U.S. Army Research Institute
for the Behavioral and Social Sciences

Research Report 1636

The Effects of Superimposing Symbology on a Simulated Night Vision Goggle Display

D. Michael McAnulty, John W. Ruffner, and David B. Hamilton
Anacapa Sciences, Inc.

February 1993

FOREWORD

The U.S. Army Research Institute Aviation Research and Development Activity (ARIARDA), an operational unit of the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI), provides research support in aircrew training to the U.S. Army Aviation Center, Fort Rucker, Alabama. Research is conducted in-house and augmented by contract support, as required. This report documents work performed by Anacapa Sciences, Inc. for ARIARDA under a Dynamics Research Corporation contract. The project was entitled "Exploration of Head-Up Display and Night Vision Attentional Issues in Rotary Wing Cockpits." The work was performed for the MANPRINT Division in support of an in-house research program on the Aviator Night Vision Imaging System Head-Up Display (ANVIS-HUD).

During low level, night vision goggle (NVG) helicopter flight, it is difficult and potentially unsafe for aviators to divert their attention from the external scene to obtain critical flight information from cockpit instruments. To address this problem, the U.S. Army is acquiring a system that superimposes flight symbology on the imagery seen through one of the NVG intensifier tubes. However, previous research indicates that superimposed symbology may distract the pilot's attention from obstacle detection, recognition, and avoidance, and may interfere with proper scanning patterns.

This report describes the results of a research project designed to determine how well rotary wing aviators can monitor the simultaneous display of NVG imagery and symbology. The results showed a relatively small decrement in performance when the two types of information were presented together. Aviator performance with the system improved with practice and was related to pilot experience and eye dominance. The aviators suggested several modifications for the symbology suite.

The results of this research can be used to modify the proposed symbology suite and to develop operational procedures for fielding the equipment. The research also identifies additional NVG imagery and symbology issues that need to be investigated.

EDGAR M. JOHNSON
Acting Director

ACKNOWLEDGMENTS

The authors express their appreciation to the following individuals who contributed to this project. From the U.S. Army Research Institute Aviation Research and Development Activity, Dennis K. Leedom provided overall guidance for the research, MAJ Dale Weiler Coordinated research subjects, and Larry Murdock conducted the statistical analyses. From Dynamics Research Corporation, Robert Simon and John Morey provided advice on the research design and reviewed a draft of the report.

Five members of the Anacapa Sciences staff assisted in conducting the research. We thank Gary Coker for designing the computer architecture, writing the computer programs, and supporting the experimental sessions; Kenneth Persin for providing technical support; Annette Swan for scoring the scene data; Tina Pridgen for providing administrative, scoring, editing, and word processing support; and David Thill for serving as an experimenter.

Finally, we thank CW2 Joseph Cogelia, who provided subject matter expertise in developing the scenarios and in identifying scoring criteria. Mr. Cogelia also served as an experimenter and compiled technical information for the project report. Although they are not mentioned by name, we also thank the aviators who enthusiastically participated in this research.

THE EFFECTS OF SUPERIMPOSING SYMBOLOGY ON A SIMULATED NIGHT VISION GOGGLE DISPLAY

EXECUTIVE SUMMARY

Requirement:

U.S. Army helicopter pilots routinely use night vision goggles (NVGs) to conduct low level flights at night. The most recent version of NVGs, the aviators night vision imaging system (ANVIS), allows the pilot to look under or around the goggles to read critical flight instruments. However, reading instruments that differ in illumination and optical distance from the external scene is difficult and potentially unsafe. To main their attention on the external scene, pilots often rely on verbal callouts from their copilots. With this method, successful crew coordination becomes increasingly important for mission performance and flight safety.

To remedy this problem, the Army is acquiring a system that superimposes instrument symbology on the imagery seen through one of the ANVIS intensifier tubes. The combination is similar to a head-up display (HUD) that projects symbolic information on a combiner lens near the aircraft windscreen. Although the ANVIS-HUD system has several benefits, previous research with HUDs and monocular helmet-mounted displays indicates that instrument symbology may distract pilots' attention from obstacle detection, recognition, and avoidance, and may interfere with proper scanning patterns. However, little research has been conducted with prototype ANVIS-HUD systems to determine potential advantages and disadvantages.

Procedure:

An apparatus was designed to present scenarios that simulated NVG imagery only, symbology only, and the imagery and symbology combined. After receiving general instructions and providing demographic data, 36 rotary wing aviators participated in five test sessions. In each session, the aviator was required to monitor and respond to predefined targets (scene features and events and symbology out-of-tolerance states) in the scenarios. During test sessions 1 and 5, the aviator viewed the scene and symbology for a helicopter flying a traffic pattern. Twice during each scenario, the screen was blanked and the aviator was

asked to recall the values of four instruments (airspeed, altitude, trim, and master caution) and the scene targets that were in view. The primary purpose of the two test sessions was to measure the effects of practice on the percentage of scene and symbology targets the aviators detected.

During sessions 2, 3, and 4, the aviator received five practice trials and then was tested using a low level flight scenario under scene-only, symbology-only, and scene-plus-symbology viewing conditions. The aviator pressed a key each time one of the instruments exceeded its defined tolerance limits. The aviator pressed a joystick button and called out the name of each scene target detected. Both the percentage of correct detections and aviator reaction time were measured. After each test session, the aviator completed a workload rating form for that scenario. After the experiment was completed, each aviator was debriefed to obtain evaluations of the simulation's realism, descriptions of the aviator's scanning patterns and viewing tendencies, and suggestions for modifying the symbology suite.

Findings:

The results indicate that aviators can detect and respond rapidly to a high percentage of scene-only and symbology-only targets. Performance was significantly better for the symbology targets than for the scene targets. There was a significant decrement in performance when the two types of information were presented simultaneously, but the decrease was small when compared to the increased amount of information available in the display. The aviators were able to divide their attention effectively between the two types of information. They reported occasionally fixating on an instrument symbol or scene feature or event, but the fixation was usually attributed to the relevance of the information in the display. The aviators' performance improved by a small but significant amount with the practice they received during sessions 2, 3, and 4. The order of training (scene-only first vs. symbology-only first) had no effect on the aviators' performance. The order of training did affect the aviators' perception of workload.

Previous experience (age, years as a rated aviator, and flight and NVG hours) and eye dominance were significantly related to performance with the ANVIS-HUD. More experienced aviators performed better with the scene information and less experienced aviators performed better with the symbology information. Presenting the symbology to the aviator's dominant eye generally produced better performance, but additional research is needed to confirm this finding. The aviators rated their workload as low (scene only) to moderate (scene plus symbology). Monitoring and responding to the symbology was the major factor in perceived workload. The aviators thought the simulated ANVIS-HUD scenarios were realistic. Most of the aviators developed a scanning strategy, but the pattern varied among them. Finally, the aviators suggested several modifications for improving the symbology suite.

Utilization of Findings:

The results of this research can be used in four ways. First, the results support fielding the ANVIS-HUD system. The anticipated problems with the system, such as the symbology distracting the aviators' attention from the scene information, were not severe and can be overcome with practice. Second, the findings can be used to modify the symbology suite design for the ANVIS-HUD. Third, the results provide initial information about the optimum configuration of the ANVIS-HUD (i.e., presenting the symbology to the dominant eye). Fourth, the results offer guidance for developing an ANVIS-HUD training program. The research also identified additional NVG imagery and symbology issues that need to be investigated.

THE EFFECTS OF SUPERIMPOSING SYMBOLOGY ON A SIMULATED NIGHT VISION GOGGLE DISPLAY

CONTENTS

CONTENTS (Continued)

LIST OF TABLES

LIST OF FIGURES

CONTENTS (Continued)

GLOSSARY OF ACRONYMS AND ABBREVIATIONS

% det - percent correctly detected
AN/AVS-6 - Army Navy/Aviator Vision System
ANOVA - analysis of variance
AN/PVS-5 - Army Navy/Pilot Vision System
ANVIS - aviator night vision imaging system
AVSCOM - Aviation Systems Command
cd/m^2 - candela per meter square
cm - centimeter
CRT - cathode ray tube
FOV - field of view
HMD - helmet-mounted display
HUD - head-up display
Hz - hertz
LLA - low level scenario A
LLB - low level scenario B
LS - leadship
m - meter
MST - master caution warning
NASA - National Aeronautics and Space Administration
NOE - nap-of-the-earth
NVG - night vision goggle
PC - personal computer
RT - reaction time
s - second
ScOn - scene only
ScSy - scene and symbology
SyOn - symbology only
TLX - task load index
TPA - traffic pattern scenario A
TPB - traffic pattern scenario B
TS - test session
VGA - video graphics array
vs. - versus

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1. AGENCY USE ONLY (*Leave blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1993, February	Interim Feb 92 - Sep 92

4. TITLE AND SUBTITLE
The Effects of Superimposing Symbology on a Simulated Night Vision Goggle Display

5. FUNDING NUMBERS
MDA903-92-D-0025
62785A
790
1211
C05

6. AUTHOR(S)
McAnulty, D. Michael; Ruffner, John W.; and Hamilton, David B.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Anacapa Sciences, Inc.
P.O. Box 489
Fort Rucker, AL 36362-5000

8. PERFORMING ORGANIZATION REPORT NUMBER
ASI958-1-400-92

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute for the Behavioral and Social Sciences
ATTN: PERI-IR
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ARI Research Report 1636

11. SUPPLEMENTARY NOTES
Contracting Officer's Representative, Dennis K. Leedom.

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

12b. DISTRIBUTION CODE
--

13. ABSTRACT (*Maximum 200 words*)

The U.S. Army is acquiring a system that superimposes instrument symbology on night vision goggle (NVG) imagery. However, previous research indicates that the symbology may distract a pilot's attention from obstacle detection, recognition, and avoidance and may interfere with proper scanning patterns. To test the effects of combining imagery and symbology, 36 helicopter pilots were presented night-flight scenarios simulating NVG imagery only, symbology only, and imagery plus symbology. The aviators were required to monitor and respond to predefined scene and symbology targets. They detected and responded rapidly to a high percentage of targets when viewing the scene-only and symbology-only scenarios. Their performance decreased significantly when the two types of information were presented together, but the decrease was small when compared to the increased amount of information available in the display. Aviator performance improved with practice and was related to experience and eye dominance. The aviators suggested several modifications for the symbology suite.

14. SUBJECT TERMS
Night vision goggles, Instrument symbology, Head-up display, Divided attention, Eye dominance, Target detection, Helmet-mounted display, Aviator workload

15. NUMBER OF PAGES
83

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

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std Z39-18
298-102

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Acting Director

Research accomplished under contract
for the Department of the Army

Anacapa Sciences, Inc.

Technical review by

Dennis K. Leedom
Dennis C. Wightman

NOTICES

DISTRIBUTION: Primary distribution of this report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, ATTN: PERI-POX, 5001 Eisenhower Ave., Alexandria, Virginia 22333-5600.

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

Research Report 1636

The Effects of Superimposing Symbology on a Simulated Night Vision Goggle Display

D. Michael McAnulty, John W. Ruffner, and David B. Hamilton
Anacapa Sciences, Inc.

ARI Aviation R&D Activity
Charles A. Gainer, Chief

Training Systems Research Division
Jack H. Hiller, Director

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

February 1993

Army Project Number
2Q162785A790

Human Performance Effectiveness and Simulation

THE EFFECTS OF SUPERIMPOSING SYMBOLOGY ON A SIMULATED NIGHT VISION GOGGLE DISPLAY

Introduction

U.S. Army helicopter pilots routinely use night vision goggles (NVGs) to conduct night missions at terrain flight altitudes. NVGs are electro-optical imaging devices that intensify the available light. Although NVGs do not "turn night into day," they allow helicopter pilots to fly effectively at low altitudes and under poor visibility conditions. Flight under these conditions would not have been possible a decade ago (Kaiser & Foyle, 1991). Nonetheless, night terrain flight with NVGs, especially in the nap-of-the-earth (NOE) mode, is demanding and often results in a high level of workload. During NVG flight, it is difficult and potentially unsafe for aviators to divert their attention from the external scene to obtain flight information from cockpit instruments (Department of the Army, 1988; Simmons, Kimball, & Hamilton, 1985).

The most recent version of NVGs, the Aviator Night Vision Imaging System (ANVIS), allows the pilot to view cockpit instruments by looking under or around the goggles. However, reading instruments that differ in illumination and optical distance from the external scene is time-consuming and difficult. Pilots often rely on verbal information about aircraft status from their copilots while maintaining their attention on the external scene. This procedure requires a high level of coordination by flight crews, which increases crew workload and potentially affects the safety of the flight. In analyzing all Army rotary wing accidents that occurred between 1983 and 1989, Zeller and Thornton (1992) found that aircrew coordination errors were 1.5 times more prevalent in night accidents than in day accidents.

To remedy these problems, the U.S. Army is acquiring a system that superimposes flight symbology (e.g., altitude, heading, airspeed) on the image presented to one of the NVG intensifier tubes. This combination is similar to the head-up display (HUD) that projects symbolic information on a combiner glass at or near the aircraft windscreen. The HUD allows the pilot to view the symbology and the external scene simultaneously. In contrast to the HUD, however, the new Army system projects the scene imagery and superimposed symbology onto a helmet-mounted display (HMD). Army literature refers to the NVG-HMD system as an ANVIS-HUD.

The ANVIS-HUD has several potential benefits. Pilots can access critical flight information without redirecting their gaze from the NVG image of the real world scene (i.e., the actual out-the-window scene) to the instrument panel. Unlike a HUD, HMD symbology is always in view, regardless of where the pilot is looking. Thus, the pilot's workload and the need for nearly continuous crew coordination may be decreased. However, the ANVIS-HUD symbology may inappropriately distract the pilot's attention from obstacle detection, recognition, and avoidance, and may interfere with proper visual scanning patterns (Larish & Wickens, 1991). The symbology may also affect the pilot's judgment of distance, altitude, and closure rates and increase the tendency to become spatially disoriented (e.g., Roscoe, 1987a).

There have been numerous studies about pilots' performance capabilities and limitations with NVGs, HUDs, and HMDs, but little is known about the potential effects of adding symbology to the NVG imagery. Two recent literature reviews (Morey & Simon, 1991a; Ruffner, Grubb, & Hamilton, 1992) summarized and integrated the previous findings and presented recommendations for research on the ANVIS-HUD system. The remaining subsections of the Introduction are based on these two reviews. The next subsection describes NVGs and the superimposed symbology system. The second subsection summarizes the most important findings and conclusions from the two literature reviews. The reviews should be consulted for further detail. In particular, Ruffner et al. provide a review of physiological optics and Morey and Simon provide a summary of theories related to human attentional processes that are not summarized in this Introduction. The final subsection describes the research issues addressed by the current experiment.

ANVIS-HUD System Description

The description of the ANVIS-HUD system is divided into two parts. The first part describes the background and components of NVGs and is based primarily on the literature of Brickner (1989), Price and McLean (1985), and Verona and Rash (1989). The second part describes the system components that generate the superimposed symbology and is based on information obtained from the U.S. Army Aviation Systems Command (AVSCOM), St. Louis, Missouri, and on Buckner (1992).

NVG Characteristics

NVGs are binocular electro-optical devices that increase night visual capabilities by enhancing the available light. The NVG system consists of two image intensifier tubes and the hardware required to mount and position the tubes on the pilot's helmet (see Figure 1). A lens on the far end of the NVG intensifier tube (the objective lens) focuses light on a sensor. Another lens on the near end (the ocular lens)

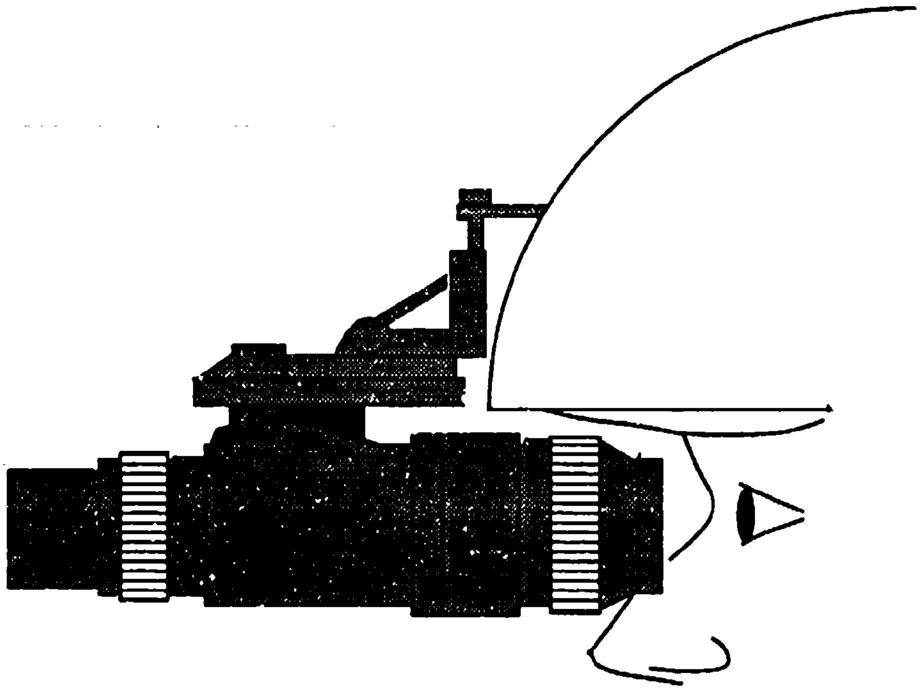

Figure 1. Night vision goggles attached to the flight helmet.

focuses the NVG viewing screen at optical infinity. For practical purposes, objects that are farther than 10 m from the viewer are focused at optical infinity. Thus, the NVG image appears to be at the natural viewing distance. The ocular lens is also used to adjust for an individual's need for corrective lenses. The lens can be adjusted from +2 to -6 diopters.

The objective lens focuses light entering the intensifier tube onto a photocathode that is sensitive to both visible and near-infrared radiation (see Figure 2). The light striking the photocathode causes a release of electrons in proportion to the amount of light. The released electrons are multiplied in a microchannel plate of small glass tubes before striking a phosphor screen. The number and velocity of the electrons striking the phosphor screen determine the amount of light produced by the system. The result amplifies the intensity of the image.

The first NVGs used by Army pilots, the Army Navy/Pilot Vision System (AN/PVS-5), were originally designed for ground vehicle operators. They use second-generation image-intensifier tubes and include a full faceplate that obscures most peripheral vision. Compared

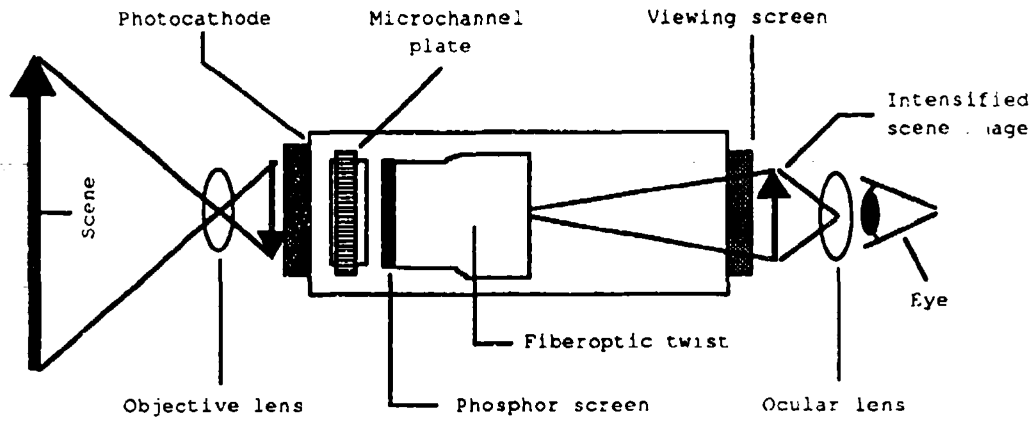

Figure 2. Schematic representation of an NVG image intensifier tube.

to natural viewing conditions, the system has limited resolution and field of view (FOV). Under optimal conditions (i.e., high brightness and contrast), the best visual acuity obtainable is 20/50. Resolution with the AN/PVS-5 is best in the center of the FOV and decreases in the periphery. When the system is properly fitted close to the eyes, it provides a circular 40° FOV. The AN/PVS-5 image has the same 1:1 magnification as when viewing with the unaided eye.

A modified version of the AN/PVS-5 was developed with a cutaway faceplate that reduces the obscuration of the pilot's peripheral vision. This change also allows the NVGs to be mounted to the helmet and flipped up when not in use. However, if the NVG tubes are moved away from the eyes to look under them, the circular 40° FOV decreases and the pilot loses the periphery of the image.

Subsequently, the Army developed the improved ANVIS goggles, known technically as the Army Navy/Aviator Vision System (AN/AVS-6). The ANVIS uses third-generation intensifier tubes, but its operation is similar to the AN/PVS-5. It has the same 40° FOV but is more sensitive to red and near-infrared light, which are characteristic of night illumination. The ANVIS has greater sensitivity, slightly improved central and peripheral resolution, and weighs less than the AN/PVS-5. Although the improved resolution of the ANVIS provides 20/40 acuity under ideal conditions, acuity usually decreases under field conditions. Tredici and Miller (1985) found that acuity under starlight conditions dropped to less than 20/80 with the ANVIS.

Although the scene image is presented to both eyes, NVG depth perception is similar to monocular vision (Wiley, 1989). However, having two images improves perceived brightness and contrast and reduces visual noise (Verona & Rash, 1989).

Superimposed Symbology System

In the ANVIS-HUD system, instrument data are taken from onboard aircraft systems and converted to a digital format. The symbology is projected from a miniature cathode ray tube (CRT) onto a semitransparent combiner lens that is mounted in front of either the left or right intensifier tube. Symbolic data can be depicted for approximately 30 instruments. The symbology suite displayed varies as a function of the aircraft type (UH-1H/V, UH-60A/L, OH-58A/C, AH-1F, and CH-47D) and the flight mode (e.g., hover, cruise) and declutter option selected.

The pilot sees a single image of the scene and symbology in the ANVIS-HUD (see Figure 3). The symbology must be bright enough to be distinguished from the external scene imagery. The pilot and the copilot can independently select a symbology suite and adjust the brightness of the symbology.

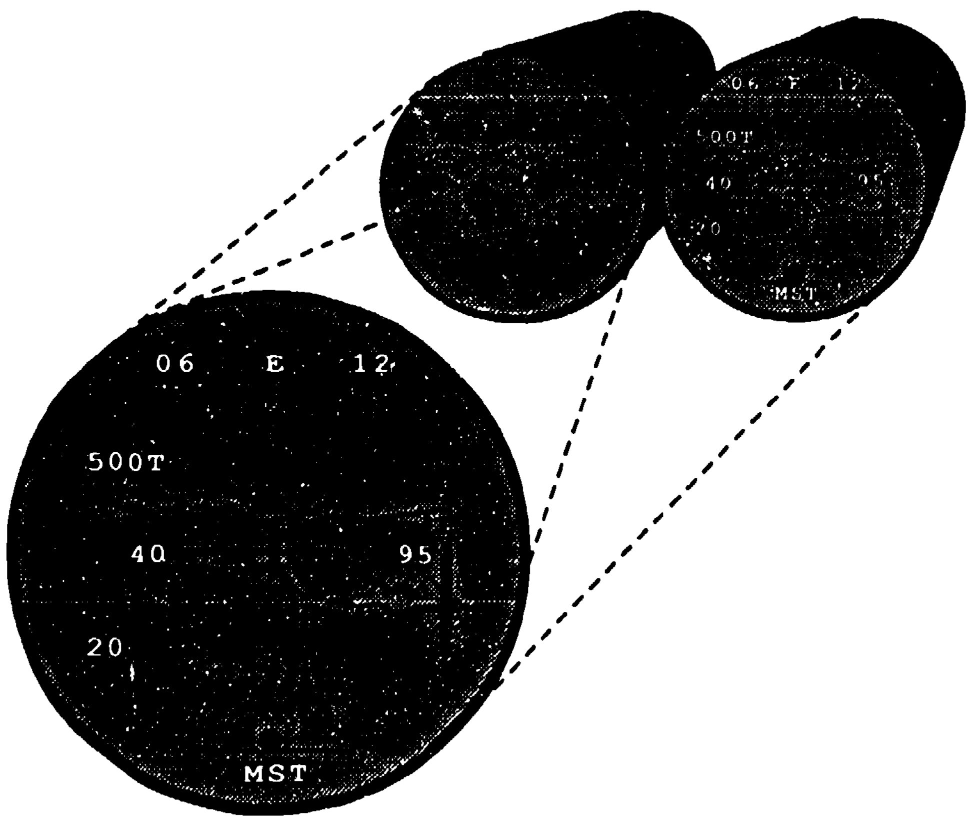

Figure 3. Visual perspective of the night vision goggle imagery with instrument symbology superimposed on one intensifier tube.

Because the symbology and external scene are displayed on the same CRT screen, both images have the same focal distance and require the same amount of accommodation. Thus, reaccommodation should not be necessary when the aviator switches attention between the two types of information. However, several researchers (McLean & Smith, 1987; Moffitt, 1989; Brickner, 1989) have reported that pilots believe the HMD symbology is closer to their eyes than the external scene image. Presently, the accommodation of a pilot wearing an ANVIS-HUD cannot be measured; therefore, it is difficult to determine if changes in accommodation occur.

Summary of Literature Reviews

Ruffner et al. (1992) conducted a general review of the literature on the capabilities and limitations of NVGs, HUDs, and HMDs. From the literature on related devices, they drew inferences about potential problems with the ANVIS-HUD and made recommendations for research. Morey and Simon (1991a) focused their review on cognitive research and theory that are related to ANVIS-HUD performance. They also recommended research on potential correlates of ANVIS-HUD performance. The following two subsections summarize the principal findings and conclusions of the two reviews.

General Literature Review

In their review of the literature, Ruffner et al. (1992) found that many of the perceptual and attentional problems associated with NVGs, HUDs, and HMDs were well documented but not well understood. They found little research that described the perceptual and performance consequences of adding symbology to NVG imagery in a HMD. However, they drew the following six major conclusions from their review of the literature.

First, NVGs provide helicopter pilots with enhanced visual capabilities compared to unaided night vision, but the effectiveness of the devices is limited by low resolution and a narrow FOV (e.g., Brickner, 1989; Weintraub, 1987). Critical flight information can be obtained from the aircraft's instruments by looking under or around the NVGs or from another crewmember. However, both of these methods may result in an increase in individual and crew workload.

Second, pilots using NVGs experience certain types of problems: They tend to overestimate distances and underestimate closure rates (e.g., Foyle & Kaiser, 1991; Roscoe, 1984), to become spatially disoriented (e.g., Brickner, 1989; Price & McLean, 1985), and to experience high levels of fatigue and workload. Factors affecting pilot performance with NVGs may interact with factors affecting pilot performance with HUDs and HMDs, causing a decrease in performance or an increase in workload.

Third, when pilots view the external scene through a collimated HUD, they tend to overaccommodate when viewing the symbology, even though the symbology and the external scene are at the same focal distance (e.g., Iavecchia, Iavecchia, & Roscoe, 1988; Roscoe, 1987a, 1987b). The symbology appears to be closer than the scene (Norman & Ehrlich, 1986). Pilots have difficulty attending to both the HUD symbology and the real world scene and must switch their attention between the two sources of information. This may be attributable to cognitive influences on depth perception or to an inability to process information in a parallel rather than a serial manner (e.g., Fischer, 1979). However, the ability to divide attention between two activities presented on the same display improves with practice (e.g., Becklen & Cervone, 1983; Stoffregen & Becklen, 1989).

Fourth, when viewing both the HUD symbology and a real world scene or a synthetic image, there is a tendency to fixate on one source of information (Fischer, Haines, & Price, 1980; Larish & Wickens, 1991). Thus, pilots experience difficulty detecting unexpected events in the symbology or the external scene, especially when under stress or high levels of workload. The variables that affect attentional fixation or cognitive capture are not well understood. However, attentional and cognitive factors appear to be equal to or more important than sensory and perceptual factors.

Fifth, many findings from HUD research may generalize to HMDs, such as the difficulty pilots have dividing attention between the symbology and the external scene. However, the effects of characteristics that are unique to the HMD (e.g., proximity to the pilot's eyes, slaving to the pilot's head, and the potential for monocular or binocular symbology and scene imagery) have not been adequately investigated.

Finally, superimposing symbology on the NVGs has the potential for decreasing individual and crew workload and for increasing the safety of NVG flight. Most of the problems reported in evaluations of prototype systems were attributed to improper fit, adjustment, or operation of the ANVIS-HUD system (Runyon, 1985; Simmons, Kimball, & Hamilton, 1985; U.S. Marine Helicopter Squadron One, 1989). Perceptual and attentional problems were only cited indirectly. Either there were few perceptual or attentional problems or their effects were small relative to fit, adjustment, or operational problems.

Review of Cognitive Factors

Morey and Simon (1991a) concentrated their review on cognitive and attentional factors associated with HUDs and HMDs. They were especially concerned with research and theory related to divided attention, time sharing ability, and selective attention. Their major conclusion was that pilots tend to process NVG imagery and superimposed symbology in series, even though both types of information are presented at the same focal distance to facilitate parallel processing (e.g., Brickner, 1989; Foyle, Sanford, & McCann, 1991; Neisser & Becklen, 1975). That is, the aviators must divide their attention between monitoring the scene imagery and the instrument symbology. They cited evidence that superimposed dual tasks were not performed as effectively as the single tasks (Damos, 1991). The literature did not support the concept of a time-sharing ability (e.g., Ackerman, Schneider, & Wickens, 1984) but did indicate that dual-task performance can be improved with practice (e.g, Hirst & Kalmar, 1987).

To use the ANVIS-HUD information effectively, aviators must selectively attend to dynamic, relevant stimuli without becoming fixated on one aspect of the display. Morey and Simon (1991a) described the major differences in the types of stimuli in the ANVIS-HUD, which have the potential for differentially attracting an aviator's attention. The scene imagery is characterized by multiple objects containing variable features and is located in transient positions. In contrast, the instrument symbology is characterized by a fixed set of objects with invariant features occupying invariant positions. They cited research indicating that aviators tend to become fixated on certain elements of the display, especially the symbology (e.g., Foyle et al., 1991; Weintraub, Haines, & Randle, 1985).

Morey and Simon (1991a) also reviewed literature about individual differences that may affect an aviator's ability to attend to the scene imagery and symbology presented in the ANVIS-HUD. They cited several studies indicating that individual differences in ability, experience, and training may affect aviator performance (e.g., Damos, 1991; Johnston, Hawley, & Farah, 1988; Kyllonen & Woltz, 1989). In addition, they identified two less obvious factors that may be relevant to ANVIS-HUD performance: hemispheric lateralization of brain function and eye dominance.

Verbal and analytical information is processed primarily in the left hemisphere and visual and spatial information is processed primarily in the right hemisphere. Because the two types of information in the ANVIS-HUD may be processed separately by the two hemispheres, Crowley (1989) has hypothesized that aviators who are well lateralized may perform better than aviators whose brain hemispheres are not as specialized. Both Crowley and Ashcraft (1989) reported that right-handed individuals demonstrate more lateralization of function than left-handed individuals. Crowley also reported some evidence of cerebral dominance in which one hemisphere, and its associated functions, predominate over the other. Right-handed individuals tend to exhibit left-brain dominance; the opposite occurs for left-handed individuals.

The potential effects of eye dominance are related to the dichoptic viewing conditions used with the ANVIS-HUD. Dichoptic viewing occurs when disparate images are presented to the two eyes. Gopher, Grunwald, Straucher, and Kimchi (1990) found that responding to dynamically changing images produced a performance decrement when the images were viewed dichoptically. In the ANVIS-HUD, one eye views only the scene imagery while the other eye views both the scene imagery and the symbology. When disparate information is presented to the two eyes, the view from one eye predominates. Porac and Coren (1976) identified numerous tests of eye dominance, but found they could be categorized as sighting, sensory, and acuity dominance. Sensory dominance was generally observed in binocular rivalry situations in which monocular images alternated in consciousness. However, they concluded that sighting dominance was the best documented of the three forms and the most highly correlated with subject behavior.

Research Objectives

The U.S. Army is acquiring a HUD applique for use with ANVIS goggles. The ANVIS-HUD system has two major advantages: It enables aviators to monitor flight instruments without having to look under or around the goggles and it reduces the need for intensive coordination between the crewmembers. Both these effects should reduce crew workload and increase mission performance and safety. However, the research literature indicates that superimposing symbology over the NVG scene imagery may negatively affect the aviator's ability to monitor and respond to both types of information. The literature also suggests that a number of factors may influence an aviator's ANVIS-HUD performance. However, very little research has been conducted that directly addresses these issues.

Therefore, this research presented simulated NVG imagery and instrument symbology to rotary wing aviators to investigate the following five basic questions about ANVIS-HUD use.

- How well can aviators monitor and respond to scene and symbology information when viewed separately?

- What are the effects on aviator performance and perceived workload when NVG imagery and instrument symbology are presented simultaneously?
- Is aviator performance and perceived workload affected by whether they are trained on scene-only information or symbology-only information first?
- Does ANVIS-HUD performance improve with practice?
- Is ANVIS-HUD performance related to individual differences in experience, eye dominance, or cerebral laterality?

Method

Overview of the Research Design

The subjects in this experiment participated in five test sessions presented under four different conditions. Sessions 1 and 5 were designed primarily to measure the effects of practice on the subjects' ability to monitor and recall scene and symbology information presented simultaneously. Sessions 2, 3, and 4 were designed to determine whether the subjects' ability to monitor either scene or symbology information was degraded when both types of information were combined. In sessions 2 and 3, the subjects were required to monitor and respond to only the scene or only the symbology information. In session 4, they were required to monitor and respond to a similar flight scenario that presented both scene and symbology information.

The conditions represented counterbalances in (a) the sequence of presentation for two traffic pattern scenarios (TPA and TPB) in sessions 1 and 5, (b) the order of training and testing the symbology only and the scene only sessions, and (c) the sequence of presentation for two low level flight scenarios (LLA and LLB) in test sessions 2, 3, and 4 (see Table 1). The same scenario was used in test sessions 2 and 3, but the subjects were not aware the scenarios were identical because only one type of information was available in each test session.

Personnel

Subjects

Requests for volunteers were sent to units and organizations at the U.S. Army Aviation Center, Fort Rucker, Alabama. The units and organizations were selected to produce highly experienced aviators, including flight instructors, and recent graduates from flight training. They were also selected to produce aviators who fly a variety of

Table 1

Test Session Counterbalances for the Four Experimental Conditions

Cond	TS1	TS2	TS3	TS4	TS5
1	TPA	ScOn-LLA	SyOn-LLA	ScSy-LLB	TPB
2	TPB	SyOn-LLA	ScOn-LLA	ScSy-LLB	TPA
3	TPB	ScOn-LLB	SyOn-LLB	ScSy-LLA	TPA
4	TPA	SyOn-LLB	ScOn-LLB	ScSy-LLA	TPB

Note. Cond = condition; TS = test session; TPA and TPB = traffic pattern scenarios A and B; ScOn = scene only information; SyOn = symbology only information; ScSy = scene plus superimposed symbology information; LLA and LLB = low level flight scenarios A and B.

helicopters in the Army inventory that are likely to use the ANVIS-HUD. The request specified that the volunteers be rated and current helicopter pilots who were qualified in NVGs but who had no experience with helmet mounted symbology. The latter requirement eliminated AH-64 pilots from participating. Volunteers were scheduled to participate in the experiment by their commanders.

Data were collected from 43 subjects during the experiment, but the data from only 36 subjects were included in the analyses. Five subjects were used to pretest the apparatus and procedures. Two other subjects did not receive adequate notice of their scheduled experimental sessions. Although they agreed to participate in the experiment, their performance was hampered by a lack of sleep and concern for other commitments. They were deleted from the data base and two additional subjects were recruited to replace them. All the subsequent data are based on the 33 male and 3 female subjects included in the analyses.

The subjects were all helicopter pilots (35 U.S. Army and 1 U.S. Air Force). Seven aviators were Department of the Army civilian flight instructors, 7 were Army aviation chief warrant officers, and the remaining 22 were commissioned officers (Captain or below). The civilian aviators all had prior Army aviation service. The primary aircraft for the aviators were UH-1 (15), UH-60 (9), OH-58 (7), and AH-1 (5).

The median age of the aviators was 27.5 years, with a range from 22 to 49 years. The median years as a rated aviator was 1.75, with a range from having just graduated from initial entry rotary wing training to 25 years. Their number of flight hours and NVG hours was similarly skewed. The median flight hours was 350, with a range of 130 to

10,000 hours, and the median NVG hours was 42.5, with a range from 15 to 1,500 hours. All the experience variables (age, years rated, flight hours, and NVG hours) were significantly correlated with each other and with wearing glasses (older aviators were more likely to wear glasses).

Eight aviators were required to wear corrective spectacle lenses to fly and during the experiment. On an eye dominance test, 23 aviators were categorized as right eyed and 13 were categorized as left eyed. Four aviators considered themselves to be left handed and had scores ranging from 1.33 to 2.33 on a handedness inventory scale where 1 = left and 5 = right. The other 32 considered themselves right handed and had handedness scores ranging from 3.83 to 5.00. There were no significant correlations between dominant hand and eye or between dominant eye and wearing glasses.

Nine aviators were assigned to each of the four experimental conditions using a block randomization process. Within experience (high vs. low) and primary aircraft groups, aviators were assigned to conditions in randomly chosen order with the restriction that a condition could not be repeated until all other conditions had been administered an equal number of times. In addition, the three female aviators (who were all low experience pilots in either the UH-1 or UH-60 helicopter) were assigned to different conditions. There were no significant correlations between experimental condition, training order, or scenario sequence and any of the demographic variables.

Experimenters

Four experimenters participated in conducting the sessions. Two experimenters were behavioral scientists who had 9 and 13 years experience conducting research with Army aviators. One experimenter was an active duty Army aviator who was working on a graduate degree in the behavioral sciences. The fourth experimenter was a retired Army aviator who had 3 years experience conducting aviation research.

Apparatus

The four types of equipment used in the experiment are described in this subsection. The first type is the optical equipment used to present the simulated ANVIS-HUD scenarios to the aviators. The second type is the computer equipment used to control the experiment and to collect aviator responses. The third type is the recording equipment used to play the videodisc and videotape scenario imagery. The final type is the equipment the aviators used to respond to the scenario stimuli.

Optical Equipment

An optical apparatus was designed and constructed to simulate the scene and symbology stimuli that are seen through an ANVIS-HUD. All stimuli were viewed biocularly in a front-surfaced mirror haploscope at a distance of 26.5 cm (see Figure 4). To simulate the collimation used in NVGs, +3.25 diopter lenses were inserted into each ray path. The lenses placed the apparent viewing distance of the displays at 1.9 m, which is the approximate viewing distance that aviators adjust the ANVIS goggles to in the laboratory and in the field (W.E. McLean, personal communication, May 27, 1992). Convergence was adjusted for normal viewing at this distance. Vertical and horizontal nonius lines were used to check the convergence in the device prior to each experimental session; all the aviators reported a fusion of the images.

Two sets of chromatic filters were inserted into the ray paths to simulate the display color of NVGs. The low optical quality of the yellow and blue mylar filters also reduced the spatial resolution of the displays to a more realistic simulation of NVGs. The field of view was restricted to 40 degrees; the display edges were hidden by a pair of opaque masks with circular apertures that measured 17.78 cm in diameter. A septum that extended from the aviator's nose to the apex of the mirrors prevented cross-path viewing. An opaque shroud prevented ambient light from entering the device and segregated the light in each path. The aviators also viewed the stimuli in a darkened experimental room. A forehead rest and an adjustable chin rest were used to center

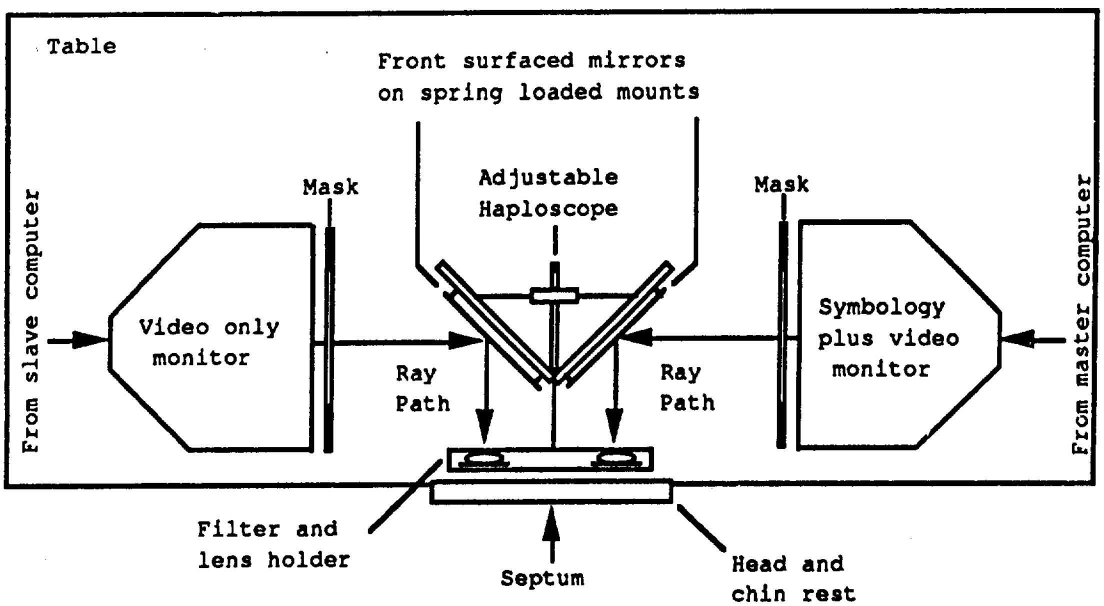

Figure 4. Schematic diagram of the optical apparatus.

the aviator's line of sight and to stabilize the aviator's head.

Computer Equipment

Hardware. The primary computer hardware in the simulated ANVIS-HUD apparatus were two Gateway 2000 33 Megahertz 386 computers, each with 4 megabytes of random access memory, and a 120 megabyte hard drive. The computers worked in a master-slave relationship (see Figure 5). The master computer controlled all aspects of the experiment, including presenting instructions, drawing the symbology, collecting aviator performance data, and controlling all devices in the system. The slave computer only carried out commands from the master computer.

Each computer was equipped with a digital video-overlay card (Super VideoWindows from New Media Graphics) capable of overlaying 800 X 600 resolution VGA images. The scene and symbology images were presented at 12 square pixels per degree of visual angle. The cards were set to output identical monochrome video images on both computers. Two Gateway 2000 CystalScan 1024NI Super VGA monitors were used to present the scene imagery and instrument symbology to the aviators. The experimenter observed the scenarios on an NEC MultiSync monitor. A Mintec VGX123 splitter (Minitronics Video Technologies) was used to deliver the signal from the master computer to the right (symbology plus video imagery) monitor and to the experimenter monitor (see Figure 5).

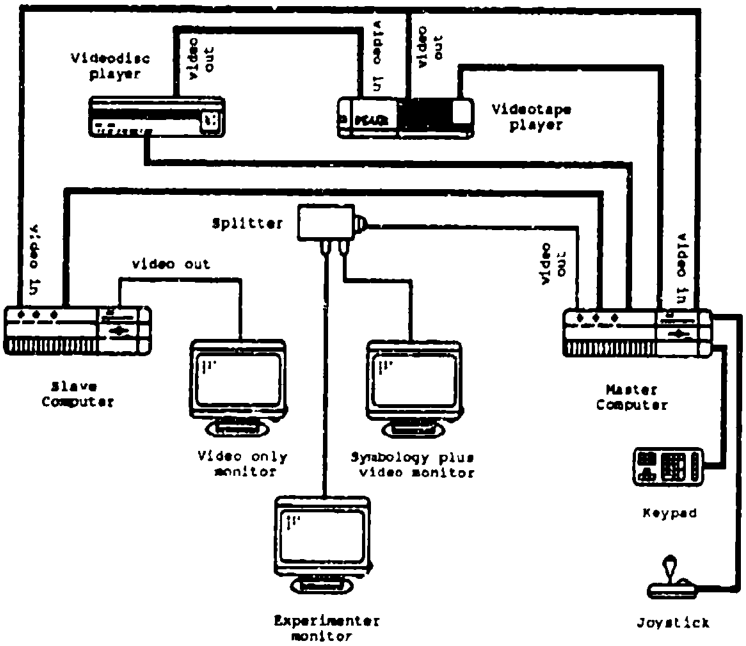

Figure 5. Schematic diagram of the computer hardware.

The brightness of the video imagery was adjusted so that the luminance of the terrain features was between approximately 5.1 and 6.9 cd/m^2, which is approximately the luminance level of the ANVIS goggles on an average night (W.E. McLean, personal communication, May 27, 1992). As a result, the luminance of the sky was between approximately 18.8 and 22.3 cd/m^2. The brightness of the symbology was adjusted so that it could be read against the sky background without being too intense against the terrain background.

Software. Custom software was written for the experiment using Borland International Turbo Pascal 6.0 and the Genus Microprogramming PCX Programmer's Toolkit. The software was written to present the instructions for each part of the experiment, to control the scenario tape presentations, to generate the symbology, and to collect aviator performance data. The program also controlled the training order and scenario sequence counterbalances for each experimental condition. Because the aviator monitors were viewed through mirrors, all the instruction screens and symbology were horizontally reversed.

Figure 6 shows an example of the instructional script. The letter I in the left hand column indicates an instruction screen presented to the aviator. Instruction screens always ended with directions for starting or continuing the session. The letter E indicates a command to the master computer to run the first segment of scenario 97 (TPA), which was located on the videotape between frames 1240.00 and 1295.00, with both scene and symbology information presented. The next instructional screen directs the aviator to report the information that was in view when the screen went blank.

Scenario Presentation Equipment

Two media were used to present the visual imagery. The practice scenarios were adapted from previously developed videodisc footage and were presented using a Pioneer LD-V6000A videodisc player. The test scenarios were videotaped for this experiment and were presented using an NEC PC-VCR videotape recorder. The output from the videodisc player was routed through the videotape player (see Figure 5) to simplify the control of the scenario presentation by the master computer.

Response Recording Equipment

Three types of recording equipment were used in this experiment. The aviators responded to symbology out-of-tolerance states with a numeric keypad (NumeriKeys by Genest Technologies). They responded to designated scene objects and events by pressing a joystick button (Winner 909 by Contrive Technologies). The master computer polled the

I In this part of the experiment, you will view both the scene and the symbology. During this scenario, you will be taking off from a landing zone, flying a traffic pattern, and making an approach back to the same landing zone.

A number of times during the scenario, the screen will go blank and you will be asked to recall your airspeed, altitude, trim, and master caution conditions and to describe objects (natural or manmade) and events (e.g., leg of the traffic pattern) in the scene.

Treat the scene and symbology as equally important.

Press the "1" key to begin the scenario.

E 97 1240 1295 BOTH

I At this time, tell the experimenter
(a) the four instrument values,
(b) any relevant scene objects, and
(c) any relevant scene events.

Press the "1" key to continue the flight.

Figure 6. An example of the instructional script.

keypad and joystick for presses at a rate of 60 Hz. The data indicated which key or button was depressed on each scenario frame. Scene callouts were recorded on a Realistic CTR-75 cassette tape recorder.

Materials

The three types of experimental materials used in this experiment are described in this subsection. The first type described is the scenarios used in the practice trials and test sessions. The second type is the general instructions that provided standardized information to each aviator and standardized directions for each experimenter. The final type of materials is the scoring instructions.

Scenarios

The scenarios used for the practice trials and test sessions were composed of visual imagery, instrument symbology, or both. The following two subsections describe the scene content and instrument symbols presented to the aviators. The first subsection describes the visual imagery for the practice trials, two traffic pattern scenarios, and two low level flight scenarios. The second subsection describes the instrument symbology used with all the scenarios except the scene only sessions. The subsections also describe the targets (scene features and events and symbology out-of-tolerance states) the aviators were instructed to monitor and respond to, the method of responding, and the data collection forms.

Visual imagery. Two media were used to present visual imagery in this experiment. There were 5 practice trials for test sessions 2, 3, and 4. The practice imagery was adapted from videodisc training scenarios developed for an Army aviation map interpretation and terrain analysis course (Miles & LaPointe, 1986). The scenes were of nap-of-the-earth flight recorded over central Germany during the summer. The footage was flown at altitudes ranging from 5 to 50 feet above ground level and at airspeeds ranging from 35 to 50 knots. The scene targets included roads, fields, streams, lakes, telephone poles and wires, and electric power lines and stanchions.

The practice trials averaged 59 s in length, with a range of 20 to 98 s. The displays were refreshed at a rate of 10 Hz. One set of five trials was presented as scene only and one set was presented as both scene and symbology. The third set did not display any visual imagery, but the symbology corresponded to the airspeed, altitude, and heading of the videodisc scene imagery.

The four test scenarios were videotapes of UH-1 flights recorded specifically for this research in Southeastern Alabama during the autumn. The videotape frames were displayed at a rate of 30 Hz.

Two of the test scenarios depicted rectangular traffic pattern flights (i.e., TPA and TPB). The flights began with a takeoff from a landing zone, which was an open field. The helicopter climbed and accelerated to an approximate airspeed of 55 knots and an approximate altitude of 155 feet above ground level as it progressed through the crosswind and downwind legs of the pattern. The helicopter descended and decelerated during the base leg and on final approach to the original landing zone. The scenario ended shortly before the helicopter landed.

As the helicopter was on its final approach, another helicopter was visible in the landing zone to the left of the flight path. In TPA, the other helicopter was visible for approximately 34 s and took off across the flight path of the helicopter from which the scene was viewed (called the ownship). In TPB, the other helicopter was in view for approximately 7 s; it was running but did not take off before the scenario ended. The helicopter in both scenarios had its anticollision light turned on.

The traffic pattern scenarios were designed to measure situational awareness. Twice during each scenario, the screen was blanked and the aviator was instructed to report four instrument values, any scene features, and any scene events (such as the leg of the traffic pattern). In TPA, the screen was blanked at 55 s and at 216 s into the flight. Each aviator's responses were recorded on a situational awareness form for TPA (see Figure 7). There were four scene targets in view at the first scenario break point and five scene targets at the final break.

In TPB, the screen was blanked at 50 s and at 173 s into the flight. Each aviator's responses were recorded on the situational awareness form for TPB (see Figure 8). There were three scene targets presented at the first scenario break point and six scene targets at the final brea...

The other two test scenarios depicted low level flights (i.e., LLA and LLB). LLA and LLB began and ended in flight. The routes were flown in an essentially straight and level profile at an average indicated airspeed of 40 knots and an average radar altitude of 100 feet above ground level. There were only minor changes in heading, airspeed, and altitude. The low level scenarios depicted flight over rural areas consisting primarily of forests, fields, and small bodies of water. The scenarios also depicted some manmade features such as roads, trails, and buildings. In both scenarios, the ownship was following another helicopter, which was called the leadship.

The aviators were required to monitor and respond to the natural and manmade terrain features and to define maneuvers by the leadship, which were called events. The maneuvers were the leadship reappearing in the field of view after an absence of 3 s or more, the leadship crossing in front of the ownship, and the leadship fuselage being profiled above the horizon. The required responses when a target was detected were to press the joystick button held in the left hand and to call out the name of the target. Each aviator's responses were recorded by the experimenter on a callout form and on a tape recorder.

LLA lasted for 342 s and contained 29 scene targets. Twelve targets were scene features and 17 were leadship events. The LLA scene callout form (see Figure 9) shows the scene targets and their times of appearance. Except for the leadship appearances, the form also shows the length of time each target was in view.

LLB lasted for 588 s and contained 40 scene targets. Only eight of the targets were leadship events, all of which were the leadship breaking above the horizon. The LLB scene callout form shows the order of presentation and length of

Situational Awareness Form

Traffic Pattern Scenario A   Subject: ________

Frame 1295 (55 seconds)

Instrument Values

Airspeed (42): __________   Altitude (100): __________

Trim (In):   In _____   Out _____

Master caution (Off):   On _____   Off _____

Terrain Features

Chicken houses _____   Field _____   Trees _____

Scene Events

Leg (Crosswind) ________________

Frame 1456 (216 seconds)

Instrument Values

Airspeed (08): __________   Altitude (23): __________

Trim (In):   In _____   Out _____

Master caution (Off):   On _____   Off _____

Terrain Features

Helicopter _____   Field _____   Trees _____

Scene Events

Leg (Final) ______________   Helicopter takeoff _____

Figure 7. Situational awareness form for traffic pattern scenario A.

Situational Awareness Form

Traffic Pattern Scenario B   Subject: ________

Frame 1595 (50 seconds)

Instrument Values

Airspeed (35): __________ Altitude (133): __________

Trim (Out): In _____ Out _____

Master caution (Off): On _____ Off _____

Terrain Features

Field _____ Over Trees _____

Scene Events

Leg (Crosswind or crosswind to downwind) ______________

Frame 1718 (173 seconds)

Instrument Values

Airspeed (16): __________ Altitude (63): __________

Trim (In): In _____ Out _____

Master caution (On): On _____ Off _____

Terrain Features

Helicopter ____ Field ____

Over trees ____ Lone Tree ____

Scene Events

Leg (Final) ____________ Helicopter running _____

Figure 8. Situational awareness form for traffic pattern scenario B.

Scene Callout Form

Low Level Scenario A

Subject: ______

_____ LS appears (0:12)

_____ LS appears (0:26)

_____ LS appears (0:33)

_____ LS crosses (0:37-0:39)

_____ LS too high (0:42-0:47)

_____ Field (0:58-1:50)

_____ LS appears (1:02)

_____ Structure (1:25-1:47)

_____ LS appears (2:18)

_____ LS crosses (2:20-2:21)

_____ Field (2:25-2:37)

_____ Road (2:32-2:36)

_____ LS appears (2:54)

_____ LS appears (3:04)

_____ Stream Bed (3:14-3:40)

_____ LS appears (3:21)

_____ LS appears (3:38)

_____ LS crosses (3:40-3:41)

_____ LS appears (3:48)

_____ Stream Bed (3:50-4:02)

_____ Field (4:11-4:25)

_____ Field (4:37-4:44)

_____ Field (4:53-5:00)

_____ LS appears (4:50)

_____ LS crosses (4:58-4:59)

_____ LS too high (5:02-5:03)

_____ Trail (5:07-5:14)

_____ Structure (5:16-5:23)

_____ Clearing (5:33-5:42)

Figure 9. Data collection form for low level scenario A (LS = leadship).

time in view for the scene features and events (see Figure 10). Although there were substantial differences in the proportions of scene features and events in the two scenarios, the rate of occurrence of scene targets was very similar: A scene target appeared on an average of every 12 s in LLA and every 15 s in LLB.

Symbology. The planned ANVIS-HUD symbology suite varies among aircraft and flight modes. Using materials provided by the U.S. Army Aviation Systems Command, St. Louis, Missouri, a symbology suite was designed for this experiment that is a composite of 12 primary instruments included in most versions

Scene Callout Form

Low Level Scenario B   Subject: _____

_____ Three fields (0:00-0:12)
_____ LS too high (0:38-1:01)
_____ Lake/pond (1:03-1:14)
_____ Field (1:08-1:24)
_____ Dirt Road (1:14-1:18)
_____ LS too high (1:16-1:22)
_____ Field (1:29-1:37)
_____ Field (1:32-1:49)
_____ Field (1:49-2:18)
_____ Trail (2:20-2:26)
_____ Trail (2:28-2:30)
_____ Trail (2:38-2:40)
_____ Road (2:40-2:46)
_____ LS too high (2:41-2:44)
_____ LS too high (3:05-3:22)
_____ LS too high (3:25-3:56)
_____ Lake/pond (3:37-3:46)
_____ Field (3:47-4:24)
_____ Trail (4:09-4:20)
_____ Dirt road (4:21-4:25)
_____ Road (4:26-4:29)
_____ LS too high (4:26-4:33)
_____ Field (4:37-5:05)
_____ Field (4:49-4:56)
_____ Stream (5:09-5:12)
_____ Stream (5:15-5:17)
_____ Field (5:34-5:46)
_____ Field (5:40-6:08)
_____ Clearing (6:46-6:59)
_____ Trail (7:31-7:35)
_____ Road (7:39-7:41)
_____ Field (7:45-8:21)
_____ Trail (8:01-8:06)
_____ Road (8:13-8:17)
_____ Lake/pond (8:39-8:59)
_____ Dirt road (8:45-9:00)
_____ LS too high (9:01-9:07)
_____ LS too high (9:17-9:22)
_____ Field/trail (9:25-9:28)
_____ Field (9:35-9:41)

Figure 10. Data collection form for low level scenario B (LS = leadship).

of the ANVIS-HUD (see Figure 11). The instrument symbols were located in the approximate positions they occupy in the ANVIS-HUD suite. The only intentional change in position was to locate the master caution warning (MST) light below the

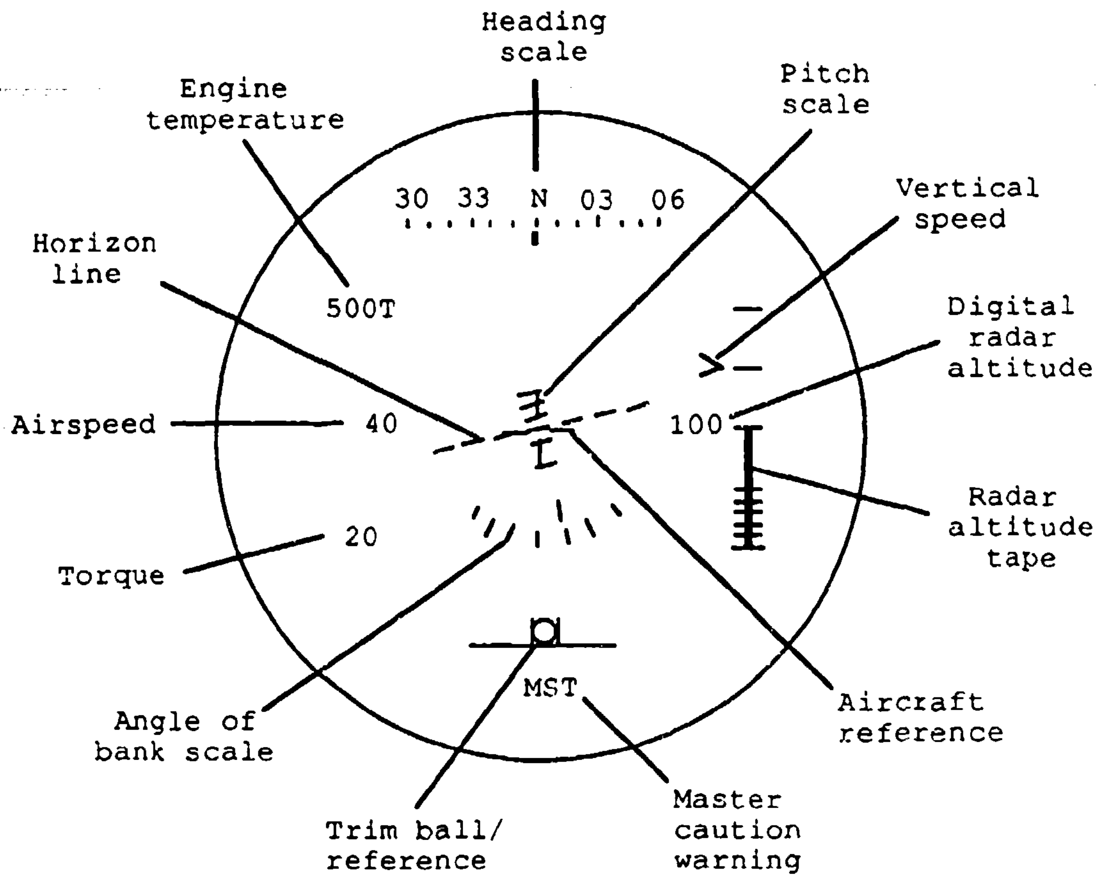

Figure 11. Symbology suite used to display instrument information.

trim ball in the composite suite rather than on the left side of the trim ball reference line. The change was made so the two symbols would not overlap if the MST light was on and the aircraft was out of trim to the left.

A computer program was written to drive the symbols during each practice and test scenario so that the instrument values would be coordinated with the visual imagery and with each other (e.g., the digital and tape radar altitudes). For example, the airspeed, altitude, and torque instruments were all at zero before takeoff in the traffic pattern scenarios. After takeoff, the instruments increased to reflect the apparent changes in airspeed and altitude. Similarly, the heading, horizon line, pitch, and bank symbols all moved in a coordinated fashion during aircraft turns. The helicopter was most likely to be out of trim during turns, but the program generated some random out-of-trim conditions if there were few turns in a scenario. Only the MST light was turned on randomly.

Although all the symbology information was available to the aviators, they were only required to monitor and respond to four instruments: airspeed, altitude, trim, and MST light. During the traffic pattern scenarios, the aviators were instructed to monitor the instruments and to recall the airspeed and altitude values and the trim and MST status when the screen went blank. The aviators' responses were recorded on the appropriate situational awareness forms (see Figures 7 and 8).

During the low level flight scenarios, the aviators were instructed to monitor the four instruments and to respond to each instrument that exceeded its defined in-tolerance state (see Table 2) by pressing the appropriate key on the numeric keypad (see Figure 12). The key press responses were selected to approximate the spatial location of the instruments in the symbology suite (i.e., airspeed on the left, altitude on the right, trim ball in the center, and MST light below all other instruments). The tolerance states were the same for the practice and test sessions except for the altitude levels, which were adjusted to reflect the apparent height above ground level in the two sets of scenarios. However, the width of the altitude tolerance band was held constant.

The average rate of out-of-tolerance states for each instrument was held constant for all scenarios. The airspeed and altitude values exceeded their tolerance limits an average of three times per minute; the trim ball and MST were out of tolerance an average of once per minute. The

Table 2

In-tolerance States for the Four Monitored Instruments

Instrument	In-tolerance state
Airspeed indicator	40 ± 5 knots
Altitude indicator	25 ± 10 feet (practice) 100 ± 10 feet (test)
Trim ball	Less than one ball width outside the reference lines
Master caution warning (MST)	Off (not visible)

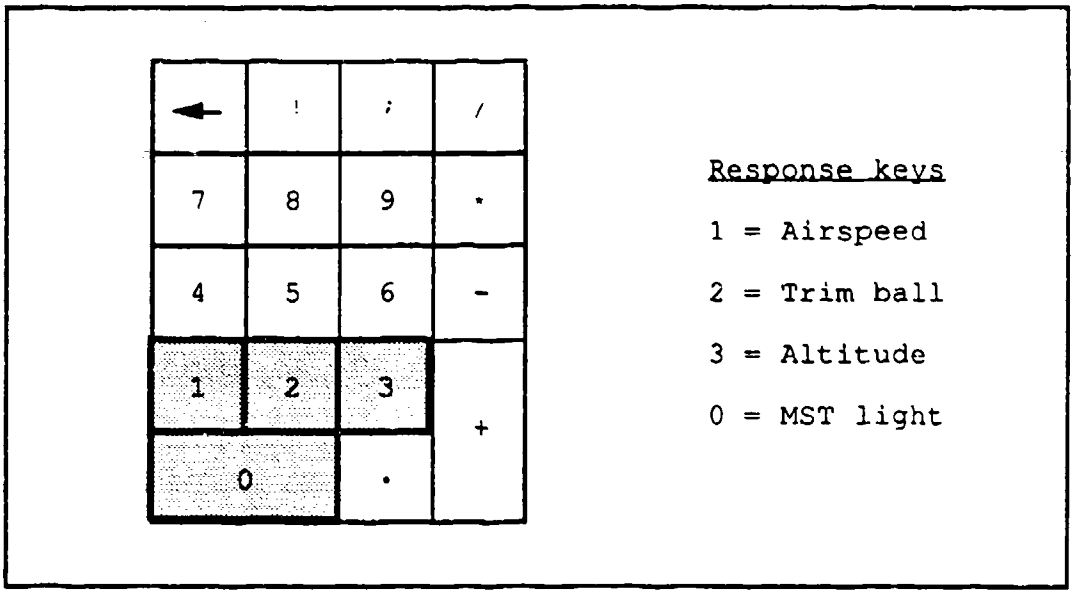

Figure 12. Numeric key press codes for responding to instrument out-of-tolerance states (MST = master caution warning).

instruments were programmed to be out of tolerance for a maximum of 3 to 6 s. During the test scenarios, the instruments returned to an in-tolerance state as soon as the aviator pressed the appropriate key. During the practice trials, which were generally flown at very low altitudes, the altitude indicator was coordinated with apparent height and did not change in response to aviator key presses. Each aviator's key presses were recorded by the master computer.

General Instructions

An instructional packet was prepared to guide the experimenter through the steps of the experiment and to provide standardized directions that the experimenter read to each aviator. The packet began with an overview of the research purpose and procedures. The next section of the instructions provided guidance for collecting administrative information (informed consent, performance feedback information) and background data (see Figure 13), and administering a handedness inventory (see Figure 14) and an eye dominance test (Morey & Simon, 1991b).

The next part of the instructions described the National Aeronautics and Space Administration (NASA) Task Load Index (TLX) workload rating scales (NASA, 1986). A sheet describing the six scales and a packet of five rating forms were provided for the aviators to study until they indicated

Background Information

Subject: ______ Last Four: __________ Date: _____________

Rank: ____ Years rated aviator: ____ Age: ____ Sex: ____

Total Flight Hours: ___________ NVG Flight Hours: ___________

Primary Aircraft: ___________ Flight/NVG Hours: ______/______

Other Aircraft: ____

Experience flying with helmet-mounted symbology? Yes No

If yes, describe: ____

Vision: Left 20/____ Right 20/____

Wear glasses or contacts? Yes No

Vision problems (if known): ______________________________________

- -

Measured Preference Hand: ___________ Eye: ___________

Reported Preference Hand: ___________ Eye: ___________

Experimental Condition: 1 2 3 4

Comments: ____

____

Experimenter: _____________ Start time: _______________

Figure 13. Background information form.

Handedness Inventory

Name: ______________________________ Last four: ____________

Directions: Indicate your preferred hand for performing each of the following 12 activities by circling the appropriate response on the 6-point scale given below. Try to answer all the questions using categories 1 - 5; use category 6 only if you have no experience with the activity.

1 = Always use left hand
2 = Usually use left hand
3 = Use both equally
4 = Usually use right hand
5 = Always use right hand
6 = Don't know

With which hand do you:	Always left	Usually left	Both equally	Usually right	Always right	Don't know
1. Write	1	2	3	4	5	6
2. Hold a nail to hammer	1	2	3	4	5	6
3. Draw	1	2	3	4	5	6
4. Use a pair of scissors	1	2	3	4	5	6
5. Use a toothbrush	1	2	3	4	5	6
6. Use a knife to carve a turkey	1	2	3	4	5	6
7. Hold a bottle to uncap it (bottle hand)	1	2	3	4	5	6
8. Hold a match when striking it	1	2	3	4	5	6
9. Use a screwdriver	1	2	3	4	5	6
10. Pour a large volume of liquid from a pitcher	1	2	3	4	5	6
11. Throw a ball	1	2	3	4	5	6
12. Use a spoon	1	2	3	4	5	6

Figure 14. The handedness inventory.

they understood the rating procedure. The scale descriptions and an example of the rating form are presented in Appendix A.

The next two parts presented the instructions for monitoring scene and symbology information, including a description of the ANVIS-HUD symbology suite, and instructions for responding to the scene features and events and symbology out-of-tolerance states that were defined as critical in the experiment. The instructions also provided a procedure for practicing the responses to the four instrument symbols.

The final parts of the instructions were directions to the experimenter for adjusting the equipment for the aviator, starting the computer administered sessions, and collecting performance data. The general instructions ended with directions for debriefing the aviator, processing the computer generated and manual data, and preparing the equipment for the next session. A form was used to standardize the debriefing (see Figure 15).

Scene Scoring Sheet and Instructions

A set of directions and forms were prepared for manually scoring the instrument and the scene performance on the traffic pattern scenarios, the scene performance on the low level flight scenarios, and the workload ratings. The percentage of symbology out-of-tolerance states correctly detected (% det) and the reaction time (RT) to those targets were scored by a computer program.

The % det for the two traffic pattern scenarios was scored on the data collection form. The scores were calculated by determining the number of scene targets or the number of symbology values correctly reported divided by the number of targets of each type. Reported airspeed was considered correct if it was within 5 knots of actual airspeed; reported altitude was considered correct if it was within 10 feet of actual altitude.

For the low level flight scenarios, separate forms were developed for scoring the % det for scene targets and the RT to a subset of those targets (see Appendix B). The scoring procedures were developed empirically after data were collected from 24 aviators. Although the experimenter check sheets for LLA and LLB contained all possible scene targets, only those targets that were detected by at least 5% of the aviators and that were in view for at least 2 s were counted for % det (24 on LLA and 32 on LLB). Only targets detected by at least 50% of the aviators were considered in calculating RT (13 on LLA and 25 on LLB). Each RT score was the arithmetic mean of the targets that were detected by

Debriefing Form

Aviator: ______ Condition: ______ Experimenter: ______

Realism of visual imagery: ______

Realism of symbology: ______

Tendency to monitor one or the other: ______

Monitor other instruments: ______

Fixation: ______

Attend upper/lower field: ______

Monitor in periphery: ______

Momentary lapses: ______

Strategy: ______

______

Practice trials helpful: ______

Improvement across parts: ______

Eyestrain: ______

Double images: ______

Brightness/contrast: ______

Suggestions: ______

______

______

REMIND NOT TO DISCUSS THE EXPERIMENT WITH OTHER AVIATORS.

Figure 15. Debriefing form.

the aviator (i.e., a maximum RT or penalty time was not assigned to undetected targets).

The directions also included instructions for scoring the NASA TLX workload ratings. Although the NASA booklet (1986) requires that subjects perform a paired comparison task to determine subscale weights for calculating a total workload rating, subsequent research (Byers, Bittner, & Hill, 1989; Hamilton, 1992) has found that simply averaging the six subscale ratings produces almost identical results. Therefore, an arithmetic mean workload score was calculated from the six ratings for each test session.

Finally, a subject data form was developed for coding the demographic data, compiling the computer scored and manually scored performance data and workload ratings, and entering all the data into a computer data base.

Procedures

All the aviators participated in individual experimental sessions that were conducted during either a morning or an afternoon period lasting from 2 to 2.5 hours. Following introductions, the experimenter took the aviator to the 3.4 m X 4.3 m experimental room. The window in the room was covered with black felt to block external light sources and visual distractions. The door was closed and signs posted to prevent noise distractions.

Administrative Procedures

The experimenter read the general instructions to the aviator and discussed any aspects that were not clear. After the overview, the experimenter asked the aviator to sign the informed consent form and obtained a mailing address if the aviator requested performance feedback. He then collected self-reported background information and instructed the aviator to complete the handedness inventory. The experimenter scored the handedness inventory by summing the circled ratings (items 2 and 7 were reversed for scoring; see Figure 14) and dividing by the number of items circled. The experimenter then confirmed that the aviator's self-reported hand preference was consistent with the handedness inventory score.

Next, the experimenter administered the eye dominance test (Morey & Simon, 1991b). The experimenter stood 2.5 m from the aviator and held a pen at eye level in his right hand. The aviator was told to hold a 30.5 cm^2 black board with both hands and to raise the board until he or she could see the pen through a 2.5 cm diameter circle in the center of the board. The experimenter noted which eye could be seen through the circle and then asked the aviator to close first one eye and then the other and report whether the pen was still in view. The experimenter instructed the aviator to lower the board while he changed the pen to his left hand. They then repeated the procedure. The experimenter recorded on the background information sheet which eye the aviator used for sighting the pen. All the aviators were consistent in their use of one eye for sighting the pen.

After the eye dominance test, the experimenter read the instructions to the aviator for filling out the workload ratings and for monitoring and responding to the scene and symbology information. The aviator was given a copy of the workload scale definitions and the rating packet, a diagram of the symbology suite, and a form showing the symbology out-of-tolerance states and the appropriate keypad presses for each instrument. The aviator then practiced pressing the appropriate key as the experimenter called out the instruments in a block random order. The keypad press practice continued until the aviator had pressed eight correct keys in a row and indicated that he or she was comfortable with the response procedure.

The experimenter then placed the aviator in front of the viewing apparatus and adjusted the chair and chin rest height to a comfortable position with the aviator's eyes aligned with the center of the lenses. The experimenter turned off the overhead light and started the computer controlled script for the aviator's assigned condition.

Experimental Sessions

Before each test session, the experimenter told the aviator to read and follow the instructions on the screen and to ask any questions before beginning the test scenario. During the practice trials that preceded test sessions 2, 3, and 4, the experimenter provided feedback about any consistent errors the aviator made. The errors included failing to respond to scene or symbology targets, responding to nontargets, or using inappropriate response techniques (e.g., pressing the wrong key or calling out two scene targets but pressing the joystick button only once).

During the two traffic pattern scenario sessions and the two low level flight scenario sessions that included visual imagery, the experimenter recorded the aviator's responses on the appropriate check sheet. He also tape recorded the aviator's scene callouts on the two low level flight scenarios. After each test session, the aviator filled out the workload rating form.

After the final test session, the experimenter debriefed the aviator for 15 - 20 minutes. The experimenter recorded on the debriefing form the aviator's responses about each topic and any comments or suggestions the aviator made about the symbology suite and the experimental procedures. The experimenter thanked the aviator for participating and requested that he or she not discuss the details of the experiment with other aviators who may participate in subsequent experiments.

Scoring Procedures

A computer program was used to score the percentage of symbology targets correctly detected and the RT for each detected target. The traffic pattern performance data, the scene performance data, and the workload ratings were scored according to the scoring instructions previously discussed. All the demographic and performance data were compiled on a subject data form and then entered into a computer data base for analysis.

Results

Overview

The results are organized into seven subsections. The Overview explains the subdivisions and describes the general statistical procedures used in the data analyses. The second subsection describes the transformation of the performance data to eliminate a methodological confound. The third and fourth subsections present the analysis of the aviators' performance during the low level flights and traffic pattern flights, respectively. The fifth subsection presents the analysis of demographic variable effects on aviator performance. The sixth subsection presents the analysis of the aviator workload ratings, and the seventh presents a summary of the debriefing comments.

The Results subsections present three types of statistics. First, descriptive statistics were computed to describe the variables being analyzed. Wherever the underlying distribution allows them, means and standard deviations are used to describe the variables. However, medians and ranges are presented for skewed demographic variables (e.g., flight hours, years as a rated aviator) and frequencies are used to describe categorical variables (e.g., sex, rank, eye preference) and to summarize the debriefing results.

Second, inferential statistics were computed to determine which experimental factors significantly affected aviator performance. As part of the data transformation procedure, *t*-tests were used to determine if there were significant differences between the recoded variables. All the analyses except the debriefing summary employed fixed-effects, factorial analyses of variance (ANOVAs) to infer causality. For the low level flight data, the type of information (scene or symbol) and the view (only or both) were within-subjects, or repeated measures, variables. For the traffic pattern data, the type of information and the test session (initial or final test) were also repeated measures. For the workload ratings, the test session (initial test, scene only, symbology only, both, or final test) was a repeated measure.

The remaining ANOVA factors (e.g., experimental condition, age, flight hours, eye preference) were between-subjects, or group, variables. A median split was used to divide all but two of the demographic variables into equal groups for the analyses. Eye preference and whether the aviators wore glasses were categorical variables and were analyzed with unequal group sizes. When the main effects with more than two levels or the interactions were significant, Newman-Keuls tests ($p < .05$) were used to determine which ANOVA cells were significantly different (Winer, 1971).

Finally, correlations were used to analyze the relationships among the variables. In some cases, the correlations were used to supplement the ANOVA results or to aid in their interpretation. In one case (handedness), a correlation was used as the primary analysis because there were too few left handed aviators in the sample and a median split would have produced illogical groupings.

Data Transformation

The first analyses of the low level flight % det and RT and the traffic pattern % det were 2 X 2 X 4 ANOVAs with experimental condition as the grouping factor. In all three ANOVAs, there was a significant two- or three-way interaction with condition. To determine if the training order (scene only first or symbology only first) or the scenario sequence (LLA or LLB, TPA or TPB) components of the experimental condition counterbalance were causing the interactions, three four-way ANOVAs were computed with order and sequence as grouping variables. In all cases, the scenario sequence was responsible for the interaction. Training order was never a significant factor.

To examine this potentially confounding interaction further, the scene and symbology % det and RT variables were recoded to analyze for differences between LLA and LLB and between TPA and TPB (see Table 3). The aviators detected fewer scene targets in LLA than in LLB, but they responded more rapidly to the LLA targets that were detected. The low level flight scenarios were not different in symbology % det, but the aviators responded more slowly to the LLA symbology

Table 3

Performance Differences Between Scenarios on the Low Level and Traffic Pattern Flights

	Scenario A		Scenario B			
Performance variable	Mean	SD	Mean	SD	*t*	*p*<
Low level scene						
% det	62.4	15.5	74.4	11.6	-3.61	.01
RT	3.4	0.8	4.2	1.1	-3.62	.01
Low level symbology						
% det	82.7	13.4	79.7	8.8	1.12	ns
RT	1.8	0.4	1.5	0.2	4.50	.001
Traffic pattern % det						
Scene	63.9	17.7	56.8	15.4	2.36	.05
Symbology	90.3	12.0	73.7	18.9	4.54	.001

Note. The degrees of freedom = 35 for all *t*-tests; SD = standard deviation; % det = percent correctly detected; RT = reaction time in seconds; ns = not significant.

targets. The % det for TPA was significantly higher than for TPB for both scene and symbology targets.

Although the scenarios were completely counterbalanced across test sessions, the interactions caused by the differences in difficulty between the scenarios may confound the analysis of demographic effects, which were measured rather than manipulated. To eliminate the differences in means and standard deviations between the A and B scenarios, each scenario variable was standardized using a z score transformation. The z scores were then multiplied by the average standard deviation of the A and B scenarios and added to the average arithmetic mean of the A and B scenarios. This conversion returned the scores to their original metric to aid in interpreting the results. That is, the % det scores ranged from approximately 11% to 100% and the RT scores ranged from approximately 1 to 6.5 s. Finally, the scores were recoded into the original variables (e.g., scene only RT or scene both RT).

The transformed data were then reanalyzed with three- and four-way ANOVAs. All the significant effects were the same as with the original data except there were no significant interactions with experimental condition, training order, or scenario sequence. These results indicate the data transformation accomplished its intended goal of removing the effects of differences in scenario difficulty without changing the other effects. Thus, the transformed data were used in all subsequent aviator performance analyses.

Low Level Flight Performance

Detection

There were significant main effects for both the type of information presented (scene vs. symbology) and for the viewing condition (only vs. both) on the aviators' ability to detect scene and symbology targets. The aviators detected 81.2% of the symbology out-of-tolerance states but only 68.4% of the scene features and events, $F(1, 32) = 56.03$, $p < 0001$.

Similar percentages were observed in the only and both viewing conditions. The aviators detected an average of 80.7% of the targets when they were viewing only the scene or symbology. Their performance decreased to 68.9% when they viewed the scene and symbology together, $F(1, 32) = 63.38$, $p < .0001$. There was no interaction between information type and viewing condition.

Reaction Time

Similar results were obtained with the aviators' speed of reaction to scene and symbology targets during the low level flights. The aviators took more than twice as long to react to the scene features and events ($M = 3.8$ s) than to the symbology out-of-tolerance states ($M = 1.6$ s), $F(1, 32) = 301.00$, $p < .0001$.

The aviators also responded significantly faster when they were viewing only the scene or only the symbology than when they viewed the scene and symbology together, $F(1, 32) = 20.91$, $p < .0001$. However, the magnitude of the differences between the two viewing conditions was smaller than between the two types of information. The mean RT to the only condition was 2.5 s and the mean RT to the both condition was 3.0 s. There was no interaction between the type of information and the only versus both viewing condition.

Traffic Pattern Detection Performance

The type of information X test session X condition ANOVA produced significant main effects for each factor but no significant interactions. To determine which component of the condition variable (see the data transformation description) caused the main effect, the data were reanalyzed with a type X test session X training order X scenario sequence ANOVA. Both types of information, $F(1, 32) = 79.03$,

$p < .0001$, and test session, $F(1, 32) = 6.05$, $p < .05$, affected the aviators' ability to detect scene and symbology targets. The aviators detected 82.0% of the symbology out-of-tolerance states but only 60.4% of the scene features and events. The aviators detected 68.3% of the targets during the initial test session. Their performance improved to 74.0% during the final test session.

Training order was not significant but the scenario sequence was, $F(1, 32) = 9.72$, $p < .0001$. Averaged across the initial and final tests, the aviators detected more scene features and events when they were presented TPA first (76.0% detected) than when they were presented TPB first (66.3% detected). This result suggests that observing the UH-1 helicopter taking off during the final approach to the landing zone on TPA may have sensitized the aviators to scan for helicopters during the approach on TPB. A frequency tabulation of the number of aviators who detected the helicopter in the landing zone on each scenario supports this interpretation. All 36 aviators detected the presence of the helicopter in TPA but only 9 of them detected it in TPB. Of the 9 detections on TPB, 8 occurred when TPB was shown in the final test session.

Demographic Variable Effects

Three sets of analyses were conducted to determine if aviator performance was affected by demographic factors. The first set included four variables that indicated the aviators' level of experience: age, years as a rated aviator, number of flight hours, and number of NVG hours. The second set included two variables related to aviator vision: whether the aviator wore glasses and the aviator's dominant eye. A final analysis evaluated the relationship between handedness and performance.

Experience Effects

The aviators were intentionally recruited to have a wide range of experience. For the ANOVAs, each variable was divided into low and high experience groups of 18 aviators each. Table 4 shows the median, minimum, and maximum value for each group on each of the experience variables. There are large differences in the median values for each group, although the maximum value for the low experience group is generally close to the minimum value for the high experience group.

Low level flight performance effects. The ANOVAs did not indicate any effects of experience on the aviators' ability to detect scene or symbology targets. Similarly, prior flight hours and NVG hours had no effect on the

Table 4

Descriptive Statistics for the Low and High Experience Groups

Variable	Low experience group Mdn	Min	Max	High experience group Mdn	Min	Max
Age in years	24.0	22.0	27.0	34.5	28.0	49.0
Years rated	0.1	0.1	1.5	8.0	2.0	25.0
Flight hours	167	130	280	2,550	420	10,000
NVG hours	23	15	35	500	50	1,500

Note. Mdn = median; min = minimum; max = maximum; NVG = night vision goggle.

aviators' speed of reaction to the targets. In the analyses of aviator RT, there was a significant interaction between age and type, $F(1, 34) = 5.48$, $p < .05$, and between years as a rated aviator and type, $F(1, 34) = 6.30$, $p < .05$. In both analyses, Newman-Keuls tests indicated the older and more experienced aviators reacted significantly faster to the scene targets than the younger aviators. The mean RTs were approximately 4.0 s and 3.6 s for the low and high groups for both age and years rated. There was a trend in the opposite direction for reacting to symbology targets (i.e., the younger aviators reacted faster), but the differences were not statistically significant.

It is not surprising that the age and years rated effects are almost identical: The correlation between the two experience variables is .95, $p < .0001$. However, when the full range of the experience variables were evaluated (i.e., not split into high and low groups), none of the relationships between experience and scene performance were significant but most of the correlations were significant between the experience variables and symbology performance. As shown in Table 5, the younger aviators with fewer years experience and fewer flight and NVG hours generally performed better than the older and more experienced aviators in detecting and reacting to symbology targets under the only and both conditions.

Traffic pattern performance effects. The ANOVAs did not indicate any effects of experience on the aviators' ability to detect scene or symbology targets during the traffic pattern scenarios. In addition, none of the correlations between experience and traffic pattern performance were significant.

Table 5

Correlations Between Experience Variables With Symbology Detection and Reaction Time Performance

Performance variable	Age in years	Years rated	Flight hours	NVG hours
Symbology only				
% det	-.46	-.38	-.36	ns
RT	.53	.55	.51	.49
Symbology both				
% det	-.41	-.39	ns	-.34
RT	.45	.52	.44	.44

Note. $N = 36$ for all correlations. Correlations > .33 are significant at $p < .05$; correlations > .42 are significant at $p < .01$ (see Roscoe, 1975, p. 265). NVG = night vision goggle; % det = percent correctly detected; RT = reaction time; ns = not significant.

Vision Effects

Corrective glasses. There were no significant ANOVA effects for whether the aviators wore corrective glasses during either the low level flight tests or the traffic pattern test. There was one significant correlation with low level flight test performance. Aviators who wore glasses detected fewer symbology targets than aviators with uncorrected vision during the scene plus symbology test ($r = -.37$, $p < .05$).

Eye dominance. During the low level flight tests, there were significant interactions between eye dominance and view for % det, $F(1, 34) = 9.63$, $p < .01$, and between eye dominance, type, and view for RT, $F(1, 34) = 7.10$, $p < .05$. There were no significant differences between left- and right-eye dominant aviators in detecting targets when viewing only the scene or only the symbology, but the right-eye dominant aviators detected more targets than the left-eye dominant aviators when viewing both the scene and symbology (see Figure 16). The correlation between eye dominance and symbology detection under the both condition was .38 ($p < .05$), indicating that right-eye dominant aviators detected more targets.

In the three-way interaction for RT, eye dominance did not affect the reaction to symbology targets, but it was a factor in reacting to scene targets (see Figure 17). The left-eye dominant aviators reacted equally well to the scene

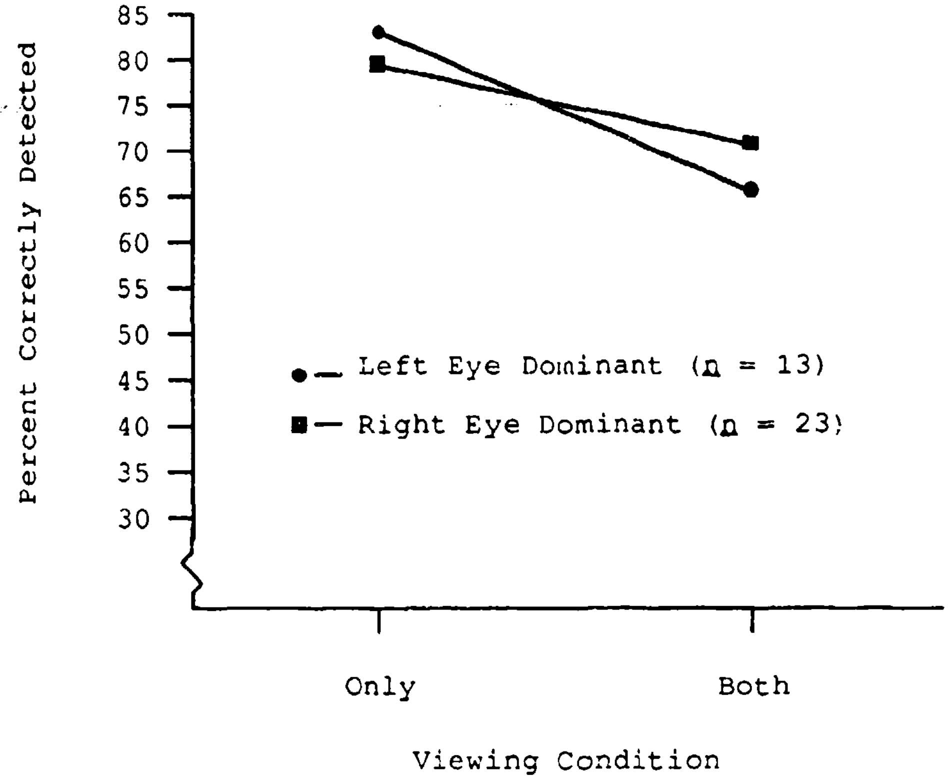

Figure 16. The percentage of targets correctly detected under the only and the both viewing condition by left- and right-eye dominant aviators.

only and scene both targets (*M* = 3.8 s). However, the right-eye dominant aviators reacted more slowly to the both targets (*M* = 4.3 s) than to the only targets (*M* = 3.3 s). There were no significant correlations between eye dominance and RT during the low level flight tests.

There were no significant ANOVA effects of eye dominance during the traffic pattern tests. In addition, there were no significant correlations between eye dominance and performance during these two test sessions.

Handedness

During the low level flight tests, handedness was significantly correlated with RT to symbology targets under the only ($r = -.44$, $p < .01$) and the both ($r = -.40$, $p < .05$) viewing condition. Aviators who were right handed tended to respond faster to the symbology targets.

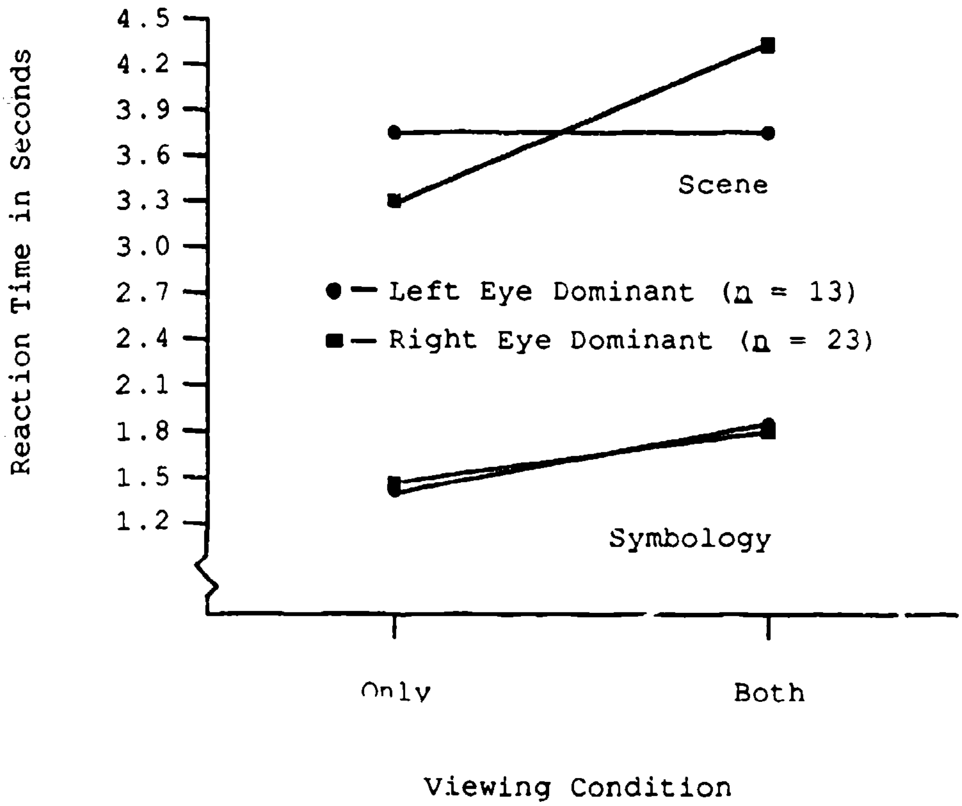

Figure 17. Reaction time to scene and symbology targets under the only and the both viewing condition by left- and right-eye dominant aviators.

Workload Ratings

A test session X experimental condition ANOVA indicated there was a significant interaction between the two factors on the mean workload rating, $F(12, 128) = 3.02$, $p < .001$. To determine which component of the experimental condition variable was interacting with the test session, two test session X training order X scenario sequence ANOVAs were run. In one ANOVA, the scenario sequence was defined as the LLA-LLB sequence. In the other ANOVA, it was defined as the TPA-TPB sequence. The significant interaction in both ANOVAs was caused by the training order, $F(4, 128) = 5.30$, $p < .001$, not the scenario sequence. Newman-Keuls tests indicated the symbology first group rated workload significantly higher than the scene first group in the scene only and the final test sessions (see Figure 18).

There were also overall differences in the mean workload rating for the test sessions. The scene only workload rating was lower than all the other sessions. The initial and final

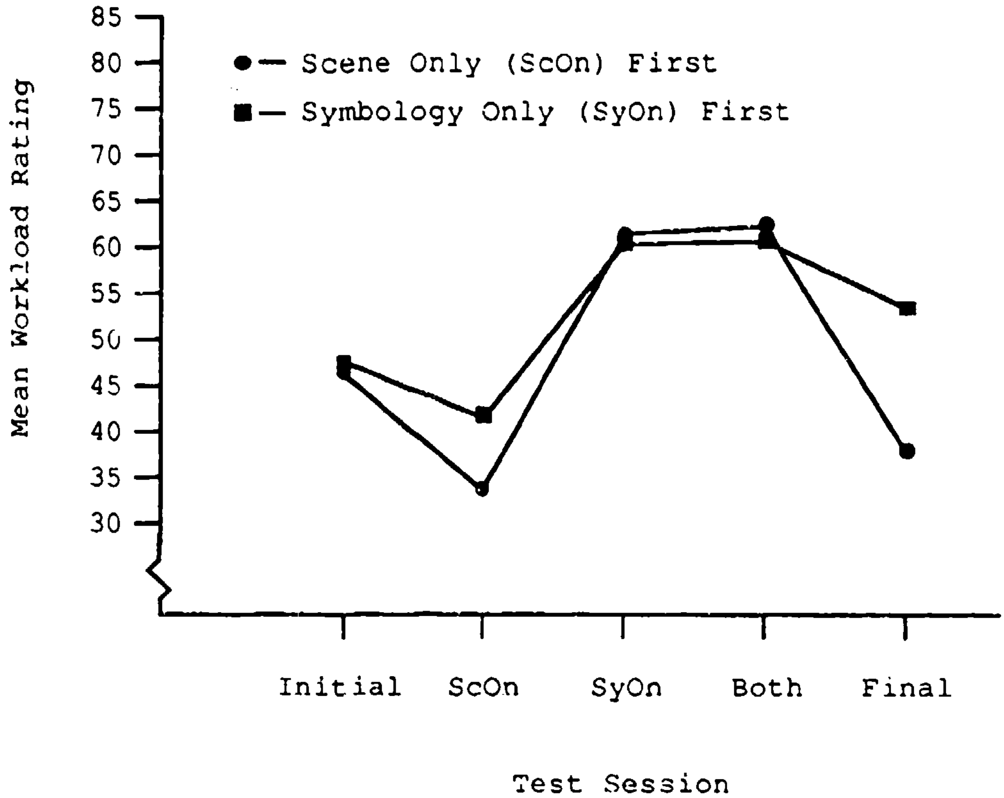

Figure 18. Mean subscale workload rating for the five test sessions for the scene only first and symbology only first groups.

session workload ratings were not different from each other but were different from all the other sessions. The symbology only and both sessions were not different from each other.

Debriefing Summary

This subsection summarizes the aviators' responses during the postexperimental debriefing. The results are summarized as the frequency of aviators responding (the number is given in parentheses), but specific or representative comments are also reported to supplement the data.

Debriefing Questions

Realism. The majority of the aviators (29) indicated that the scene imagery was highly realistic and the remainder reported that it was fairly realistic. The only complaints were about the jumpy practice trials, which were presented at a 10 Hz rate instead of the 30 Hz rate used for the test scenarios. The aviators also indicated the symbology was realistic, but less strongly: 21 said it was highly realistic and 7 said it was fairly realistic. Only one aviator did not think it was realistic: He complained that the rate of change was too fast and that the symbols were not always synchronized with the visual scene.

Viewing problems. Nineteen of the aviators reported no viewing problems with the scene or symbology. Only 2 reported eyestrain and none reported blurred or double images. Five aviators complained about the brightness and 11 complained about the contrast of the display. The primary complaint was that when the nose of the aircraft was raised, it was difficult to see the white symbology against the brighter sky background.

Monitoring tendencies. Only 1 aviator reported frequently monitoring the symbology for instruments in the symbology suite other than airspeed, altitude, trim, and master caution light. However, 26 reported checking them occasionally. The other most frequently monitored instruments were the heading scale (16), the exhaust gas temperature gauge (9), and the torque gauge (8).

Eighteen of the aviators indicated they tended to monitor the instruments more often than the scene, 4 reported monitoring the scene more often, and 11 said they tended to monitor one type of information more often as the situation dictated, such as when numerous terrain features were in view. Fifteen aviators reported frequently becoming fixated on specific instruments and 13 reported occasionally becoming fixated. The aviators fixated most frequently on the altitude (17) and airspeed (16) indicators when the instruments were near an out-of-tolerance state.

Two aviators reported having frequent momentary lapses of attention and nine aviators reported occasional lapses. Three said the lapses were usually caused by thinking about what the airspeed or altitude tolerance values were. Three other aviators reported lapses when they were transitioning between the scene and symbology information.

The aviators were inconsistent about which section of the FOV they attended to most frequently: 6 said the upper section, 6 said the middle section, 9 said the lower section, 5 said their attention varied, and 10 were unaware of any monitoring bias. Only 5 aviators said they were unable to monitor events with their peripheral vision, although 4 aviators did not comment on this question. The aviators most frequently used their peripheral vision to monitor the trim ball and MST light.

Only 3 aviators failed to develop a specific scanning strategy; 29 used a scanning strategy regularly and 4 used one occasionally. The strategy most often reported was similar to their normal instrument scan interspersed with a scan of the imagery. The strategies reported were a left-right horizontal (9), triangular (6), clockwise (5), counterclockwise (1), and an alternating clockwise and counterclockwise (1) scanning pattern.

Practice effects. The aviators considered the practice trials to be very helpful (35) or somewhat helpful (1) despite the substantial visual differences between the videodisc practice scenarios and the videotape test scenarios. The only complaint was that the altitude tolerance values should be the same on the practice trials and test sessions. With two exceptions, the aviators believed that their performance was much improved (26) or somewhat improved (8) across the test sessions. Several aviators commented that learning to use the ANVIS-HUD would require a lot of practice and that the experimental apparatus might be an effective part task trainer.

Aviator Suggestions

Symbology suite. Many of the aviators volunteered suggestions about the symbology suite. Most of the comments were about the four instruments they were required to monitor during the experiment, but some suggestions were also made about the other instruments and about desired aviator options. Their suggestions are summarized in the following paragraphs.

One aviator suggested calibrating the digital radar altimeter in increments of 5 feet rather than 1 foot so the rate of change would be slower. Three aviators suggested that the altimeter indicator should flash when a preselected altitude was exceeded. Finally, two aviators commented that the altitude tape was distracting.

Similar comments were made about the airspeed indicator. One aviator suggested calibrating the airspeed indicator in increments of 5 knots and suggested a bug to set a critical airspeed that would cause the indicator to flash. One aviator preferred an analog presentation of airspeed rather than a digital presentation. Another aviator stated a preference for a groundspeed rather than an airspeed indicator.

Two aviators suggested deleting the trim ball and another suggested having it flash whenever it exceeded predefined limits. During the general introduction about monitoring the instruments, several aviators commented that the trim condition was not critical to actual flight and that they could feel when they were out of trim (i.e., through proprioceptive sensations).

Conversely, the aviators recognized the importance of the MST and several suggested modifications to make its activation easier to detect. Four aviators suggested moving the light closer to the center of the display and two suggested making it brighter or a different color. Either during the general instructions or during the debriefing, several aviators suggested making it flash on and off to attract attention. Only one aviator thought the location used in this experiment was appropriate.

Three aviators disliked the attitude indicator but one aviator liked having it in the symbology suite. One aviator recommended moving the heading scale to the bottom of the suite and the MST to the top. Three aviators suggested deleting the exhaust gas temperature gauge and one reported the vertical speed indicator to be distracting. One aviator expressed a preference for digital presentation overall and two preferred analog presentations. Four aviators suggested having a rheostat so the aviator could control the symbology contrast. During the introduction, several aviators asked if they would have the option of selecting symbology suites for different flight modes and whether it would have a declutter mode.

Experimental design. The aviators made only a few comments about improvements to the experimental design even though they were encouraged to do so. Four aviators complained about the different altitude response criteria between the practice trials and the test sessions. Another commented that the terrain should have been more similar in the practice and test scenarios. One aviator suggested having more practice trials for the symbology responses. Only one aviator reported that the apparatus was uncomfortable and that it caused a sore neck. Two aviators suggested elevating the keypad. Finally, one aviator volunteered that the experiment was well designed and another said that the instructions were clearly explained. Several of the aviators volunteered to participate in future experiments.

Discussion

Overview

The discussion is organized into 10 subsections. The first 5 subsections discuss and draw conclusions about the primary research issues addressed by this experiment: determining aviator performance in processing scene or symbology information alone, the effects of superimposing symbology on scene information, training order effects, practice effects, and demographic effects. The next two subsections discuss the aviators' reported workload and their debriefing comments. The eighth subsection discusses some methodological issues that should be considered in designing future experiments with the ANVIS-HUD apparatus. The ninth subsection presents recommendations for future research issues. The final subsection summarizes the conclusions of the research.

Aviator Performance

Symbology Information Processing

The aviators detected a high percentage (> 80%) of symbology out-of-tolerance states during low level flight scenarios and correctly reported a similar percentage of instrument values during the traffic pattern scenarios. They also reacted quickly to the low level symbology targets, averaging less than 2 s per response. Considering that this experiment was their first exposure to a type of helmet mounted symbology and that some of the tolerance values were arbitrary, their performance in monitoring and responding to the symbology indicates that Army aviators are capable of quickly and accurately processing instrument information presented in a helmet mounted display.

Scene Information Processing

The aviators detected a lower percentage (between 60% and 69%) of scene features and events and responded more slowly to them, averaging nearly 4 s per response, than to the symbology targets. The aviators' scene performance is also marginal on an absolute basis, especially considering that the aviators were all qualified on NVGs and all stated that the scene imagery was realistic during the debriefing. However, the scene information was more difficult to process than the symbology information, which was discrete, exhaustively defined, active for a minimum of 3 s and occurred in a predefined location. In addition, the symbology out-of-tolerance states were sometimes cued by scene information (e.g., entering a turn was a cue to check the trim ball but being out of trim had no cueing value for scene information).

There are at least three reasons that the scene performance was significantly worse than the symbology performance. First, the aviators were given a general description and a lengthy but not exhaustive list of scene features and events to be detected. The list was repeated before each test session, but the complexity of the monitoring instructions manifested itself in at least three ways. Some aviators never responded to certain categories of scene targets, such as the leadship breaking the horizon.

Some aviators were confused about which aspect of a scene target was critical. This most commonly occurred when an aviator responded to the disappearance of the leadship rather than the appearance of the leadship. Finally, some aviators monitored and responded to spurious scene targets, which tended to degrade their performance on the relevant targets. Examples of these errors include responding to changes in terrain elevation and calling out the same target more than once. The experimenter tried to correct target errors during the practice trials, but not all test scene targets appeared in the practice scenarios. For example, there was no leadship in the practice scenarios.

Second, target detection was affected by many of the scene features and events being in view for only a brief period of time (< 2 s), the low contrast between some terrain targets and their background, and targets being camouflaged by other features. Target detection was also affected by the presence of multiple simultaneous targets (up to four) or targets appearing in rapid succession. Multiple or sequential out-of-tolerance states also occurred for the symbology, but it may have had a beneficial effect, especially when the symbols were located in close proximity to each other (i.e., the trim ball and MST light).

Third, scene RT was slowed by the need to identify the target, not just detect its presence. For example, a clearing in a wooded area might be perceptible in the scene but could not be clearly identified at its onset as being a field suitable for landing or as a lake or pond. Although some aviators would offer a guess about an ambiguous scene feature as soon as any part of it came into view, most aviators would defer responding until they could identify the target.

Conclusions

The statistically significant differences in scene and symbology % det and RT probably reflect differences in the complexity and real-world ambiguity of the information being monitored rather than differences in the aviators' ability to perform the two tasks. It can also be partially attributed to the precision of the monitoring instructions. Given the artificiality of some of the response criteria and the passive role of the subjects in this experiment, the results indicate that Army aviators are capable of detecting a large percentage of both scene and symbology targets and of rapidly responding to the information.

Superimposing Symbology

The aviators detected a high percentage (> 80%) of targets when they were viewing the scene only or the symbology only during the low level flight scenarios. The aviators reacted quickly to the targets under the only viewing conditions, averaging just less than 2.5 s per target. Performance under the both viewing condition was significantly degraded for both the scene and symbology information, averaging approximately 69% of targets detected and an RT of approximately 3 s.

Superimposing the flight symbology information over the scene imagery increased the total amount of information available in the helmet mounted display but reduced the accuracy and speed of processing any of the data. However, combining the scene and symbology effectively doubled the information available, yet it produced a reduction of only 17.1% in % det and an increase of only 20.7% in RT under the both viewing condition. The results of this experiment indicate that there is a cost associated with superimposing instrument symbology over scene imagery, but it may be an acceptable cost given the increase in information content in the display. Unfortunately, this experiment did not compare simulated ANVIS-HUD performance to simulated ANVIS-instrument panel (i.e., the current procedure) performance, which might have produced an even greater degradation in performance.

Aviator monitoring performance was equally degraded for scene and symbology information. The lack of an interaction between type of information and view demonstrates that the aviators were able to divide their attention effectively between the two types of information. There was no evidence in the performance data that the aviators' attention was captured by one type of information to the exclusion of the other. During the debriefing, the aviators reported some fixation, but it was not directed exclusively toward symbology or the scene and it was usually associated with relevant stimulus information (e.g., a symbol approaching an out-of-tolerance state).

Conclusions

When instrument symbology was superimposed on the scene imagery, the aviators were unable to detect as many targets or to react as quickly to the targets that were detected as they did when viewing the scene or symbology alone. However, the degradation was small considering the increase in information available to the aviator to be processed. The results also indicate that the aviators were able to divide their attention effectively between the two types of information. Because of the observed degradation, however, the amount of information presented in the symbology suite should be limited to the minimum number of critical instruments. The aviators reported occasionally attending to some of the instruments in the suite that were not important to their performance during the experiment.

Training Order

There were no significant differences in aviator performance attributed to whether they were trained and tested on scene only first or symbology only first. Training order affected only their perceived workload, which is discussed in the sixth subsection.

Practice Effects

The results from the traffic pattern flight tests showed a significant increase in the percentage of scene and symbology targets correctly detected from the initial to the final session. This effect indicates the initial test and the low level flight practice trials and test sessions improved the aviators' performance. However, the performance improvement was relatively small (8.3%), increasing from 68.3% det on the initial test to 74.0% det on the final test.

The results are difficult to interpret in terms of practice benefits. First, if the aviators were exposed to scenario TPA on the initial test, it may have sensitized them to the helicopter's presence in TPB, which would mimic a practice effect. However, the interaction between scenario order and test session was not significant, so the main effect of test session is attributed to a practice effect.

Second, the aviators practiced performing related but not identical monitoring and responding tasks during experimental sessions 2, 3, and 4. A better test of practice effects would involve practice on additional traffic pattern flights or the situational awareness (i.e., stop frame, then report symbol states and scene features) methodology. An alternative test would be to conduct an initial and final test using low level flights with the continuous monitoring and responding methodology.

Finally, the situational awareness measurement approach, which was used to assess practice effects, was not as precise as the continuous responding approach. There are at least two reasons for the lack of precision. First, the verbal report method placed a heavy demand on the aviators' working memory (e.g., Ericsson & Simon, 1980). They were asked to recall from 7 to 10 different pieces of information each time the screen was blanked. In contrast, the aviators never had more than five targets in view at a time on the low level flight scenarios, and most of these were in view for several seconds.

Second, the number of items on each situational awareness test was relatively small, so each error caused their score to change by a large amount. On the traffic pattern tests, the aviators' instrument detection scores varied in increments of 12.5% and their scene detection scores varied in increments of 11.1%. A single omission lowered an aviator's score by more than twice the amount that the average aviator's % det score increased as a result of practice. A single error on the low level flight scenarios reduced the aviator's score by as little as 1.3% (instrument detection on scenario LLB).

The situational awareness approach also measured only the simultaneous knowledge of all information and did not consider RT to any individual target. Given the total amount of information to be monitored simultaneously, a ceiling effect for practice improvement on the traffic pattern flight tests is a likely confound.

Conclusions. The results provide some evidence that aviator performance using instrument symbology superimposed over scene imagery does improve, at least slightly, with practice. However, further research is needed to corroborate this finding and to determine what performance levels can be achieved with more extensive practice.

Demographic Effects

There were several significant relationships between demographic variables and performance on the low level flight tests. These effects are discussed in the following paragraphs in sets of experience, vision, and handedness variables. There were no significant effects of any demographic variables on aviator performance during the traffic pattern tests.

Experience

Although there were some minor statistical inconsistencies between the ANOVA and correlational analyses, there was an overall pattern indicating that the less experienced aviators detected a larger percentage and reacted more quickly to the symbology targets than the more experienced aviators. Conversely, the older aviators with more years as a rated aviator performed better in reacting to the scene targets.

These effects are logical and are supported by comments made by the aviators during the experiment and the debriefing. The more experienced aviators obviously have more practice at scanning and interpreting scene information, which resulted in better scene performance. They were also more practiced at functioning without instrument information, frequently remarking that they could judge their airspeed, altitude, and trim in the aircraft without reference to the instruments. Although they monitored the symbology, their prior experience led them to concentrate more on the scene information, thus allowing the less experienced aviators to outperform them in detecting and reacting to the symbology targets.

The results may indicate that less experienced aviators may adapt to a new flight-aiding system better than aviators who have substantial experience with a previous system. However, the results are not strong enough (e.g., the maximum r^2 was only .30) to indicate that more experienced aviators cannot learn to use the ANVIS-HUD system effectively. The more important question is whether the information presented in the symbology suite is critical enough to motivate aviators to expend the additional attentional effort required to monitor it. Further research is needed to determine which instruments should be included in the symbology suite.

Vision

Corrective glasses. There was only one significant effect in the analyses related to corrected vision: Aviators who wore glasses detected fewer symbology targets than aviators who had uncorrected vision. This result is logical, given the need to focus on the symbology to interpret it. However, a single significant effect among multiple tests may be a spurious finding. The effect may also be specific to the ANVIS-HUD apparatus used in this experiment. The results are insufficient to warrant any strong concern about aviators with corrected vision reading superimposed symbology, although additional research may be indicated.

Eye dominance. Eye dominance, as measured in this experiment, was significantly related to aviator performance in detecting and reacting to targets, especially when both scene and symbology information was present. There was no difference in detecting targets under the only viewing condition, but right-eye dominant aviators detected significantly more targets in the both condition. However, right-eye dominant aviators reacted more slowly to the scene targets detected under the both condition even though they reacted more quickly to targets under the scene only condition than the left-eye dominant aviators. There was no interaction between eye dominance and RT for the symbology targets.

These results appear to be contradictory. The detection data indicate that the right-eye dominant aviators perform better; the RT data indicate the left-eye dominant aviators perform as well for symbology targets or better for scene targets than the right-eye dominant aviators. The contradiction may be at least partly an artifact of the scoring procedures. Only targets detected are included in the RT score. Delayed detecting of targets after processing symbology information would produce a larger percentage of targets but with a slower average RT. If an arbitrary maximum RT were included in the score for targets never detected, the overall results would favor presenting the symbology to the dominant eye. This finding should be confirmed in future research in which the symbology is also presented to the dominant eye of left-eyed aviators and to the nondominant eye of right-eyed aviators.

There is one significant effect that cannot be readily explained: the better RT performance by the left-eye dominant aviators in the scene only viewing condition. Symbology was not present and the scene imagery was presented to both eyes under this condition. This result implies an underlying capability difference for the two eye dominance categories or a correlation between eye dominance and some other factor, such as experience. However, there were no significant correlations between eye dominance and any of the other demographic variables. This unexplained result also should be reexamined in future research designed to investigate eye dominance effects more completely.

Handedness

There was a significant correlation between handedness and aviator RT to symbology targets, with right-handed aviators reacting more quickly than left-handed aviators. This result could be interpreted to indicate a cerebral laterality effect because the left hemisphere of the brain is associated with processing analytic, logical, and temporal information. If cerebral dominance were a significant factor, however, left-handed aviators should have exhibited superior performance with scene targets, because the right hemisphere is associated with holistic, visual, and spatial information. Furthermore, target detection should have been affected as well as RT.

An examination of the experimental equipment suggests a more parsimonious interpretation. All the aviators responded to the symbology targets by touch-typing the appropriate number key with their right fingers. Using the nonpreferred hand to make a blind, multiple-choice motoric response is likely to slow the overall RT. Although all the aviators responded to the scene targets with their left hand, they always pressed the same key. That is, it was a simple RT task that should be minimally affected by which hand was used to respond.

Conclusions

The results of this experiment indicate that experience variables such as age, years as a rated aviator, and number of flight and NVG hours may differentially affect initial aviator performance in using the ANVIS-HUD. More experienced aviators performed better in monitoring and responding to scene information and less experienced aviators performed better with symbology information. However, these effects may not occur under more realistic conditions or they may be overcome by practice.

The research did not produce strong evidence that wearing glasses is detrimental to ANVIS-HUD performance, although the data do indicate that aviators who wear glasses may detect fewer symbology targets. The experimental evidence is much stronger that eye dominance has a significant, although complex, effect on aviator performance. Overall, the results indicate that presenting symbology to the dominant eye results in better performance, but more extensive research is needed to confirm this conclusion.

Although there was a significant correlation between handedness and symbology RT, the result was attributed to the configuration of the experimental equipment rather than to a cerebral dominance effect. Thus, the potential effects of cerebral dominance on ANVIS-HUD performance has not been adequately evaluated.

Workload

Overall, the aviators' perceived workload was consistent with the objective difficulty of the test sessions and with their performance levels. The scene only test was rated as having the lowest workload. The two traffic pattern tests, which were judged to be equivalent, were rated as having moderate workload. The symbology only and both test sessions were rated as having the highest workload, but were not significantly different from each other. These results indicate that monitoring and responding to the symbology information required the most attention and was the dominant workload factor in the both condition. The nonsignificant increase in mean workload from the symbology only to the both condition may be the result of prior practice during the two only conditions.

The effect of training order on rated workload is most likely attributable to a contrast effect. The aviators who took the scene only test after the initial test had a lower baseline for rating the scene only workload than the aviators who rated it after taking the symbology only test. This contrast effect probably carried over to the final test session.

Conclusions. The overall workload ratings are very similar to the objective performance data. However, the perceived attentional demand is not only a function of the objective characteristics of each task but also of the comparative characteristics of preceding tasks.

Debriefing

All the aviators discussed the experiment during the debriefing and offered numerous insights and suggestions for fielding the ANVIS-HUD and for designing future research. Because of their subjective nature, however, the aviator comments cannot be used to draw firm conclusions but can be used to support the conclusions drawn from the performance data and to guide future research and development.

Stimulus characteristics. The aviators perceived the scene and symbology scenarios to be realistic representations of actual NVG flight under high illumination conditions. However, they indicated that many of the targets and target parameters were artificial. They had little difficulty perceiving the symbology information except when the aircraft was in a nose-high attitude.

Monitoring techniques. The aviators reported monitoring more of the symbology than was required. They occasionally became fixated on scene or symbology information, but usually because they anticipated a reportable target (e.g., an imminent terrain feature, leadship maneuver, or symbology out-of-tolerance state). Lapses of attention were rare and most often associated with recalling the arbitrary response criteria.

Nearly all the aviators developed and used a scanning strategy, but the strategy used varied among them. The portion of the visual field to which they devoted most of their attention also varied and was probably related to their scanning strategy. Most of the aviators were able to use their peripheral as well as their foveal vision to detect targets.

Practice. The aviators indicated that the practice trials for sessions 2, 3, and 4 were beneficial and generally sufficient, although more trials were suggested for the symbology only condition. They recommended that the same tolerance parameters be used for the practice and test sessions. The aviators also perceived that their performance improved across the test sessions. Several of them noted that the ANVIS-HUD would require substantial amounts of practice to attain proficiency and some suggested that the experimental apparatus would be useful for pretraining on the operational equipment.

Symbology suite suggestions. Overall, the aviators suggested that the instrument symbology used in this experiment, which was modeled on the ANVIS-HUD suite, is excessive. Some aviators suggested deleting several of the symbols (e.g., the redundant altitude tape, the somewhat redundant vertical speed indicator, the attitude indicator, the exhaust gas temperature gauge, and the trim ball). Others suggested reducing the rate of change on other symbols, such as the airspeed and altitude indicators, or having them activate only at critical parameters. Finally, the aviators suggested rearranging the symbology suite to move critical instrument information, such as heading and MST, to a more central location.

Methodological Considerations

This was the first experiment using the ANVIS-HUD simulation apparatus, scenarios, and procedures. The results of the experiment have produced information about aviators' performance under the different test conditions and about the research materials and procedures. Some of the methodological issues have been made explicit in reporting and interpreting the results, but they are all summarized in this section for consideration in future research.

First, the research equipment was functionally reliable, the stimulus materials were judged to be a realistic simulation, and the procedures produced interpretable results. Both the low level flight and traffic pattern scenarios produced usable data, but the situational awareness measurement approach was less like the conditions that are encountered in flight. That is, the flight stopped and the aviators were required to recall all the types of information rather than serially scanning for relevant scene and symbology information. The situational awareness measures also produced less precise performance data.

The differences in difficulty between the traffic pattern scenarios and the low level flight scenarios required a data transformation before the final analyses. The data transformation and the counterbalances used in this experiment were sufficient to make the results interpretable, but future research should reduce the inherent disparity in difficulty, either by modifying the scenarios or by changing the monitoring instructions, or both.

There are three other reasons to recommend modifying the monitoring instructions. First, the aviators complained that the instrument tolerance criteria were not realistic. Some suggested having only a minimum altitude criterion and a maximum airspeed criterion, which are similar to their operational flight instructions. Second, the aviators exhibited problems in employing the scene instructions, by either failing to respond to some targets, confusing the target instructions, or responding to targets that were not included in the instructions. Reducing the number of scene targets should reduce the problems encountered. Third, modifying the instructions would simplify the scoring of scene performance. Unlike the symbology scoring, which was computer generated, scoring the scene performance was labor intensive and time consuming.

Although every potentially relevant variable cannot be fully counterbalanced in any given experiment, two procedural variables that were held constant in the current experiment have limited the interpretation of the data. Symbology was always presented to the right eye and responses to symbology targets were always performed with the right hand. Presentation of the symbology to the right or left eye should be manipulated in a future experiment to investigate eye dominance effects. Symbology responses should probably be made with the preferred hand. If further research is conducted on the effects of handedness, subjects should be selected so that there is a larger number of left handed aviators. There were only four left handed aviators in the current sample, but that approximates the percentage of left handed individuals in the general population. For example, Morey and Simon (1991b) found that 10.5% of their sample was left handed.

Only one of several possible measures of eye dominance was used in this experiment. All the subjects were categorized by this test as exclusively left or right eyed, and the results indicated there were significant performance interactions with measured eye dominance. However, the test used is closely associated with sighting dominance and may not be correlated with sensory dominance, which may be more relevant to the ANVIS-HUD situation.

Finally, the aviators in this experiment were passive observers who could not affect the status of the helicopter. This role is very artificial for either a pilot or copilot. Whether the results of and the conclusions drawn from this research hold when the aviator is performing the duties or pilot on the controls or even engaged as an active copilot should be evaluated in subsequent experiments.

Future Research Recommendations

Both the research results and the methodological considerations have indicated a need for additional research. This subsection is intended to summarize and to suggest priorities for the additional research requirements based on the current findings.

First, research should be conducted to investigate further the effects of eye dominance. The research is needed to provide guidance for fielding the ANVIS-HUD and for controlling eye dominance effects in experiments designed to investigate other issues. Both the sighting dominance test used in this experiment and potentially relevant tests of sensory dominance should be used in future research.

Second, research is needed to replicate and extend the current findings when the aviators are performing a more active role in the cockpit. Performing additional tasks, whether piloting the aircraft or providing navigation or communication support, will likely reduce the attentional resources that the aviators can expend on monitoring and responding to the information available in the ANVIS-HUD. The research can also be designed to provide objective information about the need for, characteristics of, and placement of specific instrument symbols.

Third, research is needed to examine the effects of extended practice on using the ANVIS-HUD and to determine what performance levels can be achieved. This information would be valuable in the development of training programs for acquiring and maintaining ANVIS-HUD skills and in the establishment of performance standards for evaluating aviator performance. The results should also be considered when conducting additional experiments that assume the aviators are proficient in using the ANVIS-HUD. A related research need is to determine the potential training transfer from the apparatus used in this experiment or other ANVIS-HUD simulation to the operational equipment.

Summary of Conclusions

The 11 conclusions drawn from the results of this research are summarized in this subsection. The first conclusion addresses methodological issues. The next 9 conclusions address the primary research issues. The final conclusion summarizes the most important requirements for future research.

1. The scene scenarios and symbology presented with the research apparatus realistically simulates the visual stimuli of the ANVIS-HUD. The research paradigm produces interpretable results, although significant differences in difficulty between the scenarios complicated the analyses. The situational awareness measures are not as precise as the continuous responding measures. However, scoring the scene performance data obtained during the low level flight scenarios is labor intensive.

2. Rated helicopter pilots can detect a large percentage of targets and can react rapidly to critical information when viewing only scene or only symbology information.

3. When instrument symbology and scene imagery are presented together, avi ›rs can divide their attention effectively between the ‿wo types of information.

4. Presenting symbology and scene information simultaneously degrades the detection and response time to both types of information. However, the degradation is small relative to the increase in information available in the display. Fixation on instruments or scene features or events is not believed to be a serious problem and is usually caused by the relevance of the information in the display. To minimize interference and information load, only critical instrument information should be included in the display.

5. Aviator performance using displays with both symbology and scene imagery improves with only limited practice. However, substantial practice may be required to achieve maximum performance levels with the ANVIS-HUD.

6. Aviator performance using both symbology and scene imagery is not affected by the order of training on the symbology-only or scene-only tasks. However, the order of training does affect the aviators' perceived workload.

7. Aviator experience, especially age and years as a rated aviator, is related to ANVIS-HUD performance. More experienced aviators perform better in monitoring and responding to scene information and less experienced aviators perform better with symbology information.

8. Wearing corrective spectacle lenses has little effect on aviator performance, although aviators who wear glasses may detect fewer symbology targets.

9. Presenting symbology to the aviator's dominant eye generally produces better performance, but the results were equivocal. Further research is needed to evaluate the eye dominance effects.

10. The aviators' perceived workload is similar to their performance levels. Monitoring and responding to the instrument symbology dominates the perception of workload.

11. Further research is needed to confirm and extend the findings of this experiment. In particular, research is needed to investigate eye dominance effects, to determine aviators' performance capabilities when they are actively involved in flying the helicopter, and to determine the effects of extended practice on using the ANVIS-HUD.

References

Ackerman, P. L., Schneider, W., & Wickens, C. D. (1984). Deciding the existence of a time-sharing ability: A combined methodological and theoretical approach. Human Factors, 26 (1), 71-82.

Ashcraft, M. H. (1989). Human memory and cognition. New York: Harper Collins.

Becklen, R., & Cervone, D. (1983). Selective looking and the noticing of unexpected events. Memory and Cognition, 11, 601-608.

Brickner, M. S. (1989). Helicopter flights with night vision goggles - Human factors aspects (NASA-TM-101039). Washington, DC: National Aeronautics and Space Administration.

Buckner, R. L. (1992). Night vision in Army aviation. Army Aviation, 41 (7), 14-16.

Byers, J. C., Bittner, A. C., Jr., & Hill, S. G. (1989). Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? Advances in Industrial Ergonomics and Safety, Vol. 1, pp. 481-485. London: Taylor & Francis.

Crowley, J. S. (1989). Cerebral laterality and handedness in aviation: Performance and selection implications (USAFSAM-TP-88-11). Brooks Air Force Base, TX: Human Systems Division, USAF School of Aerospace Medicine.

Damos, D. L. (1991). Dual-task methodology: Some common problems. In D. L. Damos (Ed.), Multiple task performance (pp. 101-119). London: Taylor & Francis.

Department of the Army. (1988). Night flight techniques and procedures (Training Circular 1-204). Washington, DC: Headquarters, Department of the Army.

Ericsson, K., & Simon, H. A. (1980). Verbal reports as data. Psychological Review, 87, 215-250.

Fischer, E. (1979). The role of cognitive switching in head-up displays (NASA Technical Report CR-3137). Washington, DC: National Aeronautics and Space Administration.

Fischer, E., Haines, R. F., & Price, T. A. (1980). Cognitive issues in head-up displays (NASA-TP-1711). Washington, DC: National Aeronautics and Space Administration.

Foyle, D. C., & Kaiser, M. K. (1991). Pilot distance estimation with unaided vision, night vision goggles, and infrared imagery. SID International Symposium Digest of Technical Papers, XXII, 314-317.

Foyle, D. C., Sanford, B. D., & McCann, R. S. (1991). Attentional issues in superimposed flight symbology. In R. S. Jensen (Ed.), Proceedings of the Sixth International Symposium on Aviation Psychology (pp. 577-582). Columbus, OH: Ohio State University.

Gopher, D., Grunwald, A., Straucher, Z., & Kimchi, R. (1990). Tracking and letter classification under dichoptic and binocular viewing conditions. In L. Smith (Ed.), Proceedings of the Human Factors Society 34th Annual Meeting (pp. 1557-1561). Santa Monica, CA: Human Factors Society.

Hamilton, D. B. (1992). Preliminary validation of the task analysis/workload (TAWL) methodology (ASI690-353-92[A]). Fort Rucker, AL: Anacapa Sciences, Inc.

Hirst, W., & Kalmar, D. (1987). Characterizing attentional resources. Journal of Experimental Psychology: General, 116 (1), 68-81.

Iavecchia, J. H., Iavecchia, H. P., & Roscoe, S. N. (1988). Eye accommodation to head-up virtual images. Human Factors, 30, 689-702.

Johnston, W. A., Hawley, K. J., & Farah, M. J. (1988). Individual differences in attention (Final Report for Grant ASOSR-87-0212). Salt Lake City, UT: University of Utah, Department of Psychology.

Kaiser, M. K., & Foyle, D. C. (1991). Human factors issues in the use of night vision devices. Proceedings of the Human Factors Society 35th annual meeting. pp. 1502-1506. Santa Monica, CA: Human Factors Society.

Kyllonen, P. C., & Woltz, D. J. (1989). Role of cognitive factors in the acquisition of cognitive skill. In R. Kanfer, P. L. Ackerman, & R. Cudeck (Eds.), Abilities, motivation, and individual differences (pp. 239-280). Hillsdale, NJ: Lawrence Erlbaum.

Larish, I. A., & Wickens, C. D. (1991). *Divided attention with superimposed and separated imagery: Implications for head-up displays* (Technical Report ARL-91-4/NASA HUD-91-1). Urbana, IL: Aviation Research Laboratory Institute of Aviation.

McLean, B., & Smith, S. (1987). Developing a wide field of view HMD for simulators. In *Proceedings of the Society of Photo-optical Instrumentation Engineers* (pp. 79-82). Bellingham, WA: Society of Photo-optical Instrumentation Engineers.

Miles, C., & LaPointe, J. (1986). Development of a videodisc version of Advanced MITAC training exercises. In K. D. Cross & S. M. Szabo (Eds.), *Human factors research in aircrew performance and training* (Final Summary Report ASI479-080-86, pp. 95-98). Fort Rucker, AL: Anacapa Sciences, Inc.

Moffitt, K. (1989). Ocular responses to monocular and binocular helmet-mounted display configurations. In J. T. Carollo (Ed.), *Proceedings of the Society of Photo-optical Instrumentation Engineers: Helmet-Mounted Displays* (pp. 142-148). Bellingham, WA: Society of Photo-optical Instrumentation Engineers.

Morey, J. C., & Simon, R. (1991a). *Attention factors associated with head-up display and helmet-mounted display systems* (E-19298U, rev. 1). Wilmington, MA: Dynamics Research Corporation.

Morey, J. C., & Simon, R. (1991b). *Development of handedness and eye preference assessment instruments for ANVIS-HUD research applications* (E-19312U). Wilmington, MA: Dynamics Research Corporation.

National Aeronautics and Space Administration. (1986). *NASA task load index (TLX), version 1.0*. (Paper and pencil package). Moffett Field, CA: NASA Ames Research Center, Human Performance Research Group.

Neisser, U., & Becklen, R. (1975). Selective looking: Attention to visually specified events. *Cognitive Psychology*, *7*, 480-494.

Norman, J., & Ehrlich, S. (1986). Visual accommodation and virtual image displays: Target detection and recognition. *Human Factors*, *28*, 135-151.

Porac, C., & Coren, S. (1976). The dominant eye. *Psychological Bulletin*, *83*, 880-897.

Price, D. R., & McLean, W. E. (1985). Aeromedical lessons learned with night vision devices. In Visual Protection and Enhancement. AGARD/NATO Conference Proceedings, No. 379. Neuilly-Sur-Seine, France, pp. 8.1-8.10.

Roscoe, J. T. (1975). Fundamental research statistics for the behavioral sciences (2nd ed.). NY: Holt, Rinehart & Winston, Inc.

Roscoe, S. N. (1984). Judgments of size and distance with imaging displays. Human Factors, 26, 617-629.

Roscoe, S. N. (1987a). The trouble with HUDs and HMDs. Human Factors Society Bulletin, 30, 1-3.

Roscoe, S. N. (1987b). The trouble with virtual images revisited. Human Factors Society Bulletin, 30, 3-5.

Ruffner, J. W., Grubb, M. G., & Hamilton, D. B. (1992). A review of factors affecting rotary wing aviator performance with the night vision helmet-mounted display. (Research Report 1622). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A254 983)

Runyon, R. L. (1985). Qualification operational test and evaluation (QOT&E) of night vision goggles (NVG) head-up display (HUD): Test report (MAC Project Report No. 15-84-84-2). Hurlburt Field, FL: Special Missions Operational Test and Evaluation Center, Military Airlift Command. (AD B092308L)

Simmons, R., Kimball, K., & Hamilton, B. (1985). Micro-heads-up display for enhancement of night vision goggle operations. In Visual Protection and Enhancement. AGARD/NATO Conference Proceedings, No. 379. Neuilly-Sur-Seine, France, pp. 7.1-7.9.

Stoffregen, T. A., & Becklen, R. C. (1989). Dual attention to dynamically structured naturalistic events. Perceptual and Motor Skills, 69, 1187-1201.

Tredici, T. J., & Miller, R. E. (1985). Night vision manual for the flight surgeon (Report No. USAFSAM-SR-85-3). Brooks Air Force Base, TX: USAF School of Aerospace Medicine. (AD A159 720)

U.S. Marine Helicopter Squadron One. (1989). Night vision goggle head-up display assessment. Quantico, VA.

Verona, R. W., & Rash, C. E. (1989). Human factors and safety considerations of night vision systems flight (USAARL Report No. 89-12). Fort Rucker, AL: U. S. Army Aeromedical Research Laboratory.

Weintraub, D. J. (1987). HUDs, HMDs, and common sense: Polishing virtual images. Human Factors Scciety Bulletin, 30, 1-3.

Weintraub, D. J., Haines, R. F., & Randle, R. J. (1985). Head-up display (HUD) utility. II: Runway to HUD transitions monitoring eye focus and decision times. In R. W. Sweezey (Ed.), Proceedings of the Human Factors Society 29th Annual Meeting (pp. 615-619). Santa Monica, CA: Human Factors Society.

Wiley, R. W. (1989). Visual acuity and stereopsis with night vision goggles (USAARL Technical Report 89-9). Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory.

Winer, Statistical principles in experimental design (2nd ed.). NY: McGraw-Hill.

Zeller, J. L., & Thornton, R. C. (1992). Development and evaluation of an aircrew coordination training program. In D. M. McAnulty (Ed.), Human factors research in aircrew performance and training: 1986-1991 Final summary report (Technical Report 954). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A254 455)

APPENDIX A

THE NATIONAL AERONAUTICS AND SPACE ADMINISTRATION TASK LOAD INDEX WORKLOAD RATING FORM

Workload Rating Scale Definitions		
Title	Endpoints	Description
Mental Demand	Low/High	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand	Low/High	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Effort	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration Level	Low/High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?
Performance	Good/Poor	How successful do you think you were in accomplishing the goals of the task you set for yourself? How satisfied were you with your performance?

Figure A-1. Description of the six National Aeronautics and Space Administration Task Load Index workload rating scales.

NASA TLX Ratings for Test ____

Condition: ________ Last Four: ________

MENTAL DEMAND

Low High

PHYSICAL DEMAND

Low High

TEMPORAL DEMAND

Low High

EFFORT

Low High

FRUSTRATION

Low High

PERFORMANCE

Good Poor

Figure A-2. The National Aeronautics and Space Administration Task Load Index workload rating form.

APPENDIX B

SCORING FORMS FOR THE LOW LEVEL FLIGHT SCENARIOS A AND B

Underlined spaces indicate that the scene feature or event was counted in calculating the percentage of targets correctly detected (% det) and in reaction time (RT). The time shown in parentheses beside the target description is the running time from the beginning of the scenario. The key press and start times in the third and fourth columns are in fractions of seconds and are keyed to the videotape frame. Scenario A began at frame second 290 and scenario B began at frame second 2632.

Low Level Scenario A Computation Form — Page 1

Last 4: ______ View: ___ Scene Only ___ Both

Description (Times)	Detect	Press	Start	RT
LS Rotor Blades(0:12)				
LS Appears (0:26)	_____			
LS Appears (0:33)*	_____	______.__	- 323.60	= ___.__
LS Crosses (0:37-0:39)	_____			
Horizon Break (0:42-0:47)	_____			
Field (0:58-1:50)	_____	______.__	- 348.00	= ___.__
LS Appears(1:02)	_____	______.__	- 351.93	= ___.__
Structure (1:25-1:47)	_____			
LS Appears (2:18)*	_____	______.__	- 424.17	= ___.__
LS Crosses (2:20-2:21)	_____			
Field (2:25-2:37)	_____	______.__	- 436.00	= ___.__
Road (2:32-2:36)				
LS Appears (2:54)	_____			
LS Appears (3:04)	_____			
Stream Bed (3:14-3:40)	_____			
LS Appears (3:21)	_____	______.__	- 491.17	= ___.__

*If the subject did not detect the leadship appearance but did detect the leadship crossing, compute RT to the crossing detection.

Figure B-1. Low level scenario A scene scoring sheet (LS = leadship).

Low Level Scenario A Computation Form Page 2

Last 4: ______ View: ___ Scene Only ___ Both

Description (Times)	Detect	Press	Start		RT
LS Appears (3:38)*	_____	______.__	- 507.63	=	___.__
LS Crosses (3:40-3:41)	_____				
LS Appears (3:48)	_____	______.__	- 517.73	=	___.__
Stream Bed (3:50-4:02)	_____				
Field (4:11-4:25)	_____	______.__	- 542.00	=	___.__
Field (4:37-4:44)	_____	______.__	- 567.00	=	___.__
Field (4:53-5:00)	_____	______.__	- 581.00	=	___.__
LS Appears (4:55)*	_____	______.__	- 583.53	=	___.__
LS Crosses (4:58-4:59)	_____				
Horizon Break (5:02-5:03)					
Trail (5:07-5:14)	_____	______.__	- 599.16	=	___.__
Structure (5:16-5:23)	_____				
Clearing (5:33-5:42)					

- -

TOTAL OF COLUMNS _____ ______.__

% DET (TOTAL / 24) = _____ Mean RT = ______.__

*If the subject did not detect the leadship appearance but did detect the leadship crossing, compute RT to the crossing detection.

Low Level Scenario B Computation Form Page 1

Last 4: ______ View: ___ Scene Only ___ Both

Description (Times)	Detect	Press	Start	RT
Three Fields (0:00-0:26)				
Horizon Break (0:38-1:01)	_____	______.__	- 2668.30	= ___.__
Lake or Pond (1:03-1:14)	_____	______.__	- 2695.00	= ___.__
Field (1:08-1:24)	_____	______.__	- 2700.00	= ___.__
Dirt Road (1:14-1:18)	_____			
Horizon Break (1:16-1:22)	_____			
Field (1:29-1:37)	_____	______.__	- 2721.00	= ___.__
Field (1:32-1:49)	_____	______.__	- 2724.00	= ___.__
Field (1:49-2:18)	_____	______.__	- 2742.00	= ___.__
Trail (2:20-2:26)				
Trail (2:28-2:30)				
Trail (2:38-2:40)				
Road (2:40-2:46)	_____	______.__	- 2790.67	= ___.__
Horizon Break (2:41-2:44)				
Horizon Break (3:05-3:22)	_____	______.__	- 2816.77	= ___.__
Horizon Break (3:25-3:56)	_____	______.__	- 2839.00	= ___.__
Lake or Pond (3:37-3:46)	_____	______.__	- 2852.00	= ___.__
Field (3:47-4:24)	_____	______.__	- 2859.00	= ___.__
Trail (4:09-4:20)	_____			
Dirt Road (4:21-4:25)	_____			

Figure B-2. Low level scenario B scene scoring sheet.

Low Level Scenario B Computation Form

Last 4: ______ View: ___ Scene Only ___ Both

Description (Times)	Detect	Press	Start	RT
Blacktop Road (4:26-4:29)	_____	_____._	- 2898.07	= ___._
Horizon Break (4:26-4:33)	_____			
Field (4:37-5:05)	_____	_____._	- 2909.00	= ___._
Field (4:49-4:56)	_____	_____._	- 2924.00	= ___._
Stream (5:09-5:12)	_____	_____._	- 2941.33	= ___._
Stream (5:15-5:17)				
Field (5:34-5:46)	_____	_____._	- 2966.00	= ___._
Field (5:40-6:08)	_____	_____._	- 2975.00	= ___._
Cleared Area (6:46-6:59)				
Trail (7:31-7:35)	_____			
Road (7:39-7:41)	_____	_____._	- 3091.87	= ___._
Field (7:45-8:21)	_____	_____._	- 3099.00	= ___._
Trail (8:01-8:06)				
Road (8:13-8:17)	_____	_____._	- 3123.00	= ___._
Lake or Pond (8:39-8:59)	_____	_____._	- 3152.00	= ___._
Dirt Road (8:45-9:00)	_____	_____._	- 31[illegible].00	= ___._
Horizon Break (9:01-9:07)	_____	_____._	- 3172.37	= ___._
Horizon Break (9:17-9:22)	_____	_____._	- 3190.00	= ___._
Field w/Trail (9:25-9:28)	_____	_____._	- 3196.57	= ___._
Field (9:35-9:41)	_____			
TOTAL OF COLUMNS	___			_____._

% DET (TOTAL / 32) = ____ Mean RT = _____._

930407