ABSTRACT

KIM, SANG-HWAN. Examining and Explaining the Effects of Non-Iconic Conformal Features in Advanced Head-up Displays on Pilot Performance. (Under the direction of Dr. David B. Kaber.)

The primary objective of this study was to assess the impact of Synthetic Vision System (SVS) and Enhanced Vision System (EVS) depictions of terrain features on pilot performance when displayed in an advanced head-up display (HUD) during various phases of a landing approach under instrument meteorological conditions (IMCs). SVS is a display system that presents terrain features using a wireframe grid rendered polygons by integrating terrain databases with a global positioning system. EVS displays present an actual out-of-cockpit view using a forward looking infrared camera.

In the experiment as part of this study, video stimuli presenting varied HUD configurations were pre-recorded using a high-fidelity flight simulator at NASA Langley and presented to eight pilots later in a lab environment. The HUD videos from the high-fidelity simulator were combined with out-of-cockpit views from a lab simulator. The flight scenario consisted of an approach and landing on a runway (Reno, Nevada International Airport (KRNO), 16R (right)) under IMC. Each pilot completed eight trials based on a within-subjects experimental design and one additional trial to collect verbal protocols on specific display feature use. The independent variables included four display configurations (baseline, SVS-only, EVS-only, and a combination of SVS and EVS features) and two visibility conditions (IMC-day versus IMC-night). Every display configuration included tunnel features (highway-in-the sky) showing the designated flight

path. The experiment involved observing pilot performance in four segments during the approach and landing. Dependent variables included flight path control performance, pilot SA, workload, and subjective preferences. Flight path control performance was determined based on pilot errors in tracking a flight path marker in the pre-recorded videos with a super-imposed cursor using test pilots yoke controls. Pilot situation awareness (SA) was measured using SAGAT (the Situation Awareness Global Assessment Technique) in order to evaluate pilot perception, comprehension, and projection for three types of pilot SA (spatial, system, and task awareness). Workload measures were recorded using the NASA-TLX (Task Load Index) and heart-rate. In order to develop explanations of pilot behavior under the various HUD conditions, a video record of the additional test trial was reviewed by each subject using a verbal protocol analysis and probing technique.

Results revealed SVS to support overall pilot SA but to degrade flight path control performance due to confusion of visual features, EVS caused pilots to focus on path control but decreased System awareness because of visual distractions of some imagery. The combination of SVS and EVS features generated offsetting effects; however there were decrements in performance in the final landing phase due to clutter effects. In general, display configurations did not affect spatial awareness but pilot awareness of system information was impacted. The IMC-day condition produced worse flight performance than night flight due to the low visual saliency of HUD imagery in daylight. Flight performance was not different among phases of flight but different levels and types of pilot SA were affected by segment. Because the main task in the study was the tracking task, results did not reveal differences of conditions in terms of workload measures. Interestingly, patterns of pilot preference for displays did not match with the results of objective performance and SA measures. Pilots gave higher ratings of SA support and safety for the SVS and EVS displays with the lowest ratings for the combination. Ratings on annoyance increased with increases in display visual content.

The verbal protocol analysis yielded sequential and non-sequential lists of pilot tasks and behaviors and critical pilot comments. The analysis also identified the required information and alternative methods of performance for specific flight tasks in the scenario. This analysis was used to explain the experimental results and describe pilot behaviors with the SVS and EVS displays in the flight scenario.

This study assessed advanced HUD feature effects on pilot performance, using an elaborate SAGAT method for measuring pilot SA, and developed a CTA for interpreting experimental results. Further studies need to be conducted to evaluate the advanced HUDs under various flight situations using a more realistic flight simulator as a basis for optimal design. In addition, cognitive model of pilot behavior based on CTA needs to be developed for predicting performance and SA implications of HUD design.

Examining and Explaining the Effects of Non-Iconic Conformal Features in Advanced Head-Up Displays on Pilot Performance

by
Sang-Hwan Kim

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Industrial Engineering

Raleigh, North Carolina

2009

APPROVED BY:

Nancy J. Currie
Member

Brad Mehlenbacher
Member

Eric N. Wiebe
Member

Robert A. St. Amant
Minor Representative

David B. Kaber
Committee Chair

DEDICATION

To my dear wife and family for their unconditional love and support.

조건없는 사랑과 믿음을 주었던, 사랑하는 아내와 가족들에게 이 논문을 바칩니다.

BIOGRAPHY

Sang-Hwan Kim was born and raised in Seoul, Korea. He completed his high school education in Kwacheon in 1993. He then joined the SoongSil University studying for B.S. degree in Industrial Engineering. Following military service in the Korean Army during 1995-1997, he resumed his studies and graduated in 2000. He received his M.S. degree from Korea University in Industrial Engineering in 2002, where his research focused on human factors, usability, and Human-Computer Interaction (HCI) under the advisement of Dr. Rohae Myung. During the next three and half years he worked for Samsung Electronics in Korea as an HCI and usability researcher and a user interface/interaction designer.

Since 2005, he has been working towards a Doctorial degree in the department of Industrial and Systems Engineering at North Carolina State University (NCSU) under the guidance of Dr. David Kaber. While at NCSU, he was involved in a number of cognitive engineering and computational cognitive modeling research projects focusing on HCI and aviation psychology.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to Professor Dr. David Kaber for his continuing encouragement and advice through the course of this work. Dr. Kaber has helped me in every aspect of my education. He has been a true mentor and friend. I am greatly honored to have had the opportunity to work under his tutelage. I would also like to thank the members of my committee, Dr. St. Amant, Dr. Wiebe, Dr. Mehlenbacher, and Dr. Currie for their trust during my work and providing a source of confidence. I also thank Dr. Prinzel and Dr. Bailey of NASA for allowing me to use the flight simulator at the NASA Langley research center for video recordings.

I offer this dissertation as the second payment on a debt I owe to the students and my colleagues in the Industrial and System Engineering department and Ergonomics Lab. Specifically, I thank the student colleagues past and present, who helped me to go on through their scholarly example and personal concern in Cognitive Ergonomics Lab. Especially, I wish to express my special appreciation to Karl Kaufmann (referred to as the “expert pilot” in this manuscript), a colleague in the lab, for his assistance during the experiment and writings, including flight scenario development, pilot testing, cognitive task analysis, etc.

I also wish to express my appreciation to my Korean friends at NCSU and the church I have attended. I am deeply indebted to them for providing encouragement and support during a challenging period.

I owe a great debt of gratitude to my family in Korea: my parents who never doubted and devoted their whole efforts to me and to my parents-in-law for excusing me from many of my family duties and letting me concentrate on my research.

I cannot thank my wife, Sang-mi Ko, enough for her love and patience throughout all these years. Her optimism and encouragement have been invaluable at the times I felt discouraged. Her patience and tears has been the source of my motivation. Finally, I am sincerely thankful to God for his constant love to me and for allowing me to achieve my dreams.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVATIONS

ACTA	Applied (Advanced) Cognitive Task Analysis
AFFTC	Air Force Flight Technical Center
AFL	Above Field Level
AGL	Above Ground Level
ATIS	Automatic Terminal Information System
CFIT	Controlled Flight Into Terrain
CTA	Cognitive Task Analysis
DH	Decision Height
DIME	Database Integrity Monitoring Equipment
DME	Distance Measuring Equipment
EADI	Electronic Attitude Direction Indicator
EFIS	Electronic Flight Instruments
EPIC	Executive-Process/Interactive Control
EVS	Enhanced Vision Systems
FAA	Federal Aviation Administration
FAF	Final Approach Fix
FAR	Federal Aviation Regulations
FCP	Flight Control Panel
FLIR	Forward-Looking Infrared
FMS	Flight Management System
FO	First Officer
FOV	Field Of View
FPM	Flight Path Marker
FSF	Flight Safety Foundation
FTE	Flight Technical Errors
GDTA	Goal-Directed Task Analysis
GOMS	Goal, Operator, Method, Selection of rules
GPS	Global Positioning System
G/S	Glideslope
GSRL	Glideslope Reference Line
GPWS	Ground Proximity Warning System
HDD	Head-Down Displays
HITS	Highway-In-The-Sky
HMD	Head (Helmet)-Mounted Displays
HR	Heart Rate
HUD	Head-Up Displays
IAF	Initial Approach Fix

IFD	Integration Flight Deck
ILS	Instrument landing Systems
IMC	Instrument Meteorological Conditions
INS	Inertial Navigation System
KRNO	Reno/Tahoe International Airport
LAAS	Local Area Augmentation System
LOC	Localizer
LWIR	Long-Wave Infrared
MCP	Mode Control Panel
MDA	Minimum Descent Altitude
MDS	Multidimensional Scaling
MMWR	Millimeter Wave Radar
MS	Military Standard
MSL	Mean Sea Level
NASA-LaRC	NASA Langley Research Center
NASA-TLX	NASA-Task Load Index
ND	Navigation Displays
OTW	Out-of-the-Window
PFD	Primary Flight Display
RMSE	Root Mean Square Error
RNAV	Area Navigation
RWY	Runway
SA	Situation Awareness
SAGAT	Situation Awareness Global Assessment Technique
SART	Situation Awareness Rating Technique
SA-SWORD	Situation Awareness-Subjective Workload Dominance Technique
SpA	Spatial awareness
STSS	Short-Term Sensory Store
SVS	Synthetic Vision Systems
TAWS	Terrain Awareness Warning Systems
TCAS	Traffic Collision Avoidance System
VMC	Visual Meteorological Conditions
V/S	Vertical Speed
WAAS	Wide Area Augmentation Systems

1. INTRODUCTION

1.1. Challenges in Avionic Display Design / Development

1.1.1. Flight accidents under low visibility conditions

According to data from the Flight Safety Foundation (FSF), almost 60% of all commercial aircraft crashes occur during the approach and landing phases of flight. Among these accidents, Controlled Flight Into Terrain (CFIT) has been found to account for more than half of all commercial aviation fatalities to date (Etherington et al., 2000). CFIT crashes are accidents where a normally functioning, mechanically sound aircraft impacts terrain or obstacles that a pilot could not perceive due to a lack of outside visual references or impaired terrain/hazard situation awareness (SA). The ability of a pilot to determine critical information through visual perception of the outside environment may be limited by time of day and various weather phenomena, such as rain, fog and snow. Examples of CFIT accidents include the crash of KAL Flight 801 in 1997 during approach to Guam and TWA Flight 514 during approach to Dulles International Airport, Washington, D.C. in 1974 (Leiden et al., 2001). Leiden et al. examined historical CFIT accidents and found that common underlying problems included communication issues between controllers and the flight crew, loss of vertical and horizontal SA, and crew resource management issues. That is, CFIT accidents are not attributable to mechanical errors or external (normal/abnormal) situations but primarily due to human errors inside the flight cockpit.

In order to reduce the accident rate and enhance aviation safety, various systems (e.g. altitude indicators, radio navigation, instrument landing systems (ILS), ground proximity warning systems) have been developed and introduced to overcome the issues associated with limited outside visibility for the pilot. Recent advanced devices include moving map displays, incorporating Global Positioning System (GPS) capability for improved navigational accuracy, terrain awareness warning systems (TAWS) and enhanced ground proximity warning systems. However, all of aircraft information display concepts require pilots to perform various mental transformations of display data (decoding) to support flight control in a real-time environment when outside visibility is restricted (Prinzel et. al, 2002). In addition, although the TAWS technology may help to mitigate some factors causing CFIT, its use generally follows the information processing model of "warn-act" and, therefore, requires the flight crew to be reactive rather than proactive in dealing with terrain hazards. Theoretically, TAWS provides a warning when the flight crew has already lost SA, and may not be optimal given the reaction time required to adequately recognize and assess the situation and initiate an escape maneuver (Moroze & Snow, 1999).

Snow and Reising (1999) stated that what is currently needed in terms of aircraft information systems is intuitive technologies that improve pilot SA with respect to spatial orientation (relative to terrain and flight path) without requiring the pilot to divert visual attention and cognitive resources away from possible external events and primary flight references. Therefore, a proactive system that can help prevent (versus just warn a pilot of) a potential collision with terrain is needed (Prinzel et. al., 2004a)

1.1.2. Synthetic / Enhanced Vision Systems (SVS/EVS)

NASA and its industry partners have designed and prototyped novel crew-vehicle interface technologies that strive to proactively overcome aircraft safety issues due to low-visibility conditions by providing the operational benefits of clear day flight through cockpit displays, regardless of the actual outside visibility conditions (Bailey et al., 2002). These technologies include the use of non-iconic features, such as synthetic vision systems (SVS) and enhanced vision systems (EVS).

Synthetic Vision is a computer-generated display image of the out-of-cockpit scene topography based on aircraft attitude, high-precision navigation instrumentation, and data on the surrounding terrain, obstacles, cultural features, etc. SVS databases have been developed to support this display technology with real-time integrity in order to ensure accurate pilot detection of real obstacles and to plan and verify accurate flight navigation. SVS displays can also support accurate traffic surveillance (Bailey et al., 2006).

NASA's SVS display concept is presented in Figure 1.1. The display is generated by visually rendering an on-board terrain database (with airport and obstacle database information) using precise position and navigation data obtained through GPS, with augmentation from differential correction sources such as Local Area Augmentation Systems (LAAS) and/or Wide Area Augmentation Systems (WAAS). The various data sets are blended with on-board Inertial Navigation System (INS) information (Bailey et al., 2002).

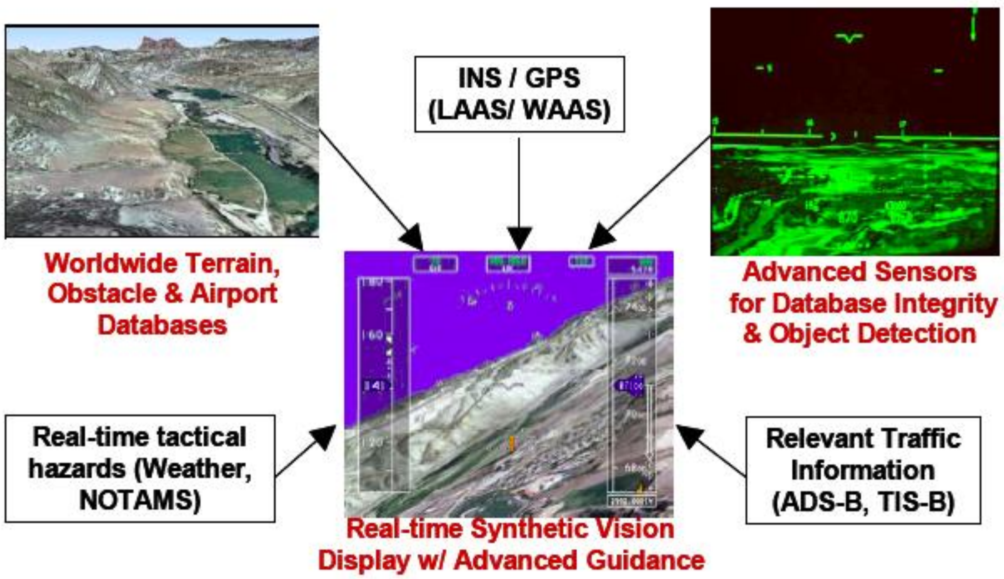


Figure 1.1. Synthetic vision system concept (Arthur et al., 2004).

It has been suggested that this display concept presents information to pilots with a level of realism that is comparable to flying under visual meteorological conditions (VMC), such as a clear and sunny day, regardless of the actual outside weather conditions (Prinzel et al., 2002). Consequently, it has also been speculated that the enhanced visibility provided with SVS may significantly improve terrain awareness and reduce the potential for CFIT incidents compared to current cockpit navigation display technologies.

Many laboratory research efforts have investigated replacing conventional attitude direction indicators, or primary flight displays, in transport airplanes with new display concepts, including SVS and EVS, in order to increase pilot SA as well as promote operational capabilities for landing in low-visibility weather conditions. To date, research has successfully demonstrated both the safety and capability benefits of SVS technologies in flight (Snow et al., 1999), landing (Prinzel et al., 2004a; Schnell et al., 2005; Bailey et al, 2006) and taxi operations (Wilson et al., 2002). (Some specific empirical studies are reviewed in the next chapter.) Thus, such display systems are expected to reduce the occurrence of accident precursors, including (Parrish et al., 2001; Bartolone et al., 2005):

- Pilot loss of vertical and lateral spatial awareness.
- Pilot loss of terrain and traffic awareness on approach.
- Unclear escape or go-around path after recognition of flight problem.
- Pilot loss of altitude awareness.
- Pilot loss of SA relating to the runway environment and incursions.
- Unclear path guidance on the surface.

All aircraft categories may also benefit from SVS applications, including general aviation aircraft, business jets, cargo and commercial airliners, military cargo and fighter jets, and rotorcraft. The greatest benefit from a safety perspective may be in commercial transport aircraft because of the frequency of flights and number of passengers. SVS images can be presented on head-down displays (HDD), head-up displays (HUD), head (or helmet)-mounted displays (HMD), and navigation displays (ND) in the cockpit.

An EVS is an electronic means by which to provide a display of the external (out-of-cockpit) scene by using an imaging sensor, such as a Forward-Looking InfraRed (FLIR) or millimeter wave radar (MMWR). Such sensors are used to penetrate weather phenomena, including fog, haze, rain, and snow. Figure 1.2 shows an example of an enhanced vision image obtained by FLIR. Like other advanced cockpit display concepts, the design and development of EVS technology was initiated for application to military aircraft. Currently, business jets incorporate EVS displays as a type of night-vision technology. Based on the development of this technology, in 2004, Section 91.175 of the US Federal Aviation Regulations (FAR) was amended such that pilots conducting "straight-in" instrument approach procedures may now operate aircraft below published Decision Heights (DH) or Minimum Descent Altitudes (MDA), when using an approved EVS presented on a HUD (Bailey, 2007).

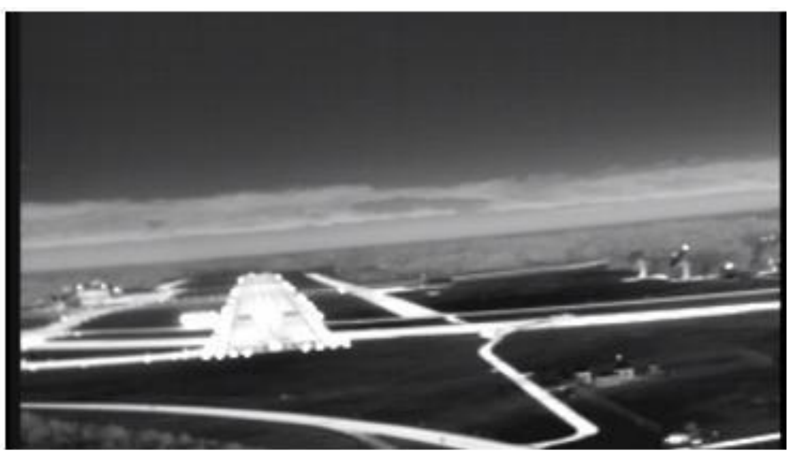


Figure 1.2. Enhanced vision image obtained by FLIR on approach to a runway under actual instrument meteorological (flight) conditions (Bailey, 2006).

The intended use of EVS mirrors SVS; both strive to eliminate low-visibility conditions that may cause major flight accidents and to provide the operational benefits of VMC, regardless of actual out-of-cockpit visibility conditions (Bailey et al., 2006). Bailey et al. (2006, 2007) stated that the use of SVS and EVS in the same display may be complementary, even though SVS and EVS had previously been perceived as competing technologies (Bailey et al., 2002). A combined system integrating both SVS and EVS imagery may reduce the disadvantages inherent to either system alone. The SVS technology has some advantages over EVS technology in providing terrain, path and obstacle awareness, which may be obscured by clouds, which an EVS sensor cannot penetrate. Another advantage of the SVS is that it can provide virtually unlimited visibility. On the other hand, EVS can provide a direct view of the external environment, independent of a database. That is, EVS can show “live” imagery of what actually lies ahead of the aircraft, while SVS cannot be used to detect dynamically changing scenery, such as other aircraft or ground vehicles. Beyond this, fixed obstructions may be missing from the terrain database used to support the SVS display, depending upon the database update rate (Ertem, 2005). A pilot using EVS can also develop an extremely high degree of confidence in the system output.

Since Bailey et al. (2002) suggested the use of both SVS and EVS concepts in the same HUD or HDD, some research (Theunissen et al., 2004; Schnell et al., 2005) investigated the usefulness of these concepts. This work has provided a basis for defining the requirements for display configurations by considering human performance implications. However, there remain a number of human factors issues relating to the development of SVS and EVS technology for effective implementation (Bailey et al., 2006). Corker and Guneratne (2002) categorized the human factors issues associated with SVS/EVS displays into three research areas: image quality, information integration, and operational concepts. Prinzel and Kramer (2006) summarized these issues as follows:

- ***Image Quality***: Field-of-view, display size, clutter, iconography, display contrast, and opacity.
- ***Information Integration***: Guidance, terrain presentation, cognitive tunneling, display integration, trend information, skill retention, and workload demand.
- ***Operational Concept***: Flight phase transition, crew interaction, failure modes, essential information, effect at various workloads, crew confidence in system, and resource management.

Among these issues, several have been investigated by experimental studies and several others are still unaddressed. These will be reviewed later.

1.1.3. Head-up display (HUD)

The purpose of a HUD is to provide primary flight, navigation and guidance information to a pilot in the forward (out-of-cockpit) field-of-view on a head-up transparent screen. Figure 1.3 shows a HUD (left, in the circle) and HDD (right). The HUD supports effective control of an aircraft by facilitating pilot simultaneous scanning of both instruments and the outside environment. Under instrument meteorological conditions (IMC) in landing, a pilot must rely on instrumentation until visually acquiring the runway. The capability to stay “head-up”, despite IMC conditions, is a significant advantage of advanced HUDs (Prinzel & Risser, 2004).

Figure 1.3. Head-up display (HUD) (left) and Head-down display (HDD) (right) in aircraft cockpit (Kramer et al., 2005).

In early use of HUDs in military aircraft, it was found that such displays could produce greater precision and accuracy than use of conventional flight instrument systems. As a result of the demonstrated benefits of HUDs, the installation of these displays in commercial aircraft has significantly expanded. HUDs are now established contents of aircraft cockpits, supporting additional operational capabilities and enhanced situational awareness, resulting in improved aircraft safety (Wood & Howells, 2001).

The symbologies used in HUDs are very similar to the symbols used in a Primary Flight Display (PFD). This eases pilot transitions from head-down instruments to the HUD symbology. Figure 1.4 shows a typical HUD (in-flight) with “primary” mode symbology.

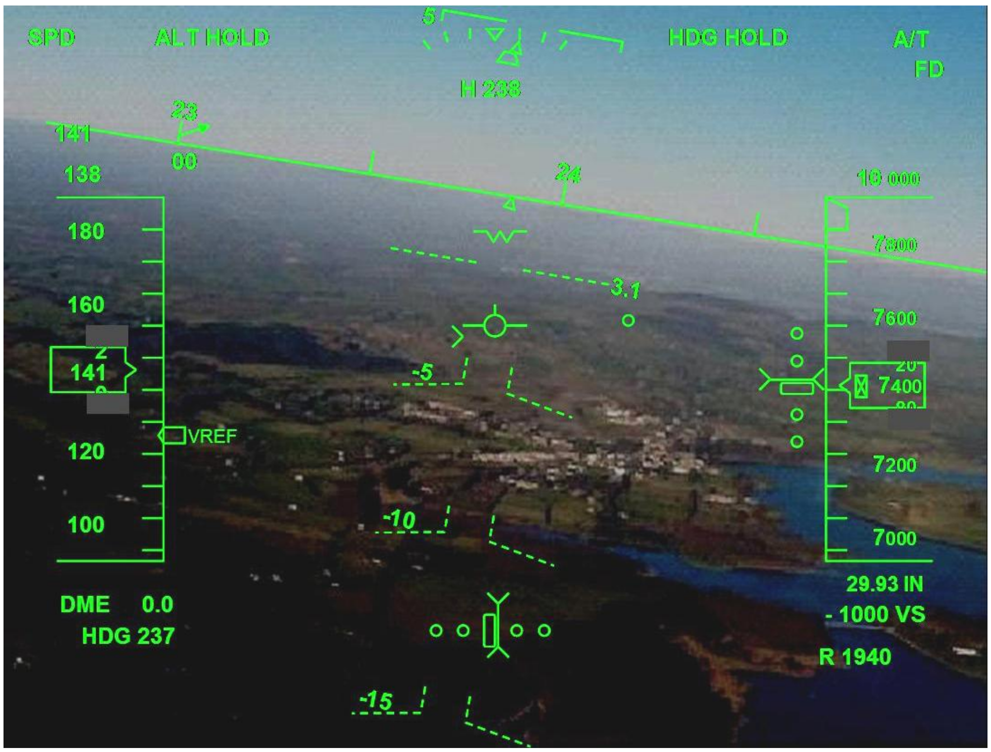


Figure 1.4. Typical HUD symbology overlaid on out-of-cockpit view.

HUDs have undergone continuous refinement for several decades and several advanced HUD formats have been investigated to date. Recent developments in HUD design include the use of pathway/tunnel/highway-in-the-sky (HITS) features, and EVS and SVS (Prinzel & Risser, 2004; Wood & Howells, 2001). The stroke of symbols and raster images making-up these features represent visual properties of the HUD (Wickens et al, 1998). The features are also typically presented as conformal displays, spatially overlaying the far visual domain (out-of-cockpit view) (Wickens, 1994). Related to this, previous research has investigated whether advanced HUDs with conformal symbology, promoting information proximity between spatial information and system information, provide an additional advantage for pilot performance over conventional HUDs or HDDs (Ververs & Wickens, 1998). General results revealed that advanced HUDs provide an advantage in the detection of events both in the symbology and the environment, as well as the benefit of reduced scanning over HDDs.

Based on the expected benefits of advanced HUD features, it may be worthwhile to retrofit HUD-equipped aircraft with SVS and EVS technologies by generation of synthetic/enhanced vision images as raster input sources to a stroke-on-raster HUD. Figure 1.5 shows examples of SVS-HUD (left) and EVS-HUD (right) concepts. As stated previously, the SVS and EVS features may be used simultaneously in the same display in a potentially complementary manner. It should be noted that both SVS and EVS features provide a 2-D perspective on terrain for pilots.

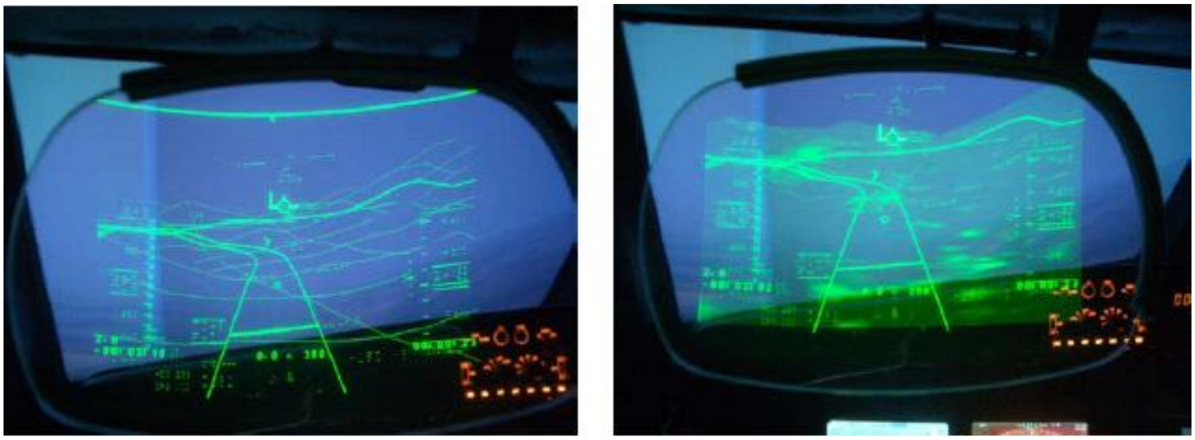


Figure 1.5. HUD with SVS terrain (left) and EVS imagery (right) (Schnell et al. 2005).

1.2. Previous Studies of Advanced Flight Displays

In this section, various historical studies relevant to SVS and/or EVS features are reviewed from a human factors perspective. In general, these studies have been conducted by assessing the effects of specific features or configurations of SVS and/or EVS technologies on human performance, in order to identify appropriate design guidelines for advanced cockpit displays. The studies can be categorized into three groups aimed at determining: how specific features of SVS affect pilot performance; why the display concept of SVS/EVS has been suggested and what corresponding studies have been conducted; and what research efforts have been conducted to investigate the effects of SVS and/or EVS in HUDs on pilot performance.

1.2.1. Use of SVS, path guidance and tunnel features

As previously stated, the use of SVS technology is expected to reduce aircraft accidents, in particular CFIT, due to the display information (terrain model) enhancing pilot SA under low visibility conditions. Several studies have been conducted to assess this expectation and to investigate the effect of specific SVS features on pilot performance. In general, studies have focused on the effects of display sizes and corresponding field of view (FOV), terrain textures, guidance images, and tunnel images on flight path tracking performance, SA, workload, or subjective display ratings.

Prinzel et al. (2002) conducted flight tests to evaluate the effects of three display concepts, including a HDD (Size A; 5.25” x 5.25”), a second HDD (Size X; 8”x 10”) and a HUD as well as two terrain texture concepts (photorealistic and generic) on SA and flight performance/error. Situation awareness was measured using the SA-SWORD (Situation Awareness – Subjective Workload Dominance) technique. Results showed that the HUD and size "A" SVS-HDD concepts significantly improved pilot SA and flight path control. However, there were no significant differences in texture concepts although subjects reported subjective preference for photorealistic texture for improved SA. In general, Prinzel et al. confirmed the hypotheses that SVS would provide safety and performance benefits over traditional EFIS (Electronic Flight Instruments) or EADI (Electronic Attitude Direction Indicators).

With a similar objective, Prinzel et al. (2004a) conducted two experiments to examine the efficacy of SVS displays and to develop FOV and terrain texture recommendations for cockpit display design. In one of their experiments, they investigated the effects of different types of displays, including a HUD and three sizes of HDD (A, X and D (6.25” x 6.25”)) for presenting SVS information, two types of textures (photorealistic and generic) and two runway conditions on performance, subjective preference ratings, workload and SA (using SA-SWORD). Results demonstrated that the different display sizes did not affect flight performance and that the use of the HUD for presenting SVS information reduced lateral path error, as compared to the HDD. Regarding the texture rendering, there was no significant effect on flight performance as a result of using different terrain texture images, i.e., photo-realistic versus generic texture.

The studies reviewed above have focused on nominal flight operations; however, other research has been conducted to examine the efficacy of SVS technology for CFIT prevention in off-nominal situations (e.g., Prinzel et al, 2003). In a first experiment by Prinzel et al. (2003), 10 display concepts, including two baseline conditions (a round-dials display and a PFD with no SVS texture), and various SVS textures were used to assess operator CFIT detection ability during incorrect altimeter setting scenarios. The SVS modes included constant color, elevation-based generic texture, photorealistic texture, and a grid fishnet. Results revealed that the use of SVS, in general, improved CFIT detection. In a second experiment, Prinzel et al (2003) evaluated four display concepts (a baseline 757 display, a Size “A” HDD with SVS, a Size “X” HDD with SVS, and a HUD with SVS) by measuring flight performance, SA and workload during a go-around situation. Situation awareness was measured using the Situation Awareness Rating Technique (SART) and SA-SWORD. Workload was measured using modified Cooper-Harper ratings. Results confirmed that the use of the SVS allowed pilots to detect CFIT more efficiently than baseline concepts. It was also revealed that a Size "X" HDD and HUD with SVS yielded lower workload and better terrain awareness.

These experiments demonstrated the general efficacy of the SVS concept. Consequently, the effects of guidance and tunnel images, combined with SVS technology, were investigated. Prinzel et al. (2004b; 2004c) conducted two experiments to compare different tunnel and guidance symbology concepts for synthetic vision display systems presented on HDDs and HUDs. They evaluated the efficacy of these concepts during complex, curved approaches under instrument meteorological conditions (IMC). In a first experiment, they focused on a SVS PFD and examined four tunnel concepts, including minimal, full or box, dynamic pathway and dynamic crow's feet, compared to a baseline (no tunnel) configuration. They also assessed three guidance symbologies, including an integrated cue circle ("ball"), a "follow me" aircraft concept ("ghost") and a "tadpole" guidance symbol, by measuring mental workload using the USAF (US Air Force) Revised Workload Estimation Scale, SA using SART and SA-SWORD, a subjective questionnaire, and flight path control (RMSE: Root Mean Square Error). The results of the first experiment revealed the baseline condition to be worse than other conditions including tunnel concepts, in terms of path control, workload and SA. On this basis, the second experiment evaluated two pathway tunnel concepts, including minimal and dynamic crow's feet and two forms of guidance, including tadpole and ghost concepts, for a HUD. Overall, the results demonstrated that presenting any kind of tunnel feature can produce better performance in terms of RMSE, workload and SA. It was also demonstrated that the concept of a dynamic crow's feet tunnel and tadpole guidance were most appropriate to use in a SVS HUD.

Schnell, Kwon, Merchant and Etherington (2004) also evaluated a SVS HDD against conventional glass cockpit displays to assess whether SVS technology could improve flight performance, SA and workload. However, they included a navigation displays (ND) in their simulation setup for providing pilots with more realistic flight situation information. They measured SA using SAGAT, mental workload using the NASA-TLX, flight technical errors (FTE) and eye movements of pilots when using three different configurations of flight decks. The configurations included a conventional PFD with ND, a SVS PFD with ND, and a conventional PFD with an exoview display. (This is a strategic/exocentric display that depicts the planned flight path in the context of the surrounding terrain. The depiction is centered on the aircraft.) Results demonstrated, in general, the use of the SVS display format to improve pilot performance by generating reduced FTEs, lower workload scores and short overall visual scan length. However, interestingly, there was no significant difference in SA across display conditions. That is, the SVS PFD with terrain representation did not seem to improve the terrain awareness of the pilot. The authors inferred that pilots relied on and trusted the pathway tunnel to the extent that they did not feel they needed to devote much attention to the aircraft-terrain situation. Schnell et al. (2004) also said that the lack of results on SA did not necessarily mean cognitive tunneling had occurred in the form of pilot reliance on the pathway tunnel image because the workload measures were lower in the SVS condition than with the conventional PFD. However, other studies have pointed-out that pathway tunnels may cause cognitive tunneling. According to experimental research (Alexander et al., 2003; Thomas & Wickens, 2004; Wickens et al., 2004), it was found that, while pathway tunnels in a SVS display can support better flight path tracking, they may degrade traffic awareness and pilot ability to detect unexpected events. The compellingness of the symbology may cause pilots to focus on virtual pathways and enter smaller but more frequent control inputs to maintain closer adherence to the flight path. Pilots may pay an undue amount of attention to the SVS display and far less to other displays, regardless of their relevance to certain tasks such as detecting outside-world unexpected events (Wickens et al., 1998; Thomas & Wickens, 2004).

1.2.2. Integration of SVS and EVS displays

While the use of SVS has been evaluated for benefits on flight safety, certification issues for operational use of the concept have arisen. Bailey et al. (2002) pointed out that the FAA (Federal Aviation Administration) has never “certified” a database which appears to support a display of primary flight information. They said the main hurdle to certification was to overcome the “hazardously-misleading information” conundrum. This implies that there may be a potential for hazardous situations when a

SVS database system creates and presents incorrect information to a pilot, who is relying on the system in an IMC situation. Theunissen et al. (2004) also noted that the Achilles-heel of any database-oriented system is that the quality of the match between the real and the synthetic world is influenced by the quality of data used to generate the synthetic world. This implies that no operations will be allowed in which an undetected error in a database used for synthetic vision can create a hazardous situation. With this in mind, Bailey et al. (2002) introduced two potential techniques to reduce the possibility of SVS displays presenting misleading information. One was development of a complementary system called Database Integrity Monitoring Equipment (DIME) and another was the additional use of EVS, real-time and non-database elements.

Historically, EVS and SVS have been perceived as competing technologies; both attempting to provide a complete picture of the terrain, the airport, and fixed/moving objects within the scene. However, neither technology completely and reliably provided the total picture outside the cockpit for pilots. Arthur, Kramer and Bailey (2005) sought to find an answer to which features should be used among EVS and SVS technologies to support reliable and accurate pilot comprehension of terrain. Although pilots indicated a preference for SVS over EVS, the authors suggested that it was necessary to develop integrated and fused EVS and SVS technologies to create “the best of both worlds”, rather than “EVS or SVS” exclusively. In this approach, the EVS imagery becomes more prominent in the absence of a “perfect” SVS display. In line with this suggestion, NASA developed a Sensor Enhanced-SVS (SE-SVS) concept, which utilizes the beneficial aspects of EVS and SVS while mitigating the negative aspects of each concept (e.g., misalignments in SVS images with terrain and low EVS image quality influenced by atmospheric conditions) (Bailey et al., 2002).

In other domains besides aviation, image fusion technologies for supporting human performance under low visibility environments have been investigated. For example, Bender, Reese and van der Wal (2003) presented an optimal image fusion algorithm based on various image sensors (e.g., Long-Wave Infrared (LWIR), Forward-Looking Infrared (FLIR) and thermal sensors) to support military drivers using vision enhancing Helmet-Mounted Displays (HMD). In general, image fusion technology has been considered to have advantages for navigation, surveillance, fire control, and missile guidance by improving position control accuracy. McDaniel et al. (1998) found a well-designed image fusion system to increase human visual scan comprehension. With this research in mind, the use of unmodified/pure SVS and EVS images in an integrated, single display format is expected to minimize operator visual scan and cognitive efforts by promoting visual proximity of terrain information with aircraft symbology information (Bailey et al., 2007).

However, Theunissen, Roefs, Koeners, Rademaker and Etherington (2004) performed a study to assess the combined effect of SVS images and symbology with EVS images on pilot preferences for desired display configurations. They manipulated three participant roles, including pilots flying with a co-pilot monitoring SVS integrity, co-pilots monitoring integrity, and pilots flying with integrity checking. They also studied two display types, including one with pilot selectable features and another with automatic feature selection. During the trials in which the display was pilot selectable, all possible combinations of seven features (sensor images, opacity of sensor images, field of view, guidance cues, runway outlines, display alignment, and obstacles) could be selected by pilots. Subjective workload ratings and questionnaire responses were collected across three legs of flight in approach and landing situations. Results demonstrated that pilots preferred the sensor image (EVS feature) to be on during the whole approach, and they also regarded synthetic overlaying of the runway outline as an important feature. It was noted that the best approach to feature integration depended on the tasks to be performed and the intended use of the resulting information.

While Theunissen et al. (2004) focused on the effects of specific symbologies in combination displays, Bailey, Kramer and Prinzel (2006, 2007) compared the general effects of EVS/SVS concepts with/without pathway tunnel images. Bailey et al. (2006, 2007) conducted an experiment to evaluate the complementary use of EVS and SVS technologies, focusing on integration and/or fusion of the display concepts during low-visibility approach and landing operations. In the experiment, four HUD display concepts were tested, including a combination of two factors with two levels, each. The first factor was the type of raster background presented, including EVS-only and a fusion of SVS/EVS imagery. (It should be noted that the term, “fusion”, in the context of this experiment, did not refer to the combination of SVS/EVS images, at the same time.) The fusion raster started out as a pure SVS image and transitioned through a fused SVS/EVS presentation (beginning at 600 feet above field level (AFL)), and ended with a pure EVS image by 500 feet AFL. Between 600 feet and 500 feet AFL, a step function modulated the fusion. The authors said that the reason for adopting this fusion approach was that it maximized image legibility and minimized image confusion during different legs of the approach. However, the authors did not provide empirical rationales for their selection of the various altitudes used as triggers in the fusion approach. It is suspected that their flight simulation (at NASA Langley Research Center) was originally programmed to transition from SVS to EVS at about 500 AGL, as an EVS camera can only give a usable image when the aircraft is quite low and close to the ground. In addition, the SVS is the only display system that can present a terrain image beyond the range of the EVS camera system. Another factor in the study was the kind of symbology presented on the HUD, including standard HUD symbology versus standard HUD symbology enhanced with pathway guidance and a runway outline (tunnel). In order to collect pilot responses, Bailey et al. (2006, 2007) measured path errors and pilot control inputs during each experimental trial. They also collected subjective questionnaire responses, SA ratings using SART and SA-SWORD, and workload ratings using the Air Force Flight Technical Center (AFFTC) revised workload estimation scale. The overall results showed that significant improvements in pilot SA, without increases in workload, could be provided by the fusion display and the pathway tunnel image. Regarding flight performance (path errors), while the raster types (EVS versus Fusion) did not affect performance, the presence of the pathway tunnel significantly decreased path errors. This result was in line with previous studies (Alexander, 2003; Mckinley et al., 2005; Wickens et al., 2004), which showed a synthetic tunnel image to improve flight performance. In other words, it can be said that the critical objective of improving flight performance might be largely facilitated by the presentation of tunnel images in a HUD.

In other research, Schnell et al. (2005) examined how EVS images should be overlapped in a HDD and how SVS/EVS features affect flight performance and SA. They used an elaborate simulation setup including various configurations of HDD-PFDs and NDs. They manipulated display combinations including: baseline displays (standard PFD and ND); baseline PFD and terrain textured ND; SVS PFD and ND; SVS PFD and ND with added EVS features on the PFD, as an overlay inset; and SVS PFD and ND with EVS features on the PFD, as a blended inset. They also presented conditions with and without the presence of a HUD. In conditions with the HUD, the display format was similar to HDD-PFD format but didn't include the combination of SVS/EVS imagery. When the PFD contained the SVS and EVS inset concepts, the HUD featured only EVS. The dependent variables were lateral and vertical flight technical errors (FTEs), workload scores (using AFFTC/AWRES (Air Force Flight Test Center workload assessment scale)), and SA (using SART) scores. Results confirmed the SVS features, including pathway guidance, to improve FTEs, SA and workload. Additional EVS feature insets in the SVS-HDD did not show significant effects. Regarding use of the HUD, results also indicated the HUD reduced FTEs. This effect was likely due to the addition of conformal features in the HUD, when compared to the reduced image of the HDD.

1.2.3. Applying SVS/EVS features in HUDs

Regarding the efficacy of using conformal feature concepts, such as SVS, in HUDs instead of in HDDs, several studies have confirmed SVS-HUDs to be comparable to SVS-HDDs in terms of flight performance, pilot SA or workload (Prinzel et al., 2003, 2004a). Such displays may even improve pilot abilities (Alexander et al., 2003; Thomas & Wickens, 2004; Wickens et al., 2004; Schnell et al, 2005) because the SVS-HUD can provide conformal guidance information. However, since the visual properties of HUDs are different from those of HDDs, including monochromatic color (HDDs present full color), size, location, and overlap of display symbology (strokes) and raster images on out-of-cockpit views, different research studies have investigated the application of specific SVS/EVS concepts in HUDs. In general, these efforts have examined the effects of HUD image features, including terrain texture, pathway tunnels and EVS images, on flight performance, workload, SA or subjective preferences. The objective has been to demonstrate the efficacy of the features as well as determine optimal HUD configurations.

Snow and Reising (1999) conducted an experiment to investigate how terrain texture concepts in a HUD affect flight performance and SA. They measured performance by calculating RMSE and SA using SA-SWORD and SAGAT for four SVS terrain texture concepts. These included grid, texture-map, partial grid, and a baseline condition that had no terrain features but presented a pathway image. The results confirmed that the SVS-HUD with the terrain feature produced higher SA than without and that grid terrain texture appeared most appropriate for use. However, there were no significant differences in flight performance across the terrain textures.

Using a similar approach, Snow, Reising, Kiggett, and Barry (1999) examined the utility of pathway-in-the-sky HUD symbology in three experiments. In a first experiment, they compared a pathway guidance HUD and a Military Standard (MS) HUD under IMC at night. The complexity of the approach (simple or complex) to landing was also manipulated. Results demonstrated that pilots were more accurate in the flight performance when flying with the pathway HUD, as compared to the MS-HUD. In a second experiment, they compared pilot performance with the pathway HUD under three different visibility conditions (VMC, partial IMC and full IMC). Results demonstrated performance under IMC to be equivalent to performance under VMC, when using the HUD with tunnel symbology. In a third experiment, Snow et al. tested the utility of the pathway guidance HUD in flying complex flight paths. They compared IMC night versus IMC day conditions and manipulated the format of a synthetic terrain model in the HUD (grid, partial grid, texture map or none). They used the SAGAT and SA-SWORD to evaluate HUD use. Results revealed a significant effect of the synthetic terrain format on pilot SA. Pilot SA was best with the grid and texture map conditions. The no synthetic terrain condition greatly decreased SA.

In an experimental study by Kramer, Arthur, Bailey and Prinzel (2005), the effect of SVS and EVS concepts in several display combinations, including HUD, HDD-PFD and ND, were examined. Four types of display configurations were manipulated: (1) All baseline displays, including HUD with no conformal images (SVS or EVS), PFD and

ND; (2) EVS featured HUD, baseline PFD and ND; (3) SVS featured HUD, PFD and ND; and (4) SVS-PFD and ND without the presence of a HUD. Response measures, including path control performance, workload and SA (using SA-SWORD) were collected and analyzed. Results revealed that the baseline combination produced higher workload than the other configurations and that the use of SVS improved pilot SA. Path control performance was not significantly different across the display combinations as all conditions involved pathway tunnels.

Regarding the use of combined terrain images from SVS and EVS technologies presented on a HUD, Kaber et al. (2007a, b) measured pilot perceptions of HUD clutter when using SVS-only, EVS-only, or combined imagery, among other display conditions. Prior research suggested that multiple iconic and non-iconic images in the same display produced clutter and negatively impacted pilot performance (Ververs & Wickens, 1998). Kaber et al. collected pilot perceived clutter ratings with HUDs, including combinations of SVS terrain, EVS images, pathway tunnels, traffic information (Traffic Collision Avoidance System; TCAS), and full (vs. reduced) overlapping PFD information. They found that all HUD features were significant predictors of perceived clutter and all two-way interactions involving SVS, EVS or TCAS features were critical. In general, the number of features in the HUD was correlated with perceived clutter ratings. It was also found that, when PRIMARY mode symbology was active, the presence of SVS, EVS or Tunnel (HITS) did not appear to influence pilot ratings. In addition to this, Alexander et al. (2008) identified pilot perceptions on HUD clutter to be primarily influenced by two major factors; one was visual density, a bottom-up (data-driven) factor, and another was information density, a top-down (knowledge-driven) factor. While the studies conducted by Kaber et al. (2007a, b) and Alexander et al. (2008) used static images of HUD configurations for assessing existence of perceived clutter, later study (Kaber et al., 2008) was conducted using an advanced flight simulator (Integration Flight Deck simulator at NASA Langley, Figure 3.1). Using this facility, Kaber et al. (2008) assessed influences of pilot experience, HUD configuration, flight segment, as well as flight workload on perceptions of display clutter and cognitive load, and flight task performance, in order to develop and validate a new measure of display clutter. They developed a multidimensional measure of display clutter, revealed relations between perceived clutter with subjective workload, and found negative effects of "low" and "high" clutter displays on and flight task performance (stability and RMSE). However, individual effects of terrain features on flight performance, SA, or workload were not assessed in the experiments by Kaber et al. (2007a; 2008) and Alexander et al. (2008).

1.2.4. Summary of previous studies on advanced display concepts

In summary, there is great deal of evidence that advanced synthetic images in cockpit displays, including SVS terrain features and tunnels, improve flight performance and (or) pilot SA, and reduce workload (Snow & Reising, 1999; Prinzel et al., 2002, 2003). In specific, HITS (tunnel) images have been revealed to be a major factor in improved pilot performance , particularly flight path tracking accuracy (Alexander et al., 2003; Bailey et al., 2006, 2007; Prinzel et al, 2004b, 2004c; Wickens et al, 2004). Since SVS features are generated from a database, they have been considered to have the potential disadvantage of providing pilots with inaccurate information, consequently, the use of combined vision system information (using both SVS and EVS) was suggested (Arthur et al. 2005; Bailey et al., 2003; Ertem, 2005). Research has investigated the utility of terrain feature combinations (SVS and EVS) in HDD design (Schnell et al., 2005) versus conventional flight instrument displays. The general finding here has been that the inset of EVS features in a SVS-PFD image did not improve flight performance. The use of non-iconic terrain features (SVS and EVS) has also been presented in HUDs and empirically evaluated. Specifically, conformal images have been presented to assure HUD information proximity with out-of-cockpit imagery. The general findings were that a SVS-HUD improved pilot SA (Snow & Reising, 1999) and that an EVS-HUD generated lower mental workload for pilots than traditional HUDs (Arthur et al., 2005; Kramer et al., 2005). However, there is a lack of research evaluating the effects of SVS or EVS alone and the combination of the two terrain features in HUDs on human performance, such as flight path control, mental workload, SA, and subjective pilot preference. Only the study by Kaber et al. (2007a, b) assessed perceived clutter ratings of various combinations of HUD features. In addition to this, although there have been many studies comparing HUD feature conditions in order to demonstrate which specific conditions improve pilot performance, there is a lack of research explaining the results in terms of human cognitive behavior. Furthermore, while some studies investigated use of HUD conditions in different flight scenarios, no study has been conducted to identify when and why it is best to use particular displays.

1.3. Human Factors Issues

Based on the limitations of previous research identified in the prior section, there is a lack of studies evaluating effects of SVS/EVS features in HUDs on human performance and explanations of effects using, for example, cognitive task analysis approaches. It is necessary to review some human factors issues in order to address these limitations. First, since the primary objective of advanced flight display technologies (SVS/EVS) is to improve pilot SA, there is a need to review SA definitions and measurement techniques in order to identify a comprehensive definition and measure of pilot SA. Second, since explanation of how specific advanced display features affect pilot performance may be achievable through cognitive task analysis, it is also necessary to be aware or what analyses have already been conducted on pilot use of advanced cockpit display technologies. Beyond this, it is important to identify how cognitive task analysis can be used in the present research.

1.3.1. Pilot situation awareness

The human factors issues suggested by Corker and Guneratne (2002), and summarized by Prinzel and Kramer (2006), in the category of "information integration," included investigating pilot workload demand when using SVS and EVS technologies in advanced cockpit displays. Regarding this issue, many empirical studies have assessed pilot cognitive load in order to compare types of display configurations. However, the main objective of applying non-iconic and conformal features, such as SVS/EVS in HUD design, is to enhance pilot SA and to reduce flight accidents, including CFIT. Unfortunately, there is no commonly accepted definition of SA and specific measurement approaches are either not defined or have serious drawbacks in terms of subject bias in responding or interference in performance. Here, definitions of SA, why SA is critical in flight situations, and how pilot SA can be measured are reviewed in detail.

1.3.1.1. Definition of SA

Flight instructors and pilots have held the notion that successful flight results when a pilot has the "big picture" in mind. Conversely, when problems occur due to pilot error, it is interpreted that the pilot missed the picture or (s)he had an incorrect picture. In the past decade, human factors researchers have attempted to transform this notion into a formal psychological construct (Uhlarik & Comerford, 2002). The most acceptable construct is the concept of SA. Uhlarik and Comerford said the concept of SA is especially compelling in the operational setting of aviation, which involves the operation and control of a complicated system in a dynamic environment. The human has to integrate widely disparate and sometimes inconsistent inter-sensory input (e.g., visual, auditory, and tactile) with elaborate cognitive models of the machine and the operating environment to control the movement of a vehicle through a medium. Based on this concept of SA, the construct has been extended to other domains beyond flying aircraft, including air traffic control (Kaber et al., 2006), operation of nuclear power plants (Hogg et al., 1995), automobile driving (Ma & Kaber, 2005), and medical procedures (Gaba et al., 1995). Uhlarik and Comerford (2002) said there are common aspects in such domains. For example: (1) the environment is dynamic and information rich; (2) the human may sometimes experience high mental workload; (3) extensive training is usually required; and (4) time is often constrained. A dynamic system can be defined as a system in which the environment is dynamically changing and in which the operator is responsible for maintaining or achieving particular states or goals in a defined time frame. The concept of SA originated from the piloting environment and it has been proven to be a critical factor for flight safety. Several case studies and analyses of existing databases have confirmed that a loss of SA is an important precursor to (aviation) performance failures (Durso & Gronlund, 1999).

Although there have been many definitions of SA in particular domains or environments, there is a lack of a generalizable or comprehensive definition of SA. In order to generalize the definition of SA, Breton and Rousseau (2001) surveyed 26 different definitions and classified them into two classes, including process-oriented versus state-oriented definitions. A process-oriented SA theory has been advocated by Smith and Hancock (1995). Smith and Hancock defined SA to be a dynamic concept that exists at the interface between the agent and its environment, like adaptation to an environment. They also proposed that SA is adaptive, externally directed consciousness. Adaptation was defined as that process by which an agent (human) channels its knowledge and behavior to attain goals as tempered by the conditions and constraints imposed by the task environment. Smith and Hancock regarded consciousness to be that part of an agent’s knowledge-generating behavior that is within the scope of intentional manipulation. Based on these analogies, they viewed SA as generating purposeful behavior directed toward achieving a goal in a specific task environment. Therefore, they asserted that assessing SA should consider the agent-environment relationship and depend on experience in the environment and development of alternative action plans.

This perspective of SA is compatible with another viewpoint of SA proposed by Sarter and Woods (1995). They suggested that: "the term situation awareness should be viewed just as a label for a variety of cognitive processing activities critical to dynamic, event-driven, and multitask fields of practice (Page 16)." Similarly, Rousseau, Tremblay and Breton (2004) said that a process-oriented definition of SA can be associated with an operator-focused approach. The operator-focused approach is concerned with the properties (basic mechanisms) of the operator as they determine SA. This theory can be applied to a piloting environment with advanced HUDs where the goal is to control an aircraft in low-visibility conditions while avoiding crashes into terrain. Behaviors may include understanding the terrain out-of-the-cockpit when SVS and/or EVS features are overlaid in a HUD. Control activities may include following a designated flight path based on previous experience. Therefore, with respect to the process-oriented definition, SA would generate instant knowledge and the actions required to achieve goals, based on information displayed in a HUD.

Competing with the process-oriented theory, Endsley (1988, 1995a) proposed a state-oriented definition of SA (Breton and Rousseau, 2001). Endsley said "Situation awareness is the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future." Endsley further claimed that SA is a state of knowledge that needs to be distinguished from the processes used to achieve that state, which is opposite to the view of Smith and Hancock (1995) and Sarter and Woods (1995). In this regard, Endsley also said the process-oriented theory of SA should be referred to as "situation assessment." This distinction between a "Process" and "State" definition of SA is of considerable importance and it has been influential in the development of measures and modeling efforts (Rousseau et al., 2004). As process-oriented theory was associated with the operator-focused approach, the state-oriented definition of SA can be associated with a situation-focused approach and characterized as driven by the properties of the situation. The situation-focused approach views SA as determined by the environment or situation in which the operator is at work (Rousseau et al., 2004).

According to Endsley's definition, SA consists of three hierarchical levels. The first level deals with perception of situation data; the second is the ability to comprehend the situation data; and the third level deals with the ability to use the data in projection of future states. With this definition in mind, it is easy to understand why achieving high SA is critical to improving flight safety, or preventing accidents in situations involving hazardous terrain and low visibility. For example, a pilot may perceive terrain along the flight path (Level 1), based on chart use, and (s)he may recognize the terrain to be hazardous based on comparison of a planned flight level with the terrain height (Level 2). Finally, based on projection of future aircraft states, the pilot can determine whether the flight will crash into the terrain, and (s)he will try to control the aircraft to avoid such an incident (Level 3). Thus, it is speculated that supporting pilot achievement of these three levels of SA through advanced cockpit display concepts, such as SVS and EVS, may improve flight performance and safety.

Another way to classify definitions of SA, aside from process versus state theories, is to consider the frameworks from which these theories emerged. Durso and Gronlund (1999) identified two frameworks. These include Endsley's (1995a) framework and Adams et al. (1995) view of SA. Uhlarik and Comerford (2002) labeled them "use of an information-processing model" and "use of the perception/action cycle," respectively. Endsley's (1995a) theory is similar to other general models of human information processing (e.g., Wickens & Hollands, 2000). That is, the information processing mechanisms in Endsley's concept include short-term sensory stores (STSS), schemata, attention, etc., also identified in Wickens and Hollands' (2000) theory. The Perception/action cycle framework (Adams et al., 1995) considers SA as the product of existing schemata that direct perceptual exploration. The perception/action cycle consists of three elements: (a) the object (i.e., information in the environment); (b) the schemata (i.e., internal knowledge from training/experience); and (c) exploration (i.e., a search of the environment). This cycle was hypothesized as: the object modifies the schema, the schema directs exploration, and exploration leads to sampling of the object. Adams et al. explained SA in terms of this perception/action cycle, but unlike Endsley, they suggested that SA should be conceptualized as both a product and a process. This concept is also compatible with the Smith and Hancock' s (1995) view of SA, because, as stated above, Smith and Hancock defined SA as, "adaptive, externally directed consciousness," and they utilized the perception/action cycle to conceptualize and define SA (Uhlarik & Comerford, 2002). However, Uhlarik and Commerford also noted that there were criticisms of each framework. Common problems among the two frameworks are that they include psychological constructs that are themselves not well-understood (i.e., attention and STSS in the information processing framework; semantic memory and schemata in the perception/action cycle framework). There is another criticism of the information processing model, specifically the process of achieving SA appears relatively static and finite, while the perception/action cycle model emphasizes the dynamic nature of SA. The perception/action cycle model also has limitations. The approach provides no suggestion as to how the product (i.e., the state of the active schemata) or the process (i.e., the state of the perceptual cycle) of SA can be measured. This criticism of the perception/action cycle model is quite critical to the use of its definition of SA in order to evaluate the affects of display alternatives on pilot performance.

1.3.1.2. Measurement of SA

Just like there are various definitions of SA, various methods to measure SA have been introduced. Fracker (1991) and Vidulich (1992) surveyed SA measurement techniques and categorized them as belonging in three major categories, including: (1) explicit measures; (2) implicit measures; and (3) subjective ratings. Explicit metrics probe the contents of subject memory to determine whether mission critical information is appropriately represented. Endsley (1995b) suggested that explicit measurement approaches could be subcategorized into three types: (1) retrospective measures, which assess SA after completing a task; (2) concurrent measures, which are used during the course of a task, such as verbal protocol analysis; and (3) measures utilizing a simulation freeze technique, like SAGAT. While Endsley (1995b) suggested that explicit measures are objective because the collected data can be compared with the true state of the situation, Fracker (1991) considered explicit measures to be subjective because the data are acquired by self-report rather than assessment of observable behavior. One disadvantage of explicit measures is that a normative model of the domain (e.g., aviation), and how operators are expected to behave may be difficult to develop because the task environment is dynamic and complex. Therefore, it is difficult to understand how an explicit measure could be developed outside of a laboratory setting (Uhlarik & Comerford, 2002).

Implicit measures of SA utilize task performance to infer SA based on the assumption that SA is correlated with performance. Endsley (1995b) also divided implicit measures into three categories including: global measures, external task measures, and embedded task measures. The advantages of using implicit measures are that they are objective, unobtrusive and easy to use (Endsley, 1995b; Fracker, 1991). On the other hand, implicit measures have the disadvantage that performance may not necessarily reflect SA in many task situations (Sarter & Woods, 1995). For example, it is possible that poor performance may be a result of something other than low SA (e.g., lack of task resources) (Uhlarik & Comerford, 2002).

Subjective rating techniques ask the operator to directly assign a value to represent the quality of SA they feel they experienced while performing a task. Rating techniques also include three subcategories (Endsley, 1995b): (1) direct self-rating, such as the 3D (dimension) or 10D SARTs (Taylor, 1989); (2) comparative self-rating (e.g., SA-SWORD); and (3) observable ratings. Advantages of subjective rating techniques are that they are easy to use, inexpensive to implement and practical because they can be used both in simulations and in the actual task environment (Metalis, 1993; Uhlarik & Comerford, 2002). However, disadvantages of subjective rating techniques have been identified to include participant ratings being affected by their performance on a trial and direct self-ratings collected at the end of task being prone to rationalizations and overgeneralizations by participants (Endsley, 1995b).

1.3.1.3. Pilot SA with SVS/EVS display features

Many previous studies assessing the effect of SVS or EVS features on pilot SA have used SART (Bailey et al, 2006, 2007; Prinzel et al., 2003, 2004b, 2004c; Schnell et al., 2005) or SA-SWORD (Kramer et al., 2005; Prinzel et al., 2002, 2004a, 2004b). These are subjective rating techniques. Relatively, few studies have used SAGAT for assessing SA (e.g., Schnell et al., 2004). Although subjective methods provide the advantage of ease of implementation, the use of SART or SA-SWORD to measure pilot SA poses limitations as identified above, including participant bias. Sarter and Woods (1995) criticized self-rating measures such as SART because they ignore the process of achieving SA and only measure SA as a product. Endsley (1995b) also asserted that participant ratings on SA may not only be affected by their performance in a test trial but their perceptions of mental workload as well. Others have said that subjective ratings may actually measure an operator’s confidence regarding SA rather than SA itself (Uhlarik & Comerford, 2002). SA-SWORD shares this shortcoming and it can only be used in contexts where a within-subjects experimental design is used (Uhlarik & Comerford, 2002).

One of the most well-known and validated direct measurement techniques is SAGAT, which was developed by Endsley (1995b). In applying SAGAT, a task simulation is frozen at random points in time and queries are posed to the operator in order to assess their perceptions of the situation at that time (Endsley, 2000). Use of SAGAT allows for evaluation of operator SA at all three levels (perception, comprehension and projection) through corresponding questions. Regarding the use of this technique, Sarter and Woods (1995) observed that halting a simulation and prompting a pilot for information concerning particular aspects of the situation was likely to disturb the very phenomena being measured. However, counter to this criticism, Endsley (1995b) demonstrated that subject accuracy in responding to SAGAT questions was not affected by the amount of time elapsed after a task freeze. Endsley also showed that subsequent task performance was affected by neither the duration nor the frequency of SAGAT freezes.

In research conducted by Snow and Reising (1999) to examine the effect of pathway guidance and synthetic terrain textures in a HUD on SA, they compared and contrasted SA-SWORD, as a subjective measure of SA, with SAGAT, as an objective measure, using experimental data. In their conclusion, they said they could not demonstrate explicit differences between the two measures. However, they did confirm that interruptions of simulations, as part of the SAGAT technique, did not affect pilot performance. This finding was in-line with Endsley's results. They concluded that the SAGAT was a useful objective SA evaluation.

1.3.1.4. Levels and types of SA

In the aviation domain, there may be multiple types of SA for supporting different tasks and cognitive behaviors. Wickens (2002) divided pilot SA into three components including: spatial awareness, system (mode) awareness, and task awareness. The concept of spatial awareness is inherent in the task of moving an aircraft through a 3-D space which can be filled with hazards. System awareness concerns a pilot's comprehension of aircraft status and mode, which may affect pilot performance (e.g., automation mode awareness). Finally, task awareness relates to a pilot's knowledge of aviation control, navigation, and communication (with a co-pilot or air-traffic controller), and systems management (e.g., managing fuel, cabin pressure, electricity).

Related to this concept of pilot SA, Bolton, Bass and Comstock (2007) introduced judgment-based measures of spatial awareness (SpA) to evaluate terrain texture and FOV features of a SVS-PFD in an experiment. Like previous studies examining the effect of SVS features, they also assumed that SVS technology could help prevent CFIT by enhancing pilot SpA. Based on Wickens' definition of SpA (Wickens, 2002) and Endsley's (1995b) concept of the three levels of SA, Bolton et al. (2007) identified three levels of SpA with respect to terrain: identification of the terrain (Level 1), its relative spatial location (Level 2), and its relative temporal location (Level 3). In their experiment, the authors investigated how SpA derived from a SVS-PFD was affected by the three leading texture types (fishnet, photo and elevation) and the two leading FOVs (30° and 60°) among display manufacturers, in all combinations. They included pilot judgments involving directional errors in relative angle, distance and height, and abeam time with respect to a terrain point. SA probes occurred at all three levels of SpA. The experimental results showed that spatial awareness was best facilitated by the elevation fishnet, photo fishnet, and photo elevation fishnet textures. Since Bolton et al. presented 5 second videos for each experimental condition to participants and asked them to assess the questions without allowing any dynamic control of the flight simulation, they were not able to use SAGAT and freezes of a dynamic task situation for SA assessment. They could only measure spatial awareness rather than system and task awareness. However, if the pilot could control a flight (simulator) in a dynamic situation, it would be worthwhile to measure all three aspects of pilot SA, according to Wickens (2002), at all three levels of SA defined by Endsley (1995b), by using SAGAT.

1.3.1.5. Summary of pilot SA

In this section, definitions of SA were reviewed, including comparison of process-oriented theory versus state-oriented theory as well as two frameworks of SA (information-processing model versus perception/action cycle model). In general, Endsley's definition of SA (1988, 1995b), categorized as a state-oriented theory and based on an information-processing model, has several advantages for the present study for examining pilot SA with specific display features. As reviewed, Endsley defined SA as "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future". Endsley said SA is a mental "snapshot" of a dynamic situation, forming a basis for decision making at a particular instant of time. The reason why this theory is considered to be appropriate for the present study is that Endsley's concept and definition have been previously applied in aviation tasks (Bolton et al, 2007; Schnell et al., 2004) and in other domains, including air traffic control (Kaber et al., 2006) and driving (Ma & Kaber, 2005). This research was facilitated, in part, by use of Endsley's measure of SA (SAGAT). That is, while other theories or frameworks of SA do not provide suggestions for measuring SA, Endsley's theory supports a validated measurement technique.

Measurement techniques for assessing SA were also reviewed. Among the various techniques, SAGAT has the greatest applicability to the present research. SAGAT allows for direct, objective assessment of SA by making comparisons of operator responses to knowledge questionnaires with the “ground truth” of a domain simulation in a dynamic environment for the three levels of SA (perception, comprehension and projection). Thus, SAGAT has been identified as a useful measure for evaluating pilot SA (Snow & Reising, 1999) and has been used in several studies (Bolton et al., 2007; Schnell et al., 2004). In aviation tasks, three types of pilot SA have been identified, including spatial awareness, system awareness and task awareness (Wickens, 2002). Consequently, in this research, SAGAT was used with queries for evaluating the three levels of SA along with the three types of pilot SA.

1.3.2. Cognitive Task Analysis (CTA)

In general, cognitive task analysis (CTA) has emerged as a useful technique for exploring how users interact with complex systems (Endsley, 1993). CTA has been used for evaluating user interfaces, in terms of human internal/external behaviors. In this section, a general overview of CTA, previous CTA efforts focused on the aviation domain, and a specific CTA method for studying complex dynamic systems will be reviewed. In this study, a cognitive task analysis approach was used to provide further explanation of the effects of non-iconic conformal features in a HUD on human performance.

1.3.2.1. Overview of CTA

CTA is analysis of the knowledge, thought processes, and goal structures of cognitive tasks (Hollnagel, 2003). CTA differs from traditional task analysis by describing the knowledge required for a task and how knowledge is used in decision making, situation recognition, or problem-solving, rather than just procedures (Kireas, 1997a; Wei & Salvendy, 2004).

CTA is appropriate for cognitively complex, or ill structured tasks that occur in dynamic, uncertain, multi-tasking, real-time operational domains (Gordon, 1995; Gordon and Gill, 1997). For this reason, CTA has been applied to various domains and tasks, including weather-related decision making in business aviation piloting (Latorella et al., 2001), SA analysis for air-to-air combat fighters (Endsley, 1993), and commercial jet aircraft piloting during instrument approaches with a HDD SVS (Keller et al., 2003).

The technique can be used to design new system interfaces or evaluate existing interface to develop expert systems and serve as a basis for operator selection and for training purposes (Wei & Salvendy, 2004). In addition to this, CTA can be used a basis for human performance modeling (HPM), which is a valuable research tool to understand new systems and their impact on human behavior and workload. Human performance models preclude the need for more costly methods of human-in-the-loop experiments or simulation (Keller et al., 2003) to evaluate systems. Zachary, Ryder and Hicinbothom (2000) said that CTA is required to build efficient cognitive models in complex real-time domains. Kieras (1997a) noted that CTA is critical for developing computational cognitive models (e.g., NGOMSL)(Kieras, 1997b) for cognitively demanding tasks.

There is currently a wide variety of CTA techniques available for research and human-machine system studies. Among the techniques, the most commonly used involve some form of interview, protocol analysis, scaling method (e.g., multidimensional scaling (MDS)), neural network modeling, computer simulation or error analysis (Redding & Seamset, 1994). O'Hare, Wiggins, Williams and Wong (1998) summarized two categories of such techniques. The first group of techniques focuses on identifying the inherent constraints in the work domain, based on decomposition of the functional structure of the task. O'Hare et al. (1998) also said the advantage of this approach is that it can be used as a basis for developing prototypes of operator interface design in complex systems not yet in operation. The second group is based on an analysis of actual user activity in an already functioning system. This approach is not only used for designing user interfaces but also for developing operator training programs.

Wei and Salvendy (2004) also classified various CTA method into four families based on their formality and analysis mechanisms. The class included:

- Family 1 – "Observations and interviews." Includes observations and unstructured /semi-structured/structured interviews.
- Family 2 – "Process tracing." Includes cognitive walkthroughs, verbal reports, and protocol analyses.
- Family 3 – "Conceptual technique." Includes diagramming, error analysis, psychological scaling/rating and ranking, sensory-motor process charting,

and questionnaires.

- Family 4 - “Formal models.” Includes multi-dimensional scaling and use of GOMS (Goal, Operator, Method, and Selection rules) and EPIC (Executive-Process Interactive Control).

1.3.2.2. Previous task analysis efforts relevant to cockpit automation.

Endsley (1993) conducted a CTA to identify the SA requirements of air-to-air fighter pilots. She reviewed three CTA methods including unstructured interviews, goal-directed task analysis, and structured questionnaires. Among the three CTA techniques, she used Goal-directed task analysis (GDTA). The goal of GDTA is to identify the information processing or situation awareness requirements of system users. She said that the outcome of GDTA is a list of critical decisions and information requirements that can be used as a basis for display design, training program development, development of situation awareness assessment measures, and operator selection. Endsley (1993) used an unstructured interview approach aas part of GDTA with experienced pilots for eliciting knowledge an combat maneuvers. Consequently, she determined the specific elements required for SA in air-to-air combat missions.

Latorella et al. (2001) conducted a CTA of business aviation piloting in order to understand challenging weather-related decisions. As a primary technique, they used the advanced cognitive task analysis (ACTA) method developed by Militello and Hutton (1998). This method is less resource intensive than traditional methods. ACTA consisted of three sub-methods including task diagrams, knowledge audits, and extended simulation-based interviews. The results of a CTA demonstrated the role of expertise in business aviation decision-making in flying through weather and how weather information is acquired and assessed for reliability by pilots. The authors also analyzed the results in order to recommend design and training interventions to improve business aviation decision-making.

Regarding the use of SVS technology, Keller and Leiden (2002) and Keller, Leiden and Small (2003) conducted a CTA with commercial jet aircraft pilots during instrument approaches using four types of cockpit interfaces, under two basic types of approaches (visual and instrument), two existing types of precision approaches (ILS and RNAV(Area Navigation)) and an SVS approach. The ultimate objective of these studies was to develop a human performance model for understanding new systems (SVS technologies) and their impact on human task performance and workload without costly experiments or simulation. As part of the modeling effort, Keller et al. (2003) analyzed and described the pilot task for the SVS condition compared to the baseline conditions (visual, instrument ILS, and RNAV), using a CTA method. The Authors decomposed pilot displays and controls, pilot responsibilities, approach tasks (sequential and non-sequential tasks), task timelines, and cognitive decisions for each display interface. Although Keller et al. analyzed pilot tasks during approaches to an airport using an SVS, the SVS was presented on an HDD and the relation between the technology and pilot task performance was not validated in their study. That is, no data was provided indicating how the SVS display affected performance. Therefore, it is necessary to use a CTA method for explaining the effects of SVS and/or EVS HUD configurations on human performance, beyond Keller and Leiden (2002) and Keller et al. (2003) studies.

1.3.2.3. Verbal protocol analysis

In order to elicit the internal information used by an operator in a specific cognitive task, verbal protocol analyses have frequently been used (Koubek et. al., 1994; Zachary et al., 2000). Verbal protocol analyses make it relatively easy to obtain data at a low cost and deliver to process key information from an operator perspective (Wei & Salvendy, 2004). In general, verbal protocol analyses are conducted based on the assumptions that the analyses can be a source of hypotheses on cognitive processes and predictions about non-verbal behaviors; and second, if an operator says something, evidently they have knowledge somewhere in their heads.

Zachary et al. (2000) said that two general methods of verbal protocol analysis have been used. The methods are categorized according to the points in time at which a subject verbalizes his/her internal behaviors. One is the retrospective approach advocated by Klein, Calderwood and MacGreger (1989), which anchors subjects on specific behaviors in the past and guides them through a verbal inspection process. Another is the think-aloud method suggested by Newell and Simon (1972), which asks the subjects to speak out their internal thoughts, decisions and intentions in real-time while interacting with a task interface.

1.3.2.4. Summary of CTA approach in aviation applications

CTA has been used as a tool for understanding user external/internal behaviors when interacting with complex systems. Many CTA methods have been developed and applied to various domains. In the aviation domain, several studies have been conducted to examine pilot cognitive activities for acquiring pilot SA requirements (Endsley, 1993), understanding weather-related decision making (Latorella et al., 2001), and modeling human performance with SVS displays in HDDs (Keller & Leiden, 2002; Keller et al., 2003). However, there is a lack of research explaining pilot cognitive and/or psychomotor behavior when using SVS and/or EVS features in HUDs. Among various existing CTA techniques (Wei & Salvendy, 2004), verbal protocol analysis may be most useful for this type of research due to explicit representation of cognitive behaviors, ease of use and cost effectiveness. As Zachary et al. (2000) indicated, verbal protocol analysis can be categorized into two categories depending on the time at which subjects make verbalizations, concurrent with task performance or retrospectively. In this study, concurrent verbal protocol analysis was applied. That is, after pilots performed the test trials using various HUD configurations, they were asked to perform an additional trial to complete a concurrent think aloud protocol and the trial was recorded. Consequently, the CTA using verbal protocol analysis was expected to help explain the effect of SVS and/or EVS in HUDs on pilot performance.

2. PROBLEM STATEMENT

2.1. Limitations of Previous Research

Previous studies indicated that non-iconic features, such as SVS/EVS images, in flight deck displays effect flight performance, SA and (or) workload such that safety can be improved. However, few research studies have been conducted on the performance, SA and workload effects of non-iconic, conformal features (SVS and EVS images) in HUDs. Furthermore, pilot preferences as to which displays to use during distinct legs of an approach and/or landing situation have not been evaluated. Moreover, few empirical studies have investigated the effects of the combination of SVS and EVS features in HUDs on pilot performance. Thus, there is a need to examine the individual additive effects of SVS and EVS features in a HUD on pilot performance during various legs of flight (especially in landing situations) by making comparisons to HUDs without such features.

Methods for measuring flight performance (e.g., FTEs, path control, or RMSE) and workload (NASA-TLX or SWORD) are well defined and can be applied accurately for assessing display design. However, measures of SA used in previous research, such as SART or SA-SWORD, have limitations because of their subjective nature. Since advanced flight displays, including SVS or EVS features, are expected to promote flight safety by improving pilot SA, careful definition and development of SA measurement approaches is critical to quantifying the SA effects of specific display features. While SART and SA-SWORD have been successfully applied in previous studies to show SA improvements when using advanced HUD features, these techniques may reflect participant bias and are not sufficient for explaining the effects of displays on the three levels of SA, defined by Endsley (1995a).

Related to this, it has been suggested that pilot SA has three types including: spatial awareness, system awareness, and task awareness (Wickens, 2002). Bolton et al. (2007) measured three levels of spatial awareness in SVS display use, but their experiment did not involve dynamic continuous controllable flight situations. Therefore, there are no studies presenting measurement of pilot SA at the three levels defined by Endsley (1995a) across the three types identified by Wickens (2002). With this in mind, it would be worthwhile to develop a method to measure pilot SA across all levels and types. SAGAT may be considered as a potential measure to address this need by posing queries to pilots associated with the various levels and types of SA. Moreover, it was expected that the use of SAGAT would allow for illustration of how non-iconic conformal features in a HUD affect pilot SA in a detailed manner.

Finally, while cognitive task analyses have been conducted to examine pilot behavior, no analyses have been conducted to explain how a SVS or EVS-display, or the combination of thereof, in a HUD affects pilot cognitive behavior, decision making, and action implementation. In other words, even though SVS/EVS technology has been revealed to improve pilot performance, the cognitive reasons for such effects have not been made explicit. There is lack of research including cognitive task analyses of pilot use of SVS/EVS HUDs to explain the effects of each feature on performance during different legs of flight. Such an analysis could provide a basis for optimizing SVS/EVS HUD design to support pilot cognition. The use of a verbal protocol technique in this research was expected to provide platform for a cognitive explanations of SVS/EVS feature effects on pilot performance, as well as a detailed description of pilot cognitive behaviors with HUDs during simulated flight performance.

2.2. Research Objectives

Given the limitations of prior research and the research needs identified above, this study aimed to address the three following objectives:

First, an experiment was conducted to assess the effects of non-iconic, conformal features (SVS/EVS) on pilot flight performance (e.g., flight path control), workload, SA, and pilot subjective preferences when using advanced HUD displays across different legs of an approach under IMC conditions. The study focused on the individual effects of each raster feature (SVS, EVS, and a combination of SVS/EVS) relative to a baseline HUD symbology condition (without any non-iconic terrain imagery), including pathway tunnel features which have been proven to increase flight path control (Alexander et al., 2003; Bailey et al., 2006, 2007; Prinzel et al., 2004b, 2004c; Wickens et al., 2004).

Second, when measuring SA in the experiment, assessment was made in terms of the three levels and three dimensions of SA identified by Endsley (1995a) and Wickens (2002), respectively, by using SAGAT. The SAGAT queries were prepared in order to assess various types of SA in piloting tasks. This elaborate and objective measure of pilot SA was expected to facilitate interpretation of the effects of the advanced HUD features on pilot cognition.

Finally, a high-level cognitive task analysis was developed to describe pilot behavior using SVS/EVS HUD features. A concurrent verbal protocol technique and probing was used as a basis for the task analysis, constituting a semi-structured interview. The aim of this analysis was to relate the new characteristics of the display concepts to pilot internal and external behaviors and, ultimately, flight performance.

3. METHODS

The methodology to address the research challenges encompassed two distinct efforts. The first was an empirical study and the second involved a cognitive task analysis to explain the experimental results in terms of cognitive behaviors.

3.1. Part I: Experimental Study

The objectives of the experiment were fourfold. The first goal was to assess the pure and additive effects of non-iconic, conformal features in HUDs on pilot performance. Typical HUD symbologies used in current commercial aircraft flight decks were integrated with SVS, EVS, and combined SVS and EVS terrain features. Pilot performance, including flight control, SA, mental/physiological workload and subjective preference data were collected. Second, the effects of HUD features on pilot performance were recorded and analyzed for different legs of flight and visibility (IMC) conditions during an approach and landing situation. Since landing situations under low visibility conditions demand a high degree of pilot cognitive activity and the required information for the task changes dynamically during different legs of flight, the effects of advanced HUD features on performance in different legs was also examined. Third, pilot SA during experimental trials was measured by SAGAT and analyzed in terms of the three levels of SA (Endsley, 1995a) and the three dimensions of flight operation (Wickens, 2002). The effects of the advanced features in HUDs were assessed in terms of spatial, system and task awareness. Lastly, videotapes, observations and verbal protocols of subjects during the experiment were used as a basis for developing a cognitive task analysis for explaining pilot internal/external behaviors.

3.1.1. Equipment and Experiment Environment

In general, conducting an experiment with a real aircraft to examine the effects of certain cockpit display concepts on human performance is quite difficult because of the issues of cost and safety. Consequently, most previous studies have been conducted using various kinds of flight simulators or prototypes in a lab environment, by fitting them to the objectives of the study. Lab settings for simulating aircraft cockpit displays range from low-fidelity, including simple functional capabilities, to high-fidelity presenting more realistic flight situations and controls (Aragon & Hearst, 2005). For example, Bolton et al. (2007) used a relatively low-fidelity simulator in their experiment. They provided subjects with short video files presenting terrain texture images using a workstation and asked subjects to estimate values associated with spatial awareness, without supporting any continuous flight control. In contrast, Bailey et al. (2006, 2007) used a high-fidelity flight simulator for their experiments, specifically the IFD (Integration Flight Deck) simulation facility at NASA Langley Research Center (LaRC), which provides pilots with a full-mission simulator.

While most previous empirical studies for confirming the efficiency of pilot use of SVS/EVS concepts were conducted with fixed-position simulators in a lab environment, Prinzel et al. (2002) confirmed the findings from simulator experiments to be comparable to findings from actual tests with real aircraft. This means a well-designed simulator experiment may be a viable process for determining results relevant to real flight situations.

In the present study, videos of expert pilot performance with an advanced HUD, including non-iconic conformal features (SVS, EVS or a combination of SVS and EVS terrain images), were recorded with the IFD simulator located at NASA LaRC. The videos were later presented to subjects as part of a PC-based lab simulation of aircraft landing in the Ergonomics lab at NC State University. That is, a high-fidelity simulator (IFD) was used to generate the videos of HUD features in flight and a low-fidelity simulator was used to assess the effects of the displays on pilot performance. The reason the videos were captured at NASA is that, the advanced HUDs to be investigated are still under development and can only be implemented in the IFD simulator. The reason for showing the videos as experimental stimuli in a low-fidelity lab simulator was that, access to the IFD simulator at NASA is extremely limited and tremendously costly (approximately $125k for a 24 pilot experiment).

The IFD simulation facility (see Figure 3.1) at NASA LaRC simulates the Boeing B-757-200 aircraft. The cab is populated with flight instrumentation and pilot controls, including the overhead subsystem panels, to replicate the B-757 aircraft cockpit. A collimated out-of-cockpit scene is produced by an Evans and Sutherland ESIG 4530 graphics system providing approximately 200 degrees horizontal by 40 degrees vertical

FOV at 26 pixels per degree. The integrated HUD (see Figure 3.1) subtends approximately 32° horizontal by 24° vertical FOV at a typical pilot viewing distance (20”) (Bailey et al, 2006).

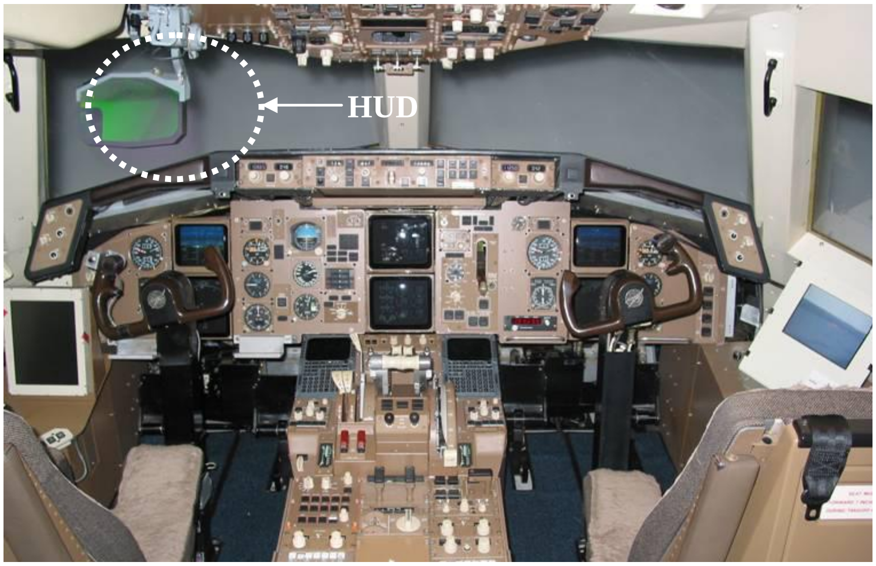


Figure 3.1. Integration Flight Deck (IFD) Simulation Facility at NASA Langly (Bailey et al. 2006).

An ex-Air Force (C-130) check pilot, who has some experience with advanced HUDs, operated the IFD simulator according to scenarios defined for an approach and landing at a major international airport (the task scenario is described in the next section), using different HUD configurations in each trial. The HUD configurations included Baseline (without any terrain features from SVS or EVS), SVS, EVS, and the combination of SVS and EVS. All configurations included tunnel features. The display content of the HUD during the trials was recorded continuously. After completing the flight trials, the recorded video files were collected and prepared for the follow-on lab tests of pilot.

The HUD videos, captured in the IFD simulator, were played-back on the PC-based (see Figure 3.2). The simulator setup consisted of a Dell OptiPlex 755 PC workstation with 4GB of RAM, 256MB of video memory and 2.4 GHz Quad processors, and flight deck controls (a yoke and a throttle quadrant), originally used for a flight simulation game (X-plane). Two 22 inch LCD monitors were also used to present the movie clips to the test pilots and an the experimenter, who manipulated the trial conditions and acted as a first officer (FO) during the flight.

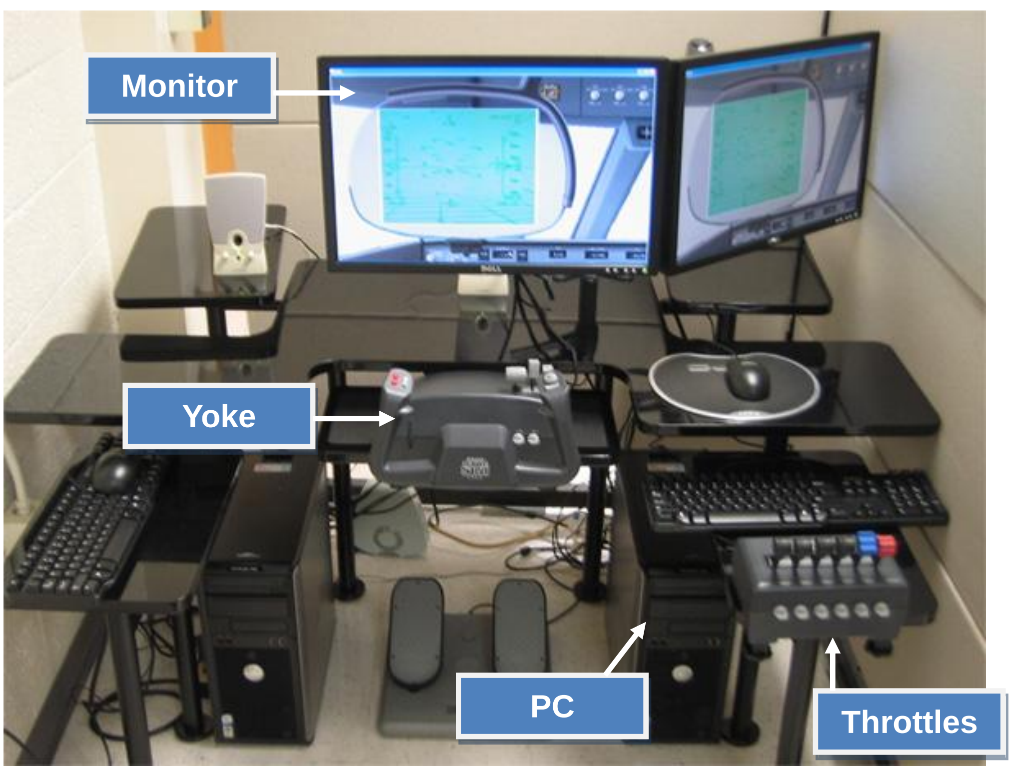


Figure 3.2. Simulator setup for experiment.

A video camcorder was used to videotape subject performance during the last test trial, the think aloud session. These videotapes were used as a basis for the cognitive task analysis and to identify specific performance strategies used by the test pilots. During the experiment, the overhead lights in the lab were covered to prevent glare on the computer screens.

3.1.2. HUD features investigated in study

Since video stimuli were recorded using the IFD simulator at NASA LaRC, the HUD symbology and non-iconic terrain features (SVS, EVS and SVS/EVS) presented in this experiment followed the format developed by NASA. This format has also been used in several prior empirical studies (Kaber et al. 2007a, 2007b, 2008). Since previous research indicated that presenting a pathway tunnel improved pilot performance, especially on flight path control (Alexander et al, 2003; Bailey et al., 2006; Prinzel et al., 2004b; Wickens et al., 2004), it was necessary to examine the pure effects of the SVS, EVS and the combination of features when all display conditions included pathway tunnel images.

Figure 3.3 shows the basic (or iconic) HUD symbologies without any non-iconic terrain features. This configuration was used as the "Baseline" condition for the study. The symbologies are similar to those on a general PFD in the commercial cockpit and included vertical and lateral path indicators. Each indicator is comprised of a scale, a path marker, and ILS deviation marker for glideslope (vertical) and localizer (lateral). The "dog bone" shaped indicators represent deviations from the designated flight path in the flight management system and the "bullet" shaped indicators represent deviations from the Ground-based ILS approach. The tunnel features consist of a series of box images defining the vertical and horizontal extent of the desired path. The box shaped tunnel features are depicted with dynamic "Crow's feet." The use of this feature was based on Prinzel et al. (2004b; 2004c) study demonstrating crow's feet to be most appropriate for use in a SVS HUD. Each box presented on the HUD is 600 feet wide by 350 feet tall and represents -1 to +1 dot of path deviation. A flight path marker (FPM) group consists of the flight path marker, airspeed error "worm," and acceleration/deceleration cue (see Figure 3.4).

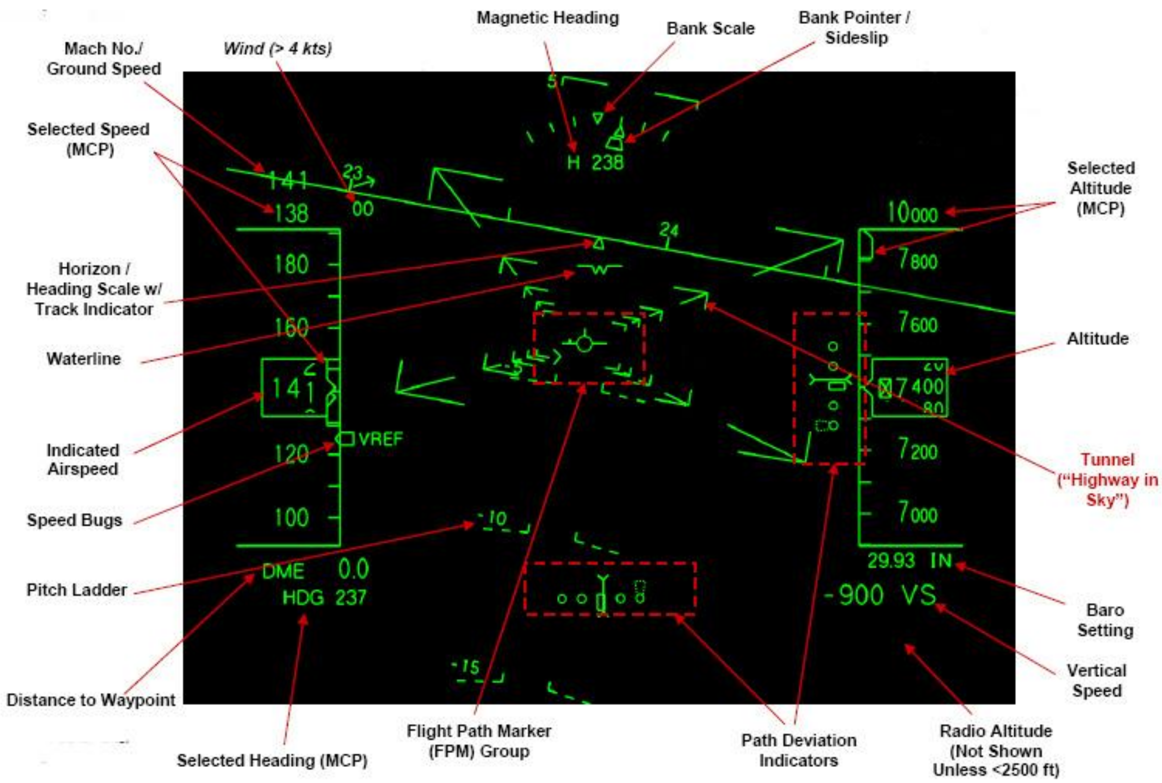


Figure 3.3. Baseline HUD symbologies

When the aircraft is less than 500 feet AGL, the tunnel features are programmed to disappear and a runway outline is shown in the display with a glideslope reference line set at 3.1 degrees. Figure 3.5 shows the runway outline and glideslope reference line in the HUD.

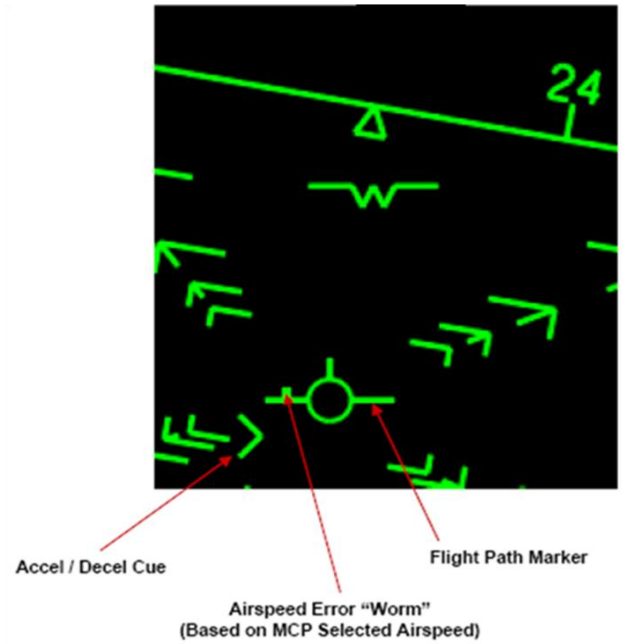


Figure 3.4. Flight path marker group.

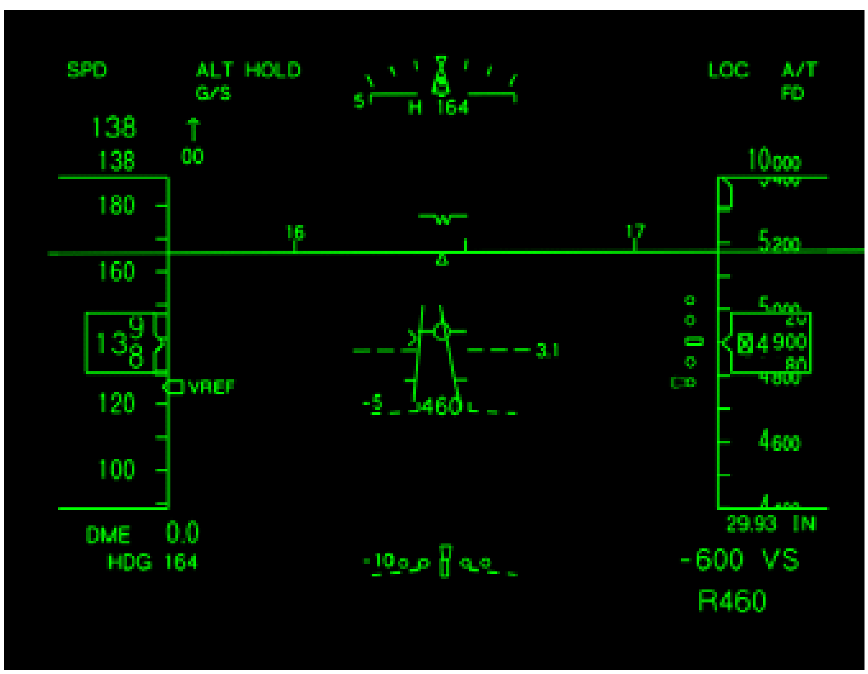


Figure 3.5. HUD with runway outline and glideslope reference line.

The SVS-HUD configuration presents terrain features using a wireframe grid. The use of the wireframe feature is based on findings from Snow and Reising's study (1999), which revealed the grid model to be most appropriate for depicting terrain in a HUD. The wireframe features in the HUD in the present study were set to represent terrain using a 500 meter line separation with a 1 pixel line width. Figure 3.6 shows a captured SVS HUD image.

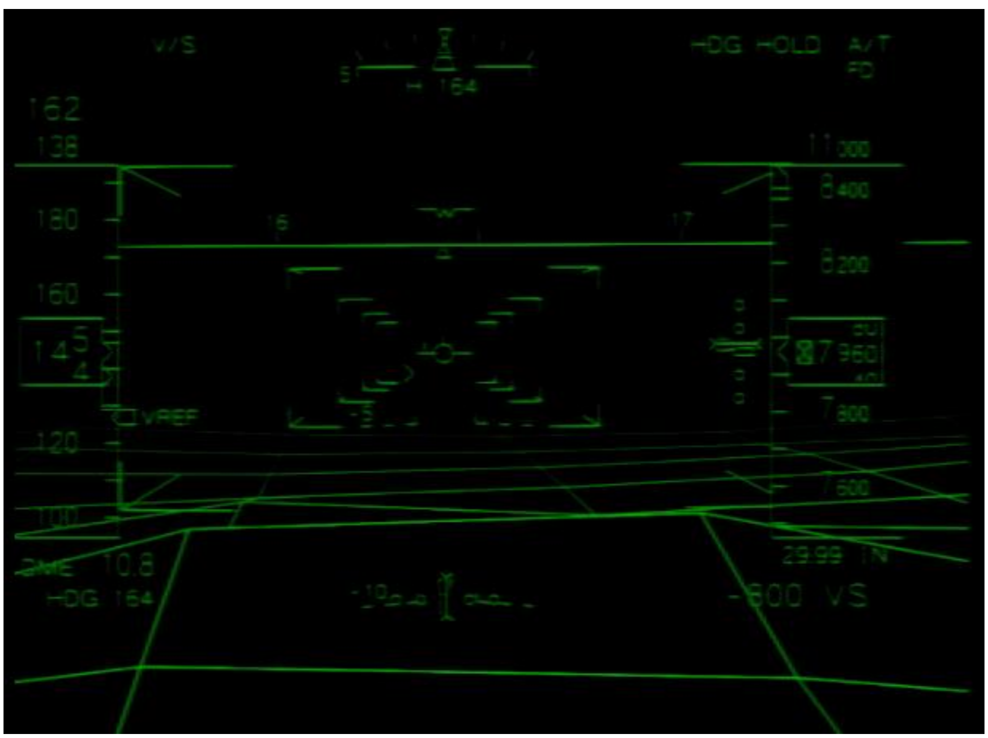


Figure 3.6. HUD with SVS terrain features.

The EVS-HUD presents an actual out-of-cockpit view using a sensor-based forward looking Infrared camera. Figure 3.7 shows the EVS-HUD and Figure 3.8 shows the combination of SVS and EVS (Combo) features in the HUD, recorded using the IFD simulator.

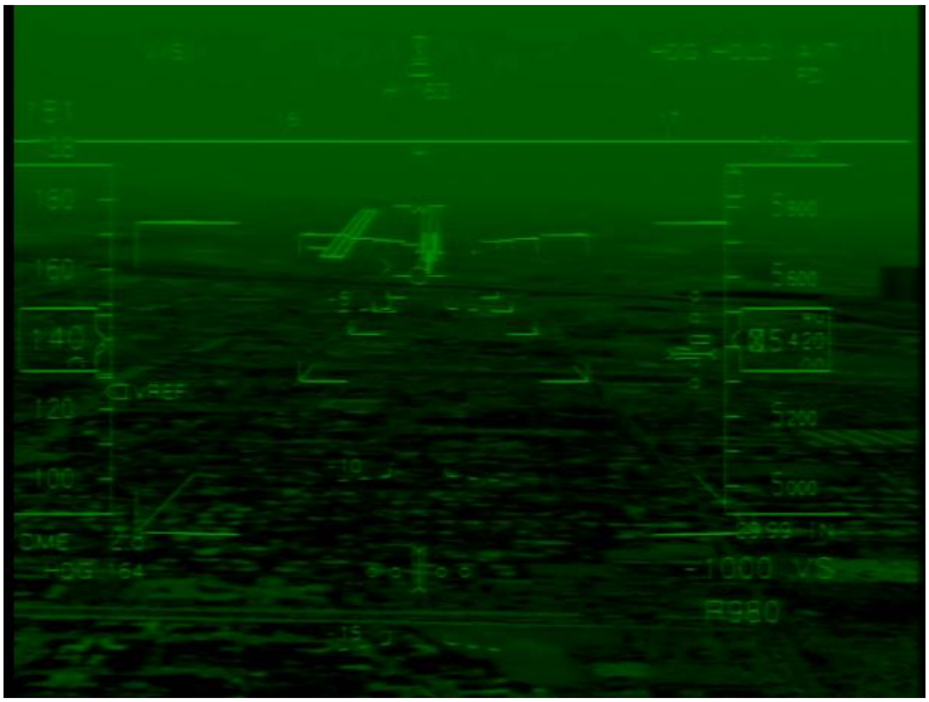


Figure 3.7. HUD with EVS features.

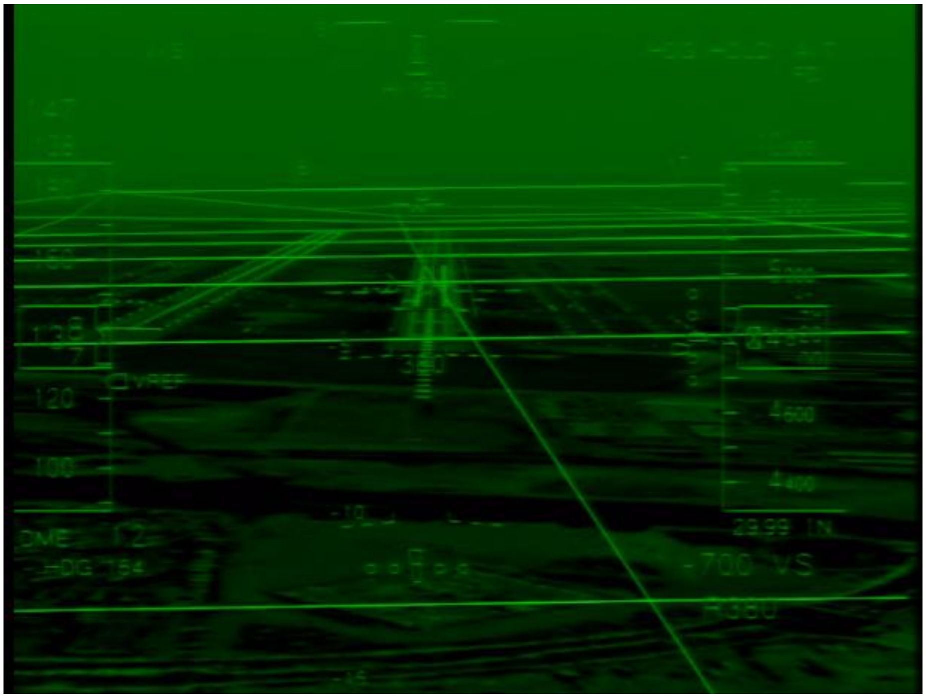


Figure 3.8. HUD with combination of SVS and EVS features.

3.1.3. Flight scenario

A realistic flight scenario was developed in which HUD images were presented in order to assess the impact of non-iconic conformal terrain features (SVS and/or EVS) on pilot performance. The scenario was used to capture video of expert pilot behavior in the IFD simulator, and was also used to guide test subjects in following a flight procedure using the movie clips in the Ergonomics Lab. When developing the flight scenario for SVS/EVS testing, several key elements were considered for inclusion. First, since the primary objective of applying EVS and/or SVS is to improve flight safety by increasing pilot terrain awareness, any test scenario should include terrain elements that are perceived as threatening to pilots. This challenge was addressed in three ways: (a) the flight task of approach and landing, an aircraft at an airport was selected. Many empirical studies have been conducted to assess the effects of display technology on these legs of flight (Prinzel et al., 2002, 2003; Schnell et al., 2004) because most critical aircraft crashes occur during approach and landing; (b) an airport and runway was selected, which is surrounded by challenging terrain. In terms of the HUD content and simulator setup used in this study, NASA has previously conducted real and simulator-based studies of the effects of advanced cockpit displays on pilot landing performance at Reno/Tahoe (NV), San Diego (SD), and Eagle/Vail (CD) airports. From these airports, Runway 16 Right (RWY 16R) at KRNO (Reno/Tahoe International Airport, NV) was selected to assess pilot behavior, as it is surrounded by challenging terrain. KRNO is located in a valley with mountains reaching approximately 4,000 to 6,000 feet above the airport, in close proximity to the approach path and airport. Figure 3.9 shows the ILS approach plate for 16R at KRNO; and (c) Finally, low visibility conditions were also applied. Two types of flight scenarios for presenting low visibility conditions were created. One scenario was an IMC-day situation (1300 local time, 2300 Zulu); that is, pilot visibility through the cockpit window was restricted by clouds or precipitation even though there was daylight. Another scenario presented an IMC-night situation (2200 local time, 0500 Zulu) in which the pilot was not able to see the terrain or runway because the plane was approaching the airport in darkness and clouds. However, in order to allow the flight to land on the runway, the runway must be in sight to the pilot through the out-of-cockpit view by the time the flight reaches the decision height.

The second key element in developing the flight scenario was that distinct legs of flight. For example, Byrne et al. (2004) divided a flight scenario for approach and landing situations into four legs. Leg 1 was from the beginning of the scenario to the first waypoint in the approach; Leg 2 was from the first waypoint through the last waypoint; Leg 3 was the last waypoint to the landing decision altitude; and Leg 4 was from the decision altitude through the end of the scenario (landing). Using a similar approach, the flight scenario in this study included four legs during the KRNO ILS RWY 16R approach and landing, as follows (see Figure 3.9 for general concept of the approach scenario and Figure 3.10 for actual runway approach plate and waypoints referred to below):

- Leg 1: Initial Approach Fix (IAP) at PYRAM (waypoint at 23.0 DME (NM) from the runway) to glideslope (G/S) intercept (at approximately 13.5 DME). This leg is flown at a constant altitude (level flight) and heading until intercepting the glideslope. In general, the pilot positions the aircraft on the Instrument Landing System (ILS) approach. In this leg, pilots were required to slow the aircraft (from 210 kts to 138 kts) by adjusting the throttles, call for setting the FCP speed, and call for extending the flaps and landing gear. The pilot must also control the aircraft to maintain the localizer track, air speed, heading, and altitude as well as complete the before landing checklist. All of these activities were reflected in the video(s) that the test pilots watched and followed.
- Leg 2: G/S intercept (approx 13.5 NM DME) to DICEY waypoint (final approach fix; approximately 5.5 DME). After the flight intercepts the glideslope, the pilot is

required to initiate a descent to maintain the glideslope on the ILS final approach from 8500 feet to 6400 feet. At the end of this period, the aircraft should be 900 feet above the normal decision height (5514 feet) for the approach and approximately 2000 feet above the runway. Thus, the pilot was required to monitor path deviation indicators (localizer and glideslope), heading, tunnel, altitude, and airspeed.

- Leg 3: FAF (Final approach fix) (DICEY) to decision height. From a piloting perspective, there was no difference in terms of how the approach had to be flown before and after DICEY. In this leg, pilots were able to see the runway through the out-of-cockpit view. They continued to maintain the localizer course and glideslope at approach speed (138 kts). The pilots verified that the aircraft was in final landing configuration and the before landing checklist was completed. At one thousand feet above the decision height, the pilot not flying (first officer) called out to the pilot flying that the DH was approaching. This marked a transition for the pilot to begin looking for the runway and to mentally prepare for either a visual landing or missed approach. The aircraft reached decision height for the normal (ILS RWY 16R) approach (1100 above field level (AGL)) at about 3.3 NM from the runway.

- Leg 4: Decision height to end of scenario (landing on the runway). In this leg, a pilot was able to see a 3.1 degree line in the HUD after descending below 500 AGL instead of the tunnel features. The pilot maintained the path of the aircraft and landed on the runway.

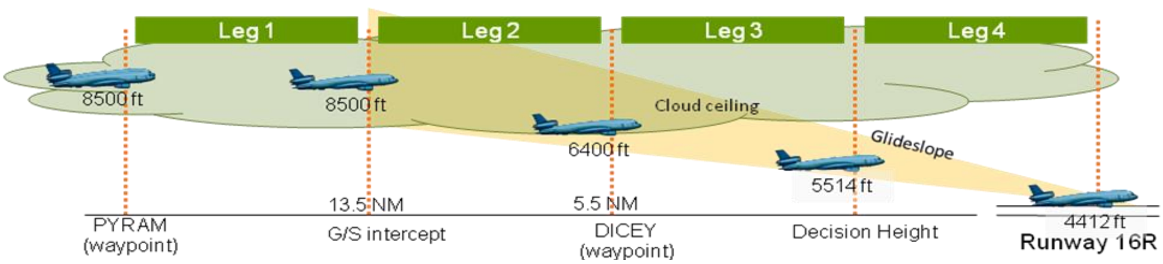
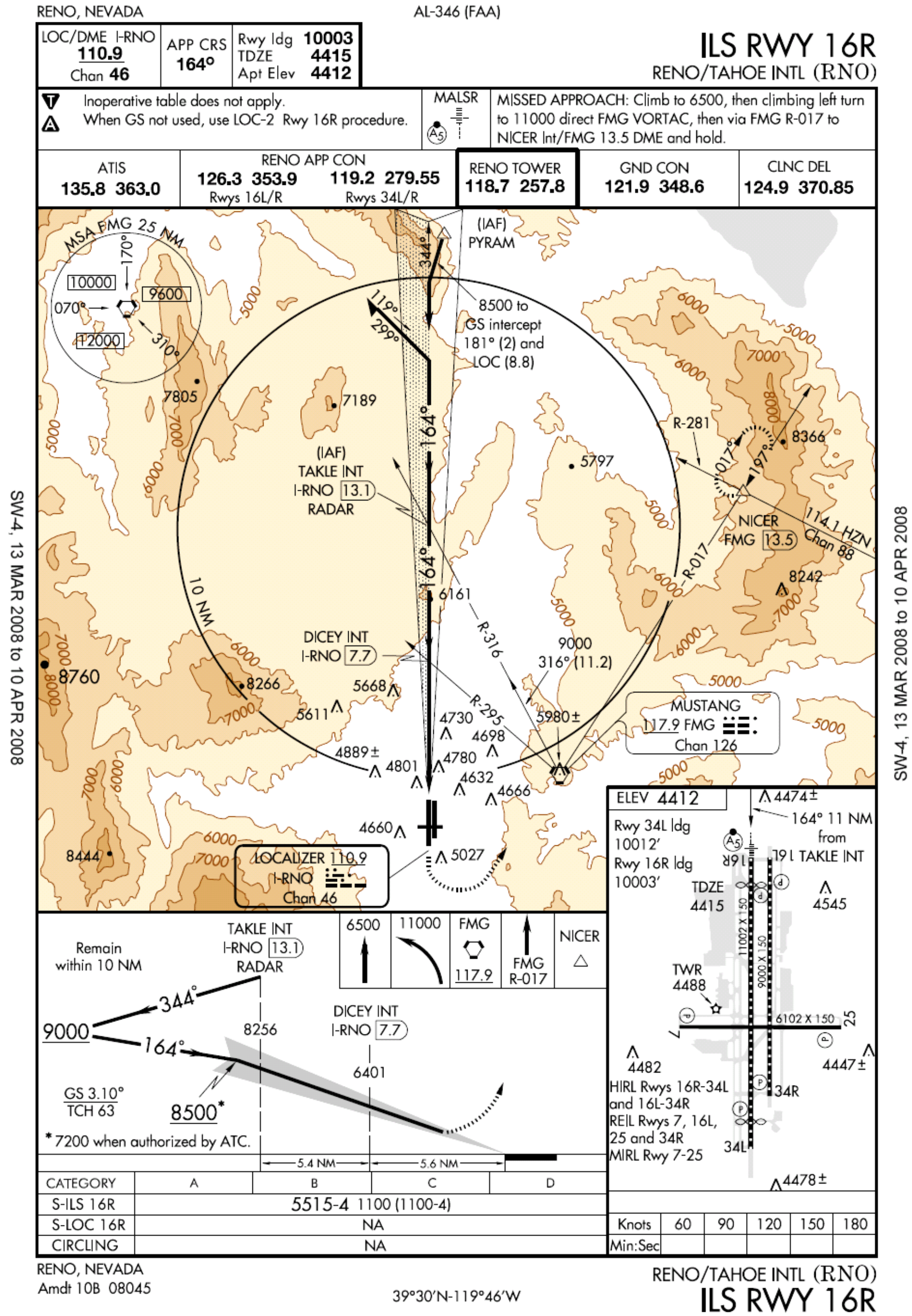


Figure 3.9. Concept of the approach scenario.



CATEGORY	A	B	C	D
S-ILS 16R	5515-4 1100 (1100-4)			
S-LOC 16R	NA			
CIRCLING	NA			

Figure 3.10. ILS approach plate for 16R at KRNO.

The reason for dividing the flight scenario into distinct segments was to assess the effects of each HUD feature on performance in each leg. In other words, pilot performance data, including path control, SA and physiological workload was collected and analyzed for each segment. The various features of the HUD were expected to have varying utility for each segment. Table 3.1 shows a summary of the characteristics of each leg of flight for the scenario.

Table 3.1. Characteristics of each leg of flight.

	Leg 1	**Leg 2**	**Leg 3**	**Leg 4**
Position	PYRAM to G/S intercept	G/S intercept to DICEY	DICEY to Decision Height (5514 ft MSL)	Decision Height to Runway
Approximate DME to runway	23.0 → 13.5 DME	13.5 → 5.5 DME	5.5 → 3.3 DME	3.3 → 0 DME
Flight Characteristics	IAF Level flight	Beginning descent	FAF	Landing
ATC Clearance	Initial clearance Slow to approach speed	Contact tower Cleared to land	N/A	N/A
Required Control	Altimeter setting Slow to approach speed (138kts) Speed bug setting	Descending Contact tower Confirm landing clearance	Descending Landing decision making	Landing
	Flaps and gear extending Complete landing checklist			
Visibility of runway	Invisible	Invisible	Visible	Visible

The third key element of the flight scenario was to provide subjects with necessary information on the aircraft status that was not available from the HUD, since only the HUD was displayed to the test subjects on the PC monitors. Thus, the scenario included an air traffic control information system (ATIS) broadcast to provide approach information at the beginning of the scenario, ATC directions to approach speed and contact tower, and landing clearances. This information was provided to subjects verbally using pre-recorded audio files during the experimental trials, according to a predefined procedure.

Lastly, the task scenarios represented a nominal landing situation. That is, the aircraft was cleared first for approach and then for landing without incident, and there were no flight tasks performed under abnormal situations, such as runway incursions or in-flight emergencies.

With these considerations in mind, the flight scenario was created and used for stimuli video recordings and test trials with the lab simulator (see Appendix A. TASK FLIGHT SCENARIO). The scenario was based on a flight scenario used for another study (Kaber et al., 2008) to assess the effects of HUD configurations, pilot expertise and flight workload manipulations on perceived clutter, perceived workload and flight performance using the IFD simulator at NASA. In addition to this, since the HUD features developed and tested by NASA have focused on implementation in commercial aircraft (e.g., B757), the flight scenario in the present study involved a commercial aircraft piloting situation, including a captain and a first officer.

3.1.4. Video stimuli preparation

Since the videos of HUD, recorded using the IFD simulator at NASA LaRC, did not include an out-of-cockpit view but only the HUD imagery on a black background, it was necessary to edit the files for use in the lab experiment. Two factors were considered in preparing the HUD videos for the experiment. One was to implement two distinctive visibility conditions (IMC-day vs. IMC-night) and another was to present the out-of-cockpit view as the flight neared to ground. These issues were addressed by synthesizing additional videos for the out-of-cockpit view with the videos of the HUD.

In order to secure videos for the out-of-cockpit view, simulated flights for approach and landing to KRNO RWY 16R in both IMC-day and IMC-night conditions were recorded using the X-plane system. The X-plane system is flight simulator software, which includes realistic three dimensional rendering of terrain and runway images of airports. The HUD videos recorded using the IFD simulator and the out-of-cockpit view videos recorded using the X-plane simulator were synthesized and rendered using a commercial video editing tool (Adobe Premiere).

According to the task flight scenario, before the flight reached 1500 ft AGL, visibility conditions were poor and pilots were not able to see the runway due to clouds (Ceiling 1500) in both the IMC-day and IMC-night conditions. Above the ceiling altitude, gray and black out-of-cockpit views were synthesized with HUD videos for the IMC-day and IMC-night conditions, respectively. A transparency effect was applied to the HUD videos. When the flight was at approximately 1500 ft AGL, the background was smoothly transitioned from the gray or black to a clear out-of-cockpit view from videos captured using X-plane. This simulated the aircraft descending through the clouds to below the ceiling. The “Cross dissolve transition effect” in the Adobe Premiere was applied for this transition. Finally, when the flight was below 1500 ft AGL, clear out-of-cockpit views were provided by synthesizing HUD content with the X-Plane video. In this step, in order to align the HUD content (e.g., runway outline or terrain features) with the out-of-cockpit view images as closely as possible, the “crop” function of Adobe Premiere was manually applied to several key frames in the video. This image synthesis was critical for examining the effect of non-iconic (terrain) features on pilot performance because the features were designed to present the pilot with conformal imagery on two visual domains using the HUD (Wickens, 1994). Figure 3.11 and 3.12 show examples of synthesized video images for IMC-day and IMC-night conditions, respectively.

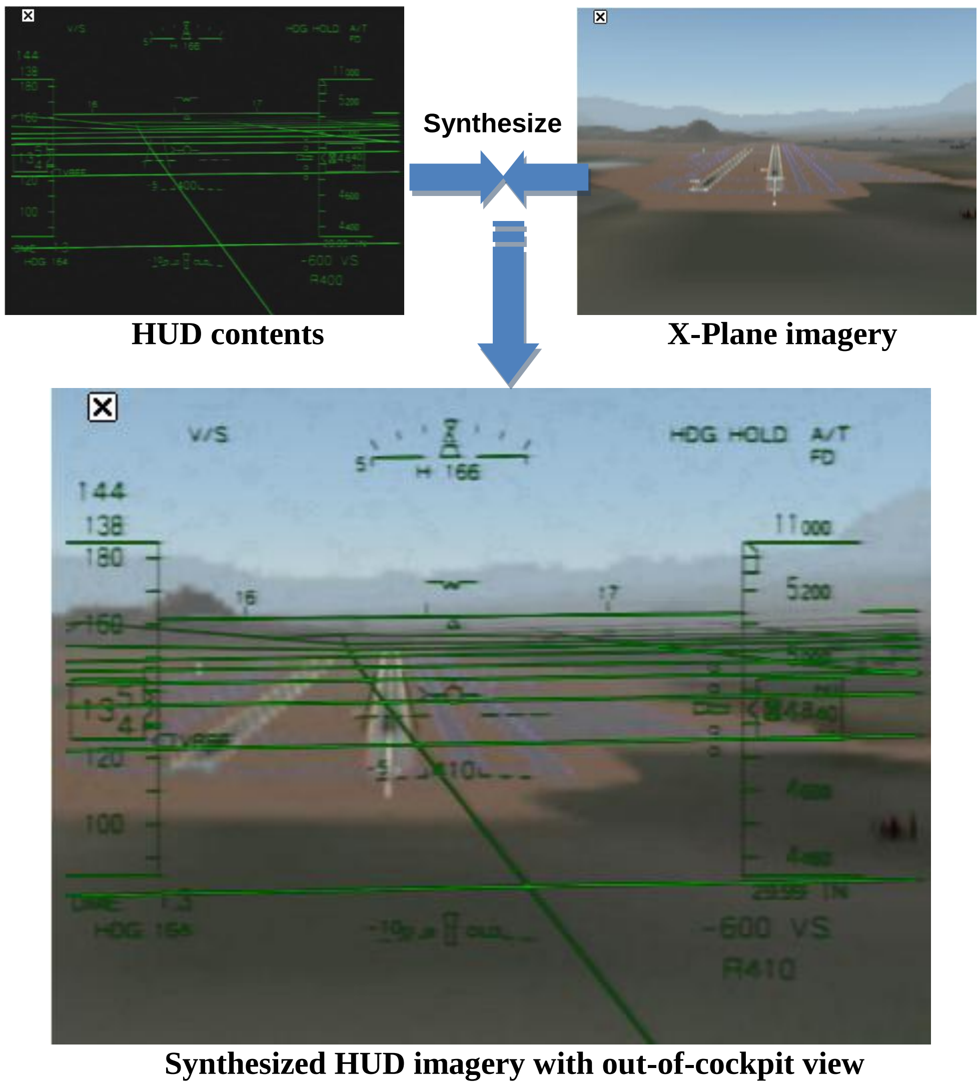


Figure 3.11. Example of imagery synthesized for the IMC-day condition.

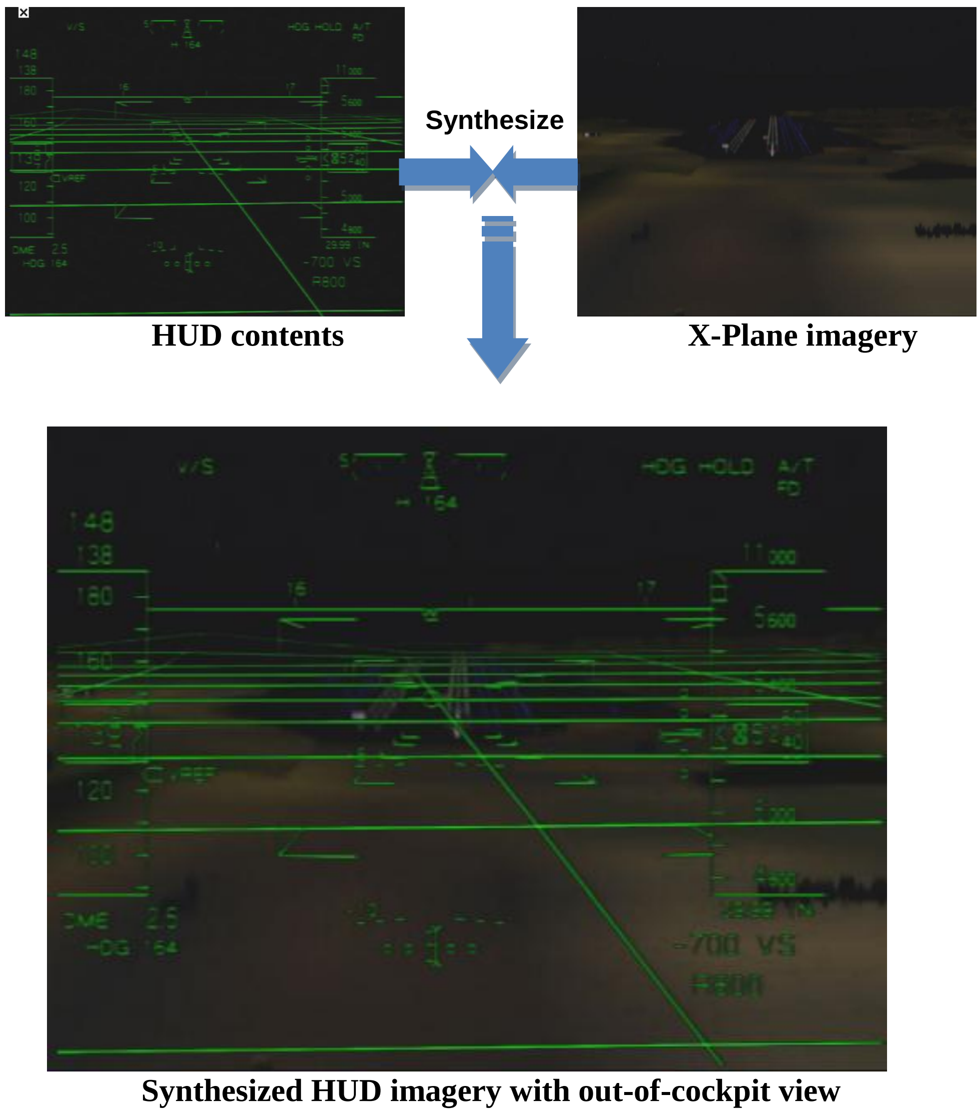


Figure 3.12. Example of imagery synthesized for the IMC-night condition.

In addition to this image synthesis, several audio files were integrated with the video stimuli in order to provide for a realistic flight simulation. ATC broadcasts were recorded separately and included with all video files according to the flight scenario. They included: the 1st ATC clearance at the beginning of the simulated flight; the 2nd ATC clearance directing a slow down to approach speed (at 19.0 DME); the 3rd ATC clearance directing a radio frequency change to contact the tower; and the 4th clearance from the tower for landing (see Appendix A: Task Flight Scenario). In addition to these audio cues for ATC broadcasts, other audio files for GPWS (ground proximity warning system) (or TAWS (terrain awareness warning system)) also were included with the video files. The GPWS were active during video recording trials using the IFD simulator and identically implemented in the lab simulator. The warning voices were presented when the flight was at 1000, 500, 400, 300, 200, 100, 50, 40, 30, 20, and 10 ft AGL and consistent throughout all trials.

Consequently, ten video files were prepared for the experiment, including two for practice trials (Baseline in IMC-night and Combo (SVS/EVS) in IMC-day) and eight for test trials (Baseline, SVS, EVS, and Combo HUD by two IMC conditions).

3.1.5. Experiment application

A java application was developed to present the video stimuli and facilitate the lab experiment. The application was launched using the PC workstation and shown on the LCD monitor as part of the lab simulator. The interface of the application consisted of a panel for playing the video stimuli, static background, and a menu for manipulating the experiment conditions (see Figure 3.13).

Since pre-recorded flight videos were presented in the application, some of the simulator flight controls were limited in use for manipulating the flight displays. Control actions of the integrated controls did not affect the status of the flight presented in the video (e.g., direction, airspeed, and altitude). However, during experimental trials, participants were required to control the yoke according to the current flight situation displayed on the monitor. That is, participant pilots were asked to use the yoke to move a cursor overlaid on the video imagery in order to track a FPM in the video. Figure 3.14 shows the super-imposed cursor for tracing the FPM. This task was implemented to immerse participants in the experiment and to simulate virtual flight path control. The tracking task represented a normal part of pilot performance in hand flying an aircraft. Since every display condition included tunnel features (i.e., HITS) or terrain features (EVS and/or SVS images), the approach of using the simulated yoke control was based on an assumption that significant path control deviations between the overlaid cursor and FPM in the pre-recorded video might reflect degradations in pilot attention and that flight performance in directing the FPM to the center of tunnel could be measured indirectly. In this regard, the Java application was programmed to record participant yoke control actions during the experimental trials. The recording frequency was set to 12 Hz. The data were analyzed to determine flight control performance in terms of path deviations (e.g., RMSE).

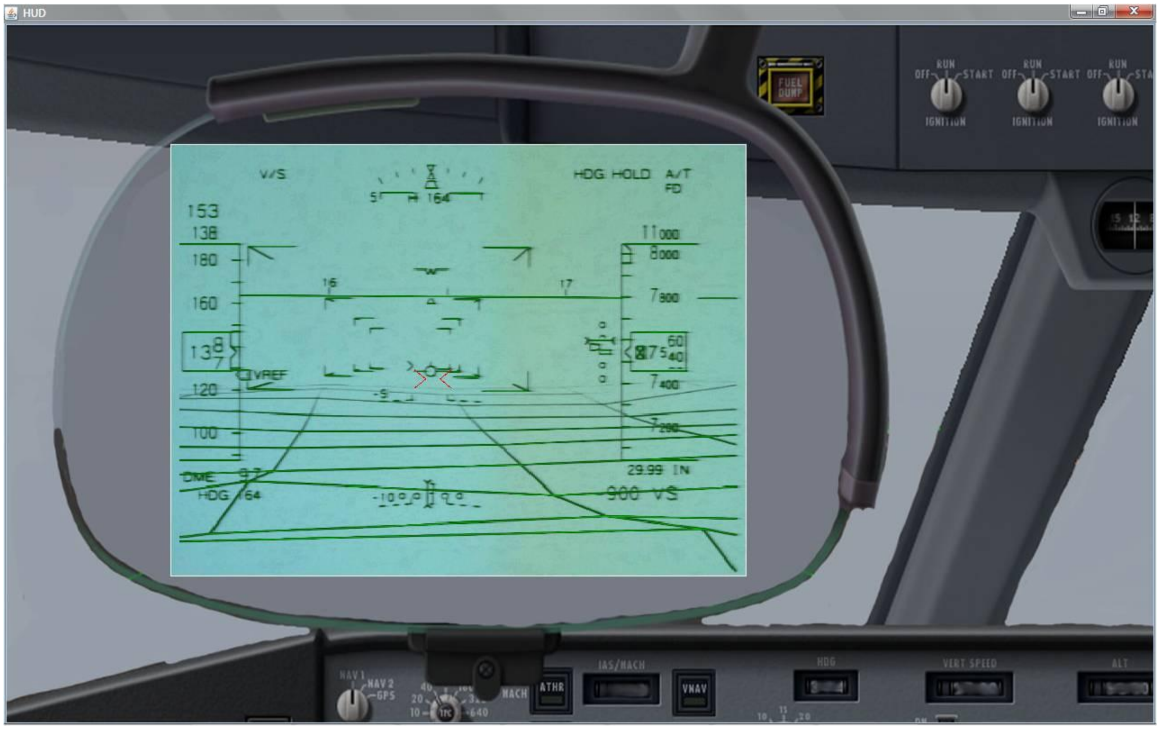


Figure 3.13. A screenshot of the Java application playing a video of the HUD.

Figure 3.14. Overlaid cursor for tracking FPM

In addition to this, participants were asked to control the throttles according to the airspeed displayed in the flight videos. Even though the controls did not affect actual airspeed in the video, positions of the throttles were programmed to provide pilots with different levels/pitches of engine sound in order to further immerse them in the simulated flight. However, the deviations among power output from throttles control and displayed speed were not captured and analyzed.

There were other functions in the Java application that also provided sound feedback, including flaps and gear extensions, which were manipulated by the FO (First Officer) at the pilots' direction. All task behaviors including callouts and decisions were recorded by the Java application with time information, as a basis for analysis.

3.1.6. Participants

Eight line pilots were recruited to participate in the lab experiment. All pilots were required to have previous experience in flying commercial aircraft (e.g., business/charter/ corporate jets) with "glass" cockpit displays. Since the SVS and EVS technologies for HUDs are still under development, experience with the use of these features was not expected. Pilots were compensated on an hourly basis for their participation in the study ($25/hour).

As part of the experimental procedure, pilots were asked to complete a demographic survey in order to establish the general characteristics of the sample population, including flight hours and experience with a HUD. All pilots were males and the average age of the pilots was 58.6 years with a standard deviation of 14.4 years. All pilots had "glass cockpit" cockpit experience with an average of 3646.3 hours (with a standard deviation of 3411.7 hours). Table 3.2 shows the mean and standard deviation of the pilot time data collected through the survey.

Table 3.2. Pilot flight time data (unit: hour).

	Mean	Standard deviation
Total instrument time	3781.3	4064.7
Total night time	3660.0	3471.8
Total flight time	11043.8	7893.1
Total time last 12 month	230.6	204.8

Among the pilots, three had experience in the use of a HUD in either actual flight (mean of 2350 hours) or in a simulator (mean of 86.7 hours). Two pilots had experience with SVS systems (mean of 4 hours) and one pilot had EVS experience (6 hours) in simulator flight.

3.1.7. Experiment design

Each pilot completed nine trials with eight of these following a completely within-subjects design and one additional trial for collecting subject verbal protocols. The verbal protocols were used in the cognitive task analysis to explain pilot internal behavior and strategy. The cognitive task analysis (CTA) technique was not used during the primary eight test trails so as not to disrupt normal pilot performance with the various display conditions. The experiment design consisted of four HUD feature configurations (baseline with tunnel, SVS, EVS, and a combination of SVS and EVS (hereafter referred to as the “Combo”)) by two visibility conditions (IMC-day and IMC-night). In order to investigate a potential additive effect of HUD features on pilot performance, structured orders of test display conditions were used rather than a randomized order of presentation. It should be noted here that the order was determined based on perceived levels of HUD clutter observed in the Kaber et al. study (2007b). Each pilot was presented with display conditions in increasing order of perceived clutter (baseline, SVS, EVS, and then combination) or a decreasing order for their first four test trials (Trial 1 – 4) or their last four trials (Trial 5 – 8). Half the participants (four pilots) began with the order of increasing clutter and the other half began with the order of decreasing clutter. The presentation order of the two visibility conditions was balanced across pilots and display configurations.

After a pilot completed the eight test trials, he was asked to complete the additional trial for verbal protocol analysis. Each pilot was randomly assigned to each of the eight HUD feature conditions (four display configurations by two IMC conditions) for the protocol analysis trial. During these trials, pilots were asked to verbalize for the experimenter any procedures they were performing for approach and landing. A video of the 9th trial was recorded for each pilot. This served as a basis for the cognitive task analysis. The experiment included a total of 72 trials across participants (8 pilots by 9 trials (8+1)).

3.1.8. Independent variables (IVs)

The main IV in this study was the feature configuration of the HUD. The four types of display configurations (Baseline, SVS, EVS, and Combo) identified above were presented to pilots. As mentioned previously, all display conditions included a pathway tunnel image (HITS).

The second IV was the visibility condition, IMC-day versus IMC-night. Previously, Snow and Reising (1999) investigated the effects of visibility conditions on pilot SA. To extend this research, the present study investigated the effects of display configurations in HUDs under different out-of-cockpit conditions on pilot performance. These weather and time of day conditions were manipulated through the flight scenario and implemented by the video synthesis approach.

As mentioned above, each of the eight experimental conditions (four HUD configurations by two visibility conditions) involved four distinct legs of flight (Legs), during the approach and landing scenario. Goodman, Hooey, Foyle and Wilson (2003) divided approach flight into four segments and they demonstrated that pilot attention to various cockpit displays differed according to the leg of the approach. Goodman et al. (2003) found, in general, pilot attention to an SVS display was predominant during break-out and runway acquisition in the third leg (among 4 legs). This implies that each specific leg of flight may dictate different information requirements behaviors by a pilot, leading to different cognitive activities. With this in mind, how each leg of flight affects pilot performance was evaluated in this study. It should be noted here that pilots were not able to manipulate the display configuration in each trial because pre-recorded videos of the HUDs were presented in the experiment. In a real aircraft or high-fidelity flight simulator (e.g., the IFD at NASA LaRC), it may be possible for pilots to change display configurations in order to acquire required information by adding or subtracting individual features based on a flight situation (e.g., phase of flight and visibility).

Figures 15 through 22 show sample captured images for each display configuration and IMC condition across the four legs of flight, from prepared videos presented to the pilots. It should be noted here that images of “moisture” clouds could be seen in Legs 1, 2 or 3 when the HUD included EVS features (see Figures 3.17, 18, 21 and 22). The features represent thermal returns of “moisture” depicted by the EVS since the sensor used in the IFD simulator could not penetrate heavy precipitation and certain fog types (Prinzel & Kramer, 2006).

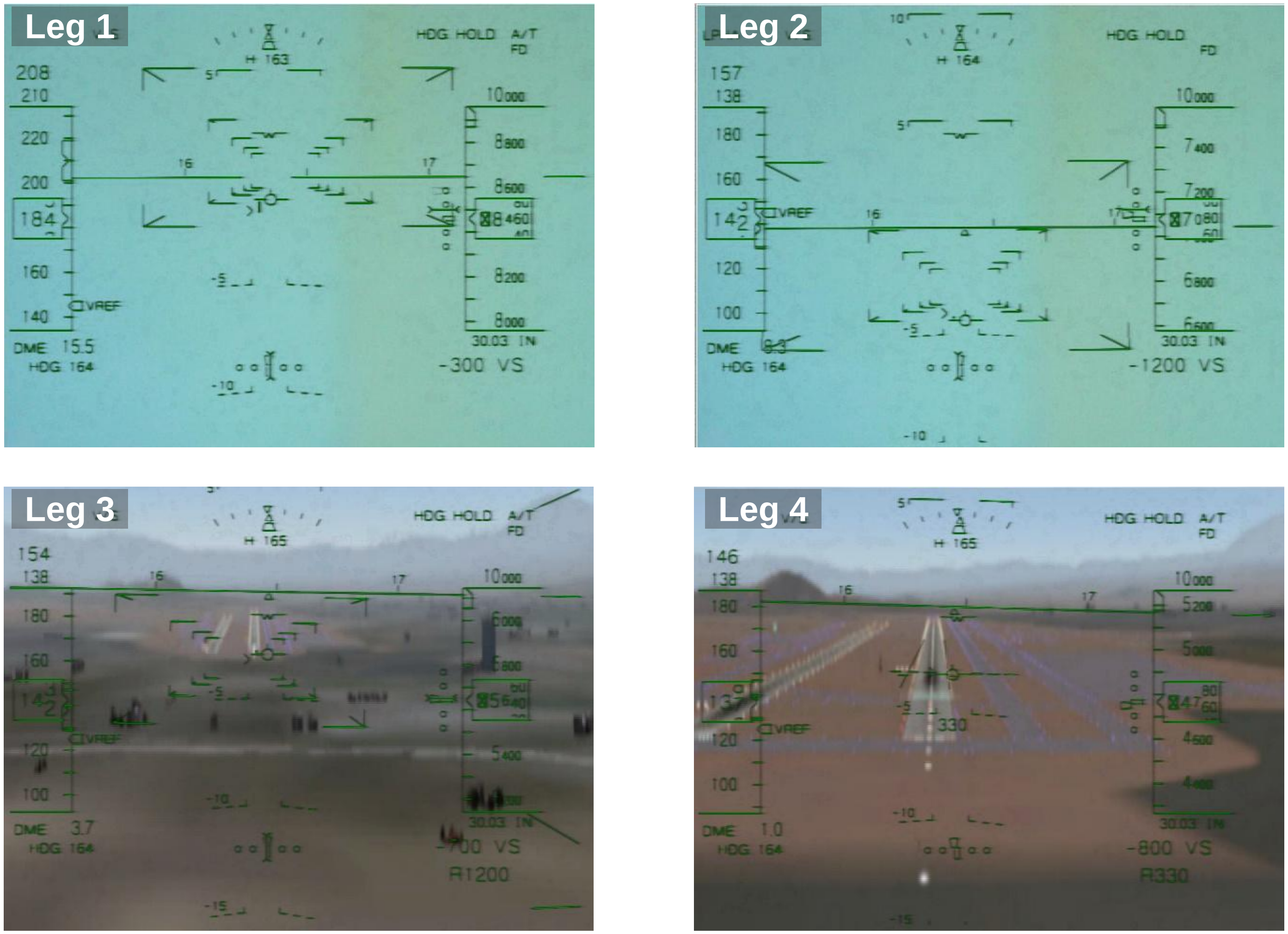


Figure 3.15. Baseline HUD for IMC-day condition.

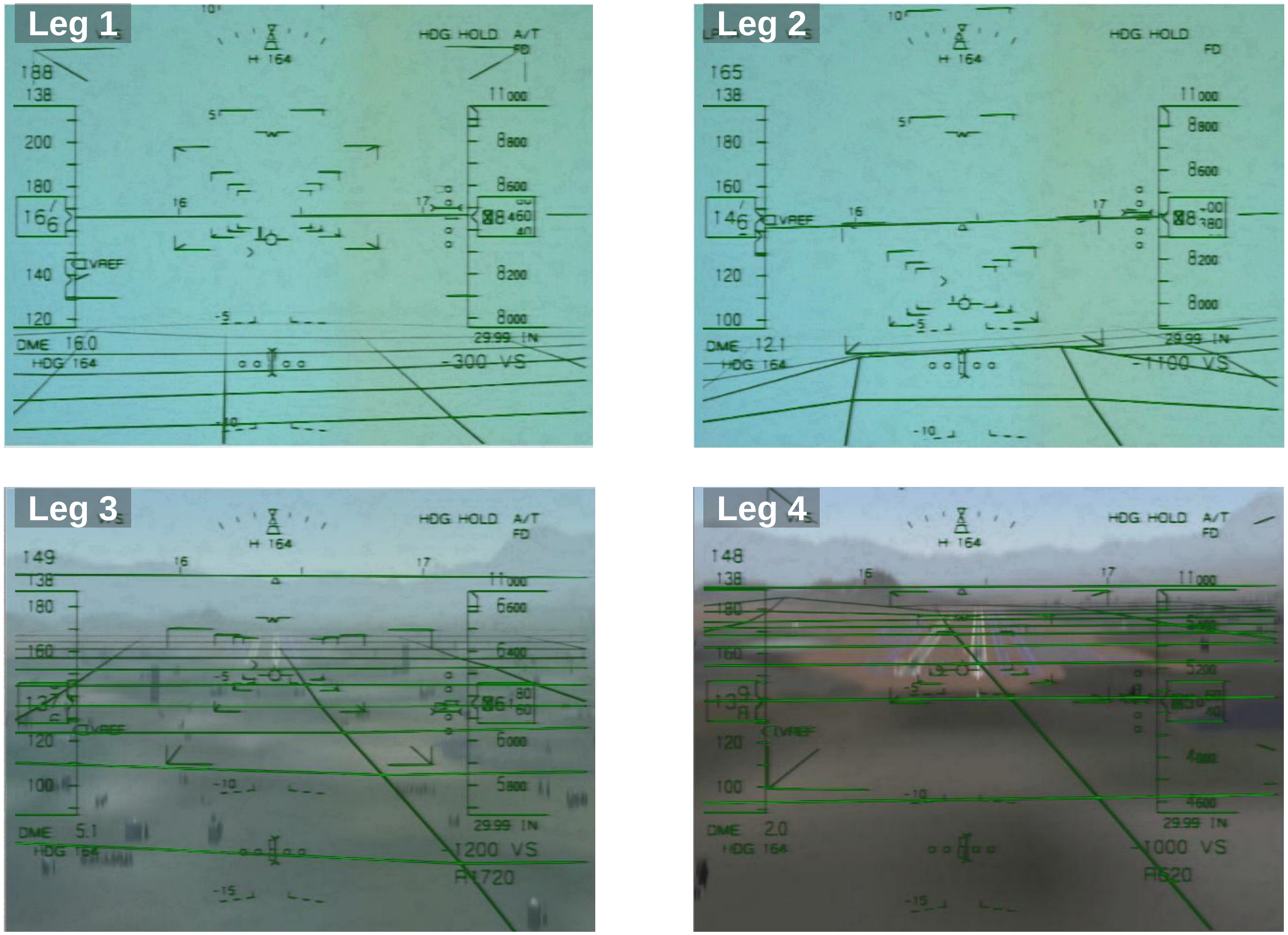


Figure 3.16. SVS HUD for IMC-day condition.

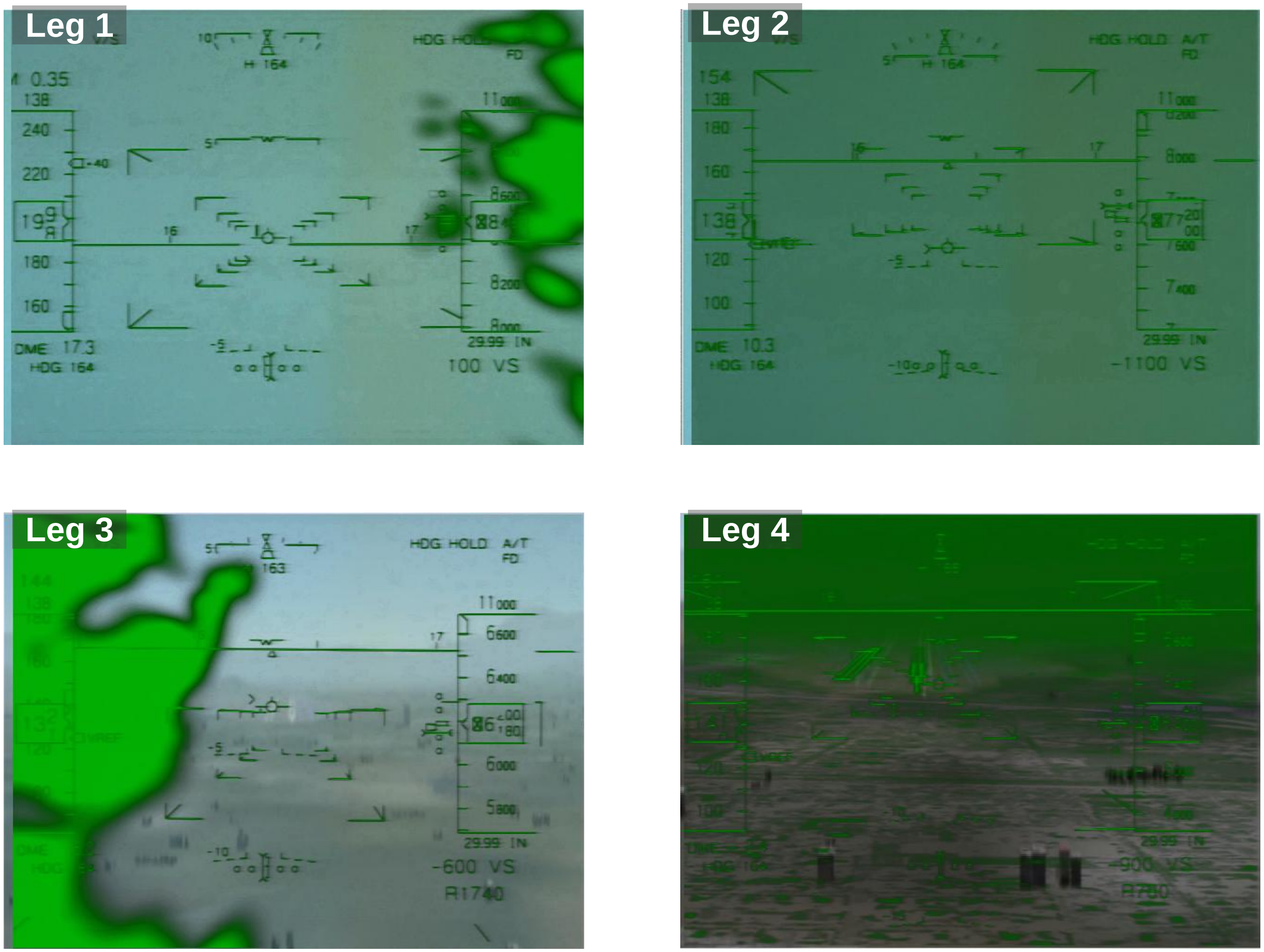


Figure 3.17. EVS HUD for IMC-day condition.

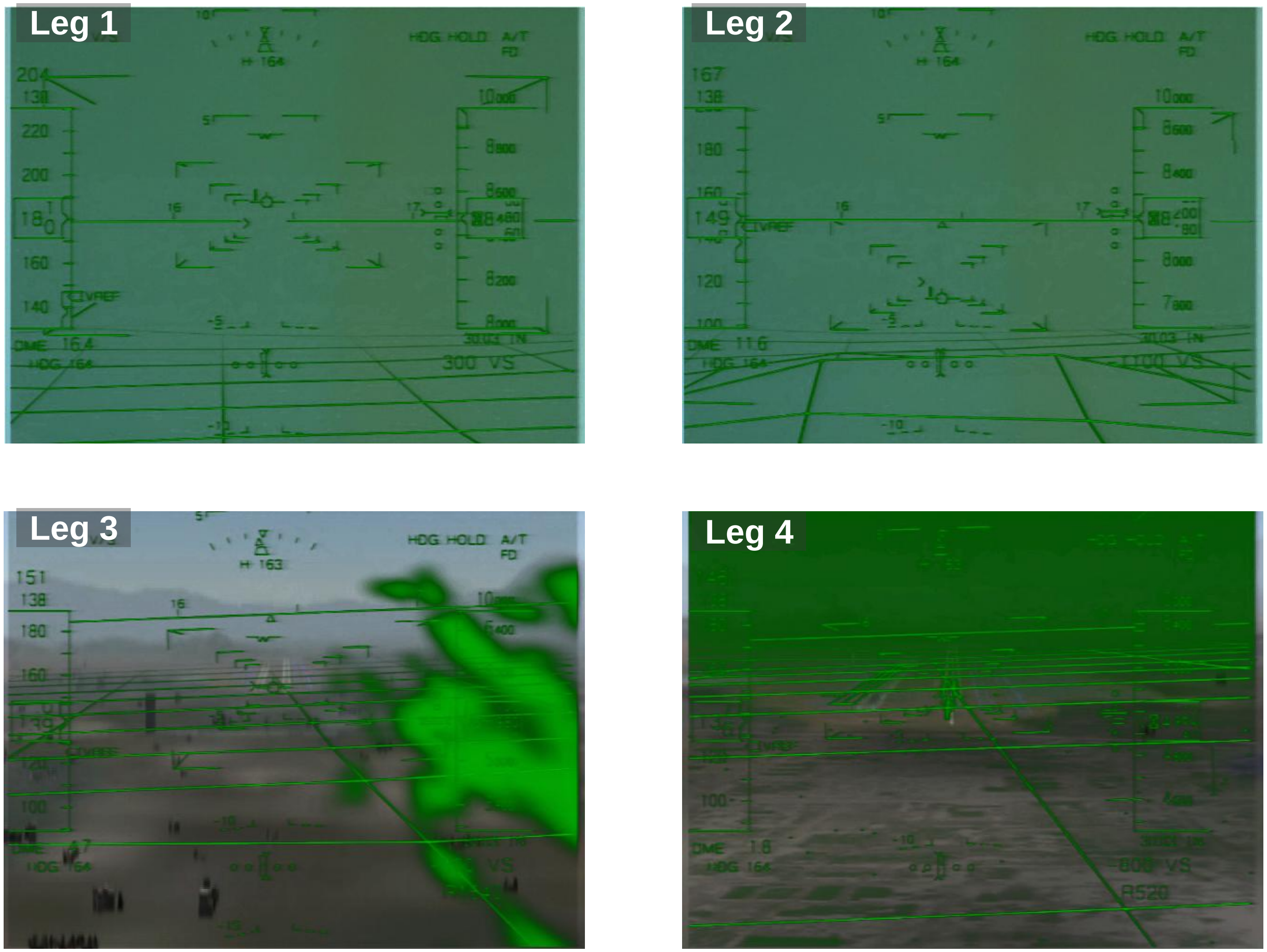


Figure 3.18. Combo HUD for IMC-day condition.

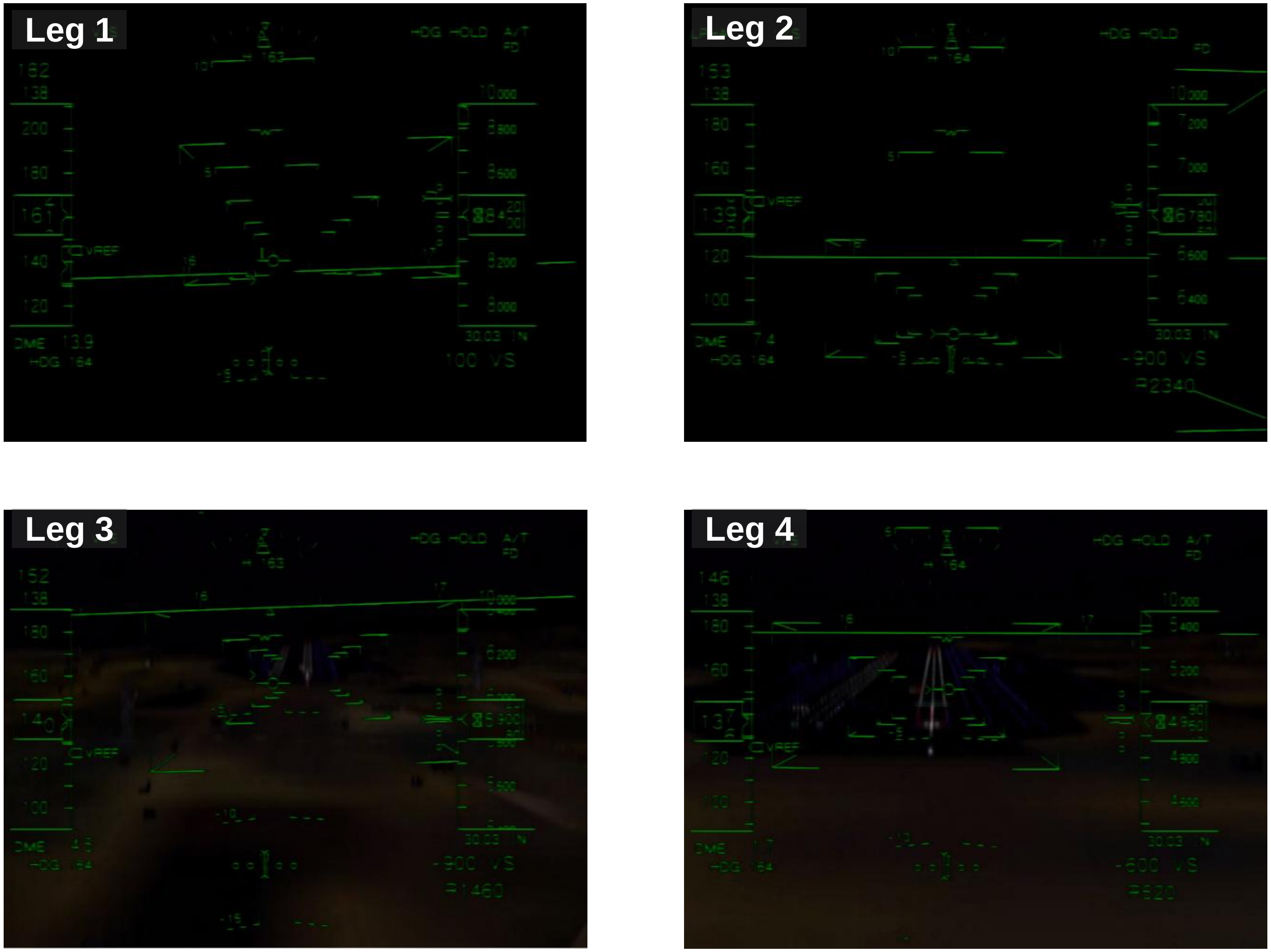


Figure 3.19. Baseline HUD for IMC-night condition.

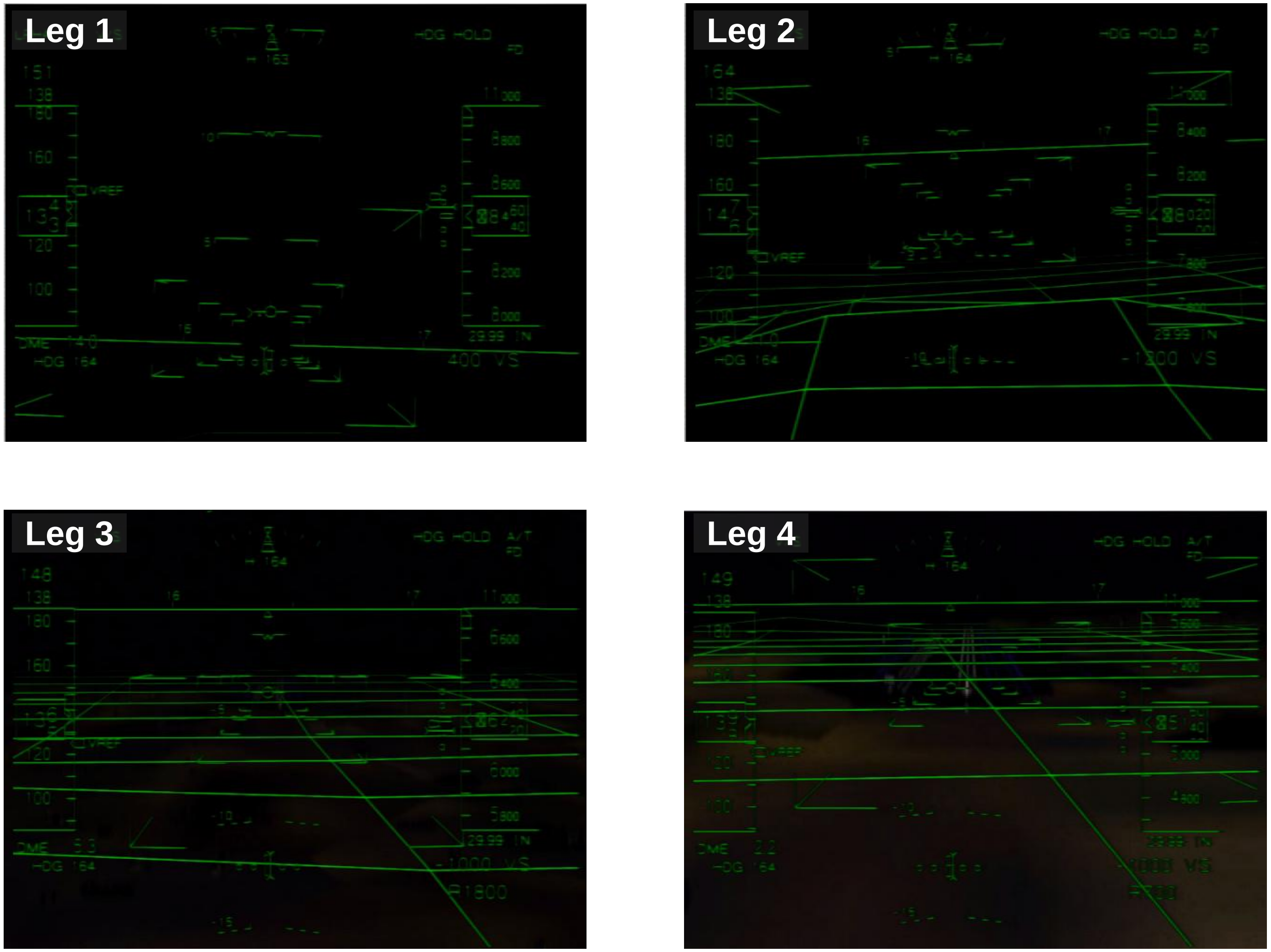


Figure 3.20. SVS HUD for IMC-night condition.

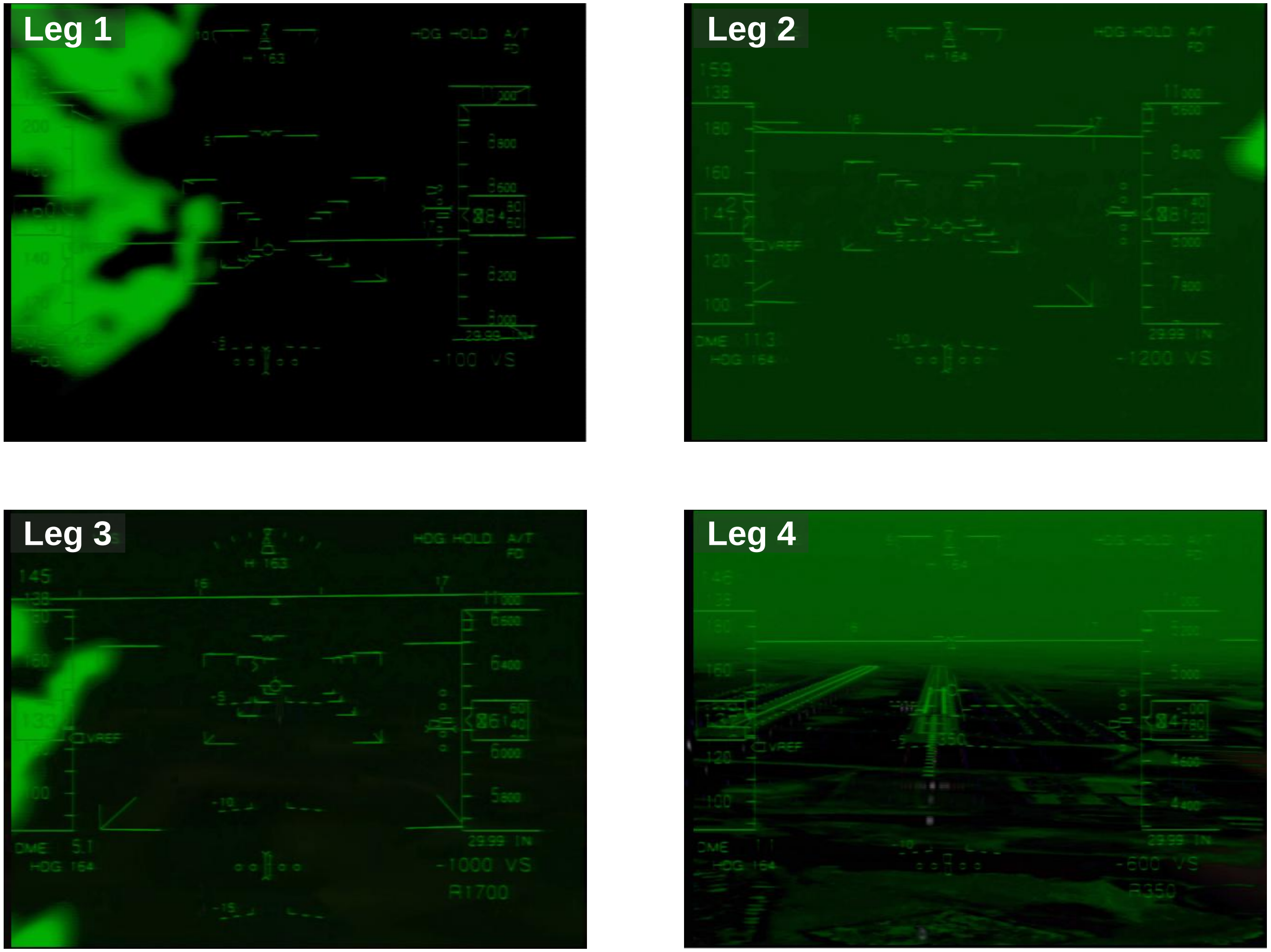


Figure 3.21. EVS HUD for IMC-night condition.

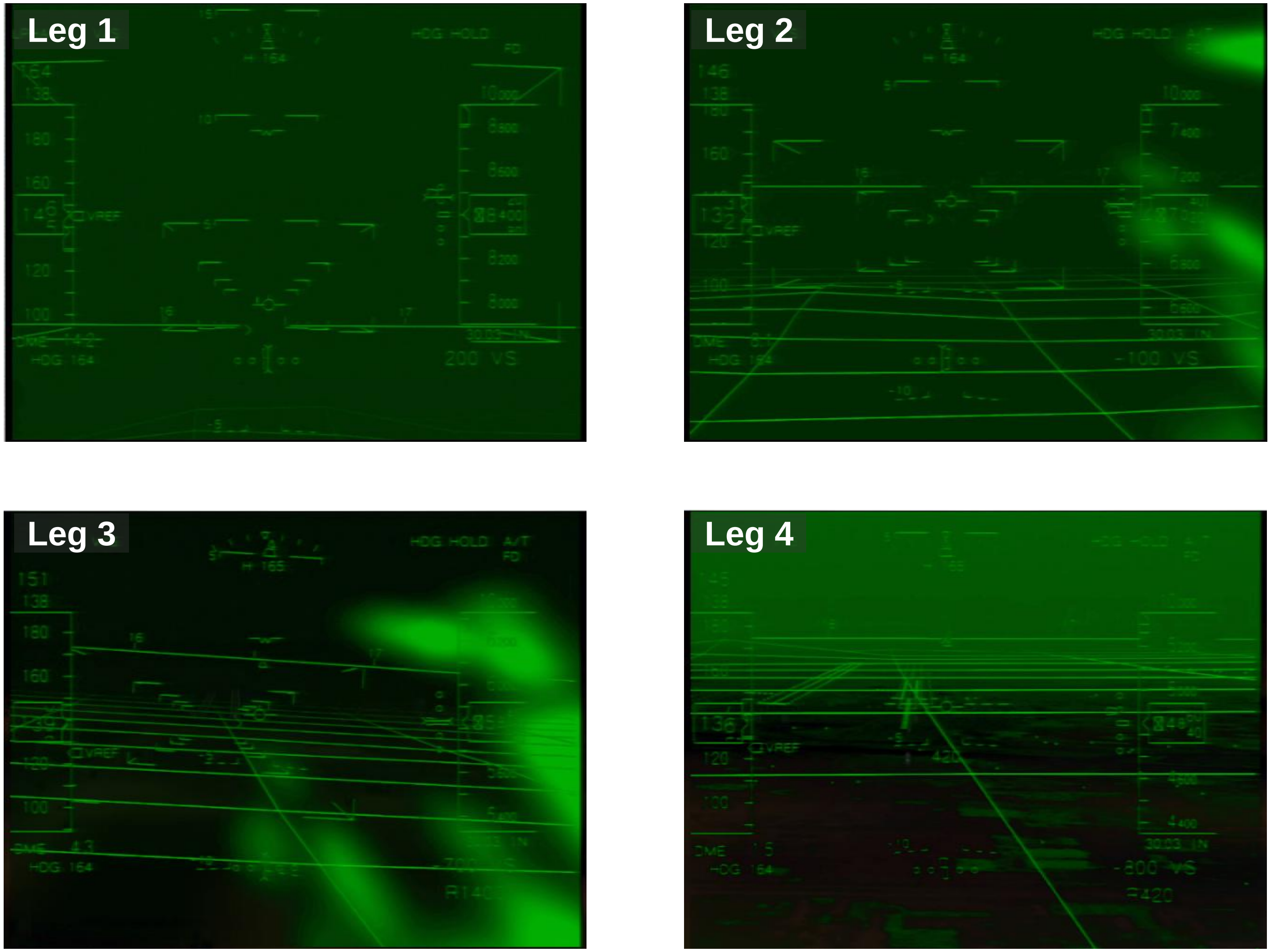


Figure 3.22. Combo HUD for IMC-night condition.

3.1.9. Dependant variables (DVs) and measures

The DVs measured in all test trials (except the 9th additional trial for collecting subject verbal protocols) included flight path control performance, SA, subjective/ physiological workload, and subjective pilot preferences.

Flight path control performance was measured on the tracking task by deviation of the overlaid cursor from the FPM presented in the video stimuli. As mentioned previously, since test pilots were shown a pre-recorded video displaying HUD content, including the FPM that appeared during the simulator flight by the expert pilot at NASA LaRC, it was impossible to allow test pilots to control the flight. Pilots were asked to control the yoke to trace the FPM displayed on the monitor with the overlaid cursor as closely as possible. After completion of each pilot's test, flight control errors were calculated by comparing the recorded tracking data for the overlaid cursor with the FPM position commended by the expert pilot when the video stimuli were created. A RMSE was computed for each trial and leg. Under the assumption that greater deviations in tracking control means subjects allocated less attention to the pathway tunnel image and the FPM, the RMSEs were used as measure of flight path control error.

Pilot SA during the flight task was assessed using SAGAT. SAGAT queries were prepared to represent the three levels of SA (Endsley, 1995b) for the three types of pilot SA (spatial, system and task awareness), as identified Wickens (2002). However, it should be noted here that, since the flight and displays investigated in this study were constrained by specific task and interface designs, definitions of the three types of pilot

SA were adapted from the definitions identified by Wickens (2002). While Wickens' categorization was focused on general flight for various interfaces and tasks, the definitions in this study were focused on the use of only the HUD for approach and landing tasks. With this in mind, the three types of pilot SA in this study were defined as follows:

- Spatial Awareness: Awareness of non-iconic information regarding spatial location (which does not require decoding of information to pilots), such as terrain features, tunnel, and path.
- System Awareness: Awareness of iconic information displayed in the HUD, indicating aircraft status. This information includes air/ground/vertical speed, MSL/radio altitude, altimeter setting, and DME to runway.
- Task Awareness: Awareness of communication on the flight deck (FO) and with ATC, as well as landing procedures (flaps, landing gear, landing checklist, and landing decision).

SA queries were formulated based on pilot information requirements for the approach and landing legs, similar to the SA queries generated and used in Snow and Reising's study (1999). SA queries were also based on a previous CTA of commercial jet aircraft piloting during ILS landings (Keller et al., 2003). The SA query set for this research was developed by analyzing this information and the required flight tasks in the scenario (see APPENDIX B: Situation Awareness Global Assessment Questions). Each query was categorized as representing one of the three levels and three types of SA.

During each trial, the simulator application was halted at random points in time during each leg and pilots were asked to complete a SA questionnaire, including nine queries (one query for each of the three levels by three types of SA) randomly selected from the overall set of queries. While pilots filled out the questionnaire, the video display was blanked.

Pilot workload was measured in two ways. Perceived workload was obtained using the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1998). The NASA-TLX measures mental, physical, and temporal demands, as well as performance, effort and frustration levels. These demands were differentially weighted based on pilot rankings of workload demand component and merged into a single workload index. The NASA-TLX ratings were collected at the end of each trial. In addition to the use of the NASA-TLX, participant heart rate (HR) data was collected as a physiological measure of workload. Svensson et al. (1997) found that pilot HR can be used as an indicator of psychophysical activation during missions. Svensson et al. (1997) also demonstrated HR to be positively correlated with pilot mental workload and the perceived complexity of the task. Lahtinen et al. (2007) also demonstrated that HR reflected the amount of cognitive load during a simulated flight for different legs of flight. With these results in mind, participant HR was measured with a heart rate monitor system (Polar Watch and Sunnto heart rate memory belt) for each trial and leg. The heart rate monitor system recorded pilot bpm (beat per minute) data in every 2 (Sunnto) or 5 second (Polar Watch). Mean bpm data in a leg for a trial was calculated and used for analyses.

In order to collect pilot subjective preferences for displays, a post-trial questionnaire was developed and presented to pilots at the end of each test trial. The questionnaire included five questions asking for subjective ratings of display usefulness for spatial awareness, system awareness, task awareness, enhancement of flight safety, and annoyance level. Ratings were made on a 7-point scale from “low” (1) to “high” (7).

3.1.10. Procedure

The procedure began with an introduction and explanation of the experiment and tasks to be completed. Following the introduction, participants were asked to review and complete an informed consent form (see APPENDIX C: Informed Consent Form), demographic data collection sheet (see APPENDIX D: Demographic Questionnaire) and payment from. Permission was also obtained from participants to videotape their test trials. Subsequently, pilots were instructed on the HUD features, flight scenario, and flight simulator system. After the instruction period, two practice trials were conducted so that participants could adapt to controlling the yoke and to become familiar with the task scenario, including the approach and landing procedures. The Baseline HUD in the IMC-night condition was used for the first practice trial and the Combo HUD in the IMC-day condition was used for the second. During the second practice trial, sample SA questionnaires were also presented to pilots at two random times to familiarize them with the questionnaires and administration procedure. After completing the two practice trials, pilots were asked to fill-out the NASA-TLX demand ranking form (See APPENDIX E:

NASA-TLX Form) and then, they were also asked to don the heart rate monitor device on their chest.

The pilots then completed eight experimental trials with several breaks. The experimenter sat on the right side of the pilot and acted as a FO, performing all other tasks except flight control and decision making, including flap and landing gear manipulations, ATC communications, landing checklist, and altitude call outs. During each trial, the simulator application froze the flight at a random time within each of the four legs of the approach and landing scenario. The experimenter asked the pilots to answer nine SAGAT queries randomly selected from the query set at each simulation freeze. Thus, participants answered four SAGAT questionnaires during each trial. After a pilot answered the SA queries, the simulation was resumed. At the end of each of the trials, pilots were asked to complete a NASA-TLX rating sheet (see APPENDIX E), and the post-trial preference questionnaire (see APPENDIX F: Post-trial Questionnaire). Pilots were asked about their subjective experience and their strategy in using the display features. This information was used as a basis for interpreting results and developing the cognitive task analysis.

Once a pilot completed the 8th experimental trial, the pilot completed an additional 9th trial. During the trial, the pilot was asked to verbalize what he was thinking and planning at that time. This trial and the pilots' comments were recorded using a video camera for further analyses.

After a pilot completed all nine trials, including the think aloud session, baseline

HR readings were collected while the pilot relaxed as much as possible in a darkened office for at least 10 minutes. The participants were then debriefed and paid for their participation. The entire procedure lasted approximately 4 hours for each pilot.

3.1.11. Specific hypotheses

Based on the literature review and the design of the experiment, several hypotheses were formulated regarding the various response measures. Regarding flight path control performance, it was expected that the tracking errors would be different for the four display conditions (Baseline, SVS, EVS, and Combo) and visibility conditions (IMC-day versus IMC-night), as well as for each of the four distinct legs of the approach and landing (**H1** (Hypothesis 1)). Flight path errors were predicted to increase as more terrain features (SVS/EVS symbologies) were overlaid on the HUD (**H1-1**) because the addition of these features might generate a clutter effect or cognitive tunneling effect for pilots. The features were also expected to distract pilot attention from the pathway tunnel and approach guidance. Such distraction could also be produced by the visibility conditions in the out-of-cockpit view. That is, IMC-day was expected to yield greater RMSEs than IMC-night because of lower saliency of symbology against the high brightness background (**H1-2**). In addition to this, as the required information for a pilot was different for each leg of flight, pilot attention patterns were expected to vary by leg, leading to a different profile of flight path tracking. Pilots were expected to produce higher RMSE values in Legs 1 and 4 than in Legs 2 and 3 (**H1-3**) because cognitive load might be higher in Leg 1 due to the need to manipulate instruments and in Leg 4 due to landing.

Regarding pilot SA as measured by SAGAT, overall SA scores and SA for each level and type were expected to vary by the four display configurations, two visibility conditions, and four legs of flight in the scenario (**H2**). SAGAT scores for the display configurations, including SVS and EVS features, were expected to be greater than displays presenting baseline features and the Combo (**H2-1**). This was because SVS or EVS only features may enhance SA beyond baseline due to terrain information, while the combination may degrade SA because of pilot distraction due to clutter effects and higher display density. For the visibility conditions, IMC-day was expected to produce lower SA than IMC-night because of low saliency of symbology against the high brightness background (**H2-2**). Among the four legs of flight, Legs 2 and 3 were expected to generate higher SAGAT scores than Legs 1 and 4 because they require more attention to terrain imagery (**H2-3**).

Subjective mental workload, measured by the NASA-TLX and the physiological measure (HR), was also predicted to be affected by the display configurations and visibility conditions (**H3**). In specific, among the four HUD configurations, workload for the baseline and combo display was expected to be higher than for the SVS-only and EVS-only configurations (**H3-1**). For example, SVS or EVS images in a HUD may support greater understanding of the out-of-cockpit situation by pilots compared to the baseline configuration, which did not present any terrain information. This, in turn, would reduce the mental workload for pilots in terms of scanning and perceiving the out-of-cockpit environment. Conversely, overlaying many display features on a HUD (e.g., the combination of SVS and EVS) may generate high perceived clutter and result in high workload ratings due to display density and low saliency of critical information. Between the two visibility conditions, IMC-day was expected to generate higher workload than IMC-night because of high display density and background brightness decreasing the saliency of the symbology against the out-of-cockpit scene (**H3-2**).

Since the NASA-TLX ratings were completed only once at the end of each test trial, it was impossible to examine the impact of flight segments on subjective workload from these ratings. Therefore, while the NASA-TLX data was used for assessing overall mental workload for display configurations and visibility conditions, the HR measure was used to assess the influence of the flight segment, display and visibility conditions on cognitive load. With this in mind, the HR data was expected to be affected by different legs of the ILS approach (**H3-3**). This was based on the findings of Lahtinen et al. (2007) revealing mean HR data to be higher in the initial and final portions of an ILS approach than in the intermediate portions. In the present study, it was expected that HR would be higher in Legs 1 and 4 than in Legs 2 and 3 due to the higher cognitive load expected in Leg 1 (associated with instrument manipulation) as well as in Leg 4 upon approach to the runway.

Finally, it was expected that the subjective preference ratings for each trial would be affected by the independent variables (displays and visibility) (**H4**). In general, preferences were expected to be correlated with the objective response measures (flight path control, SA and workload). In specific, pilot subjective ratings on questions regarding situation awareness (Q1, Q2 and Q3 (see APPENDIX F)) and flight safety (Q4) were expected to be higher (greater preference) for the SVS-only and EVS-only conditions than for the baseline and combo conditions. This was expected because the clear presentation of terrain in the SVS or EVS configurations. The combo display was expected to be less preferable because of pilot distraction due to perceived clutter (**H4-1A**). For the question on the level of display annoyance (Q5), ratings were expected to be related to flight path control performance (H1-1). Increased annoyance was expected with the presence of additional terrain features (**H4-1B**). For the visibility conditions, IMC-night was expected to be preferred (**H4-2**) because of the greater saliency of HUD information than in the IMC-day condition.

3.1.12. Data analyses

Flight path control performance was measured as RMSEs for each display configuration, visibility condition, and leg of flight (leg). Since all three IVs were within-subject variables, a three-way ANOVA (analysis of variance) was used to assess the effects of the variables on the RMSE response. The ANOVA model can be written as follows:

$$Y_{ijk} = \mu + D_i + V_j + L_k + DV_{ij} + DL_{ik} + VL_{jk} + DVL_{ijk} + \varepsilon_{ijk}$$

where,

Y_{ijk} = RMSE values

μ = gross mean

D_i = Display configurations (i=1, 2, 3, 4)

V_j = Visibility conditions (j=1, 2)

L_k = Leg of flight for approach and landing (k=1, 2, 3, 4)

ε_{ijk} = Errors

Since responses to SAGAT queries represent a binomial variable (correct or incorrect), the discrete nature of this measure violates parametric statistical test assumptions. However, Endsley (1995b) validated the arcsine function (e.g., Y' = arcsine(Y)) to be an effective transformation to account for this problem in the use of SAGAT. Applying a data transformation using the arcsine function, the normality assumption of parametric data analysis can be satisfied. With this in mind, the arcsine function was applied to the percentage of correct responses for each SA query. ANOVAs were then conducted on the transformed SAGAT scores for overall SA, SA by levels, and SA by types, with a statistical model similar to that used for RMSE data analysis.

In order to analyze the workload responses, two statistical models were used for each measure. First, since the NASA-TLX response did not account for the leg variable, a two-way ANOVA model was applied. Normalized NASA-TLX rating scores were used for the response variable in order to account for individual (internal workload scaling) differences between subjects. That is, all NASA-TLX scores for each subject were expressed as a statistical distance from the mean NASA-TLX score for that subject. The ANOVA model was as follows:

$$Y_{ij} = \mu + D_i + V_j + DV_{ij} + \varepsilon_{ijk}$$

where,

Y_{ijk} = normalized NASA-TLX scores

Second, for the physiological workload measure, individual incremental heart rates (ΔHR) for each experimental condition were calculated by making a comparison with the HR during rest (baseline HR). After normalizing the ΔHR data to account for individual differences, an ANOVA was also conducted to analyze the effect of display conditions, visibility conditions and legs of flight on HR, with a similar statistical model as for the RMSE and SA responses.

Regarding the subjective preferences measured by the post-trial questionnaire, the rating scale data was analyzed by a nonparametric statistical method because normality of the response could not be ensured. Therefore, a Friedman test as a two-way nonparametric alternative to ANOVA was conducted to assess the effect of display configuration and visibility condition on the preference ratings. Qualitative data (subject comments) from the questionnaire were used to identify reasons for patterns in preferences as well as other experimental results.

All statistical analyses were performed using SAS (Statistical Analysis Software). ANOVAs were conducted using PROC GLM and residual plots for all statistical models were examined to verify linearity, constant variance of error terms, independence of error terms, and normality of error term distribution. The pilots were also used to identify the presence of outliers, and inclusion of all significant variables in the model. If the data sets violated any of these assumptions, removal of outliers, or transformation of the response or predictor variables, was conducted prior to further analysis. An alpha level of 0.05 was used to identify any significant main effects and interactions. Further investigation of significant predictors was conducted using Duncan's Multiple Range tests with an alpha criterion of 0.05.

3.2. Part II: Cognitive Task Analysis

The objective of the CTA effort was to interpret and explain the results from the experiment in terms of human cognitive and psychomotor behavior. While previous empirical studies have demonstrated that specific HUD features (such as SVS and/or EVS) improved pilot performance, including measures of path tracking, SA and workload, there has been a lack of research to explain the effects of these technologies on pilot cognition and the origins of the performance results. Therefore, it is important to develop a theoretical framework to explain HUD feature effects in terms of cognitive processing. In this study, a concurrent think aloud method was used for the CTA. First, protocols were recorded and collected from the participant pilots. Second, protocols were transcribed and prepared for analysis. Third, the content of protocols in form of task lists was analyzed. Finally, the results of the CTA were used to interpret empirical study data.

3.2.1. Procedure

Collection and analysis of pilot verbal protocol data during and after the 9th trial followed the general procedure identified by Bainbridge and Sanderson (1990). The procedure in the present study included three sequential steps, as follows:

- *1st step: Collection and storage of protocol data*. In order to accurately capture what the pilot was doing and what information he needed to understand, the pilot was asked to speak out his/her internal thoughts, intentions or decisions (concurrent think-aloud) while he performed the 9th trial with the assigned HUD and visibility configuration. On occasion, the experimenter asked questions in the form of a probing in order to cause pilots to elaborate or clarify the verbal protocols provided. The sample questions are presented in Table 3.3. The questions were formulated to relate to the various response measures collected during the eight test trials. This combination of verbal protocol and probing interview is regarded as a viable method for systematic knowledge elicitation (Shadbolt & Burton , 1990).
- *2nd step: Preparation of protocol data for analysis.* After the experiment was completed, the videotapes of pilot performance in the 9th trial were transcribed for each experimental condition. Based on the transcription, a general protocol structure was identified and the protocol was divided into meaningful phrases.
- *3rd step: Analysis of explicit and implicit content.* In this step, three structures of protocols were summarized in the form of lists, including sequential/non-sequential tasks, target objects, and critical comments. In addition to these, several rules or

patterns for information acquisition were identified.

Table 3.3. Sample questions for verbal protocol analysis.

	Questions
Pilot SA	− How well are you aware of the status of the aircraft? − Which HUD feature do you mostly focus on, excluding the tracking task? − Would you tell me the order of priority of your usage of HUD features? − Which features do you focus on? − What do you think you are not aware? − Which aspects of the HUD may obstruct your awareness of the status of the aircraft? − How well are you aware of the terrain features? − If you are uncomfortable in your understanding of the terrain image, what makes you confused?
Pilot workload	− Would you tell me how complex the task is? − Which HUD features create workload for you? − Which features cause you to devote effort to performing the task? − How much do you think about your performance? − Is the tracking task difficult for you?

3.2.2. Outcomes of the CTA

The CTA on the flight task for approach and landing using the various HUD configurations under IMC conditions yielded three types of outcomes. First, a list of sequential events and tasks was developed for the flight scenario considering actual pilot behaviors. This approach was based on previous CTA studies (Keller & Leiden, 2002; Keller et al., 2003) that identified pilot tasks for approach to an airport with a commercial aircraft using various types of cockpit interfaces (ILS, RNAV and SVS-HDD). In the present study, the list was expected to provide understanding of: which sequential external/internal behaviors were performed in each leg of the flight; what information triggered such behaviors; who is responsible for handling the information and actions (pilot or FO) since the flight situation was assumed as a dual-piloted commercial aircraft; and which objects were required for manipulating the information and actions.

Second, a list of non-sequential tasks performed with the sequential tasks was identified. This list was also developed based on the format of CTA conducted by the Keller et al. (2003). The list was expected to provide information on specific actions in the non-sequential tasks (e.g., airspeed control, altitude control, and heading control) and to identify which objects were required for the task. In addition to this, several alternative behaviors for achieving a specific task were identified.

Finally, a list of critical comments was summarized from the verbal protocols and probes. This list was used to explain the results from the experiment, along with the list of sequential/non-sequential tasks.

Consequently, the results of the CTA were matched with the results of the empirical study in order to explain observed flight performance for the various HUD configurations. Such a descriptive explanation of experimental results through the CTA provided a rationale for optimal SVS/EVS HUD design. That is, while comparisons of various display technologies through lab experiments tell us whether alternative "A" is better than "B", the proposed CTA was expected to provide a platform for a cognitive explanation of SVS/EVS features effects on pilot performance as well as description of cognitive behaviors with HUDs during simulated flight performance.

4. RESULTS

4.1. Part I: Results of Experiment

4.1.1. Analysis of tracking difficulty

Prior to analyzing the effect of the display, weather condition, and leg of flight on tracking performance (RMSE), it was necessary to examine the level of difficulty of the tracking task, as presented in the videos, and determine whether differences in RMSE were generated by the pilots or by variations in the video. First, raw position data (x, y) for the FPM in the videos was recorded for each display condition across the four legs of the scenario. The sampling rate of the position data was 12 Hz and the observations were used as a basis for calculating the RMSE for pilot flight path control. Second, the raw position data for the FPM were resampled for each second in the flight scenario order to magnify any changes in position. Third, FPM position changes, Δx (=x_t-x_{t-1}) and Δy (=y_t-y_{t-1}), were calculated for each display configuration across all legs of the flight. Finally, time series analysis (using SAS "PROC REG" including time parameters in the model) was used to identify any significant changes in Δx and Δy for the experimental conditions.

Results demonstrated that changes in the x position of the FPM were not affected by display condition (p=0.9947) or leg (p=0.9686). Results also revealed the display and leg conditions to be insignificant (p=0.9836 and p=0.4316, respectively) in changes in the y position of the FPM. Since the same video clip for each display configuration was used for both visibility conditions (IMC-day vs. IMC-night), the tracking task difficulties for each visibility condition were identical. From these results, it was confirmed that there were no significant differences in tracking task difficulty among experiment conditions (display, IMC, and leg).

4.1.2. Flight path control performance

Based on residual analyses, a double log transformation was applied to the RMSE response variable (Y'=log (log (RMSE))) in order to satisfy the statistical assumptions of the ANOVA including normality, constant variance, and independence of residuals. In addition to this, based on the residual analysis, a single outlier was removed from the dataset (RMSE for Leg 4 of the 2nd trial (EVS HUD) for Pilot 2) because it was observed that the pilot lost track of the flight task in the trial. That is, during the 4th leg, he aimed the display cursor at the runway and glideslope reference line, instead of the FPM.

ANOVA results showed that several main and interaction effects were significant in terms of RMSE response (see Table 4.1). The display configuration influenced pilot path control ($F(3,223)=14.37$, $p<0.0001$). A post-hoc analysis showed that pilots made greater errors when the SVS was active ($M=15.7$) versus the Combo ($M=14.3$), Baseline ($M=12.9$) and EVS ($M=11.3$) conditions. The Combo and Baseline produced higher RMSE than the EVS condition (see Figure 4.1). This indicated that the SVS HUD was associated with less effective flight path control than the other displays, while the EVS HUD was most effective.

Table 4.1. ANOVA results on RMSE data.

Source	df	Type III SS	Mean Square	F	P
**Display	3	0.324819	0.108273	14.37	<.0001
**IMC	1	0.458698	0.458698	60.89	<.0001
Leg	3	0.021173	0.007058	0.94	0.4236
IMC*Leg	3	0.052812	0.017604	2.34	0.0746
**Display*IMC	3	0.211148	0.070383	9.34	<.0001
**Display*Leg	9	0.176666	0.01963	2.61	0.007
Display*IMC*Leg	9	0.07069	0.007854	1.04	0.407
Error	223	1.67998204	0.00753355		

** - significant at $p < 0.05$ level.

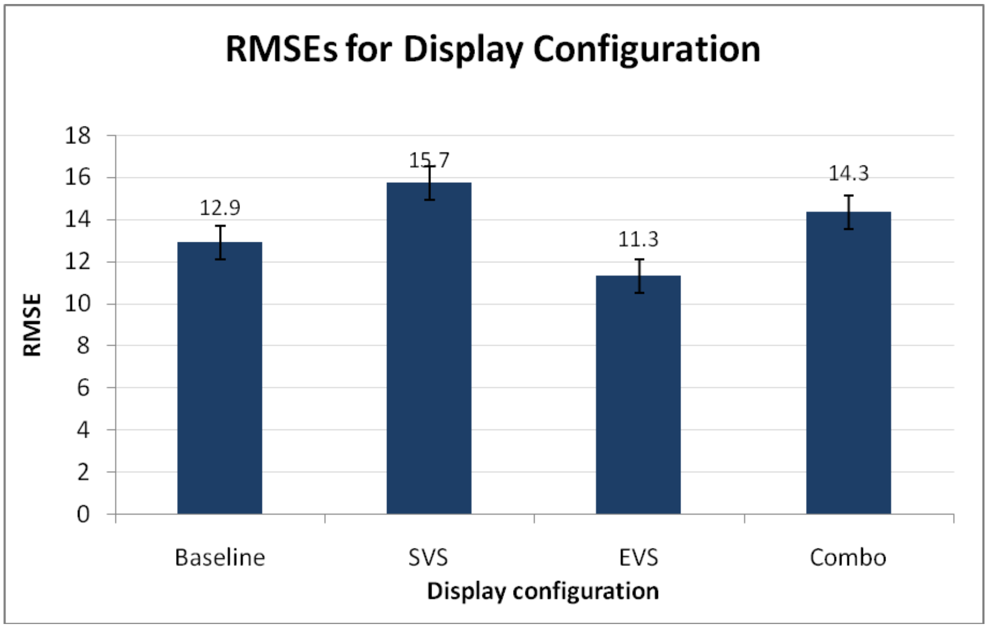


Figure 4.1. RMSEs for each display configuration.

ANOVA results also revealed that there was a significant effect of the IMC condition on tracking performance ($F(1,223)=60.89$, $p<0.0001$). The IMC-day condition (M=14.3) was associated with greater errors in the tracking task than the IMC-night condition (M=12.0). This was in line with Hypothesis 1-2.

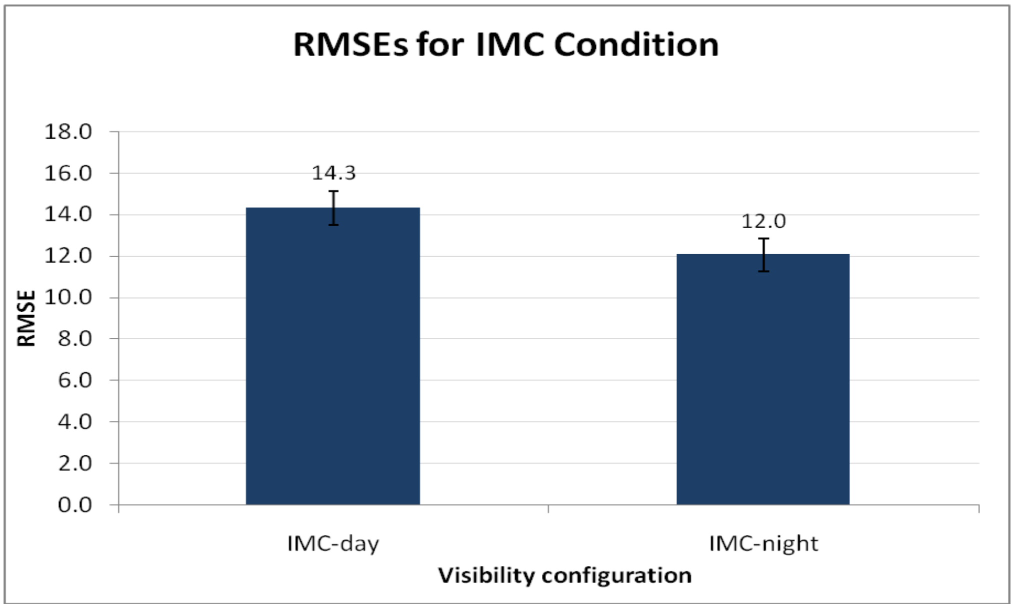


Figure 4.2. RMSEs for IMC conditions.

There was no significant effect of leg ($F(3,223)=0.94$, $p=0.4236$) and no interaction of IMC condition and leg ($F(3,223)=2.34$, $p=0.0746$) on RMSE. However, ANOVA results revealed an interaction effect among the display and IMC conditions ($F(3,223)=9.33$, $p<0.0001$). Figure 4.3 shows the interaction plot. The Figure indicates the SVS-HUD under the IMC-day condition produced higher RMSE than the same HUD under the IMC-night condition. In general, IMC-day conditions were associated with higher RMSE than IMC-night conditions.

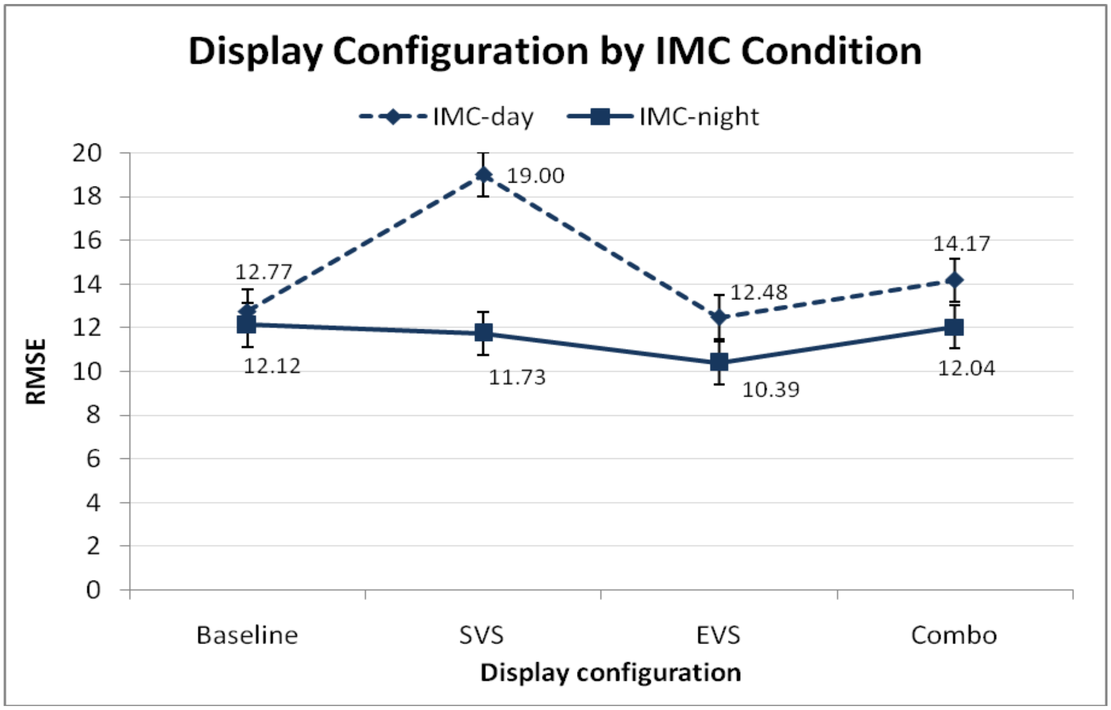


Figure 4.3. RMSEs for display configuration by IMC condition.

ANOVA results also revealed a significant interaction effect between display and flight leg on RMSE (F(9,223)=2.61, p=0.007). Figure 4.4 presents the RMSEs for each display configuration for the four legs. In general, the SVS-HUD yielded higher RMSEs and the EVS-HUD produced lower RMSEs across legs. However, the Combo-HUD in Leg 4 generated higher tracking error than the other displays in other legs.

There was no three-way interaction effect among the experimental manipulations

(F(9,223)=1.04, p=0.4070).

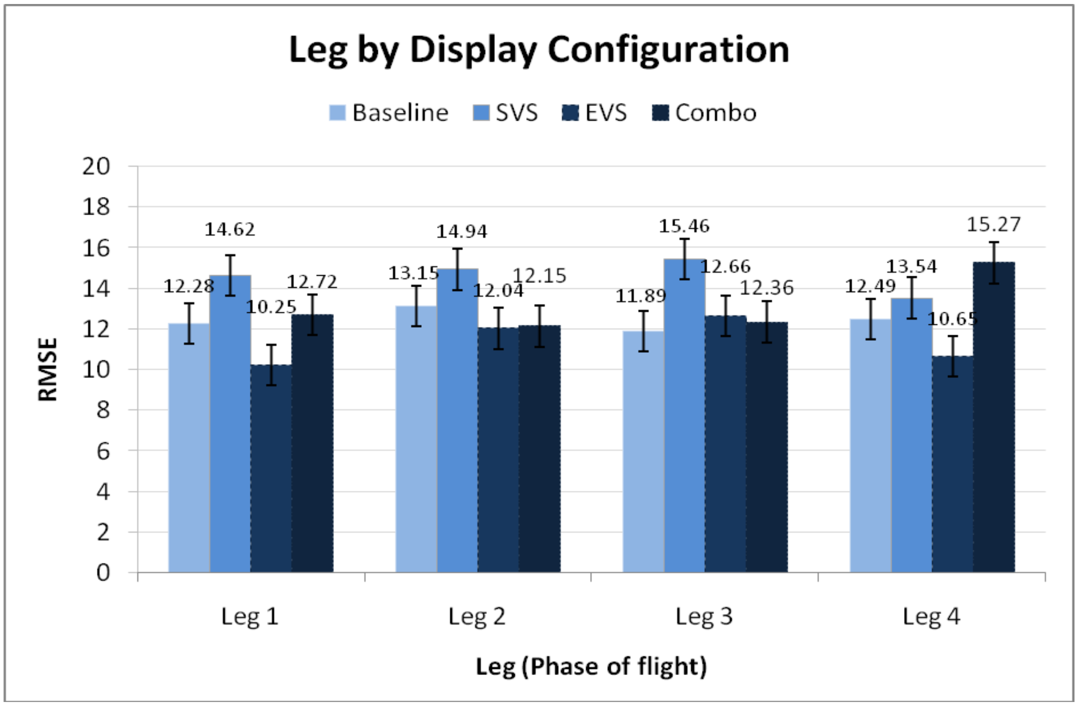


Figure 4.4. RMSEs for leg by display configuration.

4.1.3. Pilot SA

Results of an ANOVA revealed significant main and interaction effects of display configuration, IMC condition, and leg of flight on SAGAT scores, including overall SA, and for various levels and types of SA. Table 4.2 shows a summary of F-test results on the SAGAT scores.

Table 4.2. Summary of F-test results on SAGAT scores.

Independent Variables	Overall SA	Levels of SA			Types of SA		
		Level 1	Level 2	Level 3	Spatial Awareness	System Awareness	Task Awareness
Display	F(3,224)=3.04 p=0.0300**	F(3,224)=2.37 p=0.0712	F(3,224)=1.42 p=0.2374	F(3,224)=1.73 p=0.1614	F(3,224)= 1.1 p= 0.3517	F(3,224)= 4.55 p= 0.0041**	F(3,224)= 0.82 p= 0.4823
IMC	F(1,224)=1.22 p=0.2701	F(1,224)=0.44 p=0.5096	F(1,224)=5.99 p=0.0151**	F(1,224)=0.56 p=0.4555	F(1,224)= 2.3 p= 0.1306	F(1,224)= 3.99 p= 0.0469**	F(1,224)= 3.39 p= 0.0668
Leg	F(3,224)=9.75 p<0.0001**	F(3,224)=2.28 p=0.0799	F(3,224)=1.15 p=0.3292	F(3,224)=10.62 p<0.0001**	F(3,224)= 9.37 p< 0.0001**	F(3,224)= 10.36 p < 0.0001**	F(3,224)= 11.12 p< 0.0001**
Display*IMC	F(3,224)=0.76 p=0.5188	F(3,224)=0.88 p=0.4532	F(3,224)=1.45 p=0.2289	F(3,224)=0.14 p=0.9334	F (3,224)= 0.57 p=0.6334	F (3,224)= 0.99 p=0.3978	F (3,224)= 1.68 p=0.1711
Display*Leg	F(9,224)=1.91 p=0.0509	F(9,224)=1.32 p=0.2282	F(9,224)=1.45 p=0.1686	F(9,224)=1.66 p=0.1009	F (9,224)= 0.55 p=0.8386	F (9,224)= 2.31 p=0.0167**	F (9,224)= 0.84 p=0.5827
IMC*Leg	F(3,224)=1.77 p=0.1544	F(3,224)=0.66 p=0.5748	F(3,224)=0.56 p=0.6412	F(3,224)=4.28 p=0.0059**	F (3,224)= 0.97 p= 0.4064	F (3,224)= 2.84 p= 0.0386**	F (3,224)= 0.97 p= 0.4091
Display*IMC*Leg	F(9,224)=0.62 p=0.7786	F(9,224)=0.63 p=0.7722	F(9,224)=1.08 p=0.3754	F(9,224)=0.86 p=0.5604	F (9,224)= 1.05 p=0.4030	F (9,224)= 0.42 p=0.92237	F (9,224)= 0.22 p=0.9911

** - significant at p < 0.05 level.

4.1.3.1. Overall SA score

ANOVA results revealed significant effects of HUD configuration ($F(3,224)=3.04$, $p=0.03$) and leg of flight ($F(3,224)=9.75$, $p<0.0001$) on overall SA score. A post-hoc analysis categorized the display into two groups: one group consisted of SVS (M=58.7%), Baseline (M=55.5%), and the Combo (M=54.5%) condition, which produced higher SA scores; and another group consisted of the Baseline, Combo, and EVS (M=49.8%) conditions, which were associated with lower SA scores (see Figure 4.5).

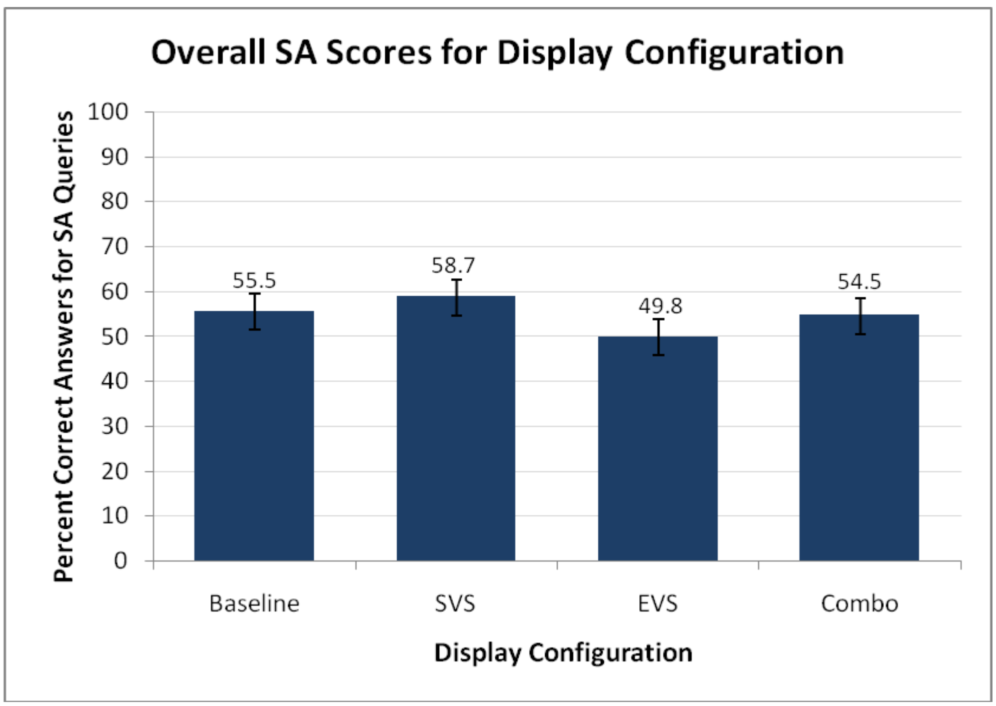


Figure 4.5. Overall SA scores for each display configuration.

A post-hoc analysis for the effect of leg of flight on overall SA score also yielded two distinctive groups of legs. That is, the SA scores for Leg 1 (M=62.4%) and Leg 3 (M=57.4%) were higher than for Leg 2 (M=49.5) and Leg 4 (M=48.9). Figure 4.6 shows the overall SA scores for the four legs of flight.

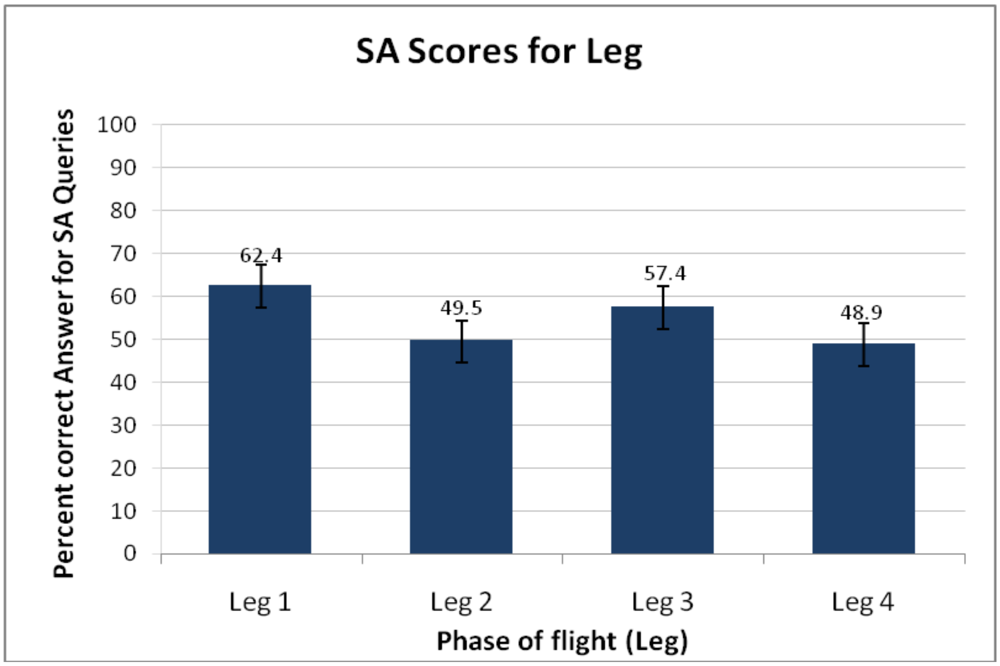


Figure 4.6. Overall SA scores for each leg.

4.1.3.2. SA score by levels

ANOVA results revealed no significant main or interaction effects on Level 1 SA. However, there was a significant effect of IMC condition on Level 2 SA ($F(1,224) = 5.99$, $p=0.0151$), indicating IMC-night produced higher Level 2 SA scores (M= 54.4%) than the IMC-day condition (M=44.2%). ANOVA results also revealed the effect of leg to be significant on Level 3 SA scores (F(3,224)=10.62, p<0.0001). A post-hoc analysis on this effect yielded the same groups as the post-hoc analysis result on overall SA scores. That is, Leg 3 (M=72.6%) and Leg 1 (M=68.1%) were associated with higher Level 3 SA scores than Leg 2 (M=50.7%) and Leg 4 (M=47.9%).

The interaction effect between IMC and leg was found to be significant on Level 3 SA (F(3,224)=4.28, p=0.0059). Figure 4.7 presents the interaction plot of Level 3 SA scores for the IMC conditions by leg. From the Figure, it can be observed that the IMC-day condition during Leg 3 resulted in higher pilot projection scores compared to other legs and IMC-night condition.

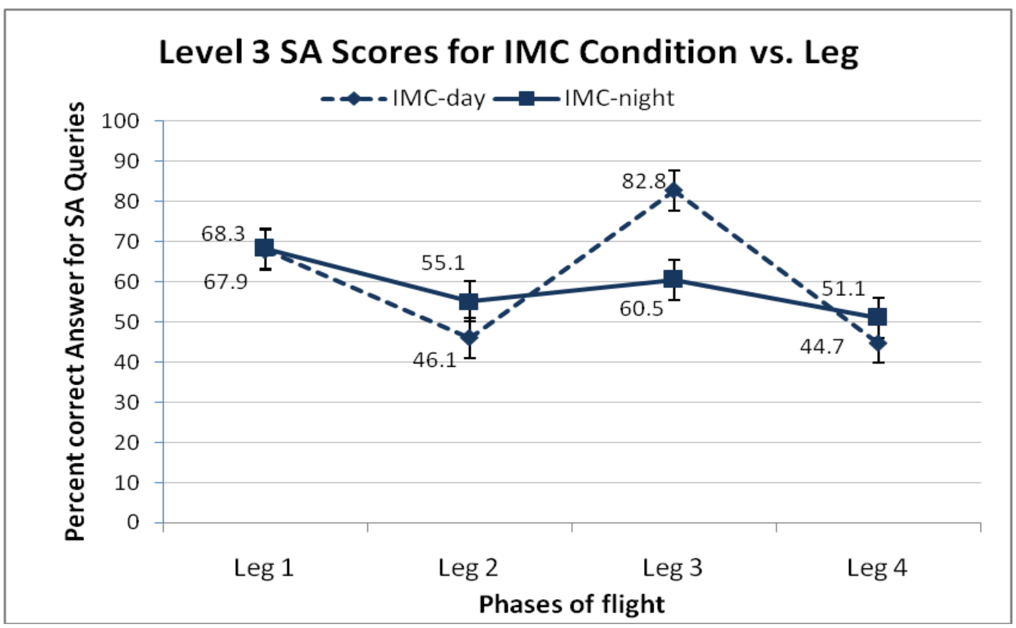


Figure 4.7. Level 3 SA scores for IMC condition by leg.

4.1.3.3. SA score by types

Regarding the effect of display configuration on the three types of pilot SA, ANOVA results revealed a significant effect on pilot system awareness ($F(3,224)=4.55$, $p=0.0041$). Table 4.3 shows the post-hoc analyses for the effect of display configuration on the three types of pilot SA and Figure 4.8 presents the associated graph. From the Table and Figure, it can be observed that system awareness was degraded by the EVS-HUD while the other display effects were comparable.

ANOVA results also revealed the effect of the IMC condition to be significant on system awareness ($F(1, 224)=3.99$, $p=0.0469$). IMC-day (M=58.5%) was associated with higher SA scores for system awareness than the IMC-night condition (M=49.9%). Mean SA scores for IMC-night were higher than for IMC-day for spatial awareness and task awareness, though not significantly. Figure 4.9 shows the SA scores for the three types of pilot SA by IMCs.

Table 4.3. Results of post-hoc analyses for display effect on three types of pilot SA.

	Spatial Awareness	**System Awareness**	**Task Awareness**
Group 1	Combo (M=51.8%) Baseline (M=47.9%) SVS (M=47.2%) EVS (M=40.5%)	Baseline (M=61.5%) SVS (M=61.2%) Combo (M=51.4%)	SVS (M=85.3%) EVS (M=83.3%) Combo (M=79.4%) Baseline (M=78.2%)
Group 2		Combo (M=51.4%) EVS (M=41.9%)	

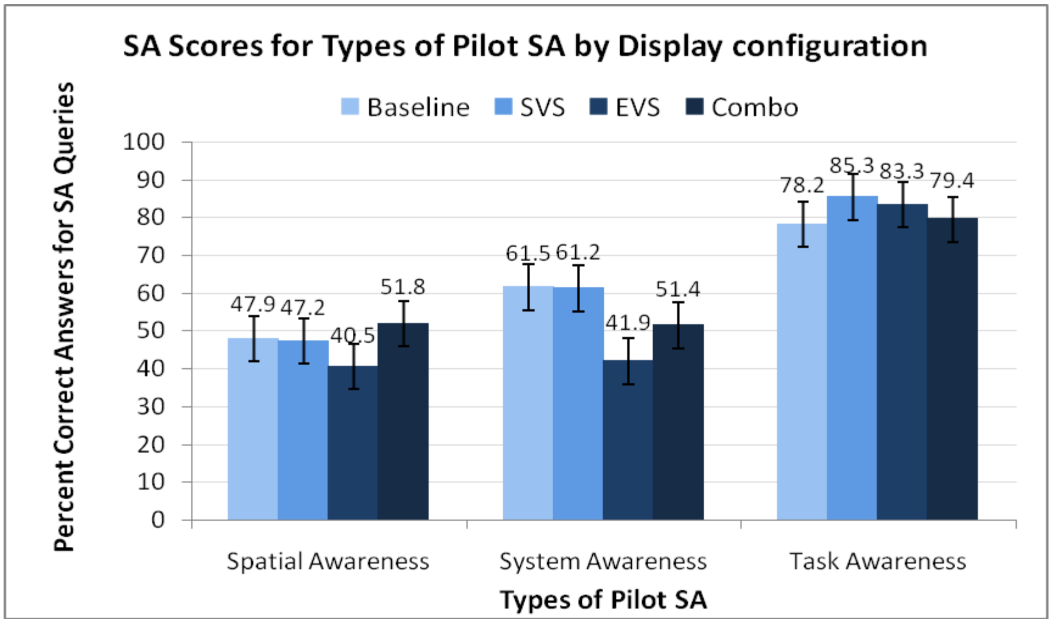


Figure 4.8. SA scores for types of pilot SA by display configuration.

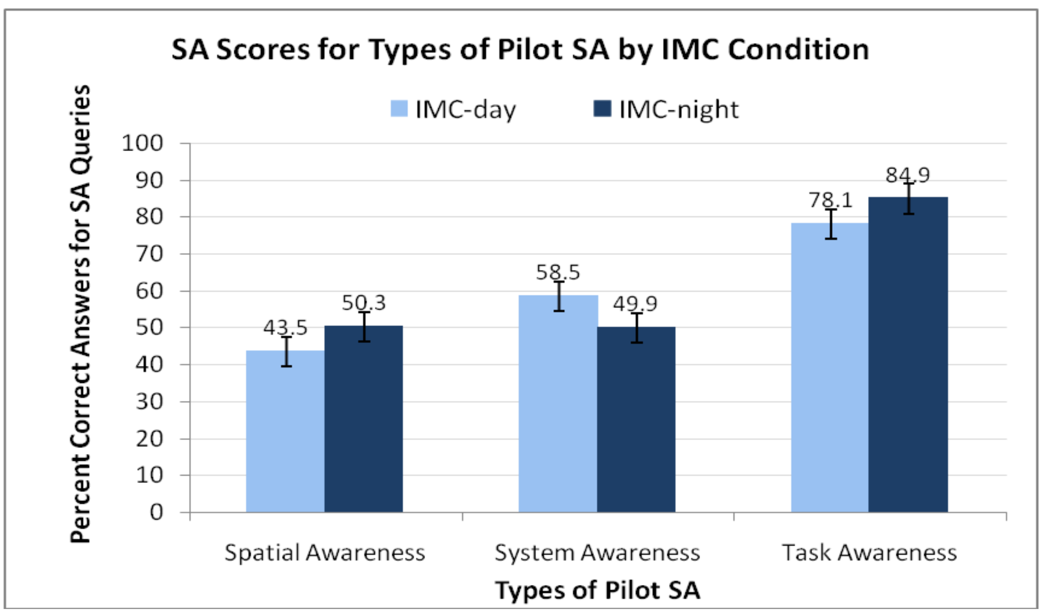


Figure 4.9. SA scores for types of pilot SA by IMC condition.

The effect of leg of flight was revealed to be significant for all types of pilot SA, including spatial awareness ($F(3,224)=9.37$, $p<0.0001$), system awareness ($F(3,224)=10.36$, $p<0.0001$), and task awareness ($F(3,224)=11.12$, $p<0.0001$). Table 4.4 summarizes the post-hoc analyses on the three types of pilot SA and Figure 4.10 presents the associated plot.

Table 4.4. Results of post-hoc analyses for leg effect on three types of pilot SA.

	Spatial Awareness	**System Awareness**	**Task Awareness**
Group 1	Leg 3 (M= 61.5%) Leg 4 (M= 53.9%)	Leg 1 (M= 73.1%)	Leg 1 (M= 93.6%)
Group 2	Leg 1(M= 38.7%) Leg 2 (M= 31.8%)	Leg 3 (M= 49.4%) Leg 4 (M= 46.7%) Leg 2 (M= 45.4%)	Leg 2 (M= 85.1%) Leg 3 (M= 77.6%)
Group 3			Leg 4 (M= 64.8%)

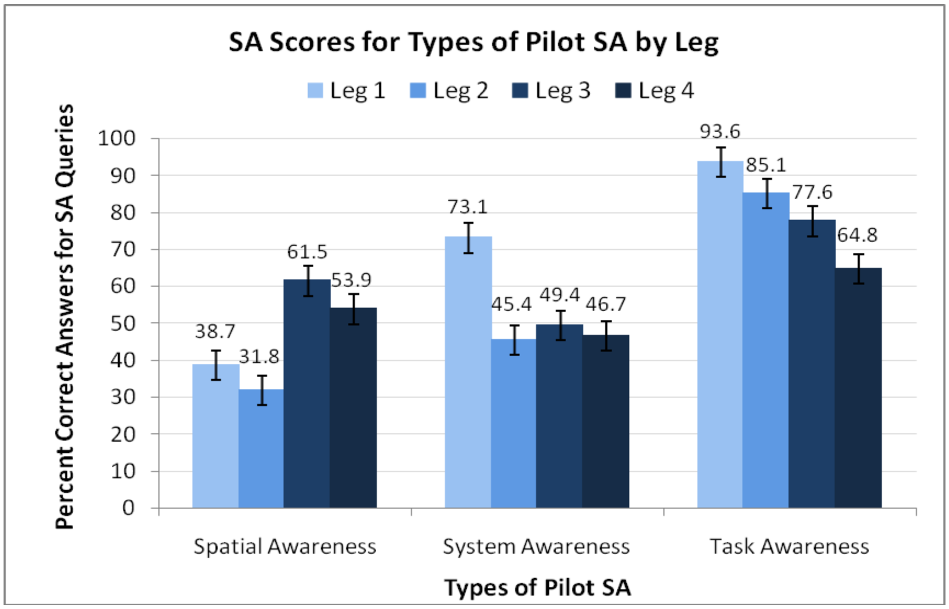


Figure 4.10. SA scores for types of pilot SA by leg.

From Table 4.4 and Figure 4.10, it can be observed that: spatial awareness was higher in Legs 3 and 4 than Legs 1 and 2; system awareness was higher in Leg 1 than the other Legs; and Task awareness decreased as the flight flew through the sequential legs.

ANOVA results revealed two kinds of two-way interactions on pilot system awareness, while SA scores for spatial and task awareness were not affected by any interactions. The interaction between display configuration and leg was significant (F(9,224)=2.31, p=0.0167) and Figure 4.11 shows the interaction plot. From the Figure, it can be observed that the SVS feature generated higher system awareness than the other display configurations at the IAF and beginning of descent (Legs 1 and 2). However, after the actual out-of-cockpit view was presented (on Legs 3 and 4), the Baseline display produced higher awareness than the other display configurations.

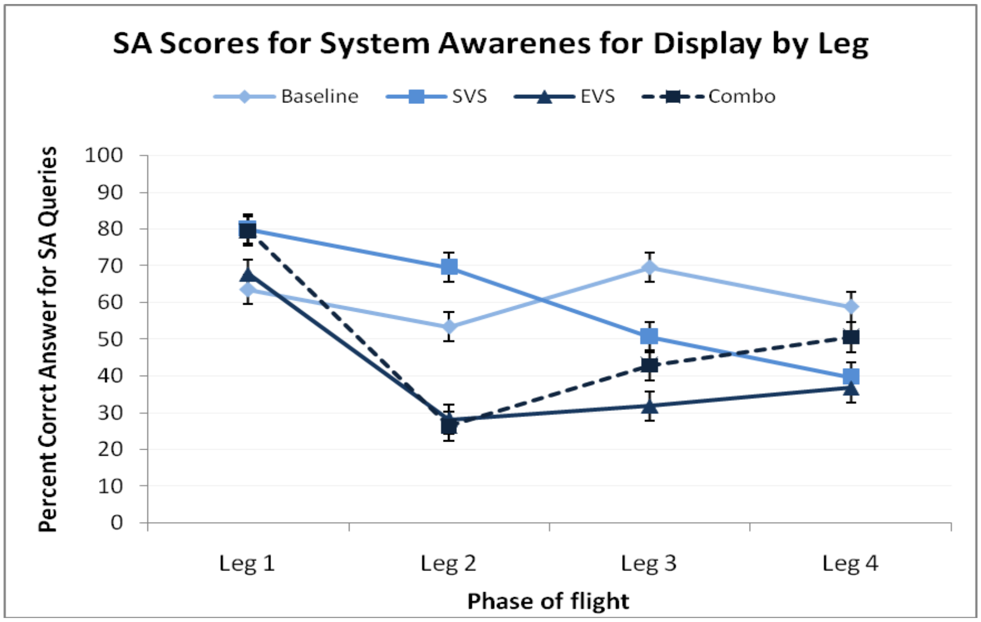


Figure 4.11. SA scores for system awareness for display configuration by leg of flight.

The interaction between IMC condition and leg was also significant on system awareness (F(3,224)=2.84, p=0.0386). Figure 4.12 presents the interaction plot between IMC condition and leg, indicating, in general, IMC-day was associated with higher system awareness during Legs 1 and Leg 3 than in the IMC-night condition.

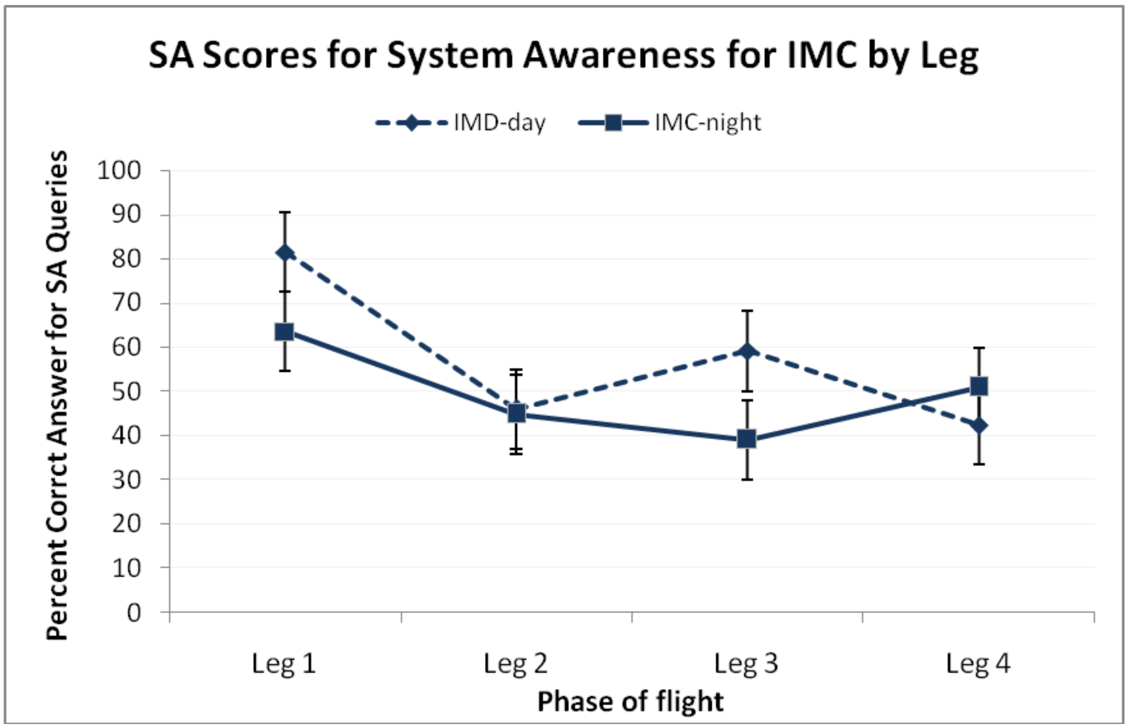


Figure 4.12. SA scores for system awareness for IMC condition by leg of flight.

4.1.4. Pilot workload

As described previously, two measures of pilot workload were collected in this study. An ANOVA on pilot workload ratings, z-scored NASA-TLX ratings, failed to reveal any significant main effects of display configuration, IMC condition, or any interaction between display and IMC. Results of a power analysis on the ANOVA results revealed the NASA-TLX response to have sufficient sensitivity for revealing differences in cognitive load among the experimental manipulations (beta values were less than 0.2). Similar to this, an ANOVA revealed no significant main or interaction effects of display, IMC and leg on the physiological workload response using individual incremental heart rates (ΔHR). Only trial order was revealed to be significant ($F(7,218)=24.61$, $p<0.0001$).

4.1.5. Subjective preference ratings

Most results of the non-parametric tests (Friedman test as a two-way non-parametric alternative to the ANOVA) for the effects of display configuration and IMC condition on each of the five subjective survey questions revealed no significant differences in pilot ratings (see Table 4.5). However, the test revealed the ratings on annoyance level to be marginally significantly different between display configurations ($F(3,56)=2.44$, $p=0.0740$). A post-hoc analysis revealed two groups of display configurations. One consisted of the Combo (M= 4.8), EVS (M= 4.0), and SVS (M=3.4) conditions, which were associated with higher annoyance levels. Another group consisted of EVS, SVS and Baseline (M=2.38), which yielded lower annoyance levels. Figure 4.13 shows the graph of mean ratings of annoyance for each display configuration. It suggests that pilots experienced increasing discomfort as features were added to the Baseline configuration.

Table 4.5. Summary of F-test results on five subjective preference questions.

Source	Question 1- Spatial Awareness	Question 2- System Awareness	Question 3- Task Awareness	Question 4- Flight Safety	Question 5- Annoyance Level
Display	F(3, 56)= 0.79, p=0.5067	F(3, 56)= 0.19, p=0.3207	F(3, 56)= 0.63, p=0.5991	F(3, 56)= 1.32, p=0.2751	F(3, 56)= 2.44, p=0.0740
IMC	F(1, 56)= 0.00, p=0.9782	F(1, 56)= 0.85, p=0.3617	F(1, 56)= 0.10, p=0.7588	F(1, 56)= 0.37, p=0.5466	F(1, 56)= 1.28, p=0.2628

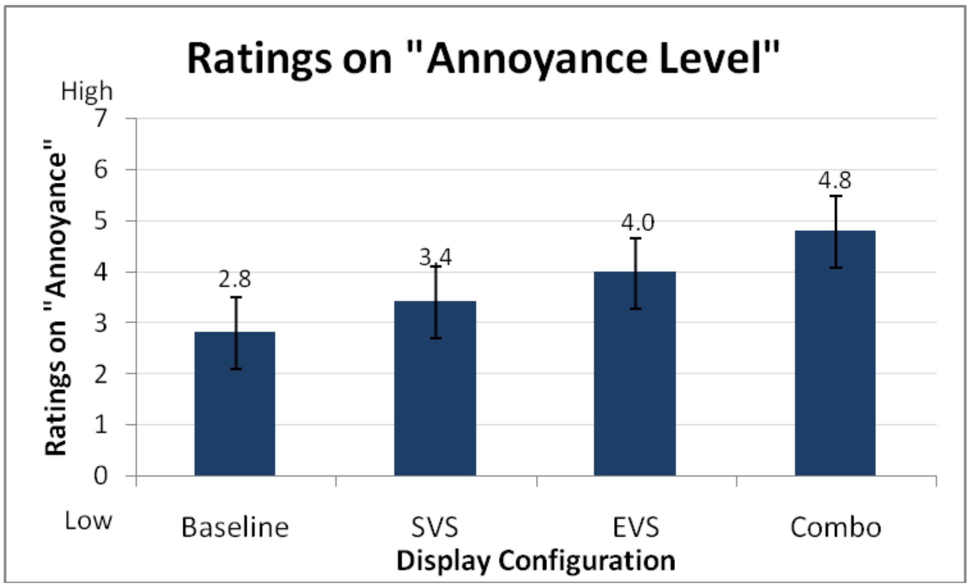


Figure 4.13. Subjective ratings on “Annoyance level” for display configuration.

Even though the ratings were not significantly different at the α=0.05 level, the means and trend of ratings were examined in order to understand how pilot feelings about the display configurations under day and night IMC condition. Table 4.6 provides the mean ratings for each question (Q1 to Q5) for the eight experimental manipulations. Across all questions, it can be noted that pilots provided lower preference ratings on the Combo display for both the IMC day and night conditions. For questions regarding display preferences to support the three types of situation awareness (Q1, Q2 and Q3), the use of SVS or EVS was generally rated comparable to or lower than the baseline configuration for the IMC-day condition; however, the same displays produced higher preferences than baseline for the IMC-night condition. In both IMC conditions, EVS was less preferred than SVS. In addition to this, between the two IMC conditions, it can be observed that the preference for use of the baseline display was higher in the IMC-day condition, while the use of the Combo configuration was preferable in the IMC-night condition. This may be attributable to the brighter image of the out-of-cockpit view in daylight for the Combo display, which had lower contrast, higher clutter and, consequently, lower discriminability of the HUD features than in the night condition.

Regarding the question on the capability of the HUD to enhance flight safety (Q4), pilots thought SVS would increase flight safety for IMC-day and EVS was rated to increase safety for IMC-night conditions. From the pilot comments, some suggested the use of colorful features (instead of monochrome images) might be helpful. In addition to this, some pilots suggested decluttering the HUD below a specific altitude by removing terrain and some iconic features. This suggestion was also supported by the CTA results on the importance of specific features for flight tasks.

Table 4.6. Mean ratings on five subjective preference questions.

	IMC-day				IMC-night			
	Baseline	**SVS**	**EVS**	**Combo**	**Baseline**	**SVS**	**EVS**	**Combo**
Q1	5.75	5.75	5.56	4.56	5.50	5.69	5.50	5.13
Q2	5.88	5.88	5.44	4.81	5.69	5.94	5.75	5.13
Q3	5.06	4.86	4.56	3.63	4.38	4.56	4.56	4.06
Q4	6.63	6.88	6.13	4.50	6.13	5.63	6.19	5.50
Q5	2.71	3.29	4.00	5.88	2.88	3.50	3.93	3.69

Q1- How useful was the HUD configuration for understanding the aircraft's position relative to the terrain and approach path? (1-not useful at all, 7-exteremly useful)

Q2- How useful was the HUD configuration for understanding current flight status (e.g., airspeed, altimeter setting, and designated heading)? (1-not useful at all, 7-exteremly useful)

Q3- How useful was the HUD configuration for using your knowledge of aircraft control, navigation, and communication (with ATC or FO)? (1-not useful at all, 7-exteremly useful)

Q4 - Do you think the HUD configuration would enhance aviation safety in low visibility conditions? (1-not at all, 7-yes)

Q5- Was the HUD configuration you used in the previous trial annoying? (1-not at all, 7-yes)

4.2. Part II: Results of CTA

As mentioned previously, the results of the CTA, based on the additional test trial with pilots thinking aloud, included three types of outcomes: lists of sequential and non-sequential tasks, and critical comments from pilots. In addition to this, patterns of pilot attention shifting about the HUD were examined.

4.2.1. Sequential events and tasks

Tables 4.7 through 4.10 provide results of the analysis of sequential events and tasks for the four legs of flight across all display configurations and IMC conditions. There were no specific differences in the sequential tasks among the experimental manipulations, suggesting the display configurations and IMC conditions do not affect sequential flight tasks. Each table consists of several hierarchical categories of events, identifying triggers for specific flows of tasks according to the events. In general, trigger events were based on the flight scenario and confirmed by observations of common pilot behaviors in the videos. Examples of trigger events include ATC broadcasts, altimeter call outs, and critical changes in flight status (e.g., intercepting glideslope or having the runway appear in sight through the out-of-cockpit view). Each event or task then consisted of an associated flow of elementary actions. The tables also include indications of the operator who handles the actions (e.g., ATC, pilot or FO) and objects required for each of the actions.

Table 4.7. Sequential events and tasks for Leg 1.

Event/Task Description	Operator	Object
1.1. Receive approach clearance from ATC.		
• At the beginning of simulated flight. • 23 DME inbound.		
ATC communication: *"Wolfpack1. Reno Approach: Abeam PYRAM, maintain 8500 ft until established, cleared ILS 16R approach. Maintain 21 0 for slower traffic. Altimeter setting is [2999 or 3003]."*	ATC	Audio
Response back to ATC: *"Reno Approach, Wolfpack1, maintain 8500 until established , cleared ILS 16 right, maintain 210. Altimeter [2999 or 30 03], we have Delta."*	FO	Verbal
Check altitude <8500 ft>.	Pilot	Altitude
Check airspeed setting <210>.	Pilot	Selected speed
Check airspeed <210 kts>.	Pilot	Airspeed
Check altimeter <designated altimeter>.	Pilot	Barometer
1.1.1. If altimeter is not correct, set Altimeter.		
• to <designated altimeter>.		
Say "*set altimeter to <designated altimeter>*."	Pilot	Verbal
Sets altimeter.	FO	Altimeter
Says "*Altimeter set.*"	FO	Verbal
Check Altimeter.	Pilot	Barometer
1.2. Receive ATC clearance to slow to approach speed.		
• 19 DME inbound.		
ATC communication: *"Wolfpack,. Reno Approach. Slow to approach speed."*	ATC	Audio
Response back to ATC: *"Reno Approach, Wolfpack 1. Slow to approach speed."*	FO	Verbal
1.2.1. Speed bug setting.		
• Set speed bug to approach speed (138kts).		
Say "*set speed bug to 138.*"	Pilot	Verbal
Sets speed bug to 138.	FO	Speed bug
Says "*speed bug set.*"	FO	Verbal
Confirm airspeed setting.	Pilot	Selected speed

Table 4.7 Continued.

1.2.2. Flap deployment.		
• According to current airspeed and pilot preference.		
Say “*flaps <1, 5, 15, 20, 25, or 30>.*”	Pilot	Verbal
Sets flaps <1, 5, 15, 20, 25, or 30>.	FO	Flap
Says “*flaps <1, 5, 15, 20, 25, or 30>.*”	FO	Verbal
Hear flaps extending.	Pilot	Audio
1.2.3. Landing gear deployment.		
• According to pilot preference.		
Say “*Gear down.*”	Pilot	Verbal
Deploys gear.	FO	Landing Gear
Says “*Gear moving.*”	FO	Verbal
Hear landing gear moving.	Pilot	Audio
Says “*Three green lights*”	FO	Verbal
1.2.4. Landing Checklist.		
• After completing gear down and flaps 30. • Should be completed before decision height (Leg3).		
Say “landing Checklist.”	Pilot	Verbal
Checks flaps, gear and speed brake.	FO	Levers
Says “flaps 30, gear down, speed brake armed. Checklist completed.”	FO	Verbal

Table 4.8. Sequential events and tasks for Leg 2.

Event/Task Description	**Operator**	**Object**
2.1. G/S Intercept.		
• 13.5 DME inbound. • Tunnel image starts curving downward.		
Watching for G/S intercept.	Pilot	G/S indicator
Confirm descending.	Pilot	Altimeter & V/S
2.2. Receiving ATC clearance, asked to contact tower.		
• 9.0 DME inbound.		
ATC Communication: *"Wolfpack1, Reno Approach. Approaching DICEY, contact to wer."*	ATC	Audio
Response back to ATC: *"Reno Approach, Wolfpack1,switching to tower."*	FO	Verbal
Switching frequency to tower.	FO	Radio
Contact tower: *"Reno Tower, Wolfpack1, DICEY inbound on ILS 16R."*	FO	Verbal
2.3. Receive landing clearance from tower.		
• 8.0 DME inbound.		
Tower communication: "*Wolfpack1, cleared to land 16 R. Wind 165 at 10 kts."*	Tower	Audio
Response back to tower: *"Wolfpack1, cleared to land."*	FO	Verbal
Confirm landing clearance: Say *"clear to land."*	Pilot	Verbal
2.4. 1000ft Call out.		
• About 6.5 DME inbound.		
Says, "1000 to go."	FO	Verbal
Hear call out.	Pilot	Audio

Table 4.9. Sequential events and tasks for Leg 3.

Event/Task Description	Operator	Object
3.1. Call out.		
• 500ft - about 5.0 DME (6020ft MSL) • 400ft – about 4.6 DME (5920ft MSL) • 300ft – about 4.3 DME (5820ft MSL) • 200ft - about 4.0 DME (5720ft MSL) • 100ft – about 3.8 DME (5620ft MSL)		
Says, *“<XXX> ft to go.*”	FO	Verbal
Hear and remember call out.	Pilot	Audio
3.2. Watch for runway in out-of-cockpit view.		
• From approximately 5.0 DME .		
Looking for runway.	Pilot	Through HUD
If runway is in sight, say “*Runway in sight.*”	Pilot	Verbal
3.3. Landing decision.		
• About 3.6 DME (5520ft MSL).		
Says “Minimums.”	FO	Verbal
If determined to land, say “*Landing*” or “*Continue.*”	Pilot	Verbal
If pilot determined to land, says “*Landing*” or “*Continue.*”	FO	Verbal

Table 4.10. Sequential events and tasks for Leg 4.

Event/Task Description	**Operator**	**Object**
4.1. Voice alert from terrain awareness warning system (TAWS).		
• “1000”, “500”, “400”, “300”, “200”, “100”, “50”, “40”, “30”, “20”, and “10” according to radio altitude.		
Hear *“<XXX>.”*	Pilot	Audio
4.2. See runway outline and glideslope reference line.		
• From 500ft above (and lower).		
Looking at actual runway and overlapped runway outline symbology.	Pilot	HUD
Looking at glideslope reference line set to 3.1 degree.	Pilot	HUD
4.3. Landing.		
• At runway.		
Hear wheel touch on runway.	Pilot	Audio
Decrease throttles.	Pilot	Throttle

4.2.2. Non-sequential tasks

While the analysis of sequential events and tasks revealed required pilot behaviors as the flight approached the airport, the analysis of non-sequential tasks revealed required pilot behaviors in controlling the aircraft, regardless of the task time. Table 4.11 shows the list of non-sequential tasks and includes seven categories of tasks with hierarchical sub-tasks. Each task description identifies elementary behaviors with information on objects needed to perform the behaviors. It should be noted that since the flight task in the experiment did not include control of rudder pedals, pilot behavior with the rudder that would normally occur was not analyzed.

Throughout the verbal protocols and probes analyses, it was revealed that some tasks have several alternative methods of performance based pilot preferences, from training or experience as well as the status of the flight. These alternatives were identified in form of options in the table. As an example, four methods were observed for determining altitude deviations (see "3.2. Determining altitude deviation" in Table 4.11): first, during Leg 1 while maintaining level flight at 8500 ft MSL, pilots could calculate the altitude deviation by comparing the current altitude in the altitude tape with the cleared altitude (8500ft MSL) in memory or selected altitude on the HUD (see 3.2.A); second, after the glideslope was intercepted (Leg 2), pilots could obtain information about altitude deviations by looking at the glideslope deviation indicator (see 3.2.B); third, pilots could identify deviations from the designated flight path by recognizing the relative position of the FPM against the center of the tunnel feature, in all legs of flight (see 3.2.C); and finally, when the flight was below 500 ft AGL and the glideslope reference line was presented in the display with the runway outline feature (see Figure 3.4), the vertical path deviation could also be obtained by looking at where/how much the FPM deviated from the glideslope reference line.

It should be noted here that the 1st task item ("1. Track the flight path marker") identifies the unit behaviors for tracking the FPM in the video using the cursor controlled with the simulator yoke. Since, in an actual flight situation, pilots are required to control the flight in order to direct the FPM to the center of tunnel or ILS signals, instead of tracking the FPM with a cursor, the actual pilot control behaviors are also presented in the table in parentheses.

Table 4.11. Non-sequential tasks.

Task Description	Object
1. Track the flight path marker (FPM).	
• This task is ongoing throughout all legs of flight in order to immerse pilots in the simulated flight. • (Pilot behaviors for an actual flight are presented in parentheses.)	
View FPM and cursor. (View FPM and tunnel.)	FPM & cursor
If cursor is on left side of FPM (if FPM is on left side of tunnel or center of LOC), turn yoke in clockwise direction.	Yoke
If cursor is on right side of FPM (if FPM is on right side of tunnel or center of LOC), turn yoke in counterclockwise direction.	Yoke
If cursor is above FPM (if FPM is above tunnel or center of G/S), push forward on yoke.	Yoke
If cursor is below FPM (if FPM is below tunnel or center of G/S), pull back on yoke.	Yoke
Confirm position of cursor (FPM) relative to FPM (tunnel).	FPM & cursor
2. Airspeed control.	
2.1. Acquire current airspeed.	
View airspeed tape and recognize current airspeed.	Airspeed tape
2.2. Determine airspeed deviation (2 options).	
2.2.A. Use of information from airspeed tape.	
• This is one of the methods for determining airspeed deviation, using the airspeed tape.	
View selected airspeed or recall target airspeed.	Selected speed
Compare selected airspeed to current airspeed.	Cognition
2.2.B. Use of airspeed error worm.	
• This is another method for determining airspeed deviation, using the airspeed error worm in the FPM group.	
View airspeed error worm at FPM.	FPM
Recognize direction and magnitude of the airspeed error.	Cognition
2.4. Adjust airspeed.	
If the current airspeed is higher than the designated speed, decrease throttles.	Throttles
If the current airspeed is lower than the designated speed, increase throttles.	Throttles
2.5 Confirm airspeed change.	
View airspeed tape or airspeed error worm.	Airspeed tape or FPM

Table 4.11. Continued.

3. Altitude control.	
3.1. Acquire current altitude.	
View altitude tape and recognize current altitude.	Altitude tape
3.2. Determine altitude deviation (4 options).	
3.2.A. Use of information from altitude tape.	
• This method is viable for Leg1, level flight.	
View selected altitude or recall designated altitude.	Cognition
Compare designated altitude to current altitude.	Cognition
3.2.B. Use of G/S indicator.	
• This method is viable for Leg 2 through Leg 4, using the G/S indicator.	
View G/S deviation indicator.	G/S indicator
Recognize position of G/S indicator relative to center of scale.	Cognition
3.2.C. Use of tunnel feature.	
• This method is viable across all legs.	
View tunnel features and FPM.	Tunnel & FPM
Recognize vertical position of FPM relative to center of tunnel.	Cognition
3.2.D. Use of glideslope reference line (GSRL).	
• This method can be used when the flight is below 500 ft AGL at which runway outline is shown in HUD.	
View glideslope reference line, which is set to 3.1 degree and the FPM.	GSRL & FPM
Recognize vertical position of the FPM relative to GSRL.	Cognition
3.3. Adjust altitude.	
If the current altitude is higher than the designated altitude, then push yoke forward.	Yoke
If the current altitude is lower than the designated altitude, then pull yoke back.	Yoke
3.4 Confirm altitude change	
View altitude tape, G/S indicator, or position of FPM relative to tunnel.	Altitude tape, G/S indicator, tunnel

Table 4.11. Continued.

4. Heading control.	
4.1. Acquire current heading (2 options).	
4.1.A. Use of heading information on top of HUD.	
View magnetic heading information on top of HUD.	Magnetic heading
4.1.B. Use of heading scale on horizon	
View horizon line including track indicator.	Track indicator
Assume current heading based on position of track indicator.	Cognition
4.2. Determine heading deviation (3 options).	
4.2.A. Use of localizer (LOC) indicator.	
View LOC deviation indicator.	LOC indicator
Recognize position of LOC indicator relative to center of scale.	Cognition
4.2.B. Use of tunnel feature.	
See tunnel features and FPM.	Tunnel & FPM
Recognize horizontal position of FPM relative to center of tunnel.	Cognition
4.2.C. Use of runway outline features	
• This method can be used when the flight is below 500 ft AGL when the runway outline is shown in HUD.	
View glideslope reference line and runway outline	Runway outline
Recognize horizontal position of FPM relative to runway outline.	Cognition
4.3 Adjust heading.	
If the current heading is on right side of designated heading, then turn Yoke counterclockwise.	Yoke
If the current heading is on left side of designated heading, then turn Yoke clockwise.	Yoke
4.4 Confirm heading change	
View LOC indicator or position of FPM relative to tunnel	LOC indicator or tunnel

Table 4.11. Continued.

5. Acquire current vertical speed (V/S) information (2 options).	
5.A. Use of V/S indicator.	
View V/S indicator.	V/S indicator
5.B. Use of pitch ladder for inferring V/S.	
View pitch ladder, horizon and FPM.	Pitch ladder
Infer current V/S based on current airspeed and vertical position of FPM between horizon and pitch ladder.	Cognition
6. Acquire DME information (2 options).	
6.A. Use of DME indicator.	
View DME indicator.	DME indicator
6.B. Use of approach plate	
• This method is useful after descending (given that the flight is on course).	
Acquire current altitude.	Altitude tape
Look at approach plate.	Approach plate
Infer current DME based on current flight's altitude.	Cognition
7. Acquire Radio altitude (3 options).	
7.A. Use of current MSL altitude.	
• This method can be used across all legs, but is preferred above 2500ft AGL.	
Acquire current altitude.	Altitude tape
Recall ground level .	Cognition
Calculate radio altitude (=current altitude – ground level).	Cognition
7.B. Use of Radio altitude indicator.	
• This method can be used when the flight is below 2500 ft AGL at which the radio altitude information is shown in lower right corner of the HUD.	
View radio altitude.	Radio altitude
7.C. Use of Radio altitude information below runway outline.	
• This method can be used when the flight is below 500 ft AGL and the runway outline is shown in the HUD.	
View radio altitude along with runway outline.	Runway outline

Based on analysis of non-sequential tasks, inferences can be made on how specific options (unit behaviors and required information and objects) affect pilot workload and SA in task performance. As an example, the task of acquiring current vertical speed involves two methods (see 5th task in Table 4.11). One method (5.A) is the use of the V/S indicator, in the lower right corner of display (see Figure 3.3), another method (5.B) is pilot inference of the current speed based on the vertical position of the FPM between the horizon and pitch ladder features (also see Figure 3.3). In the first method, a pilot can acquire a precise value of vertical speed but they may lose awareness of position information from the FPM and experience increased workload due to attention shifts between the V/S indicator and the FPM. In contrast, while the second method may allow a pilot to focus his attention on or around the FPM so that he does not lose awareness of the aircraft position, this method may not provide a precise V/S value, and requires relatively complex cognitive activity by the pilot.

Like the results on the sequential tasks analysis, the non-sequential behaviors among display configurations and IMC conditions did not reveal differences in pilot performance due to the non-iconic features, such as the terrain imagery, because most tasks in the analysis required information from HUD iconic features. However, it was observed that pilot behaviors varied by the leg of flight. In the Table 4.11, the 7th task allowed for three distinctive methods for acquiring radio altitude. The first method (see "7.A. Use of current MSL altitude") can be used when the flight is over 2500ft AGL (in general, during Leg 1 and 2 in the scenario) and pilots calculated the radio altitude through internal cognitive activity using the values of current MSL altitude and the ground elevation. Once the flight was below 2500 ft AGL (after Leg 3), the value of the radio altitude was shown in the lower right corner of the HUD (see Figure 3.4) and pilots could assess the value by shifting their attention (see "7.B. Use of Radio altitude indicator"). Finally, when the flight approached the runway (below 500ft AGL) and the runway outline was shown in the HUD, the value of the radio altitude displayed on below the runway outline feature (also see Figure 3.4 or look ahead to Figure 4.15). The pilot could use this display along with other critical information (FPM, runway outline, and glideslope reference line) without significant attention shifting. This indicates that the display was designed to present pilots with information (in specific locations) according to the relevance of the information in different legs of flight. This yielded different pilot internal/external behaviors.

Consequently, the CTA results can be used to analyze "what" and "how" information presented in the HUD affects pilot behaviors in a detailed manner. In addition to this, this analysis approach can also be used to evaluate and design related cockpit display interfaces.

4.2.3. Critical comments from pilots

From the analysis of the think aloud and probing, pilot critical comments were identified in order to interpret the experimental results and to gain further insights on the use of the features in advanced cockpit displays. Table 4.12 shows the summary of the critical comments from pilots. In the table, the comments are categorized into meaningful groups with sub-groups. Pilot numbers are listed for each comment. Information about which experimental manipulation was assigned to each pilot is also presented at the end of the table.

Table 4.12. Critical comments from pilots.

1. Features not aware.
1.1. Ground speed.
– "It is not relevant to me." (P1, P5, P7).
1.2. Nose of flight.
– "I'm not concerned about the nose because it doesn't matter and the focus is on the FPM." (P4).
1.3. Vertical speed, heading, and DME.
– "I don't look at the vertical speed and heading very much because I know the tunnel features keep me on course." (P1). – "I don't look up for heading because it is too far away from the FPM." (P6). – "Since there is LOC, I don't care about the heading as much." (P7). – "I check it seldom and in terms of safety, those are not important indicators." (P7).
2. Features creating workload and annoyance (except tracking task).
2.1. Thermal features (moisture) from EVS.
– "All the crap (moisture from EVS) coming at me and the tunnel is annoying. The tunnel creates clutter on the screen." (P3). – "The clouds are annoying." (P4). – "The thermal features are very distracting and they create workload in a bad way." (P7).
2.2. Wireframe features from SVS.
– "The grid (wireframe) creates a lot of workload. It seems to overload me. It's quite confusing with the tunnel." (P6).

Table 4.12. Continued.

3. Use of ILS (G/S and LOC) vs. Tunnel for keeping the aircraft on path.
3.1. Preference for the Tunnel over ILS aviation indicators.
– "Since there is a tunnel, the localizer and G/S indicator are not important." (P1). – "I always fly with the tunnel. The G/S indicator is too indistinct and is too close to the altitude tape." (P2). – "I'm not aware of the G/S and LOC indicators much. I see them, but I'm not paying attention to what they are telling me. I'm using the tunnel and airspeed." (P6).
3.2. Preference for ILS deviation indicators over Tunnel.
– "I don't even see the tunnel. I don't look at it because I have all the information I need from the FPM, G/S and LOC. I don't need the tunnel at all." (P3). – "From my past experience, I crosscheck the G/S and LOC more than the tunnel." (P5).
3.3. Use of both features (Tunnel and ILS).
– "I'm using it (G/S and LOC) very much. The tunnel is good and the FPM is telling me where I am. I can double check across features. The tunnel is a nice back-up to your G/S and LOC indicators." (P4). – "I see those (G/S and LOC) pretty regularly to make sure that the tunnel is not lying to me and make sure the raw data is there." (P7). – "I'm in the boxes (tunnel) and I back-up with seeing the G/S and LOC." (P8).
3.4. Suggestions on tunnel features.
– "Tunnel would be nice if you have a turn and you can see the turn coming ahead. On a straight-in glideslope and localizer like this, you don't need it. It is just more distraction." (P3). – "It's nice to have the tunnel. But it would be more useful for a curved approach." (P4). – "I find the graphics to be confusing. I think the tunnel should be a different color." (P6).

Table 4.12. Continued.

4. Use of terrain features.
4.1. General criticisms on terrain features.
− "I don't care about terrain features because I know where it is via the approach plate." (P4). − "Even if there were changes in the terrain, I would have difficulty in noticing them." (P6). − "It is fairly irrelevant. It is not important to me as long as I have G/S and LOC armed and I know the G/S and LOC were tested and safe. I trust G/S and LOC. They are criteria for a safe approach." (P7). − "I shift my attention to understand the terrain features, but simply, because I know the boxes (tunnel) are going to keep me away from the terrain. I don't need to worry much about collision into terrain." (P8). − "The moisture is a kind of distraction. From the wireframe, I see there is raised feature around at 12 o'clock." (P8).
4.2. Criticisms on wireframe features (SVS).
− "I'm not aware them (terrain using wireframe) because it is just a bunch of lines. If it looks like mountains, I would worry about it." (P4). − "Lines (wireframe) are not helpful because they look like other tunnels." (P4).
4.3. Criticisms on thermal features (EVS).
− "I do not like the moisture feature at all." (P6). − "The thermal image distorts the HUD, and washes out the symbology and FPM. I couldn't see it at the moment. I don't like it." (P7). − "Moisture feature makes me confused. It messes up a little bit." (P8).
4.4. Suggestion on terrain feature
− "I would put in color terrain features. For example, red for rising." (P6). − "They (terrain features) should have an indicator on terrain to reveal its status because I don't understand it. I can't tell how high it is." (P6).
4.5. Terrain features need to disappear as a flight approaches the runway
− "This terrain image (wireframe) should go away (below minimums). They are doing me no good at all. (I'd like to suggest) wiping them out completely." (P2). − At 200 ft AGL, "It (HUD including EVS in day) is horrible. Now I would turn everything off." (P3). − "When I start to get in the runway, then I start to focus on the runway and I tend to ignore other features. Eliminate the grid terrain." (P6). − "Now this is becoming very cluttered. Everything is kind an overlapped." (P8).

Table 4.12. Continued.

5. Others	
5.1. Glideslope reference line	
− "It is useful. That's what I am used to." (P1). − "I like it because it is a good cue. And the rising ground is a good cue." (P5).	
5.2. EVS features cause pilots to concentrate on the FPM	
− "I'm trying to keep myself focused on the FPM whenever I get those clouds coming through, because it is hard to see and make out where it is" (P3).	
Pilot assignment	
P1- Baseline, IMC-day	P5- Baseline, IMC-night
P2- SVS, IMC-day	P6- SVS, IMC-night
P3- EVS, IMC-day	P7- EVS, IMC-night
P4- Combo, IMC-day	P8- Combo, IMC-night

In general, pilots were not aware of the information on ground speed and the nose of the aircraft because they were not considered relevant to the flight (1.1 and 1.2). In addition to this, some pilots did not use the vertical speed, heading and DME indicators due to reliance on the suggested flight path (ILS or tunnel). The locations of this information also required significant attention shifts, and was considered irrelevance to flight safety by some (1.3).

Regarding display features, which created workload for pilots, most pilots said the tracking task was the dominant factor in workload. Other than the tracking task, the moisture images from the EVS (2.1) and grid lines as part of the SVS (2.2) were identified as generating workload because of distraction and confusion with tunnel features, respectively.

In order to keep the flight on the approach path by manipulating vertical (altitude) and horizontal (heading) controls, pilots revealed three types of preferences, including using the tunnel features (3.1), using the ILS indicators (3.2), or using both features (3.3) to double check each other. These individual preferences may be attributable to training, personality and/or reliance on technology (e.g., presentation of tunnel features based on a database with the GPS system). This variability was reflected in the CTA results of non-sequential task analysis (Table 4.11) in terms of different methods for achieving flight tasks. However, some pilots suggested that the tunnel feature would be more useful for a curved approach than for a straight approach because it can cause distractions to pilots (3.4).

Related to the use of terrain features, in general, most pilots were not satisfied with the SVS and/or EVS, though the SVS was considered to be slightly more useful than the EVS and Combo conditions. Pilots believed that keeping the flight on the approach path with either the tunnel or ILS indicators would prevent collisions with terrain (4.1). In addition, grid line features as part of the SVS were not considered realistic and produced visual confusion with the tunnel features (4.2). Furthermore, thermal features from the EVS, in particular moisture imagery, were distracting and washed-out critical information on the HUD (FPM and other indicators) (4.3). Pilot suggested using color coding and additional information to indicate the height of terrain ahead (4.4). In addition to this, pilots also suggested that all terrain features should be turned off at specific flight positions, as the aircraft approached the runway (e.g., runway in sight, landing decision, or 500ft AGL) because the features produced serious clutter effects when overlapped with the visible runway in the out-of-cockpit view (4.5).

Regarding the other comments, pilots thought the presence of the glideslope reference line display (when the flight was below 500 ft AGL) was useful (6.1). Interestingly, EVS thermal features were cited as causing pilots to focus on the major task (tracking the FPM) (6.2).

4.2.4. Patterns of pilot attention shifting

Although there were individual differences among pilots, the general profiles of pilot attention to the HUDs were captured from the think aloud and probing techniques during verbal protocol trials. The majority of pilots focused on center of the display (FPM and tunnel features) 60-90% of the time, followed by the airspeed 10-30% of the time, and the altitude 10-30% of the time. Most pilots did not pay significant attention to the other features. The typical order of attention shifting was FPM and tunnel, airspeed, back to FPM and tunnel, and then to altitude. When shifting attention across the display, pilots quickly verified ILS indicators (localizer and glideslope) based on the position of the FPM against the tunnel features and then glanced at DME, vertical speed, and terrain features. It was observed that pilots tended to perceive the FPM and tunnel features as a single group when the FPM was inside the tunnel. Figure 4.14 illustrates the pattern of typical pilot attention shifting.

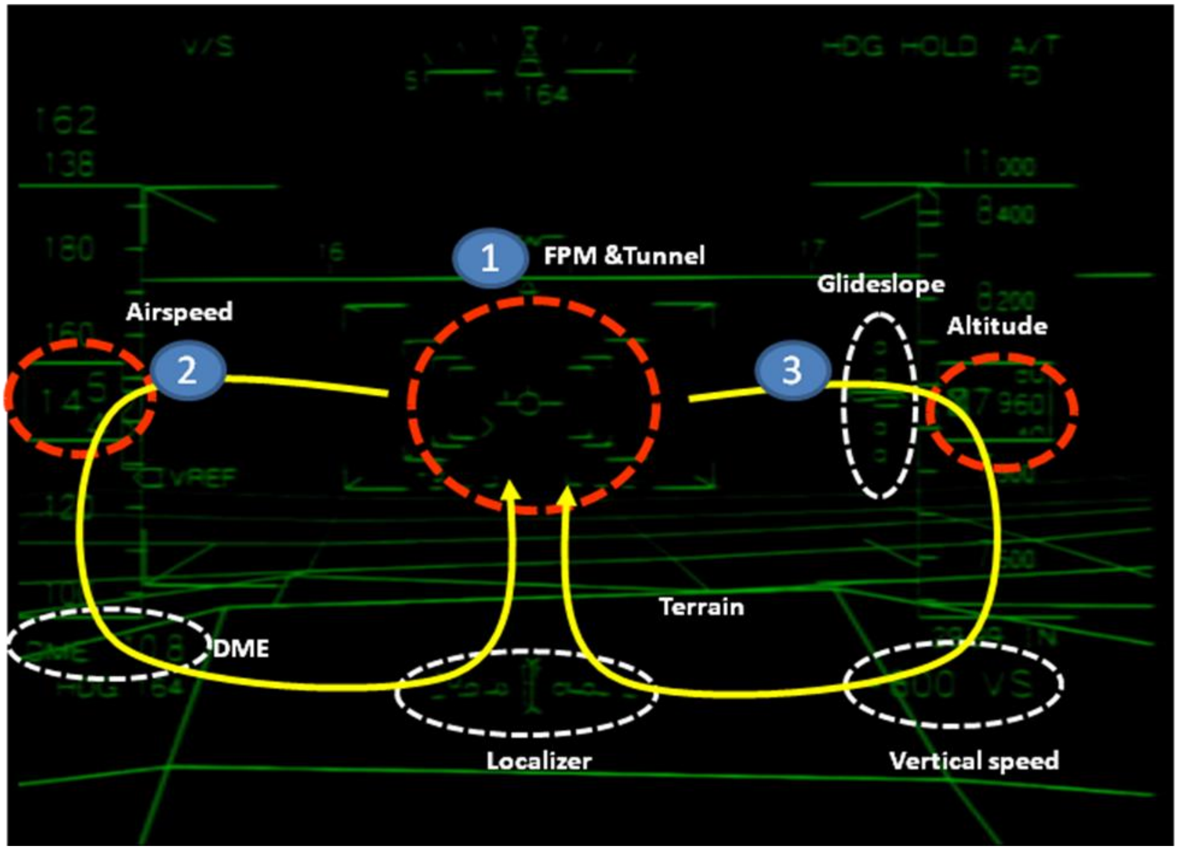


Figure 4.14. Pattern of typical pilot attention shifting.

When the flight was flying below 500 ft AGL, the tunnel features disappeared and the runway outline feature was shown with the glideslope reference line (and radio altitude). The pattern of typical pilot attention shifting was changed. Pilots mostly focused on the center of display, including the FPM, runway outline, glideslope guideline, and radio altitude while checking airspeed on the left side of the display. Figure 4.15 depicts the pattern of typical pilot attention shifting when the flight was below 500 ft AGL.

Although the patterns of pilot visual attention were determined based on verbal protocol results rather than a more objective measure (e.g., eye-tracking), results revealed general pilot behaviors in perceiving flight information on the displays. In addition to this, an expert pilot, who had experience with the advanced HUD features used in this study, reviewed the results and confirmed that the attention pattern results were representative of scanning during real flight operations.

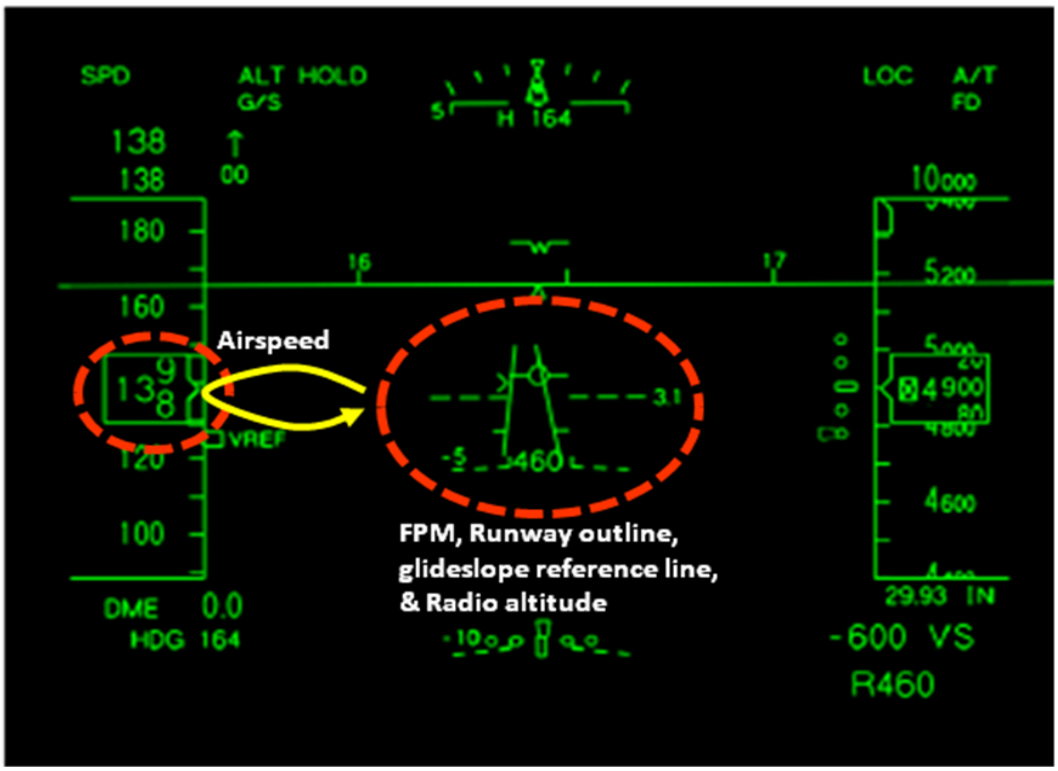


Figure 4.15. Pattern of typical pilot attention shifting below 500 ft AGL.

5. DISCUSSION

5.1. Flight path control

Errors in the tracking task were evaluated as a measure of flight path control performance. It was assumed that higher RMSEs would be associated with less attention to the FPM indicative of degraded path control performance. Flight path control performance was affected by display configuration and visibility condition (IMC) and two interaction effects (display by IMC condition as well as display by leg). These results were in-line with Hypothesis 1, which was a general expectation for the HUD content to drive variation in flight performance under the various environment conditions.

In general, EVS generated lower RMSEs, the Baseline and Combo condition were comparable to each other, and SVS induced the greatest flight path control. This was not in-line with Hypothesis 1-1, which stated that path deviation errors to be lowest with the Baseline display and then increase with the SVS, EVS, and Combo display. These findings can be explained by pilot comments during the "think aloud" session. First, the grid lines depicting terrain features, generated by the SVS were often confused by pilots with the FPM and the tunnel features, which also consisted of lines. This confusion may have diverted pilot attention from the FPM in order to discriminate other features from the SVS imagery and caused higher tracking errors. Second, the thermal returns (e.g., moisture images) from the EVS did not appear to produce pilot confusion. Interestingly, EVS imagery appeared to compel pilots to focus more on the FPM in the display and this produced lower tracking task error. Finally, since the SVS increased

RMSEs and the EVS decreased RMSEs, the Combo was comparable to the Baseline display configuration, which did not include any terrain features.

Between the two visibility conditions (IMCs), IMC-day was associated with higher RMSEs than IMC-night. This result was in agreement with Hypothesis 1-2, suggesting that low saliency of the dynamic symbology (FPM) against the high brightness background (daylight) degraded tracking performance. It is possible that pilot confusion from the use of the SVS was further magnified with low saliency of the FPM during daytime conditions in dense cloud cover.

Although Hypothesis 1-3 was not supported, as no differences in RMSEs were observed among the four legs of flight across display conditions and IMCs, the interaction of leg and display was significant. In general, the use of SVS induced higher RMSEs in Legs 1, 2 and 3, and the Combo produced higher RMSEs in Leg 4, which was the final landing leg (see Figure 4.4). These findings can also be explained by pilot comments. That is, the use of both SVS and EVS features with actual terrain (or runway) features visible in the out-of-cockpit view may have caused higher display clutter than other display configurations. During Leg 4, this clutter effect may have distracted pilots from focusing on flight control to the runway. Patterns of typical pilot attention shifting also support this inference. The patterns of attention in the HUD was changed during Leg 4 (see Figures 4.14 and 4.15) after the tunnel features disappeared and the runway outline presented to pilots. When the aircraft was flying under 500ft AGL, pilots highly focused on the center of the HUD, including FPM, runway outline, glideslope reference line, and radio altitude, rather than other system information across the display. The use of terrain features in HUD (SVS and EVS) caused distractions to pilot for concentrating the features around the FPM. With this reason, pilots suggested removing any terrain features at specific points in the flight, including landing, provided the runway was visible in the out-of-cockpit view.

5.2. Pilot SA

Pilot SA measured using SAGAT (including overall SA, levels of SA, and types of SA) was affected by the experiment manipulations. Even though this was expected (H2), there were several results that contradicted the detailed hypotheses.

Regarding HUD configuration, overall SA scores were higher for SVS use and lower for EVS use, while the effect of each of these configurations was not different from the Baseline and Combo conditions. In specific, the use of EVS degraded system awareness (see Figure 3), which concerned pilot understanding of iconic information in the display (e.g., airspeed, altitude, DME and vertical speed). Decrements in SA while using the EVS may be attributed to: (1) the thermal features frequently washing-out iconic features, presenting system information; and (2) pilot focus on the FPM to perform the tracking task to the neglect of attending to system information. Therefore, pilots using EVS features produced higher tracking performance (lower RMSEs) but had lower system awareness. This suggests a cognitive tunneling effect due to the presence of thermal features from the EVS in the HUD.

The visibility conditions were found to have no effect on overall SA scores; however, there were significant effects on specific levels and types of SA. That said, some results were contradictory. While IMC-night was associated with higher level 2 SA scores (in-line with H2-2), it induced lower system awareness scores (not in-line with H2-2). This suggested that night flying increased pilot comprehension of overall flight with decreased understanding of system information in the HUD. In general, mean scores for other levels and types of SA across display configurations and legs were higher in the IMC-night condition.

Legs of flight were found to affect pilot SA, although the pattern was not the same as hypothesized (H2-3). Legs 1 and 3 were associated with higher overall SA and level 3 SA than Legs 2 and 4. Results on the three types of pilot SA provided more detailed information on the effects of the legs of flight. As shown in Figure 4.10, spatial awareness was higher in Legs 3 and 4 than Legs 1 and 2. These results suggested that pilots gained more understanding of spatial information during Legs 3 and 4 since the flight was below the "ceiling" and the actual terrain and runway were visible through the out-of-cockpit view, instead of clouds or darkness. SA for system awareness was higher in Leg 1. This may have been due to more tasks requiring pilots to aircraft flight parameter using iconic features in the HUD. Pilots had to slow the aircraft to approach speed (210 kts to 138 kts) while checking the airspeed and controlling the throttles, frequently. As shown in Table 4.7 (the list of sequential events and tasks for Leg 1), pilots also manipulated the flaps and landing gear controls, depending upon airspeed and DME information (Keller et al., 2003), with frequent confirmation of this information. Therefore, the importance of the system information caused pilots to achieve high levels of awareness in Leg 1. SA scores for task awareness by legs were related to the number of tasks occurring in each leg. That is, pilot awareness about communications and landing preparations (flaps, landing gear, and landing checklist) decreased as the flight approached on the runway in order pilot to concentrate on flight maneuver for landing. This trend agreed with the findings of the CTA results. In specific, the number of sequential events and tasks decreased as the flight progressed (see Tables 4.8 – 4.10).

There were no significant interaction effects on overall SA scores; however, there were significant effects on several levels and types of pilot SA. First, display configuration caused differences in system awareness among the different legs of flight. While the use of SVS produced higher system awareness in Legs 1 and 2, the Baseline configuration yielded higher system awareness for Legs 3 and 4, where the actual terrain and runway could be seen through the out-of-cockpit view (see Figure 4.11). This finding may be attributable to display clutter effects as terrain features overlapped the out-of-cockpit view degrading pilot understanding of system information in the display. Second, although IMC-night produced higher SA scores than IMC-day, IMC-day was associated with higher level 3 SA and system awareness (see Figures 4.7 and 4.12, respectively). This may be due to the level of visibility of the out-of-cockpit view. The out-of-cockpit views (especially, the runway) for Leg 3 were clearer in IMC-day than IMC-night because of daylight, while the views of the runway for Leg 4 were not significantly different between the two IMC conditions. It is possible that features visible in the out-of-cockpit view yielded higher pilot system awareness for Leg 3 in this study.

Among the three types of pilot SA, spatial awareness and task awareness were not affected by display configuration and visibility condition, while the effects of most experimental manipulations were significant on system awareness. Only the segment of flight was found to effect spatial awareness due to visibility of actual terrain images through the out-of-cockpit view. Segment also affected task awareness due to the number of tasks differing across legs. This suggests that the use of non-iconic conformal features (SVS and/or EVS) used in this study may be more effective for facilitating SA on system status information as compared to providing pilots with spatial information as compared to.

5.3. Pilot workload

Both subjective and physiological pilot workload measures (using the NASA-TLX and heart rate, respectively) were not affected by the experimental manipulations, including display configurations, IMC conditions and legs of flight. These findings were not in-line with Hypothesis 3 and its sub-hypotheses (H3-1, H3-2 and H3-3). In addition to this, while a previous study (Kaber et al., 2008) found subjective workload ratings measured by NASA-TLX to be affected by levels of display clutter, the present study could not reveal the impact of display conditions on pilot workload. This unexpected result may be attributed to constraints on the experiment and nature of the task. That is, the major task required of the pilots was not actual flight control, but a tracking task on the FPM. As mentioned in the CTA results section, pilots in this experiment commented that the major factor generating workload was the tracking task. The difficulty of the tracking task among the video stimuli did not differ (see section 4.1). It is likely that the two workload measures did not vary for this reason and tracking task involved limited cognition (vs. perception).

5.4. Pilot subjective preferences for displays

Even though non-parametric analyses revealed HUD configuration in ratings for questions regarding SA and flight safety (Q1-Q4) to be statistically insignificant, in general, the mean values of the ratings matched with expectation (H4-1A). This suggests that pilots felt the SVS and EVS technology would increase spatial, system and task awareness and flight safety, while they did not prefer the use of the combination display. The ratings on questions for annoyance level were significantly different for display conditions, indicating the more features in HUD, the greater the level of pilot frustration. This finding was in-line with Hypothesis 4-1B.

Hypothesis 4.1A was based on expectation of an effect of display on SA (H2-1). Hypothesis 4-1B was based on expectation of an effect of display on flight path control performance (H1-1). Results on the objective response measures did not validate H1-1 or H2-1; however, subjective preference ratings did (H4-1A and H4-1B). This suggests that there were differences between pilot perceived utility of displays and actual performance measured with objective measures in pilot use of the displays. Therefore, it is important for empirical studies like this to use both objective measures and subjective surveys evaluating display designs in for these kinds of domains with cognitively complex tasks.

5.5. CTA results

Analyses on the contents of the think alouds and probes produced three types of results. Among the results, the analysis on sequential events and tasks was used to explain the effects of legs of flight on pilot SA. That is, the number of tasks and amount of information for each leg of flight affected system awareness and task awareness. The non-sequential task analysis demonstrated how pilots control the aircraft, beyond sequential events, and which alternative behaviors can be employed to perform the specific task along with the relevant information. Since the two lists of pilots tasks were analyzed at a high level, the effects of display configurations or IMC conditions could not be explained in a detailed manner. However, pilot comments during the think-aloud session provided useful information for interpreting the effect of the display configurations on performance. For example, the use of SVS terrain features produced confusion with the tunnel and FPM features; EVS caused a cognitive tunneling effect that compelled pilots to concentrate on the FPM instead of iconic information; and use of terrain features should not be considered after the runway is visible in the out-of-cockpit view because of the potential to create display clutter effects.

6. CONCLUSION

6.1. Outcomes related to objectives

The objectives of this study were to: (1) assess the effects of non-iconic conformal features (SVS and/or EVS) in an advanced HUD on line pilot response (flight path control, SA, workload and preferences), across distinct legs of an approach and landing scenario; (2) measure pilot SA using an elaborate SAGAT methodology in order to identify the effects on pilot SA in detail (by levels and types of SA); and (3) conduct a high-level cognitive task analysis to explain experimental results and describe pilot behaviors using the advanced HUD.

6.1.1. The effects of an advanced HUD features on pilot performance

In this study, the videos of HUD content for approach to runway 16R at KRNO captured using the IFD simulator at NASA LaRC, were prepared and then presented to pilots in a lab simulator. Pilots were asked to perform a tracking task to trace the FPM in the video stimuli, while following a flight scenario and understanding the flight situation. Quantitative descriptions of the impacts of the advanced HUD configurations (SVS, EVS, and combined SVS and EVS), flight visibility conditions (IMC-day versus IMC-night) and distinct legs of flight on pilot responses, including flight path control, SA, workload and subjective preferences were developed. Since several prior studies confirmed that the use of tunnel features improved pilot performance, the separate and combined effects of non-iconic conformal features (SVS and EVS) were examined when all HUD configurations included tunnel features. In general, the SVS increased overall SA but degraded flight path control (shown by tracking errors) due to visual confusion from SVS grid lines with the tunnel. The EVS condition increased flight performance; however, there were decrements in SA, especially understanding of system information, due to distractions from thermal images. Once real terrain and the runway are visible through the out-of-cockpit view (technically, VMC), pilot awareness on the spatial situation was increased, but the display provided clutter effects, leading to a degradation of flight path control.

Contrary to the original purpose of SVS and EVS technology, the non-iconic conformal terrain features examined in this study were not proven to facilitate increased pilot understanding of terrain/spatial information. In addition to this, while the tunnel features and ILS information were displayed in the HUD, pilots did not regard the terrain features as critical information in terms of flight safety. This is supported by the CTA results (see section 4.2.3) and not surprising, as Schnell et al. (2004) study also demonstrated that pilots relied on and trusted the tunnel to the extent that they did not feel the need to devote much attention to the aircraft-terrain situation.

6.1.2. The use of an elaborate SAGAT methodology for assessing pilot SA

In order to assess pilot SA in the experiment, a more elaborate SA measurement technique was used. SAGAT was extended to include SA queries covering three levels of SA (according to Endsley (1995a, b)) as well as three types of pilot SA (according to Wickens (2002)). This approach provided useful and detailed information for explaining independent variable effects on pilot SA.

SART and SA-SWORD, measurement techniques used in many previous studies for evaluating advanced cockpit interfaces, have been noted to suffer from participant bias in SA ratings. In use of SART or SA-SWORD, participants rating may be affected by impression of their preference from performance or workload in a trial. In the present study, patterns of subjective preferences for displays, in terms of situation awareness (Q1, 2 and 3 in post-trial questionnaire), were in-line with Hypothesis 4-1A which was based on the results of previous studies using SART or SA-SWORD. Therefore, it can be inferred that the previous findings on SA using SART and SA-SWORD may not present actual pilot SA but pilot preferences. The empirical results on pilot SA in this study using the extended SAGAT methodology did not match with pilot preference ratings. The SAGAT results revealed the effects of experimental conditions to be variable for levels and types of pilot SA. Consequently, the development and use of SAGAT for measuring SA at different levels and along different types (in aviation tasks) can serve as a novel and potentially useful framework for SA measure development.

6.1.3. The use of CTA for aviation applications

From the CTA, including the verbal protocol and probe responses collected during the final test trial, descriptive explanations of the experimental results in terms of pilot cognitive behaviors were produced. That is, results of the CTA facilitated the interpretation of how the non-iconic terrain features included in the HUD affected pilot behaviors in terms of cognitive activity. Previous research efforts have focused only on revealing whether display alternative “A” is better than “B”, or vice versa. In particular, the CTA results provided information including: practical comments to interpret specific effects of HUD features; relations between pilot tasks and situation awareness; and several alternative methods for achieving a task with different information or objects in a cockpit interface.

6.4. Caveats

There are some limitations in generalizability of this study that should be noted with respect to using the results as a basis for designing or making decisions on optimal advanced HUD features. The caveats of this study include the use of a low-fidelity flight simulator for the experimental testing. Pre-recorded videos of HUD content were played for line pilots using a PC-based simulator. Pilots were asked to perform tracking of a FPM in the video, instead of performing actual flight control. Flight path control performance was assessed strictly based on deviations in the tracking task. In an actual flight situation, pilots might demonstrate different behaviors compared to the findings in the present study. In addition to this, the use of the tracking task did not reveal differences in pilot workload among the experimental conditions. This may have been due to the fact that the tracking task was not cognitively complex. Also, the use of rudder pedals was not allowed and not analyzed by the CTA because of the limited simulator fidelity while the pedals can be critical controls in actual flight. Another limitation due to the use of pre-recorded videos was the restriction on airspeed control. Since the airspeed in the video stimuli could not be controlled by test pilots in the lab simulator, and deviations among power outputs from the throttles and displayed speed were not captured, pilot performance in airspeed control was not examined in the present study. In real flight, or when using a high-fidelity simulator (e.g., IFD), airspeed control performance may be an important indicator of the impact of specific cockpit interfaces.

Related to this, the syntheses of the pre-recorded videos from the IFD and X-plane simulator produced another limitation. Specifically, there were distortions in the prepared video stimuli compared to the original display properties. For example, video post-processing might have increased the intensity of thermal images from the EVS in the displays, in particular, for IMC-day conditions. An expert pilot who is familiar with the HUD implemented in IFD simulator also pointed-out that EVS features in the synthesized videos stood-out more than in the actual images in the IFD simulator. However, the expert pilot commented that the differences in feature intensity might be negligible in terms of pilot performance.

The simulator setup also limited pilot information sources. Only the information presented on the HUD was used for the simulated flight. In a real flight context, pilots use not only HUD information but many other displays on the cockpit panel or head-down areas, including the PFD, ND, FCP (flight control panel), or FMS (flight management system). Pilots integrate various information across these displays during a flight, to maintain SA. This may result in different pilot performance than the results in this experiment.

Although SAGAT was used as an objective measure of pilot SA, the technique posed some limitations in terms of producing a high-resolution or real-time assessment of pilot SA. Since only one SAGAT freeze occurred in each of the four legs during a trial in this experiment, the method may not have revealed short-time span changes in the SA profile. Pilot perception, comprehension and projection may vary dynamically during a flight. The low resolution of the SAGAT freezes might have produced a limited representation of display effects on pilot SA. In order to address this resolution issue, more frequent simulation freezes for SAGAT queries or real-time probes for SA measurement (Jones & Endsely, 2004) could be used in future studies.

Finally, the flight scenario used in this experiment did not involve critical events that can happen in real flight or provide the opportunity to assess the utility of terrain features in such situation. For example, the effects of SVS/EVS features may be different in flight under off-nominal conditions such as when a pilot must go-around at decision height or there is a runway incursion. Moreover, even though the approach and landing at KRNO 16R was selected as an extreme scenario from the terrain perspective, there were no significant challenges from a flight path control perspective due to the existence of the ILS and tunnel features.

In addition to this, the scenario used in this study and the findings are most relevant to dual-piloted commercial aircraft in which pilots and FOs share the flight tasks. The sample of test pilots used in this study included eight line pilots that were highly experienced in flying commercial aircraft (including the use of “glass” cockpit displays). Some pilots had experience in the use of basic or advanced HUD features. Therefore, the results of the study are most generalizable to commercial flight operations by high experience pilots. If novice pilots or pilots of other types of aircraft (e.g., fighters or helicopters) are involved in experiments or actual flights, there may be different consequences in terms of pilot performance.

Even though limitations of the findings were attributable to use of a low-fidelity simulator compared to the IFD, the present study may generalize to actual commercial flight operations for several reasons: (1) prior research (Prinzel et al., 2002) has indicated that finding from lab simulations were comparable to results from actual tests with real aircraft; (2) lab simulator and video stimuli in the present study were designed to closely approximate use of the IFD simulator; and (3) the present study used a relatively high-fidelity lab simulator compared to equipment used by others (e.g., the Bolton et al. (2007) study presented short video clips of terrain textures to participants without requiring any flight control or producing any out-of-cockpit view).

6.5. Future research directions

On the basis of this study, directions of future research include investigating advanced HUD design and the use of non-iconic conformal features for flight safety under off-nominal conditions. First, the effects of SVS and EVS features under various flight situations should be further evaluated. In the present experiment, there were effects of HUD features in specific segments of flight (an approach and landing) that led to differences in pilot performance. HUD features may also influence performance under situations, like runway-incursions. In addition to this, pilots commented that the tunnel features would be more useful on a curved approach prior to intercepting the ILS signal. Therefore, the effects of tunnel features with terrain images might differ with the shape of the approach, and it would be worthwhile to investigate this possibility.

Second, this type of research should be conducted in a more realistic simulator or real aircraft and in a more realistic flight context. This may facilitate the assessment of real flight control performance instead of RMSE from a tracking task. The use of high-fidelity simulators or actual aircraft may also allow pilots to use a full suite of displays including a HUD and other HDDs, as well as experience communication and coordination with a co-pilot.

Regarding the design of advanced HUDs, more studies should be conducted in order to determine the optimal features to include. Although previous research efforts revealed the optimal format for SVS, tunnel and FPM group features in a HUD to be a grid wireframe (Snow & Resign, 1999), "dynamic crow's feet" and "tadpole" (Prinzel et al., 2004b, 2004c), respectively, the present study demonstrated the combination of such features in the same display to generate visual confusion for pilots. Therefore, it would be interesting to investigate the effects of these features with a variety of visual display properties including different brightness, line widths, and contrast. As pilots commented, the utility of color in the HUD might be of considerable value, if technology can support it.

Based on empirical studies, including this research, it would also be interesting to develop a flight context-based "Adaptive HUD" interface. Although it may be possible for pilots to manipulate display configurations in order to achieve most appropriate feature combination or display properties (e.g., brightness and contrast) under particular flight conditions, this may cause additional cognitive workload for pilots. With this in mind, it would be desirable to present pilots with pre-determined optimal display feature sets according to dynamic changes in a flight situation. For example, the HUD interface used in this study presents pilot with a runway outline and glideslope reference line instead of tunnel features when the flight is below 500feet AGL. This display transition, based on altitude, was preferred by most pilots. Another example of an adaptive HUD may be the concept of a “fusion” display investigated by Bailey et al. (2006, 2007), which gradually transitions from SVS to EVS based on altitude. Based on findings in the present study, several approaches for an adaptive HUD can be suggested as follows: EVS could be automatically turned off when a sensor detects significant moisture features, in order to avoid pilot distraction; all terrain features could be removed when a flight approaches on the runway and the out-of-cockpit view of the runway becomes clear (especially, in daylight); and brightness and contrast of each feature in the HUD could be automatically adjusted based on ambient light sensor readings in order to provide pilots with the best readability of display information.

The use of an elaborate SAGAT method for pilot SA measurement is worthy of being applied in other research after modifications. This novel approach assessed pilot SA on three levels of SA and in terms of three types (spatial awareness, system awareness, and task awareness) and provided valuable detailed information for interpreting pilot performance. However, the SA queries used in this study were limited to understanding of a restricted flight situation due to the constraints of the simulator. If this approach is applied in other avionics research measuring pilot SA, queries should be extended to cover all aspects of pilot SA, as identified by Wickens (2002). It might be expected that the technique would provide detailed explanations of effects on pilot SA, as was obtained in this study.

Finally, while an experimental study requires considerable time and cost in using human subjects, development and use of cognitive models of pilot behavior for assessing interface would be interesting. The CTA results can serve as a basis for developing a cognitive model. Such a model could provide a basis for optimizing advanced HUD design rather than expensive and time consuming pairwise comparisons of various display technologies. Zachary et al. (2000) and Kieras (1997a) said that CTA is critical for developing cognitive models for cognitively demanding tasks. With this in mind, results could be used for GOMS modeling (Goal, Operator, Method, and Selection of rules (Card et al, 1983)) and, further, computational cognitive models (e.g., GOMSL (Kieras, 1999)). For example, the hierarchical task items identified in Tables 4.7 through 4.11 could be translated into "Goals" and "Methods" for accomplishing the goals in a GOMS model. Unit behaviors for each task could be regarded as "Operators" and several options (shown in Table 4.11) for specific tasks could be directly converted into "Selection rule" statements in a cognitive model. In addition to this, the flow of sequential events and tasks may dictate the flow of the cognitive model of pilot behavior. The sequential and non-sequential tasks might be implemented in a single model in form of threads of cognition, based on Salvucci and Taatgen (2008) theory of concurrent multitasking. However, since the results of the CTA in the present study and GOMS modeling are at high-levels of specificity for describing human behaviors, the use of this approach may not be feasible for describing the effects of non-iconic features in HUD.

This may require the description of low-level behaviors, such as the profile of pilot attention shifting based on visual display properties. Therefore, in order to develop applicable cognitive models for understanding and predicting pilot behaviors using advanced features in a HUD, a model should include enhanced characteristics for describing low-level behaviors. In order to model pilot low-level behaviors, eye tracking could be used to collect patterns of pilot attention shifting, which were inferred from verbal protocol in the present study. Another method for low-level behavior modeling would be based on measurements of objective visual display properties (e.g., density, luminance, occlusion, and contrast) affecting pilots perceptions. Once a model is successfully developed with low-level behavior descriptions, it would allow predictions of pilot performance under various display conditions, without costly experiments.

5. REFERENCES

Adams, M. J., Tenney, Y. J. & Pew, R. W. (1995). Situation awareness and the cognitive management of complex systems. *Human Factors*, 37 (1), 85-104.

Alexander, A. L., Wickens, C. D., & Hardy, T. J. (2003). Examining the effects of guidance symbology, display size, and field of view on flight performance and situation awareness. In *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Alexander, A. L., Stelzer, E. M., Kim, S-H., & Kaber, D. B. (2008). Bottom-up and Top-down Contributors to Pilot Perceptions of Display Clutter in Advanced Flight Deck Technologies. In *Proceedings of Human Factors and Ergonomics Society 52nd Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Aragon, C. R. & Hearst, M. A. (2005). Improving aviation safety with information visualization: a flight simulation study. In *Proceedings of the SIGCHI conference on Human factors in computing systems,* 441-450.

Arthur, J. J., Prinzel, L. J., Kramer, L. J., Parrish, R. V., & Baily, R. E. (2004). *Flight Simulator Evaluation of Synthetic Vision Display Concepts to Prevent Controlled Flight Into Terrain (CFIT)*. (Tech. Rep.: NASA TP-2004-213008). Hampton, VA: NASA Langley Research Center.

Arthur, J. J., Kramer, L. J., & Bailey, R. E. (2005). Flight test comparison between Enhanced Vision (FLIR) and Synthetic Vision Systems. In *Proceedings of Enhanced and Synthetic Vision (SPIE) 2005*, Vol. 5802, 25-36.

Bailey, R. E., Parrish, R. V., Kramer, L. J., Harrah, S., & Arthur, J. J. (2002). Technical

Challenges in the Development of a NASA Synthetic Vision System Concept. In *Proceedings of the North Atlantic Treaty Organization (NATO) Symposium on Enhanced and Synthetic Vision Systems*.

Bailey, R. E., Kramer, L. J., & Prinzel, L. J. (2006). Crew and Display Concepts Evaluation for Synthetic/Enhaced Vision Systems. In *Proceedings of the SPIE, Enhanced and Synthetic Vision Conference 2006*.

Bailey, R. E., Kramer, L. J., & Prinzel, L. J. (2007). Fusion of Synthetic and Enhanced Vision for All-Weather Commercial Aviation Operations. *NATO-OTAN*, 11.11-11.26.

Bainbridge, L. & Sanderson, P. (1990). Verbal protocol analysis. In J. R. Wilson & E. N. Corlett. (Eds.), *Evaluation of human work: A practical ergonomics methodology*, 2nd Ed., Taylor & Francis, 169-201.

Bartolone, A. P., Glaab, L. J., Hughes, M. F., & Parrish, R. V. (2005). Initial development of a metric to describe the level of safety associated with piloting an aircraft with synthetic vision systems (SVS) display. In *Proceedings of the SPIE, Enhanced and Synthetic Vision 2005*.

Bender, E. J., Reese, C. E., & van der Wal, G. S. (2003). Comparison of addictive image fusion vs. feature-level image fusion techniques for enhancing night driving. In *Proceedings of the SPIE.*

Bolton, M. L., Bass, E. J., & Comstock, J. R. (2007). Spatial Awareness in Synthetic Vision Systems: Using Spatial and Temporal Judgments to Evaluate Texture and Field of View. *Human Factors*, 49(6), 961-974.

Brenton, R. & Rousseau, R. (2001). *Situation Awareness: A review of the concept and its measurement*. Technical Report No. 2001-220. Defense Research and Development, Canada: Valcartier.

Byrne, M. D., Kirlik, A., Fleetwood, M. D., Huss, D. G., Kosorukoff, A., Lin, R., & Fick, C. S. (2004). A closed-loop, ACT-R approach to modeling approach and landing with and without synthetic vision system (SVS) technology. In *Proceedings of the Human Factors and Ergonomics Society 48th Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Erlbaum.

Corker, K. M., & Guneratne, E. (2002). *Human Factor Issues and Evaluation of Commercial and Business Aircraft Synthetic Vision Systems* (NASA Contractor Final Report No. 21-1214-2882).

Durso, F. T. & Gronlund, S. D. (1999). Situation Awareness, In R. Nickerson, R. Schvaneveldt, S. Dumais, M. Chi, & S. Lindsay. (Eds.), *The Handbook of Applied Cognition*. Wiley, 284-314.

Endsley, M. R. (1988) Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society 32nd Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M. R. (1993). A survey of situation awareness requirements in air-to-air combat fighters. *The International Journal of Aviation Psychology*, 3(2), 157-168.

Endsley, M. R. (1995a). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37, 32-64.

Endsley, M. R. (1995b). Measurement of situation awareness in dynamic systems. *Human Factors*, 37, 65-84.

Endsley, M. R. (2000). Direct measurement of situation awareness: validity and use of SAGAT. In M. R. G. D. J. Endsley (Ed.), *Situation awareness analysis and* *measurement* (pp. 147-173): Mahwah, NJ: Lawrence Erlbaum Associates.

Ertem, M. C. (2005). An Airborne Synthetic Vision System with HITS Symbology Using X-Plane for a Head Up Display. In *Proceedings of the Digital Avionics Systems Conference, DASC 2005.*

Etherington, T. J., Vogl, T. L., Lapis, M. B., & Razo, J. G. (2000). Synthetic vision information system. In *Proceedings of the 19th Digital Aviation Systems Conference.* Philadelphia, PA, 2A4/1-2A4/8.

Fracker, M. L. (1991). *Measures of situation awareness: Review and future directions*. (AL-TR-1991-0128). Wright-Patterson AFB OH: Armstrong Laboratory, Crew Systems Directorate.

Gaba, D. M., Howard, S. K. & Small, S. D. (1995). Situation awareness in anesthesiology. *Human Factors*, Vol. 37, 20-31.

Goodman, A., Hooey, B. L., Foyle, D. C., & Wilson, J. R. (2003). Characterizing visual performance during approach and landing with and without a synthetic vision display: A part task study. In *Proceedings of the 2003 Conference on Human Performance Modeling of Approach and Landing with Augmented Displays*. Moffett Field, CA.

Gordon, S. E. (1995). Cognitive task analysis using complementary elicitation methods, In *Proceeding of the Human Factors and Ergonomics Society 39th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Gordon, S. E. & Gill, R. T. (1997). Cognitive task analysis. In C.E. Zsambok and G. Klein (Eds.), *Naturalistic Decision Making* (pp. 131-140). Mahwah, New Jersey: Laurence Erlbaum Associates.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index):

Results of empirical and theoretical research. In P. A. Handcok & N. Meshkati (Eds.), *Human Mental Workload* (Vol. 1, pp. 139-183): Elsvier, Amsterdam.

Hogg, D. N., Folleso, K., Strand-Volden, F., & Torralba, B. (1995). Development of a situation awareness measure to evaluate advanced alarm systems in nuclear power plant control rooms, *Ergonomics*, 11, 2394-2413.

Hollnagel, E. (2003). Prolegomenon to cognitive task design. In E. Hollnagel (Ed.), *Handbook of Cognitive Task Design* (pp. 3-16). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Jones, D. G. & Endsely, M. R. (2004). Use of Real-Time Probes for Measuring Situation Awareness. *The International Journal of Aviation Psychology*, 14(4), 343-367.

Kaber, D. B., Perry, C. M., Segall, N., McClernon, C. K. & Prinzel, L. P. (2006). Situation awareness implications of adaptive automation for information processing in an air traffic control-related task. *Int. J. of Industrial Ergonomics*, 36, 447-462.

Kaber, D. B., Alexander, A., Hsiang, S., Kim, S.-H., & Stelzer, E. (2007a). Psychophysical Modeling of Perceived Clutter in Advanced Head-Up Displays. In *Proceedings of the NASA Aviation Safety Technical Conference*, St. Louis, Missouri, October 10-12.

Kaber, D. B., Alexander, A., Stelzer, E., Kim, S-H, Kaufmann, K., Cowley, J., Hsiang, S., & Bailey, N. (2007b). *Testing and validation of a psychophysically defined metric of display clutter*. (Final Rep.: NASA Langley Research Center Grant#NNL06AA21A). Hampton, VA: NASA Langley Research Center.

Kaber, D. B., Kim, S-H., Kaufmann, K., Veil, T., Alexander, A. L., Stelzer, E. M., Prinzel, L., & Allamandola, A. (2008). Modeling the Effect of HUD Visual Properties, Pilot Experience and Flight Scenario. Poster in *Proceedings of the NASA Aviation*

Safety Technical Conference. Denver, Colorado, October 21-23.

Keller, J., & Leiden, K. (2002). *Information to support the human performance modeling of a B757 flight crew during approach and landing*. NASA Contractor Report.

Keller, J., Leiden, K. & Small, R. (2003). Cognitive task analysis of commercial jet aircraft pilots during instrument approaches for baseline and synthetic vision displays. In D.C. Foyle, A. Goodman & B.L. Hooey (Eds.), *Proceedings of the 2003 Conference on Human Performance Modeling of Approach and Landing with Augmented Displays*, (NASA/CP- 2003-212267). 15-69. Moffett Field, CA: NASA.

Kieras, D. E. (1997a). Task analysis and the design of functionality. In A. Tucker (Ed.), *Handbook of Computer Science and Eng*ineering, (2nd Ed). Boca Raton, CRC Press.

Kieras, D. E. (1997b). A guide to GOMS model usability evaluation using NGOMSL. In M. Helander, T. Landauer & P. Prabhu. (Eds.), *Handbook of Human-Computer Interaction*, 2nd Ed. completely revision ed., 733-766.

Kieras, D. E. (1999). A guide to GOMS model usability evaluation using GOMSL and GLEAN3. In M. Helander, T. Landauer & P. Prabhu. (Eds.), *Handbook of human-computer interaction*, 3rd Ed., Amsterdam: North-Holland, 733-776.

Klein, G.. A., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(3), 462-472.

Koubek, R. J., Salvendy, G., & Noland, S. (1994). The use of protocol analysis for determining ability requirements for personal selection on a computer-based task. *Ergonomics*, 37 (11), 1787-1800.

Kramer, L. J., Arthur, J. J., Bailey, R. E., & Prinzel, L. J. (2005). Flight testing an integrated synthetic vision system. In *Proceedings of the SPIE, Enhanced and Synthetic Vision 2005*.

Lahtinen, T. M. M., Koskelo, J. P., Laitinen, T., & Leino, T. K. (2007). Heart rate and performance during combat missions in a flight simulator. *Aviation, Space, and Environmental Medicine*, Vol. 78, No. 4, 387-391.

Latorella, K., Pliske, R., Hutton, R., & Chrenka, J. (2001). *Cognitive Task Analysis of Business Jet Pilots' Weather Flying Behaviors: Preliminary Results*. NASA TM-2001-211034.

Leiden, K., Keller, J. W., & French, J.W. (2001). *Context of Human Error in Commercial Aviation*. (Technical Report). Boulder, CO: Micro Analysis and Design, Inc.

Ma, R. & Kaber, D. B. (2005). Situation awareness and workload in driving while using adaptive cruise control and a cell phone. *International Journal of Industrial Ergonomics*, 35(10), 939-953.

McDaniel, R., Scribner, D., Krebs, W., Warren, P., Ockman, N., & McCarley, J. (1998). Image Fusion for Tactical Application. In *Proceedings of the SPIE-Infrared Technology and Applications,* 3436, 685-695.

McKinley, J. B., Heidhausen, E., Cramer, J. A., & Krone, N. J. (2005). Flight Testing of an Airborne SVS with Highway-in-the-Sky on a Head-Up Display. In *Proceedings of the Digital Avionics Systems Conference*, *DASC 2005*, 2, 13.C.1-1-13.C.1.1-9.

Metalis, S. A. (1993). Assessment of pilot situational awareness: Measurement via simulation. In *Proceedings of the Human Factors Society 37th Annual Meeting*, 113-117. Santa Monica, CA: The Human Factors and Ergonomics Society.

Militello, L. G. & Hutton, R. J. (1998). Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands, 42 (11), 1618-1641.

Moroze, M. L., & Snow, M. P. (1999). Causes and remedies of controlled flight into terrain in military and civil aviation. In *Proceedings of the 10th International Symposium on Aviation Psychology*. Columbus, OH, Ohio State University, 2, 1143-1149.

Newell, A. & Simon, H. (1972). *Human problem solving*. Englewood Cliffs, N.J: Prentice Hall.

O'Hare, D., Wiggins, M., Willams, A., & Wong, W. (1998). Cognitive task analyses for decision centered design and training. *Ergonomics*, 41 (11), 1698-1718.

Prinzel, L. J., Kramer, L.J., Comstock, J.R., Bailey, R.E., Hughes, M.F., & Parrish, R.V. (2002). NASA synthetic vision EGE flight test. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Prinzel, L. J., Hughes, M. F., Arthur, J. J., & Kramer, L. J. (2003). Synthetic vision CFIT experiments for GA and commercial aircraft: "A picture is worth a thousand lives". In *Proceeding of the Human Factors and Ergonomics Society 47th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Prinzel, L. J., Comstock, J. R., Glaab, L. J., Kramer, L. J. & Arthur, J. J. (2004a). The efficiency of head-down and head-up synthetic vision display concepts for retro- and forward-fit of commercial aircraft. *International Journal of Aviation Psychology*, 14(1), 53-77.

Prinzel, L. J., Kramer, L. J., Arthur, J. J., Bailey, R. E. & Comstock, R. J. . (2004b). Comparison of head-up and head-down "Highway-In-The-Sky" tunnel and guidance concepts for synthetic vision displays. In *Proceeding of the Human Factors and Ergonomics Society 48th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Prinzel, L. J., Arthur, J.J., Kramer, L.J., & Bailey, R.E. (2004c). Pathway Concepts Experiment For Head-Down Synthetic Vision Displays. In *Proceedings of the SPIE, Enhanced and Synthetic Vision 2004*. 5424-02.

Prinzel, L. J., & Risser, M. (2004). *Head-Up Displays and Attention Capture* (No. NASA/TM-2004-213000).

Prinzel, L. J. & Kramer, L. J. (2006). Synthetic Vision Systems. In W. Karwowski. (Eds.), *International Encyclopedia of Ergonomics and Human Factors* (pp. 1264-1271). 2nd Ed., Talyor & Francis.

Redding, R. E. & Seamster, T. L. (1994). Cognitive task analysis in air traffic controller and aviation crew training, in N. Jonston, N. McDonald & R. Fuller (Eds.), *Aviation Psychology in Practice* (Aldershot: Avebury), 190-222.

Rousseau, R., Tremblay, S., & Breton, R. (2004). Defining and modeling situation awareness: a critical review. In S. Banbury & S. Tremblay. (Eds.). *A cognitive approach to situation awareness: Theory and application*, 3-21.

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review*, 115, 101–130.

Sarter, N. B. & Woods, D. D. (1995). How in the World Did We Ever Get into That Mode? Mode Error and Awareness in Supervisory Control. *Human Factors*, 37(1), 5-19.

Schnell, T., Kwon, Y., Merchant, S., & Etherington, T. (2004). Improved flight technical performance in flight decks equipped with synthetic vision information system display. *International Journal of Aviation Psychology*, 14, 79-102.

Schnell, T., Ellis, K., & Etherington, T. (2005). Flight simulator evaluation of an integrated synthetic and enhanced vision system for terrain avoidance. In *Proceedings of the 24th Digital Avionics Systems Conference, DASC 2005*.

Shadbolt, N. & Burton, M. (1990). Knowledge elicitation: a systematic approach. In J. R. Wilson & E. N. Corlett, (Eds). *Evaluation of human work*, 2nd Ed, Taylor & Francis, 406-440.

Smith, K. & Hancock, P. A. (1995). Situation awareness is adaptive, externally directed consciousness. *Human Factors*, Vol. 37(1), 137-148.

Snow, M. P., & Reising, J. M. (1999). Effect of pathway-in-the-sky and synthetic terrain imagery on situation awareness in a simulated low-level ingress scenario. In *Proceedings of the 4th Annual Symposium on Situation Awareness in the Tactical Air Environment*, Patuxent River, MD, 198-207.

Snow, M. P., Reising, J. M., Kiggett, K. K., & Barry, T. P. (1999). Flying complex approaches using a head-up display: Effects on visibility and display type. In *Proceedings of the 10th International Aviation Psychology Symposium*.

Svensson, E., Angelborg-Thanderez, M., Sjoberg, L., & Olsson, S. (1997). Information complexity-mental workload and performance in combat aircraft. *Ergonomics*, 40(3), 362-380.

Taylor, R. M. (1989). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In *Proceedings of the AGARD AMP Symposium on Situational Awareness in Aerospace Operations*, CP478.

Theunissen, E., Roefs, F. D., Koeners, G.J.M., Rademaker, R. M., & Etherington, T. J. (2004). Integration of iImaging sensor data into a synthetic vision display. In

Proceedings of the 23rd Digital Avionics Systems Conference. Salt Lake City, UT.

Thomas, L. C., & Wikens, C. D. (2004). Eye-tracking and individual differences in off-normal even detection when flying with a synthetic vision system display. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Uhlarik, J., & Comerford, D. A. (2002). *A review of situation awareness literature relevant to pilot surveillance functions*. (Tech. Rep.: Federal Aviation Administration, DOT/FAA/AM-02/3).

Ververs, P. M. & Wickens, C. D. (1998). Head-up displays: Effects of clutter, display intensity, and display location on pilot performance. *International Journal of Aviation Psychology*, 8(4), 377-403.

Vidulich, M. A. (1992). Measuring situation awareness. In *proceedings of the Human Factors and Erogonmics Society 36th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Wei, J. & Salvendy, G. (2004). The cognitive task analysis methods for job and task design: Review and reappraisal. *Behaviour & Information Technology*, 23(4), 273-299.

Wickens, C. D. (1994). Conformal symbology, attention shifts, and the head-up display. In *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Wickens, C. D. (2002). Situation awareness and workload in aviation. *Current Directions in Psychological Science,* (11), 128–133.

Wickens, C. D., Fadden, S., Merwin, D., & Ververs, P. M. (1998). Cognitive factors in aviation display design. In *Proceedings of the 17th Digital Avionics Systems*

Conference (DASC), The AIAA/IEEE/SAE.

Wickens, C. D. & Hollands, J. G. (2000). *Engineering psychology and human performance*, 3rd Ed., Prentice Hall.

Wickens, C. D., Alexander, A. L., Horrey, W. J., Nune, A. & Hardy, T. J. (2004). Traffic and flight guidance depiction on a synthetic vision system display: The effects of clutter on performance and visual attention allocation. In Proceedings *of the Human Factors and Ergonomics Society 48th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Wilson, J. R., Hooey, B. L., Foyle, D. C., & Williams, J. L. (2002). Comparing pilot's taxi performance, situation awareness and workload using command-guidance, situation-guidance and hybrid head-up display symbologies. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Wood, R. B. & Howells, P. J. (2001). Head-Up Displays. In C. R. Spitzer. (Eds.), *The Avionics Handbook* (pp. 4-1-4-24). Boca Raton, FL: CRC Press.

Zachary, W. W., Ryder, J. M., & Hicinbothom, J. H. (2000). Building Cognitive Task Analyses and Models of a Decision-making Team in a Complex Real-Time Environment. In J. M. Schraagen, S. F. Chipman, V. Shalin (Eds.). *Cognitive Task Analysis*, Erlbaum.

APPENDICES

APPENDIX A: Task Flight Scenario

[for IMC-Day conditions]
You are flying a day flight from New York, NY (JFK) to Reno, NV (RNO).
Your ETA is 1300 local time (2300Z).

[for IMC-Night conditions]
You are flying a night flight from New York, NY (JFK) to Reno, NV (RNO).
Your ETA is 2200 local time (0500Z).

Your equipment is a B-757 with EVS (Enhanced Vision System) and SVS (Synthetic Vision System) displays on-board. These systems are used to support a HUD (head-up display) for the pilot flying. You will fly the ILS Runway 16R approach to RNO, beginning abeam the IAF (PYRAM) and ending with either a landing or missed approach. Your first officer is an experienced pilot, but new to your company and not yet proficient with company specific procedures and crew coordination standards. Prior to beginning the approach, you will have an opportunity to brief him on the approach and your expectations for crew coordination.
[Provide pilot with instrument approach plate and weather conditions for study.]

[Read to pilot while (s)he reviews plates.]
You will start out at an altitude of 8500 ft abeam PYRAM (3400 ft) on localizer on a heading of 164. This will be your IAF (Initial Approach Fix). You will fly the ILS RWY 16R approach as depicted on the approach plate. There is a PAPI system on the left side of RWY 16R (and it can only be used within 2nm of the runway threshold). The decision height for the field is 5515ft MSL (1100 AGL). You can assume at the start of the scenario that your vehicle is on course from PYRAM. Your autopilot and flight director are inoperative, and you will be hand-flying the approach. However, your FO will manipulate flaps and landing gear for you and the throttles are operated automatically. All navaids and FMC are already set for the approach.

[Experimenter to give printed copy of weather strip to pilot.]
TAF KRNO XX0400Z XX0410Z 16510KT 4SM HZ OVC015 TEMPO0410 1/2SM FG OVC001 (where XX = Date of session)

- Wind 165 at 10 kts
- Visibility 4 SM (statute miles) in haze
- Ceiling 1500 and overcast
- Temporary conditions
  - 0.5 SM in fog
  - Ceiling 100 and overcast.

	[Experimenter asks pilot to brief planned approach and crew coordination expectations.] Like we did in training trials, would you please brief the planned approach and crew coordination?
Discussion about role	**[Pilot should state that FO will:** • **communicate with ATC;** • **operate the FCP (speed bug), altimeter, flaps and gear;** • **confirm aircraft control settings as required (e.g., flaps 15 selected);** • **complete the landing checklist; and** • **make callouts as instructed (e.g., 1000 to go, 500 to go).]** **[The pilot should confirm that he will inform the FO of:** • **his decision at decision height; and** • **when he has the runway in sight.]**
Flight Preparation	**[for IMC-Night conditions]** Your 757 is beginning the ILS RWY 16 approach at PYRAM at night. Altitude is 8500 ft. with poor visibility due to cloud. HDG is 164 degrees. Speed is 210 knots with Flaps 0. **[for IMC-Day conditions]** Your 757 is beginning the ILS RWY 16 approach at PYRAM in daylight. Altitude is 8500 ft. with poor visibility due to cloud. HDG is 164 degrees. Speed is 210 knots with Flaps 0. **[Experimenter prepares appropriate trial setup in application, based on experiment design and trial allocation.]** **[Experimenter advises pilot of display configuration.]** Your HUD configuration for this trial includes...[Baseline / SVS / EVS / SVS&EVS]
Initial Clearance	**[Experimenter role-playing ATC to provide ATIS broadcast.]** “Reno/Tahoe international information. **[for IMC-Day Conditions]** Delta. 2300 Zulu weather. **[for IMC-Night Conditions]** Delta. 0500 Zulu weather.

Ceiling 1500, overcast. Visibility 4 miles, haze.
Temperature 15, dew point 12.
Wind 165 at 10 knots.
Temporary conditions are:
0.5 SM in fog
Ceiling 100, overcast.
Altimeter 3007.
ILS Runway 16R in use. Landing 16R, Departures 16L.
Unlit tower 150 feet AGL 2 nautical miles south of airport.
Advise on initial contact you have Delta.

Would you please confirm when you are ready to begin this trial by saying “ready”.

ATC clearance #1 (23.0 DME)	**[ATC to provide clearance (at beginning of trial). It will be played automatically by the experiment prototype application]** *“Wolfpack1. Reno Approach: Abeam PYRAM, maintain 8500 ft until established, cleared ILS 16R approach. Maintain 210 for slower traffic. Altimeter setting is [2999 or 3003].”* **[Experimenter (FO) responds.]** “Reno Approach, Wolfpack1, maintain 8500 until established, cleared ILS 16 right, maintain 210. Altimeter [2999 or 3003], we have Delta.”
ATC clearance #2 (19.0 DME)	**[ATC to provide clearance at 19.0 DME]** *“Wolfpack,. Reno Approach. Slow to approach speed.”* [Experimenter (FO) responds.] “Reno Approach, Wolfpack 1. Slow to approach speed.” **[Pilot is expected to call for FCP speed setting (138kts), call for extending flaps and landing gear.]** **[Experimenter should respond to pilot calls (e.g., flaps, landing gear or speed setting)]**

ATC Clearance #3 (9.0 DME)	**[ATC clearance will be played automatically.]** *"Wolfpack1, Reno Approach. Approaching DICEY, contact tower."* **[Experimenter (FO) responds.]** "Reno Approach, Wolfpack1, switches to tower."
ATC Clearance #4 (8.0 DME)	**[Experimenter (FO) tunes Reno tower (118.7)]** "Reno Tower, Wolfpack1, DICEY inbound on ILS 16R." **[ATC will be played]** "*Wolfpack1, cleared to land 16 R. Wind 165 at 10 kts.*" **[Experimenter (FO) responds]** "Wolfpack1, cleared to land." **[Provide callouts to pilot according to assigned callout type]**
6515ft MSL	**[for callout type A: "1000 to go" "500 to go", "400 to go", "300 to go", "200 to go", "100 to go", and "Minimum" according to MSL altitude from decision height.]** **[for callout type B: "500 to go" according to MSL altitude from decision height.]**
5515ft MSL (Decision Height)	**[Experimenter (FO) callouts.]** "Minimums" **[Pilots are expected to say "continue" or "landing."]**
After Landing	Now, you have completed a test trial.

APPENDIX B: Situation Awareness Global Assessment Questions

Spatial Awareness

Level 1

- Was the flight path marker above the horizon?
- Where was the flight from the center of the tunnel?
- Did you see the actual runway (or rabbit) with your bare eyes?
- Were you seeing the runway outline in the display at the time the simulator stopped?
- Estimate the current pitch of the aircraft at the time the simulator stopped.
- Does your flight path marker currently overlap the terrain image?
- Where was the flight path marker relative to the runway?
- Where was the aircraft from the center of the tunnel?
- Give a description of the terrain at 10 o'clock, halfway to the horizon on the display.
- Give a description of the terrain at 6 o'clock, halfway to the bottom of the display.

Level 2

- Was the aircraft ascending or descending at the time the simulator stopped?
- Was the aircraft moving away from or toward the ILS (localizer and glideslope)?
- In which direction do you need to move the flight path to align with the runway?
- Was the aircraft moving away from or toward the center of the tunnel?
- In what direction from your aircraft was the nearest significant terrain feature you passed?
- In what clock position is the nearest terrain in front of you?

Level 3

- If your flight continues as it is now, which direction will the aircraft fly relative to the tunnel?
- If your flight continues as it is now, would the risk that your aircraft collides into terrain increase or decrease?
- From your current position, what is the safest route if forced off the approach by traffic at 12 o'clock?
- What control movements do you need to maintain in order to make the aircraft at the center of the tunnel?
- If your flight continues as it is now, which direction will the aircraft deviate relative to the glideslope?
- If you are forced to go-around from your current position, which direction would you need to direct the aircraft?

- What control movements do you need to correct to the ILS (localizer and glideslope)?

System Awareness

Level 1

- What is the current airspeed bug setting?
- What was your ground speed at the time the simulator stopped?
- What was the current aircraft heading at the time the simulator stopped?
- What was the DME to runway at the time the simulator stopped?
- What was your air speed at the time the simulator stopped?
- What was your MSL altitude at the time the simulator stopped?
- What was your vertical speed at the time the simulator stopped?
- Has your aircraft intercepted the glideslope?
- What was the altimeter setting value in display?

Level 2

- How far is the aircraft above decision height?
- How long has it been since you began your descent?
- Was the aircraft's vertical speed increasing or decreasing at the time the simulator stopped?
- Was the aircraft's accelerating or decelerating at the time the simulator stopped?
- How much has the altitude deviated from target MSL altitude?
- How much has the airspeed deviated from the airspeed bug setting?
- If you want to be descending at -700 FPM (feet per min), should you increase or decrease pitch?

Level 3

- When will your aircraft intercept the glideslope?
- In order to keep the target airspeed, would you increase or decrease the throttle setting?
- How far do you need to descend to see the glide slope reference line in center of the display?
- When will your aircraft reach the runway?
- When will your aircraft reach the published decision height?
- Estimate the MSL altitude after 10 sec, if your flight continues as it is now.
- Estimate the airspeed after 10 sec, if your flight continues as it is now.
- Estimate the heading after 10 sec, if your flight continues as it is now.

Task Awareness

Level 1

- Has your FO set the airspeed bug?
- Have you received ATC clearance to contact tower?
- What was the last callout from your FO?
- What was the last voice warning from the terrain awareness warning system?
- What was the flap position at the time the simulator stopped?
- Has your FO set the altimeter?
- Has your FO contacted tower?
- Did your FO made callout “1000 to go”?
- Have you received landing clearance from the tower?

Level 2

- At what DME did you ask to extend flaps?
- At what DME did you ask to complete landing checklist?
- How long has it been since you received “clear to land” from tower?
- How long has it been since you received your last ATC clearance?
- At what DME did you call for landing gear down?
- What did you last communicate with your FO?

Level 3

- What is your next task?
- What will you next ask your FO?
- Which will be next action of your FO?
- What do you expect your decision for landing will be, based on current weather?
- What will be the next voice warning from terrain awareness warning system?
- What do you expect us your next ATC clearance?
- When should you make landing decision?

APPENDIX C: Informed Consent Form

North Carolina State University

INFORMED CONSENT FORM for RESEARCH

Title of Study: Examining and Explaining the Effects of Non-Iconic Conformal Features in Advanced Head-up displays on Pilot Performance.

Principal Investigators: Sang-Hwan Kim Faculty Sponsor: Dr. David Kaber

We are asking you to participate in a research study. The purpose of this study is to assess the impact of Synthetic Vision System (SVS) and/or Enhanced Vision System (EVS) terrain features on pilot performance when displayed in an advanced head-up display (HUD) during various phases of a landing approach.

INFORMATION

If you agree to participate in this study, you will be asked to perform simulated flight tracking tasks using a PC-based flight simulator, you will be presented with videos of HUD on arrivals and approaches to Reno-Tahoe International Airport (KRNO), including four different prototypes of HUD configurations under two instrument meteorological conditions (IMC). The experimental procedure is as follows: (1) complete a short demographic survey; (2) complete two training sessions for tracking a simulated flight path in order to become familiar with the simulator and control systems; (3) fill-out a perceived workload survey to rate the load of the practice session; and (4) complete nine test trials, each simulating an arrival and approach to KRNO. You will be presented with a flight scenario to follow including aircraft status, clearance from ATC, and required actions. All required charts will be provided. During the first eight test trials, the simulator will be randomly stopped four times and you will be asked to answer queries for assessing your situation awareness. After completing each of the first eight test trials, you will fill out the workload and subjective preference rating form. During the test trials, your tracking control performance will be record by the simulator. You will also be required to wear a Polar-watch heart monitor system on your chest to collect your heart rate. During the 9^{th} test trial, you will be asked to speak out your thoughts or intentions for performing the task and your behavior and speech will be videotaped. While the video is being played, you will be asked to verbalize what you were thinking and intending during the test. (5) Following the test session, you will relax without doing any tasks for approximately 10 min in order for us to measure your baseline heart-rate. You will then be debriefed on the study. The entire procedure will occur in one session and last approximately 4 hours.

RISKS

The risks associated with participation in this study are minimal. They include potential visual strain and/or fatigue from viewing the simulator displays on conventional PC-monitors for an extended period, soreness of the hands from use of the simulated aircraft yoke. The visual workload is expected to be less than that experienced in actual flight of a commercial aircraft because of the controlled laboratory lighting. Any fatigue or strain is expected to be less than that associated with real flight and any symptoms are reversible. Regarding the measurement your heart rate using the Polar watch system, the sensor of the system will be worn on bare skin at your chest. The sensor does not produce an electrical signal interfering with your heart signal frequency and there is no physiological risk. The sensor will be cleaned with alcohol and a swab before each and every test trial.

In the event that you indicate fatigue or discomfort during the described experiment, a rest period will be provided. If abnormal physiologic conditions persist, you will be excused from further participation in the experiment.

BENEFITS
There are no direct benefits of this study to you. You may derive some indirect benefits including an understanding of human factors research methods for aviation system design and insight into advanced HUD features including SVS and EVS technologies.

CONFIDENTIALITY
The information is our study records will be kept strictly confidential. Data on your responses will be stored in locked cabinets in the ISE Ergonomics Lab or on password protected computer workstations. The data will only be made available to persons conducting the study. We will make a video recording of your performance for analysis purposes and all tapes will be destroyed after the research is complete. You will be represented as a number in all test data. No reference will be made to you, individually, in oral or written reports, which could link you to the study.

COMPENSATION
For participation in the study, you will be paid an honorarium of $100.00. If you withdraw from the study prior to its completion, your compensation will be prorated to match the amount of time you participated.

EMERGENCY MEDICAL TREATMENT
If you need emergency medical treatment during the study session(s), the researcher(s) will contact the University's emergency medical services at 515-3333 for necessary care. There is no provision for free medical care for you if you are injured as a result of this study.

CONTACT
If you have questions at any time about the study or the procedures, you may contact Dr. David Kaber, at the Department of Industrial Engineering, Box 7906, North Carolina State University, or (919) 515 3086. If you feel you have not been treated according to the descriptions in this form, or your rights as a participant in research have been violated during the course of this project, you may contact Mr. Matthew Ronning, Assistant Vice Chancellor, Research Administration, Box 7514, NCSU Campus (919/513-2148).

PARTICIPATION
Your participation in this study is voluntary; you may decline to participate without penalty. If you decide to participate, you may withdraw from the study at any time. If you withdraw from the study before data collection is completed your data will be returned to you or destroyed at your request.

CONSENT
"I have read and understand the above information. I have received a copy of this form. I agree to participate in this study with the understanding that I may withdraw at any time."

Subject's signature____ **Date** ________________

Investigator's signature____________________________________ **Date** ________________

APPENDIX D: Demographic Questionnaire

We would like to know about you and your flight experience. Please answer each of questions below as accurately as you can.

Last Name: ____
First Name: ____
Phone: ____________________________ Age:____________
Email: ____
Age: _________ Gender: ☐ Male ☐ Female

Are you experienced with "glass cockpit" experience? ☐ Yes ☐ No
Total Aircraft Hours for "glass cockpit": __________Hours
Total Simulator Hours for "glass cockpit":_________Hours

Are you experienced with Head-up Display (HUD)? ☐ Yes ☐ No
Total Aircraft Hours for HUD:_______Hours Total Simulator Hours for HUD:_______ Hours

Are you experienced with Synthetic Vision Systems (SVS)? ☐ Yes ☐ No
Total Aircraft Hours for SVS:_______Hours Total Simulator Hours for SVS:________ Hours

Are you experienced with Enhanced Vision System (EVS)? ☐ Yes ☐ No
Total Aircraft Hours for EVS:_______Hours Total Simulator Hours for EVS:________ Hours

Grade (Check all that apply)	Airplane Category Ratings (Check all that apply)
☐ ATP ☐ Commercial	☐ Single Engine ☐ Multi Engine ☐ Land ☐ Sea ☐ Instrument

Total Instrument Time: _______
Total Night Time: __________
Total Flight Time: __________
Total Time Last 12 Months: _________

Please describe the nature of your company's flight standards and the FAR Part you operate under (for example: The company operates under a Part 135 certificate, but company standards and insurance requirements dictate the use of Part 121 PIC qualifications and scheduling restrictions): ________________________
____
____

APPENDIX E: NASA-TLX Forms

Subjective Comparison of Demand Factors:

Examining and Explaining the Effects of Non-Iconic Conformal Features
in Advanced Head-up displays on Pilot Performance.

For each of the pairs listed below, circle the scale title that represents the more important contributor to workload in the display.

Mental Demand	or	Physical Demand
Mental Demand	or	Temporal Demand
Mental Demand	or	Performance
Mental Demand	or	Effort
Mental Demand	or	Frustration
Physical Demand	or	Temporal Demand
Physical Demand	or	Performance
Physical Demand	or	Effort
Physical Demand	or	Frustration
Temporal Demand	or	Performance
Temporal Demand	or	Frustration
Temporal Demand	or	Effort
Performance	or	Frustration
Performance	or	Effort
Frustration	or	Effort

Definition of Task Demand Factor

Mental demand

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

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

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?

Performance

How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

Frustration level

How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Effort

How hard did you have to work (mentally and physically) to accomplish your level of performance?

Subjective Comparison of Demand Factors

Please place an "X" along each scale at the point that best indicates your experience with the display configuration.

<u>Mental Demand</u>: How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc)? Was the mission easy or demanding, simple or complex, exacting or forgiving?

Low High

<u>Physical Demand</u>: How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the mission easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Low High

<u>Temporal Demand</u>: How much time pressure did you feel due to the rate or pace at which the mission occurred? Was the pace slow and leisurely or rapid and frantic?

Low High

<u>Performance</u>: How successful do you think you were in accomplishing the goals of the mission? How satisfied were you with your performance in accomplishing these goals?

Low High

<u>Effort</u>: How hard did you have to work (mentally and physically) to accomplish your level of performance?

Low High

<u>Frustration</u>: How discouraged, stressed, irritated, and annoyed versus gratified, relaxed, content, and complacent did you feel during your mission?

Low High

APPENDIX F: Post-trial Questionnaire

Please answer the following by circling a number that best completes the sentence based on your experience in the scenario.   Please elaborate in the comments area

1. On a scale from 1-7, how useful was the HUD configuration for understanding the aircraft's position relative to the terrain and approach path?

|___1___|___2___|___3___|___4___|___5___|___6___|___7___|

Not useful at all   Extremely useful

Comments: ____
____
____
____

2. How useful was the HUD configuration for understanding current flight status (e.g., airspeed, altimeter setting, and designated heading)?

|___1___|___2___|___3___|___4___|___5___|___6___|___7___|

Not useful at all   Extremely useful

Comments: ____
____
____
____

3. How useful was the HUD configuration for using your knowledge of aircraft control, navigation, and communication (with ATC or FO)?

|___1___|___2___|___3___|___4___|___5___|___6___|___7___|

Not useful at all Extremely useful

Comments: ____

____

____

____

4. Do you think the HUD configuration would enhance aviation safety by in low visibility conditions?

|___1___|___2___|___3___|___4___|___5___|___6___|___7___|

Not at all Yes

Comments: ____

____

____

____

5. Was the HUD configuration you used in the previous trial annoying? If so, please state the reasons and make any suggestions for improving this display.

|___1___|___2___|___3___|___4___|___5___|___6___|___7___|

Not at all Yes

Comments: ____

____

____

____