AD-A274 707

NAVAL HEALTH RESEARCH CENTER

PERFORMANCE ASSESSMENT IN SUSTAINED OPERATIONS

T. F. Elsmore

P. Naitoh

S. Linnville

DTIC
ELECTE
JAN 14 1994

94-01506

94 1 13 010

Report No. 92-30

Performance assessment in sustained operations

using a computer-based synthetic work task

DTIC QUALITY INSPECTED 8

Timothy F. Elsmore[1]

Paul Naitoh[2], and Steven Linnville[2]

Accesion For		
NTIS CRA&I		☑
DTIC TAB		☐
Unannounced		☐
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and / or Special	
A-1		

---

[1] Walter Reed Army Institute of Research

[2] Naval Health Research Center

Table of Contents

Summary

Problem.

Prediction of the effects of stressors on military performance from laboratory tests frequently involves use of procedures for isolating and measuring psychological constructs assumed to underly military jobs. This approach to performance assessment has been criticized for being oversimplified, ignoring variables known to operate in the control of complex behaviors in operational environments. To address this shortcoming, the "synthetic work" approach takes some of the complexities of operational environments into consideration.

Objective.

The primary objective of the work described in this paper was demonstration of the feasibility of using a microcomputer-based synthetic work test, SYNWORK1, in laboratory and field studies.

Approach.

The SYNWORK1 synthetic work test was incorporated into laboratory studies involving sleep deprivation and exposure to simulated SONAR noise and was employed in a field study conducted by the US Army during Operation Desert Storm.

Results.

Learning of the task was comparable under both laboratory and field conditions. In general, subjects were highly motivated to do well, and continued to improve gradually during up to 250 minutes of exposure to the task. In a laboratory study, SYNWORK1 was shown to be sensitive to the circadian cycle and to sleep deprivation. However, little or no degradation in performance was observed in well rested subjects exposed to noise, or to aircrews under heavy operational loads. The SYNWORK1 software performed well on laptop computers under adverse field conditions.

Conclusions.

The synthetic work approach provides a useful adjunct to existing methods for assessing the impact of stressors on performance in both laboratory and field settings.

Introduction

Military medical research has the global mission of preserving the fighting strength. Many of the questions asked in the pursuit of this end relate to the performance of individual soldiers, sailors, and airmen. While the ultimate measure of the success or failure of a military operation is, of course, whether or not the assigned mission is carried out in a timely manner, researchers must devise means of predicting operational performance from tests carried out under laboratory conditions or in field exercises. These tests are designed to simulate critical aspects of operational environments in the hope that experimental results will predict operational performance, and thus advise policy-making decisions.

Typical questions asked of military performance researchers take the form, "What is the effect of stressor X on performance Y". ("Stressor" is used here to mean any environmental or physiological manipulation which is suspected of deleteriously affecting performance.) Unfortunately, this simple question is not easily answered. The researcher must ask additional questions, such as, "In what age group, what sex, with what degree of training and education, at what time of day, ...". The combinatorial explosion of possible experiments precludes simple answers. The usual approach to this problem has been to break military performances down into component abilities required to perform the task, then assess effects of stressors of interest on these component abilities. Numerous performance assessment batteries (PAB's) of this sort have been devised (AGARD, 1989; Anger, 1990; Englund, Reeves, Shingledecker, Thorne, Wilson, and Hegge, 1985; Kennedy et al., 1981; Perez, Masline, Ramsey, and Urban, 1987), and are currently being used in a variety of situations. Data generated by these PAB's can either be used directly to infer probable operational consequences of a stressor, or as input to computerized models of military tasks or systems (cf. Hegge, 1991).

Modern military operations require military personnel to function in complex situations, simultaneously attending to multiple tasks. Chiles (1982) points out that this area has been largely ignored by behavioral researchers:

> It is an unfortunate, but inescapable, fact that a body of established, reliable data does not exist in this area of the behavioral sciences. Most of the published laboratory research findings on tasks like those found in operational systems have little or no potential relevance, because the tasks were performed alone, as single tasks by themselves without time-sharing requirements. (Chiles, 1982, p. 51)

Typical test batteries discussed above are designed to *isolate* aspects of performance, thus losing most of the dynamic properties and complexity found in operational situations. Another approach is to build test systems that more closely approximate actual military tasks

or systems. Full-blown system simulators or "part" simulators yield data that is specific to the system being simulated, and of great interest to users of that system. However, such systems are relatively rare, expensive, and usually dedicated to training rather than research purposes.

A somewhat more general approach using "synthetic-work tasks" (Alluisi, 1967; Chiles, 1982), attempts to simulate *functional* aspects of a situation requiring simultaneous attention to several tasks. Chiles, Alluisi, and Adams (1968) described such a Multiple Task Performance Battery (MTPB), in which various combinations of six tasks[3] could be simultaneously presented. The original MTPB was a one-of-a-kind hardware device, as have been other synthetic-work consoles (see Hartman, Benes, and Storm, 1980), thus limiting their application. The ready availability of powerful standard personal computers has enabled the construction of PC-based synthetic-work tasks with the potential for broad application. This paper describes one such program, and presents data from two laboratory studies, and one wartime field study.

The SYNWORK1 program

This computer program was designed as a prototype for the implementation of synthetic work tasks on personal computers. It contains two characteristics of real work tasks that are commonly lacking in computer-based tests of performance, the presentation of concurrent tasks and explicit assignment of outcomes for component task performance. The program was not designed to simulate any particular military job. It does however, contain elements common to a number of watch-standing jobs. The program was explicitly designed to allow researchers to readily manipulate the difficulty of the component tasks, as well as the payoff matrix for component task performance.

Task structure

During a test session, the PC screen is divided into four quadrants or "windows", each assigned to a different task. The screen background is dark blue, with double magenta lines delineating the windows. Boxes within windows are drawn in light blue, and informational text is displayed in light red. A small window in the center of the screen is used for displaying a composite "score" for performance on all of the subtasks within the synthetic work environment. The subject is instructed to maximize this score. Figure 1 illustrates the screen.

Upper left window, sternberg memory task: Each session, a list of letters (the "positive list") is chosen from the alphabet and displayed in bright white upper case letters

---

[3] Warning lights, meter monitoring, mental arithmetic, tracking, problem solving, pattern discrimination
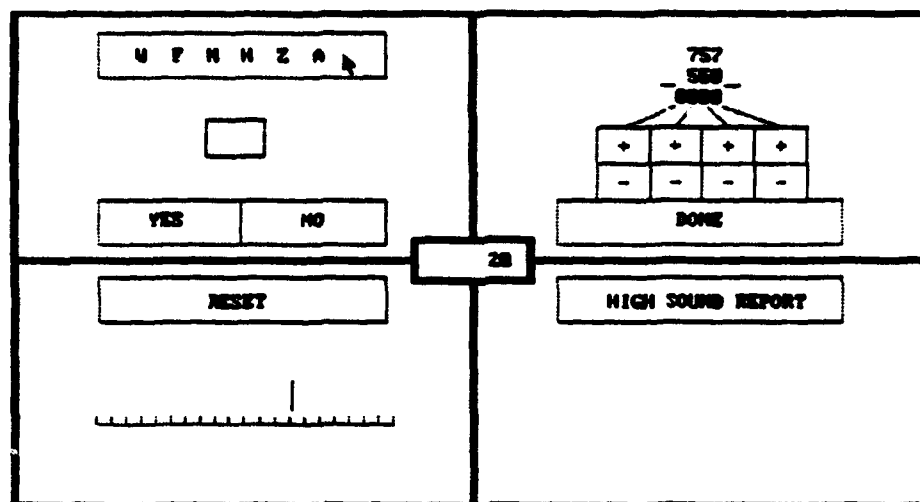
3

Figure 1. The SYNWORK1 screen. Upper left quadrant: memory task, upper right: addition, lower left: visual monitoring, and lower right: auditory signal detection.

in a box at the top of the window. The positive list was displayed for only 5 sec, after which it is replaced by the words "RETRIEVE LIST". When this message was displayed, clicking the mouse on the list box resulted in display of the list for another 5 sec. An equal-sized list (the "negative list") is also selected at the start of each session. At the beginning of each trial, a sample letter is displayed in the box in the center of the window. The subject's task is to indicate, by clicking the mouse on either the YES or the NO box at the bottom of the window, whether the letter is a member of the positive list or not. Ten points are awarded for each correct response, and deducted for each error.

Upper right window, arithmetic task: An addition task presents two randomly selected numbers less than 1000. The numbers are displayed in bright white characters. The subject's task is to adjust the answer by clicking on "+" and "-" boxes below each character of the answer, which is initially set to 0000. Clicking on a box labeled DONE at the bottom of the window results in the presentation of a new problem, addition of 10 points for correct answers, and deduction of 10 points for errors. There are no time limits for completion of this task.

Lower left window, visual monitoring: A bright white pointer moves from the center of a graduated scale, towards either end at a rate of 5 pixels per second. At this speed, the end of the scale is reached in 20 s. Clicking the mouse on a box labeled RESET at the top

4

of the window resets the pointer to the center. The subject's task is to prevent the pointer from reaching the end of the scale. During initial training sessions, the subjects are instructed to reset the pointer when it had traversed about half the distance to the end of the scale. Points are awarded for each reset, with the number of points being proportional to the distance of the pointer from the center at the time of reset. Ten points are deducted when the pointer reaches the end of the scale, and ten additional points are deducted each second the pointer remains at the end of the scale. Thus, the penalty for allowing the pointer to reach the end is severe.

Lower right window, auditory monitoring: Every 5 seconds, a brief tone is sounded in the headphones. The tone is either of two frequencies, low (931 Hz) or high (1234 Hz). The subject's task is to click the mouse in a box at the top of the window labeled HIGH SOUND REPORT following a high tone. The probability of a high tone is 0.2 (i.e. 20% of the tones are high). Correct responses are those that occur following a high tone, prior to the next scheduled tone. All other responses are incorrect. Ten points are awarded for each correct response, and deducted for each error.

Technical details

The SYNWORK1 program was programmed using Microsoft's QuickBasic 4.5 language, augmented with share-ware subroutines for mouse and file handling (Hamlin, 1990) and for timing (Graves and Bradley, 1988, 1991). The timing routines provide accuracy to ±0.1 msec, though actual timing precision may be as poor as ±3 msec, depending on the processing speed of the computer. Even the worst case, however, is more than an order of magnitude better than the ±55 msec that is obtainable by programs using the BIOS timer.

SLEEP DEPRIVATION EFFECTS ON SYNTHETIC WORK

The SYNWORK1 program was initially incorporated into a battery of performance tests administered to subjects in a study of the effects of a nap during 64-hours of sleep deprivation. The data from this study establish that the method is sensitive to fatigue, one of the most common stressors in operational military environments.

Method

Subjects

The subjects were nine male enlisted personnel in either the Navy or the Marine Corps, recruited from a variety of sources. Most had completed High School, but none had completed four years of college.

Apparatus

Four IBM/AT-compatible computers (Unisys 386, 20 Mhz, color VGA monitor, 80 Mbyte hard disk, Microsoft mouse) were used in this study. The computers were located in a single room, with testing stations separated by low partitions.

Testing protocol

Two training sessions were conducted on day 1. Beginning at 0900 at day 2 and continuing until the end of the experiment, a battery of computer-based performance tests was administered every three hours. The battery required approximately 70 minutes to complete, and SYNWORK1 was the last test of the battery. SYNWORK1 sessions were 15 minutes in duration.

Results

The SYNWORK1 program collects data in a manner that permits extremely detailed analysis of the data. The time of each important event during an experimental session is recorded, thus permitting almost total reconstruction of the experimental session. Data analysis programs are available which extract summary measures from the raw data files, and prepare tables suitable for direct importation into standard statistical data analysis programs. Only selected measures will be presented in this paper.

Overall Measures

In Figure 2, the average group composite score (i.e. the overall score for the session), ±s.e.m., is plotted for each of the 23 sessions in the study. Performance continued to improve for the first 6 sessions, reaching its maximum by the 1800h session on day 1. Performance dipped markedly on both days 2 and 3, with minimum scores during the 0600h session on both days. A one-way repeated measures (day, and time of day repeated factors) analysis of variance for scores on days 2 and 3 showed significant effects for both day, $F(1,8) = 10.98$, $p < .011$, and time of day, $F(2.3,18.2^4) = 3.44$, $p < .048$. Thus, performance continued to deteriorate throughout the entire period of sleep deprivation, with the lowest scores in the early morning hours and recovery later in the day.

The rate at which the mouse button is clicked may be taken as an index of the level of effort being exerted by the subject. Figure 3 plots this measure in terms of responses per second. There was a very strong correlation (0.929) between the Response Rate curve and the Composite Score curve. However, analysis of response rates from days 2 and 3 showed no statistically significant effect of day on rate, but did show an effect of time of day,

---

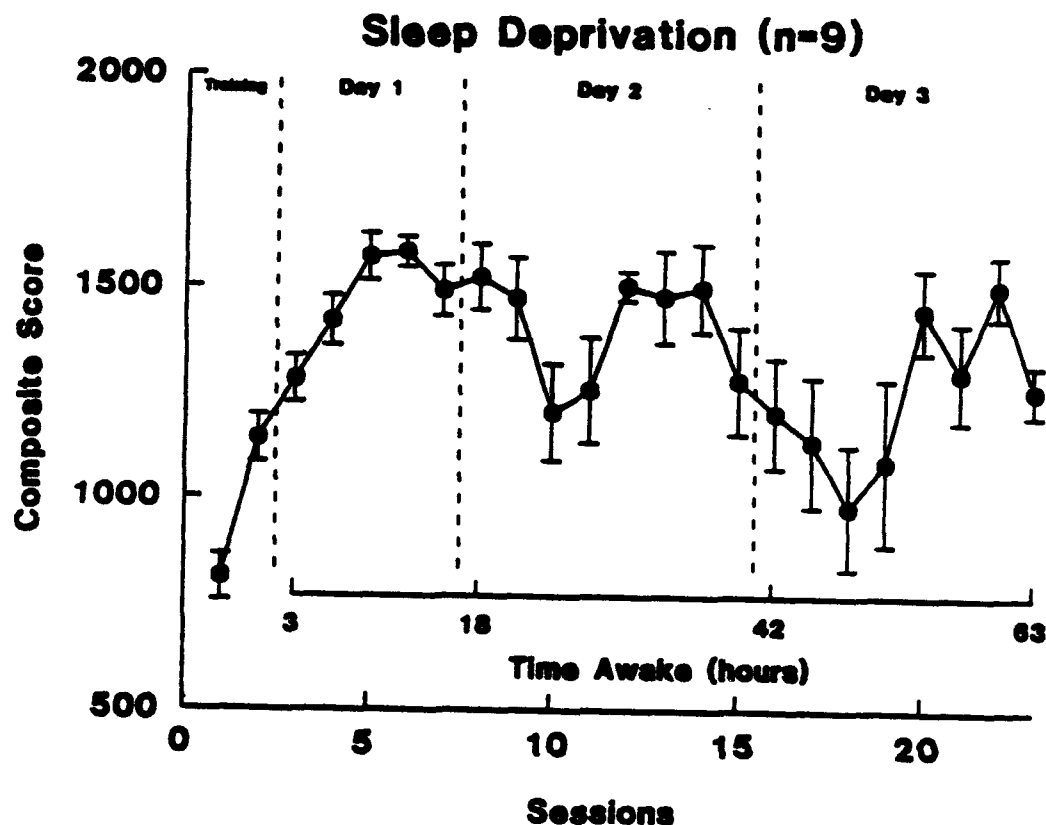[4] Greenhouse-Geisser $\epsilon = .3294$

Figure 2. Composite score on SYNWORK1 during a 64-hour sleep deprivation study as a function of sessions. Points are means of nine subjects ±s.e.m.

$F(2.4, 19.1^5) = 3.60$, $p < .040$. Thus, although the circadian rhythm continued to exert its effect on response rate throughout the sleep deprivation period, overall response rate did not deteriorate as much as the composite score. In fact, the highest average response rate occurred after 60 hours of sleep deprivation.
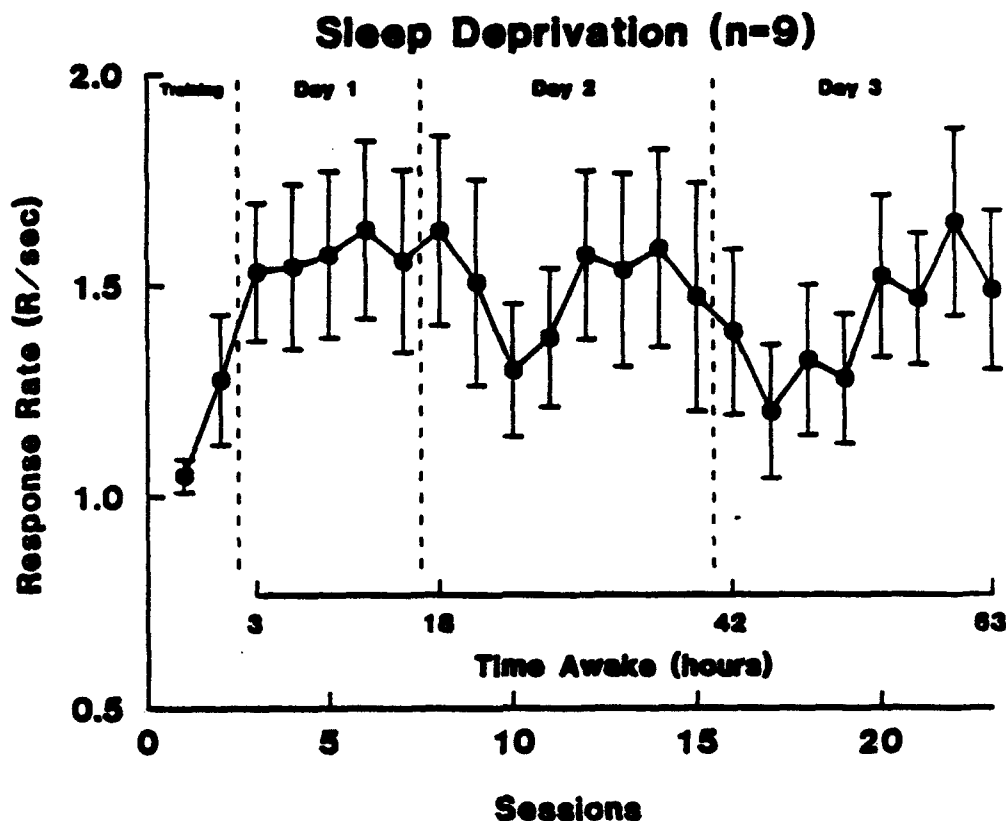
---

[5] Greenhouse–Geisser $\epsilon = .3414$

Figure 3. Overall response rate on SYNWORK1 during a 64-hour sleep deprivation study as a function of sessions. Points are means of nine subjects ±s.e.m.

Separate task measures

A more detailed analysis of the data is presented in Figure 4. In this figure, a selected measure of performance from each of the four sub-tasks within SYNWORK is plotted in a separate frame. The top two frames of this figure present accuracy measures from the memory task and the addition task. The lower left frame presents the number of resets from the visual monitoring task, and the lower right frame presents the percent of the high tones reported on the auditory task. Repeated measures analysis of variance of these four data sets showed no significant effects for time of day, but did show significant effects for day for the addition and auditory vigilance tasks. Visual inspection of Figure 4 shows clear indication of time-of-day effects in the group means for most of the tasks, but also
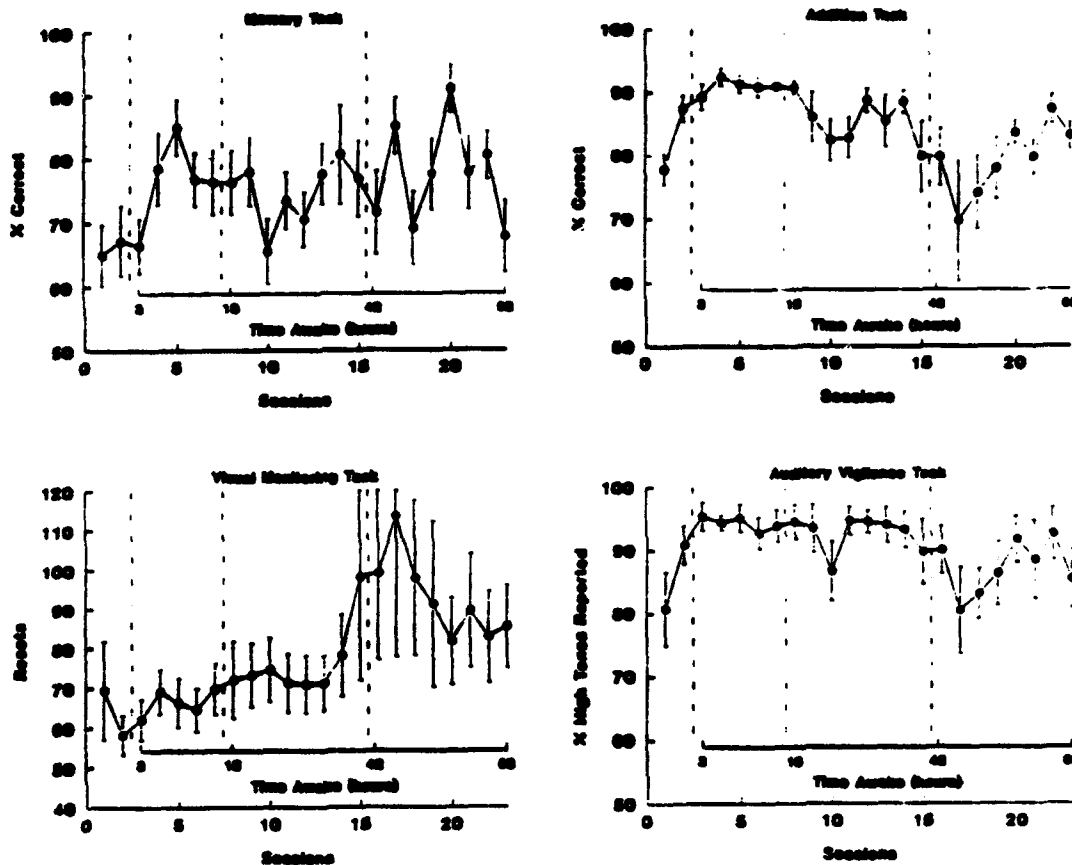
8

Figure 4. Selected measures from SYNWORK1 subtasks during a 64-hour sleep deprivation study as a function of sessions. Points are means of nine subjects ±s.e.m.

increased variance as sleep deprivation progressed, suggesting large individual differences in response to sleep deprivation.

Individual subject data

To illustrate individual differences, Figure 5 presents data from individual subjects for the composite score and percent correct from the addition task. Numbers between the two columns of the figure are subject numbers. Scales for the two columns are shown to the right and left of the plots for subject 816. This figure shows substantial differences between subjects in their response to sleep deprivation. For example, subjects 620 and 810 appeared
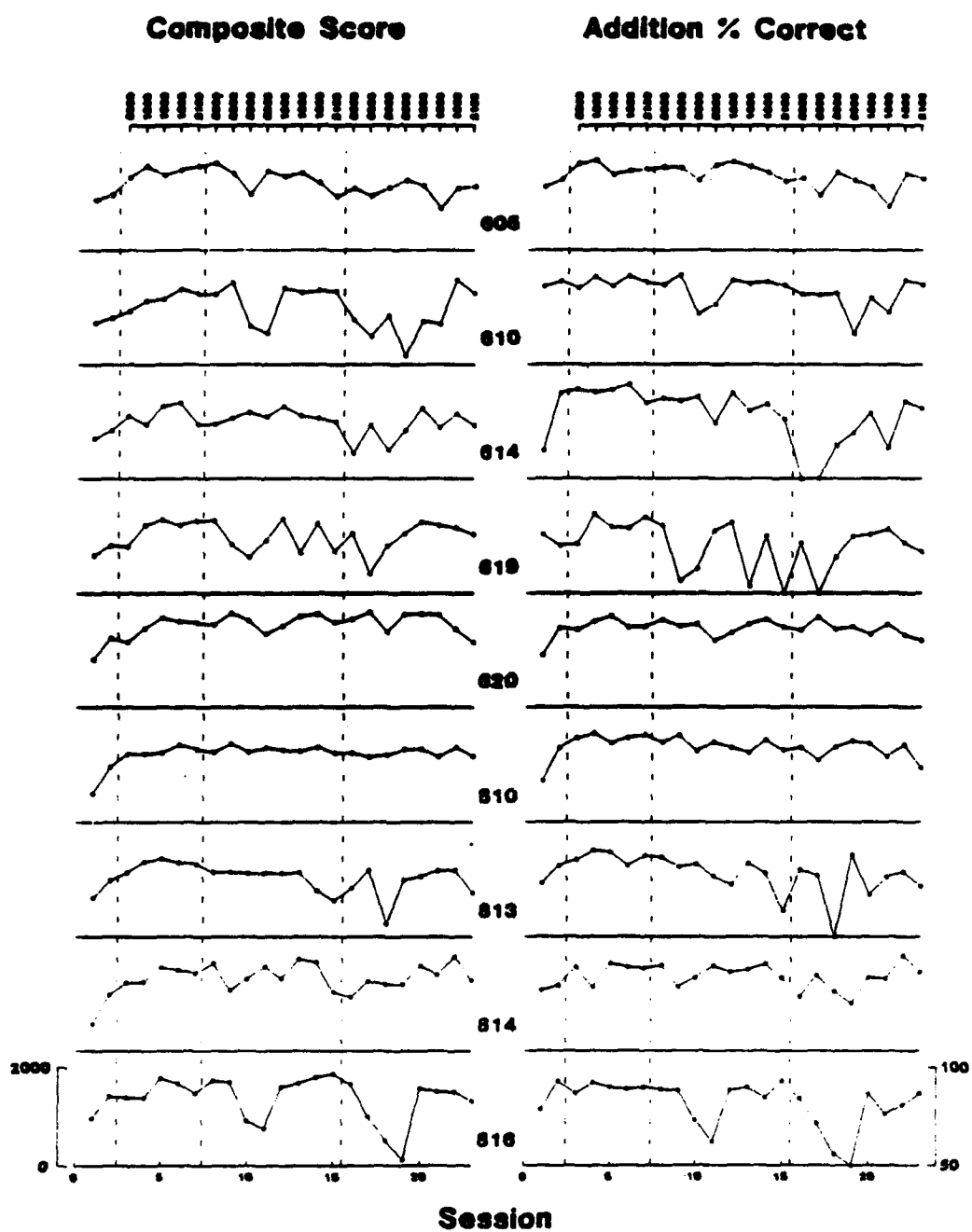
9

Figure 5. Composite score and percent correct on the addition subtask for all sessions and all subjects.

to be relatively unaffected, while subjects 610 and 816 showed very profound time-of-day effects on both of these measures.

Individual differences are illustrated even more strikingly in Figure 6. Again, scales are shown for subject 816. For visual monitoring, the performance of 6 of the 9 subjects did not vary throughout the experiment, whereas 3 of the subjects showed substantial increases either transiently (614 and 619) or steadily throughout the study (813). Both Figures 5 and 6 clearly show that the most severe performance decrements tended to occur in the early morning hours. In most cases, performance recovered to some degree in the afternoon and evening, indicating that even in severely sleep-deprived subjects, more or less normal circadian rhythms in performance continue to be observed (cf. Brown and Graeber, 1982).

The increase in resets shown in Figure 6 reflects a decrease in efficiency on this task, since the optimal performance is to wait as long as possible between resets. Figure 7 illustrates this shift in strategy for subject 619. In this figure, cumulative mouse button clicks are plotted as a function of time in the session. Thus a steady rate of response would appear as a line of constant slope. Downward pips on the record represent correct responses, and upward pips represent errors. The left plot is from session 16, for which performance on the visual monitoring task was normal. The right plot is from session 17, in which this subject had approximately 4 times his normal number of resets. Approximately 10 minutes into the session, indicated by the arrow, the subject began repeatedly resetting the pointer, reflected by the high slope of the cumulative record, and the large number of downward deflections.

More detailed aspects of performance breakdown for subject 619 are illustrated in Figure 8 which shows the temporal distribution of mouse clicks in each of the four tasks for 619 for sessions 16 and 17. Performance on the memory task was comparable on both sessions. Relative to session 16, addition performance on session 17 was very intermittent, with a long gap of five minutes with no activity at all. During this period, there was a very large increase in activity on the visual monitoring task. Performance on the auditory detection task was also greatly decremented on session 17, to about 50% of the activity seen in session 16.

Discussion

The data presented here demonstrate that performance on the SYNWORK1 multi-task synthetic work program is sensitive to sleep deprivation. The composite score declined as time without sleep increased, with the decrement being most evident in morning sessions, followed by recovery later in the day.

More detailed measures of performance appeared to be less sensitive to sleep deprivation than composite score. However, this seems to be a reflection of differences between subjects in terms of how they coped with the task, that is, individual differences in
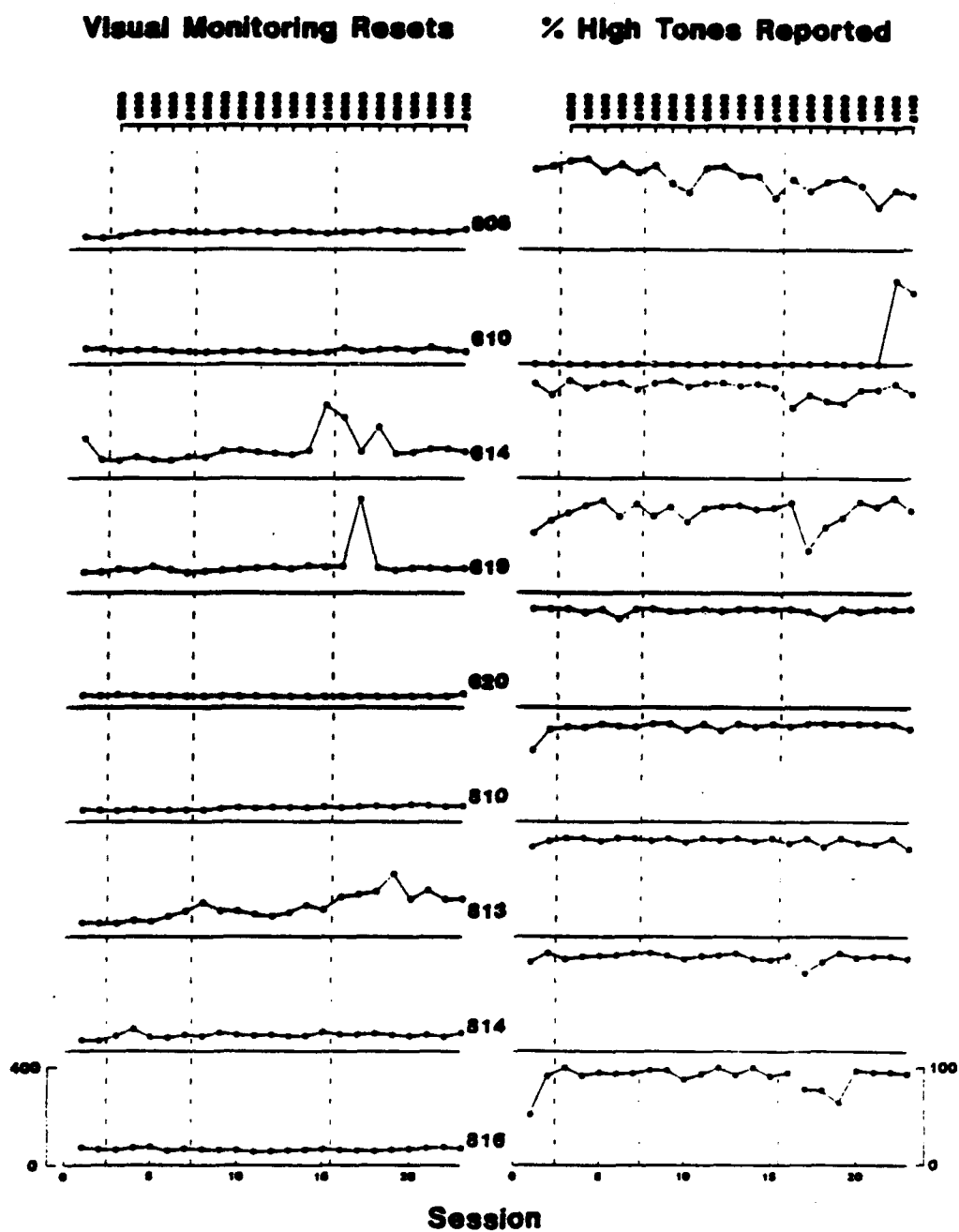
11

Figure 6. Visual monitoring resets and percent high tones reported for all sessions and all subjects.
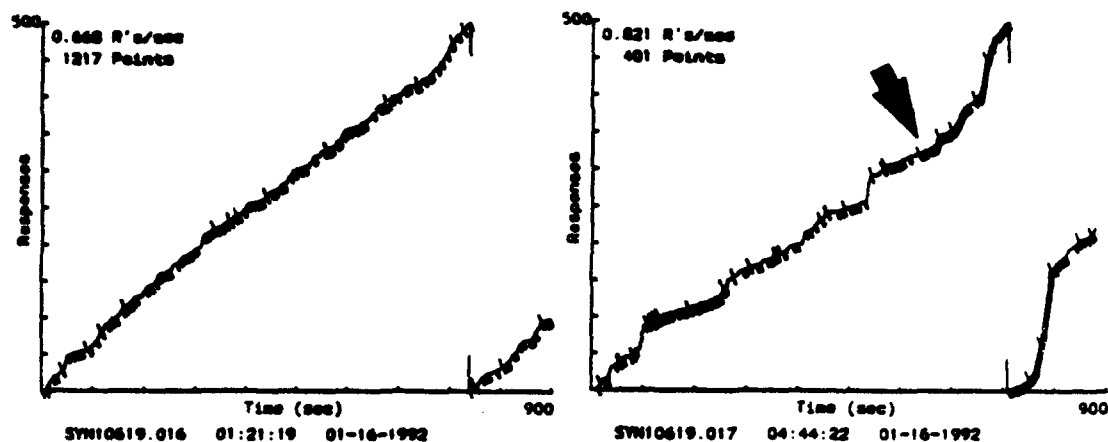
12

Figure 7. Cumulative mouse button presses as a function of time for sessions 16 and 17 for subject 619. Downward pips indicate point awards, and upward pips represent errors.

strategy. SYNWORK1 was explicitly designed to permit this type of variation to be observed, since this is an aspect of real-world job performance that is not measured in traditional Performance Assessment Batteries. Of the four sub-tasks within SYNWORK1, response omission was penalized on only one, the visual monitoring task. Thus, when severely sleep deprived, three of the nine subjects concentrated their efforts on this task, while devoting fewer resources to the other component tasks. These effects were manifested as local breakdowns in performance. Presumably, more consistent performance could be maintained by providing tighter contingencies for all tasks (e.g. point penalties for response omissions).

The present study demonstrates that SYNWORK1 can serve as a research framework for investigating variables that determine how complex, multi-component jobs may be expected to deteriorate under stress. Since the program does not simulate any particular military job, the results must be considered as indicators of functional performance class breakdown. Clearly, if more system-specific information is desired, other synthetic work tasks or part-task simulators will be required.

SYNTHETIC WORK EFFECTS IN A SIMULATED SONAR ENVIRONMENT

As an illustration of the application of SYNWORK to another type of stressor, the procedure was modified slightly for a study on exposure to simulated sonar noise. Sailors aboard surface ships and submarines may be subjected to extended periods of exposure to sonar "pinging" sounds. The health and performance effects of this noise exposure are

13
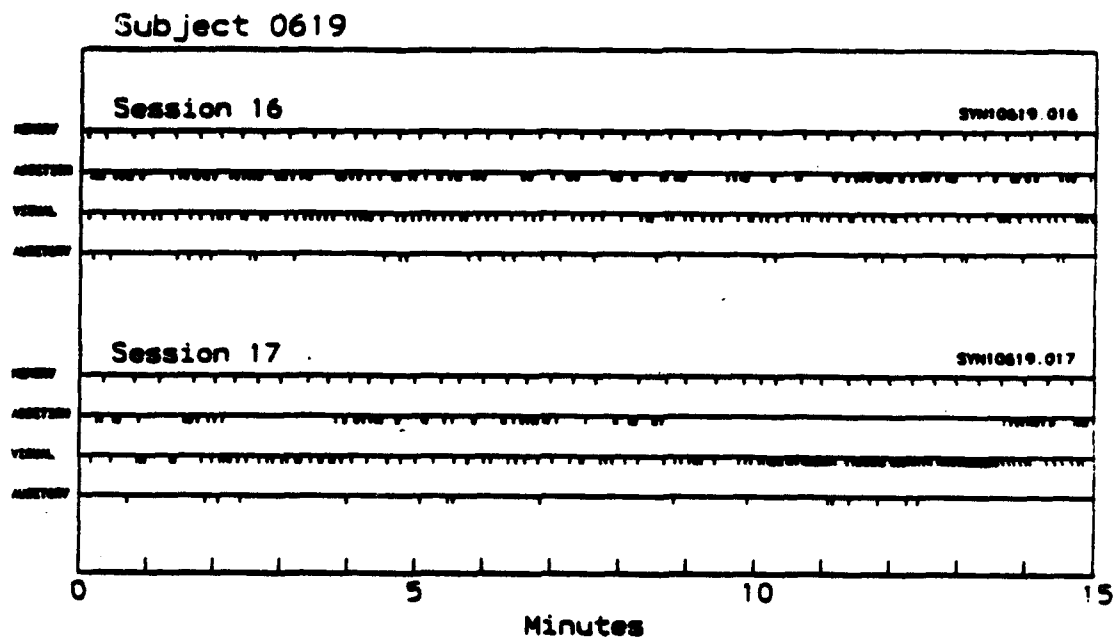
Figure 8. Simulated strip-chart recordings for sessions 16 and 17 for subject 619. Each line shows mouse button clicks in one of the four sub-tasks.

generally unknown. The purpose of this study was to evaluate effects of 24 hours of simulated sonar pings on performance.

Method

The subjects were 38 enlisted military volunteers who lived in a self-contained laboratory for the duration of the study. Due to the auditory nature of the study, the SYNWORK1 task was modified slightly by replacing the auditory detection with a functionally similar visual target detection task. In this task, a filled circle appeared in the lower right window on 80% of the trials, and a target appeared on the other 20%. The subject's task was to indicate when the target appeared. After initial training on the synthetic work task, subjects were tested three times a day for four days. Test sessions were 19 minutes in duration.

On the third day, simulated sonar pings consisting of six consecutive 1-second pure tones were presented at intervals of 30 seconds. The pings were broadcast through multiple speakers located throughout the experimental space. Sound intensity was calibrated to be 83 dB for 9 subjects and 89 dB for 29 subjects. Pinging began at 0800h on day 3 and

14
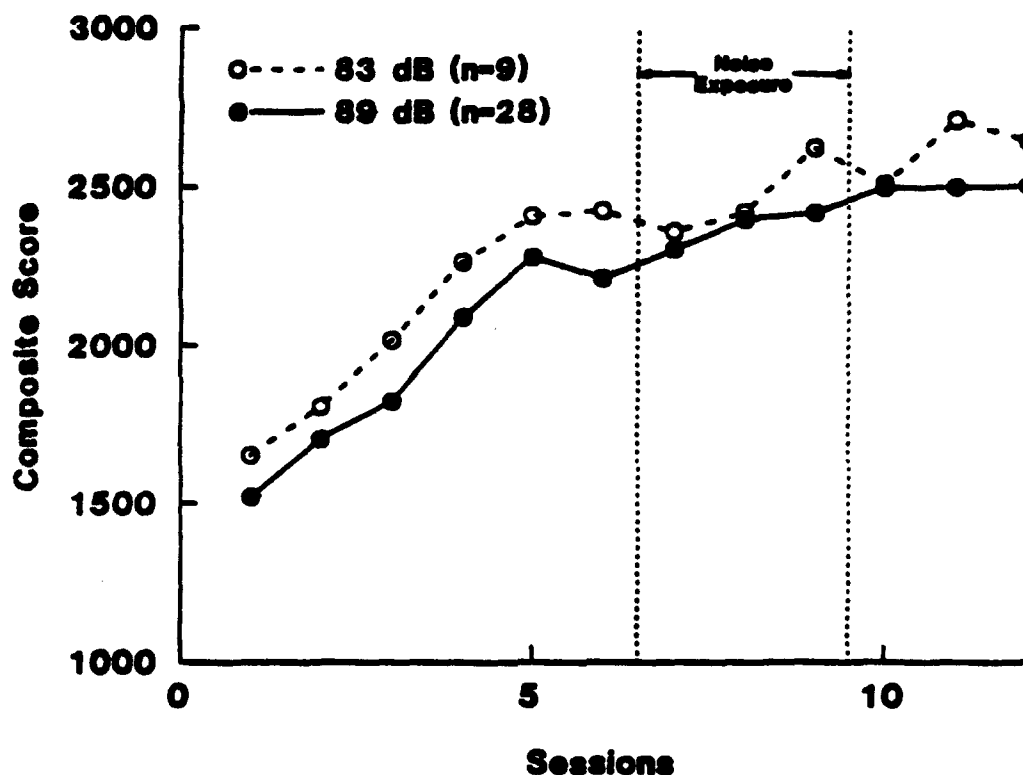
Figure 9. Mean composite score for both groups in the Ping study. Sessions 7, 8, and 9 were run during 24 h. of continuous exposure to simulated sonar noise.

continued for 24 hours.

Results

In general, no adverse effects of noise exposure were detected in either the 83 dB or the 89 dB group. All subjects showed normal acquisition of the task. Figure 9 shows the average composite score for both groups. Performance continued to improve for the duration of the study.

Figure 10. Selected measures from SYNWORK1 subtasks during the Ping study.

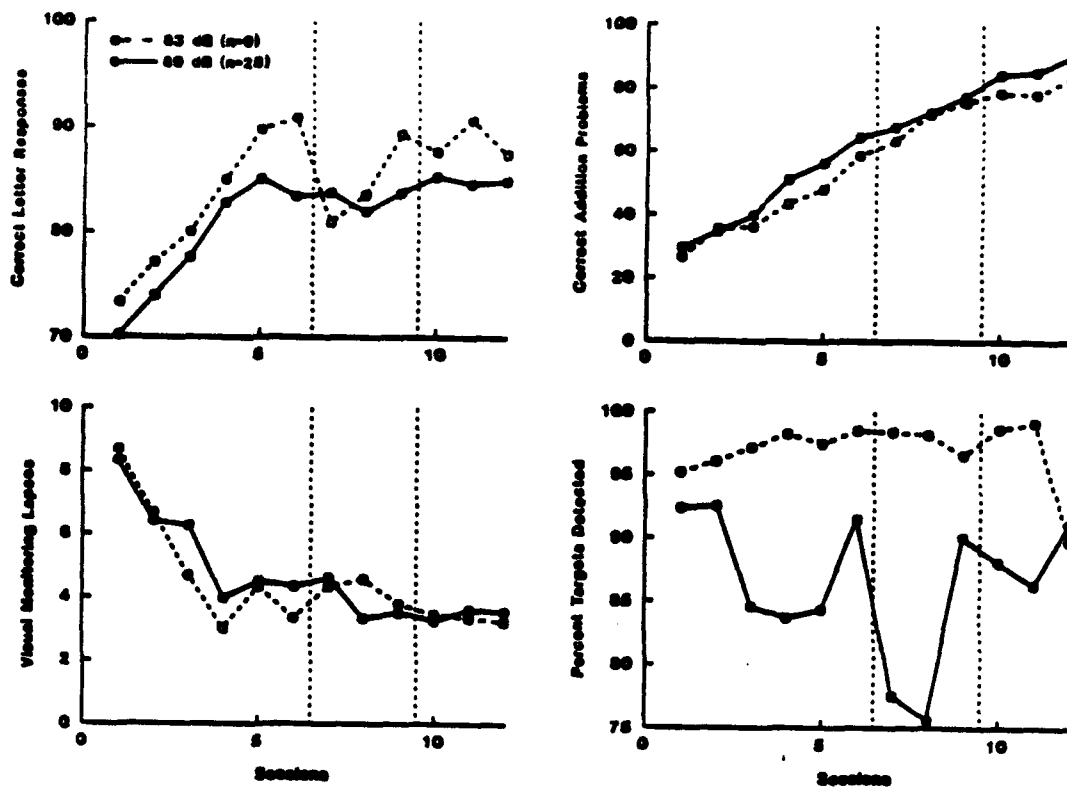Selected measures from each of the four tasks are plotted in Figure 10. Two-way repeated measures analyses of variance showed no significant group effects for any of these measures. While there are suggestions of effects on the memory and target detection tasks, extremely large individual differences were observed. This figure does illustrate, however, that the improvement in overall score in the first few sessions was largely a reflection of improvements in three of the four tasks. Performance on the target detection task started out at a relatively high level, and, if anything, deteriorated as the experiment progressed.

Discussion

Previous studies have reported that acute exposure to noise may result in significant decrements in performance (Boff and Lincoln, 1988; DeJoy, 1984). There are several possible explanations for the fact that in the present study, no deterioration in SYNWORK performance was observed during 24 h of exposure to simulated sonar noise. One possibility

16

is that the noise parameters (intensity, frequency) used in this study may not have been severe enough to produce effects on performance. Secondly, adaptation to the noise may have occurred during the 24-hour exposure. Finally, the SYNWORK task may simply not be sensitive enough to detect noise-induced changes. It is not possible to discriminate among these possibilities without additional research, and further speculation is beyond the scope of the present paper.

SYNTHETIC WORK IN OPERATION DESERT STORM

In the fall of 1990, the U.S. Army Medical Research and Development command initiated several projects designed to assess possible medically-related problems encountered by US troops during operations Desert Shield and Desert Storm. As part of this effort, a team of researchers was deployed to study personnel in the 101st Airborne Division[6]. This team chose SYNWORK1 to monitor possible deterioration in performance of aircrews during a period of intense operational activity.

Method

Subjects

The subjects were soldiers serving in the 101st Airborne Division of the U. S. Army during the Persian Gulf War of 1991. All subjects in the study were volunteers. The primary criteria for inclusion in the study were availability and willingness to participate. The two groups investigated included helicopter aircrew members from the 1st Battalion, and officer and enlisted personnel in the Division Tactical Operations Center. During the testing, both groups were under a heavy workload, with the 1st Battalion flying nighttime armed reconnaissance missions, and the TOC processing prisoners of war. Subjects were all males ranging in age from 19 to 44 and in rank from E4 to O4.

Apparatus

Portable NEC 386SX laptop computers were used to run SYNWORK1. The machines were equipped with a hard disk, backlit LCD EGA screen and a Microsoft mouse. The computers were modified with a headphone jack for presentation of the auditory stimuli through headphones. During tests, the PC was placed on a field table located in a tent with dim fluorescent lighting, with the subjects seated on a folding chair. The 1st Battalion testing area was located in a patient holding area which was relatively quiet, providing few

---

[6] The team leader was MAJ John Leu of the Walter Reed Army Institute of Research; other team members were MAJ Mary Mays of the US Army Research Institute of Environmental Medicine, and CPT Kathy Popp of WRAIR.

distractions. In contrast, the TOC subjects were tested a tent with a great deal of bustle, noise, and chaos. Due to extremely dusty conditions, it was necessary to frequently clean the mice.

Since the subjects in this study were from the 101st Airborne Division, they were instructed to view the synthetic work task as being analogous to the flight deck of an aircraft.

Testing protocol

Test sessions were scheduled at the convenience of the subjects. The first two sessions were usually five minutes in duration, and were regarded as training sessions. The remainder of the sessions were 15 minutes long. One to four sessions were conducted per day. All tests were conducted during February, 1991, prior to the start of the ground offensive.

Results

Only data from subjects who completed at least 8 sessions will be presented. The SYNWORK1 program yields a great deal of data. Since this work was exploratory in nature, only figures summarizing the performance of the two groups of subjects will be presented. Figure 11 shows data from 10 subjects from the 1stBN, and Figure 12 shows the same variables for 6 subjects from the TOC. In these figures, each frame shows a different measure plotted as a function of session. Recall that the first two sessions were only 5 minutes long, and the remainder were 15 minutes. Each point is a score from a single subject. To illustrate changes in average performance, the lines represent smoothed fits to the group data using the method of distance weighted least squares (McLain, 1974).

The top two frames show that the composite score and overall response rate continued to increase at a relatively constant rate for the entire 12 sessions. The rapid rise on the third session was due to the fact that session was three times the length of the preceding two. Overall rate, which is the number of mouse button clicks per second, showed an orderly rise throughout the study. These frames, as well as most of the other frames, show a great deal of similarity between the two groups.

The second row shows data from the Sternberg memory task. Most subjects rapidly acquired the task. Examination of the data, however, showed that over half of the subjects did not memorize the list each session, clicking on the "RETRIEVE LIST" box each time a new probe letter appeared. There were no differences between the two groups in this regard. The 1stBN appeared somewhat faster in the speed with which the Sternberg task was performed, as reflected by the right panel in row two of Figures 11 and 12, although because of the small number of subjects and large between-subject variability, the difference is not statistically significant.
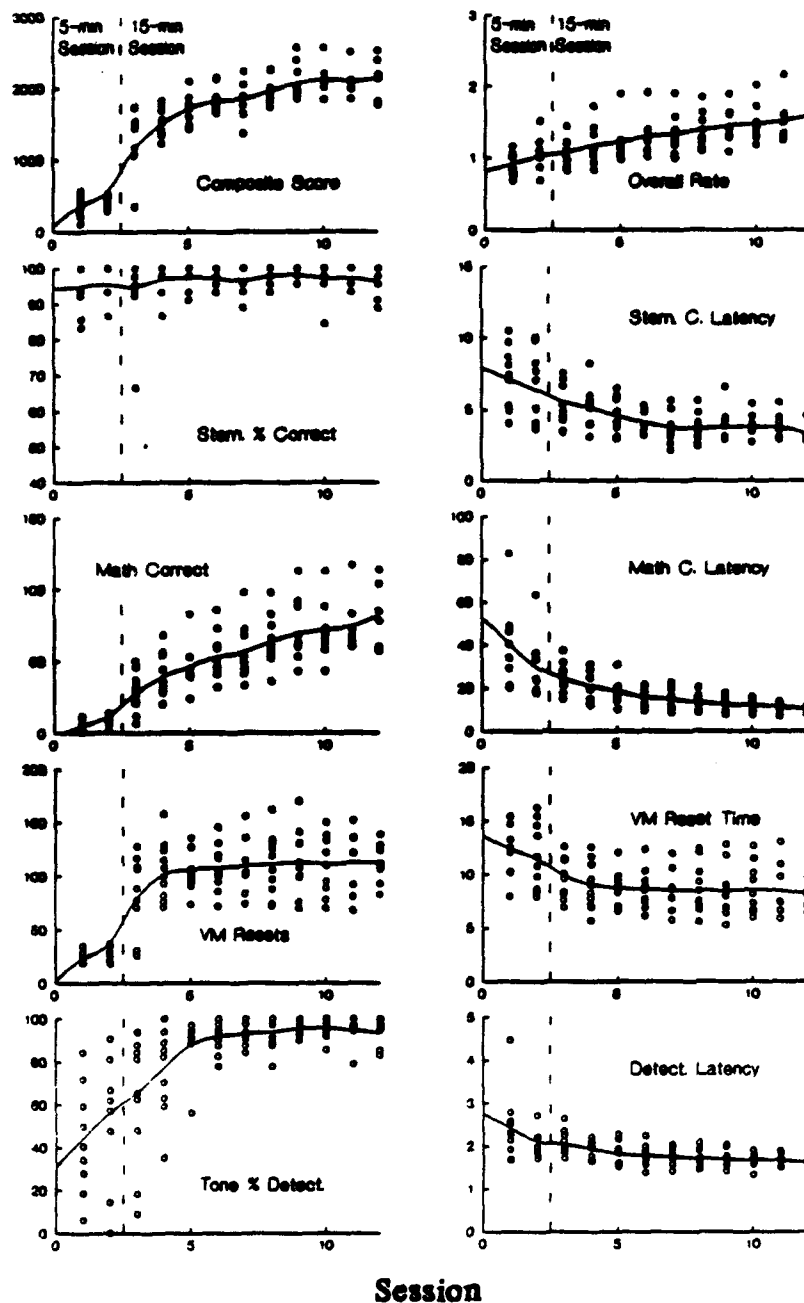
18

Figure 11. Selected measures from the 1st Battalion of the 101st Airborne Division in the Desert Storm study. The smooth functions are fitted to the individual subject data.
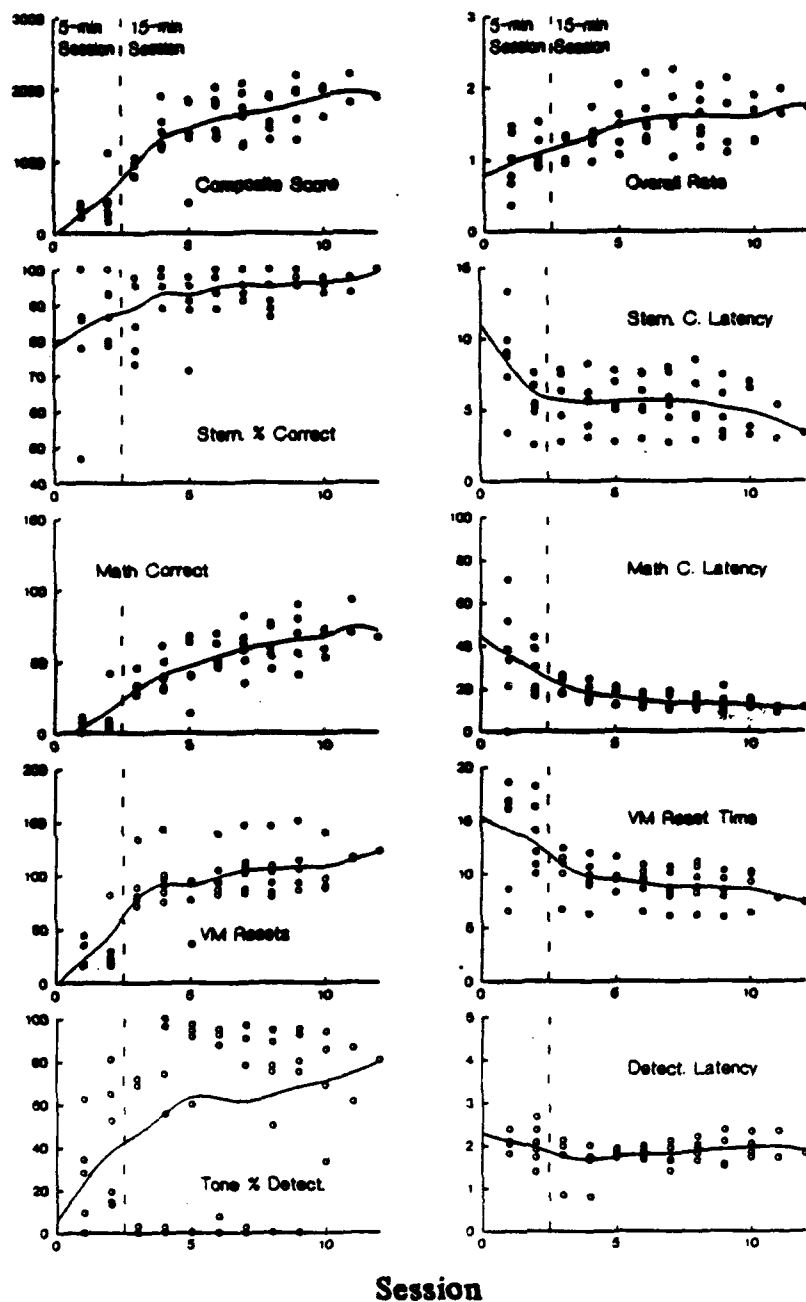
Figure 12. Selected measures from the Tactical Operations Center of the 101st Airborne
Division in the Desert Storm study.

The third row shows performance on the math task. Both groups showed consistent increases in numbers of correct problems, and decreases in the time to complete a correct problem, with completion latency approaching a limit by around the fifth session.

The fourth row shows performance on the visual monitoring task, in which all measures became constant on the fourth session. This constancy probably reflects the extraordinarily simple nature of this task. The reset time settled down at slightly less than 10 seconds. At the speed setting that was used, the pointer reached the end of the scale in 20 seconds, thus the subjects reset the pointer to the center when it had traversed less than half the distance to the end.

The final row shows performance on the auditory detection task. There was a very clear difference between the two groups on this task, reflecting the different testing environments. The performance of the 1st Battalion improved dramatically over the first six sessions, remaining quite stable thereafter. In contrast, performance of the TOC group was highly variable, and much less accurate. Subjects from this group reported that it was too hard to hear the tones, and several of them simply stopped attending to this portion of the task.

Discussion

The data collected in this study demonstrate the feasibility of the synthetic work approach for collecting data in field studies. Data collected in two different situations with two different populations showed a remarkable degree of consistency. The one major difference in performance between the two groups, i.e. performance on the auditory task, was directly attributable to the conditions under which the data were collected.

The acquisition data (approximately sessions 1-5) from both groups was comparable to data obtained from subjects in laboratory studies at the Naval Health Research Center in San Diego (see below). Performance improved rapidly for the first three to five sessions, and more slowly thereafter. In studies of total sleep deprivation, performance on all aspects of SYNWORK begins to deteriorate during the first night of lost sleep. The fact that this did not occur in the present study suggests that, although the subjects were under a heavy operational load, their ability to perform at acceptable levels was unimpaired. In the absence of more appropriate experimental and statistical controls, however, such a conclusion is clearly premature.

The investigators report a degree of subject acceptance of the task that was atypically high for field studies. The game-like quality of SYNWORK1 contributed to a strong motivation on the part of most of the subjects to do well, and improve from session to session. In fact, there was strong competition to earn high scores, particularly among the TOC subjects. This feature of the synthetic work approach suggests that use of performance tasks of this nature in field studies might serve a multiple function. Not only will it maximize

the likelihood that an acceptable sample of behavioral data is collected, it may also serve to assure the physical presence of subjects for other purposes, e.g. interviews or collection of blood and urine samples.

The fact that these data were collected under the harsh conditions present in the desert in Saudi Arabia is a testimony to the hardiness of state-of-the-art computing equipment. The only significant problem that was encountered was the fouling of the mice by the pervasive powder-like sand. An effort should be made to identify alternative pointing devices, e.g. a sealed joystick or optical mouse, that could be used in future deployments to harsh environments.

General Discussion

Comparison of SYNWORK data from different studies

Acquisition of performance on SYNWORK1 was surprisingly similar in all three studies reported in this paper. In Figure 13, to correct for different session lengths, composite scores were converted to points per minute, and the time scale was converted to time on task. Each point represents a group mean. Performance of all three groups was similar for the first 75 min. of exposure to the task, though the Desert Storm subjects were consistently higher than subjects run in the two studies at NHRC. Perhaps this difference is attributable to the generally higher rank (training and education) of the Desert Storm subjects, but this is merely speculation.

Both the Desert Storm and ping subjects continued to improve as long as they were exposed to the task. Performance of the sleep deprivation subjects, however, leveled off and declined as they became sleep deprived. They showed some recovery during the second day, but even with extended exposure to the procedure, sleep-deprived subjects never improved beyond their initial acquisition level. The data shown in Figure 13 tend to strengthen the assertions made in the discussions of both the ping and Desert Storm studies that significant performance degradation did not occur in those studies. Certainly, the level of performance in both of those studies remained substantially higher than that of sleep-deprived subjects with a similar amount of experience with the task.
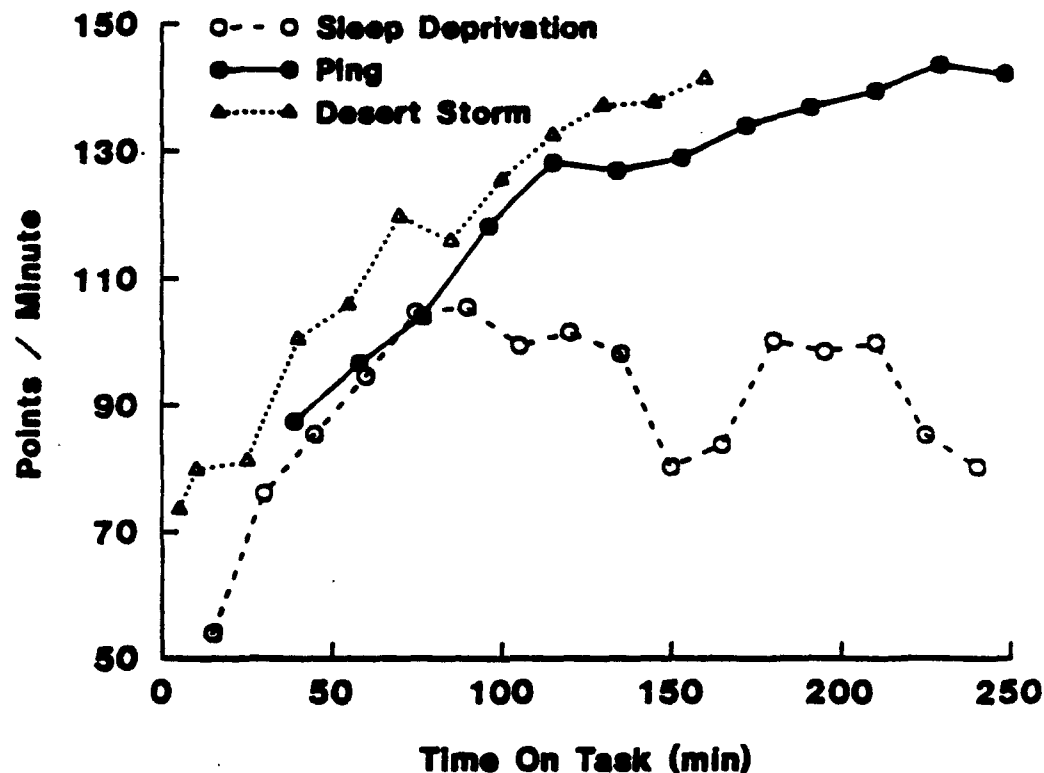
22

Figure 13. Comparison of overall SYNWORK1 performance from the Sleep Deprivation, Ping, and Desert Storm studies.

Relation of PAB'S to SYNWORK: "structural" vs "functional" performance assessment.

In answering the question, "What is the effect of stressor X on performance Y", researchers usually do not have a very firm grip on the "Y". As a result, test batteries typically reflect a "structural" approach. That is, tests are selected to sample a range of behavioral characteristics from which it is assumed complex performances are constructed. Thus it is presun.ed that the results of PAB tests are relevant to a large number of "Y's". As a rule, PAB's yield data showing reliable relationships between stressors and performance. Thus, when "Y" is equated to fundamental behavioral elements, PAB data are sufficient to answer the question, "Does stressor X affect performance Y". When "Y" is equated to an operational task however, a leap of faith is required when asserting that a stressor-induced decrement in performance on a PAB test implies that the same stressor will

23

adversely impact performance, (or conversely, that no decrement on a PAB test implies that the stressor will *not* affect operational performance).[7]

Performance assessment batteries, unlike the real world, typically ignore, or avoid behavioral consequences. In behavioral pharmacology with non-human subjects, it has been repeatedly demonstrated that the effect of a drug depends on the schedule of reinforcement maintaining the behavior (cf. Elsmore, 1976; Galbicka, Fowler, and Ritch, 1991; Hughes and Branch, 1991). A systemic variable such as fatigue might be expected to operate in a similar fashion. In practice, most PAB tests are conducted in extinction, that is, by not providing subjects with feedback regarding the appropriateness of their performance. To the extent that this practice fails to simulate schedules of reinforcement typical of operational tasks, data on stressor effects collected with PAB's may have limited applicability in predicting operational performance. Similarly, PAB's usually do not include elements of multitasking found in operational environments (although dual-task tests are occasionally used). The synthetic work approach attempts to capture some of the complexity of operational tasks by explicitly addressing some of these functional issues.

Response consequences. The awarding of points for correct responses and punishment of errors by point penalties simulates the reinforcement contingencies typical of military tasks (or indeed, non-military tasks as well). Those aspects of the performance on SYNWORK with programmed consequences were under good control, and those aspects without programmed consequences were not. For example, in the Desert Storm study, there was no penalty for clicking on the "RETRIEVE LIST" box in the Sternberg task, and as a result, many of the subjects frequently did so, thus changing the nature of the task from being a memory task, to a simultaneous matching-to-sample task. Similar results have been found in our laboratory studies. In the Ping and sleep deprivation studies, charging a 10-point penalty for looking at the list of letters completely eliminated this pattern of behavior. Penalizing errors elicited emotional outbursts such as, "...it cheated me!". In at least one case in the Desert Storm study, a subject quit because of penalties he thought were unwarranted. These results are totally consistent with what is known of the effects of punishment on

---

[7] One approach to the solution this problem is through the use of computerized sequential network modeling using such tools as Saint, Micro Saint, and Aura. These models simulate the operations of specific systems such as a tank, shipboard defense system, or sonar system, and are developed by psychologists in extensive consultation with system experts. Attempts are made to identify critical points in task performance where memory processes, simple and choice reaction times, etc. come into play. Empirical distributions of data from laboratory studies of stressor effects are then inserted at these points, and Monte Carlo simulations of system operations used to estimate stressor effects on system operation under stress. Unfortunately, few good system models exist, and this technology is not widely available.

behavior. Although no manipulations were made of the relative payoff of the different tasks in the studies reported here, it is likely that the allocation of resources among the tasks would be highly sensitive to such manipulation. Indeed, a large body of experimental literature suggests this would be a most fruitful avenue for future studies (e. g. Catania, A. C., 1963; Hursh, 1980).

Multi-tasking. The relative priorities assigned to the various tasks are most likely a result of the consequences assigned to the various tasks. For example, there were only very loose temporal constraints on the timing of the Sternberg response, which resulted in this task being relatively low priority as shown by an average response time of almost 5 seconds. It is highly likely that the priority subjects assign to this task, or any of the tasks within SYNWORK, could be altered by shifts in the payoff matrix for the overall task.

Savu (1991) demonstrated that there are clear interactions among the tasks within SYNWORK when task difficulty is manipulated. In particular he demonstrated a significant deterioration in performance on the math task when the speed of the visual pointer was increased. Similarly there is no doubt that other manipulations, such as an increase in the difficulty of math problems might have similar effects. Indeed, a large literature exists on the effects of workload on multiple-task performance (e.g. Chiles, 1982). The SYNWORK1 program was designed to permit such manipulations by allowing changes in a variety of task parameters. Thus the program, as constructed, can serve as a vehicle for an extensive program of research on variables controlling performance in a complex multi-tasking environment.

SYNWORK as simulation

In the programming of SYNWORK, there was no intention to create a part simulator for any particular task. In the present study the subjects were instructed to consider the task a simulation of a helicopter flight deck. In fact, in the Desert Storm study, several subjects commented that, functionally, the simulation was not too bad, suggesting that perhaps some of the functional characteristics of this particular task had been isolated. Others have commented that SYNWORK is not a bad analog for other tasks, such as watch-standing in a hospital intensive-care unit. These observations suggest that in the construction of simulations of operational tasks, it may be that the features of contingency and concurrency are at least as critical as the specific nature of the sub-tasks comprising the simulation in making a task realistic.

Conclusion

How then do we answer the question, "What is the effect of stressor X on performance Y?" PAB tests tell us about effects on the elements that comprise "Y". For

25

many purposes, particularly those situations in which the emphasis is on stressor effects on the individual operator, PAB tests may be the method of choice. Through functional tests such as SYNWORK1 we can economically begin to assess the validity of conclusions drawn from PAB-based tests since these assessment procedures begin to add the elements of contingency and concurrency lacking in PAB's. If more system-specific information is required, part-task and high-fidelity system simulators may be required.

References

AGARD (1989). Human performance assessment methods. AMP Working Group 12 and AGARD Lecture series 163, AGARDograph No. 308, NATO.

Alluisi, E. A. (1967). Methodology in the use of synthetic tasks to assess complex performance. *Human Factors*, 9, 375-384.

Anger, W. K. (1990). Worksite behavioral research: Results, sensitive methods, test batteries, and the transition from laboratory data to human health. *NeuroToxicology*, 11, 629-720.

Boff, K. R. and Lincoln, J. E. (1988). *Engineering data compendium: Human perception and performance*. Armstrong Aerospace Medical Research Laboratory, Wright-Patterson AFB, OH.

Brown, F. M. and Graeber, R. C. (1982). *Rhythmic Aspects of Behavior*, Hillsdale, NJ: Laurence Earlbaum Associates.

Catania, A. C. (1963). Concurrent performances: reinforcement interactions and response independence. *Journal of the Experimental Analysis of Behavior*, 6, 253-263.

Chiles, W. D. (1982). Workload, task, and situational factors as modifiers of human performance. In *Human Performance and Productivity: Stress and Performance Effectiveness*, E. A. Alluisi and E. A. Fleischman, Eds. Lawrence Earlbaum Associates: Hillsdale, NJ, Pp. 11-29.

Chiles, W. D., Alluisi, E. A., and Adams, O. S. (1968). Work schedules and performance during confinement. *Human Factors*, 10, 143-196.

DeJoy, D. M. (1984). The nonauditory effects of noise: Review and perspectives for research. *Journal of Auditory Research*, 24, 123-155.

Elsmore, T. F. (1976). The role of reinforcement loss in tolerance to chronic delta-9-tetrahydrocannabinol effects on operant behavior of rhesus monkeys. *Pharmacology, Biochemistry, and Behavior*, 5, 123-128.

Englund, C. E., Reeves, D. L. Shingledecker, C. A., Thorne, D. R., Wilson, K. P., and Hegge, F. W. (1985). *The Unified Tri-Service Cognitive Performance Assessment Battery (UTC-PAB): Design and Specification of the Battery*. (JWGD[3] MILPERF Report No. 85-1) U. S. Army Medical Research and Development Command, Fort Detrick, MD.

Galbicka, G, Fowler, K. P., and Ritch, Z. (1991). Control over response number by a targeted percentile schedule: Reinforcement loss and the acute effects of d-amphetamine. *Journal of the Experimental Analysis of Behavior*, 56, 205-215.

Graves, R. E. and Bradley, R. (1988). More on millisecond timing and tachistoscope applications for the IBM PC. *Behavior Research Methods, Instruments & Computers*, 20, 408-412.

Graves, R. E. and Bradley, R. (1991). Millisecond timing on the IBM PC/XT/AT PS/2: A review of the options and corrections for the Graves and Bradley algorithm. *Behavior Research Methods, Instruments & Computers*, 23, 377-379.

Hamlin, Thomas III (1990). Advanced function library for the BASIC compiler.

Hartman, B. O., Benes, R. A., and Storm, W. F. (1980). *A review of USAFSAM studies employing multiple-task performance devices*. (Report SAM-TR-80-16) USAF School of Aerospace Medicine, Brooks Air Force Base, TX.

Hegge, F. W. (1991). Performance assessment and risk analysis. *Proceedings of the 1991 Medical Defense Bioscience Review*, U. S. Army Medical Research Institute of Chemical Defense: Aberdeen Proving Ground, MD, 329-337.

Hughes, C. E., and Branch, M. N. (1991). Tolerance to and residual effects of cocaine in squirrel monkeys depend on reinforcement-schedule parameter. *Journal of the Experimental Analysis of Behavior*, 56, 345-360.

Hursh, S. R. (1980). Economic concepts for the analysis of behavior. *Journal of the Experimental Analysis of Behavior*, 34, 219-238.

Kennedy, R. S., Bittner, A. C., Carter, R. C., Krause, M., Harbeson, M. M., McCafferty, D. B., Pepper, R. L., and Wilker, S. F. (1981). *Performance evaluation tests for environmental research (PETER): Collected papers*. (NBDL - 80R008) Naval Biodynamics Laboratory), New Orleans, LA.

Mclain, D. H. (1974). Drawing contours from arbitrary data points. *The Computer Journal*, 17, 318-324.

Perez, W. A., Masline, P. J., Ramsey, E. G., and Urban, K. E. (1987). *Unified Tri-Services Cognitive Performance Assessment Battery: Review and Methodology*. (AAMRL-TR-87-007) Armstrong Aerospace Medical Research Laboratory, Dayton, OH.

Savu, T. G. (1991). A synthetic work environment experiment. Unpublished master's thesis, University of Oklahoma.

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATE COVERED

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
Performance assessment in sustained operations using a computer-based synthetic work task	Program Element: Work Unit Number: 63706N M0096.002-6201 62233N MM3P30.002-6005

6. AUTHOR(S)
Timothy F. Elsmore, Paul Naitoh, Steven Linnville

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION Report No. **92-30**
Naval Health Research Center P. O. Box 85122 San Diego, CA 92186-5122	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER
Naval Medical Research and Development Command National Naval Medical Center Building 1, Tower 2 Bethesda, MD 20889-5044	

11. SUPPLEMENTARY NOTES

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

13. ABSTRACT (Maximum 200 words)

Prediction of the effects of stressors on military performance from laboratory tests frequently involves use of procedures simulating critical functional aspects of military jobs. This paper presents a computerized "Synthetic Work" test, SYNWORK1, as a prototypical alternative to traditional Performance Assessment Batteries. This test simulates complex operational environments by requiring subjects to work on four simultaneous tasks, and assigns explicit consequences for work on each component task. As illustrations of the utility of this technique, data are presented from three applications of SYNWORK1, a laboratory study of sleep deprivation, a test of extended exposure to a simulated sonar "ping", and the use of the same task during army operations in Operation Desert Storm.

14. SUBJECT TERMS			15. NUMBER OF PAGES 30
Performance, cognitive function, sustained operations, sleep deprivation, noise, simulation			16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited