Examining the Effects of Conformal Terrain Features in Advanced Head-Up Displays on Flight Performance and Pilot Situation Awareness

Sang-Hwan Kim[1] and David B. Kaber[2]

1 Department of Industrial and Manufacturing Systems Engineering, University of Michigan-Dearborn, Dearborn, MI 48128
2 Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, NC 27695

Abstract

Synthetic vision systems (SVS) render terrain features for pilots through cockpit displays using a GPS database and three-dimensional graphical models. Enhanced vision systems (EVS) present infrared imagery of terrain using a forward-looking sensor in the nose of an aircraft. The ultimate goal of SVS and EVS technologies is to support pilots in achieving safety under low-visibility and night conditions comparable to clear, day conditions. This study assessed pilot performance and situation awareness (SA) effects of SVS and EVS imagery in an advanced head-up display (HUD) during a simulated landing approach under instrument meteorological conditions. Videos of the landing with various HUD configurations were presented to eight pilots with a superimposed tracking task. The independent variables included four HUD feature configurations (baseline [no terrain imagery], SVS, EVS, and a combination of SVS and EVS), two visibility conditions, and four legs of the flight. Results indicated that SVS increased overall SA but degraded flight path control performance because of visual confusion with other display features. EVS increased flight path control accuracy but decreased system (aircraft) awareness because of visual distractions. The combination of SVS and EVS generated offsetting effects. Display configurations did not affect pilot spatial awareness. Flight performance was not different among phases of the approach, but levels and types of pilot SA did vary from leg to leg. These results are applicable to development of adaptive HUD features to support pilot performance. They support the use of multidimensional measures of SA for insight on pilot information processing with advanced aviation displays. © 2012 Wiley Periodicals, Inc.

Keywords: Head-up display (HUD); Synthetic vision systems (SVS); Enhanced vision systems (EVS); Flight performance; Pilot SA

1. INTRODUCTION

According to data from the Flight Safety Foundation, almost 60% of all commercial aircraft crashes occur during airport approach and runway 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). Leiden, Keller, and French (2001) examined historical CFIT accident reports and found that the majority are not attributable to mechanical errors or external (normal/abnormal)

Correspondence to: Sang-Hwan Kim, Department of Industrial and Manufacturing Systems Engineering, University of Michigan-Dearborn, 2230 HPEC, 4901 Evergreen Rd., Dearborn, MI 48128-1491. Phone: 313-436-9140; e-mail: dysart@umich.edu

Received: 5 January 2011; revised 7 November 2011; accepted 7 November 2011

View this article online at wileyonlinelibrary.com/journal/hfm

DOI: 10.1002/hfm.20501

situations but primarily due to human errors inside the flight cockpit, such as communication errors between controllers and the flight crew, loss of vertical and horizontal situation awareness (SA), and crew resource management errors.

To reduce the accident rate and enhance aviation safety, various systems (e.g., altitude indicators, radio navigation, ground proximity warning systems, and instrument landing systems [ILS]) have been developed and introduced into aircraft and airport facilities to address low visibility conditions for pilots. Snow and Reising (1999) previously stated that aircraft information systems need to include intuitive technologies supporting pilot SA, specifically spatial orientation (relative to terrain and flight path), without requiring diversion of visual attention and cognitive resources from external events and primary flight references. With this in mind, next-generation (NextGen) concepts for the National Airspace System integrate technologies to support flight safety through improved pilot terrain awareness. Information displays represent a subset of these technologies, including synthetic vision systems (SVS) and enhanced vision systems (EVS), for the aircraft cockpit. SVS displays present computer-generated images of the external scene topography from the perspective of the flight deck that are derived from aircraft attitude, high-precision navigation solutions, and a database of terrain, obstacles, and relevant cultural features. An EVS is an electronic means by which to provide a display of the forward external scene topography (the natural or man-made features of a place or region) especially in a way to show their relative positions and elevation through the use of imaging sensors, including forward-looking infrared (FLIR). Terrain features are revealed by these systems to pilots as realistic "non-iconic" imagery in head-down or head-up displays (HDD/HUD) and are expected to enhance flight safety by improving pilot SA. When these features are presented in a HUD as overlapping imagery on an actual out-of-cockpit view, the imagery is required to be conformal; that is, the terrain model or infrared imagery matches the actual terrain view by creating intuitive, visual-like information scaled and aligned to mimic the external scene. Such HUDs also promote information proximity for a pilot; that is, the terrain imagery and primary flight displays (PFDs) are in line with a pilot's out-of-cockpit view (Ververs & Wickens, 1998; Wickens & Hollands, 2000).

Previous research has also demonstrated that SVS imagery, including highway-in-the-sky (HITS) or tunnel displays, improves flight performance and/or pilot SA, and reduces workload (Prinzel et al., 2003; Snow & Reising, 1999). In specific, HITS images in SVS have been found to be major factors in improving pilot flight path tracking accuracy (Alexander, Wickens, & Hardy, 2003; Bailey, Kramer, & Prinzel, 2007; Prinzel et al., 2004; Wickens et al., 2004). However, it also has been found that pathway tunnels may cause a cognitive tunneling effect (Alexander et al., 2003; Thomas & Wickens, 2004; Wickens et al., 2004). That is, although pathway tunnels can support better flight path tracking, they may degrade, for example, pilot traffic awareness and ability to detect unexpected events. Because SVS features are generated from a database, they have the potential disadvantage of providing pilots with inaccurate information relative to the actual state of the terrain. Consequently, the use of a combination of SVS and EVS displays was suggested (Arthur, Kramer, & Bailey, 2005). Research has investigated the utility of SVS and EVS terrain features rendered in an HDD versus pilot use of conventional flight instrument displays (Schnell, Ellis, & Etherington, 2005). Results confirmed the SVS features, including pathway guidance, to improve flight tracking performance, SA, and workload. Additional EVS feature insets in the SVS-HDD did not show significant effects. However, there has been a lack of research evaluating the effects of the combination of SVS and EVS terrain features in HUDs on pilot performance and SA.

Regarding pilot SA, it has been posited that there are multiple types of SA supporting different types of tasks and cognitive behaviors in the aviation domain. Wickens (2002) divided pilot SA into three dimensions: spatial awareness (SpA), system (mode) awareness, and task awareness. The concept of SpA is inherent in the task of moving an aircraft through a three-dimensional space that can be filled with hazards. SA concerns a pilot's comprehension of aircraft status and mode, which may affect pilot performance. Finally, task awareness relates to a pilot's knowledge of aviation control, navigation, communication, and systems management. Endsley (1995) divided SA into three levels of cognition, which are relevant to aircraft piloting, including: perception of elements in the airspace, comprehension of their meaning relative to pilot goals and the current state of ownership, and projection of element status in the near future as a basis for effective flight path control. Related to these concepts of SA, Bolton, Bass, and Comstock (2007) introduced judgment-based measures of pilot SpA to evaluate terrain texture and field-of-view features of an SVS–PFD in an experiment. Based on Wickens's definition of SpA (Wickens, 2002) and Endsley's (1995) concept of the three levels of SA, Bolton et al. (2007) identified three levels of SpA with respect to terrain, including identification of the terrain (Level 1), relative spatial location of the terrain (Level 2), and relative temporal location of the terrain (Level 3). The experimental results showed that SpA was best facilitated by fishnet textures.

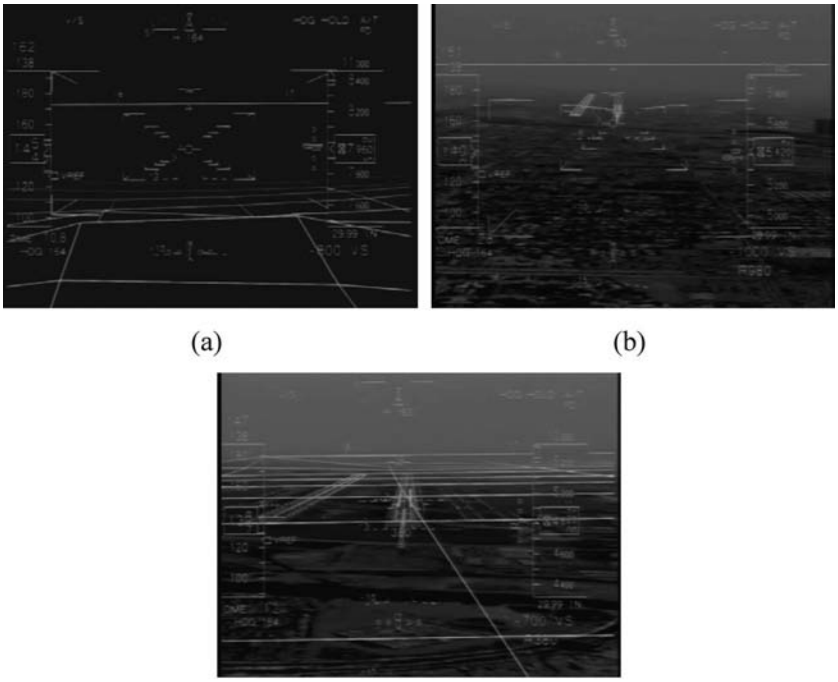
In the present study, an experiment was conducted to assess the effects of SVS and EVS imagery (non-iconic, conformal display features) on pilot flight path control performance and SA when using an advanced HUD during a simulated landing approach under instrument meteorological conditions (IMC; conditions for which SVS and EVS are most relevant). The study focused on the independent and combined effects of each feature (SVS, EVS) relative to a baseline HUD symbology condition (without any non-iconic terrain imagery), including pathway tunnel features. When measuring SA in the experiment, assessment was made in terms of the three levels and types of pilot SA identified by Endsley (1995) and Wickens (2002), respectively. This detailed analysis of pilot SA with advanced HUDs has not been conducted in prior research. Based on the prior research described earlier in text, we speculated that both the SVS and EVS HUD features would promote pilot SpA across levels of SA and, in turn, support accurate flight path control. However, the combination of terrain features was expected to generate display clutter (cf., Kaber et al., 2009) and potentially disrupt pilot visual scan, flight SA, and path control.

2. METHOD

2.1. Stimuli and Flight Simulator

Videos of expert pilot performance with an advanced HUD, including non-iconic conformal features (SVS, EVS, or a combination of both [hereafter referred to as the "Combo"]), were recorded using the Integration Flight Deck (IFD) simulator at NASA Langley Research Center. An ex-Air Force check pilot, with some HUD experience, operated the IFD simulator (a high-fidelity representation of a Boeing 757 aircraft) according to a flight scenario involving an approach and landing to runway 16R at Reno-Tahoe International Airport (KRNO). Different HUD configurations were used in several trials, including Baseline (without any terrain imagery), SVS, EVS, and the Combo condition. The IFD simulator presented SVS terrain features using a wireframe grid in the HUD. The use of the wireframe feature was based on findings from Snow and Reising's study (1999), which revealed the grid model to be most appropriate for depicting terrain. The wireframe features in the present study were set to represent terrain using a 500-meter line separation with a 1-pixel line width. Figure 1a shows a captured SVS HUD image. The EVS-HUD presented an actual out-of-cockpit view using a sensor-based FLIR camera. Figure 1b shows the EVS-HUD, and Figure 1c shows the combination of SVS and EVS (Combo) features, recorded using the IFD simulator. As shown in Figure 1, a and b, 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 moving "Crow's feet." The use of this feature was based on the Prinzel et al. (2004) study demonstrating crow's feet to be most appropriate for conveying a pathway in the sky in an SVS HUD. Each box presented on the HUD was 600 feet wide by 350 feet tall (in air space). When the aircraft was less than 500 feet above ground level, the tunnel features were programmed to disappear, and a runway outline was shown on the display with a glideslope reference line set at 3.1 degrees. Figure 1c includes the runway outline and glideslope reference line in the HUD. All HUD configurations used in the study included tunnel features and glideslope reference lines.

After completion of the expert pilot flight trials, the recorded video files were prepared and organized for follow-on lab tests using a PC-based simulation of aircraft landing. A low-fidelity lab simulator was used because of access and cost issues associated with the IFD simulator. The videos of the HUD content did not include an out-of-cockpit view but only the HUD imagery on a "black" background. To simulate IMC-day and IMC-night conditions, videos of the out-of-cockpit view for an approach and landing to KRNO runway 16R were recorded using the X-plane system. X-plane is a PC-based flight simulator software, which includes realistic three-dimensional rendering of terrain and runway images for airports. The HUD videos recorded using the IFD simulator and the out-of-cockpit view videos recorded using X-plane were synthesized and rendered using a commercial video-editing tool. Figure 2 shows an example of synthesized video images for the IMC-day conditions. In addition to the image synthesis, several audio files were



(a) (b)

(c)

Figure 1 HUD configurations with (a) SVS terrain features, (b) EVS features, and (c) combination of SVS and EVS features.

integrated with the video stimuli to create a realistic flight simulation (e.g., air traffic control [ATC] directions and engine power up and down). A total of ten video files were prepared for the experiment, consisting of two for practice (Baseline HUD under IMC-night and Combo under IMC-day) and eight for test trials (Baseline, SVS, EVS, and the Combo HUD under the two IMC conditions).

The lab simulator setup consisted of a PC workstation and aircraft-like controllers, including a yoke and a throttle quadrant (see Figure 3 for setup). A Java application was developed to present the video stimuli and record pilot tracking accuracy. Given the use of prerecorded videos, some of the flight controls were limited in functionality. For example, pilot control actions at the throttle quadrant did not affect the airspeed or altitude, as presented on the simulator displays.

2.2. Participants and Tasks

Eight commercial line pilots participated in the lab experiment. All pilots were required to have previous experience in flying aircraft with "glass" cockpit displays. The mean total flight hours across pilots was 11,044 with a standard deviation of 7,893 hours. Among the pilots, three had experience in the use of a HUD, either during actual flight ($M = 2{,}350$ hours) or in a simulator ($M = 86.7$ hours). Two pilots had experience with SVS systems ($M = 4$ hours), and one pilot had EVS experience (6 hours) in simulator flight. Each pilot completed the two practice trials and eight test trials following a completely within-subjects experiment design. One additional experiment trial was conducted to record subject verbal protocols in the approach as a basis for understanding pilot internal behaviors and strategy (i.e., a cognitive task analysis [CTA] was conducted). The expert pilot who flew the NASA IFD simulator for the HUD video recordings role-played a first officer (FO) for participants during all trials.

As part of testing, participants used the yoke to move a cursor to track the recorded aircraft flight path marker (FPM) shown on the monitor. Figure 4 shows a screenshot of the flight simulator display, including the HUD and the superimposed cursor for tracking the FPM. A Java application was developed to control the cursor in the tracking task with the yoke. The cursor exhibited a first-order (velocity) response to yoke control inputs. Maintaining the yoke in an off-center position produced a constant rate of change in the cursor position. When the yoke was returned to the center position, the cursor did not move. Such velocity control systems are frequently encountered in manual control,

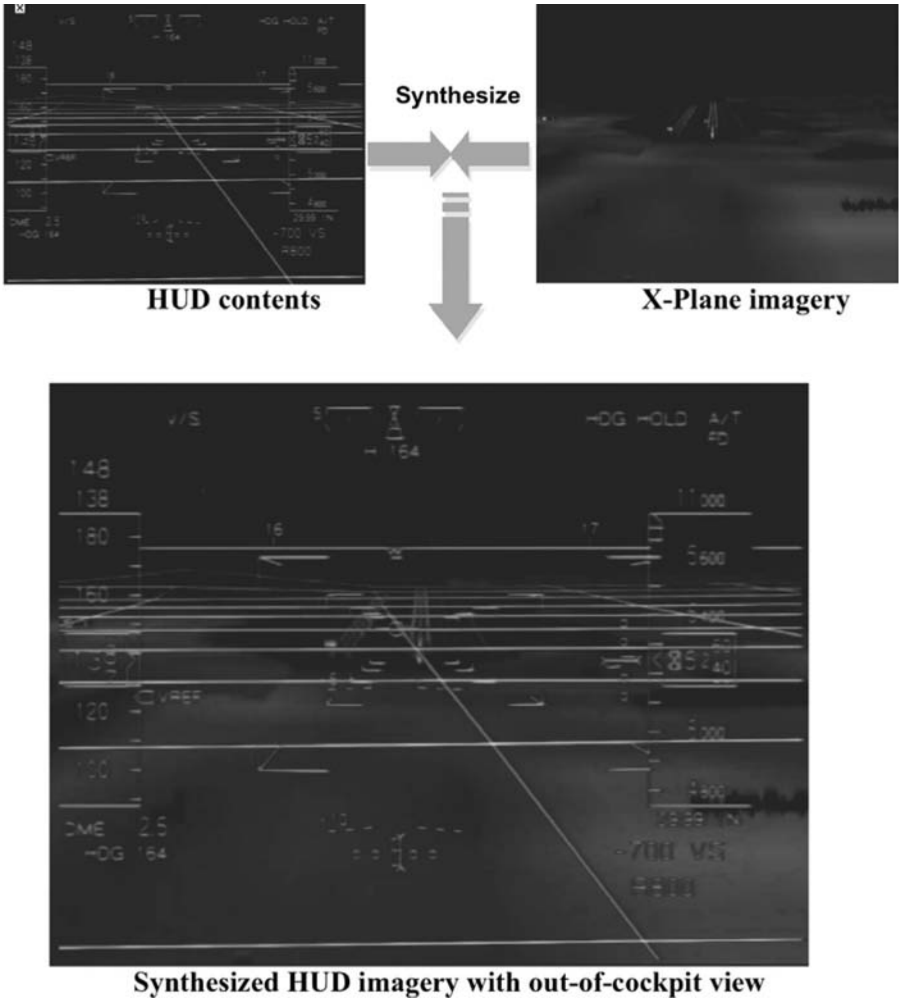


Figure 2 Example of imagery synthesized for the IMC-night condition.

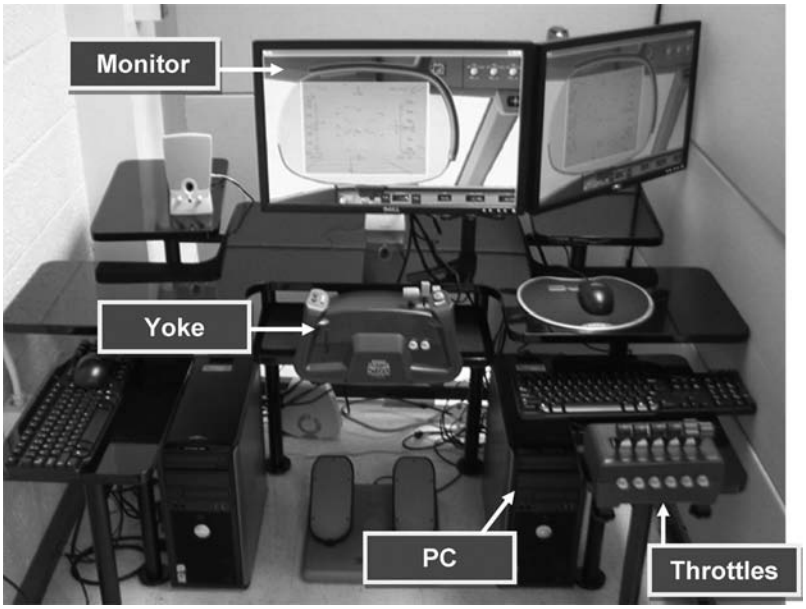


Figure 3 Simulator setup for experiment.

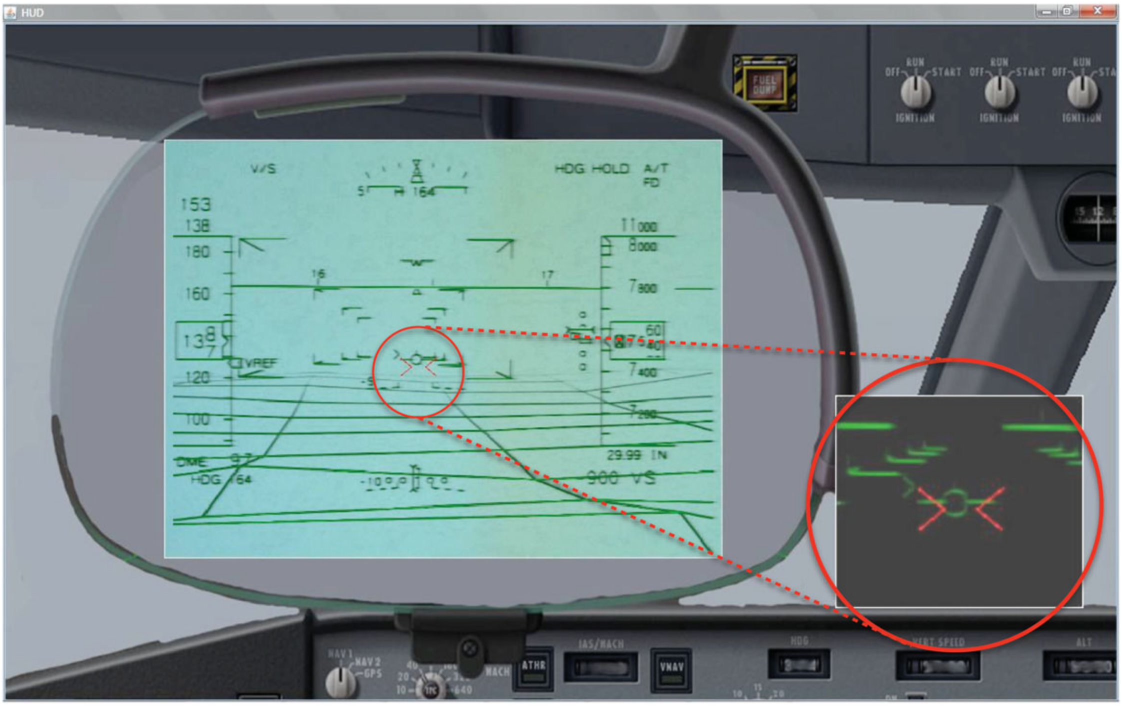


Figure 4 A screenshot of the Java application playing a video of the HUD and overlaid cursor for tracking FPM (in circle).

including aircraft control (Wickens & Hollands, 2000). Although the tracking task did not affect actual aircraft flight maneuvers, the task represented a normal part of pilot performance in hand flying an aircraft. The tracking task was also designed to encourage immersion in the experiment scenario and to virtually mimic flight path control. Because every condition included tunnel features (i.e., HITS), the approach of using the yoke control for the tracking task was intended to reveal any decrements in pilot attention and flight performance (i.e., driving the overlaid cursor to the FPM at the center of the tunnel) because of the presence of terrain features (SVS and/or EVS imagery).

2.3. Experiment Design

Independent variables included the HUD configuration (baseline, SVS, EVS, and the Combo with all configurations including the tunnel feature) and the visibility condition (IMC-day and IMC-night). To investigate a potential additive effect of HUD features on pilot performance, structured orders of test display conditions were used rather than a randomized presentation. Each pilot was presented with display conditions in increasing order of visual density (active pixels [Kaber et al., 2009; Kim et al., 2011]), specifically baseline, SVS, EVS, and then the Combo, or in a decreasing order of density. The two orders were manipulated among the first four test trials (Trials 1–4) and the last four trials (Trials 5–8) for each participant. Half the participants 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. As an additional independent variable, pilot performance was recorded in four legs of flight across approach and landing. Figure 5 illustrates the general concept of the approach scenario across the four legs of flight. Table 1 presents additional detailed information on the legs, including aircraft position, flight characteristics, required control actions by pilots, and visibility of the runway.

Figure 6 shows example screenshots of the video stimuli for the four display configurations, two IMC conditions, and four legs of flight. Because the EVS sensor used in the IFD simulator was not capable of penetrating heavy precipitation and certain types of fog, Figure 6c reveals FLIR returns on "moisture" clouds (which were perceived as display clutter by many pilots).

Pilots were required to complete eight experimental trials involving four legs of flight. Thus, the

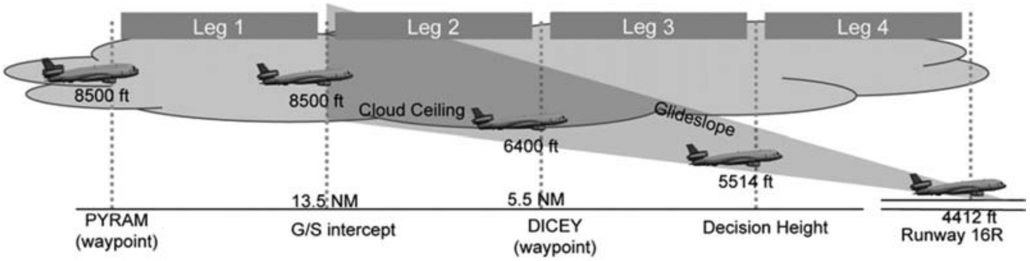


Figure 5 Concept of the approach scenario.

TABLE 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 Flaps and gear extending Complete landing checklist	Descending Contact tower Confirm landing clearance	Descending Landing decision making	Landing
Visibility of runway	Invisible	Invisible	Visible	Visible

PYRAM & DICEY = Waypoints for KRNO 16R approach.
G/S = Glideslope; IAF = Initial Approach Fix; FAF = Final Approach Fix.
MSL = Mean Sea Level; DME = Distance Measuring Equipment

experiment included a total of 64 trials across all pilots (8 pilots × 8 trials) and 256 observations (64 trials × 4 legs) on measures.

2.4. Dependent Measures

Dependent variables were flight path control performance and pilot SA. Performance was assessed in terms of deviation errors (root mean square error [RMSE]) in the tracking task (tracing the FPM in the prerecorded video with the superimposed cursor). The RMSE was calculated based on the positional (pixel) differences between the cursor and FPM in the Java flight simulation display. Greater deviations in tracking control were an indicator that participants allocated less attention to the pathway tunnel image and the FPM to view other features of the HUD. It should be noted here that an analysis of FPM position changes in the original IFD videos revealed no significant differences in tracking task difficulty among experimental conditions (HUD configuration, visibility condition, and leg). Therefore, higher RMSEs between the superimposed cursor and FPM were considered to represent pilot attentional distractions leading to higher flight path control error.

Pilot SA was measured using the Situation Awareness Global Assessment Technique (SAGAT; Endsley, 1995) to assess the display, visibility, and leg affects on the three levels of pilot SA (perception, comprehension, and projection). SAGAT was selected among SA measures because prior investigations of the effect of SVS and EVS technologies used subjective measures and SAGAT is validated operational measure grounded in SA theory. Although the subjective SA measurement techniques used in previous studies, such as the Situation Awareness-Subjective Workload Dominance Technique (SA-SWORD; Prinzel et al., 2004) and the Situation Awareness Rating Technique (SART; Bailey

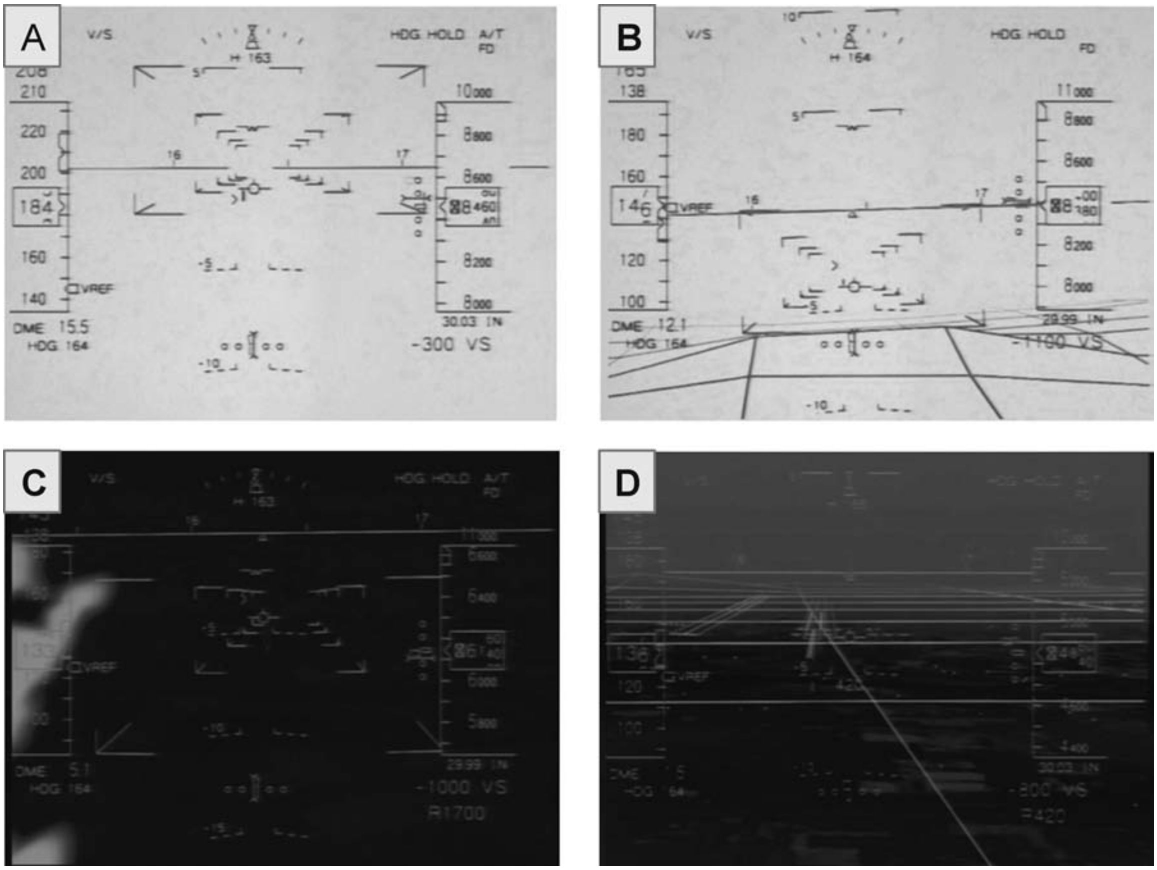


Figure 6 Examples of (A) Baseline-day for Leg 1, (B) SVS-day for Leg 2, (C) EVS-night for Leg 3, and (D) Combo-night for Leg 4.

et al., 2007, Schnell et al., 2005), provide the advantage of ease of implementation, they pose a number of limitations. Endsley (1995) said that rating methods may be affected by subject performance on trials and that direct self-ratings collected at the end of the task may be prone to rationalization and overgeneralization by participants. In addition to this, self-ratings are typically summative measures collected at the end of trials and may be biased by subject memory limitations. In contrast, the 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 in real time, across multiple levels of SA (Endsley, 1995). For these reasons, SAGAT has been identified as a useful measure for evaluating pilot SA (Snow & Reising, 1999) and has been successfully used in several studies (e.g., Bolton et al., 2007). There are, however, some limitations of this methodology, including potential intrusion into operator performance as a result of simulation freezes for administering queries. Related to this, Endsley (1995) observed that expert pilots were able to recall and effectively resume prior simulation performance with freezes lasting as long as 6 minutes. Other limitations of SAGAT include the fact that it conceals or removes display information normally available to participants when freezes occur and questions are posed. Some have suggested that this methodology is akin to a test of working memory versus SA. Jones and Kaber (2004) offered that the approach actually provides an indication of what information elements were present in a participant's state of awareness at the time of a freeze and were considered important relative to task goals. Consequently, SAGAT was used with queries for evaluating pilot SA in the present study.

SAGAT is considered an online questionnaire methodology. Participant responses to queries are recorded during freezes, and grading of responses by experimenters occurs based on the ground truth of the simulation. The SA queries used as part of the SAGAT in this study were also targeted at the three types of pilot SA defined by Wickens (2002), including Spatial, System, and Task awareness. The definitions of the

TABLE 2. Sample SA Questions for Types and Levels of Pilot SA

Types	Levels	Questions
Spatial Awareness	**1**	– 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.
	2	– 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?
	3	– From your current position, what is the safest route if forced off the approach by traffic at 12 o'clock? – If you are forced to go-around from your current position, which direction would you need to direct the aircraft?
System Awareness	**1**	– What was your air speed at the time the simulator stopped? – What was your MSL altitude at the time the simulator stopped?
	2	– How far is the aircraft above decision height? – If you want to be descending at –700 FPM (feet per min), should you increase or decrease pitch?
	3	– When will your aircraft intercept the glideslope? – Estimate the airspeed after 10 seconds, if your flight continues as it is now.
Task Awareness	**1**	– Have you received ATC clearance to contact tower? – What was the last voice warning from the terrain awareness warning system?
	2	– How long has it been since you received "clear to land" from tower? – What did you last communicate with your FO?
	3	– What will you next ask your FO? – What do you expect your next ATC clearance?

three types of pilot SA were adapted for the present experiment, including:

- SpA: awareness of non-iconic display information regarding spatial location (which does not require decoding by pilots), such as SVS/EVS terrain features and tunnel or path information;
- 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; and
- Task Awareness: awareness of communication on the flight deck with the FO regarding landing procedures (flaps, landing gear, landing checklist, and landing decision) and communication with ATC regarding clearances (Abbott & Rogers, 1993).

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, Leiden, & Small, 2003). Table 2 presents example SA queries for the three types and three levels of pilot SA. Expert pilots, who were familiar with the SAGAT methodology, were asked to review the queries in the context of the experimental task, as a basis for validation. They evaluated numerous candidate queries and selected those considered appropriate for the categories of SA in the study. For example, regarding SpA, queries were chosen to assess pilot awareness of aircraft orientation with respect to terrain as well as tunnel or path information. Other queries were also posed regarding potential evasive maneuvers to determine whether the specific features of the HUDs supported pilot SpA and maintenance of safe flight routes. During each trial, the simulator was frozen at random points in time in each leg. Pilots were asked to complete an SA questionnaire with nine queries (one query for each of the three levels by three types of SA) randomly selected from the overall pool of queries. While pilots filled out the questionnaire, the video display was blanked. The answers to each query depended on the ground truth of the simulation when it was frozen at random points in time under the various experimental conditions. The criteria for grading the SAGAT queries were also validated by the expert pilots involved in the study.

2.5. Specific Hypotheses

Based on 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 vs. IMC-night), as well as for each of the four distinct legs of the approach and landing ((Hypothesis 1 [**H1**]). 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 (Kaber et al., 2009). 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, 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 (Byrne et al., 2004). 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 because of the need to manipulate instruments and in Leg 4 because of 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 SVS- and EVS-featured display configurations were expected to be greater than the Baseline and the Combo conditions (**H2-1**) because SVS- or EVS-only features may enhance SA beyond baseline because of terrain information (Prinzel et al., 2003; Snow & Reising, 1999), whereas the combination may degrade SA because of pilot distraction due to clutter effects and higher display density (Kaber et al., 2009). 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**).

3. RESULTS AND DISCUSSION

3.1. Flight Path Control Performance

Based on residual analysis, a double log transformation was applied to the RMSE response variable to satisfy the statistical assumptions of analysis of variance (ANOVA). The ANOVA results revealed a significant effect of display configuration on RMSE ($F(3{,}223) = 13.96$, $p < .0001$). A post hoc analysis using Duncan's multiple range tests 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 did the EVS condition. In general, the EVS generated lower RMSEs, the Baseline and Combo were comparable to each other, and the SVS induced the greatest flight path control errors. This finding was not in line with H1-1, which stated that path deviation errors would be lowest with the Baseline display and then increase with the SVS, EVS, and Combo display. This result also implies that an SVS-featured HUD may increase path deviation errors by approximately 40 percent, as compared to an EVS-featured HUD. These findings were explained by pilot comments recorded during the "think aloud" session. Grid lines used depict terrain features generated by the SVS were often confused by pilots with the FPM and the tunnel features, which were constructed with similar lines. This confusion diverted pilot attention from the FPM in attempting to discriminate other features from the SVS imagery and caused higher tracking errors. Opposite to this, the thermal returns (e.g., moisture images) from the EVS did not produce pilot confusion in visualizing other HUD features. Instead, the features appeared to compel pilots to focus more on the FPM in the display, and this yielded lower tracking error. Because the SVS increased RMSEs and the EVS decreased RMSEs, it made sense that the Combo condition was comparable to the Baseline display configuration in terms of performance, which did not include any terrain features.

ANOVA results also revealed a significant effect of the IMC condition on tracking performance

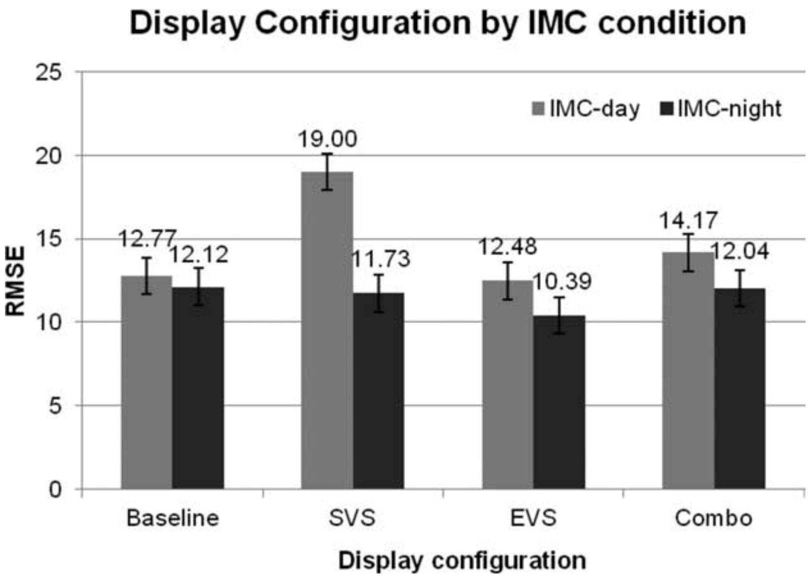


Figure 7 RMSEs for display configuration vs. IMC condition.

($F(1{,}223) = 56.85$, $p < .0001$). The IMC-day condition ($M = 14.3$) was associated with higher deviations in the tracking task was than the IMC-night condition ($M = 12.0$). This result was in agreement with H1 and H1-2. However, ANOVA results also revealed an interaction effect among the display and IMC settings ($F(3{,}223) = 8.12$, $p < .0001$). Figure 7 shows the interaction plot indicating that the SVS-HUD under the IMC-day condition produced significantly higher RMSE than the same HUD under the IMC-night condition. These results imply that a pilot using the advanced HUD under IMC-day conditions may generate approximately 19 percent greater path control errors than under IMC-night conditions. In particular, the use of the SVS-HUD under IMC-day may degrade path-tracking performance by roughly 83 percent as compared to the EVS-HUD under IMC-night. The bright daylight with dense cloud cover produced a display background that substantially reduced the salience of the dynamic symbology, including the FPM, and degraded tracking performance. It is possible that pilot confusion of display features in using the SVS was further magnified by the low salience of the FPM.

There was no significant effect of leg on RMSE ($F(3{,}223) = 0.94$, $p = .4236$). This finding was not in line with H1-3. However, ANOVA results revealed a significant interaction effect between display and the leg of the flight on RMSE ($F(9{,}223) = 2.49$, $p = .0098$). Figure 8 presents the RMSEs for each display configuration during 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 all other legs. The use of both SVS and EVS features with actual terrain (or runway) features visible in the out-of-cockpit view caused higher display clutter than the other display configurations. During Leg 4, this clutter effect may have distracted pilots from focusing on flight path control to the runway.

The effects of trial order and test display condition order on flight path control were also tested. Results revealed no significant effects of trial order (1–8) or increasing or decreasing order of display (pixel) density on the RMSE response measure.

3.2. Pilot SA

Because responses to SAGAT queries represent a binary variable (correct or incorrect), data transformation using the arcsine function was applied to the percentage of correct responses for each SA query, as suggested and validated by Endsley (1995). ANOVAs were then conducted on the transformed SA scores for overall SA, SA by levels, and SA by types. ANOVA results revealed a significant effect of HUD configuration on overall SA scores ($F(3{,}224) = 3.04$, $p = .03$), which was expected in the general hypothesis on SA (H2). Contrary to our Hypothesis 2-1, post hoc analysis revealed SA to be higher for SVS ($M = 59.7\%$) and lower for EVS ($M = 49.8\%$) conditions, and the effect of each of them was not different from the Baseline ($M = 55.5\%$) and Combo ($M = 54.5\%$) conditions.

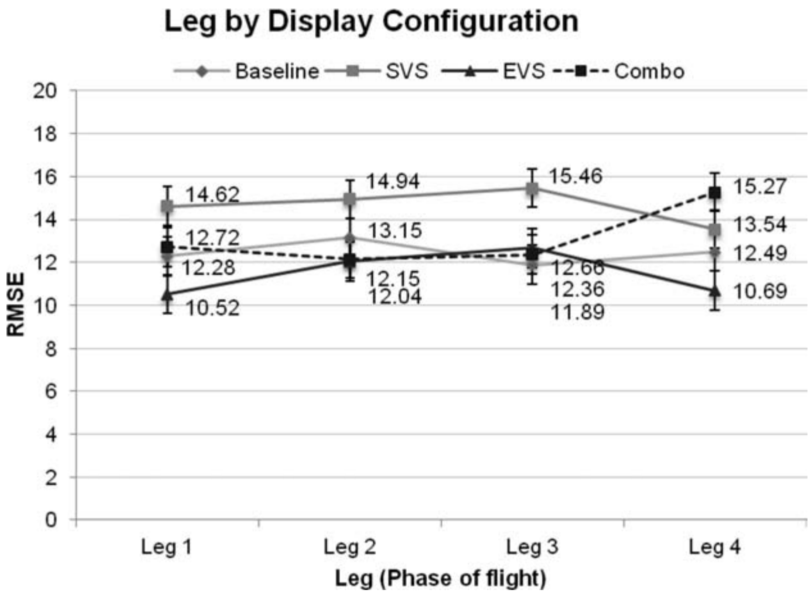


Figure 8 RMSEs for leg by display configuration.

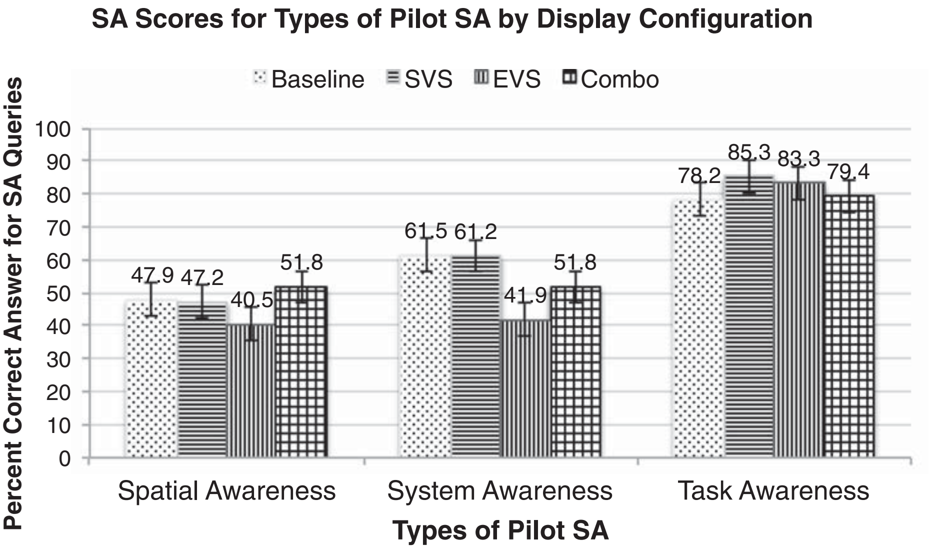


Figure 9 SA scores for types of pilot SA by display configuration.

Specifically, the use of EVS significantly degraded system awareness ($F(3{,}224) = 4.55$, $p = .0041$) (see Figure 9), which concerns pilot understanding of iconic information in the display. This means that the pilot awareness on system information when using the EVS-HUD may be up to 32 percent less than when using the SVS-HUD. Decrements in SA while using the EVS may be attributable to: (1) the thermal features frequently washing out iconic features presenting system information; and (2) the pilot increasing 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 finding suggests a "cognitive tunneling" effect due to the presence of infrared features from the EVS. It is likely that the SVS condition produced higher SA, as the three-dimensional graphical model drew pilot attention away from the FPM group and to other aircraft status indicators at the periphery of 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 with some results being contradictory in nature. While

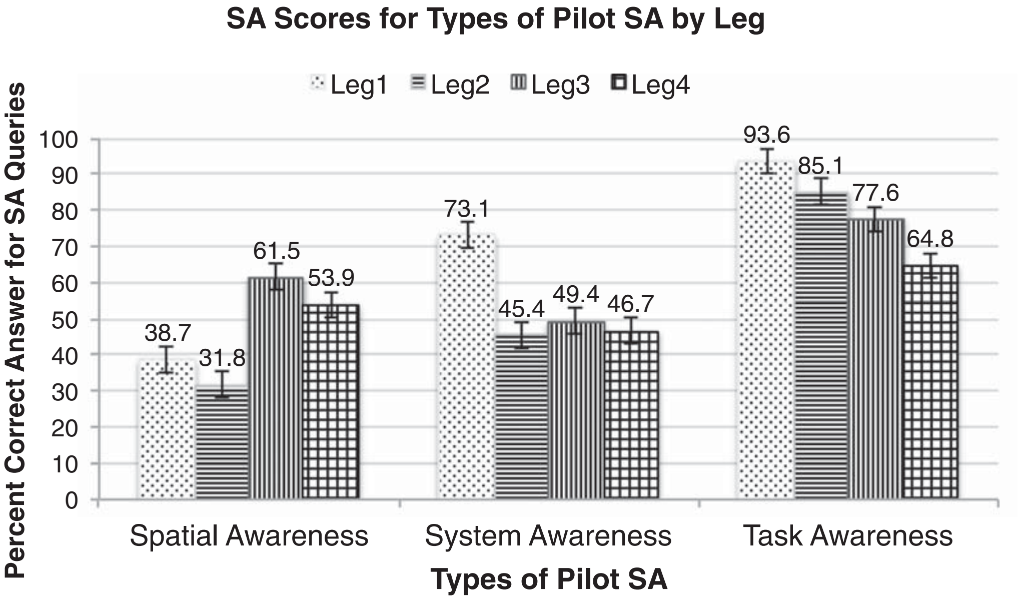


Figure 10 SA scores for types of pilot SA by leg.

IMC-night was associated with higher Level 2 SA scores ($F(1{,}224) = 5.99$, $p = .0151$) (in line with H2-2), in general, it induced lower system awareness scores ($F(1{,}224) = 3.99$, $p = .0469$) (not in line with H2-2). This result suggested that night flying increased pilot comprehension of spatial information on cockpit communications but decreased understanding of aircraft system information via the HUD. In general, mean SA 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 H2-3. Legs 1 and 3 were associated with higher overall SA and Level 3 SA than were Legs 2 and 4 ($F(3{,}224) = 9.75$, $p < .0001$). Results on SA scores by the three types of pilot SA provided more detailed information for the legs of flight. As shown in Figure 10, SpA was higher in Legs 3 and 4 than in Legs 1 and 2 ($F(3{,}224) = 9.37$, $p < .0001$). Pilots had greater understanding of spatial information when 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). System awareness was higher in Leg 1 because of more flight tasks requiring pilots to check system information using iconic features in the HUD. Pilots had to slow the aircraft to approach speed (210 kts to 138 kts) by controlling the throttles and monitoring the HUD. They also manipulated the flaps and landing gear controls depending on speed and DME (distance to runway) (Keller et al., 2003), with frequent confirmation of this information through the HUD and FO. Therefore, the importance of the system information caused pilots to achieve high levels of awareness in Leg 1. Another potential reason for the higher system awareness in Leg 1 might have been due to pilots attempting to orient to the current state of the system at the beginning of each test trial. Pilots are trained to first check the state of the flight before turning to navigation and other types of tasks. SA scores for task awareness by leg were related to the number of tasks occurring in each leg; that is, pilot task awareness decreased as the flight approached on the runway in order for the pilot to concentrate on flight maneuver for landing. The higher scores for task awareness, as compared with system awareness or SpA, might have been because the tasks for landing were identical across trials regardless of display conditions and the pilots were highly trained in landing procedures.

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 induced differences in system awareness among the different legs of flight ($F(9{,}224) = 2.31$, $p = .0167$). Although the use of the SVS was associated with higher system awareness in Legs 1 and 2, the Baseline configuration was superior in Legs 3 and 4, where the actual terrain and runway could be seen through the out-of-cockpit view (see Figure 11). This finding may be attributable to display clutter as synthetic terrain

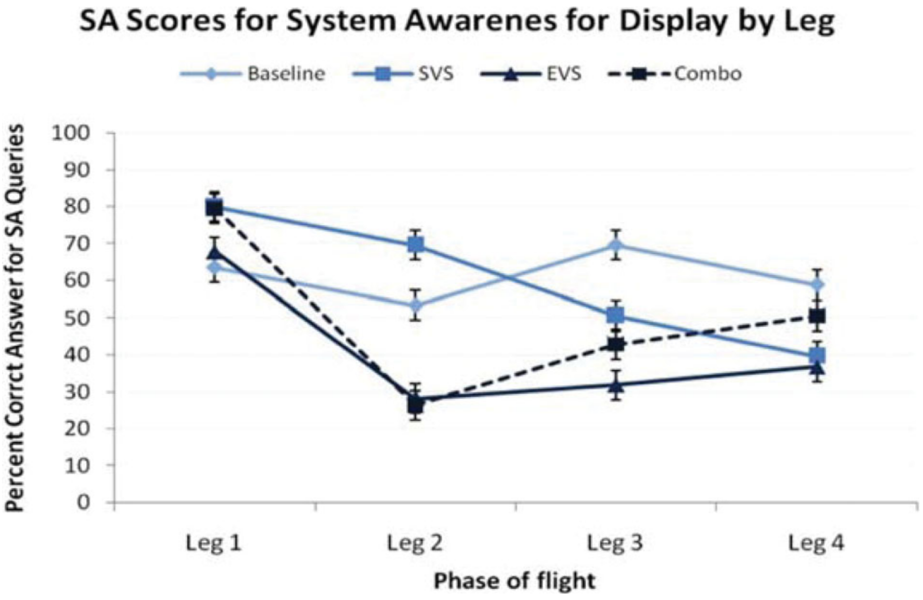


Figure 11 SA scores for system awareness by display configuration versus leg of flight.

features overlapped the out-of-cockpit view degrading pilot understanding of system information in the display. Second, the visibility conditions caused differences in Level 3 SA and system awareness in specific legs of flight. Although IMC-night produced higher overall SA scores than IMC-day did, IMC-day was associated with higher Level 3 SA and system awareness in Leg 3. This result was likely due to visibility in the out-of-cockpit view below the ceiling being greater for IMC-day than IMC-night. Pilot awareness of the runway in Leg 4 was not significantly different between the two IMCs because of the runway lights at nighttime. It is possible that the visible out-of-cockpit features under the IMC-day condition yielded higher pilot system awareness in Leg 3.

Although the effects of most experimental manipulations were significant for system awareness, contrary to expectation, pilot SpA and Task awareness were not affected by display configuration and visibility condition. Only the segment of flight was found to affect SpA because of visibility of actual terrain images through the out-of-cockpit view. Segment also affected task awareness because of the number of tasks differing among legs. This finding suggests that use of the non-iconic conformal features (SVS and/or EVS) examined in this study may be less effective for providing pilots with spatial information as compared to facilitating SA on system status information.

As with flight performance, the effects of trial order and test display condition order on pilot SA were also tested. Results revealed no significant effects of trial order (1–8) or increasing or decreasing order of display (pixel) density on the various SA responses.

4. CONCLUSION

This study assessed the effects of terrain features (SVS and/or EVS) presented in an advanced HUD on pilot flight control and SA under different visibility conditions (IMC) and legs of a simulated approach and landing scenario. In general, the SVS display (graphical terrain model based on GPS data) increased overall SA but degraded flight path control. The EVS display (thermal imagery of actual terrain) increased flight performance with decrements in SA, especially comprehension of system information, because of pilot distraction with the novel HUD feature. That is, visual confusion because of the similarity of SVS grid lines for depicting terrain features, tunnel features, and the FPM group might have caused pilots to focus more on display areas for system information rather than on path control under limited attentional capacity. Opposite to this, visual distraction from thermal features in the EVS condition, overlapping system information, might have caused pilots to focus on flight control tasks rather than on gathering and comprehending system information. This may imply a cognitive trade-off in strategic attention allocation between flight control and system information acquisition when using SVS and EVS technologies in HUDs. Because each system has certain pros and cons, the Baseline and Combo displays may appear comparable in terms of pilot performance under identical flight circumstances. When real terrain and the runway were visible through the out-of-cockpit view (technically VMC), pilots' SpA increased as they were able to see the terrain with the naked eye; however, the terrain information from the SVS and/or EVS created display clutter issues and led to degradations in flight path control, especially upon landing when SVS and/or EVS features overlapped the actual terrain scene. This finding was supported by pilot comments recorded during the "think aloud" session. Many suggested that the SVS/EVS terrain features not be presented after the runway was visible in the out-of-cockpit view because of the potential to create display clutter effects.

Although the presence of the SVS degraded flight path control, if an autoflight mode of control was used, this might serve to offset any performance problems. Consequently, greater pilot SA could be achieved than under other display configurations with the same flight path control performance. However, the advantage of the SVS in generating higher SA would apply only to Legs 1 and 2 of the approach, where the actual out-of-cockpit view is not visible. As revealed through the experiment, when a pilot is able to see the actual terrain scene with the naked eye, aircraft system awareness with the SVS is worse than with the Baseline configuration.

Although the pathway features were displayed in the HUD, pilots did not regard the terrain features as critical information in terms of flight safety. Even though previous studies (e.g., Snow & Reising, 1999 and Schnell et al., 2005) demonstrated that SVS improved flight performance, the improvement might be attributable to the flight pathway feature included in SVS as suggested by Alexander et al. (2003) and Wickens et al. (2004). The study by Schnell et al. (2004) also demonstrated that pilots relied on and trusted the pathway to the extent that they did not feel the need to devote much attention to the aircraft–terrain situation. However, the present study assessed the pure effects of terrain features of SVS and EVS when the pathway feature was presented in all display configurations.

To assess pilot SA in the experiment, an elaborate adaptation of SAGAT was used. The online questionnaires included queries targeting the three levels of SA identified by Endsley (1995) as well as the three types of pilot SA defined by Wickens (2002). This approach proved sensitive and useful for explaining the effects of HUD configurations, visibility conditions, and phase of flight on pilot information processing. Interaction effects of display by leg and visibility condition by leg, which were not apparent in overall SA scores, were revealed through pilot projection of future aircraft states and system awareness. Consequently, formal development of an extended version of SAGAT for measuring SA at different levels of cognitive processing in different types of (aviation) tasks may provide a useful framework for additional research on NextGen cockpit technologies.

The caveats of this study include use of a low-fidelity flight simulator in the lab experiment. Prerecorded videos of HUD content were played for pilots using a PC-based simulator. Pilots were asked to perform a tracking task based on the video stimuli, instead of actually controlling the simulated aircraft. Flight path control performance was inferred based on errors in the tracking task. That said, tracking deviations were indicative of degraded pilot attention to flight guidance, which would ordinarily lead to flight technical errors in real operations. It is possible that pilots would demonstrate different behaviors under real flight circumstances, as compared to the findings in the present study. The simulator setup also limited pilot information sources. Only information presented via the HUD was used for the simulated flight. In a real flight context, pilots rely on many cockpit panel displays, HDDs, or HUDs. Pilots integrate various information across these displays to achieve and maintain SA. This may result in a different performance than the results of the present experiment. Although pilots were specifically exposed to the Baseline and Combo display conditions during training, this did not prove to bias the test trial results in any way. Results revealed no significant effects of trial order and test display condition order on flight path control and pilot SA.

The flight scenario used in this study did not involve critical events that can happen in real flight or provide the opportunity to assess the utility of terrain features in such situations. For example, the effects of SVS/EVS features may be different in flight under non-nominal conditions, such as when a pilot must go-around at decision height or when there is a runway incursion.

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. In the present experiment, there were effects of HUD features during specific legs of flight (in approach and landing), which led to differences in pilot performance. The effects of SVS and EVS features under various flight situations should be further evaluated. In addition, this type of research should be conducted in a more realistic simulator or real aircraft and in a more realistic flight context.

In addition, it would be interesting to develop a flight context–based "Adaptive HUD" interface by considering the results of this study. Although it may be possible for pilots to manipulate display configurations to achieve the most appropriate feature combination under particular flight conditions, this may cause additional cognitive workload, particularly in high time-stress situations. With this in mind, it would be desirable to present pilots with predetermined optimal display feature sets according to dynamic changes in a flight. Beyond this, the use of an elaborate SAGAT method for pilot SA measurement is worth exploring in other empirical research on such adaptive display concepts. The novel approach used to assess pilot SA in the present study at three levels of cognition and in terms of three types of tasks may provide valuable detailed information for interpreting pilot performance and as a basis for refining display design.

ACKNOWLEDGMENT

This research was completed while the first author, Sang-Hwan Kim, worked as a research assistant at North Carolina State University. This research was supported by NASA Langley Research Center under Grant No. NNL06AA21A. The opinions and conclusions expressed in this article are those of the authors and do not necessarily reflect the views of NASA.

References

Abbott, T. S., & Rogers, W. H. (1993). Functional categories for human-centered flight deck design. In Proceedings of the 12th Digital Avionics Systems Conference (pp. 66–74), October 25–28, 1993, Fort Worth, Texas.

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.

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), 5802, 25–36.

Bailey, R. E., Kramer, L. J., & Prinzel, L. J. (2007). Fusion of synthetic and enhanced vision for all-weather commercial aviation operations. In North Atlantic Treaty Organization (NATO) Human Factors and Medicine Symposium on Human Factors and Medical Aspects of Day/Night All Weather Operations: Current Issues and Future Challenges, Tech. Rep. No. NATO RTO-HFM-141, Neuilly-sur-Seine, France, RTO, pp. 11–1 -11–18.

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.

Byrne, M. D., Kirlik, A., Fleetwood, M. D., Huss, D. G., Kosorukoff, A., Lin, R., et al. (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.

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

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.

Jones, D. G., & Kaber, D. B. (2004). Situation awareness measurement and the situation awareness global assessment technique (Chapter 42). In N. Stanton (Eds.), Handbook of Human Factors Methods. London: Taylor & Francis, 419–428.

Kaber, D. B., Alexander, A., Kaufmann, K., Kim, S.-H., Naylor, J. T., & Entin, E. (2009). Testing and validation of a psychophysically defined metric of display clutter (Final Report: NASA Langley Research Center Grant. #NNL06AA21A). Hampton, VA: NASA Langley Research Center.

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.

Kim, S.-H., Prinzel, L. J., Kaber, D. B., Alexander, A. L., Stelzer, E. M., Kaufmann, K., et al. (2011). Multimodal measure of display clutter and pilot performance for advanced head-up display. Aviation, Space, and Environmental Medicine, 82(10), 1013–1022.

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.

Prinzel, L. J., Arthur, J. J., Kramer, L. J., & Bailey, R. E. (2004). Pathway concepts experiment for head-down synthetic vision displays. In Proceedings of the SPIE, Enhanced and Synthetic Vision 2004. Vol. 5424-02. Jacques G. Verly (Ed.). Orlando, Florida.

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 Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting. Santa Monica, CA: Human Factors and Ergonomics Society.

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, October 30 – November 3, 2005, Washington, D.C. 2005.

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.

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.

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.

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.

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

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.

Wickens, C. D., & Hollands, J. G. (2000). Engineering psychology and human performance (3rd ed.). New Jersey: Prentice Hall.