AFHRL-TR-83-49

AD A140469

(12)

AIR FORCE HUMAN RESOURCES

PHYSIOLOGICAL ASSESSMENT OF AIRCRAFT PILOT WORKLOAD IN SIMULATED LANDING AND SIMULATED HOSTILE THREAT ENVIRONMENTS

By

Ernest Lindholm
Cary Cheatham
John Koriath

Arizona State University
Department of Psychology
Tempe, Arizona 85287

Thomas M. Longridge

OPERATIONS TRAINING DIVISION
Williams Air Force Base, Arizona 85224

April 1984
Final Report

LABORATORY

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

MILTON E. WOOD, Technical Director
Operations Training Division

ALFRED A. BOYD, JR., Colonel, USAF
Commander

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER AFHRL-TR-83-49	2. GOVT ACCESSION NO. AD-A140469	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) PHYSIOLOGICAL ASSESSMENT OF AIRCRAFT PILOT WORKLOAD IN SIMULATED LANDING AND SIMULATED HOSTILE THREAT ENVIRONMENTS		5. TYPE OF REPORT & PERIOD COVERED Final 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Ernest Lindholm, John Koriath Cary Cheatham, Thomas M. Longridge		8. CONTRACT OR GRANT NUMBER(s) F33615-80-C-0020
9. PERFORMING ORGANIZATION NAME AND ADDRESS Arizona State University Department of Psychology Tempe, Arizona 85287		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2313T315
11. CONTROLLING OFFICE NAME AND ADDRESS HQ Air Force Human Resources Laboratory (AFSC) Brooks Air Force Base, Texas 78235		12. REPORT DATE April 1984 13. NUMBER OF PAGES 44
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Operations Training Division Air Force Human Resources Laboratory Williams Air Force Base, Arizona 85224		15. SECURITY CLASS (of this report) Unclassified 15.a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of this abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

event-related potentials
flight simulation
heart rate
information processing
physiological measures
respiration
skin conductance
workload

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

In two experiments, physiological metrics of cockpit workload were investigated in highly realistic flight simulators. In Experiment 1, non-pilot males were trained on a simulated landing task and a secondary, tone discrimination task while heart rate, skin conductance, and brain event-related potentials were continuously quantified. The results showed that heart rate was a more stable measure of workload than was skin conductance. Heart rate increased during each final approach to landing, and mean heart rate decreased as the subjects gained mastery over the task as a function of practice. Four ERP components (N1, P2, N2, P3) were statistically evaluated. As workload increased, N2 became more negative and P3 became less positive; also, as workload increased, the

DD Form 1473 EDITION OF 1 NOV 65 IS OBSOLETE
1 Jan 73

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Item 20 (*Continued*)

latency difference between P3 and N1 increased. Finally, a within-subject regression analysis was employed to express the extent to which the four ERP components were intercorrelated. This measure proved to have considerable power to predict how well individual subjects would perform on the landing tasks. In Experiment 2, rated male pilots flew a simulated mission involving threat by surface-to-air missiles (SAMs). Heart rate, respiration activity, and ERPs were quantified by means of a custom-designed, miniaturized recording system. The pilots were informed of the level of SAM threat by tones sounded in the headset. The results showed that heart rate and respiration activity increased as SAM threat increased. The ERP analysis showed that N2 and P3 amplitude and P3 latency increased with threat level. The autonomic results are discussed within the framework of activation theory, and, regarding the ERP results, it is suggested that N2 might be more important for workload and information processing studies than is P3.

Dist | S
A-1

SUMMARY

Objectives

The principal objective was to assess pilot workload in complex flight environments which closely approximate real-world situations through use of non-intrusive, physiological metrics. A secondary objective was to develop a reliable, miniaturized physiological recording system for obtaining these metrics.

Background

Traditional approaches to workload assessment (e.g., dual-task methodology) are referred to as "intrusive" techniques because the act of measuring workload through a secondary task intrudes upon the operator's ability to perform the primary task. This has prompted interest in physiological measures of workload because they are non-intrusive and provide measures of workload based on the internal state of the operator.

Approach

Heart rate was suggested as a workload metric nearly two decades ago, but its use for that purpose has met with mixed research results. More recently, attempts have been made to relate changes in the brain event related potential (ERP) to workload, but the tasks lacked realism and were not directly relevant to pilot training. The current effort evaluates both autonomic and central nervous system metrics of workload in tasks that are directly relevant to pilot training and performance.

Specifics

In Experiment 1, 12 male subjects (non-pilots) were trained on a computer-simulated aircraft landing task. A simple tone discrimination task (secondary task) was performed alone and in combination with the landing task. Heart rate, skin conductance responses, ERPs, and eye movements were continuously monitored. Since eye movements, including blinks, can contaminate ERP records, the eye movement variable was a necessary control for interpretation of ERP data. In addition to the autonomic variables, four major components of the ERP were subjected to analysis. Experiment 2 utilized the Advanced Simulator for Pilot Training; 20 rated pilots flew simulated attack missions while exposed to surface-to-air missiles (SAMs). The mission goal was to bomb a target protected by SAMs. Throughout each mission, the pilot received tones ever the headset which signaled the level of SAM threat. Heart rate, respiration activity, and ERPs were continuously recorded by a custom-designed minature telemetry device. With these experiments, it was possible to relate both autonomic and central nervous system response to workload in two quite different tasks.

The results for Experiment 1 showed that heart rate was a more reliable autonomic indicator of workload than was skin conductance and that three of the four ERP components were related to workload. These results extend previously published results in the following ways: (a) heart rate is a reliable metric of workload and can detect decreases in workload as a function of practice, as well as momentary increases in workload as typified by final approach to landing, and (b) several components of the ERP seem sensitive to workload changes; thus, it appears unwise for researchers to focus on just one component such as the P3. Of particular interest were the results of a within-subject regression analysis that, in effect, expressed the extent to which the various components of the ERP were intercorrelated. This measure showed considerable power in predicting an individual's performance on the landing task.

The results of Experiment 2 showed that both heart rate and respiration activity increased in an orderly fashion as the level of SAM threat increased. As the Experiment 1, ERP components other than P3 were reliably related to workload, strengthening the suggestion that future research should not be limited to P3.

Conclusions/Recommendations

From the results of the two experiments, it is concluded that non-intrusive, physiological measures can be used to assess workload changes in highly realistic, simulated aircraft environments. Of the autonomic variables explored, heart rate yielded the most consistent results and, because of its ease of measurement, should be considered the variable of choice. In general, the ERP waveform is complex and appears to contain more information relevant to workload experiments than had been suggested by earlier reports. For example, results show that the earlier N2 component might represent a brain process more primary than P3. Finally, a miniature physiological recording system was developed that proved to be remarkably free of movement artifact and could be easily adapted for use in an actual aircraft.

PREFACE

The basic research reported herein was conducted in support of the Aircrew Training Thrust of the Air Force Human Resources Laboratory at Williams AFB, Arizona.

The authors wish to express their gratitude to the staff of the Air Force Human Resources Laboratory for their technical support of the simulator research and, in particular, to Mr. Brett Butler for his assistance in the handling of the data tapes. Special thanks are extended to Mr. Norwood Sisson, Department of Psychology, Arizona State University, for designing the interface associated with the biotelemetry system. Captain Dan Boone at the USAF School of Aerospace Medicine served as co-monitor for this effort.

TABLE OF CONTENTS

INTRODUCTION

Although several methods for assessing operator workload have been developed over the years (Wierwille and Williges, 1978) dual-task methodology appears to be the most widely used. A problem with this approach is that it is, by its nature, intrusive. That is, the demands of the secondary task intrude upon, or interfere with, the operator's ability to perform the primary task (Rolfe, 1971). A related problem is that multiple tasks require multiple motor responses, so it is frequently not clear whether a performance decrement is due to high mental workload or to response interference (McLeod, 1978). Considerations such as these have stimulated research on measures that are more non-obtrusive and non-interfering so that the act of measurement does not disrupt the performance being evaluated (Wickens, 1979).

Physiological measures have the advantage of being non-intrusive; no responses other than those normally emitted by the operator are required. Workload is quantified, not in terms of secondary or primary task performance decrements, but in terms of autonomic and central nervous system responses that reflect variations in physiological function introduced by variations in workload.

The physiological approach is not particularly new. It has been known for some time that the heart rate of experienced pilots flying either commercial or military aircraft tends to peak during takeoffs and landings (Nicholason, Hill, Borland, & Krzanowski, 1973; Roman, Older, & Jones, 1967; Ruffel Smith, 1967). The heart rate increases cannot be due to "anxiety" associated with these relatively high risk flight segments since similar heart rate increases are observed during a computer simulated landing in which there is no threat to life or property (Lindholm & Cheatham, 1983).

Another physiological measure of potential importance for workload assessment is the event-related potential (ERP) recorded from the surface of the scalp. Just as heart rate reflects autonomic nervous system functioning, the ERP reflects central nervous system functioning. The ERP is a series of voltage oscillations in the brain wave immediately following the presentation of a stimulus to any sensory modality. The voltage oscillations (called "components") are usually grouped into "early" components (those occurring within about 250 msec of stimulus onset) and "late" components (those occurring between 250 and 500 msec following stimulus onset). One late component that has attracted considerable attention is the P3, or P300, which is a positive oscillation occurring about 300 to 450 msec following stimulus onset. P3 amplitude and/or latency has been related to several psychological variables including stimulus relevance, subjective probability, and decision-making (Pritchard, 1981). Recent evidence demonstrates that the P3 elicited by tone stimuli is reduced in size when subjects are placed in a dual-task situation: Isreal, Chesney, Wickens, and Donchin (1980) required subjects to count target tones covertly, and P3 was quantified. When a visual tracking task was combined with the tone task, the magnitude of P3 decreased. Similar results were reported by Isreal, Wickens, Chesney, and Donchin (1980) when the tone-counting task was combined with a computer-simulated air traffic control display. The authors suggest that the reduced P3 is a reflection of increased sensory processing load placed on the central nervous system during dual-task conditions.

The experiments reported here represent an extension of earlier efforts to apply psychophysiological techniques to the study of pilot workload. These efforts differ from others in two important respects: (a) the tasks used are highly relevant to pilot training and performance and (b) several physiological responses are quantified in continuous fashion so that the internal state of the pilot can be described in considerable detail.

In Experiment 1, conducted in the Arizona State University (ASU) Laboratory, the major task was a computer simulation of a Navy A-7 aircraft landing on a carrier; a simple tone discrimination task performed alone and in combination with the carrier landing task was used to examine ERP changes. The results showed that ERP components other than P3 are sensitive to workload fluctuations and that heart rate changes also track workload changes. A new and potentially very important relationship was discovered which related intercorrelations of ERP components to carrier landing task performance in a predictive manner.

Experiment 2 utilized the Advanced Simulator for Pilot Training (ASPT) located at Williams AFB. Rated pilots of differing experience levels flew a simulated hostile threat mission with the goal of bombing a target and avoiding threats represented by surface-to-air missiles (SAMs). Level of threat was signalled to the pilot by presenting different frequency tones to indicate the four threat levels of "safe" (no SAM threat), "acquisition," "track," and "launch" (highest SAM threat). The results showed that heart rate and respiration rate increased as a function of threat level. Analysis of the ERPs evoked by the tones showed, as in Experiment 1, that components other than P3 are sensitive indicators of workload variations.

EXPERIMENT 1

METHOD

Subjects

The 12 males who volunteered to participate in this study were scheduled to begin Undergraduate Pilot Training at Williams AFB a few weeks after this experiment was conducted. All had some flight experience in single reciprocating engine aircraft, but none had piloted jet aircraft.

Tasks

Tone Discrimination Task- A Digital Equipment Corporation PDF 11/34a computer was programmed so that the digital-to-analog outputs drove a voltage controlled oscillator (EXACT Model 126). Tones were presented binaurally through Sennheiser Model HD 400 earphones. Tone duration was 200 msec at 65 dB.

A "run" on this task was defined as follows. A reference tone of 1500 Hz was presented 10 times at a repetition rate of once per 3 seconds. Thirty seconds later, 24 comparison tones were presented at a rate of one per 5 seconds. The 24 comparison tones consisted of six repetitions of four tones (1000, 1250, 1750, and 2000 Hz) which were block randomized. Subjects were instructed to respond to tones higher than the reference tone by saying the word "tone" into a microphone. The microphone was held in a brace worn around the subject's neck and was adjusted to within 5 cm of the lips. In pilot work with this task, subjects were instructed to respond to high tones on some runs and low tones on other runs. This response set manipulation had no demonstrable effect on the ERP, reaction times, or error rates; thus, it was omitted in this experiment.

Each subject received five runs; reaction times were measured to the nearest 4 msec by computer software and a hardware clock. Subjects were reminded frequently that they should respond as quickly as possible but without error. Errors were scored if the subject responded to a tone lower than the reference tone (error of commission) or failed to respond to the higher tones within 1500 msec of stimulus onset (error of omission). The pitch of two of the tones (1250 and 1750 Hz) was close to that of the reference tone of 1500 Hz; these conditions were defined as the "hard" discrimination. The other two comparison tones (1000 and 2000 Hz) were further in pitch from the reference; these conditions were defined as the "easy" discrimination.

Carrier Landing Task- The PDP 11/34a computer was used to generate display images on a Digital Equipment Corporation model VT-11 video graphics device with a screen size of 33 cm by 25 cm. The display was an out-the-window simulation of a Navy A-7 aircraft landing on a carrier deck. The lower one-third of the display consisted of altimeter, vertical speed indicator, rad.. compass, distance to the carrier in nautical miles and tenths of nautical miles, and percent engine power. The upper two-thirds of the display consisted of a simulated true horizon and the aircraft carrier complete with rudimentary superstructure, well defined landing area, the carrier wake, and the Fresnel Optical Landing System (FOLS, or "meatball"). The entire display changed in real time in response to movements of joystick and throttle with a refresh and update rate of 30 Hz. This software was originally developed by the Navy to provide adjunct training for carrier pilots and has undergone several modifications to permit on-line quantification of physiological variables.

The problem was programmed as follows: The aircraft was released from freeze at 3.8 nautical miles (7.04 km) from the carrier at an altitude of 1550 feet (4.7 km) above sea level. The heading was ideal for a straight-in approach. The throttle was frozen at 87% power to simulate the power-on approach typical of carrier landings and to equate approach velocity for all subjects. This power setting produced a nominal airspeed of 120 knots with full flaps. With these parameters, flight duration from freeze release to carrier deck was 120 seconds for a straight-in approach and was longer if the subject failed to maintain the proper heading.

Each flight could terminate in one of six ways, listed here in order from poorest to best performance:

1. Splash. Aircraft reached 0 feet altitude and impacted with water.
2. Time-Out. Subject lost orientation and could not reach the carrier deck within 150 seconds.
3. Ramp Strike. Aircraft struck stern of carrier below landing area.
4. Crash. Aircraft contacted carrier deck while in a state of excessive roll (greater than 10 degrees) or excessive vertical speed (greater than -1000 feet per minute (-3.05 km per minute).
5. Bolter. Aircraft contacted carrier deck but attitude was incorrect causing a bounce and subsequent miss of tail-hook cables.
6. Landing. Aircraft contacted carrier deck in designated landing area with roll less that 10 degrees and vertical speed between 0 and -2.3 km per minute.

Each subject received a "flight termination" score from 1 to 6 for each flight according to the preceding scheme. Additionally, a flight "approach score" was calculated for each subject and each flight in the following manner. Each 5 seconds during the flight, the computer sampled and stored the instantaneous values of vertical speed, roll, and heading. RMS deviations from ideal values were calculated: ideal roll was zero; ideal heading was the starting heading; and ideal vertical speed was -2.3 km per minute calculated from a knowledge of starting altitude and distance, height of the carrier deck, and sink dynamics of the A-7 with full flaps. Finally, the RMS scores were subtracted from 100 so that the subject could be told simply that a score of 100 is perfect. For each flight, means were calculated on-line so that the subject could be given immediate feedback; e.g., "You scored 60 on roll, 80.3 on heading, and 70.7 on vertical speed so your average approach score was 70.3 for that last flight."

Subjects were given instructions to help them perform well. They were cautioned to watch the altimeter to avoid splash, avoid large stick movements, keep the carrier visually lined up at all times, and pay close attention to the FOLS, or meatball. The latter is simulated on the carrier deck just to the left of the landing area. It is displayed as a horizontal line and ball. If the ball is above the line, the approach is high, and the stick should be moved forward. If the ball is below the line, the approach is low, and the stick should be pulled back. The meatball is too small to be resolved easily at the 3.8-mile starting distance, but becomes very clear during the last half of the flight when all aspects of the carrier are perceptually much larger.

Physiological Recording

ERPs were recorded from the vertex (Cz in International 10-20 system) referenced to the right mastoid. The left mastoid served as ground. Eye movements and blinks were monitored by placing electrodes on the lateral canthus and superior ridge of the left eye. Heart rate (actually quantified as inter-beat-interal, or IBI, in msec) leads were placed on the left lateral rib cage and the sternum. Skin conductance leads were placed on the middle finger of the left hand, referenced to the back of the same hand. Beckman silver-silver chloride elctrodes were used for all placements. The vertex lead was held in place with Grass electrode paste and a gauze sponge; all other leads utilized Beckman double adhesive collars and Beckman electrode cream. Electrode impedance (measured at 30 Hz) was less than 5 k-ohms for the vertex and mastoids, and less than 30 k-ohms for all other leads.

Potentials were led to a Beckman Type 611 Dynograph with bandpass of .16 to 30 Hz for the vertex and eye movement channels, 5.3 to 30 Hz for the heart channel, and DC to 30 Hz for the skin conductance channel. The high level outputs of the dynograph served as inputs to PDP 11/34a analog-to-digital converters. Ocular activity and ERPs were recorded to each of the tones, starting 500 msec before tone onset and continuing for 1500 msec thereafter. Sample rate was 250 per second. The skin conductance channel was sampled four times per second, and IBI was measured on-line by a machine language routine that detected R-waves. The subject wore a lightweight junction box around the neck, and this served to connect the primary leads to the Dynograph. In this manner, subjects could disconnect from the Dynograph and walk around freely during breaks.

Procedure

All subjects received the tone discrimination task first followed by a 10-minute break, then the carrier landing task followed by a 10-minute break, and finally the combined tasks. There were five runs of the tone task alone and 10 runs or "flights" of the carrier task alone. The tones were presented during this phase, but subjects were told not to respond to the tones. Their task was to fly the airplane as best they could, and the tones could be totally ignored. In the combined task condition, there were 10 more flights and subjects were required to fly the airplane and respond to the tones simultaneously.

Note that each flight is 120 seconds in duration and the 24 tone discrimination trials spaced 5 seconds apart also occupy 120 seconds. Thus, in both the carrier alone and combined task conditions, 24 ERPs are elicited during each flight. The only difference is that subjects do not respond to the tones during the carrier-alone condition.

EXPERIMENT 1

RESULTS AND DISCUSSION

Statistical tests included analysis of variance (ANOVA) and step-wise regression, implemented through the BMDP statistical package.

Tone Task Performed Alone

Since the tone task required only simple psychophysical judgments and the ability to make a simple response, practice effects were not expected. Indeed, the only expected effect was that the "hard" discrimination (1500-Hz reference versus 1750-Hz target) should require more processing time than the "easy" discrimination (1500-Hz reference versus 2000 Hz-target). These expectations were confirmed by a repeated measures ANOVA with two levels of trial blocks (first 12 and last 12), two levels of discrimination difficulty (easy and hard), and five levels of runs (the five runs). The grand mean RT was 809 msec for the hard discrimination and 715 msec for the easy discrimination ($p<0.002$). None of the other main effects or interactions was statistically significant. Errors of omission or commission occurred so rarely that statistical tests were not attempted.

Carrier Task Performed Alone

This task is difficult for persons without jet aircraft experience due, presumably, to the high performance characteristics of the A-7. The most common error during early flights was gross and violent stick movements that resulted in total loss of aircraft control and an early splash. However, subjects quickly learned to inhibit these maladaptive responses and gained control over the aircraft. The performance changes as a function of practice (flights) are summarized in the top portion of Table 1. The increase in performance (approach score) from first to second halves of training was significant ($p<0.03$). The remainder of Table 1 summarizes the ways in which flights terminated as a function of practice. The percentage of disastrous terminations (splashes, ramp strikes, and crashes) decreased with practice ($p<0.05$) while the percentage of landings and bolters increased with practice ($p<0.01$).

Taken together, the data in Table 1 and the associated ANOVAs simply make the point that performance on the carrier-landing task improved with practice; approach scores increased, disasters were less frequent, and bolters and landings became more common as a function of practice.

Combined Task (Carrier landing and tone tasks performed together)

When the tasks were combined, subjects maintained high performance on the carrier landing task as shown in Table 2. In fact, there was significant improvement in the flight approach score, relative to the last five flights of the carrier-alone condition ($p<0.005$). The percentages of disasters and of bolters and landings showed no significant changes relative to the last five flights of the carrier-alone condition. However, subjects tended to treat the tone discrimination task as a secondary, low priority task. RTs and discrimination errors increased in the combined task condition relative to the tone-alone condition, and the increase was greater as the subject flew closer to the carrier landing area ($p<0.01$). This effect has been reported previously (Lindholm, Cheatham, Longridge, & Buckland, 1982) using the same combination of tone task and landing task, and it is also in agreement with anecdotal reports of instructor pilots who claim that student pilots tend to ignore auditory messages when they are practicing a new maneuver in the aircraft.

ERP Analysis

ERP variables of interest were N1 latency (N1L), P2 latency (P2L), N1P2 amplitude (trough to peak), N2 latency (N2L), P3 latency (P3L), N2P3 amplitude (trough to peak), and P3-N1 latency shift (the subtracted difference between P3L and N1L).

Table 1. Approach scores and terminations by category for the carrier landing task performed alone.

	First 5 Flights	Last 5 Flights
Approach Score	36.1	52.1
Percentage of:		
splashes	32.0	20.0
ramp strike & crash	25.0	22.0
bolters	36.0	42.0
landings	7.0	17.0

Table 2. Approach scores and terminations by category for the combined task condition.

	First 5 Flights	Last 5 Flights
Approach Score	61.1	60.8
Percentage of:		
splashes	7.0	12.0
ramp strike & crash	20.0	15.0
bolters	55.0	61.0
landings	18.0	12.0

A software routine compared each ERP record with the corresponding eye movement record and discarded trials when eye movements produced artifacts in the ERP. Approximately 20% of all trials were discarded, distributed equally across conditions. For all remaining records, the latencies and amplitudes of the four prominent components were measured by a software routine (Cheatham and Lindholm, unpublished) which accepts, as input parameters, the grand mean of N1 and P3 latencies. The routine then defines N1 and P3 as the highest amplitude components closest to these means on each trial, and the P2 and N2 as the intervening peak and trough, respectively. Amplitude differences of N1P2 and N2P3 were calculated by simple algebraic subtraction. Validation of the software routine for single trial data was accomplished by having two members of the ASU laboratory staff independently identify the four components by visual inspection of hundreds of trials over a 2-month period, and these judgments were compared with the determinations made by the software routine. Agreement was above 90%.

Among-subject analyses utilized repeated measures ANOVA (BMDP P2V) to test for differences in the ERP components as a function of conditions. Although earlier reports have related only changes in P3 amplitude to variations in task workload, the present results show a more complete picture of workload effects on the brain wave. For example, Figure 1 shows that N2 amplitude is least negative in the tone-alone condition, most negative in the carrier-alone condition, and intermediate for the combined-task condition. The overall change in N2 as a function of condition was substantial ($F(2/22)=18.3$, $p<0.001$), as was the quadratic trend ($F(1/11)=18.1$, $p<0.002$). A similar effect was seen for P3 amplitude as shown in Figure 2. P3 amplitude was most positive during the tone-alone condition, became slightly negative during the carrier alone condition, and was slightly positive for the combined task condition. Again, both the overall effect and the quadratic trend were significant ($F(2/22)=19.2$, $p<0.001$; $F(1/11)=15.1$, $p<0.003$, respectively). Thus, the introduction of the high workload carrier landing task moved both the N2 and the P3 in a more negative direction. The effect, however, was not monotonic across conditions. There is reason to believe that the combined-task condition failed to represent a higher workload than the carrier-alone condition. As mentioned earlier, in the combined-task condition, subjects maintained performance on the carrier-landing task and allowed performance on the tone task to degrade. Also, they were well practiced on the carrier task by the time the combined-task condition was introduced. With this in mind, it becomes plausible to argue that lowest workload was represented by the tone-alone condition, highest workload was represented by the carrier-alone condition (since they were unfamiliar with the task), and the combined-task condition represented an intermediate level of workload.

These results suggest a reinterpretation of the earlier data of Isreal, Chesney, Wickens, and Donchin (1980), and Isreal, Wickens, Chesney, and Donchin (1980). They reported that P3 decreased in amplitude when visual information processing workload was increased. While the present results replicate their findings, it is apparent that the N2, which precedes P3 by about 100 msec, is at least as sensitive to visual workload increments as is P3. The N2 results reported here suggest that N2 might represent an earlier and more important brain process than P3;

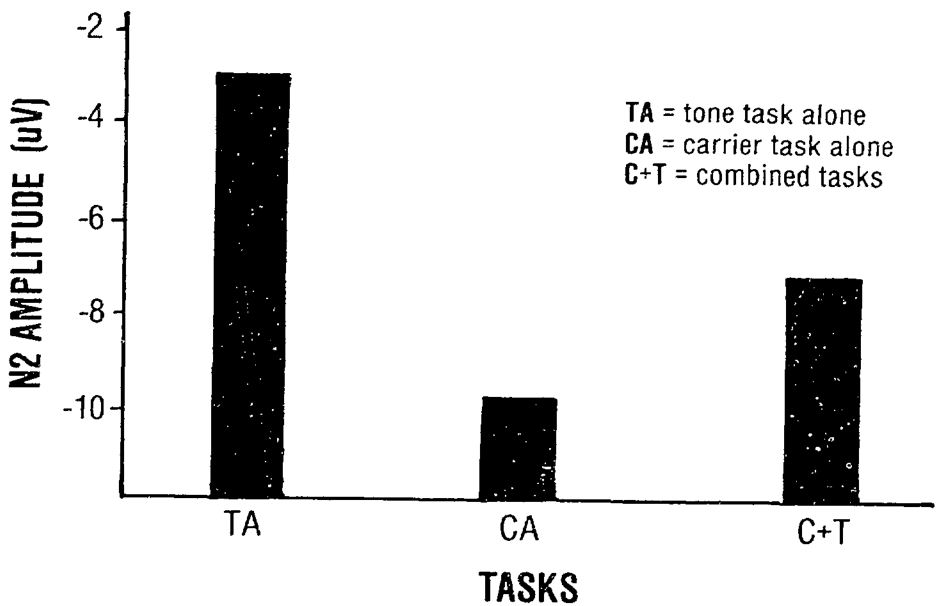

Figure 1 — N2 amplitude as a function of conditions.

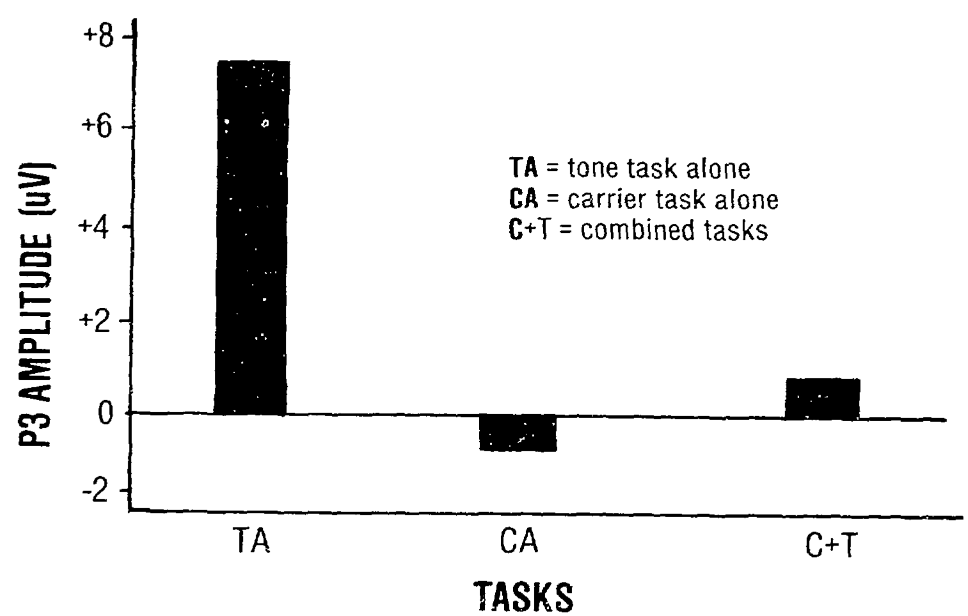

Figure 2 — P3 amplitude as a function of *conditions.*

indeed, P3 might be dependent on N2. A similar suggestion has been made by Ritter (1978).

Another line of evidence indicates that P3 should be considered as just one component and not uniquely important. The analyses showed that N1 tended to occur earlier, and P3 tended to occur later when the carrier landing task was introduced. Neither latency difference was significant by itself, but the pattern of results suggested the obvious metric of a P3-N1 latency difference; that is, a derived measure formed by subtracting N1L from P3L for each tone trial, producing a new measure, "P3-N1 Shift." The relationship of P3-N1 shift to conditions is shown in Figure 3. Again, the function is U-shaped, but in this case, not significantly so. The overall increase in P3-N1 shift across conditions was significant ($F(2/22)=4.6$, p 0.03), but the quadratic trend was not significant.

Figure 4 summarizes the changes seen in the brain wave when subjects perform under low workload (tone-alone) and high workload (carrier-alone) conditions. It is important to remember that subjects still hear the tones during the carrier-alone conditions but are told to ignore the tones. In effect, the tones become a "probe" stimulus to examine the brain's response to the tones when the tones are defined as task-irrelevant. Figure 4 is a composite tracing from several overlays and does not represent the data for any one subject. It is offered in an attempt to summarize the ANOVA results. One major point is that when workload changes from low (simple tone-discrimination task) to high (introduction of carrier-landing task), the latency of N1 shortens and the latency of P3 lengthens. This is the P3-N1 shift. Also, with high workload, late component amplitudes become more negative (less positive), and obviously, the change in N2 precedes the change in P3.

Relationships among ERP components (within subject effects)

Since the ASU laboratory measures four major components of ERP simultaneously, it becomes possible to investigate sequential dependencies that might exist between the latencies and amplitudes of N1, P2, N2, and P3. Examination of many bivariate correlation matrices (Pearson "r") suggested that the intercorrelations are very complex and the intercorrelations show different patterns for different subjects. Thus, multiple regression techniques were employed so the intercorrelations could be reduced to a single number for each subject. Specifically, for each subject, N1, P2, N2, and P3 latencies, and the amplitudes of N1 and P2 were regressed on late component amplitude (N2-P3 trough-peak). This criterion variable was chosen because earlier literature (e.g., Pritchard, 1981; Isreal et al, 1980, 1980) emphasized the importance of late component amplitude variations as measures of information processing demands and workload. This derived measure is referred to as "late component correlation" (LCC) since the R-squared from the regression analysis expressed the extent to which the latencies and amplitudes of ERP components correlate with late component amplitude. LCC scores were calculated for each of the 12 subjects for the data from tone-task alone, carrier-task alone, and combined-tasks conditions. The LCC score for each subject was then treated as a subject variable and

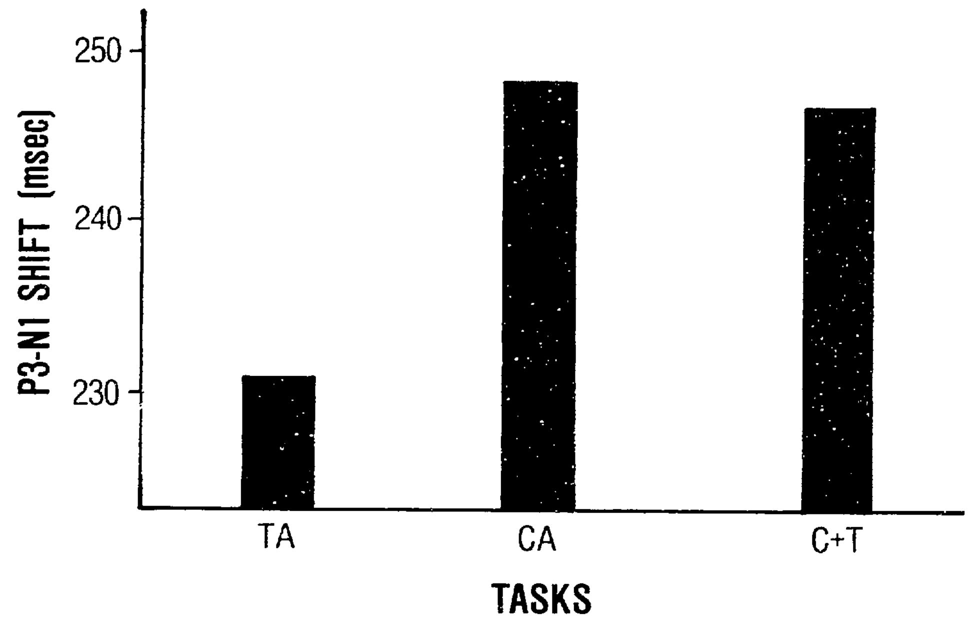

Figure 3 — P3-N1 Shift as a function of conditions. TA=tone alone, CA=carrier landing task alone, C+T = combined tasks.

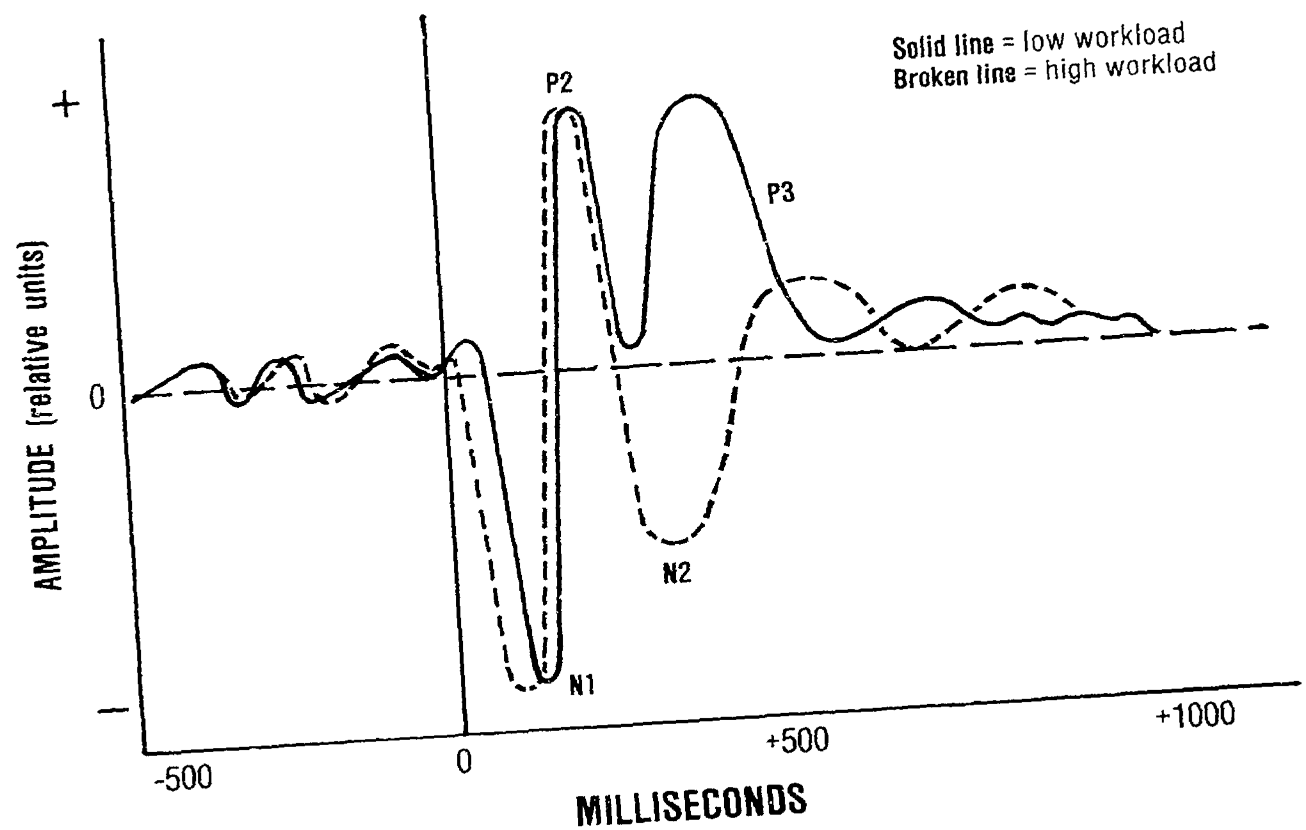

Figure 4 — ERP Waveforms as a Function of Workload

correlated with performance on the tone and carrier-landing tasks. LCC was not significantly correlated with tone-task performance, nor with carrier-landing performance in the combined task condition. However, of considerable interest was the finding that LCC calculated from the tone-task-alone data was related to performance on the carrier-landing task performed alone. The relationship was linear and quite strong, indicating that LCC was capable of predicting subsequent performance on the carrier-landing task. Obviously, there is a danger that these results are peculiar to the particular sample of 12 subjects. Therefore, some older data on five entirely different subjects who performed the same tasks were re-analyzed with the goal of replicating the LCC relationships. The combined results are shown in Figure 5. The "B" subjects on the scatter-plot represent the current 12 subjects, and the "A" subjects are the five from an earlier study. Note that the A and B subjects are intermixed, indicating that the effect is not peculiar to a particular group of subjects.

Interpretation of LCC

The LCC scores used in the Figure 5 analysis were calculated for each subject across all "good" trials (120 minus trials with ERP artifacts). Thus, LCC represents the extent to which the shape and form of the brain wave were similar from trial to trial. As such, it can be viewed as a measure of consistent, time-dependent relationships among late component amplitude, late component latency, and early component latency and amplitude. Exactly why this measure should predict performance on the subsequently administered carrier landing task is not obvious. Perhaps LCC is a measure of information processing efficiency; certainly, LCC appears to be a true individual difference variable since a given subject's LCC score changed little across the tone-alone, carrier-alone, and combined-task conditions. Other experiments are currently being designed to test the relationship between LCC and performance on tasks other than the carrier-landing task.

It is possible that LCC is related to the "string measure" discussed by Blinkhorn and Hendrickson (1982). These authors presented a series of tones to subjects who were not required to respond in any way. The ERPs elicited by the tones were analyzed by measuring the total amount of excursion, as if the brain wave were a piece of string straightened out. This measure correlated substantially (at least +0.7) with a measure of intelligence (Raven's Progressive Matrices Test). The authors are cautious in their interpretation, but the implication is that the string measure is an indicator of brain complexity, with high complexity being related to high intelligence.

Autonomic measures (IBI and skin conductance)

For the purposes of analysis, IBIs and skin conductance responses were averaged into four blocks of six tone trials each. Thus, each block represents 30-second segments of the 2-minute tasks. The IBI results are summarized in Table 3. The numbers in parentheses are the heart rates corresponding to each mean IBI. One of the major results is that heart

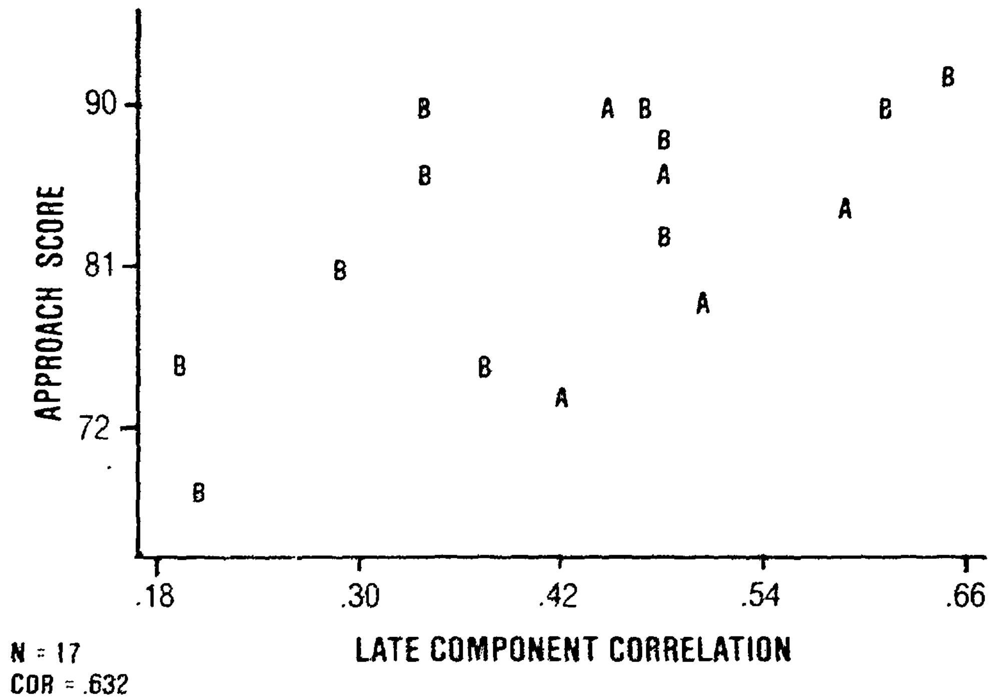

Figure 5 — Late Component Correlation calculated from the tone alone ERP data correlated with subsequent performance on the carrier landing task.

Table 3. Inter-beat interval (IBI) changes over 4 blocks of 6 tone trials each for the 3 conditions.

Condition	TRIAL BLOCK 1	2	3	4
Tone task alone	[a] 851(70.5)	876(68.4)	897(66.8)	901(66.6)
Carrier task alone	886(67.7)	856(70.1)	829(72.4)	773(77.6)
Combined tasks	876(68.5)	860(69.8)	843(71.2)	801(74.9)

[a] IBI in msec; corresponding heart rate in beats per minute is shown in parentheses.

rate slows significantly ($p<0.01$) over blocks during the tone task performed alone. A second major result is that heart rate increases significantly as the subjects fly closer to the carrier landing area in either the carrier-alone condition ($p<0.001$) or the combined-task condition ($p<0.001$). Both of these effects were previously reported using the same tasks but a different population of subjects (Lindholm, et al, 1982; see also Lindholm & Cheatham, 1983). The IBI changes are interpreted as reflecting workload differences among the tasks. That is, the tone-alone condition is low workload; the task is simple and workload decreases over blocks because the subject has heard the tones on previous blocks and has a recent memory for the different pitches. By contrast, the carrier task is high workload, and becomes increasingly higher as the landing area is approached. As mentioned in the Introduction, increased heart rate on final approach to landing is commonly observed in pilots flying actual aircraft, and the same effect is observed here, using a simulated landing. These results taken together leave little doubt that heart rate changes track workload fluctuations in a highly reliable manner.

Skin conductance responses did not change significantly as a function of any of the task variables. This is a departure from earlier findings (Lindholm et al, 1982; Lindholm & Cheatham, 1983) in which skin conductance amplitude did increase during final approach on the carrier landing task. However, even in those reports, skin conductance changes were not as stable as heart rate changes; also skin conductance is more difficult to quantify and is affected by more sources of artifact than is heart rate. Since one purpose of these experiments is to determine the relative worth of different physiological measures for workload studies, it is suggested that heart rate should be the autonomic measure of choice.

Summary

Although the ERP has been investigated in previous workload studies (Isreal et al, 1980; 1980), interest focused on only one component, the P3. The present results suggest that P3 is not uniquely important and might even be secondary in importance to the preceding N2. Reliable indicators of increased workload, summarized in Figure 4, include the P3-N1 latency shift and a negative shift in the amplitudes of both N2 and P3. The implication is clear that components other than P3 should be vigorously investigated in future workload studies.

Potentially of extreme importance is the finding that a subject's LCC score has power to predict how well that person will perform in the subsequently administered carrier-landing task. Prediction of performance is one of the most elusive areas of human research, and one of the most important. Further work with the LCC measure is certainly indicated.

Finally, heart rate seems firmly established as a reliable metric for the type of workload typified by final approach to landing. The results reported here replicate earlier findings (e.g., Lindholm & Cheatham, 1983) and agree well with field studies investigating pilots in actual aircraft.

In the following experiment, many of the same physiological measures are used, but the task is very different and rated pilots served as subjects. No secondary task was employed, but tone stimuli were used to provide the pilot with information critical to mission success. The task is more realistic than is the carrier-landing task and provides information relevant to how physiological indices of workload may be applied in operational environments.

EXPERIMENT 2

Experiment 1 utilized a reasonably simple flight simulator and inexperienced pilots. In Experiment 2, Air Force rated pilots at two levels of experience flew a simulated dangerous mission in the Advanced Simulator for Pilot Training (ASPT) located at Williams AFB. Heart rate, respiration rate, ERPs and eye movements were recorded continuously during 10 simulated missions of 3 minutes each.

EXPERIMENT 2

METHOD

Subjects

The subjects were 20 pilots who were in transition training to the F-16 (T-Course pilots) and 14 pilots who had just finished Undergraduate Pilot Training (B-Course pilots). Complete data were obtained on 11 T-course and 7 B-course pilots. The mean ages of these were 30.5 years for the T-course and 25.6 years for the B-course. Mean flight time was 1617 hours for T-course and 913 hours for B-course.

Apparatus

The F-16 version of ASPT was programmed in the following manner. The aircraft was released from freeze at an altitude of 294 m, 5 km from a south mountain pass. The pilots had been previously briefed, with terrain maps, that the target (a rectangular building 30.5 m by 15.2 m by 15.2 m) was located 18 km beyond the south mountain pass in a west by northwest direction. They were to follow a road as a landmark, maintaining as low an altitude as possible. A reasonable pop-up point was marked by an intersection of that road with another. Following bomb release, the pilots were instructed to make a 180- degree turn around a small mountain located adjacent to and north of the target. Egress was through a north mountain pass. The mission track was thus U-shaped with a roughly west heading for ingress and an east heading for egress. The ingress and egress paths were approximately equal in length (18 km).

The briefing also clearly described the locations of four surface-to-air missile (SAM) sites located directly in the ingress path and approximately 5 KM from the target. Thus, it was virtually impossible for the pilot to pop without attracting SAM activity.

Physiological Recording

A custom-designed six-channel biotelemetry system manufactured by UFI, Morrow Bay, California, was used so that no hard wire connections were needed between the ASPT cockpit and the recording device. This system consists of three component groups: (a) a miniature moderator/transmitter weighing about one pound (0.5 kilogram), carried on the pilot's person, (b) a receiver/demodulator rack-mounted within 50 feet (15 m) of the pilot, and (c) a four-channel amplitude-modulated (AM) tape deck; special circuity permitted all six biotelemetry channels to be recorded on a single tape channel in pulse width modulation form. This tape deck was rack-mounted with the receiver/demodulator. Channels 2 and 3 were used for special event pulses, and channel 4 was reserved for voice commentary. A multiple trace oscilloscope was also rack-mounted so that the various channels of data could be monitored.

Additional electronic circuitry was used to interpret logic pulses from ASPT so that important events (e.g., freeze/release from freeze bomb release) could be stored on Channels 2 or 3 of the tape. This circuitry also triggered oscillators so that tones could be sent to the pilot over his headset.

In this experiment, the tones were used to signal the level of SAM threat: 1000 Hz tones signalled "safe" (no threat), 1250-Hz tones signalled "acquisition" (SAM radar has acquired aircraft), and 1750-Hz tones signalled "launch" (a SAM has been launched against the aircraft). Each tone was 250 msec in duration and was presented each 3 seconds. Thus, the pilot had frequent information concerning the level of SAM threat.

Recording leads for heart rate, brain wave activity, and eye movement were attached as described for Experiment 1. Additionally, respiration was measured by having the pilot wear a lightweight dust mask of the type found in any hardware store. A thermistor was affixed inside the mask so that inhalations and exhalations were reflected as cooling and warming, respectively, of the thermistor. This simple device produces very clear respiration records.

Once the leaders were attached and connected to the modulator/transmitter, the pilot was asked to wander about while the equipment was checked. This biotelemetry system yields records remarkably free of movement artifact except when the pilot makes gross head movements (e.g., looking over shoulder); such movements interfered greatly with brain wave recording. Following final equipment check, the pilot entered the ASPT cockpit and the simulated missions began.

A duplicate receiver/demodulator and tape deck resided in the ASU laboratory. At the end of data collection at Williams AFB, only the magnetic tapes needed to be transported back to the ASU laboratory for decoding and statistical analysis. Although all the physiological data and some important events were stored on the AM tapes, many other data sources were not. For example, altitude, airspeed, g-force, and several other variables were accumulated by the DATARECORD capability of ASPT and stored on digital tape. It was agreed that this information would be decoded and re-formatted by Williams AFB personnel in a manner that could be read and interpreted by the ASU laboratory.

Procedure

Each simulated mission was about 3 minutes in duration. In a single session 10 to 12 missions were administered, with 30 seconds to 60 seconds between missions. Between missions, the pilot and experimenter communicated over an intercom to check for problems and questions. The biotelemetry signals were continuously monitored on the oscilloscope to ensure continued proper functioning of the equipment. The pilot was told his bomb score (bomb-to-target miss distance in meters) after each mission, or if he was hit by a SAM, the particular SAM site was identified for him. Pilots were fully informed concerning the purposes of the experiment and the majority showed keen interest in the physiological techniques and the hardware.

EXPERIMENT 2

RESULTS AND DISCUSSION

Data limitations

The data for some subjects had to be discarded in whole or part because of hardware problems with the ASPT and artifacts in the physiological records. The ASPT is a highly visual environment and leads to considerable eye and head movement; as a result, the brain wave data were seriously contaminated for several subjects. Still, complete data sets were available for 11 T-course and 7 B-course pilots.

Mission success

None of the pilots in this study were highly familiar with the F-16; therefore, it was expected that skills relevant to mission success would show improvement over the 10 simulated missions. Four aspects of improvement from first half to second half of training were examined: (a) bomb-to-target miss distance, (b) percentage of bomb releases, (c) percentage of hits by SAMs, and (d) percentage of time spent in low SAM threat and high SAM threat conditions.

Bomb-to-target miss distance was a highly variable score and did not improve with practice ($p<0.80$). B-course pilots tended to have a shorter mean miss distance (87 m) than T-course pilots (466 m) over all missions combined; although this difference appears large, it was not significant ($P<0.10$). The high variance in the bomb-to-target score was apparently due to at least two factors. First, most pilots were unfamiliar with the F-16 "heads-up display" (HUD) and weapons release procedure; secondly, the SAM threats were taken very seriously and all pilots seemed more concerned with avoiding launches and sustained tracks with than accurate bomb scores.

Percentage of bomb releases did increase reliably as a function of practice, from 61% during the first half of practice to 77% during the second half of practice ($p<0.03$). T-course and B-course pilots did not differ significantly on this measure. Similarly, the percentage of missions on which the pilot was hit by a SAM decreased significantly with practice from 33% during the first half of training to 10% during the second half of training ($p<0.003$). The B-course pilots tended to get hit less often than did the T-course pilots (6% SAM hits for B-course, versus 14% SAM hits for T-course pilots in second half of training), but this difference was not statistically significant ($p<0.20$).

Figure 6 shows the percentage of time spent in the lowest threat level (safe) as a function of halves of training, and Figure 7 shows the percentage of time spent in the highest threat level (launch) as a function of halves of training. As is apparent from these figures, the younger, less experienced B-course pilots outperformed the older, more experienced T-course pilots on this metric. ANOVAs performed on the data shown in Figures 6 and 7 revealed that B-course pilots spent more time in safe and less time in launch than did T-course pilots ($p<0.01$ in both comparisons). Both groups improved with practice ($p<0.01$), and there was no interaction between groups and halves of training.

Some notion of why the B-course pilots performed better on this measure can be obtained from examination of variables closely related to threat level: altitude, airspeed, and the amount of chaff used. ANOVAs performed on these variables showed that B-course pilots did fly at lower altitudes when they were not engaged in pop-up. Mean altitude during the safe condition was 254 feet (77.4 m) for B-course and 433 feet (132 m) for T-course pilots (altitude difference significant, $p<0.05$). However, the groups did not differ significantly with respect to altitude at which maximum SAM threat was encountered (992 feet or 302 m for B-course and 850 feet or 259 m for T-course, $p<0.3$). Regarding airspeed, B-course pilots tended to fly more slowly (mean over all threat levels was 536 knots for T-course and 492 knots for B-course), but not significantly so ($p<0.10$); similarly, B-course pilots tended to use more chaff than did T-course, but this difference also was not significant ($p<0.10$).

The general picture that emerges from these results is that the B-course pilots were somewhat better at this task than were the older, more experienced T-course pilots. B-course pilots flew at lower altitudes and avoided serious SAM threats a greater percentage of the time. They also tended (not significant but strong trend) to have better bombing accuracy, they tended to use chaff more often, and they tended to get hit less often by SAMs. Finally, both groups improved with practice as shown by the increased percentage of bomb releases, decreased percentage of hits by SAMs, increased percentage of time spent in the low SAM threat condition, and decreased percentage of time spent in the high SAM threat condition from first to second halves of training.

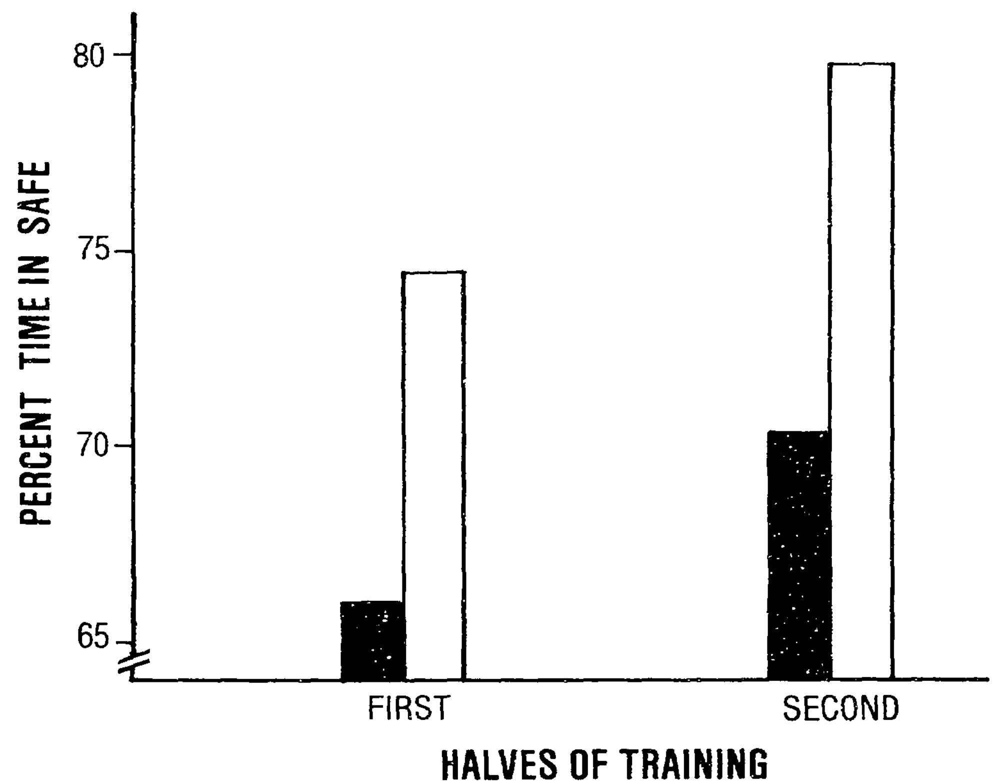

Figure 6 — Percent time spent in the lowest threat level (safe) as a function of practice. Solid bars = T-course pilots, open bars = B-course pilots.

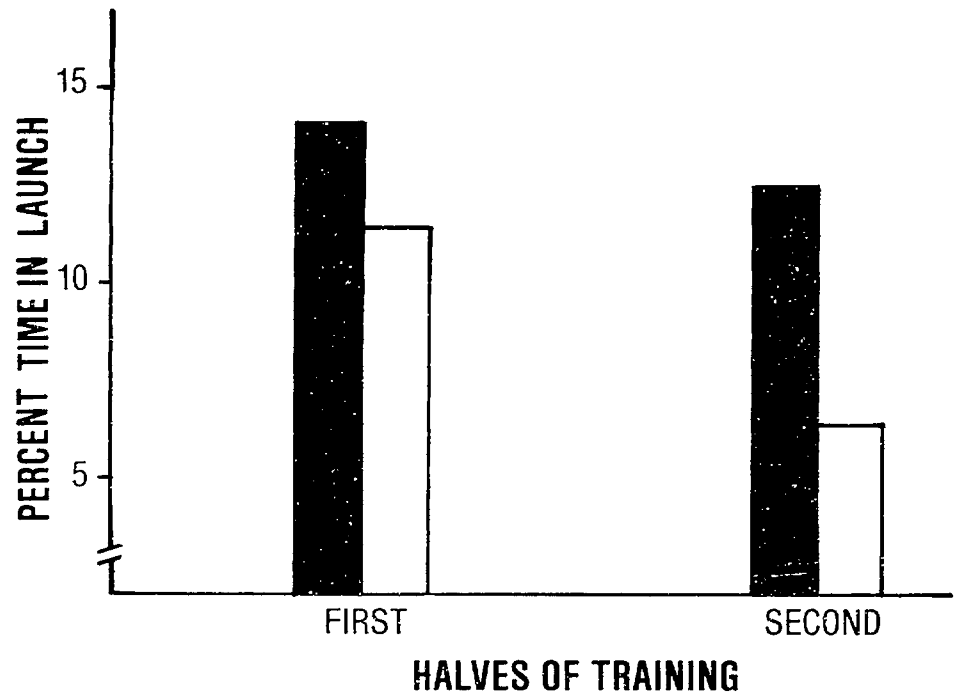

Figure 7 — Percent time spent in the highest threat level (launch) as a function of practice. Solid bars = T-course pilots, open bars = B-course pilots.

ERP results

As in Experiment 1, N2 amplitude appeared to represent some brain process importantly involved in processing critical auditory information. The major results for N2 amplitude are shown in Figure 8 where N2 amplitude is plotted against threat level (there were no differences between T-course and B-course pilots, so the function represents the mean of all subjects). N2 amplitude became more negative as the tone signalled higher threat levels. The linear component was significant ($p<0.001$) and the quadratic component approached significance ($p<0.07$). These results, in conjunction with visual inspection of Figure 8, suggest that N2 amplitude was roughly the same for the two lowest threat levels (safe and acquisition), but became more negative at the two higher threat levels.

P3 amplitude also changed with threat level, but in a less orderly fashion. These results, shown in Figure 9, indicate that P3 amplitude generally increased as threat level increased. The ANOVA performed on these results indicated that only the linear trend was significant ($p<0.05$). Thus, while visual inspection of Figure 9 suggests a non-monotonic function, the statistical results argue for monotonicity.

This relationship between N2 and P3 amplitude is somewhat different than that reported for Experiment 1. In Experiment 1, both late components became more negative (less positive) when workload was increased by introduction of the high visual workload carrier landing task. In Experiment 2, the visual workload was always high and relatively constant; the tones in this case were not secondary task stimuli which could be ignored, but signals that were vital to successful completion of the mission.

Another line of evidence suggests that the tones of Experiment 1 and Experiment 2 were being processed in different ways. The P3-N1 shift observed in Experiment 1 was not present in Experiment 2; rather, P3 latency increased with threat level whereas N1 latency remained essentially invariant. The P3 latency results are summarized in Figure 10. The associated ANOVA revealed only a linear effect ($p<0.002$), suggesting that P3 latency increased monotonically and linearly with threat level.

Autonomic results (heart rate and respiration)

The major results for heart rate (inter-beat interval) are summarized in Figure 11. The older, T-course pilots had higher mean heart rates over all threat conditions than did the younger, B-course pilots ($p<0.02$). Both groups showed increased heart rate as a function of increasing threat ($p<0.001$). The functions shown in Figure 11 appear roughly parallel, and analysis bears out this visual impression. That is, there was no significant interaction of groups by threat ($p<0.40$).

The respiration results are shown in Figure 12. The ordinate represents "respiration activity" in polarity changes each 3 seconds. A polarity change would occur when an inspiration changed to an expiration

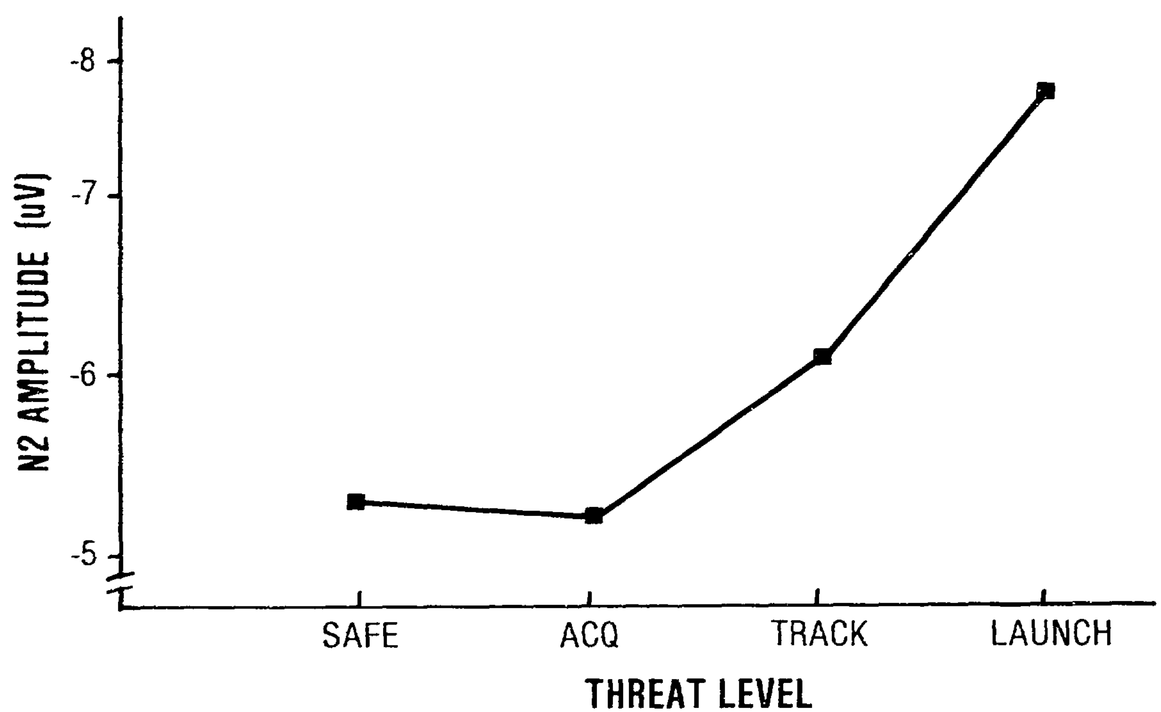

Figure 8 — N2 amplitude as a function of threat level.

Figure 9 — P3 amplitude as a function of threat level.

Figure 10 — P3 latency as a function of threat level.

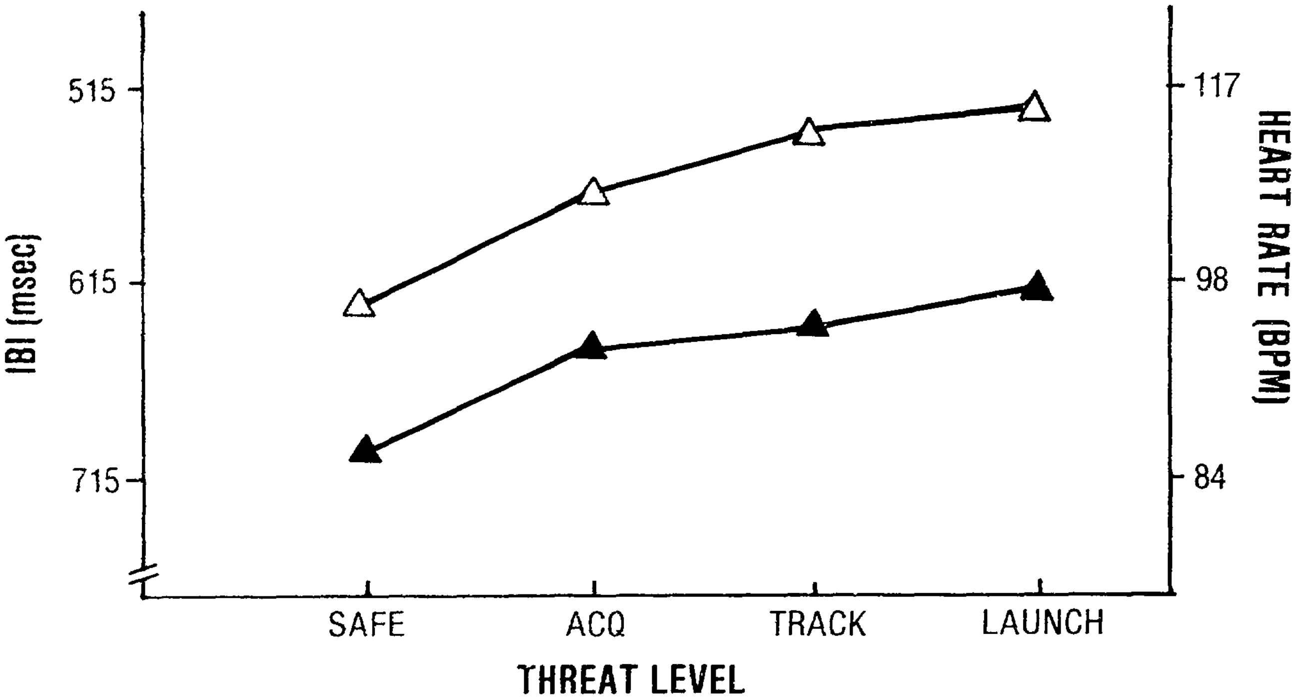

Figure 11 — IBI as a function of threat level. Closed triangles = T-course pilots, open triangles = B-course pilots. Heart rate plotted on right ordinate.

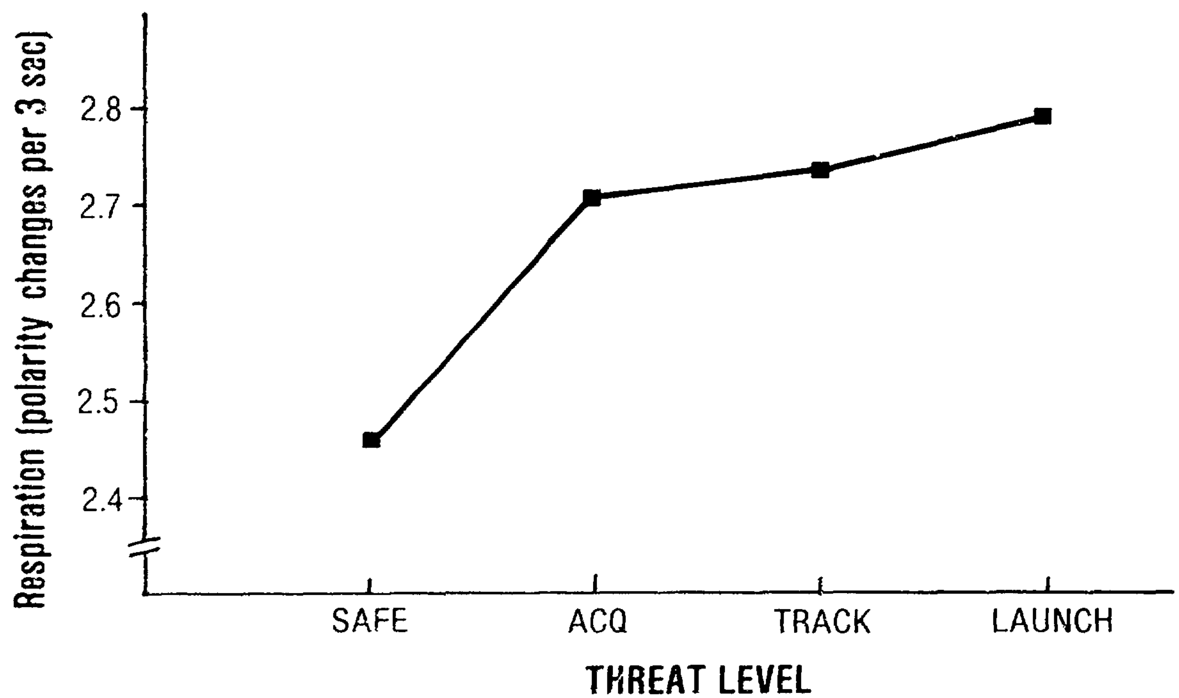

Figure 12 — Polarity changes on the respiration channel as a function of threat level. A polarity change would be caused by an inspiration and expiration, thus polarity change is a close approximation to breathing rate.

or vice versa. The activity measure is preferable to breathing rate since polarity changes also reflect "catches of the breath" (brief inspirations during expiration or brief expirations during inspiration) which are likely to occur in high workload situations upon receipt of important information.

In this case, the groups did not differ significantly; thus, the function represents the mean for all subjects. The thermistor polarity changes increased with increasing threat level ($p<0.001$) and the major component was linear.

Thus, as threat level increased, so did two important autonomic measures of workload: heart rate and respiration activity. The greater overall heart rate for the T-course pilots may be due to age alone, but it is not possible to make that statement with confidence from the available data. Still, there seems little other explanation. They were more experienced and should have found the mission easier than did the B-course pilots; they flew at higher altitudes, thus reducing (somewhat) their moment-to-moment workload. Whatever the reason for the higher mean heart rates in the T-course pilots, the more interesting observation is that both T-course and B-course pilots showed heightened autonomic activation when the tone signalled higher threat levels.

GENERAL DISCUSSION

The study of cockpit workload and pilot performance can progress only so far using dual-task methodology. The dual-task approach, or for that matter, any intrusive technique, runs the risk of measuring a pilot's capability for tolerating interference rather than a pilot's capability to perform multiple tasking elements when all tasking elements are important. The different uses of tone stimuli in Experiment 1 and Experiment 2 serve as an example. In Experiment 1, the tones were viewed by subjects as secondary task stimuli. During the last portion of the final approach to landing, subjects tended to ignore the tones or respond to them slowly, as evidenced by the increase in errors and RTs. Further, the tones in Experiment 1 did not elicit significant changes in autonomic activity; neither heart rate nor skin conductance covaried with tone presentation. By contrast, the tones in Experiment 2 carried vital information concerning the degree of SAM threat. The tones were an integral part of the task and could not be considered low priority. Under these conditions, there was considerable autonomic involvement. Heart rate and respiration increased as a direct function of threat level. Although the term "threat" is used freely in this report, it is understood that there is no actual threat because the SAMs are not real. However, increased simulated threat increased pilot workload in that the pilot had to lower altitude, use chaff, and/or mentally calculate a strategy (e.g., "How long has that track tone been on? Am I nearly out of range or will the next tone be a launch? Should I lower altitude even more and run the risk of crashing, or should I wait?").

The Experiment 2 environment provides a plausible prototype for operational use of physiological measures in workload studies. The telemetry modulator/transmitter could be easily adapted for use with an aircraft in-flight recorder rather than broadcasting the signals to a receiver. This would mean that physiological responses could be easily time-locked to the flight parameters and events normally stored on the in-flight recorder. The entire mission could be reconstructed from the recorder, complete with moment-by-moment fluctuations of the physiological state of the pilot.

There seems no doubt that heart rate should be accepted as the simplest, most consistent, and most easily interpreted physiological metric of workload. The heart rate measure has been used for nearly two decades, and with impressive results. For example, Roman (1965) addressed the question of perceived risk to life during final approach by using two equally qualified pilots in a dual-seat aircraft. One pilot was in control, the other pilot was a passenger. Presumably, perceived threat to life would be equated for the two pilots, or perhaps, the passenger pilot would feel more anxiety because he had to trust the skills of the controlling pilot. However, the heart rate of the controlling pilot increased more than the heart rate of the passenger pilot. Similar results were reported later by Roscoe (1978), and both sets of authors argued that increased heart rate reflects increased mental workload rather than "anxiety." Their position was strengthened by Lindholm and Cheatham (1983) who showed that heart rate increased on a simulated landing task (the same task used in Experiment 1) where, obviously, there was no chance of an actual crash involving bodily harm. Since the Lindholm and Cheatham heart rate results were replicated in the present Experiment 1, the argument is strengthened even further. Also, Lindholm and Cheatham (1983) suggested that heart rate changes and workload variations can be interpreted within the theoretical framework of activation theory proposed by Duffy (1972): as workload increases, sympathetic autonomic activity must increase in order to meet the demands placed on the human system. One clear sign of increased sympathetic activity is an increase in heart rate, and certainly all the studies relating heart rate to workload yield results compatible with this theoretical model.

The value of heart rate was also demonstrated in Experiment 2 where specific tone stimuli provided information critical to successful mission completion. Heart rate (and respiration activity) increased as the simulated threat level increased. Increased threat level placed greater demands on the the pilot's ability to deal effectively with all elements of the task (fly low, watch for the pop-up point, get the HUD adjusted, etc) and, as such, represented increased mental workload. Taken together, results from previous as well as present studies recommend heart rate strongly as a reliable, non-intrusive measure of cockpit workload.

The ERP is more complex than is heart rate and represents central nervous system functioning rather than autonomic nervous system functioning. Although autonomic variables have been studied for more than a century, the history of brain wave research has been slower to develop due in no small part to the technical difficulties of recording and quantification. Modern, but still early, attempts to understand the complex ERP viewed the ERP waveform from stimulus onset to about 0.5 second following stimulus onset as a series of components. These components were assumed to be relatively independent and representative of different brain processes. These implicit assumptions, that different ERP components are indicative of different information processing stages, are apparent in the writings of Hillyard and coworkers (e.g., Hillyard & Picton, 1979), who view N1 and P2 as largely related to attention, and Donchin (e.g., Donchin, 1981) who views P3 as the critical ERP component for the study of information processing and decision-making. Donchin (1981) has suggested that P3 is an electrical manifestation of central nervous system (CNS) "subroutines" which become active during certain critical stages of central processing. The evidence, however, is largely based on a single paradigm, the so-called "odd-ball" paradigm in which subjects are instructed to detect a low probability stimulus among a string of higher probability stimuli. The closest real-world analogue to this paradigm is the classic vigilance task. Thus, a possible problem with Donchin's analysis of the importance of P3 is the restricted paradigms used. A second problem is that the intense interest in P3 has tended to overshadow interest in N2. Certainly, the results reported here and in previous investigations (Lindholm, Ruppel & Buckland, 1979); Lindholm et al, 1982) argue that P3 should not be considered as uniquely important in the study of human information processing and operator workload. The present results suggest that N2 might be a more important sign of CNS processing than is P3. These results also suggest that ERPs are different for tone stimuli which may be considered by the subject as "low priority, secondary task stimuli" (Experiment 1) as opposed to "stimuli critical for successful task completion" (Experiment 2). If researchers are ever to understand the full importance of ERP components in workload studies, or the more basic question of the meaning of ERPs, it seems clear that future research should follow this lead of examining the entire brain wave rather than concentrating on just one of the several components. Also, investigators should not overly use one single paradigm because this constricts generality.

The greatest consistency in the ERP waveform across the two experiments reported here was in the amplitude of the N2 component, which became more negative as workload increased. By contrast, P3 amplitude decreased as workload increased in Experiment 1, and P3 amplitude increased as workload increased in Experiment 2. P3 latency increased with workload in Experiment 2, but in Experiment 1, the latency change was evidenced only by the P3-N1 shift. It is difficult to interpret these differences with the limited amount of data available on multiple component responses, but one speculation might be offered based on the different task demands of the tone stimuli in the two experiments: in Experiment 1, the tone required only a simple RT response and a binary decision ("Yes, I should respond" or "No, I should not respond"). Also, the tone had no effect; this was a simple discrimination and autonomic activation did not vary with tone presentation. Under these conditions, both N2 and P3 became more negative (less positive) when the visual task was introduced. By contrast, the tones in Experiment 2 were an integral part of the main task, could not be considered as "secondary," and did alter autonomic activity considerably. Under these conditions, N2 became larger in the negative direction and P3 became larger in the positive direction as workload (threat) increased. This might suggest that P3 amplitude is, at least in part, affected by emotions associated with the stimulus, whereas N2 is more a reflection of stimulus evaluation independent of affect.

CONCLUSIONS

In two experiments, the physiological state of naive subjects and Air Force rated pilots was quantified while they flew aircraft simulators. In both experiments, heart rate closely tracked workload changes and, because of its ease of measurement and quantification, is suggested as the measure of choice in future workload studies. Respiration activity, measured in Experiment 2, provided results very similar to those for heart rate and should be quantified whenever possible. The brain wave (ERP) shows great promise as a measure of central nervous system activity and appears to reflect brain events relevant to information processing and decision-making. Although much earlier work has concentrated on the P3 or N1, the present results suggest that N2 might represent a process more closely related to decision-making than does P3.

It is strongly urged that future ERP studies abandon the narrow approach of investigating only one component at a time. Instead, simultaneous quantification of multiple components should become standard procedure so that relationships among components can be better understood.

References

Blinkhorn, S.F. & Hendrickson, D.E. Average evoked responses and psychometric intelligence. Nature, 1982, 295, 596-597.

Cheatham & Lindholm (unpublished) cited on p.15

Donchin, E. Surprise!...Surprise? Psychophysiology, 1981, 18, 493-513.

Duffy, E. Activtion. In N.S. Greenfield & R.A. Sternbach (Eds.), Handbook of Psychophysiology. New York: Holt, Rinehart, and Winston, 1972.

Hillyard, S.A., & Picton, T.W. Event-related potentials and selective information processing in man. In J.E. Desmedt (Ed.), Cognitive components in cerebral event-related potentials and selective attention. New York: Karger, 1979.

Isreal, J.B., Chesney, G.L., Wickens, C.D., & Donchin, E. P300 and tracking difficulty: Evidence for multiple resources in dual task performance. Psychophysiology, 1980, 17, 259-273.

Isreal, J.B., Wickens, C.D., Chesney, G.L., & Donchin, E. The event-related brain potential as an index of display-monitoring workload. Human Factors, 1980, 22, 211-224.

Lindholm, E., & Cheatham, C.M. Autonomic activity and workload during learning of a simulated aircraft carrier landing task. Aviation, Space, and Environmental Medicine, 1983, 54, 435-439.

Lindholm, E., Cheatham, C.M., Longridge, T., & Buckland, G.H. Physiological and dual-task assessment of workload during tracking and simulated flight. Final report submitted fall, 1982. AFOSR, Bolling AFB, contract #F49620-79-C-0197.

Lindholm, E., Ruppel, M., & Buckland, G.H. Attention and task complexity as indicated by physiological indices. AFHRL TR-79-47, AD-A080851. Williams AFB, AZ: Flying Training Division, Air Force Human Resources Laboratory, December 1979.

McLeod, P. Does probe RT measure central processing demands? Quarterly Journal of Experimental Psychology, 1978, 30, 83-89.

Nicholason, A.N., Hill, L.E., Borland, R.G. & Krzanowski, W.J. Influence of workload on the neurological state of a pilot during the approach and landing. Aerospace Medicine, 1973, 44, 146-152.

Pritchard, W.S. Psychophysiology of P300. Psychological Bulletin, 1981, 89, 506-540.

Ritter, W. Latency of event-related potentials and reaction time. In D. Otto (Ed.), Multidisciplinary perspectives in event-related brain potential research. EPA-600/9-77-043, U.S. Government Printing Office, Washington, D.C., 1978.

Rolfe, J.M. The secondary task as a measure of mental load. In W.D. Singleton, J.G. Fox, & D. Whitfield (Eds.), Measurement of man at work. London: Taylor and Francis, 1971.

Roman, J.A. Flight research progran: II. Risk and responsibility as factors affecting heart rate in test pilots. Aerospace Medicine, 1965, 42, 1200-1207.

Roman, J.A., Older, H., & Jones, W.L. Flight research program: VII. Medical monitoring of Navy carrier pilots in combat. Aerospace Medicine, 1967, 38, 133-139.

Roscoe, A.H. Stress and workload in pilots. Aviation, Space, and Environmental Medicine, 1978, 49, 630-636.

Ruffel Smith, H.P. Heart rate of pilots flying scheduled airline routes. Aerospace Medicine, 1967, 38, 1117-1119.

Wickens, C.D. Human workload measurement. In N. Moray (Ed.), Mental workload: Its theory and measurement. New York, Plenum Press, 1979.

Wierwille, W.W., & Williges, R.C. Survey and analysis of operator workload assessment techniques. Blacksburg, VA: Systemetrics, Technical Report S-78-101, 1978.