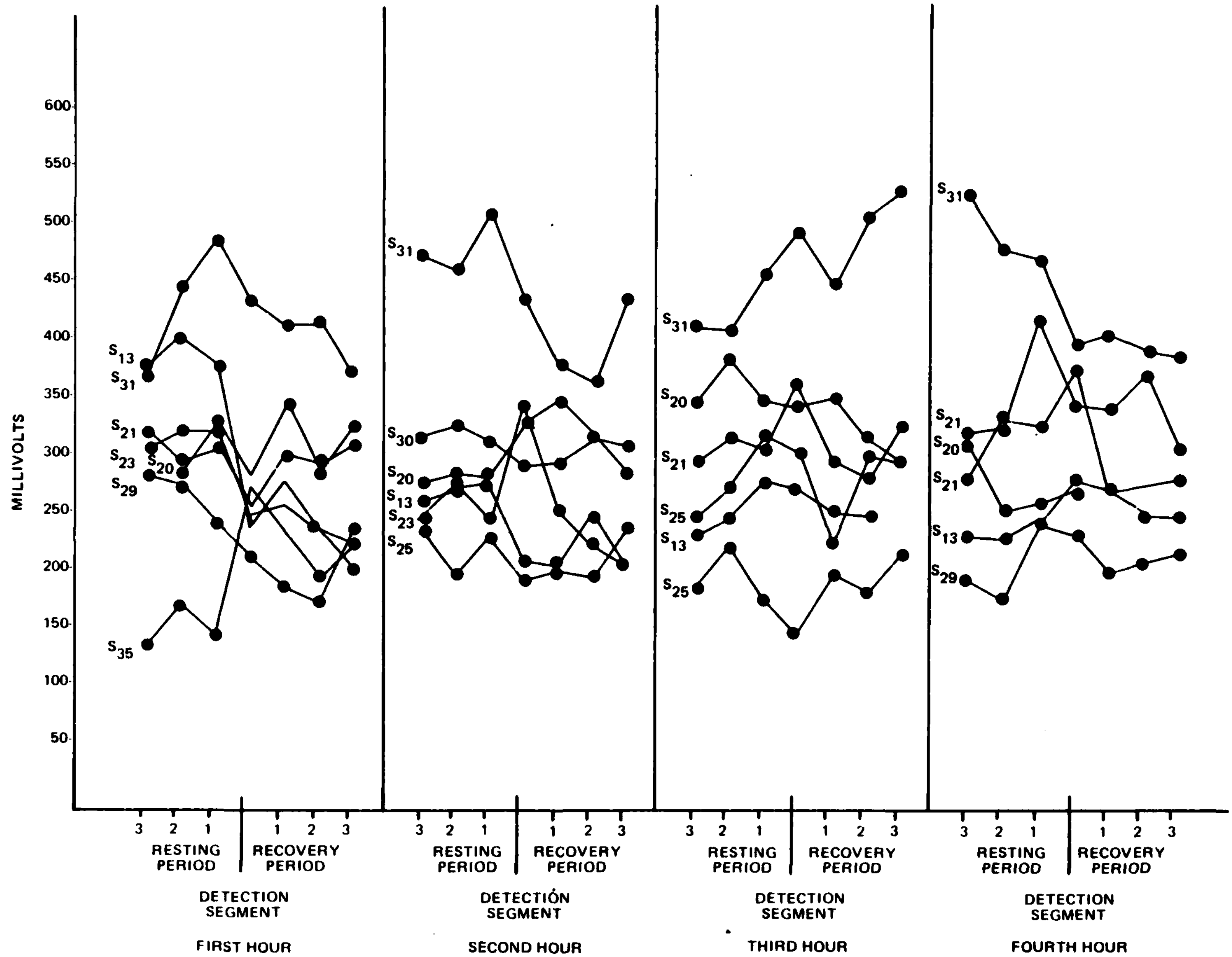

Figure E-4. Measurement of EMG in millivolts for each Subject in the No Sleep/Afternoon Group, Before, During, and After Detection of an Incident

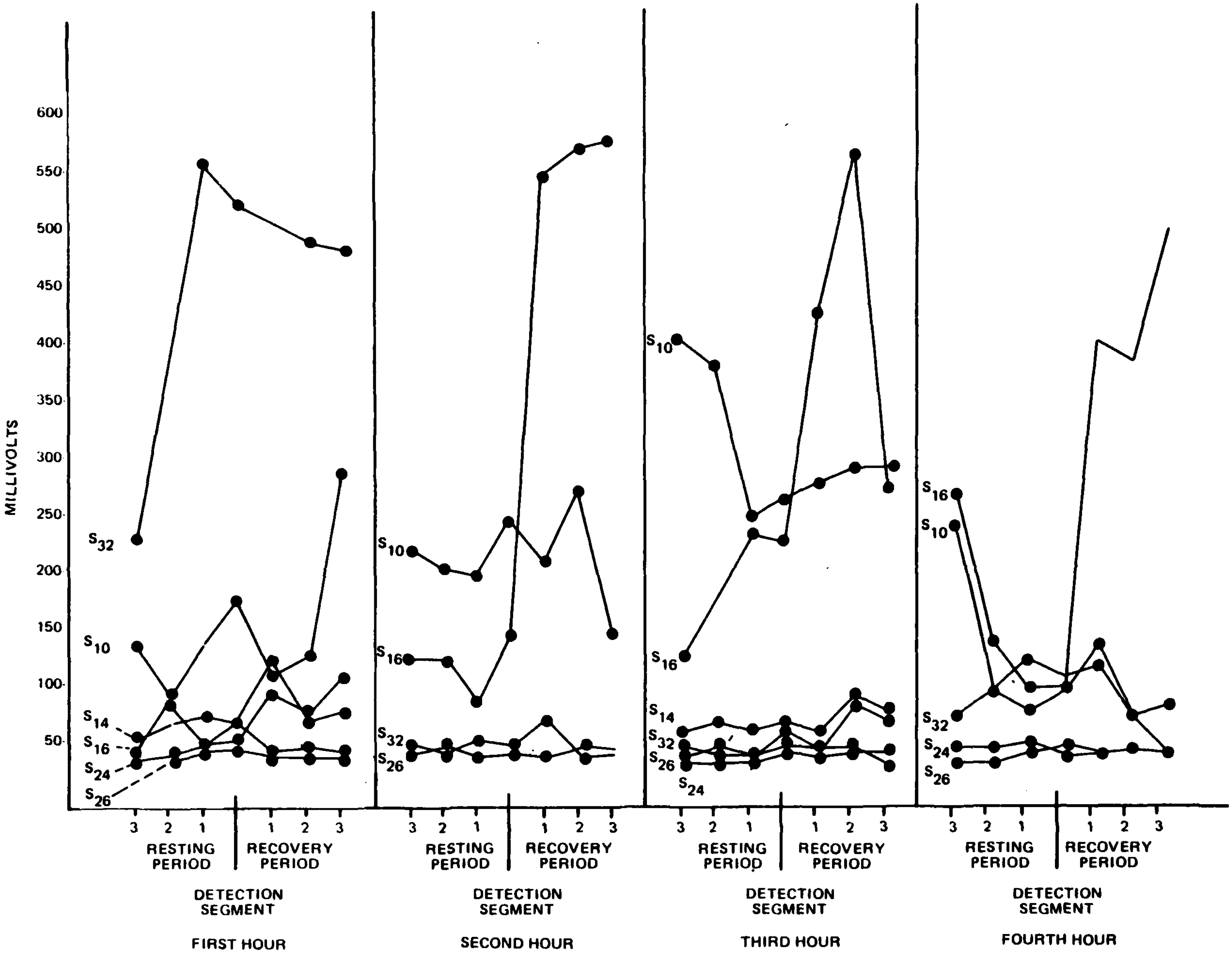

Figure E-5. Measurement of Physical Activity in millivolts for each Subject in the Full Sleep/Morning Group, Before, During and After Detection of an Incident

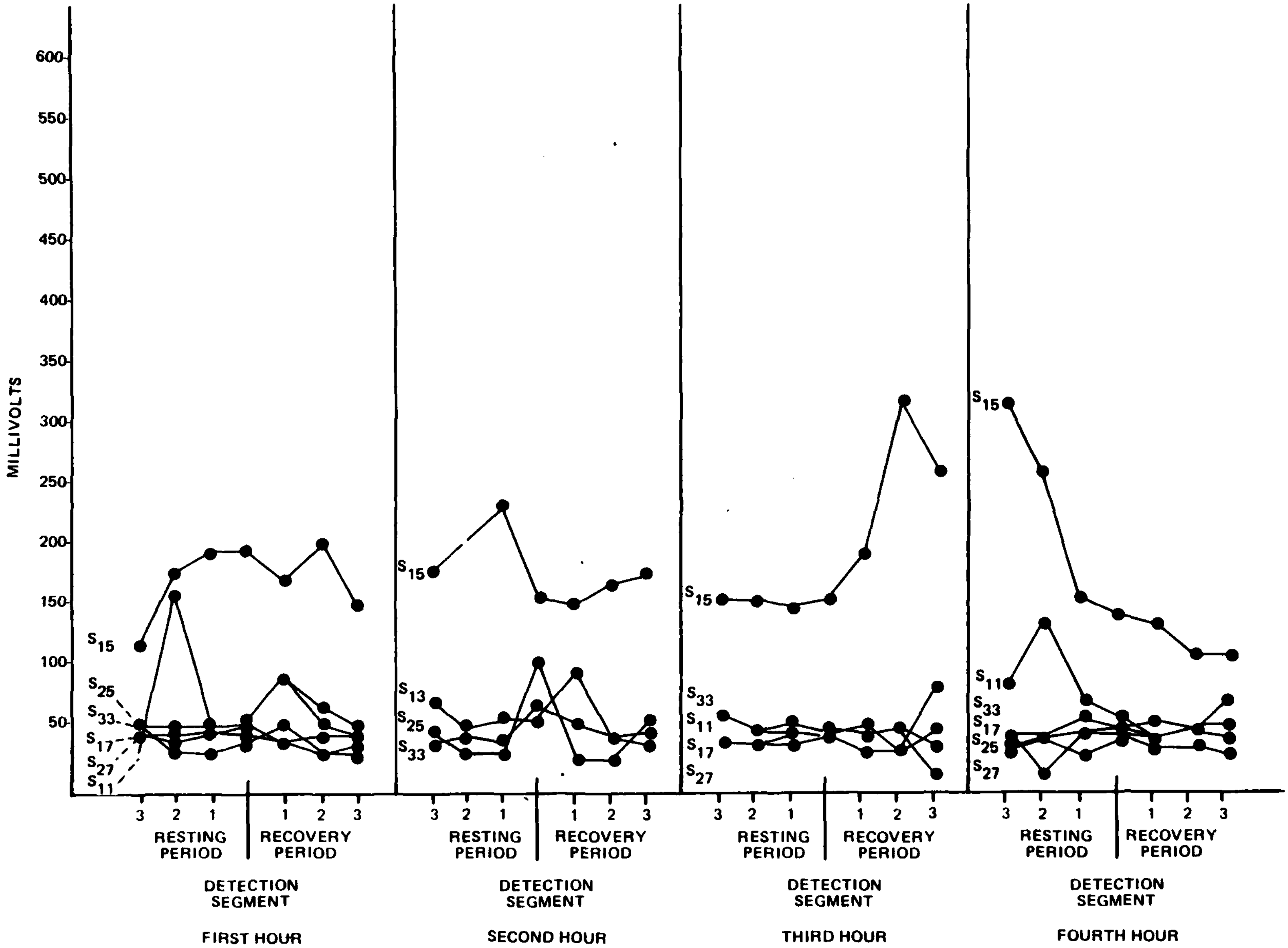

Figure E-6. Measurement of Physical Activity in millivolts for each Subject in the Full Sleep/Afternoon Group, Before, During, and After Detection of an Incident

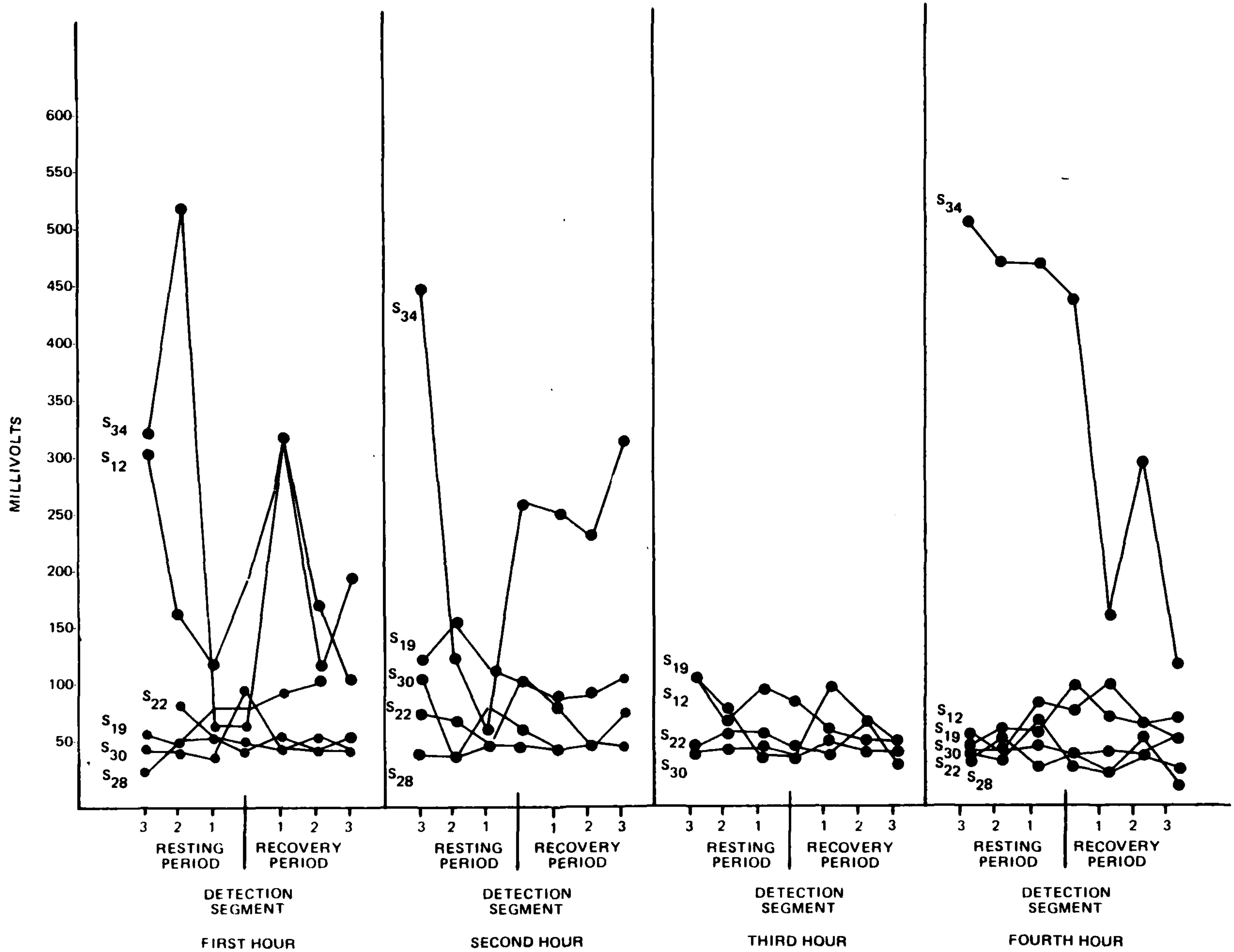

Figure E-7. Measurement of Physical Activity in millivolts for each Subject in the No Sleep/Morning Group, Before, During and After Detection of an Incident

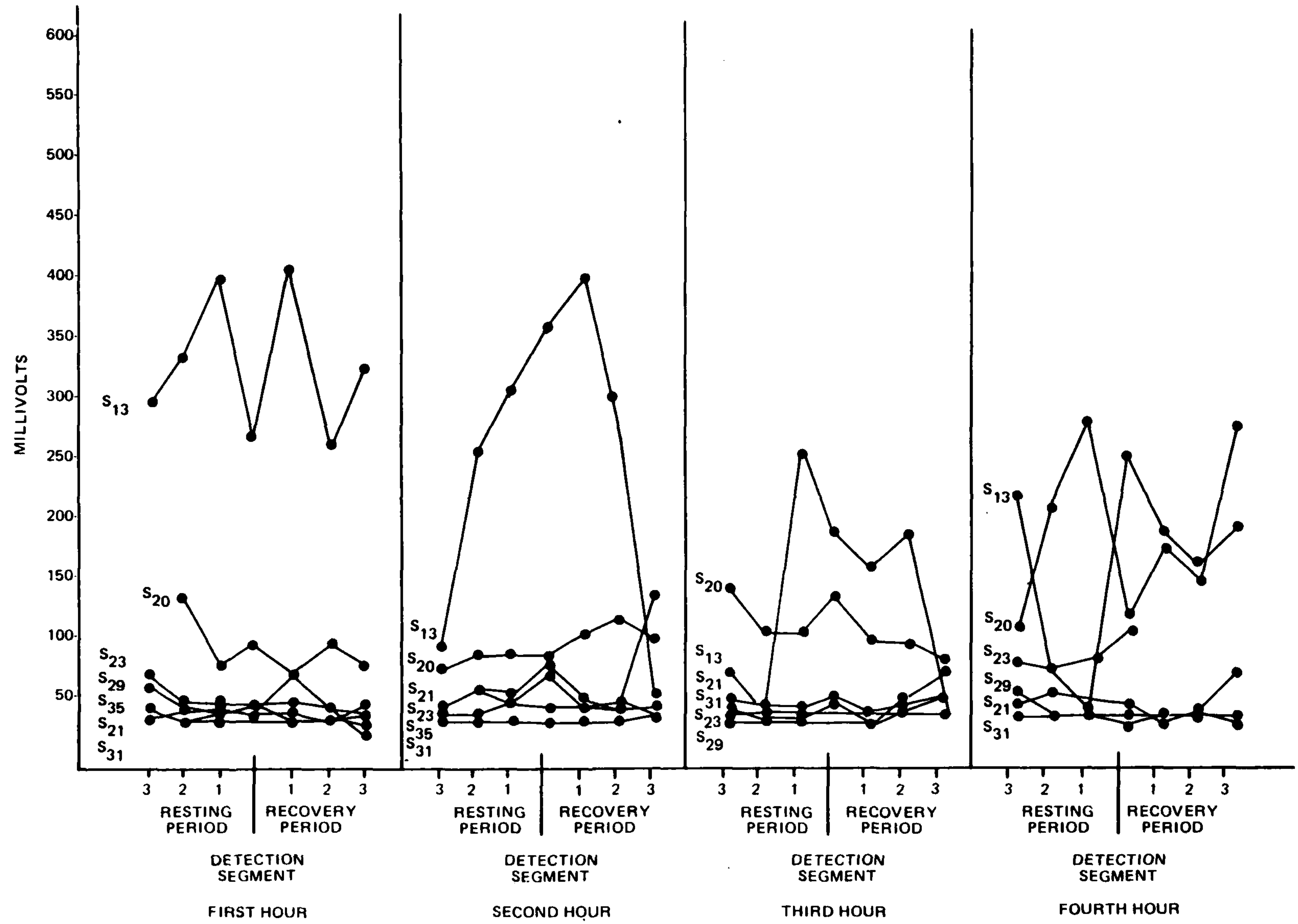

Figure E-8. Measurement of Physical Activity in millivolts for each Subject in the No Sleep/Afternoon Group, Before, During, and After Detection of an Incident

APPENDIX F

DESCRIPTION OF PHYSIOLOGICAL RECORDING TECHNIQUE

Continuous measurements of electrocardiagram (EKG), electromyogram (EMG), eye movement (EOG), and gross body movement were made by means of a small, portable tape recorder strapped to the waist of each participant. (Eye movement data were not used as a dependent measure in this study, however.) The Oxford Medilog Series 4-2 Analog Recording System was utilized. This tape recorder measures 4½ x 3½ x 1½ inches and weighs approximately 14 ounces. Because of the small size, the tape recorder remained reasonably unobtrusive. The recording speed is 24 mm/second and recording time is approximately two hours on a standard audio cassette.

Electrical signals from the forehead and eye muscles of the participants were picked up using silver electrodes approximately ¼ inch in diameter. A potassium gel was applied to the electrodes to facilitate electrical conductivity and the electrodes were fixed to the appropriate areas using a colodion solution and paper adhesive tape. The EMG recordings were made by attaching two electrodes to the forehead, in line with the participant's eye approximately one inch above the eyebrow. The electrodes were configured in a vertical line, approximately ½ inch apart.

Electrical signals from the heart were measured with Hewlett-Packard pre-gelled adhesive electrodes, Part #14245B. Two electrodes were used, one placed in the cardiac notch under the left pectoral muscle, the other placed directly under the sternum. Physical movement was assessed by means of a small plastic box (2 in. x 1 in. x ½ in.) containing a transducer wire with a weight at one end. This box was worn like a wristwatch on the subject's left wrist.

Data collected on cassette tapes were replayed for reduction of an Oxford Medilog PB-2 Playback Unit which is designed to replay cassette tapes at approximately five (5) times recording speed. The playback unit is equipped with four amplifiers, each of which corresponds to one of the tape recorder channels. Each amplifier is provided with a high and low frequency cutoff filter and a gain control. All channels were set similarly for reducing this data with the high frequency filter set at "6," the low frequency filter set at "4" and the gain at "3."

The signal from the playback unit was sent through an Ambulatory Monitoring (Inc) CI-102 Coupler/Isolator. This allowed the signal from the tape recorder to be monitored directly while the participant was wearing it, without danger of feedback from the alternating current power source entering the electrodes and harming the participant. This "on-line" monitoring was useful for ensuring that the recorder was operating properly and for establishing the participants resting baseline prior to the beginning of the experimental task.

The tape recorded data were compiled and statistically reduced using an Autogenics Systems (Inc) Autogen 5600 Data Acquisition Center (DAC). This device inputs electrical signals from the tape playback unit and displays digital, summations of the information, with the level of electrical activity being sampled twice per second. Four channels of the DAC were employed, one for each channel on the tape recorder. Each input channel was isolated from the tape player by an Autogenics I-5000 Optically Isolated A/D Converter as specified by the manufacturer. The DAC is a programmable unit and averages can be taken of the twice-per-second samples over any desired period. For this project, the average was compiled over two minutes and six seconds of tape time, which corresponded roughly to five minutes and thirty seconds in real time. Ten of these averages were collected automatically at which time the tape recorder had to be stopped manually and the DAC reprogrammed. Means and standard deviations for each of these two minute-six second periods were computed by the DAC. After ten periods had been collected and reduced, the DAC automatically activated an Autogenics P-5000 Thermal Printer which listed the means and standard deviations for each period, along with other summary statistics. Digital compilations of EMG, EOG, and physical movement were computed in this way. Data from the EKG record were found to be inappropriate for this type of analysis and a different method had to be utilized. This data was fed from the playback unit directly to a Hewlett Packard 7414A Strip Chart Recorder which produced an analog record of heart rate activity. This record was then reduced by visual inspection. This involved calculating the period of tape time which corresponds to one minute real time and measuring the length of the strip generated during this period. The number of heart beats was then counted over the appropriate period. Although the strip chart recorder is equipped with a gain control, no specific gain setting was utilized, since frequency, rather than amplitude of heart beats, was the critical variable.