NASA/TP–2008–215321

Head-Worn Display Concepts for Surface Operations for Commercial Aircraft

Jarvis (Trey) J. Arthur III, Lawrence J. Prinzel III, Randall E. Bailey, Kevin J. Shelton, Steven P. Williams, Lynda J. Kramer and Robert M. Norman

June 2008

The NASA STI Program Office . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA Scientific and Technical Information (STI) Program Office plays a key part in helping NASA maintain this important role.

The NASA STI Program Office is operated by Langley Research Center, the lead center for NASA's scientific and technical information. The NASA STI Program Office provides access to the NASA STI Database, the largest collection of aeronautical and space science STI in the world. The Program Office is also NASA's institutional mechanism for disseminating the results of its research and development activities. These results are published by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers, but having less stringent limitations on manuscript length and extent of graphic presentations.
- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.
- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.
- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.
- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.
- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services that complement the STI Program Office's diverse offerings include creating custom thesauri, building customized databases, organizing and publishing research results . . . even providing videos.

For more information about the NASA STI Program Office, see the following:

- Access the NASA STI Program Home Page at ***http://www.sti.nasa.gov***
- E-mail your question via the Internet to help@sti.nasa.gov
- Fax your question to the NASA STI Help Desk at (301) 621–0134
- Phone the NASA STI Help Desk at (301) 621–0390
- Write to:
  NASA STI Help Desk
  NASA Center for AeroSpace Information
  7115 Standard Drive
  Hanover, MD 21076–1320

NASA/TP–2008–215321

Head-Worn Display Concepts for Surface Operations for Commercial Aircraft

Jarvis (Trey) J. Arthur III, Lawrence J. Prinzel III, Randall E. Bailey, Kevin J. Shelton, Steven P. Williams, Lynda J. Kramer and Robert M. Norman

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

June 2008

Acknowledgments

The authors would like to acknowledge the contributions of the technical review committee for their comments in improving this paper. Denise R. Jones (chair), Monica H. Barnes and Elliot (Tod) T. Lewis

Available from:

NASA Center for AeroSpace Information (CASI)
7115 Standard Drive
Hanover, MD 21076–1320
(301) 621–0390

National Technical Information Service (NTIS)
5285 Port Royal Road
Springfield, VA 22161–2171
(703) 605–6000

Contents

List of Tables

List of Figures

Nomenclature

ADS-B	Automatic Dependent Surveillance-Broadcast
ATCT	Airport Traffic Control Tower
AMASS	Airport Movement Area Safety System
ANOVA	Analysis of Variance
ATC	Air Traffic Control
AvSP	Aviation Safety Program
CFIT	Controlled Flight Into Terrain
CRM	Crew Resource Management
dB	Decibels
EFB	Electronic Flight Bag
EMM	Electronic Moving Map
EP	Evaluation Pilot
EV	Enhanced Vision
EVO	Equivalent Visual Operations
FAA	Federal Aviation Administration
FLIR	Forward Looking Infrared
FOV	Field-of-View
ft	Foot/Feet
FTE	Flight Technical Error
HDD	Head-Down Display
HMD	Helmet-Mounted Display
HUD	Head-Up Display
HWD	Head-Worn Display
Hz	Hertz
ICAO	International Civil Aviation Organization
IIFD	Integrated Intelligent Flight Deck
ILS	Instrument Landing System
LaRC	Langley Research Center
MANOVA	Multivariate Analysis of Variance
msec	Millisecond
NAS	National Air Space

NASA-TLX	NASA Task Load Index
ND	Navigation Display
NextGen	Next Generation Air Transportation System
nmi	Nautical Mile
NRA	NASA Research Announcement
NTSB	National Transportation Safety Board
ORD	FAA Identifier for Chicago O'Hare International Airport, Illinois
PFD	Primary Flight Display
RFD	Research Flight Deck
RIPS	Runway Incursion Prevention System
RMS	Root Mean Square
RNO	FAA Identifier for Reno/Tahoe International Airport, Nevada
RNP	Required Navigation Performance
RVR	Runway Visual Range
SA	Situation Awareness
SAGAT	Situation Awareness Global Assessment Technique
SART	Situation Awareness Rating Technique
SA-SWORD	Situation Awareness Subjective Workload Dominance
SSQ	Simulator Sickness Questionnaire
SBIR	Small Business Innovative Research
S/EVS	Synthetic / Enhanced Vision Systems
SNK	Student-Newman-Keuls
SV	Synthetic Vision
SVS	Synthetic Vision System
SWORD	Subjective Workload Dominance
T-NASA	Taxiway-Navigation And Situation Awareness
TSE	Total System Error
VFR	Visual Flight Rules
VISTAS	Visual Imaging Simulator for Transport Aircraft Systems
VMC	Visual Meteorological Conditions
VRS	Voice Recognition System
2-D	2-Dimensional
3-D	3-Dimensional
4-D	4-Dimensional

Abstract

Experiments and flight tests have shown that a Head-Up Display (HUD) and a head-down electronic moving map (EMM) can be enhanced with Synthetic Vision for airport surface operations. While great success in ground operations was demonstrated with a HUD, the research noted that two major HUD limitations during ground operations were its monochrome form and limited, fixed field-of-regard. A potential solution to these limitations found with HUDs may be emerging with Head Worn Displays (HWDs). HWDs are small display devices that may be worn without significant encumbrance to the user. By coupling the HWD with a head tracker, unlimited field-of-regard may be realized. The results of three ground simulation experiments conducted at NASA Langley Research Center are summarized. The experiments evaluated the efficacy of head-worn display applications of Synthetic Vision and Enhanced Vision technology to improve transport aircraft surface operations. The results of the experiments showed that the fully integrated HWD provided greater pilot performance with respect to staying on the path compared to using paper charts alone. Further, when comparing the HWD with the HUD concept, there were no differences in path performance. In addition, the HWD and HUD concepts were rated via paired-comparisons the same in terms of situation awareness and workload.

1 Introduction

The Integrated Intelligent Flight Deck (IIFD) project of NASA's Aviation Safety Program (AvSP), comprises a multi-disciplinary research effort to develop flight deck technologies that mitigate operator-, automation-, and environment- induced hazards. Toward this objective, the IIFD project is developing crew/vehicle interface technologies that reduce the propensity for pilot error, minimize the risks associated with pilot error, and proactively overcome aircraft safety barriers that would otherwise constrain the full realization of the Next Generation Air Transportation System (NextGen) [1]. Part of this research effort involves the use of synthetic and enhanced vision systems and advanced display media as enabling crew-vehicle interface technologies to meet these safety challenges.

While NextGen concepts envision the capability to handle up to a 3-fold increase in air traffic, the National Transportation Safety Board (NTSB) continues to have runway incursion prevention on its top six most-wanted list for aviation safety [2]. In the 4-year period between 2001 and 2004, 1,395 runway incursion events were reported to the FAA which is a rate of almost 1 runway incursion event per day [3]. Also during this time, over 60% of the FAA towers reported at least one runway incursion event. These statistics and events are cause for alarm. The worst aviation accident in terms of fatalities occurred in 1977 when two fully loaded Boeing 747 airplanes collided on a runway at Tenerife airport. Moreover, each year there are reports of close "near-miss" runway incursions that happen with sufficient regularity at the world's busiest airports to pose perhaps the most significant hazard confronting aviation today.

One such airport plagued with runway incursions is Chicago O'Hare International Airport (FAA identifier: ORD). Chicago O'Hare is a complex airfield for surface operations and represents one of the world's busiest and most challenging airports for surface operations. Current runway incursion safety mitigations employ a "layered" approach using technology, training, and awareness. The ORD airport authority has identified "hot spots" which are areas where incursions are likely to occur. Within these areas, special ground traffic and aircraft handling are designed so that nominal operations minimize incursion potential. These hot spots are published and disseminated to aircraft and ground crew operators to heighten vigilance when operating in and near these areas. ORD operates an Airport Movement Area Safety System (AMASS) which provides warnings of runway incursions to controllers. Even with these protocols and technology implementations, there have been several close calls [3]. For example, there was a runway incursion between two Boeing 747s at O'Hare on April 1, 1999. From the NTSB meeting on June 13, 2000:

> *"On April 1, 1999, just after 2 o'clock in the morning, Korean Air flight 36 and Air China 9018, both Boeing 747s, nearly collided on runway 14 Right at the Chicago O'Hare International Airport. Air China had just landed and was rolling out on runway 14 right when the tower controller instructed Korean Air to taxi into position and hold. After Air China exited the runway at taxiway T10, the tower controller instructed the flight to turn left on taxiway Kilo and cross runway 27 left. The tower controller then cleared Korean Air for takeoff. As the airplane was rolling down the runway, Air China deviated from its assigned taxi route and taxied on to runway 14 Right. The Korean Air captain saw the 747 taxiing on to the runway but it was too late to stop. Instead, Korean Air 36 lifted off earlier than normal and banked left to avoid striking Air China. The two aircraft, carrying 382 people, missed colliding by about 80 feet."*

It should be noted that AMASS was not installed at O'Hare at the time of this incident between Korean Air and Air China. However, using data from flight recorders, investigators at the NTSB were able to play back the data through an AMASS simulator. The results of the simulation showed that the AMASS alert happened approximately 6 seconds before the potential collision. It was determined that 6 seconds was not enough time for the controller to notice the alert, determine the affected aircraft, decide on a course of action and then transmit the instructions on two different radio frequencies (tower and ground).

The potential consequences and the high historical occurrence of such runway incursions prompted NASA to initiate research to reduce the increasing rate of runway incursions. This work has yielded research and development leading to concepts for awareness for the flight crew in two NASA programs, Taxiway-Navigation And Situation Awareness (T-NASA) [4–7]) and Runway Incursion Prevention System (RIPS) [8–12]. Significant improvements in surface operations safety and efficiency were found under the T-NASA and RIPS research and development efforts. To date, implementation of these concepts for US and foreign operators is meeting with varying degrees of success despite the lack of safety mandates by the FAA for equipage.

The present study was an extension of this previous research to evaluate if emerging synthetic vision and head-worn display technologies can provide further safety and operational improvements and to investigate the potential that these concepts may have in smoothing the path to or providing a motivation for voluntary equipage among users. The current airspace system already shows a significant safety problem in surface operations and runway incursions. This problem will only get worse if proactive technologies are not developed soon as the National Air Space (NAS) traffic density grows by as much as 3-fold [1]. For instance, automated surface management systems are being developed that utilize dynamic algorithms to calculate the most efficient movement of all surface traffic in order to increase the efficiency with which airport surfaces are utilized. If these systems are to be implemented, pilots will be required to comply with 4-Dimensional (4-D) taxi clearances, which requires an aircraft to be at a specific location at a specific time. It is prudent to solve the current problems of surface operation safety and efficiency and prepare to accept the new and novel concepts of operation and traffic conditions emerging under NextGen [13].

2 Background

This research primarily involves the confluence and interplay of two technologies - synthetic vision and head-worn displays - for surface operations. Relevant research is summarized in the following to provide background and motivation for the experimental research detailed herein.

2.1 Synthetic Vision System

Previous research has shown that while the capability may be available to take-off and land aircraft in near zero visibility and zero ceiling weather, the operational tempo and safety within the airport terminal area is significantly degraded due to limitations in surface operations. These surface operations include taxiing and maneuvering aircraft and vehicles on the runways, taxiways and aprons.

Synthetic Vision (SV) is a computer-generated image of the external scene topography that is generated from aircraft attitude, high-precision navigation, and data of the terrain, obstacles, cultural features, and other required flight information (Fig. 1). Since 1999, NASA and its industry partners have developed and deployed SV technologies for commercial and business aircraft which have been shown to provide significant improvements in terrain awareness and reductions in the potential for Controlled Flight Into Terrain (CFIT) incidents/accidents compared to current generation cockpit technologies [14, 15].

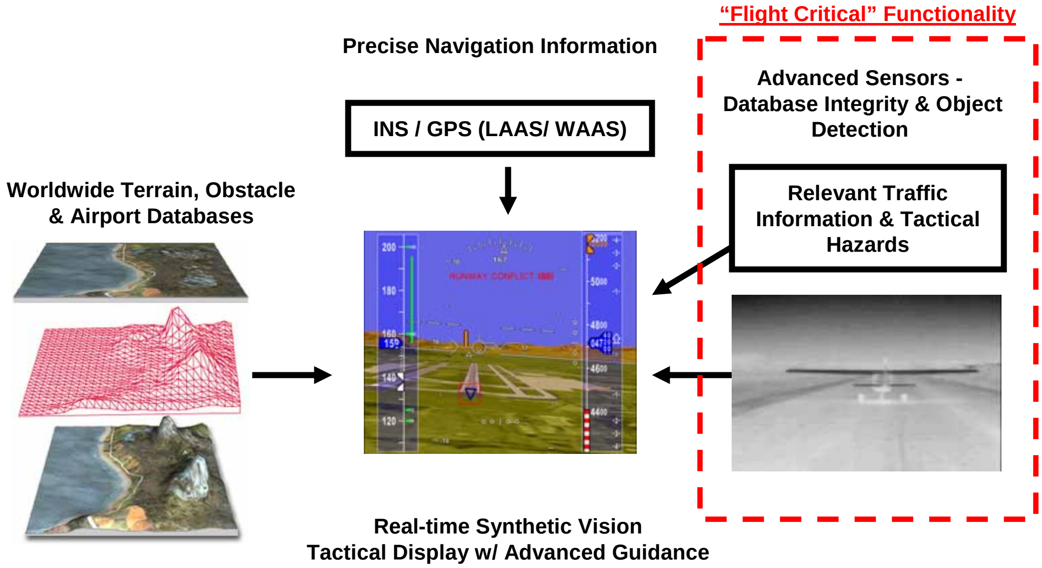


Figure 1. The Synthetic Vision System concept.

In comparison, Enhanced Vision (EV) (or Enhanced Flight Vision System) is an electronic means to provide a display of the external scene through the use of an imaging sensor, such as a Forward Looking Infrared (FLIR) or millimeter wave radar. Both SV and EV are "vision-based" technologies intended to create, supplement, or enhance the natural vision of the pilot.

SV, by virtue of being weather-independent and unlimited in field-of-regard, is particularly advantageous during flight phases which may be obscured by clouds and precipitation that an EV sensor cannot penetrate or when large fields-of-regard are necessary for mission success. The accuracy of the SV image compared to the real-world is dependent upon the position navigation solution accuracy, the attitude sensing accuracy, and the database fidelity, both in terms of how accurately the database information represents the real-world and what information about the real-world may be omitted, intentionally or unintentionally. With SV, the designer controls the SV scene lighting, terrain coloring and virtual camera angles. Unlimited field-of-regard is easily provided in a SV image because there are no physical properties or limitations associated with changing the view-point orientation, direction, or Field-of-View (FOV) within the computer-generated image. In contrast, a camera system's FOV is a physical characteristic of the lens and its placement is dependent on suitable mounting points.

Conversely, EV provides a direct view of the vehicle external environment; independent of the derived aircraft navigation solution or of a database. Very little stands between the EV image shown to the pilot and the real-world; thus, an EV pilot gets an extremely high degree of confidence in the system that is "bolted" to the airframe. The field-of-regard provided by the EV may be limited by the physical problems associated with the installation, alignment, and fusion of multiple EV sensors. Typically, a single EV sensor system is used to provide the FOV equivalent to the Head-Up Display (HUD) (nominally 30°H x 24°V for current transport category HUD installations). The EV image is critically dependent upon the EV sensor characteristics and the corresponding external environment, including the properties of the atmosphere and the properties of the "real-world." Under conditions of smoke, haze, and night, a FLIR/EV, for instance, provides orders-of-magnitude improvement over the pilot's natural vision; greatly enhancing the pilot's situation awareness and reducing the pilot's workload. A long-wave (8-12 microns wavelength) FLIR can generate an outstanding view on a night approach [16], where roads and runways are clearly demarcated from vegetation because of thermal differences. But atmospheric moisture, such as clouds, will be clearly visible and thermal cross-over must always be considered in a FLIR EV operation. Thermal cross-over is the natural phenomenon that normally occurs twice daily when temperature conditions are such that there is a loss of contrast between two adjacent objects on infrared imagery.

Experimental evidence has shown that a "perfect" Synthetic Vision System (SVS), which includes decision aids for alerting for real-time obstacle detection (i.e., finding objects not included in the stored database) and database integrity (i.e., identifying in real-time if the stored database is inaccurate or if the navigation accuracy is not sufficient to support the application) is superior to SV concepts without these decision aides or EV concepts [17]. These findings were validated during a recent flight test comparison against EV technologies [18]. Significant efforts are being devoted to developing real-time database and obstacle detection technologies to support the "perfect" SVS application [19, 20], but there may always be "imperfections" in the technology at least for the near-future (e.g., non-transponding traffic, finite precision in sensors, and low-probability of detection "targets").

The complementary capabilities of SV and EV have been well-recognized [18] with the premise that "the strengths of the enhanced system can compensate for the deficiencies in the synthetic system and that the strengths of the synthetic system can compensate for the deficiencies in the enhanced vision system." [21] Methods and capabilities to create complementary integration and/or fusion of synthetic and enhanced vision technologies have been developed and tested [22]. The optimal fusion of Synthetic / Enhanced Vision Systems (S/EVS) technology is emerging as a cornerstone to the development of advanced fight deck information systems which can provide the flight crew with significantly improved spatial awareness, increased awareness of outside terrain and obstacle features, enhanced manual flight performance, and reduced pilot workload. S/EVS technology can potentially provide unlimited field-of- regard awareness for terrain, obstacles, traffic, and airspace constraints and establish one cornerstone to an Equivalent Visual Operations (EVO) capability [1].

Figure 2. The T-NASA display suite.

2.2 Taxiway-Navigation And Situation Awareness

The T-NASA concept (Fig. 2) was developed to improve the efficiency and safety of airport surface operations in Instrument Landing System (ILS) Category IIIB weather (no decision height, less than 1200 Runway Visual Range (RVR)). T-NASA uses a suite of cockpit displays - a HUD and an Electronic Moving Map (EMM) concept, implemented on a Navigation Display (ND) or Electronic Flight Bag (EFB). The concepts have been shown to provide the following benefits, in various degrees of measure and success [6]:

- Eliminated hold location errors and failure to hold errors
- Allowed increased taxi speeds
- Eliminated taxi navigation errors in low-visibility and night conditions
- Enabled better awareness of airport traffic
- Improved pilot-ATC communication of clearance
- Improved captain / first officer intra-cockpit pilot communication

Under the T-NASA concept, the EMM includes a labeled airport layout, ownship position, positions of other traffic, graphical route guidance, text clearance window, and ground speed and heading indicators. The EMM depicts the cleared taxi route graphically, via a magenta path, and textually, via a text box on the bottom of the map. Hold short instructions are portrayed with a yellow hold bar, and the portion of the route beyond the holding position is displayed in yellow. Airport traffic is depicted in real-time, and pilots can choose to view aircraft icons with or without data tags. All information is dynamic and updated in real-time.

The EMM is designed with the primary purpose of aiding navigation and situation awareness; it is not designed to support the control of the aircraft. As such, the map purposely lacks specific detail regarding the aircraft's position relative to the centerline, location of wheels, speed or braking parameters, or an accurate depiction of aircraft size and wingspan. In contrast, the HUD uses "scene-linked" symbology (Fig. 3) for conformal display against the out-the-window environment (when visible) which theoretically leads to efficient cognitive processing of both the symbology and the environment, and mitigates problems of attentional tunneling and symbology fixation. The taxi symbology contains taxiway centerline markers and taxiway edge cones. Virtual signage aids in augmenting cleared-path awareness. Taxiway information provides enhanced situation awareness for taxi navigation.

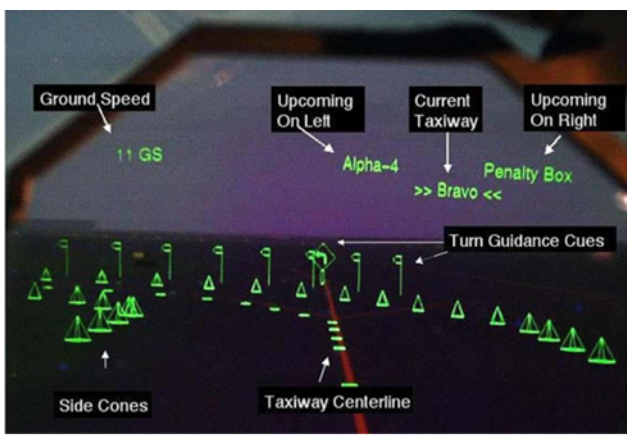


Figure 3. The T-NASA HUD symbology.

Simulation data were analyzed to pinpoint the mechanisms by which T-NASA technology components could mitigate classes of surface operations navigation errors (pilot deviations) [23]. A taxonomy of 3 error classes (planning, decision, and execution) was used. The simulation data replicated current-day operations and also included trials with T-NASA technologies including datalink, EMM concepts, and HUD concepts. The error decomposition showed that pilots committed navigation errors on 17% of current-day operations trials (in low-visibility and night), distributed roughly equally across the three error classes. When using T-NASA technologies, the error data showed a unique set of contributing factors and mitigating solutions:

- Planning errors were mitigated by technologies that provided an unambiguous record of the clearance (datalink and the EMM, which possessed a text-based clearance).

- Decision errors were mitigated by technologies that provided both local and global awareness including information about the distance to and direction of the next turn, current ownship location, and a graphical depiction of the route (as provided by the EMM and HUD together).

- Execution errors were best mitigated by the HUD, which removed ambiguity from the environment and depicted the cleared taxi route.

Further enhancements to the T-NASA concept has evolved based on follow-on research and testing. In particular, tactical turn guidance, in the form of so-called "breadcrumbs" or Taxi Director, have shown to significantly aid in tactical surface operations guidance, particularly for aircraft which require judgmental over-steering to keep the main gear on the taxiway during turns (Fig. 4, [9]). Without non-conformal guidance information, the conformal information such as the centerline and edge markings would not be drawn on the limited HUD field-of-view or the information that was provided would be difficult to interpret.

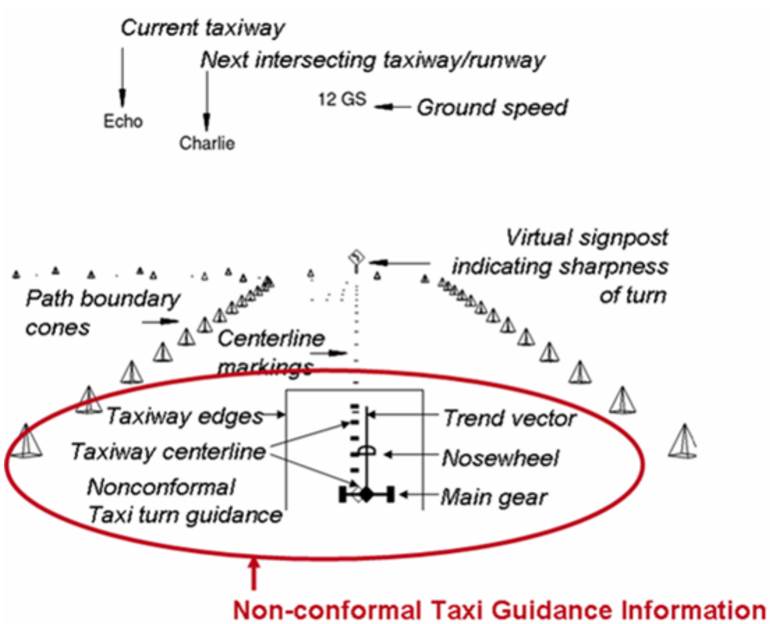


Figure 4. The RIPS HUD concept.

The key to preventing runway incursions is to ensure that pilots know [24]:

- Where they are located
- Where other traffic is located
- Where they are cleared to go on the airport surface

The T-NASA concept and its instantiations contribute significantly toward these elements. However, if the flight crew becomes disoriented, is unaware of close traffic or deviates from the cleared route, the flight crew and Air Traffic Control (ATC) should be alerted to the situation. NASA's RIPS program developed methodologies for flight deck alerting, targeted toward the prediction of runway incursions to provide immediate alerting for the principal participants in the operation (i.e., the flight crews). The T-NASA concept provides guidance and situation awareness information to mitigate many factors contributing toward runway incursions, but a final protective "wrapper" was felt to be warranted nonetheless [12]. Since the objective of the present research was to focus evaluations on proactive surface operations situation awareness, none of the concepts tested included flight deck alerting (i.e., RIPS concepts).

While success in efficiency and safety were measurable with the T-NASA concept, several notable deficiencies [25] have emerged:

- Lack of universal enthusiasm about EMM displays: "I don't want a system that depends on other aircraft, satellites, etc. I want to see a real picture."
- Concern about increasing head-down time during surface ops: "I don't want a display that keeps me heads-down while taxiing. Even at night and in poor weather I see things out the window (lights on other aircraft, runway markers)."
- Questions of data integrity created by a system that might fail to detect and include other aircraft, ramp vehicles and other obstacles on the displays: "I fear that all moving vehicles (e.g., trucks, luggage carts) would be not be displayed."
- Fear that the display will be too cluttered if it captured all moving vehicles.

The T-NASA system showed the importance of taxiway awareness, route awareness, and taxi guidance on head-up and Head-Down Displays (HDDs) to increase taxi times, approaching those achievable in Visual Flight Rules (VFR) operations, and to prevent pilot error in surface operations, including missing or misunderstanding taxi clearances, making wrong turns, and becoming lost on the airport surface. The EMM implemented on a head-down display provided strategic information and textual display for clear, continual, and unambiguous reference to the taxi clearance. The HUD showed great success in ground operations by providing tactical information and allowed the pilot to remain head-up. The HUD presentation of information was critical to enable the pilots to remain head-up, improving the cross-check of information and the outside world, and ensuring that the aircraft remained on the taxiways. However, this research also noted that two of the major HUD limitations during ground operations were its monochrome form and its limited, fixed field-of-regard. A monochromatic display precludes the use of color for information decluttering and information cuing. The display area of a HUD must be carefully designed to provide the pilot with enough information without saturating it with clutter. Further, HUD imagery, while allowing head-out operations, is restricted to its forward FOV and the use of conformal imagery is, consequently, limited. These HUD limitations, coupled with the importance of keeping the aircrew head-up, naturally points to an ideal application for Head-Worn Displays (HWDs) [26, 27].

2.3 Head-Worn Displays

HWDs are small, light-weight, full-color display devices that can be worn on the head without significant encumbrance (Fig. 5). Helmet-Mounted Displays (HMDs) and HWDs are not new technology, particularly for military operations, but component miniaturization and maturation are progressing to the point where HWDs can be considered in commercial and business aircraft operations (i.e., the costs are reaching affordable levels and their use should be as unobtrusive to the pilot as wearing sunglasses). Advances in display devices (e.g., head-worn devices) have been studied by NASA researchers as an alternate and practical method for delivering SVS concepts to the cockpit [26]. By coupling the HWD with a head tracker, unlimited field-of-regard can be realized. Unlike fixed FOV sensors, the camera position and orientation for Synthetic Vision can be defined via software; thus, an unlimited field-of-regard is achieved since the Synthetic Vision scene is viewable from any virtual camera angle.

NASA recently conducted a preliminary evaluation of a HWD as a viable technology to support an EVO concept [28]. The HWD provided a full color, 40° FOV, full-overlapped binocular display and was coupled with a head-tracker. Approach and landing operations, commensurate with commercial airline operations and procedures, were evaluated. Technical difficulties with the HMD negatively influenced the results, but nonetheless, these data showed the promise of the technology, the pitfalls for commercial applications, and the potential that new, emerging technologies may provide. Subsequent to this work, a trade study was conducted by NASA to assess the emergence of small, very lightweight HWDs and to begin the build-up of system level requirements to meet commercial and business aircraft applications. This work, as well as

Figure 5. A Synthetic Vision enhanced HWD with a head tracker has unlimited field-of-regard.

other HWD research, is being used by NASA to evaluate the potential of HWDs for emerging NextGen operating concepts, such as EVO.

The previous work in leveraging HWD technology as an enabler of EVO focused on approach and landing operations. While this work is relevant and more studies will continue, as previously stated, the worst civilian aircraft accident in history occurred during the collision of two Boeing 747 aircraft at Tenerife during runway/surface operations in fog. Further, the NTSB has included runway incursion on its “most-wanted list” of transportation safety improvements since the list began in 1990 [2].

In fact, Rediess [29] proposed a HWD concept and developed an architecture under a Small Business Innovative Research (SBIR) contract [30]. Similar concepts have been proposed for Air Traffic Controllers. The work by Rediess, however, deviated from the present work and did not address some very key elements:

- The encumbrance of the HWD to the commercial and business aircrew is paramount to success. Without unobtrusive technology, the application is likely unacceptable for commercial aviation use.
- Synthetic and Enhanced Vision Technologies have advanced to the point where these systems are not only realizable, but extremely capable and beneficial.
- Latency is critical [31]. Any potential that latency may induce visual illusion or spatial disorientation will doom the application.

In addition to the proposed HWD, Rediess did significant work involving datalink technology for communications between ground control and the aircrew. While datalink technology is clearly the path to the future, the importance of aural communications and its criticality in today's operations cannot be dismissed. This communication modality should be utilized even in the future since the visual modality can be overloaded and voice communications may provide an avenue for error detection and enhanced situation awareness. NASA is investigating methodologies and technologies which capitalize on datalink communications. The goal is to ensure that modality changes retain the good features of previous auditory channels (e.g., retaining "party-line" situation awareness), maximize the benefit of new technologies (e.g., improved non-native English language communications, automatic route entry into a Flight Management System) and create new and improved functionalities (e.g., a "Culture-Neutral/Language-Neutral" flight deck) [32]. NASA has filed a patent application [33] for a crew-vehicle interface system concept that couples a HWD and speech interface system for greatly improved surface operations, particularly in single-crew operations.

A plethora of human factors issues are brought into the design and usage of HWDs that must be carefully considered [34, 35]. Since encumbrance - essentially the weight and volume of the display device affixed to the pilot's head - has been identified by NASA as a key component to acceptability of HWDs for commercial and business aircraft operations, two of the first HWD design parameters considered in this work are the FOV provided by the HWD and whether the HWD provides biocular (or binocular) or monocular viewing. The weight and size of the HMD (or HWD) is nominally in proportion to the FOV provided. The weight and complexity of the HWD is greatly increased as a design changes from monocular to biocular.

Normal, unaided human vision consists of an overlapped divergent binocular FOV where the overall horizontal FOV is approximately 200° of visual angle (when viewed straight ahead), with each eye's monocular field around 120° of visual angle [36]. The two partially overlapping monocular fields produce a FOV consisting of three regions:

- A central binocular region, which both eyes can see, is approximately 120°, and

- Two lateral monocular regions, seen exclusively by each eye, are each approximately 40°.

Previous NASA research has evaluated many of these factors for military applications of HMDs such as the trade-offs between color-cuing and monocular, binoptic, and stereoscopic cuing information [37]. For commercial and business applications, the current emphasis is to minimize pilot encumbrance with the present efforts focused on a monocular HWD.

When a monocular HWD design is employed, a number of perceptual conflicts result because the image is presented to only one eye, thereby creating a visual field that is different from the other. The key differences [34] relevant to this work are:

- Brightness differences
- Color differences
- Motion differences

These perceptual conflicts manifest themselves into so-called "binocular rivalry" whereby the two visual fields are either fused, one visual field is seen to the occlusion of the other, or the two visual fields alternate. Binocular rivalry has been studied extensively and the general rules pertinent to this work regarding binocular rivalry [34, 38] are that:

- The higher contrast visual field dominates, so the higher contrast field is viewed longer.
- The brighter field dominates.
- The rate of alternation between the two fields tends to increase as the size of the items in the two fields increases.
- Alternation is not under complete voluntary control.
- The alternation of the visual field typically takes 1 to 4 seconds.
- Time on task reduces the alternation.

Generally, performance data during target acquisition tasks have not revealed any significant effects in binocular rivalry due to participants' eye dominance.

If the size of the HWD FOV is small, binocular rivalry issues vanish since the visual fields for the two eyes become nearly identical and the "trigger mechanisms" for binocular rivalry diminish. However, a small FOV can be detrimental to the visual task performance as demonstrated in testing with military pilots using HMDs [36, 39, 40].

FOV requirements for HMDs (HWDs) have been found to be task dependent. A FOV as small as 5° may be suitable for target reticle applications, but at least 20° has been found necessary for flight applications. During night flight operations, peripheral vision and sensor information may be critical - requiring a much wider FOV, nearing 40° horizontal subtended angle [34].

A large FOV allows users to use eye- and head-movements cooperatively to find vital information. Eye-movements are faster than head-movements, and coordination of head and eyes to acquire visual targets is a natural activity [41]. A large FOV allows the use of peripheral vision information. In addition, a large FOV allows for conformal display information and is less susceptible to clutter concerns.

2.4 Present Study

The advantages of the full color, head-tracked HWD can directly address the HUD limitations shown in RIPS [8–12] and T-NASA [4–7] testing. As such, a study was conducted to determine the efficacy of a head-tracked HWD in a complex taxiing task in a fixed-based simulator. In addition, the study was used to obtain pilot comments on the concept and future enhancements required for using a head-tracked HWD for surface operations.

In many cases, research questions, such as the use of color and acceptable levels of user "encumbrance" were not experimentally varied because of logistical constraints in the design and development of appropriate display hardware. In these cases, available, off-the-shelf hardware were used and the resultant data provide "point-design" information that will be used to start building the research database for these parameters.

An initial usability study examined the efficacy of head-worn display concepts for surface operations. Various display concepts were developed to elicit feedback from pilots to improve the concept (if viable) for future experiments. Full-color HWD display concepts were evaluated during surface operations to address previously witnessed display technology limitations.

From the results of the first usability study, two experimental studies were conducted to determine the efficacy of using HWDs to enhance taxi operations. For both experiments, full-color HWD display concepts were evaluated in surface operations to address previously witnessed display technology limitations. Previous research has shown that a HUD can significantly enhance Situation Awareness (SA) for surface operations. However, due to the HUDs fixed field-of-regard and limited FOV, intuitively portraying turns on the HUD can be difficult [5]. Further, information clutter is a driving constraint for the monochromatic HUD.

3 Usability Study

The intent of the usability study was to generate top level data regarding the concept of surface operations using a HWD and the influence of color HWD design parameters. The results of the usability study were used to reduce the experiment matrix for future HWD experiments.

3.1 Simulation Facility

For the HWD usability study, two HWD devices (Fig. 6) were used: 1) 800H x 600V pixel, full color display with optional see-through capability; and, 2) a glasses-mounted, full color 640H x 480V pixel, non-see-through display. An optical head tracker provided the head orientation data. The pilot controls included a tiller, throttles, and differential toe brakes. The Evaluation Pilot

(EP) interacted with the system via a Voice Recognition System (VRS). The VRS was not a requirement for the evaluation of HWD technology but provided an expedient way to allow pilot interaction. A total of six VRS commands were used to change the display modes and change display range (see Appendix A).

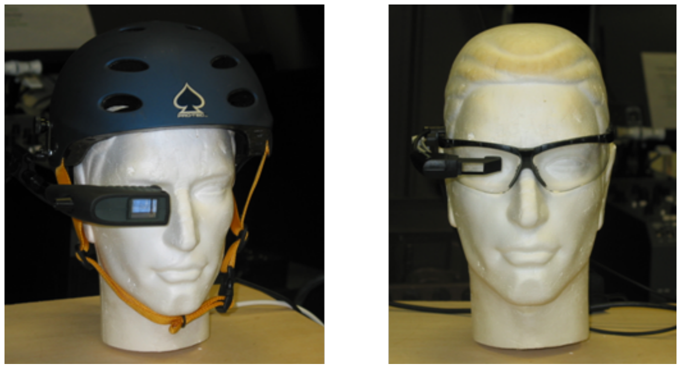


Figure 6. The HWDs tested in the usability study.

The usability study was conducted in the Visual Imaging Simulator for Transport Aircraft Systems (VISTAS) III part-task simulator at NASA Langley Research Center (LaRC) (Fig. 7). The VISTAS III configuration was a single pilot fixed-base simulator consisting of a 144° x 30° out the window visual, a 36 inch wide x 15 inch tall HDD, and pilot input controls. The simulated aircraft in VISTAS III was a medium- to long-haul commercial passenger aircraft classified by International Civil Aviation Organization (ICAO) Aeroplane Design Code D [42].

Figure 7. VISTAS III part-task simulator.

3.2 Evaluation Pilots

Eight EPs participated in the usability study and the EPs operated without a crew member. The EP population consisted of six commercial pilots and two test pilots. Of the six commercial pilots, two were captains and four were first officers. The six commercial pilots had an average of over 13,000 flight hours. The EPs were given a 30-minute briefing to explain the display concepts and the evaluation tasks. After the briefing, a 1-hour training session was conducted to familiarize the EPs with the VISTAS III simulator, the HWD devices, and the piloting task. Following training, 2.5 hours of data collection was conducted. The total time for an EP was approximately 4 hours.

3.3 Evaluation Task

The EPs conducted taxi operations at Reno / Tahoe International Airport in Nevada (FAA identifier: RNO) based on simulated datalink taxi clearances. The HWD device, the display concept, and weather conditions were varied. Pilots were instructed to taxi at a speed they thought appropriate for the task. Two taxi routes were used in the study: 1) a runway to gate route (Fig. 8, dashed route on left), and 2) a gate-to-runway route (Fig. 8, dotted route on right). The runway-to-gate taxi route was to exit Runway 16R via Taxiway November, turn onto Taxiway Bravo, and turn on Taxiway Golf proceeding to the gate. The gate-to-runway route was to taxi from the Mercury Aviation Center apron onto Taxiway Charlie, cross Runway 25 to Taxiway Papa, and hold short of Runway 34L. The weather condition for the out the window scene was varied between clear day with unlimited visibility and fog with 1000 RVR.

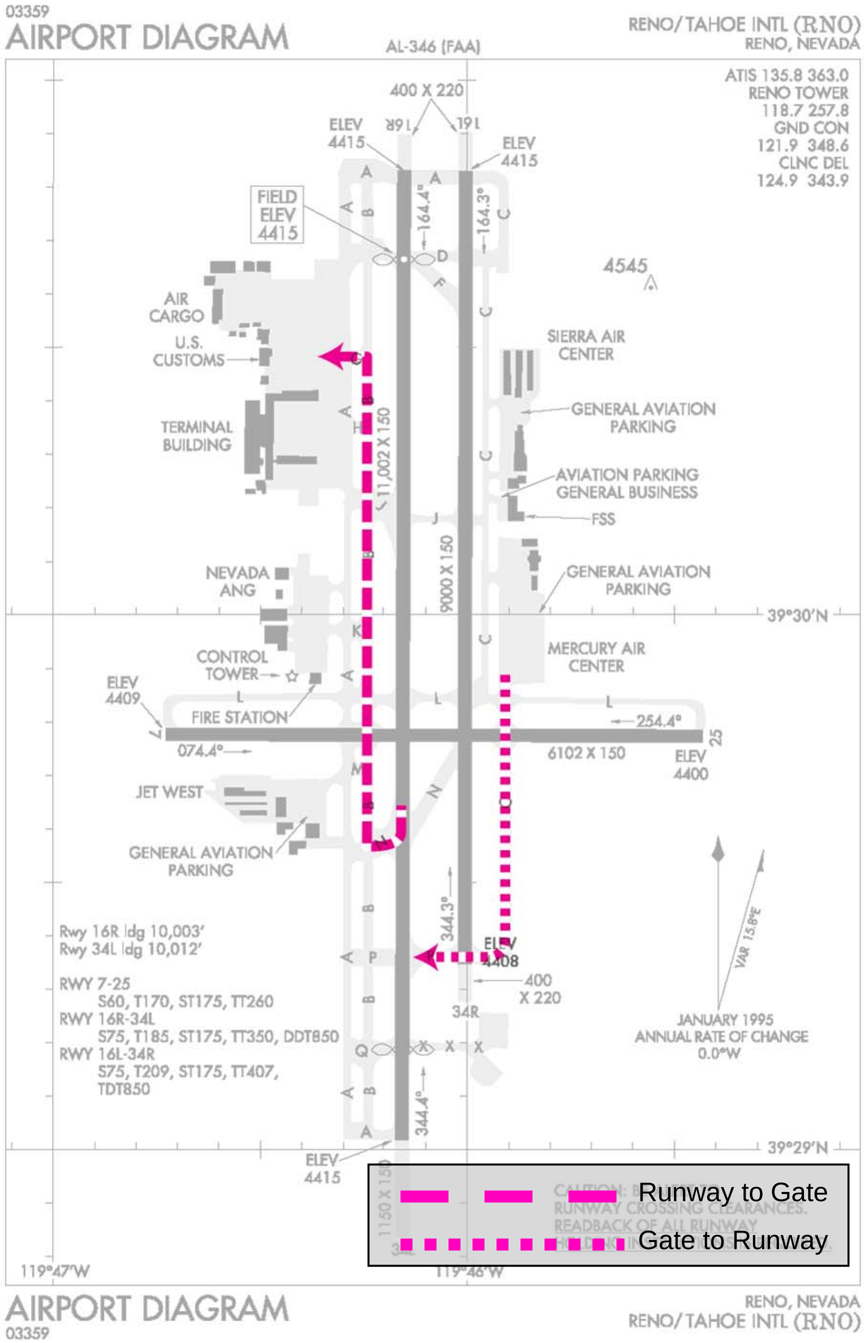


Figure 8. Taxi routes for the usability study.

3.4 Display Conditions

Three different display conditions were used: 1) a Paper Chart, 2) an advanced EFB-type display consisting of a 2-Dimensional (2-D) track up moving map with the cleared taxi route and ownship symbol and 3) HWD concepts.

For the HWD concepts, two HWD devices were tested. The first HWD was nominally see-through with 800x600 pixel resolution. The display could also be made non-see-through by closing an opaque door on the display. The FOV was approximately 23° horizontal x 16.5° vertical. The second HWD had 640x480 pixel resolution with an approximate FOV of 14° horizontal x 10.4° vertical and was non-see-through. EPs were instructed to position the display in front of either eye according to their preference. Table 1 and Figure 9 summarize the HWD display condition variations tested. Items in parenthesis in Table 1 represent a one-character abbreviation for Figure labels.

Table 1. HWD Display Conditions and associated configuration abbreviations for the usability study.

HWD	**Mode**	**See-Through**	**FOV**	**Head Tracker**
800x600 (8)	Multi (M)	Open (O) / Closed (C)	Conformal (F)	ON (K)
640x480 (6)	Single (S)	N/A (-)	HUD (H)	OFF (N)

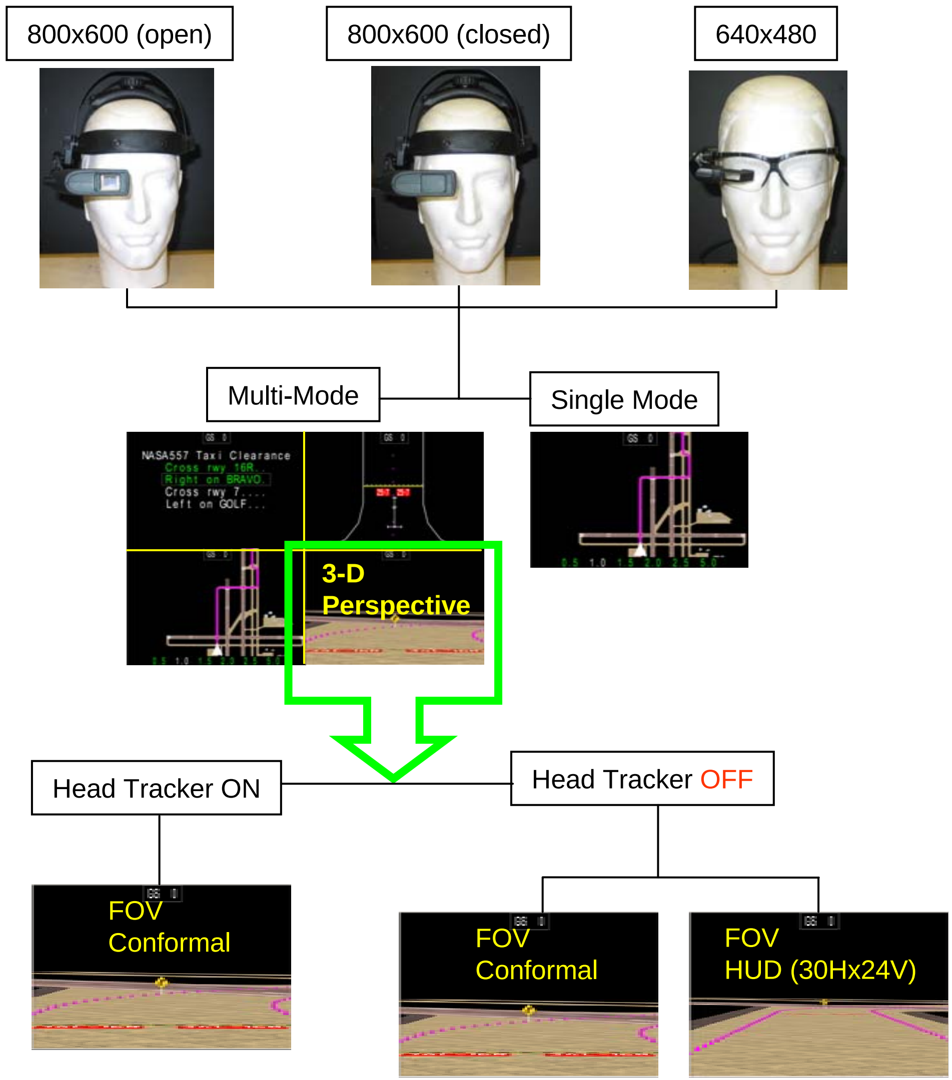


Figure 9. HWD display conditions for the usability study. Note that the display conditions for the Paper Chart and EFB are not shown.

For each HWD, there were two display concepts tested: 1) Single Mode and 2) Multi-Mode (Fig. 10). Single Mode consisted of a 2-D moving map, plan view display. This mode did not require head tracking. Using the VRS for the Multi-Mode display, the EPs could choose between the four different display modes: 1) a text display of the taxi clearance, 2) a 2-D moving map (the Single-Mode presentation), 3) a zoomed-in 2-D moving map for precision surface guidance and 4) a 3-Dimensional (3-D) perspective display. The display modes are described in detail in Appendix A.

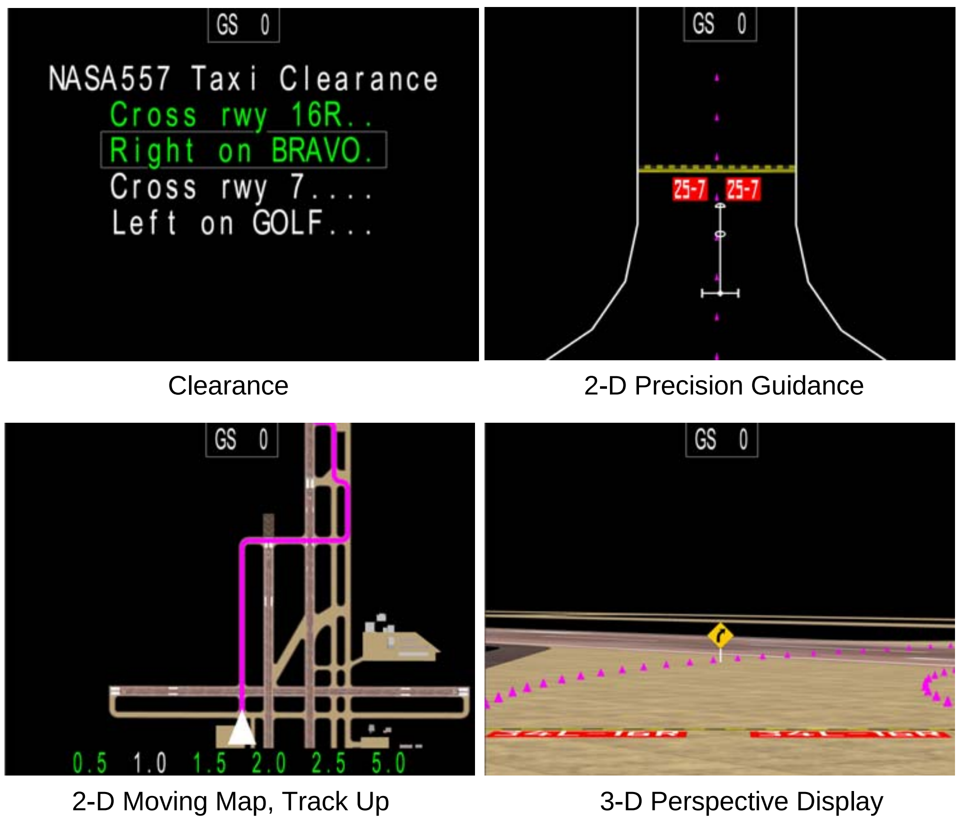


Figure 10. The four display concepts included in the Multi-Mode surface operations concept. In the Single Mode case, only the 2-D moving map (lower left) concept was displayed to the pilot.

The 3-D perspective display was evaluated with and without head-tracking. For the head-tracked condition, the FOV presented on the display corresponded with that of the display device (i.e., it was conformal, thus the virtual-world objects would overlay real-world objects). For the non-head-tracked case, two FOV conditions were tested: 1) the FOV of the display device (i.e., the same FOV as conformal, but non-head-tracked); and, 2) a fixed field-of-regard approximately the same as a HUD FOV (30° horizontal x 24° vertical).

The usability evaluation was conducted in four blocks and the run order was the same for each EP (Table 2 and Appendix B). For each of the four blocks, the weather condition (Clear or 1000 RVR) and display mode (Multi-Mode and Single Mode) were held constant. For Block 1, the outside weather was clear and unlimited visibility and the display mode was pilot-selectable (Multi-Mode). Block 2 was the same as Block 1, however, the outside weather was 1000 RVR. The taxi clearance for all of the data runs in Blocks 1 and 2 was the runway-to-gate route described previously. For Blocks 3 and 4, the display concept was held constant to the Single Mode presentation. For Block 3, the weather condition was clear and unlimited visibility; while for Block 4, the weather condition was set to 1000 RVR. In Single Mode, there was neither head tracking nor any FOV variations; thus, the only variation was the HWD display device (800x600 pixel display (open), 800x600 pixel display (closed) and the 640x480 pixel display). The taxi clearance for all data runs in Blocks 3 and 4 was the gate-to-runway route described previously.

Table 2. Experimental Data Collection Blocks.

Block	**Weather**	**Display Mode**	**Taxi Route**
1	Clear	Multi-Mode	Runway-to-Gate
2	1000 RVR	Multi-Mode	Runway-to-Gate
3	Clear	Single Mode	Gate-to-Runway
4	1000 RVR	Single Mode	Gate-to-Runway

3.5 Results

The usability study was used to confirm the efficacy and acceptance of using a HWD for surface operations and to explore possible experiment matrix conditions for future experiments. As such, the results of the HWD usability study were principally subjective. The pilot comments were sought to improve the concepts and help focus more rigorous follow-on research (Experiments I and II in this paper). The questionnaires used in the usability study are shown in Appendix C, Figures C1 through C5.

3.5.1 Multi-Mode Display Concept with Clear Weather (Block 1)

After the completion of Block 1, the EPs rank-ordered their preferred concept. Using the Friedman test, the rank ordering of the Multi-Mode display types by the EPs was significant (Fig. 11) for the clear weather case, $\chi^2(8, n = 8) = 19.600, p < 0.05$. The highest ranked Multi-Mode display type was the 800x600 pixel display (open) with a conformal FOV for the 3-D perspective mode and the head tracker on (config. 8OFK, refer to Table 1 for an explanation of the configurations). The lowest ranked display type was the non-head-tracked 800x600 pixel display (closed) with the conformal FOV for the 3-D perspective display mode (config. 8CFN).

3.5.2 Multi-Mode Display Concept with 1000 RVR Weather (Block 2)

After the completion of Block 2, the EPs again rank-ordered their preferred concept. Using the Friedman test, rank ordering of the Multi-Mode display types by the EPs was again significant (Fig. 11) for the 1000 RVR weather case, $\chi^2(8, n = 8) = 19.162, p < 0.05$. The most preferred Multi-Mode display type was the head-tracked 800x600 pixel display (open) with the conformal FOV (config. 8OFK) . The least preferred display type was the non-head-tracked 800x600 pixel display (closed) with the conformal FOV for the 3-D perspective display mode (config. 8CFN).

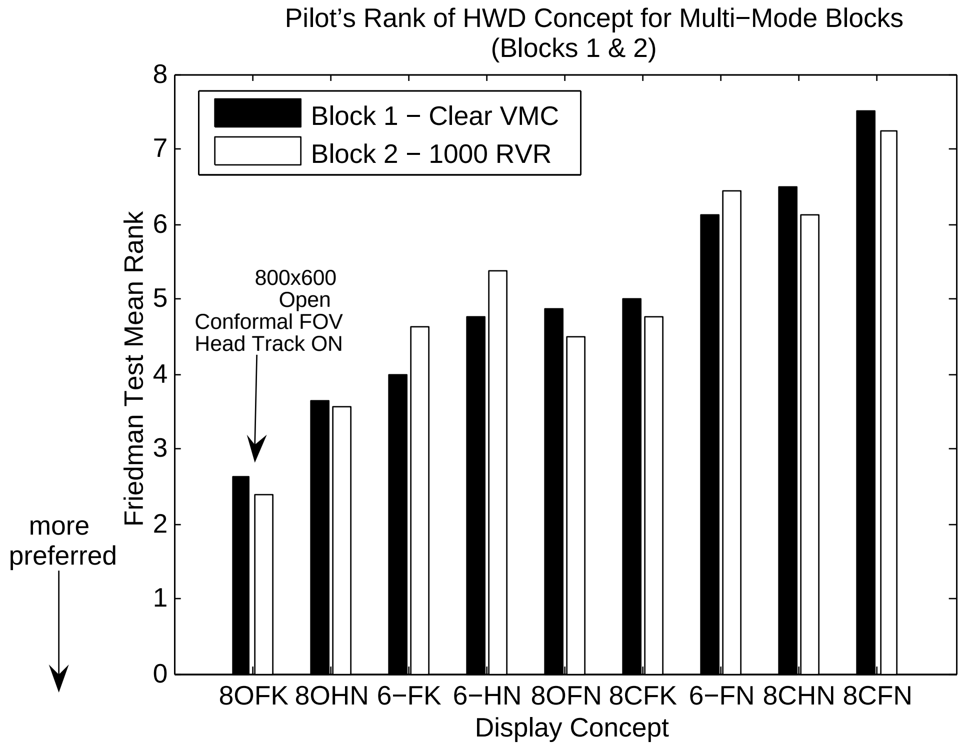


Figure 11. Display concept ranking for Blocks 1 and 2. The x-axis labels: 6 = 640x480 HWD; 8 =800x600 HWD; O = Open; C = Closed; F = Conformal FOV; H = HUD FOV; K = Head Tracker On; N = No Head Tracking. Note that Open or Closed does not apply to the 640x480 HWD.

3.5.3 Single Mode Display (Blocks 3 and 4)

Regardless of the visibility condition, the EPs preferred the 800x600 pixel display (open) and the 640x480 pixel display over the 800x600 pixel display (closed) (Fig. 12). For unlimited visibility, the rank order was significant, $\chi^2(8, n = 2) = 6.750, p < 0.05$. The 800x600 pixel display (open) and the 640x480 pixel display had the same highest ranking of 1.62 and the 800x600 pixel display (closed) had the lowest ranking of 2.75. For 1000 RVR visibility, the rank order was again significant, $\chi^2(8, n = 2) = 12.000, p < 0.05$. The 800x600 pixel display (open) and the 640x480 pixel display having the same highest ranking of 1.5 and the 800x600 pixel display (closed) having the lowest ranking of 3.0.

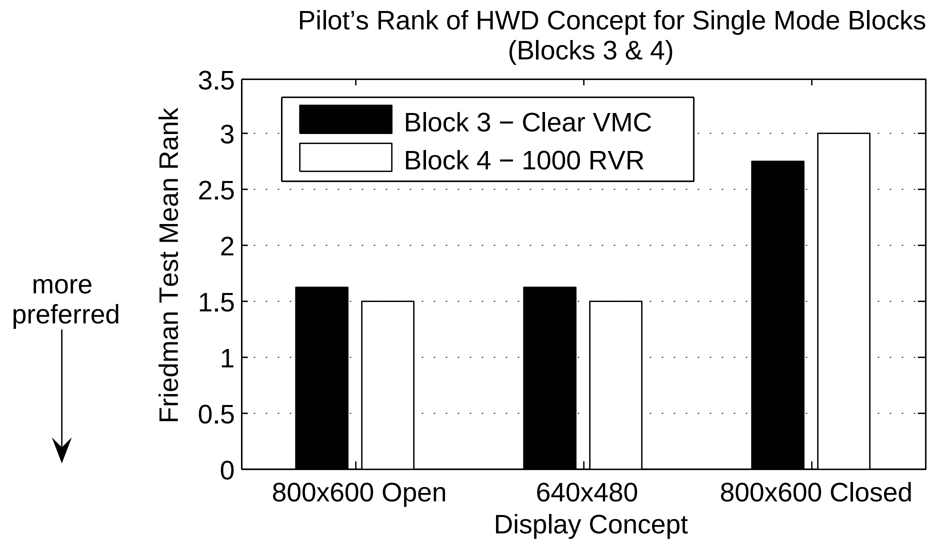


Figure 12. HWD display concept ranking for Blocks 3 and 4.

3.5.4 Situation Awareness

Upon completion of their trials, the EPs rated the HWD configurations for the situation awareness provided using a paired-comparison test. For this test, SA was defined as: The pilot's awareness and understanding of the dynamic environment and degree to which he or she is aware, and can successfully conduct and comply with taxi clearances under various weather conditions. Analysis of the Situation Awareness Subjective Workload Dominance (SA-SWORD) [43] data showed statistically significant differences across all display concepts for subjective SA, $F(8, 56) = 5.954, p < 0.05$. There were three subsets (Fig. 13) based on Student-Newman-Keuls (SNK) at $\alpha = 0.05$. In this paired-comparison test, the non-head-tracked display comparison was collapsed. The HUD FOV and Conformal FOV display concepts were treated as one, with the EP using his/her preferred FOV.

The 800x600 pixel display (closed), Single Mode (config. 8CS) had significantly less SA than all other display types tested (Fig. 13). Conversely, the 800x600 pixel display (open), Multi-Mode with the conformal FOV for the 3-D perspective display (config. 8OFK) had significantly higher SA than the following display types: 1) 800x600 pixel display (closed), non-conformal FOV, Single Mode (config. 8CS); 2) 800x600 pixel display (closed), non-conformal FOV, Multi-Mode (config. 8CHN); 3) 640x480 pixel display, non-conformal

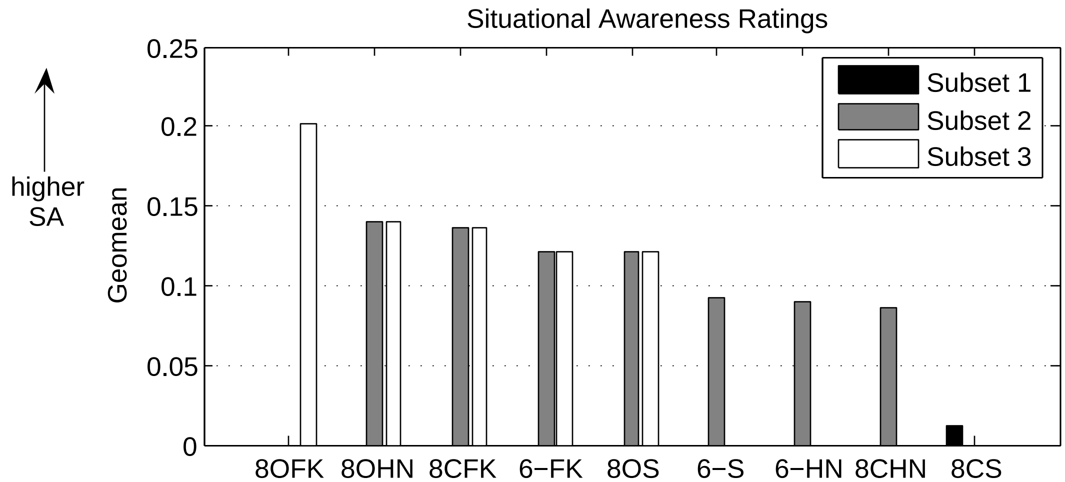


Figure 13. Situation Awareness per display concept. The x-axis labels: 6 = 640x480 HWD; 8 =800x600 HWD; O = Open; C = Closed; F = Conformal FOV; H = HUD FOV; K = Head Tracker On; N = No Head Tracking; S = Single Mode. Note that Open or Closed does not apply to the 640x480 HWD.

FOV, Multi-Mode (config. 6-HN); 4) 640x480 pixel display, non-conformal, Single Mode (config. 6-S) but no significant differences with the following display types: 1) 800x600 pixel display (open), non-conformal, Single Mode (config. 8OS); 2) 640x480 pixel display, conformal, Multi-Mode (config. 6-FK); 3) 800x600 pixel display (closed), conformal, Multi-Mode (config. 8CFK); and 4) 800x600 pixel display (open), non conformal, Multi-Mode (config. 8OHN).

The EPs were asked if the display type provided adequate awareness of their ownship position with respect to runways, taxiways, and stationary objects. EPs were asked to rate each of the display types (including Paper Charts (config. PC) and EFB) using a scale of 1 (low) through 10 (high). The Display Type was statistically significant for awareness of ownship position, $F(10, 70) = 7.923, p < 0.05$. There were four overlapping subsets (Fig. 14) when using the SNK method. Paper Charts were rated significantly lower (4.5) than all the display media formats. The 800x600 pixel display (open), conformal, Multi-Mode (config. 8OFK) was rated significantly higher (8.75) than 1) the EFB moving map (6.13); 2) 800x600 pixel display (closed), non-conformal, Multi-Mode (6.38), (config. 8CHN); and 3) 800x600 pixel display (closed), conformal, Single Mode (6.38), (config. 8CS) but with no appreciable differences with the remaining formats.

At the conclusion of the data collection runs, the EPs were asked to rank order the concepts in terms of overall effectiveness of the concept. They were asked to provide their most preferred Multi-Mode HWD concept (known as the EP-preferred Multi-Mode HWD) and their most preferred Single Mode HWD concept (known as the EP-preferred Single Mode HWD). A post-test paired comparison questionnaire was conducted in terms of runway taxi

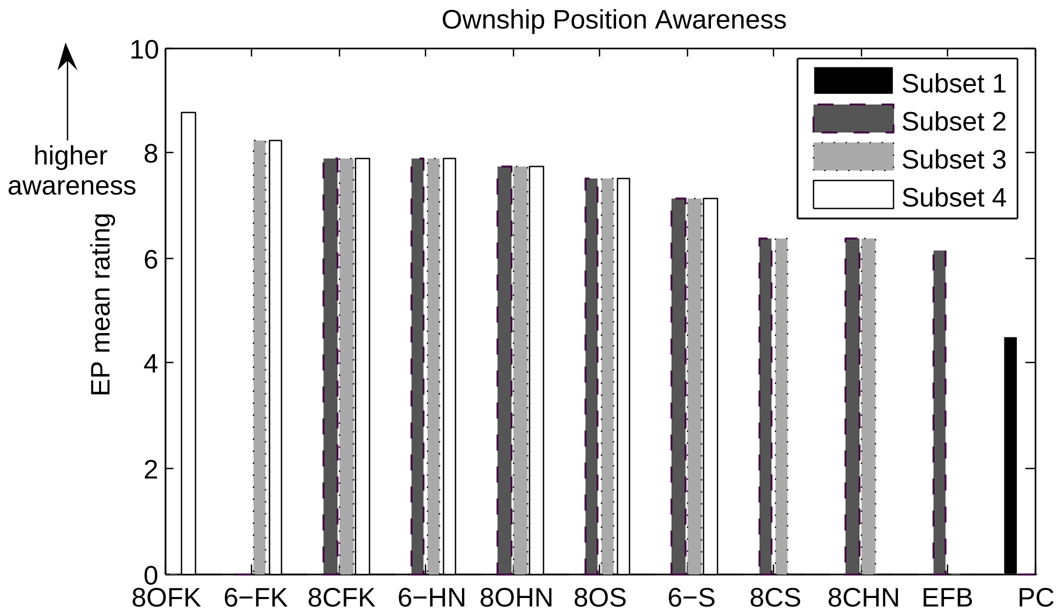


Figure 14. EPs rating of ownship position awareness for each display concept. The x-axis labels: 6 = 640x480 HWD; 8 =800x600 HWD; O = Open; C = Closed; F = Conformal FOV; H = HUD FOV; K = Head Tracker On; N = No Head Tracking. Note that Open or Closed does not apply to the 640x480 HWD.

and surface situation awareness using the displays tested (EP-preferred Multi-Mode HWD and EP-preferred Single Mode HWD) and current / near-term airline equipage (paper charts and EFB moving map). The results showed that there were statistically significant SA differences depending upon display media type, $F(3, 21) = 40.898, p < 0.05$. Post-hoc tests show three unique subsets (Fig. 15). Comparing the EP-preferred Multi-Mode HWD and the EP-preferred Single Mode HWD, the EPs preferred HWD concept was the Multi-Mode as it gave significantly greater surface SA than the three other types (preferred Single Mode, EFB moving map, paper charts). EPs ranked the Single Mode as having significantly higher SA than EFB moving map and paper charts. There were no appreciable differences in SA between EFB moving map and paper charts.

3.5.5 Workload

The EPs were given a paired comparison questionnaire regarding workload during runway taxi/surface operations using the displays tested (EP-preferred Multi-Mode HWD and EP-preferred Single Mode HWD) and current / near-term airline equipage (paper charts and EFB taxi-map). The results showed that there were statistically significant workload differences depending upon display media type, $F(3, 21) = 24.906, p < 0.05$. Post-hoc tests showed two unique subsets (Fig. 16). The EPs rated paper charts as generating signifi-

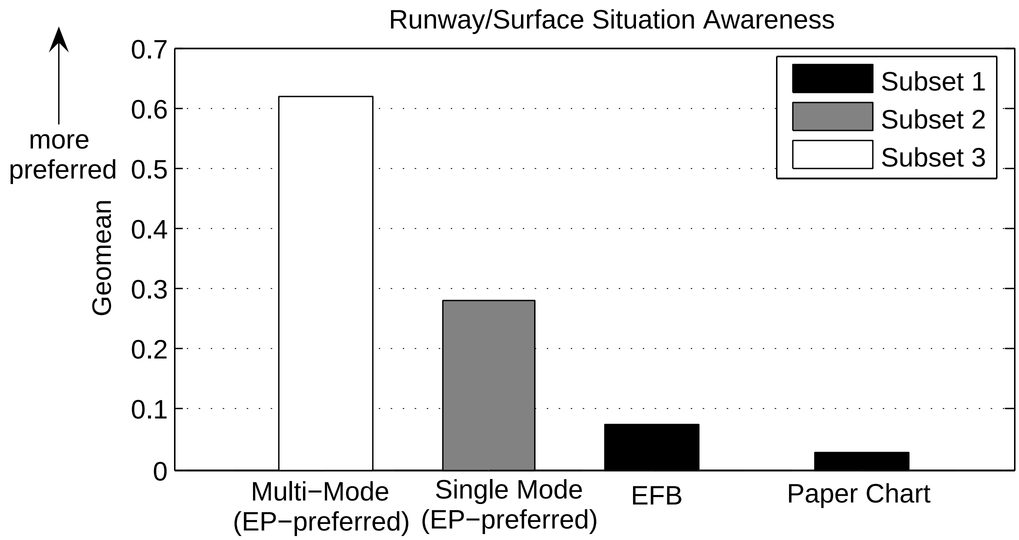


Figure 15. EPs rating of their surface situation awareness.

cantly more workload than the other three display media types (EP-preferred Multi-Mode, EP-preferred Single Mode, and EFB with moving map). There were no appreciable differences in workload among the EP-preferred Multi-Mode, EP-preferred Single Mode, and EFB moving map display media types.

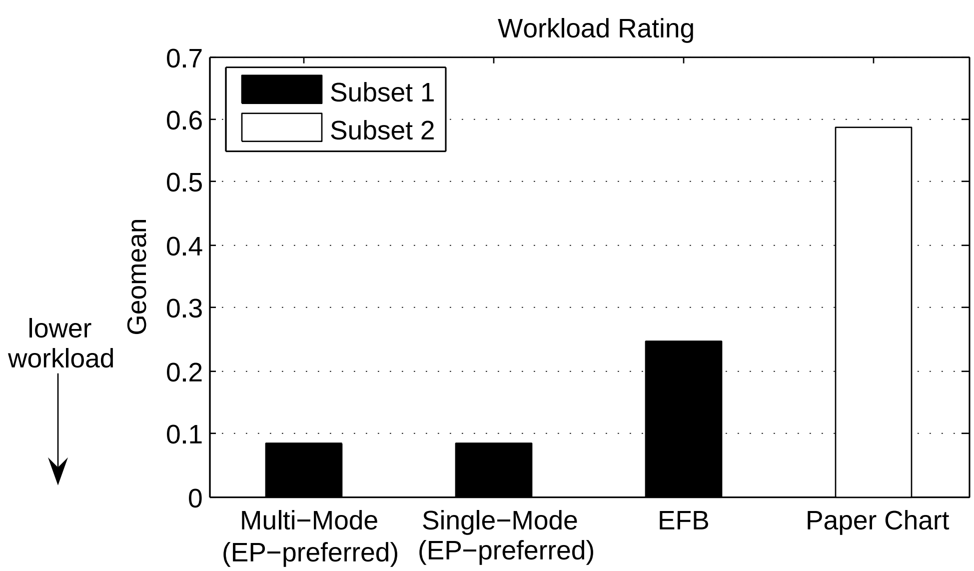


Figure 16. Workload ratings for the usability study.

3.5.6 Display Media

At the end of the test, the EPs were asked to rank-order the various display concepts they had evaluated by preference. Analysis of the data showed that display media type rankings were significant, $\chi^2(3, n = 8) = 22.950, p < 0.05$. The EPs display media ranking order was Multi-Mode (1.1), Single Mode (1.9), EFB (3.0), and Paper (4.0). For the Multi-Mode concept, the EPs were asked to rate the 4 modes (Fig. 10) based on which mode was most beneficial in terms of performance, SA and workload. Analysis of the results using the Friedman test showed that the Multi-Mode concept rankings were significant, $\chi^2(3, n = 8) = 13.050, p < 0.05$. The rank-order results shown in Figure 17 highlight the 2-D moving map mode (1.8) was most preferred followed by the 2-D precision guidance mode (2.0), the 3-D perspective mode (2.4), and the clearance text mode (3.9).

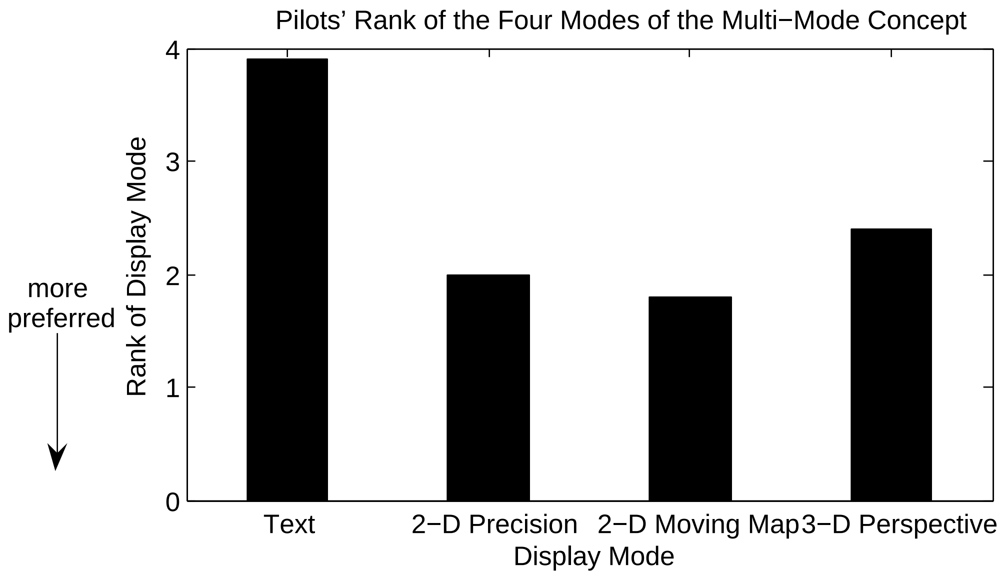


Figure 17. EP ranking of Display Modes of the Multi-Mode concept.

3.6 Usability Discussion

The usability study was designed to assess the efficacy of head-worn display concepts for surface operations. The results demonstrate that providing pilots with the ability to virtually see well beyond visual range can significantly increase situation awareness on the airport surface. The EPs reported significantly higher situation awareness with the HWD concepts compared to an electronic moving map or paper charts of the airport environment. Furthermore, the study provided tremendous insight into future design and development of head-worn displays, including hardware considerations and methods for integration of display modes.

3.6.1 Hardware Considerations

The usability study highlighted two significant hardware considerations. Nearly all EPs rated the 800x600 pixel display higher because it had higher resolution than the 640x480 display and it was see-through. The higher resolution improves the readability of the display especially for text and numbers. This finding is consistent with past research findings that higher resolution displays are generally preferred by pilots [44]. Additionally, the EPs preferred not to have their forward vision blocked; even by the small 640x480 pixel display. The see-through capability allowed the EPs to continue their nominal out the window surveillance of the airport environment during taxi. Also, the see-through display provided the EPs with confidence that the display was aligned with the scene. Therefore, the 800x600 pixel HWD (open), conformal FOV with head tracking was the HWD concept chosen for the follow-on experiments. For surface operations, it is important for a HWD to be see-through because, for all practical purposes, the HWD will always be providing an "augmented reality" not a "virtual reality" condition. The concept of operating in 0 ft ceiling and 0 RVR is a goal and not an actual weather condition (i.e., fog so dense that you could not see your hands in front of your face). Some visibility is almost always available so the HWD will provide symbology of information that can always be used by the pilot to compare to the "real-world".

3.6.2 Mode Integration

With regard to display mode integration, the EPs reported that the four modes (Clearance, 2-D Precision Guidance, 2-D Moving Map, and 3-D Perspective) each have relative merits in supporting taxi operations. For complex operational environments, the datalink textual clearance mode was reported to be of significant value in ensuring compliance with ground instructions. However, by itself, this clearance mode would likely not be enough to improve aviation safety. Only when the clearance mode is combined with the other modes does its potential become evident. In fact, this observation was witnessed for each of the modes. For example, the EPs reported that the 2-D moving map was of substantial benefit, but it did not provide them with precision guidance. However, the 2-D precision guidance only provided the EPs local guidance and not global situation awareness. The 3-D perspective helped give the EPs a sense of immersion; that is, the feeling that they were looking outside the cockpit into the real-world. This resulted in high situation awareness. However, the EPs still felt that the precision and the "big picture" were missing. Therefore, each of the modes contributed something unique, and the display concept was limited in its efficacy only because the EPs did not like to have to continuously switch between modes to extract the necessary information that each mode provided separately. As a consequence, the EPs unanimously stated their preference for a more integrated display that would blend the various display modes together to reduce the workload and improve accessibility to display information. These remarks resulted in the refined concepts (Fig. 18) taken forward to the follow-on experiments.

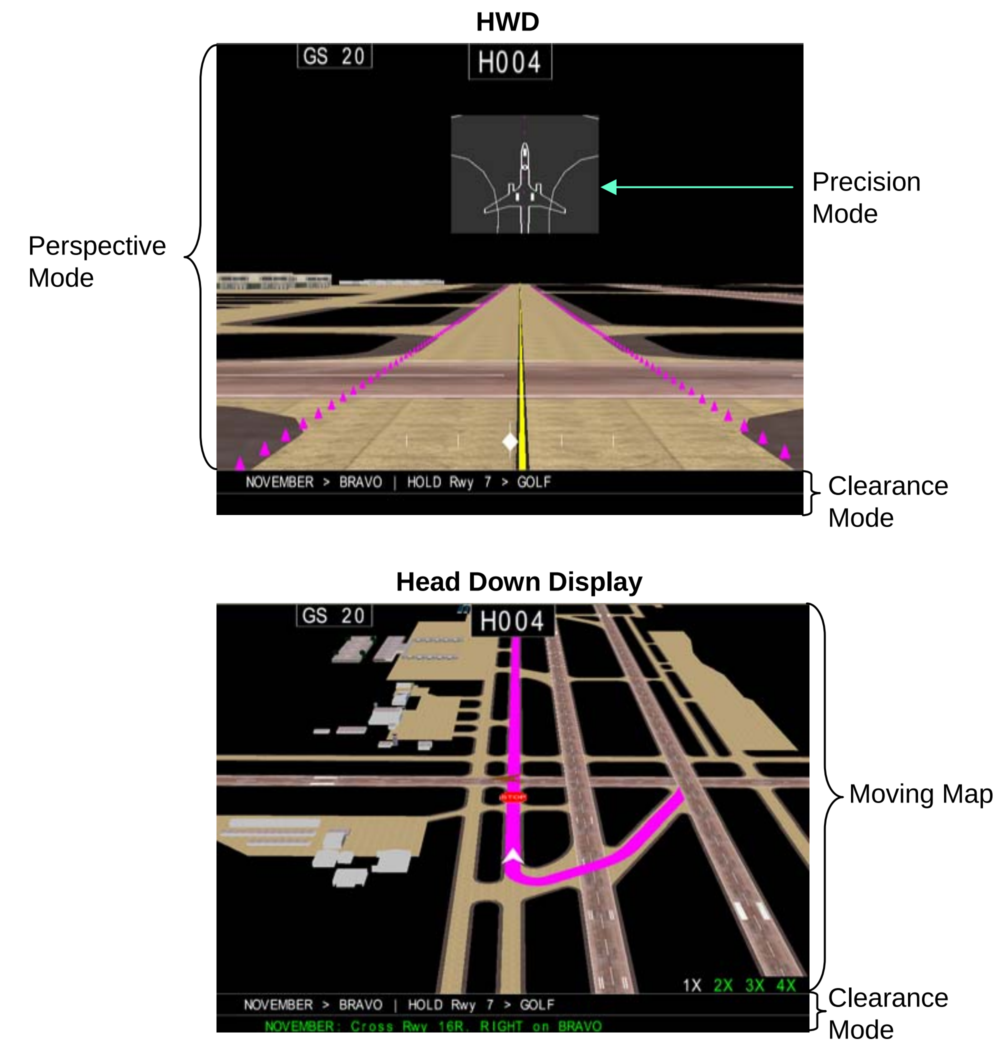


Figure 18. Improved and integrated Multi-Mode display concept.

4 Experiment I

Based on the results of the aforementioned usability study, two experimental studies were conducted to determine the efficacy of using HWDs to enhance taxi operations. The usability study showed that pilots preferred the open, 800x600 HWD display; therefore, it was used exclusively in both experiments. This section describes the first experiment.

4.1 Simulation Facility

Experiment I was conducted in the Research Flight Deck (RFD) simulator (Fig. 19) at NASA LaRC. The RFD configuration was a fixed-based, dual-pilot simulator consisting of a collimated 200° out-the-window scene. The out-the-window scene consisted of the airport, including taxiways and runways with appropriate markings, airport lighting, model aircraft representing traffic and simulated weather/lighting conditions. The visual acuity of the out-the-window scene was 20/80. The RFD was equipped with a 30°H x 24°V HUD on the captain's side. The HWD, worn only by the captain, was an 800H x 600V pixel, full color display with see-through capability, 60 Hertz (Hz) refresh and a pilot selectable brightness knob.

Figure 19. Research Flight Deck.

The captains placed the HWD near their right eye so that it was visible by glancing up which maintained unimpeded stereoscopic vision for out-the-window monitoring. The resulting display was conformal to the real-world (out-the-window scene) if the pilot tilted his or her head down. This procedure was also used to minimize binocular rivalry. An optical head tracker provided the head orientation data. The RFD had eight Size D (6.4 inch square viewable area) head-down displays: captain and first officer Primary Flight Displays

(PFDs) and NDs, two engine displays in the center of the instrument panel and two outboard auxiliary displays. For both experiments, the first officer's outboard auxiliary display was used as a repeater display of the captain's HWD or HUD depending on the scenario. The captain's auxiliary display showed basic aircraft status and was not utilized in the experiment. The pilot controls were a tiller, throttles, rudder pedals (nose wheel steering) and differential toe brakes. The simulated aircraft for both experiments was a medium- to long-haul commercial passenger aircraft, classified as an ICAO Aeroplane Design Code D [42].

A semi-automated ATC system was used. The ATC messages were broadcast via a standard voice audio instruction (pre-recorded audio files) and a simulated datalink message. Messages from the tower ATC were triggered 2 seconds after the beginning of the data run. The ground ATC messages were controlled by the researcher. When the first officer requested the ground clearance, the researcher would press a button to trigger the audio and datalink ground clearance message. The ground clearance would be repeated if requested by the EPs. During the data run, automated ATC communications were played to simulate typical radio party line chatter.

4.2 Evaluation Pilots

Sixteen commercial flight crews (a captain and first officer) participated in the experiment. Each flight crew flew for a major U.S. air carrier. The EPs were paired with others from the same company to ensure crew coordination and cohesion with regard to surface operation procedures. The captains had an average of over 16,000 flight hours with 22.3 years of commercial flying. The first officers had an average of over 11,000 flight hours with 13.5 years of commercial flying. Forty-four percent of the captains required corrective lenses. The EPs were given a 45-minute briefing to explain the display concepts and the evaluation tasks. After the briefing, a 1-hour training session was conducted to familiarize the EPs with the RFD simulator, HUD, the HWD, and the piloting task. Only the captain used the HUD or HWD. An eye dominance test was performed after the training briefing, revealing that all captains were right eye dominant. The HWD was viewed with the right eye. Following training, 5 hours of data collection was conducted. The total experiment time for a crew was approximately 8 hours.

4.3 Evaluation Task

The EPs conducted simulated taxi operations at Chicago O'Hare International Airport. The display condition and weather were experimentally varied. A total of 27 different taxi scenarios (Appendix D) were used in the study. The EPs were provided an enlarged airport diagram (Appendix E) to use at their discretion during any scenario. Three of the 27 scenarios were rare-events [45].

Rare event scenarios offer the opportunity to evaluate the display concepts in off-nominal situations. All taxiing tasks involved exiting the active runway and taxiing to the airport movement area boundary. The weather state for the out-the-window scene was varied between night-time with unlimited visibility Visual Meteorological Conditions (VMC), and daytime with 700 RVR. For the final run, the visibility was reduced to 500 RVR. The EPs were instructed to choose an appropriate taxi speed for the task and to avoid other aircraft. The EPs were briefed to follow their company guidelines for taxi speeds and procedures. Further, the EPs were instructed that the safety of the aircraft should never be compromised.

Before each data trial, the flight crews were briefed on their current location and expected runway turnoff. Each trial began with an initial speed of 10 or 15 knots followed by an immediate call from the Airport Traffic Control Tower (ATCT) local controller. Once cleared of the runway, the first officer switched to the ground frequency and called the ground controller for clearance. The ground controller (i.e., researcher button press) provided the taxi instructions along with a datalinked message of the cleared route. In addition, other pre-recorded aircraft traffic taxied around the surface. Crews were instructed that the traffic was pre-recorded; therefore, they should give way to all traffic. Further, they were briefed that the ground controller would not provide traffic awareness cues.

4.4 Display Conditions

Four display conditions were tested (see Fig. 20 and Appendix F):

1. A head-down EMM without routing or traffic information. There was no head-up display with this condition. This condition is hereafter referred to as the "Baseline,"

2. A HWD condition with a head tracker that displayed a virtual airport environment but without traffic, routing or clearance information. The Baseline EMM was displayed head-down. This condition is hereafter referred to as the "Intermediate HWD,"

3. A HUD condition with an advanced EMM head-down display. The scene-linked HUD symbology consisted of a 3-D depiction of the cleared route by highlighting the taxiway edge lines and centerlines. No traffic symbology was presented on the HUD. An advanced EMM containing iconic traffic, clearance and routing information was shown head down. This condition is hereafter referred to as the "Advanced HUD" and

4. An advanced HWD concept with virtual traffic and routing information and an advanced EMM head-down display. This condition is hereafter referred to as the "Advanced HWD."

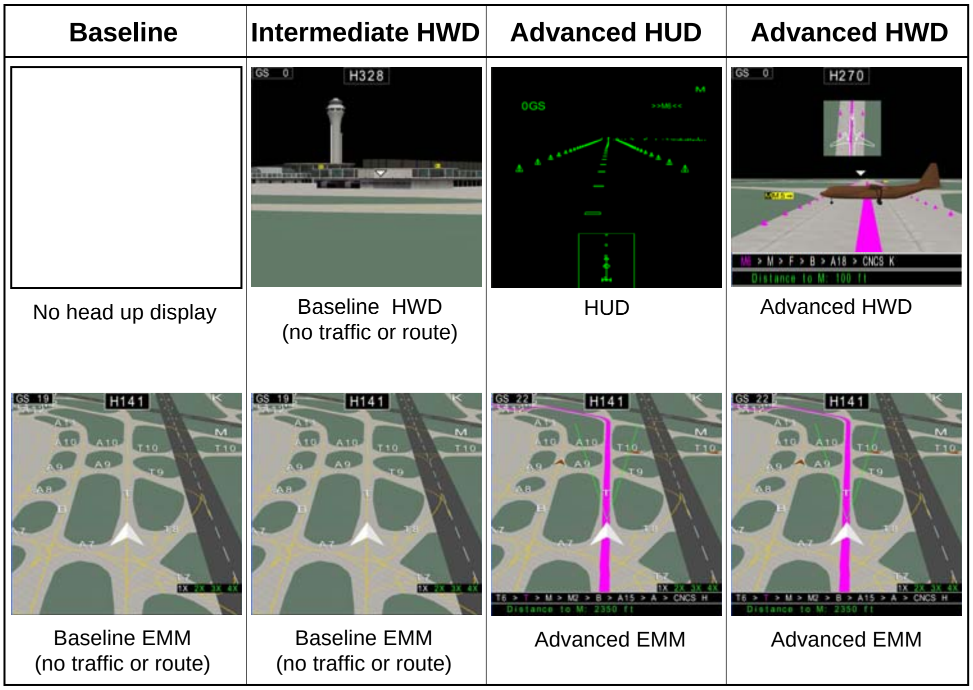


Figure 20. Thumbnail pictures of the four display concepts: Baseline, Intermediate HWD, Advanced HUD and Advanced HWD.

All display concepts employed a head-down EMM display shown on both the captain and first officer navigation displays. The EMM consisted of a perspective, track-up view of the airport showing an ownship symbol, ground speed, heading, airport movement areas, taxiway/runway centerlines, airport surface labels, and current range selection. Both the captain and first officer had independent range controls for the EMM, which consisted of four zoom levels. In addition to the perspective track-up mode, the EP could select a north-up mode that showed the entire airport view from directly above.

Two EMM display concepts were used: 1) a Baseline EMM that contained an ownship symbol, ground speed, heading, taxiways with centerlines and labels and runways, and 2) an advanced EMM that contained the same information as the baseline EMM with the addition of a route display, clearance information, distance to the next taxiway, and traffic icons. For the advanced display concepts (Advanced HUD and Advanced HWD), symbology was displayed on the EMM that depicted the FOV of the head-up device. On the advanced EMM shown in Fig. 20, the FOV symbology is shown as green wedge-shaped lines on the ownship symbol. For the HUD, the angle of the wedge symbology was 30°, as this is the horizontal FOV of the HUD. For the HWD concepts, the wedge symbology angle was 22°. Also, this FOV symbology moved as the captain moved his/her head; thus, the portion of the virtual airport the captain was viewing could be correlated on the EMM. “Hold short” symbology was depicted as a typical roadway stop sign.

The Intermediate HWD concept provided a conformal (head-tracked) virtual airport view from the pilot's eye perspective. This concept, while head tracked, was considered a "minimal" SV concept since it portrayed only a database presentation of the airport. The virtual airport consisted of the ORD airport, buildings, airport movement areas and runway/taxiway centerlines. Taxi signage was displayed in the HWD. This signage was modeled to appear to be actual airport surface signage; however, the HWD signage was placed on the side of an upcoming turn and did not necessarily correlate with the actual out-the-window sign placement. The HWD displayed the ground speed, heading and an aircraft-heading pointer. The aircraft-heading pointer was used to aid the EP in determining the aircraft heading during head movement. This concept did not contain traffic, route or clearance information.

The Advanced HUD display concept was based on RIPS [8,11] and T-NASA [4, 46] concepts without incursion alerting (Fig. 21). The head-up display showed current ground speed in digital format, the current taxiway (shown as $>> D <<$ in Fig. 21), next cleared taxiway (shown as $A17$ in upper right of Fig. 21), centerline markers and virtual cones on the taxiway edge. Additional cues were given for turns. These cues consisted of turn flags and virtual turn signs (similar to roadway turn signs) [47]. Runway holding positions were displayed as a single solid line at the hold short locations. Also, a virtual stop sign was placed at the middle of the hold short line. A non-conformal taxi director display provided an intuitive display of the relationship between the taxiway centerline and the aircraft's landing gear. The captain could remove all the symbols from the HUD display by pressing a declutter button. The auto-throttle disconnect button was used for declutter because it was conveniently located and auto-throttles were not used in the experiment. A second press of the auto-throttle disconnect would restore all of the HUD symbology. The captain also had control of the brightness level of the HUD via a rotary knob.

The Advanced HWD concept contained all of the information in the Intermediate HWD concept with addition of traffic and routing information. The Advanced HWD employed a 3-D generic aircraft model to depict traffic, the cleared route was shown as a magenta overlay on the taxiway centerline, text was displayed for the cleared route and for the distance to the next taxiway, and virtual taxiway edge cones depicted the edge lines of the cleared route. Like the HUD, virtual turn signs were used as an additional turn cue and hold short cues were denoted by virtual stop signs. Similar to the HUD, a non-conformal insert depicted a plan view of the runway, together with the airplane outline and location of the gear (Appendix F, Fig. F6). The EP could remove this non-conformal display by pressing the auto-throttle disconnect button. A second press of the auto-throttle disconnect button would remove all symbology in the HWD. A third press would bring all symbology

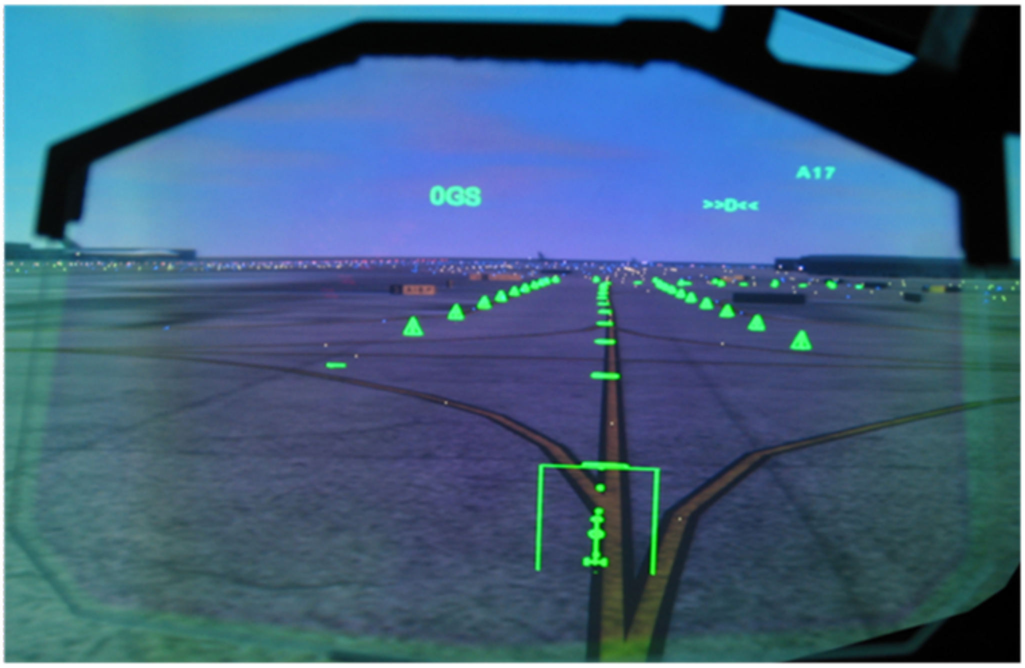


Figure 21. HUD showing centerline and edge line cones representing the cleared route. Bottom box shows the taxi director non-conformal display.

back to the HWD. Also, the captain could control the brightness of the display via a rotary knob. The Advanced HWD format represents the most complex but most preferred configuration based on the usability study.

The HWD concepts required alignment (known as bore-sighting) before the start of the data trials. The eye piece was positioned slightly above the captain's right eye. An alignment grid was displayed on the out-the-window visuals and the HWD. The captain bore-sighted the HWD by aligning the grids through head movement. Once the grids were aligned, the captain verbally called "alignment" and the bore-sight parameters were recorded. The conformal virtual airport view was then provided whenever the captain slightly tilted his/her head down; thus, the actual taxiways aligned the virtual taxiways.

4.5 Quantitative Results

4.5.1 Taxi Performance for All Trials

The rare event trials were excluded from the taxi performance analysis. Three taxi performance measures were evaluated for this experiment: Root Mean Square (RMS) path error, delta taxi speed (difference, in knots, from the average taxi speed for a specific route and the actual taxi speed during a specific data run on that path), and delta time-to-taxi (difference, in seconds, from the average time-to-taxi for a specific route and the actual time-to-taxi during a specific data run on that path). The "delta" taxi speed and time-to-taxi measures were developed to isolate the effects of display condition and visibility for each run by eliminating the route effects (i.e., long versus short routes and simple versus complex routes).

A Multivariate Analysis of Variance (MANOVA) on RMS path error, taxi speed, and time-to-taxi yielded significant display condition effects ($F(9, 589) = 4.57, p < 0.001$) and visibility condition effects ($F(3, 242) = 6.40, p < 0.001$) according to Wilk's Lambda. There were no significant effects for the interaction between display and visibility. Follow-up univariate Analysis of Variances (ANOVAs) indicated that RMS path error was not significantly ($p > 0.05$) affected by display condition but taxi speed ($F(3, 244) = 7.48, p < 0.001$) and time-to-taxi ($F(3, 244) = 12.25, p < 0.001$) were. Post-hoc SNK tests revealed three unique subsets for taxi speed and three overlapping subsets for time-to-taxi (Fig. 22). For these related measures, the EPs were able to taxi quicker (indicated by positive delta taxi speed and negative delta time-to-taxi) with the Advanced HUD and Advanced HWD concepts with taxi guidance and routing information than those without (Baseline and Intermediate HWD). Similarly, univariate ANOVAs showed that visibility effects were significant for taxi speed ($F(1, 244) = 17.79, p < 0.001$) and time-to-taxi ($F(1, 244) = 11.89, p = 0.001$) but not for RMS path error. On average, the EPs taxied 0.8 knots faster resulting in 13 seconds less time-to-taxi during the Day 700 RVR condition than the Night VMC condition, but these small differences result in little operational advantage.

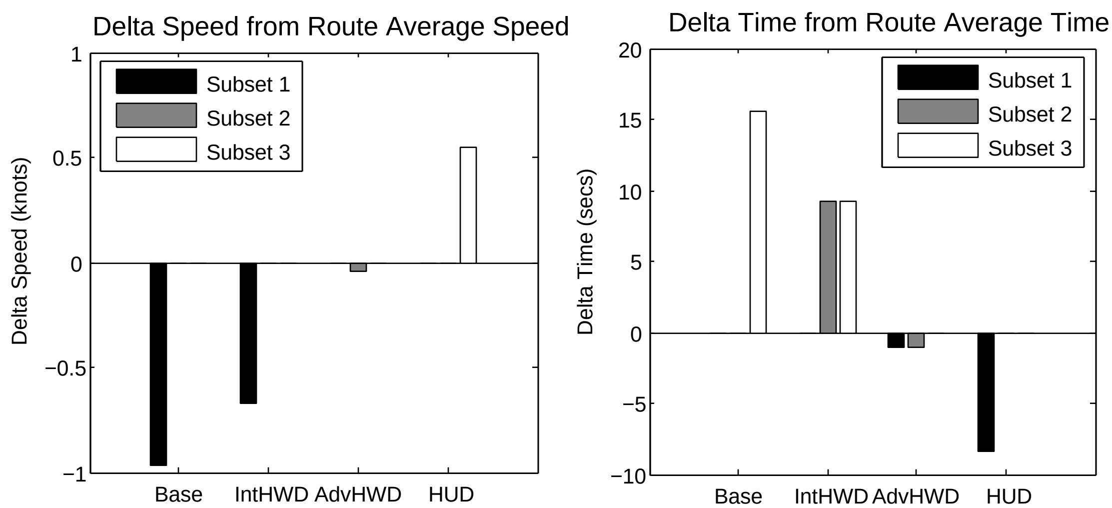


Figure 22. Delta taxi speed and time for all trials for Experiment I.

4.5.2 Navigational Errors

Navigational errors, when they occurred, were divided into two categories: major and minor [6]. A major navigation error is defined as a loss of navigational awareness, which resulted in a wrong turn or a failure to turn. A minor navigation error is defined as failure to remain on route but was immediately noticed and corrected by the crew. A blunder error that involved a conflict with another aircraft, was accounted for in a different measure and not captured as a navigational error. A total of 32 navigational errors were made, where 22 were classified as major errors and 10 were classified as minor. All of the major errors occurred with the Baseline and Intermediate HWD concept. Figure 23 shows the navigational errors made per display concept and weather event (night VMC or day 700 RVR).

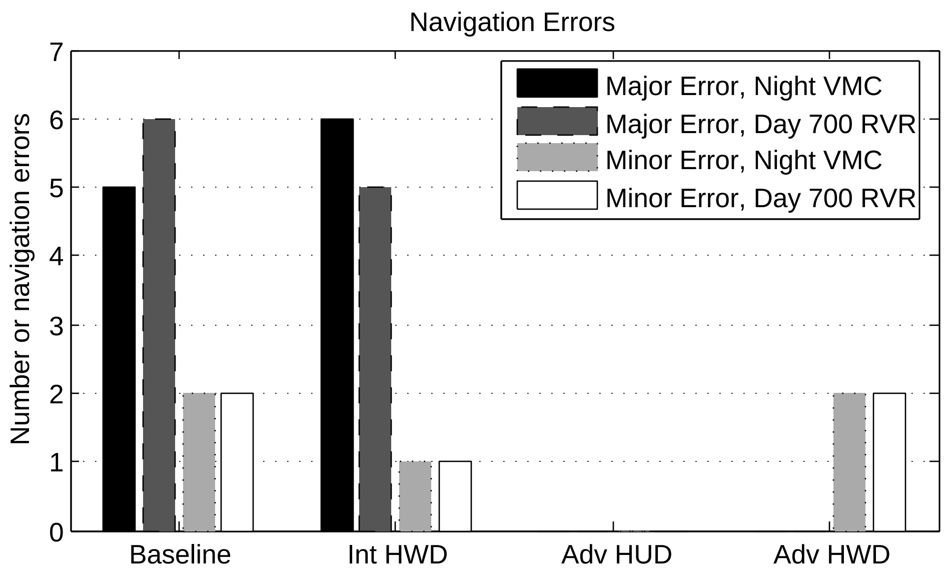


Figure 23. Navigational errors per display concept and weather event.

Figure 24 shows the errors made for each of the taxi route scenarios. Routes 13 - 24 were taxi routes that were used in previous experiments [48]. At least one crew made a navigational error on one of those 12 routes. Two-thirds of all the navigational errors were made with these routes.

Route 20 accounted for 25% of all errors. Most of the errors for route 20 were with display concepts in which a route was not displayed. The nature of the error involved turning the wrong way while taxiing along the route. To follow the route given by the ground controller, the captain had to turn away from the terminal area. It wasn't until crews made the wrong turn that they realized they could not follow the clearance. For the scenarios that had the route displayed, the proper turn was clearly displayed.

4.5.3 Taxi Conflict Events

A taxi conflict event was defined as a collision with another aircraft or making a turn in front of another aircraft and creating a close call. A total of 17 taxi

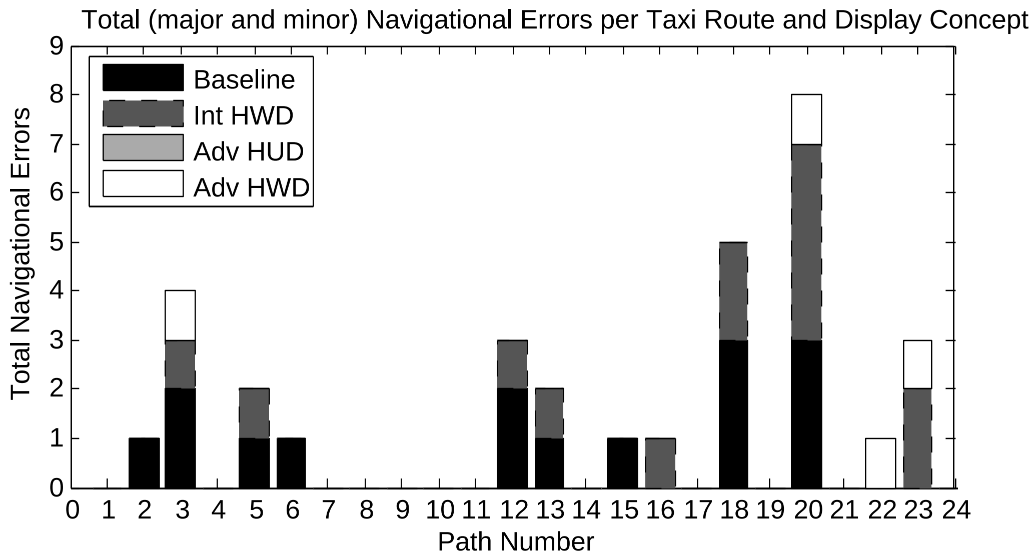


Figure 24. Navigational errors per route for each display concept.

conflict events were recorded (Fig. 25). An ANOVA was performed on the number of taxi conflicts committed by the crew with display concept (Baseline, Intermediate HWD, Advanced HUD, Advanced HWD), and visibility (Night VMC, Day 700 RVR) as the main factors. There were no significant differences ($p > 0.05$) among the main factors or their interactions for this measure.

4.5.4 Rare Events

Of the 27 different runs that each flight crew experienced, three data runs were rare-event scenarios. The sequence of these rare-event scenarios in the data trials was based on the severity of the rare-event. High severity rare-events occurred late in the data trials as to minimize the confound on the other experimental runs [45]. All 16 crews experienced each of the 3 rare events once.

The first rare event represented an ATC error in which the controller verbally instructed the EPs to turn right even though the ground controller's datalinked route depicted a left turn. The results from this trial showed that all crews that did not have a route displayed, either on the EMM or HWD device, turned the wrong way on the intended route in accordance with verbal instruction. All crews with the displayed route immediately called ground and asked for clarification before making the turn.

The second rare event had traffic that was not broadcasting its position, thus, the non-transponding traffic did not appear on the head-down Advanced EMM or the Advanced HWD. The non-transponding aircraft did not cause a conflict event because it was obvious enough to be easily detected in the out-the-window scene. Neither the Baseline concept nor the Intermediate HWD

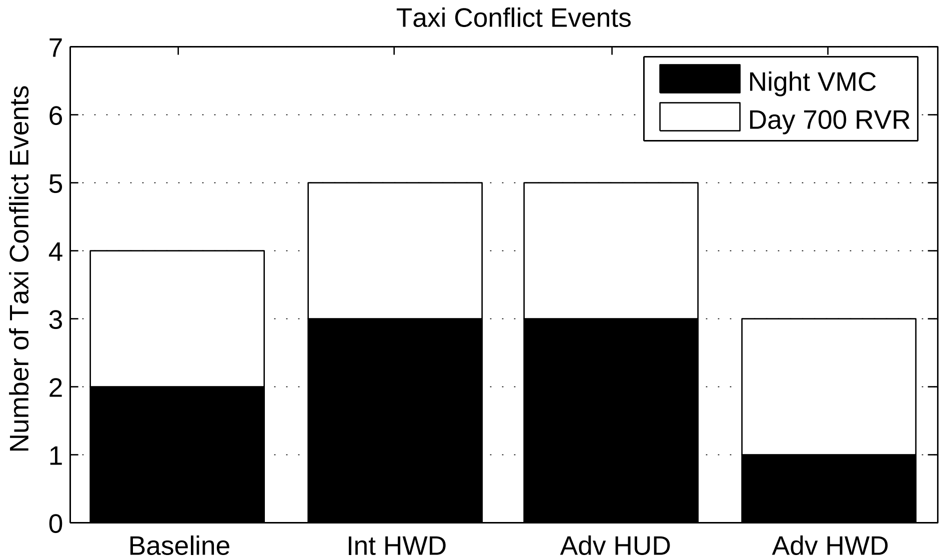


Figure 25. Taxi conflict events per display concept and weather condition.

concept was included as a display condition for this rare event because traffic is not displayed on these concepts. All crews noticed the traffic out-the-window and 9 of the 16 crews noticed the traffic was not displayed head-down. The remaining 7 crews did not notice that the traffic was not displayed on the EMM or the Advanced HWD.

The final rare event and final run of the day created a potential nose-to-nose traffic conflict. The nose-to-nose rare event was designed to provide insight into traffic awareness between the different display concepts. A common occurrence at ORD is when the terminal area is congested, aircraft may be given a “double back” clearance to create spacing and clear other taxiways. This event did not create a collision scenario but instead represented a more likely and common situation where two aircraft become “stuck” which would require an aircraft tug to separate the two airplanes. A nose-to-nose situation can significantly reduce airport efficiency to resolve the conflict. To create this rare event, crews were given a ground controller instruction to turn onto a taxiway that was already occupied by a small commuter jet. The visibility was reduced to 500 RVR for this scenario so that the traffic was difficult to see, but still detectable out-the-window. The small commuter jet was on the left (captain’s) side. The scenario also had two aircraft on the first officer’s side to serve as a potential distraction. To further increase the workload of the first officer, a complex ground clearance was given close to the point of conflict. Therefore, the prevention of the nose-to-nose situation depended mostly on the captain’s awareness.

This rare event display condition was evenly distributed between each of the four display concepts; therefore each of the four display concepts had four rare event data points. For display concepts that did not have path or traffic information (Baseline and Intermediate HWD), 7 of the 8 crews were not able to avoid a nose-to-nose condition. The one flight crew that avoided the nose-to-nose was able to turn out as they noticed the traffic before fully committing to the turn. For the display concepts that had iconic traffic display (Advanced HUD and Advanced HWD), all but two crews were able to avoid the nose-to-nose situation. Both nose-to-nose conditions occurred with the Advanced HWD concept. Exit interviews with these flight crews revealed that the traffic represented on the HWD was not quickly discernible against the background because the aircraft was colored brown to conform with Automatic Dependent Surveillance-Broadcast (ADS-B) color symbology convention [49]. The brown colored traffic icons provided little contrast against the rest of the display symbology; therefore, traffic tended to "blend-in" to the background. This result highlights the need for research into the design requirements for HWDs. One reason for conducting Experiment II (discussed in Section 5) was to identify the significance of this limitation.

4.5.5 Taxi Performance for Nominal Trials

In the above taxi performance analyses, large standard deviations existed due to the inclusion of navigation error runs and taxi incursion runs. For Experiment I, 80 trials out of 432 (16 crews each executing 27 data trials) total trials had an off-nominal event. An off-nominal event was defined as meeting any one of the criteria below:

1. The data trial was a rare event scenario.
2. At any time during the data trial, the path error was greater than ± 50 feet (± 15 meters).
3. The data trial contained either a major or minor navigation error.
4. The data trial contained a taxi conflict event.

A second analysis of the taxi performance was conducted but runs that met the above criteria were excluded from the analysis.

A MANOVA on RMS path error, taxi speed, and time-to-taxi yielded significant display condition effects ($F(9, 397) = 4.96, p < 0.001$) and visibility condition effects ($F(3, 163) = 8.15, p < 0.001$) according to Wilk's Lambda. There were no significant effects for the interaction between display and visibility. Post-hoc SNK tests revealed that the EPs had significantly higher (statistically, but not operationally) RMS path error with the Advanced HUD (mean=7 ft) than with the other three concepts (mean=6 ft). There were no significant differences between the Baseline, Intermediate HWD and Advanced HWD for this measure. SNK tests showed three overlapping subsets for the taxi speed and time to taxi measures (Fig. 26). For these related measures, the

EPs were able to taxi quicker (indicated by positive delta taxi speed and negative delta time-to-taxi) with the advanced concepts with taxi guidance and routing information than those without (Baseline and Intermediate HWD). On average, the EPs taxied 0.75 knots faster resulting in 10 seconds less time-to-taxi during the Day 700 RVR condition than the Night VMC condition but these small differences result in little operational advantage.

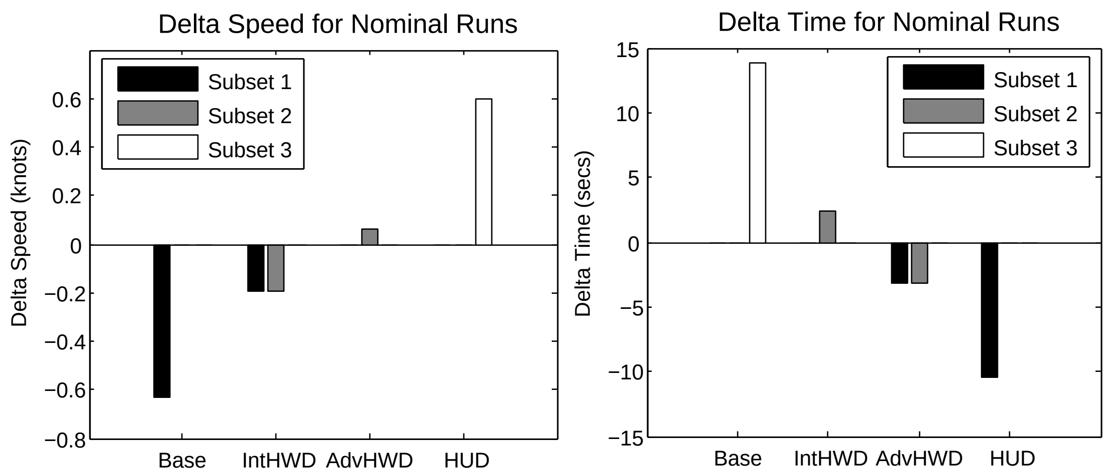


Figure 26. Delta taxi speed and time for nominal trials for Experiment I.

4.5.6 Required Navigation Performance

Similar to Required Navigation Performance (RNP) [50] standards being used for aircraft flight procedures, required airport surface performance has been proposed [51]. For Experiment I, the visibility conditions were such that RNP for an ICAO Code D [42] aircraft is within ±7.2 feet (±2.2 m) of the route centerline 95% of the time. Nominally, to determine the RNP capabilities of an aircraft, the navigational errors and their bounds need to be known. For the analysis in this paper, all sources of error with the exception of Flight Technical Error (FTE) are constant as the EPs used the same simulator throught the exeriment. Therefore, FTE was assumed to be the Total System Error (TSE) [50]. Also, for this paper, FTE is defined to be the path error measured from a defined point on the aircraft to the centerline of the intended/cleared route.

Figure 27 shows the RNP values for all run conditions (including off-nominal conditions) for each display concept. This treatment recasts the taxi performance analysis of Section 4.5.1. Taxi performance was not within ± 7.2 feet of route centerline 95% of the time for any of the display conditions. The large RNP values for the Baseline and Intermediate HWD are attributed to the crews making wrong turns, thus being hundreds of feet from the intended route.

The data were also analyzed using trials with no navigational anomalies.

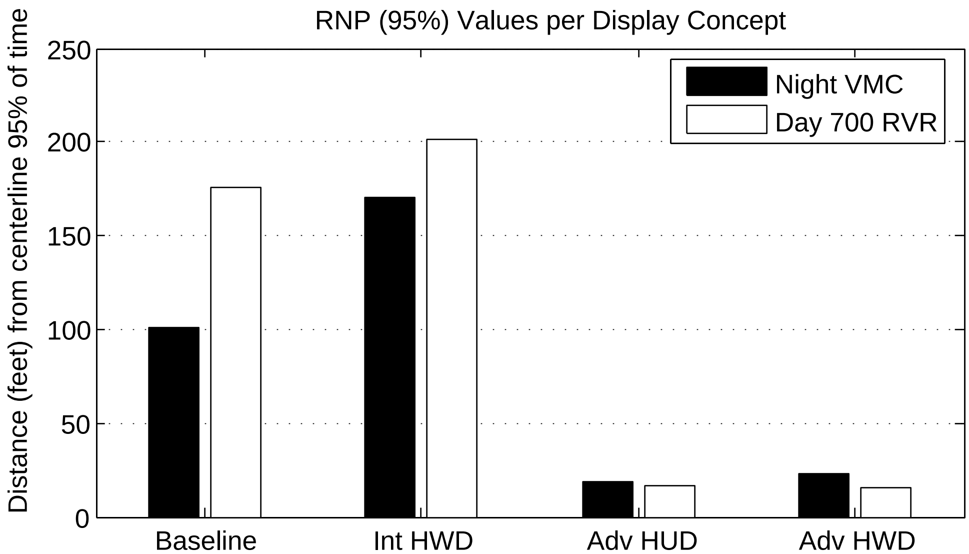


Figure 27. RNP for all data trials for Experiment I.

Figure 28 shows the results of the RNP with only nominal trials. This treatment recasts the taxi performance analysis of Section 4.5.5. The data were separated across day/night visibility condition and also used two different measurement references of the path error. The path error was measured from a point derived from desktop simulations to provide judgmental over steering guidance. Based on the desktop simulations, the over steer point was found to be 40% of the distance from the nose gear to the main gear. This 40% Guidance Point was drawn as a symbology on the EPs display (See Appendix F, Figure F5). By keeping this 40% Guidance Point symbology on the route centerline symbology, the aircraft's main landing gear would stay on the taxiway. This 40% Guidance Point was displayed head up to the EP on the Advanced HUD and Advanced HWD concepts. For the simulated aircraft in this experiment, the 40% Guidance Point was 20 feet behind the nose gear on the longitudinal axis. The EPs were instructed that the 40% Guidance Point would provide proper judgmental oversteering for the task.

Since the current literature references the path error (and RNP) from the nose gear, the path error was also computed from the nose gear location for comparison. The nose gear path error was computed post test. Using only nominal data trials, no display condition met the surface RNP requirements, whether measured from the 40% Guidance Point or the nose gear.

An ANOVA revealed that the visibility condition was significant, $F(1, 10) = 7.752, p = 0.019$. The daytime 700 RVR (mean=13.375 ft) resulted in significantly (but not operationally) lower lateral RNP than night (mean=15 ft). This is operationally insignificant as the difference between the conditions is less than 2 feet.

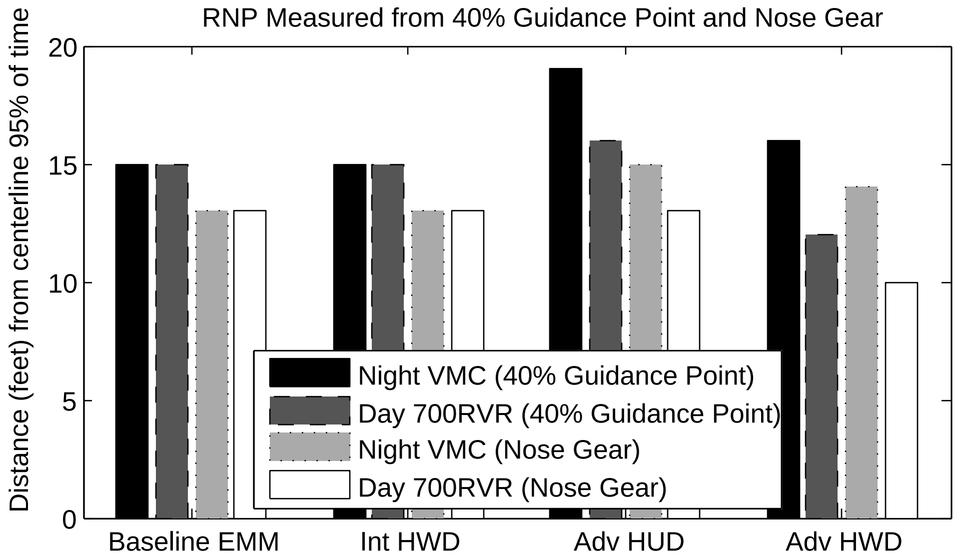


Figure 28. Nominal trial RNP for Experiment I.

The Display Condition was significant, $F(3, 10) = 3.838, p = 0.046$, for RNP performance. Post-hoc tests on the Display Condition showed two overlapping subsets: the Advanced HUD had significantly worse (statistically, but not operationally) lateral RNP during surface operations than the Advanced HWD, but had no differences with the Baseline and Intermediate EMM.

The path error computed from the different guidance points was highly significant, $F(1, 10) = 16.56, p = 0.002$. Using the nose gear to calculate the path error resulted in significantly lower lateral RNP (mean=13.0 ft) than when using the 40% Guidance Point (mean=15.375 ft) to calculate RNP. This result suggests that the EPs were tracking the taxiway centerline with the nose gear which is consistent with their training.

4.5.7 Latency

System time delays, or latencies, inherent to the head-worn (and helmet-mounted) displays have been shown to critically influence performance, utility, usability, and acceptability. Consistent definitions and measurement techniques as well as preliminary latency (time delay) requirements have been proposed for HWD S/EVS applications [31, 52]. Based upon the most stringent requirements for HMD applications of S/EVS (i.e., demanding tasks using a high resolution, large field-of-view head-worn display), the required system latency might be as low as 20 msec.

An in-situ latency measurement technique - the "windshield washer" test [31] - was used to measure the HWD latency. This test used the conformal runway scene (i.e., space-stabilized, boresighted symbology) and a grid pattern of 5° lines projected on the simulator out-the-window scene. The grid pattern was used for simulator visual scene alignment and calibration. In the presence of HWD latency, the boresight symbol cannot remain perfectly space-stabilized. The test required that the user smoothly oscillate his head in azimuth at a rate which caused the runway centerline symbol to touch the outer, target box. The aircraft was headed directly down the runway and positioned on the runway centerline. The head-movement rate data divided by the size of the target box defines the equivalent time delay (at a frequency). (Note that the latency in the head movement data is immaterial to this computation; only the average rate is needed for the equivalent delay calculation.)

The head-tracker allows "prediction" or lead compensation to be applied to the tracker data, attempting to minimize the system latency. The prediction method is not described by the head-tracker manufacturer. The latency data was taken at three prediction values: a) none (0 msec); b) 25 msec prediction; and, c) 50 msec prediction. The experiment was run using the 50 msec prediction value. The selection of a prediction value is a compromise between system latency (i.e., time delay) and noise or jitter in the HWD-displayed scene. At the 50 msec prediction value, the latency was minimized while the jitter and noise of the head-tracker output was judged during the "pre-test" trials to be "reasonable."

In Figure 29, the measured latency values are shown. Using the "windshield wiper" test, the measured latency values are unequivocally "end-to-end." It represents the total lag between the pilot's head movement until the conformal imagery is finally drawn. This latency includes not just the head-tracker, but also the communications delay between the head-tracker and the SV image generator (RS-232) and the SV image generator computational time, to name just two elements. The test was conducted three times to measure repeatability, and it also used a 5° target spacing and 2.5° spacing.

Without "prediction" from the head-tracker, the end-to-end latency was approximately 70 msec. Less than 2 msec of variation was measured between the three samples. With 25 msec prediction, the measured latency was approximately 52 msec with very little variation in measurements. With 50 msec prediction, the measured latency varied considerably depending upon the target spacing used. The measured latency was 45 msec for the 5° spacing and only 30 msec for the 2.5° spacing. This plot also includes the theoretical latency from the 70 msec nominal value if the prediction method was "perfect." The data do not match these "perfect" predictions.

The latency results do, however, match expectations because of the nature of the test. Using the "windshield wiper" technique, the frequency of the head movement (i.e., the input) to measure latency depends on the target spacing and the inherent system latency. For instance:

- For the no-prediction test, relatively slow head-movements were needed to achieve a 5° target spread; thus, it was easy to get repeatable latency data. Also, without prediction, while there was a lag in the display due to

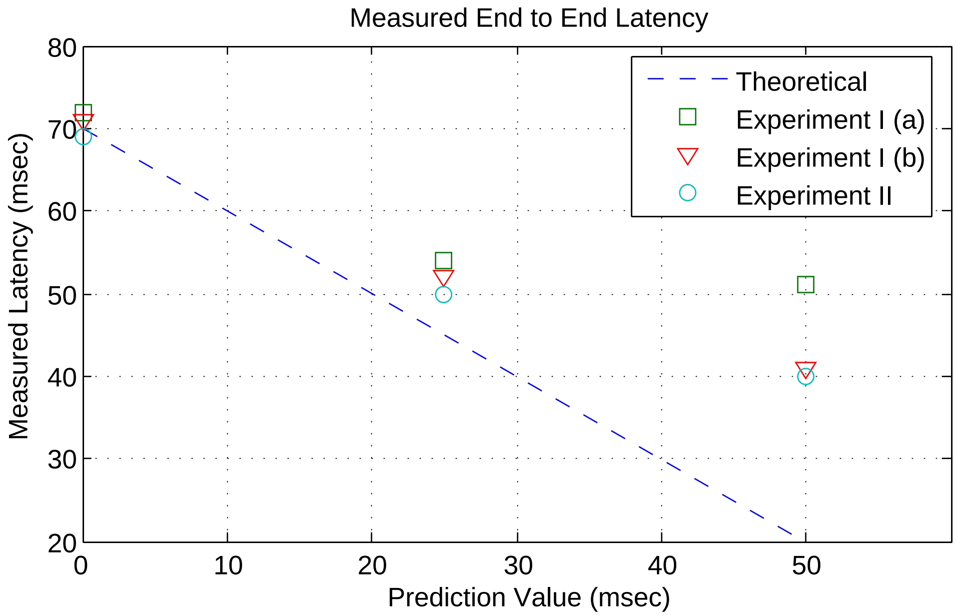


Figure 29. Measured end to end latency.

the inherent latency. There wasn't amplitude modification between the head-movement and the resultant conformal image. Therefore, a 5° head-movement resulted in a 5° image displacement.

- When prediction was used, the predictor modifies the system phase response (i.e., it provided compensation for the lag) but it also introduced amplitude modification. While the exact prediction method for the head-tracker in this test wasn't known, the frequency response of a "typical" prediction method [53] is shown in Figure 30 to illustrate the effect. The predictor attempts to keep the phase response close to 0° phase across a large frequency range; thus, minimizing the lag (delay) between the head-movement and resultant image response. However, the amplitude response of the predictor is no longer 0 dB (amplitude ratio of 1), so a 5° head movement will result in a image response of greater than 5° for frequencies where the amplitude response is greater than 0 dB.

- For a typical predictor, the higher input frequency yielded a greater resultant amplification. Since the "windshield wiper" test relied on the amplitude response of the image, a non-unity frequency response from a predictor will distort the delay (phase) measurement.

- This amplitude distortion was particularly evident in the "50 msec prediction" data in Figure 29. For the 2.5° target, lower system latency was measured, but for the 5.0° target, more latency was measured. This difference was due to the greater amplitude magnification with the higher

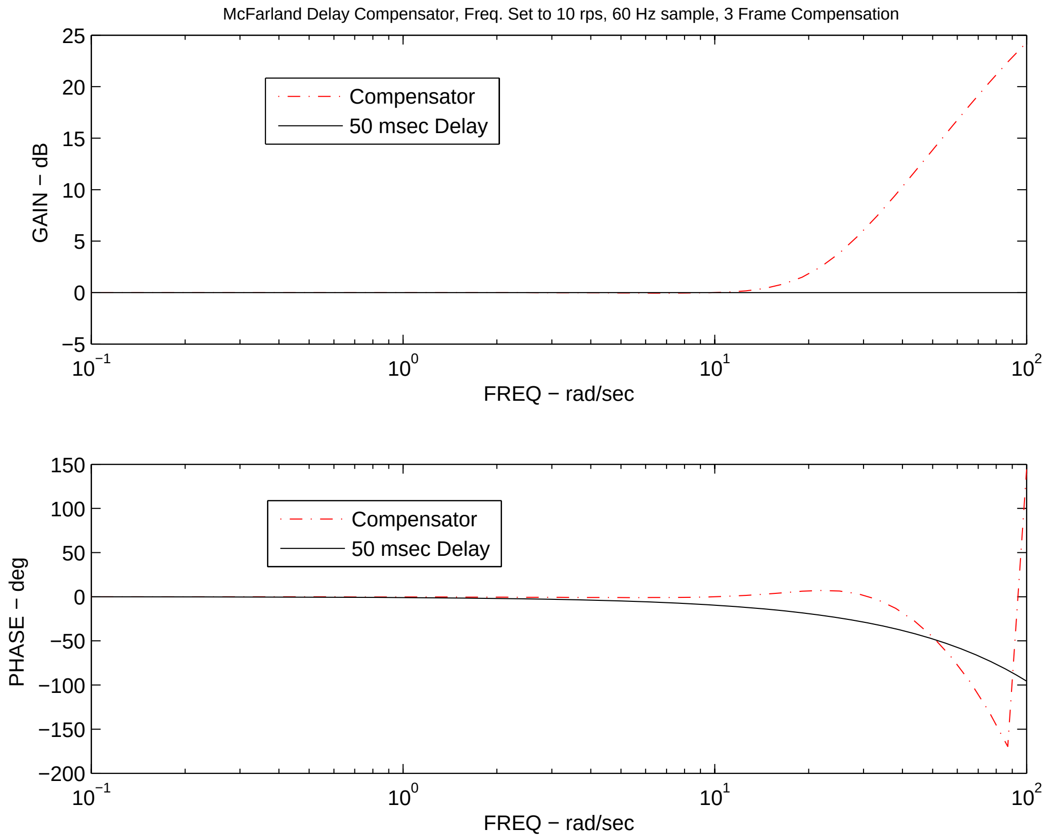


Figure 30. Typical delay compensator frequency response.

input frequency needed for the 5.0° target. It was also due to the predictor having less phase compensation at the higher frequencies.

- The higher propensity for noise and jitter in the image with prediction was also due to this amplitude distortion in the predictor.

A single latency value cannot be associated with the system used in this experiment due to the amplitude and phase characteristics with a predictor in the loop, but the data showed that the tested latency was between 30 and 50 msec.

4.5.8 Head Tracking

A comparative evaluation of the EP's head movement with and without a HWD was not performed in the experiment. The experiment was designed to be a "pure" comparison of head-down, head-up, and head-worn displays, so no head-tracking apparatus was worn by the EP during the evaluation of head-up and head-down display concepts. Non-contact head-tracking methods were neither economically nor practically viable for implementation in this experiment. Subsequent experiments will address whether a HWD affects a pilot's head-movement compared to present-day operations.

The captain's head movement when wearing the HWD was analyzed. Collapsing the data across all captains and all maneuvers, the percent of time the captain was looking in azimuth and elevation is shown in Figures 31 and 32, respectively. Both figures show a high concentration of time within ±45° of the aircraft centerline, particularly for elevation. The EP's head was positioned within ±25° of elevation 98% of the time. The EP's head was positioned within ±50° of azimuth 97% of the time. The azimuth data shows a slightly longer "tail" reflecting the need for the EPs to clear a target and maneuver at angles exceeding 45° off-boresight, but the data is clearly different than that shown for a typical fighter aircraft head-worn display application [34]. As could be expected, the need for off-boresight capability is necessary for commercial applications.

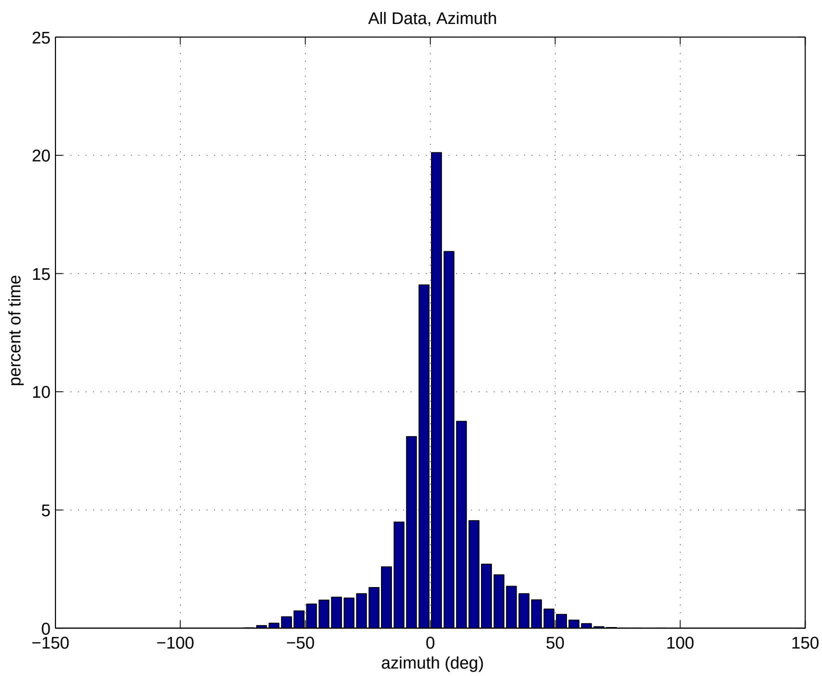


Figure 31. Percentage of time for head direction in Azimuth (deg).

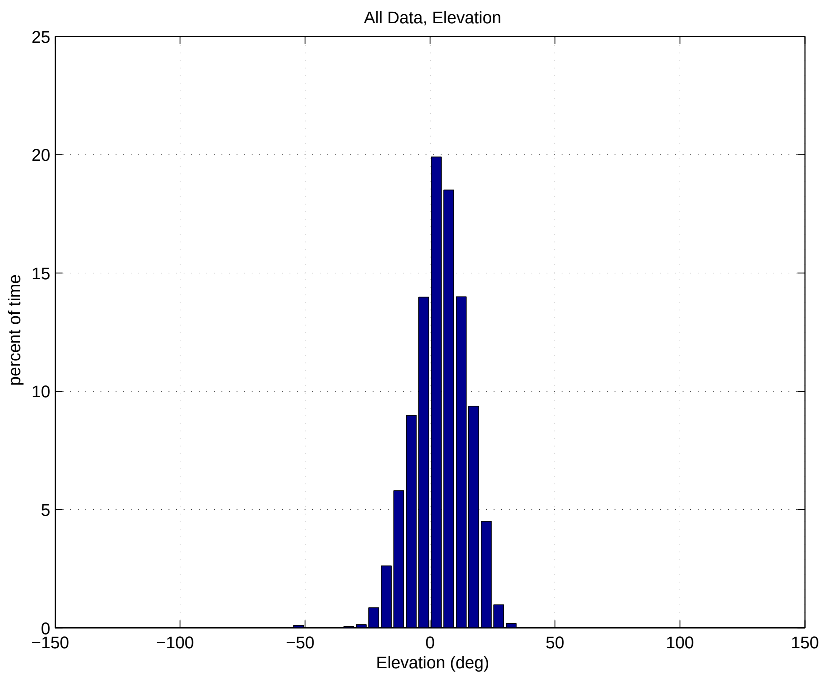


Figure 32. Percentage of time for head direction in Elevation (deg).

The rates of head movement when wearing the HWD were analyzed. Collapsing the data across all captains and all maneuvers, the captain's head movement rates in azimuth and elevation, expressed by percentage time, is shown in Figures 33 and 34, respectively. Both data show a high concentration of time within ±45 degrees/sec, particularly for elevation. The rate of head movement was within ±30 degrees per second in elevation 97% of the time. The rate of head movement was within ±60 degrees per second in azimuth 97% of the time. Note that the total percentage of time when rates were greater than 100 deg/sec or less than -100 deg/sec are plotted in Figure 33 and 34, at the ±100 deg/sec point (for readability). The azimuth rate data show that the pilots typically used relatively slow head-movements to clear for traffic in maneuvers but about 3% of the time, the pilots exceeded 100 deg/sec head-movement. Maximum head rates in azimuth of 200 to 300 deg/sec were found. These data indicate much slower head movement for commercial as opposed to military applications, but a small, yet significant percentage of time still requires very fast head-rates and tracking capability. This same characteristic was not found in the elevation data.

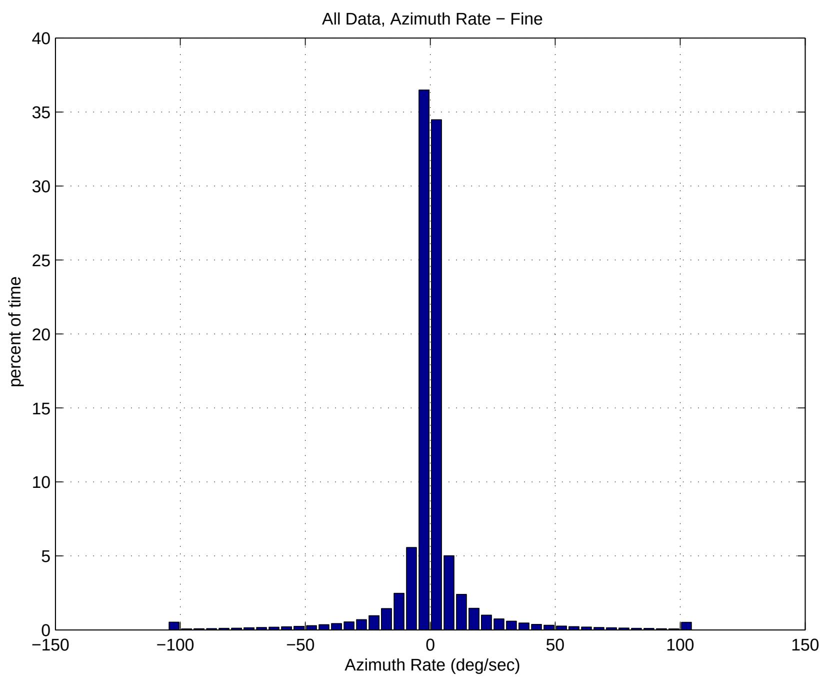


Figure 33. Percentage of time for head movement - Azimuth rates (deg/sec).

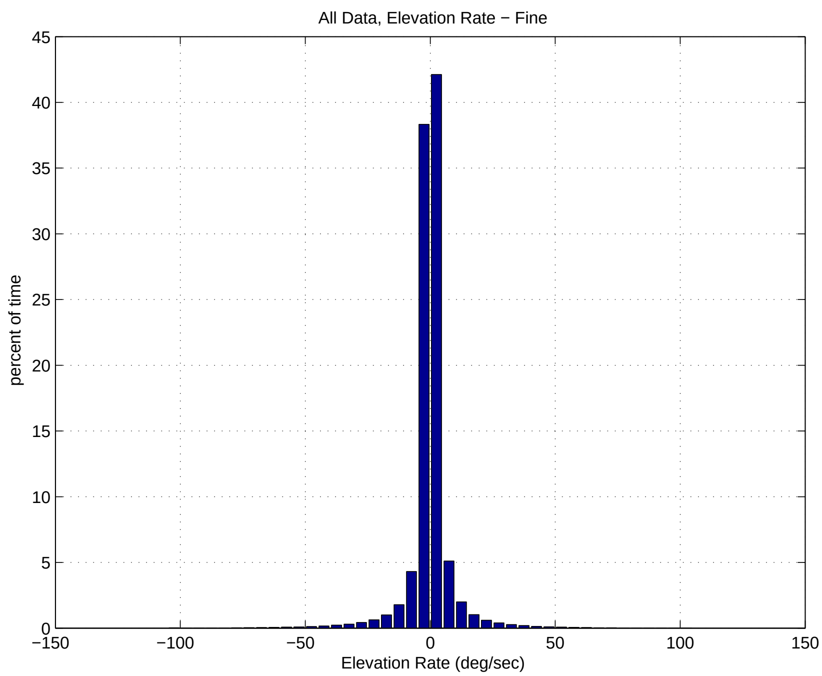


Figure 34. Percentage of time for head movement - Elevation rates (deg/sec).

4.6 Post-Run Questionnaire Results

Several questionnaires were given at the end of each data run. At the end of the day, paired-comparison questionnaires were given to both the captain and first officer. The post-run questionnaires used in this experiment are in Appendix C, Figures C6 through C10.

4.6.1 NASA Task Load Index

An ANOVA was performed for the dependent variable of task load index (mental workload) from flight crew ratings on the NASA Task Load Index (NASA-TLX) scale [54]. No significant differences were found for the display condition, $p > 0.05$.

4.6.2 Situation Awareness Rating Technique

An ANOVA was performed for the dependent variable of situation awareness derived from flight crew ratings (0-100) on the Situation Awareness Rating Technique (SART) [55]. SA is defined as Understanding - (Demand - Supply). Analysis found a significant effect for display condition, $F(3, 15) = 4.16, p < 0.05$. A SNK test revealed two unique subsets: (1) Advanced HWD (133.04) and Advanced HUD (130.23) - highest SA; and, (2) Baseline (104.07) and Intermediate HWD (112.30) - lowest SA. The SA provided by the Advanced HWD and Advanced HUD was not significantly different from each other, nor were the Baseline and Intermediate HWD significantly different from each other (Fig. 35).

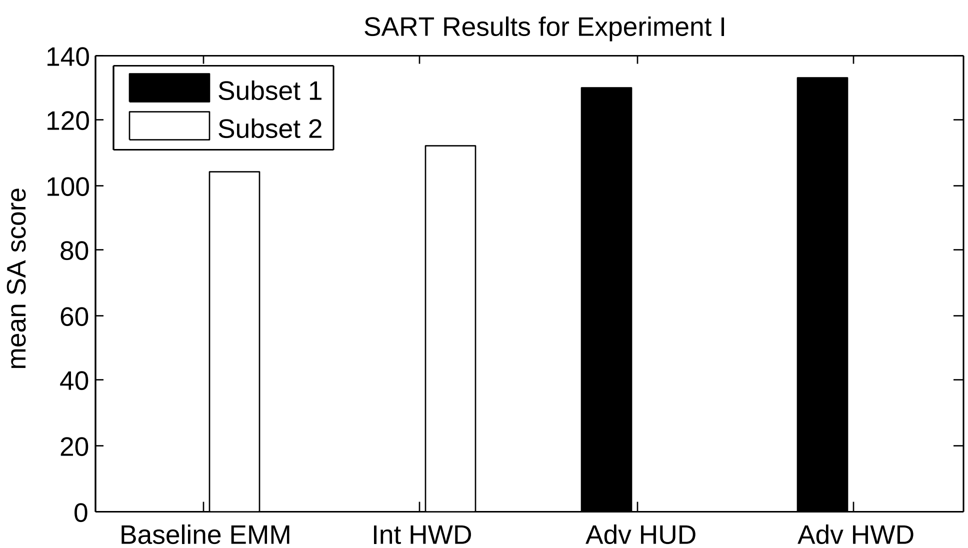


Figure 35. SART results for Experiment I. Higher scores denote higher SA.

4.6.3 Taxi Situation Awareness Questions

Flight crews were administered a Likert post-run experimental questionnaire (1 to 5 scale; 1 = "not at all"; 5 = "very much") to rate the contribution of the display conditions to taxi efficiency, overall navigation awareness, route awareness of local controller clearance, route awareness of ground controller clearance, surface traffic awareness, directional awareness, and taxi safety. An ANOVA revealed significant effects for all dependent variables, $p < 0.05$. Post-hoc SNK tests were performed on these dependent variables resulting in two unique subsets: (1) Advanced HUD and Advanced HWD, and (2) Intermediate HWD and Baseline. Means for each dependent variable are presented in Figure 36.

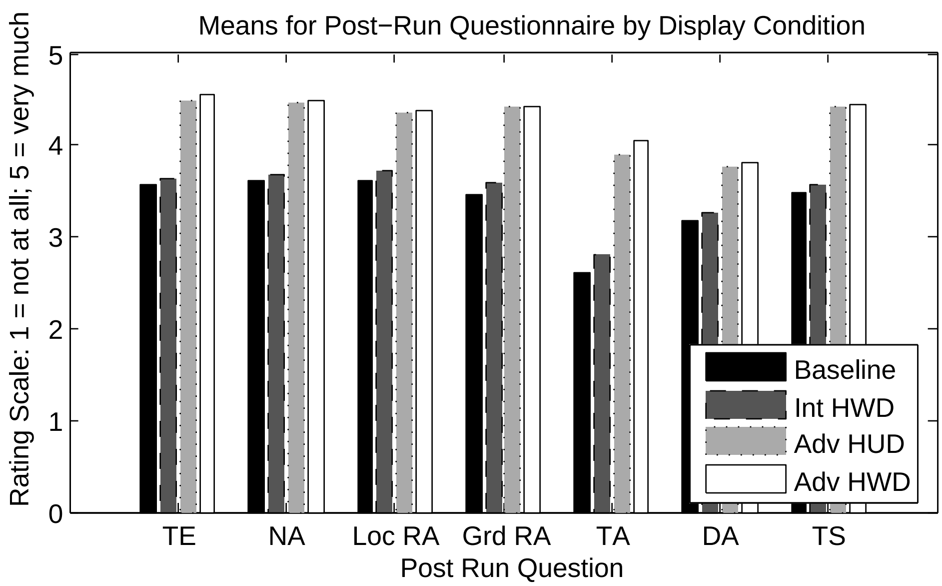


Figure 36. Means for post-run questionnaire by display condition. TE = Taxi Efficiency; NA = Overall Navigational Awareness; Loc RA = Route Awareness of the Local Controller Clearance; Grd RA = Route Awareness of the Ground Controller Clearance; TA = Surface Traffic Awareness; DA = Directional Awareness; TS = Taxi Safety.

4.6.4 Situation Awareness Global Assessment Technique

During training, the EPs were instructed that certain data collection trials would be Situation Awareness Global Assessment Technique (SAGAT) trials. At a certain point during the SAGAT trial, the simulation was "frozen" in time and the displays were blanked out. The EPs were then asked 12 questions (Table 3) regarding their situation awareness. The crews were not aware of which trials contained SAGAT questionnaires. For each of the 16 crews, 4 trials contained a SAGAT questionnaire.

The SAGAT data was analyzed using a Cochran's Q test to evaluate for statistical differences between display types. Table 3 presents the statistical results of the analyses collapsed across pilot type. Correct responses to each question were based on an "acceptable tolerance band around the actual value" (p. 170, [56]).

Of the 12 SAGAT questions, 5 questions were significantly different among the display concepts. These were Questions 1-4 and Question 11. Figure 37 shows correct crew responses to the questions with significant differences. From Figure 37, display concepts with a path displayed (i.e., advanced display concepts) had significantly more correct responses for 5 of the 12 questions.

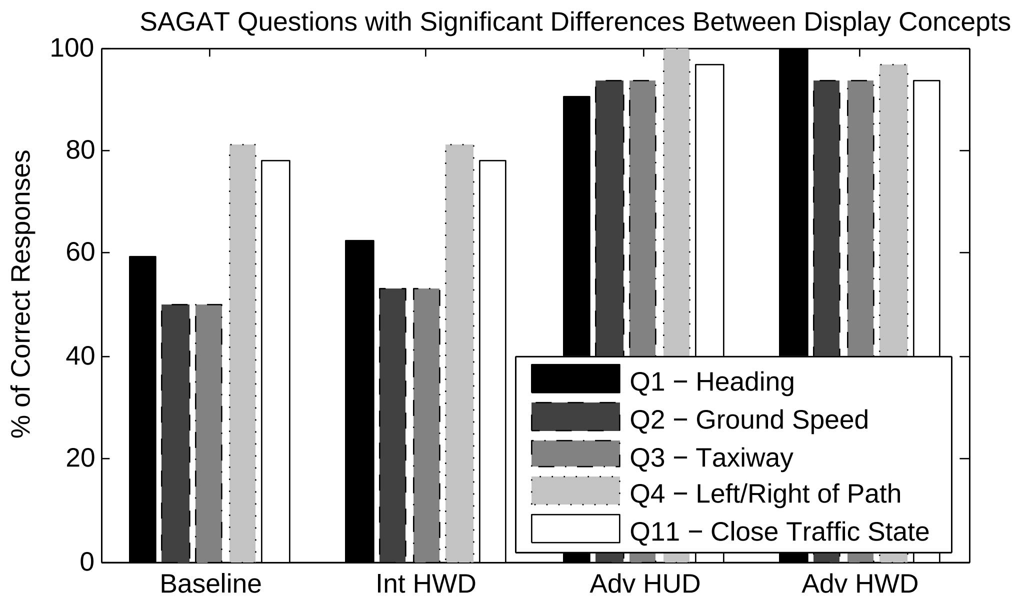


Figure 37. SAGAT Results for Significant Questions

For all concepts, the heading and ground speed were displayed in the same location on the EMM. The ground speed and heading were displayed on all head up display concepts (Intermediate HWD, Advanced HUD and Advanced HWD). As the Intermediate HWD had this data head up, this suggests that the workload for the Baseline and Intermediate HWD is such that the data cannot be scanned or retained.

Questions 3 and 4 regarded the current taxiway and path error awareness. This information was directly displayed in the advanced concepts but had to be estimated by the pilot for the Baseline and Intermediate HWD concept. Note that there were no statistically significant differences among the concepts for the question regarding the crew's knowledge of the full taxi clearance. It can be inferred that workload is higher with the Baseline and Intermediate HWD concepts and that this tactical information is not immediately known while the strategic knowledge of the full taxi clearance is maintained. The workload is such that the information is not scanned or retained.

Table 3. Situation Awareness Global Technique Results.

SAGAT Question	**Cochran's Q**	**Frequencies Ratio = correct/incorrect**			
		Baseline	Int HWD	Adv HUD	Adv HWD
Q1. What was your present heading?	$Q(3,32) = 21.6, p < 0.001$	19/13	20/12	29/3	32/0
Q2. What was your present ground speed?	$Q(3,32) = 27.07, p < 0.001$	16/16	17/15	30/2	30/2
Q3. Which taxiway is the aircraft presently located?	$Q(3,32) = 27.07, p < 0.001$	16/16	17/15	30/2	30/2
Q4. Is the aircraft currently X left X center X right of taxiway centerline?	$Q(3,32) = 11.18, p < 0.01$	26/6	26/6	32/0	31/1
Q5. What was your full taxi clearance?	$Q(3,32) = 5.148, p > 0.05$	18/14	23/9	24/8	26/6
Q6. Which taxiway did the ground controller next expect the aircraft to taxi to from present location?	$Q(3,32) = 3.17, p > 0.05$	28/4	28/4	31/1	27/5
Q7. Which direction were you to next turn the aircraft onto taxiway?	$Q(3,32) = 2.63, p > 0.05$	25/7	25/7	29/3	26/6
Q8. Estimate the time of taxi route completion from present location to concourse in seconds.	$Q(3,32) = 1.50, p > 0.05$	28/4	28/4	30/2	30/2
Q9. How many aircraft were you aware of (perceived out-the-window and/or display concept(s))?	$Q(3,32) = 4.13, p > 0.05$	23/9	25/7	27/5	29/3
Q10. Where was the location of the nearest aircraft to your aircraft?	$Q(3,32) = 4.13, p > 0.05$	23/9	25/7	27/5	29/3
Q11. What was the heading and ground speed of nearest aircraft to your ownship heading and ground speed?	$Q(3,32) = 8.20, p < 0.05$	25/7	25/7	31/1	30/2
Q12. Please provide a description of the objective of the nearest aircraft to your ownship.	$Q(3,32) = 2.90, p > 0.05$	23/9	25/7	28/4	27/5

One question regarding traffic was significant. Crews still rely on and are trained to scan for traffic out the window. However, regarding other traffic speed and heading, the advanced concepts showed statistically significant improvement. From Figure 37, crews with the advanced concepts answered Question 11 correctly 90% of the time as opposed to the Baseline and Intermediate HWD concepts about 80% of the time. This is probably not operationally significant, but is further evidence of the increased situation awareness and lower workload of the advanced concepts.

4.6.5 Simulator and HWD-Induced Sickness

HWDs have been found to induce symptoms of motion or simulator sickness. Further, immersion into a virtual reality environment can cause symptoms similar to simulator sickness, but with differing magnitudes of severity [57]. The (potential) causes are many [58]. The occurrence of simulator sickness in operation or training with HWDs for commercial aviation applications would be problematic and detrimental to the commercial applications of HWDs if this were the case [59].

During the usability study prior to this experiment, some cases of simulator sickness were experienced but objective data of the sickness severity was not taken. For this experiment, simulator sickness questionnaires [60] were used to track the presence and trends in simulator sickness as it might be induced by the HWD or the simulator through the course of the experiment. The simulator sickness questionnaire considers 16 symptoms (Table 4), belonging to the three factors of oculomotor disturbance, disorientation, and nausea. The EPs then rated how they felt for each symptom: None, Slight, Moderate, Severe. The computation of a Simulator Sickness Questionnaire (SSQ) "score" was obtained by weighting the three factors and the 16 symptoms according to Kennedy [60] to yield a total score.

Table 4. Simulation Sickness Questions

Nausea	**Oculomotor**	**Disorientation**
General Discomfort	General Discomfort	Difficulty Focusing
Increased Salivation	Fatigue	Fullness of Head
Sweating	Headache	Blurred Vision
Nausea	Eye Strain	Dizzy (Eyes Open)
Difficulty Concentrating	Difficulty Focusing	Dizzy (Eyes Closed)
Stomach Awareness	Difficulty Concentrating	Vertigo
Burping	Blurred Vision	

The display configurations were experimentally blocked. After training, the first block consisted of evaluations using "conventional" displays (head-down and head-up display concepts), followed by a second block consisting of evaluations using the HWD concepts. This pattern was repeated twice more.

Simulator sickness questionnaires were administered at the end of each data run.

For 11 out of the 16 crews, no simulator sickness symptoms were noted in any of the runs. For the other 5 crews, there were 8 cases of significant simulator sickness. For this paper, total SSQ scores of 15 and above [57] were considered to be an indicator of significant simulator sickness. Five simulator sickness cases were experienced by the captain (who was wearing the HWD) - relatively severe in two instances. The other three cases of simulator sickness symptoms were experienced by the first officer. The first officer was not using a HWD. The reasons for the first officer sickness was probably due to a combination of simulator latency, the lack of motion cues, fatigue over the course of the experiment, and the need to repeatedly go head-down to write taxi clearances and then go head-up to monitor taxi performance.

In the most severe case, the session had to be stopped because of pilot discomfort. The contributing factor was not felt to be related to the HWD in this case. A couple of "mid-run" simulator resets were unfortunately performed and the sudden visual "rush" from these resets caused the pilot to experience severe symptoms. On Run 17 (out of a planned 27 runs), the pilot rated a total SSQ score of 29.9 with a Nausea score of 57, followed on the next run with a total SSQ score of 45 with a Nausea score of 76. Prior to these runs, no symptoms were reported. The test was subsequently stopped for this crew. More careful resets were subsequently enforced in the simulation protocol (i.e., stopping the aircraft before the reset and warning the pilots of an up-coming reset so they could close their eyes, if desired).

For the remaining four affected crews, three of the crews reported significant simulation sickness symptoms (i.e., total SSQ scores over 15). The mean of the total simulator sickness scores by the captain within the display block is shown in Figure 38. For these EPs, simulator sickness was generally "triggered" by the first HWD evaluation block. The severity of the sickness symptoms also generally increased over time, analogous to other research findings [58]. For these three captains with significant SSQ scores, the symptoms were primarily related to oculomotor disturbances which is consistent with simulator sickness; however, for the more severe symptoms noted by the captains of Crew 2 and Crew 3, commensurate nausea and, to a lesser extent, disorientation symptoms were also noted.

The SSQ data suggests a "good news / bad news" scenario for the commercial viability of HWDs. The good news is that 75% of the evaluation pilots who used the HWDs experienced no or mild symptoms of simulator sickness. The bad news is that in 12.5% of the evaluation pilots who used the HWD did experience notable and in one case, relatively severe symptoms of simulator sickness. Although there is no guidance as to what an acceptable level of the affected pilot population would be, 12.5% intuitively seems *unacceptable*.

The general and individual-specific differences that influence the propen-

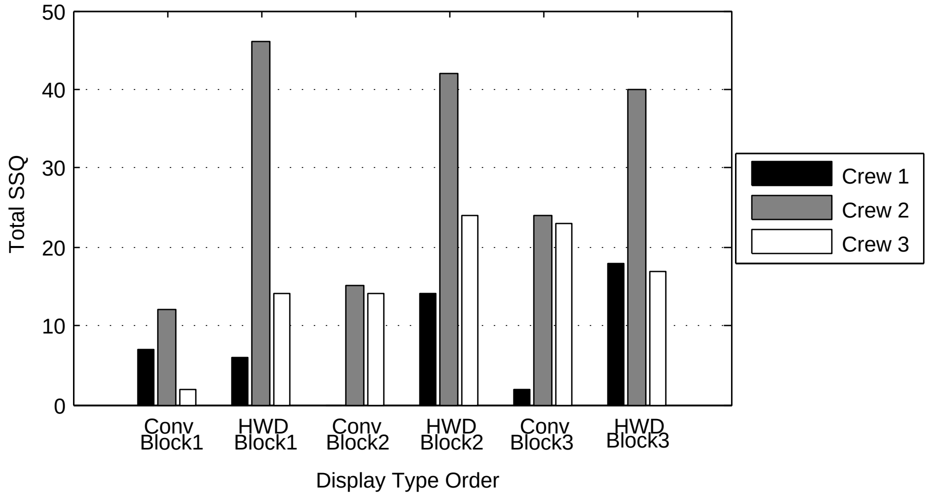


Figure 38. Mean total Simulator Sickness Questionnaire (SSQ) score within an Evaluation Block.

sity for simulator sickness when using an HWD needs to be further explored. Potential influences that the research team felt may have contributed in this experiment were the EP's most recent flight experience, their fatigue and general health at the time of the simulation, and their propensity for developing virtual "presence" and visual-vestibular conflict. Further, the lack of motion cues in the simulator may have contributed to the simulator sickness results (e.g., several first officers experienced simulator sickness symptoms even though they didn't wear the HWD). Most importantly, however, future investigations must evaluate whether there is any propensity for "simulator sickness" with these devices in operational use versus simulator evaluations.

4.7 Post-Test Paired Comparison Results

A MANOVA statistical procedure was performed on four paired comparison scales administered to the captain and first officer of each flight crew. The paired comparison scales asked the pilot to factorially evaluate each of the four display concepts in comparison to one another on four constructs: Situation Awareness (SA-SWORD) [43], Mental Workload (Subjective Workload Dominance (SWORD)) [61], Taxi Efficiency, and Taxi Safety. The analyses were conducted separately for captain and first officer responses. Significant results reported are at the $p < 0.01$ significance level.

4.7.1 Situation Awareness

For the captain ratings, there was a significant main effect found for situation awareness, $F(3, 45) = 54.49$. A post-hoc test (Table 5) revealed that (a) the Baseline moving map display provided significantly lower in situation awareness than the other three display concepts; (b) the Intermediate HWD provided significantly lower situation awareness than the Advanced HUD and Advanced HWD; and (c) no significant differences between the Advanced HUD and Advanced HWD concepts.

Table 5. Significance Table for Situation Awareness (SA-SWORD) Paired Comparison Ratings

Display Comparisons	CAPTAIN		FIRST OFFICER	
	Significant	Non-Significant	Significant	Non-Significant
EMM - Int HWD	X			X
EMM - Adv HUD	X		X	
EMM - Adv HWD	X		X	
Int HWD - Adv HUD	X		X	
Int HWD - Adv HWD	X		X	
Adv HUD - Adv HWD		X		X

For the first officer ratings, there was a significant main effect found for situation awareness, $F(3, 45) = 32.38$. A post-hoc test (Table 5) indicated that the first officers rated (a) the Advanced HUD and Advanced HWD to be significantly better for taxi efficiency than both the Intermediate HWD and moving map display which (b) were not significantly different from each other and (c) no significant differences between the Advanced HUD and Advanced HWD concepts.

4.7.2 Mental Workload

For the captain ratings, there was a significant main effect found for mental workload, $F(3, 45) = 5.28$. A post-hoc test (Table 6) revealed (a) the Advanced HUD and Advanced HWD to be significantly lower in Mental Workload than both the Intermediate HWD and the Baseline moving map display which (b) were not significantly different from each other and (c) no significant differences between the Advanced HUD and Advanced HWD concepts.

Table 6. Significance Table for Mental Workload Paired Comparison Ratings (SWORD)

Display Comparisons	CAPTAIN		FIRST OFFICER	
	Significant	Non-Significant	Significant	Non-Significant
EMM - Int HWD		X		X
EMM - Adv HUD	X		X	
EMM - Adv HWD	X		X	
Int HWD - Adv HUD	X			X
Int HWD - Adv HWD	X			X
Adv HUD - Adv HWD		X		X

For the first officer ratings, there was a significant main effect found for mental workload, $F(3, 45) = 7.988$. A post-hoc test (Table 6) reported differences only between the Baseline moving map display and the Advanced HUD and Advanced HWD display concepts. No other effects were significant.

4.7.3 Taxi Efficiency

For the captain ratings, there was a significant main effect found for taxi efficiency, $F(3, 45) = 23.655$. A post-hoc test (Table 7) revealed (a) the Advanced HUD and Advanced HWD to be significantly better for taxi efficiency than both the Intermediate HWD and the Baseline moving map display which (b) were not significantly different from each other and (c) no significant differences between the Advanced HUD and Advanced HWD concepts.

Table 7. Significance Table for Taxi Efficiency Paired Comparison Ratings

Display Comparisons	CAPTAIN		FIRST OFFICER	
	Significant	Non-Significant	Significant	Non-Significant
EMM - Int HWD		X	X	
EMM - Adv HUD	X		X	
EMM - Adv HWD	X		X	
Int HWD - Adv HUD	X		X	
Int HWD - Adv HWD	X		X	
Adv HUD - Adv HWD		X		X

For the first officer ratings, there was also a significant main effect found for taxi efficiency, $F(3, 45) = 48.63$. A post-hoc test (Table 7) revealed (a) the Baseline moving map display to be significantly poorer for taxi efficiency than the other three display concepts; (b) the Intermediate HWD to be significantly poorer than the Advanced HUD and Advanced HWD; and (c) no significant differences between the Advanced HUD and Advanced HWD concepts.

4.7.4 Taxi Safety

For the captain ratings, there was a significant main effect found for taxi safety, $F(3, 45) = 23.859$. Post-hoc tests (Table 8) revealed (a) the Advanced HUD and Advanced HWD to be significantly higher in reported surface operations and taxiing safety than both the Intermediate HWD and moving map display which (b) were not significantly different from each other, and (c) no significant differences between the Advanced HUD and Advanced HWD concepts.

Table 8. Significance Table for Taxi Safety Paired Comparison Ratings

Display Comparisons	CAPTAIN		FIRST OFFICER	
	Significant	Non-Significant	Significant	Non-Significant
EMM - Int HWD		X	X	
EMM - Adv HUD	X		X	
EMM - Adv HWD	X		X	
Int HWD - Adv HUD	X		X	
Int HWD - Adv HWD	X		X	
Adv HUD - Adv HWD		X		X

For the first officer ratings, there was also a significant main effect found for taxi safety, $F(3, 45) = 45.19$. A post-hoc test (Table 8) revealed (a) the Baseline moving map display to be rated significantly lower in perceived taxi safety than the other three display concepts; (b) the Intermediate HWD to be significantly lower in perceived taxi safety than the Advanced HUD and Advanced HWD; and (c) no significant differences between the Advanced HUD and Advanced HWD concepts.

4.8 Discussion of Experiment I Results

Crews were asked to perform fairly complex taxi maneuvers for various display and weather conditions. From Figure 23, it can be seen that the best navigational performance was with the Advanced HUD display concept regardless of the weather condition. The Advanced HWD display concept had 4 minor navigational errors, where the crew knew immediately that they made a mistake and informed the ground controller and requested instructions. However, for the Baseline and Intermediate HWD concepts, the crew didn't realize a mistake had been made. The crews continued on the wrong course unaware they were no longer on the cleared path resulting in major navigation errors. With the exception of the SA paired comparison result, there were no significant differences between the Baseline EMM and Intermediate HWD concepts. Pilots commented that the Intermediate HWD concept was advantageous because it presented the taxi signage in a head-up and familiar format.

From the post-test analysis, it can be seen that there were no significant differences between the Advanced HUD and Advanced HWD concepts for either the captains or the first officers for all of the paired comparison questionnaires. For the first officers, who did not have a head-up device, significant differences for SA were reported between all concepts that compared concepts that displayed route/traffic information (Advanced HUD and Advanced HWD) with display concepts that did not display route or traffic information (Baseline and Intermediate HWD) (see Table 5). From the first officer's perspective, the two displays that varied across all display concepts were the EMM and the head-up repeater display. Further, first officers commented that the repeater display was a distraction and ignored it most of the day.

For the nose-to-nose rare event scenario, all but two of the crews who had traffic displayed were able to avoid the nose-to-nose situation. The rare event showed that having traffic displayed was a significant enhancement to the crew's situation awareness. Crews that avoided the nose-to-nose were able to notify the ground controller of the traffic on their cleared path and ask for new instructions. The two crews that did not notice the traffic stated that the color of the traffic symbol made it difficult to see the traffic icon, especially at large range scales. They also said it was hard to remember which scenarios had displayed traffic. Half of the data runs did not display traffic (Baseline and Intermediate HWD) and these runs were randomly assigned in the run sequence. Therefore, a follow-on experiment was warranted and thus, designed to address the readability/discernibility of traffic.

5 Experiment II

Experiment II was conducted closely following Experiment I to principally investigate several issues uncovered in Experiment I. Also, some experiment configurations were modified to take advantage of the additional test opportunity. From the rare event results of Experiment I, the traffic icons were found to be difficult to distinguish, particularly on the head-down display. In Experiment I, a brown color was used for traffic icons to conform with ADS-B color-coding standards for ground traffic. For Experiment II, a cyan color was used to increase contrast. Further, the traffic icon sizes were scaled as a function of the range scale.

In addition to this change, the taxi director insert display on the advanced HWD concept was modified to improve readability and computational frame rates. Also, from Figure 24, most errors in Experiment I were made with Route 20. Therefore, the ground controller clearance for Route 20 was changed to provide the direction of the turn in the route prior to where most crews mistakenly turned the wrong way.

5.1 Evaluation Pilots

Twelve commercial flight crews (a captain and first officer) participated in the experiment. The 12 crews for Experiment II did not include anyone who had participated in Experiment I. Each flight crew flew for the same company to ensure crew coordination and cohesion with regard to surface operation procedures. The captains had an average of over 15,000 flight hours with 29 years total flight time and the first officers had an average of over 9,000 flight hours with an average of 26 years total flight time. Two-thirds of the captains required corrective lenses. The crews were given a 45-minute briefing on the display concepts and the evaluation tasks. After the briefing, a 45-minute training session was conducted to familiarize the EPs with the RFD simulator, the HUD, the HWD device, and the piloting task. Only the captain had a HUD or HWD; the first officer had a head-down repeater display of the captain's head-up device. An eye dominance test was performed after the training briefing. Of the 12 captains, 11 were right eye dominant. The HWD was viewed with the right eye for all EPs. The HWD is compatible with eyeglasses. Following training, 2.5 hours of data collection was conducted. Twelve of the 24 scenarios from Experiment I were replaced with very short taxi routes, thus allowing for shorter data collection time. The total experiment time for each crew was approximately 4 hours.

5.2 Evaluation Task

The evaluation task for Experiment II was the same as Experiment I with the following exceptions:

1. A total of 25 (instead of 27) taxi scenarios were used in the study.
2. Twelve of the 24 non rare event routes from Experiment I were shortened due to time constraints.
3. There was only one rare event scenario, the nose-to-nose scenario from Experiment I.

5.3 Display Conditions

The display conditions (Fig. 39 and Appendix G) for Experiment II were the same as Experiment I with the following exceptions:

1. The Baseline display condition was replaced with a paper chart and existing cockpit displays, and
2. The Intermediate HWD condition was replaced with a head-down only display, an Advanced EMM. The Advanced EMM included iconic traffic, clearance information and the cleared route.

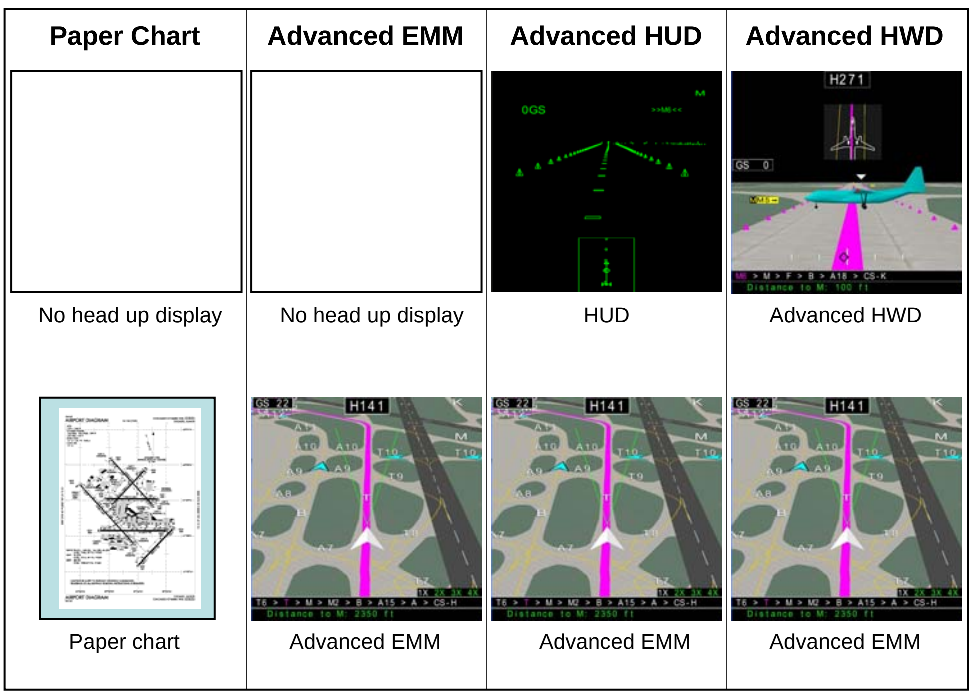


Figure 39. Thumbnail pictures of the four display conditions for Experiment II: Paper Chart, Advanced EMM, Advanced HUD and Advanced HWD.

There was no head-up or head worn display with either the Advanced EMM condition or the Paper Chart condition. The traffic icons on the Advanced EMM and the Advanced HWD were modified to improve the EPs ability to acquire the traffic (Appendix G, Figs. G1 and G4). The color was changed from brown to cyan. For the Advanced EMM, the traffic chevrons were scaled in size as a function of the range scale. As the range scale increased (zoomed out), the traffic chevron was increased in size to improve readability. Also, for the Advanced HWD, directional and strobe lights were added to the 3-D traffic models. Note that the out-the-window scene models had directional and strobe lights.

In addition, the taxi director insert display on the Advanced HWD was modified to improve readability and frame computation speed. In Experiment I, the airport database was rendered in the insert window. For Experiment II, the airport model was replaced with a simple model of the runway and taxiway outlines. The result was a greatly simplified airport database, which improved rendering speed as well as readability by providing only essential information. The gain in computational speed allowed the virtual scene to be rendered within a 60 Hz frame. Additionally, the simplified airport database had greater contrast between the cleared route and the background, thus readability was improved in preliminary testing.

5.4 Quantitative Results

5.4.1 Taxi Performance for All Data Runs

A MANOVA on RMS path error, taxi speed, and time-to-taxi yielded significant effects for display condition ($F(9, 363) = 4.18, p < 0.001$), visibility condition ($F(3, 149) = 5.71, p = 0.001$) and their interaction ($F(9, 363) = 3.60, p < 0.001$) according to Wilk's Lambda. Follow-up univariate ANOVAs indicated that time-to-taxi was not significantly ($p > 0.05$) affected by display condition but taxi speed and RMS path error were. Post-hoc SNK tests revealed two unique subsets of the display conditions for taxi speed and for RMS path error (Fig. 40). The EPs taxied significantly slower and had more path error for the Paper Chart condition than with the Advanced EMM, Advanced HUD, and Advanced HWD. On average, the EPs had 5.4 ft less path error and taxied 0.4 knots quicker during Day 700 RVR conditions than in Night VMC, but these improvements have little operational significance.

5.4.2 Navigational Errors

As in Experiment I, navigational errors were divided into two categories: major and minor. Also, as with Experiment I, a blunder error, which involved a conflict with another aircraft, was accounted for in a different measure and not included as a navigational error. A total of 14 navigational errors were made, where 7 were classified as major errors and 7 were classified as minor. Most of the major errors occurred with the Paper Chart condition. Figure 41 shows the navigational errors made per display condition and weather event (Night VMC or Day 700 RVR). Figure 42 shows the number of navigational

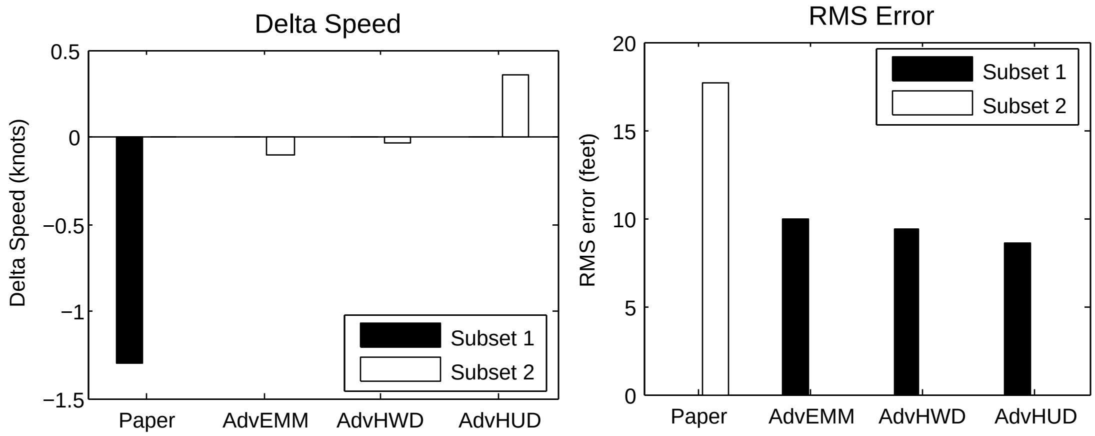


Figure 40. Delta taxi speed and RMS error for all trials for Experiment II.

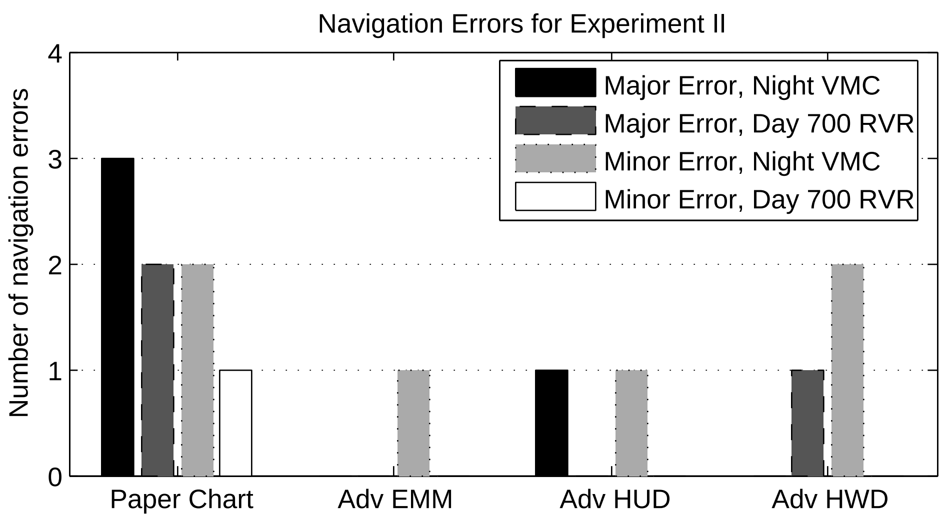


Figure 41. Navigational errors for Experiment II.

errors made per taxi route.

Cochran's Q statistical analyses revealed non-significant results for the number of navigation errors associated with the main effects of display type, $Q(3, N = 12) = 3.00, p > 0.10$; and visibility condition, $Q(1, N = 12) = 1.40, p > 0.10$, due largely to the low power of the data analysis.

5.4.3 Taxi Conflict Events

A taxi conflict event was defined as a collision with another aircraft or making a turn in front of another aircraft and creating a close call. A total of 2 taxiway conflict events occurred, one with the Advanced EMM condition and one with the Advanced HUD condition. Because of the low number of observations and the expected zero values, a statistical analysis was not conducted on the data.

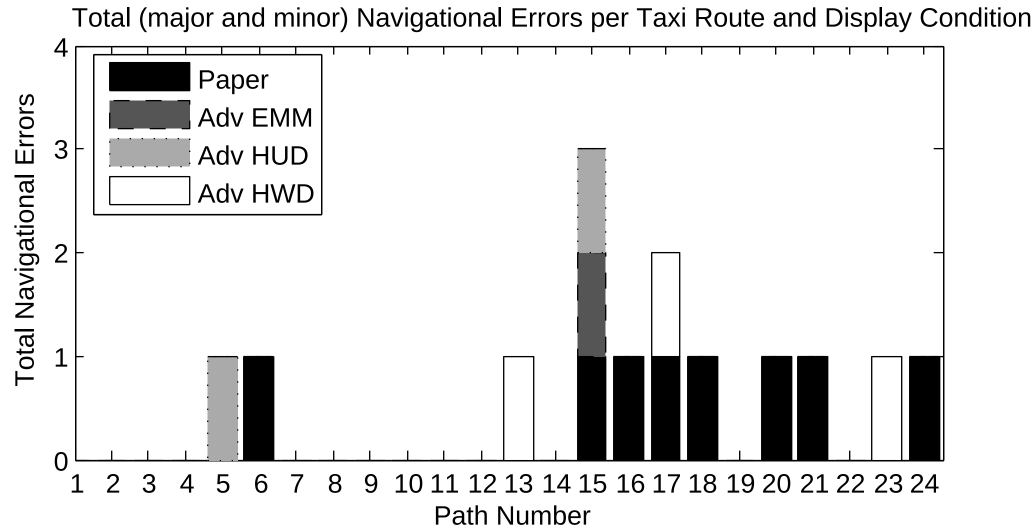


Figure 42. Navigational errors per route for Experiment II.

5.4.4 Rare Event

Experiment II included one rare event scenario. The rare event was the final run of the day and created the same nose-to-nose traffic conflict used in Experiment I. No EPs from Experiment I were used; thus, the EPs in Experiment II were unaware of the scenario. Again, this event did not create a collision scenario but instead represented a more likely and common situation where two aircraft could come nose-to-nose. Crews were given a ground controller instruction to turn onto a taxiway that was occupied by another aircraft (a small commuter jet). The visibility was reduced to 500 RVR for this scenario so that the traffic was difficult to see, but was detectable in the out-the-window scene.

The rare event display condition was evenly distributed between each of the four display conditions; therefore each of the four display conditions had 3 rare event data points. For the Paper Chart condition, which did not have path or traffic information, all three crews got into a nose-to-nose situation. For the display conditions that had iconic traffic display (Advanced EMM, Advanced HUD and Advanced HWD), all crews were able to avoid the nose-to-nose situation.

5.4.5 Taxi Performance for Nominal Runs

A second analysis of the taxi performance measures was conducted with the off-nominal runs excluded from the analysis. A MANOVA on RMS path error, taxi speed, and time-to-taxi yielded significant effects for display condition ($F(9, 300) = 4.62, p < 0.001$) and the interaction between display condition and visibility ($F(9, 300) = 1.95, p = 0.045$) according to Wilk's Lambda.

There were no significant ($p > 0.05$) visibility effects. Subsequent univariate ANOVAs showed that the display effects ($F(3, 125) = 3.80, p = 0.012$) were only for the taxi speed measure and the interaction effects ($F(3, 125) = 4.20, p = 0.007$) were only for the time-to-taxi measure. Post-hoc SNK tests revealed that on average 1) the EPs taxied significantly (statistically, but not operationally) quicker (about 1 second) with the Advanced HUD compared to Paper, but with no appreciable differences between the Advanced EMM or Advanced HWD and, 2) the EPs had no significant differences when taxiing with the Advanced HWD, Advanced EMM, or Paper.

5.4.6 Required Navigation Performance

As with Experiment I, the proposed surface RNP [51] requirements were used. For Experiment II, the visibility conditions were such that RNP for an ICAO Code D [42] aircraft should remain within ±7.2 feet (±2.2 meters) of the route centerline 95% of the time. Figure 43 shows the RNP values for all run conditions. None of the display concepts were within ± 7.2 feet of route centerline 95% of the time.

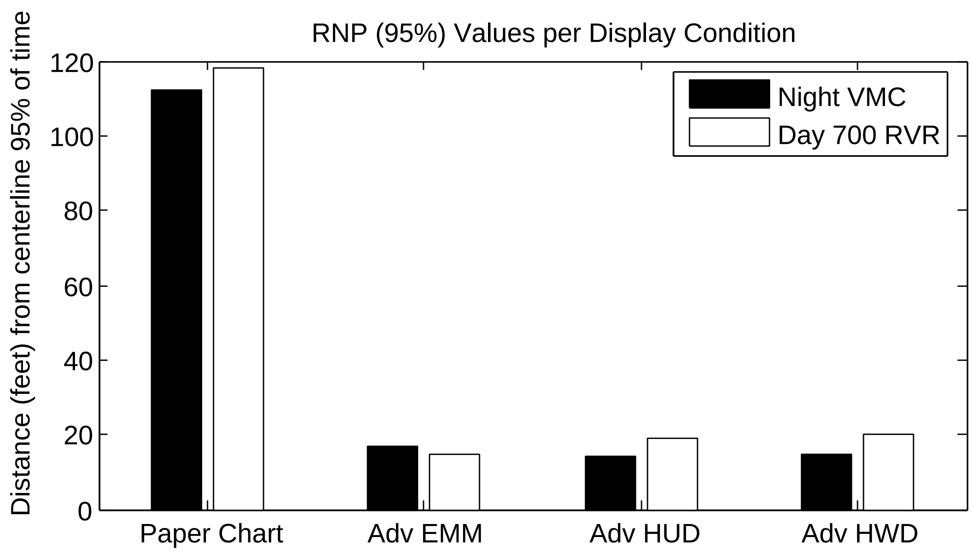


Figure 43. RNP for Experiment II.

Figure 44 shows the results of the RNP with only nominal trials. Using only nominal data trials, no display condition met the surface RNP requirements and an ANOVA showed that time of day was not significant, $F(1, 10) = 7.563, p = 0.094$. Display condition was significant, $F(3, 10) = 3.719, p = 0.05$. Post-hoc tests on display condition show two overlapping subsets: 1) Advanced EMM (mean=12.5 ft), Advanced HUD (mean=13.5 ft) and Advanced HWD (mean=13 ft), and 2) Advanced HWD (mean=13 ft), Advanced HUD (mean=13.5 ft), Paper Chart (mean=15.8 ft). The Paper Chart condition had significantly worse (statistically, but not operationally) lateral RNP during surface operations than the Advanced EMM. However, there were no statistical differences between the Advanced HUD, the Advanced HWD and the Paper Chart condition.

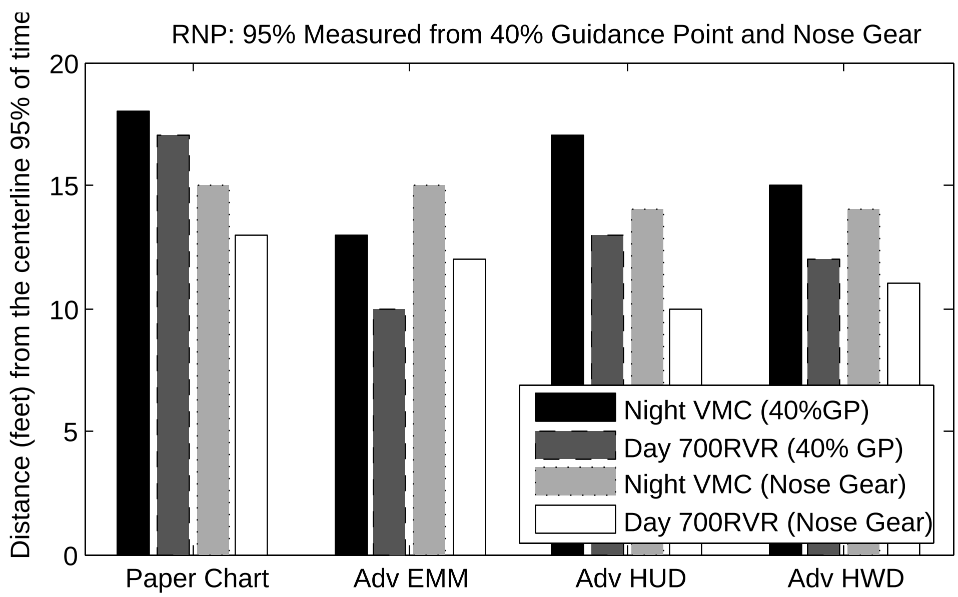


Figure 44. RNP with only nominal trials for Experiment II.

The point of computation of the path error was highly significant, $F(1, 10) = 14.944, p = 0.003$. Using the nose gear to calculate the path error resulted in significantly lower lateral RNP (mean=12.25) than when using the 40% Guidance Point (mean=15.125). Though not operationally significant, this suggests that in turns, pilots tended to keep the nose centered on the taxiway centerline as they do in straight taxi segments which is consistent with their training.

An ANOVA was performed on the surface RNP measurements made from the nose gear only with time of day (Night VMC and Day 700 RVR) and display condition as the factors. Neither factor was significant, $p > 0.05$.

5.5 Post-Run Questionnaire Results

The same questionnaires from Experiment I were administered for Experiment II with the exception of the SAGAT probes (due to time constraints). The post-run questionnaires used in this experiment are shown in Appendix C, Figures C6 through C9.

5.5.1 NASA Task Load Index

An ANOVA was performed for the dependent variable mental workload (task load index) from flight crew ratings on the NASA-TLX [54] scale. No significant differences were found among the display conditions, $p > 0.05$.

5.5.2 Situation Awareness Rating Technique

An ANOVA was performed for the dependent variable situation awareness derived from flight crew ratings (0-100) on the SART. SA is defined as Understanding - (Demand - Supply). Analysis found a significant effect for display condition, $F(3, 15) = 3.77, p < 0.05$. A SNK test revealed two unique subsets: (1) the Advanced HWD (135.25), the Advanced HUD (142.16), and the Advanced EMM (142.38) which have the highest SA and no significant differences between them and (2) the Paper Chart condition (82.0) which had the lowest SA (Fig. 45).

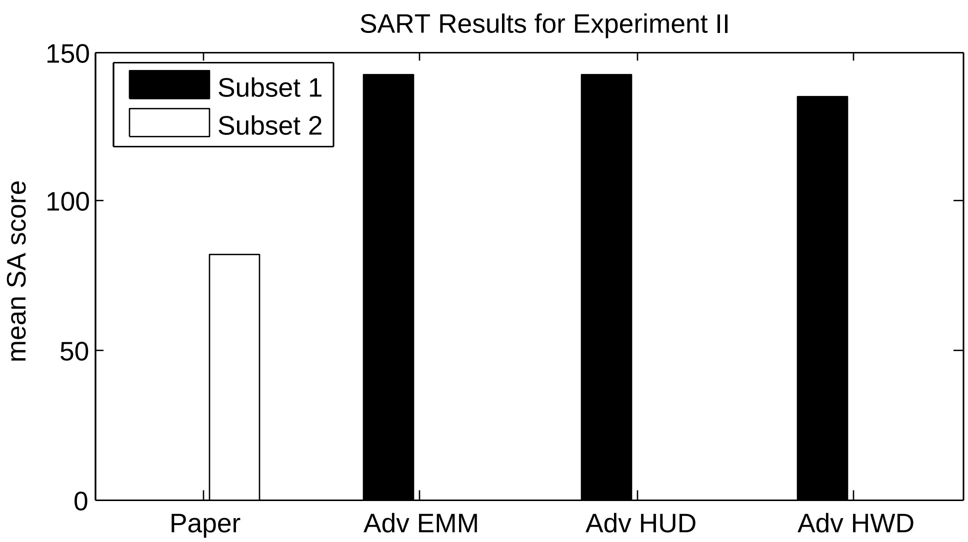


Figure 45. SART results for Experiment II. Higher scores denote higher SA.

5.5.3 Taxi Situation Awareness Questions

Flight crews were administered a Likert experimental questionnaire (scale 1 to 5 scale; 1 = “not at all”; 5 = “very much”) after each run that asked the EPs to rate the display condition’s contribution to taxi efficiency, overall navigation awareness, route awareness of local controller clearance, route awareness of ground controller clearance, surface traffic awareness, directional awareness, and taxi safety. Ratings were similar to Experiment I in that advanced concepts (path, clearance and traffic displayed) were rated significantly higher than the Baseline and Intermediate HWD display concepts. An ANOVA revealed significant effects for all dependent variables, $p < 0.05$. Post-hoc SNK tests were performed on these dependent variables resulting in two unique subsets: (1) the Advanced EMM, the Advanced HUD, and the Advanced HWD (no significant differences between them), and (2) the Paper Chart condition. Only the EP ratings of display contribution to taxi efficiency were found not to be significant. Means for each dependent variable are presented in Figure 46.

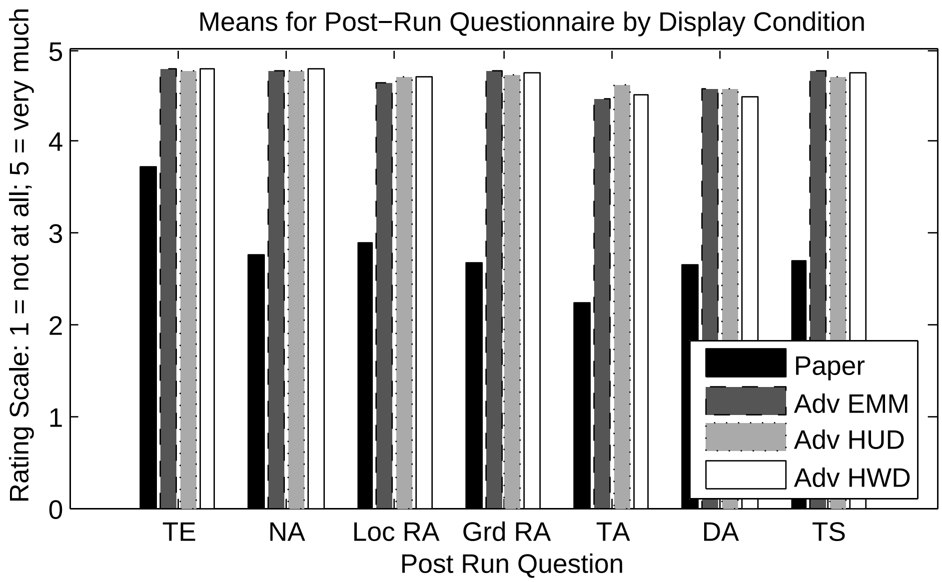


Figure 46. Means for post-run questionnaire by display condition. TE = Taxi Efficiency; NA = Overall Navigational Awareness; Loc RA = Route Awareness of the Local Controller Clearance; Grd RA = Route Awareness of the Ground Controller Clearance; TA = Surface Traffic Awareness; DA = Directional Awareness; TS = Taxi Safety.

5.5.4 Simulator and HWD-Induced Sickness

For Experiment II, nine of 24 EPs (38%) reported significant (SSQ score 15 and above) simulator sickness. Five of the nine cases were reported by the captains and four by first officers. As with Experiment I, the first officer never wore the HWD. This is a higher percentage of simulator sickness cases than with Experiment I, however, of the 9 simulator sickness cases, 4 of those EPs baseline SSQ score was at or above 15. In other words, before data collection started, the EPs were reporting significant simulator sickness symptoms. Discounting the four EPs who began the trials with significant simulator sickness symptoms, the percentage of EPs with simulator sickness was 25%, which is the rate reported from Experiment I.

An analysis of simulator sickness was performed. Before and after each block of data trials with the HWD, the same SSQ from Experiment I was given. In addition, the EPs were given a baseline questionnaire before data collection and an end of the day questionnaire. To isolate the effects of the HWD, the difference between the start and end of a HWD block was used. Often, there was no difference from before to post HWD block despite the high SSQ ratings, owing to the fact that the EPs sometimes began the block with a high rating (not due to wearing the HWD). This method then allows for the analysis of the HWD independent of the other runs since it was within block. The baseline questionnaire is confounded with the other runs but still gives an indication of overall simulator sickness experienced, collapsed across all data runs.

There were no significant differences found. From Table 9, in some cases the EPs reported better (negative) ratings after block (perhaps psychologically because they had completed the block, particularly for Block 3). The total SSQ score was very slight and actually improved for the captains and was slight for the first officers. None of the means were much higher than 2+ (positive) suggesting very little simulator sickness for HWD runs (2+ would be considered very good [57]). Based on the data, there was no simulator sickness effect for the HWD and suggests something different to explain high ratings for some EPs at the beginning of some HWD blocks (before HWD runs) that changed very little over the course of that block or subsequent blocks.

Table 9. Simulator Sickness Results: Experiment II

				95% Confidence Interval	
Block	Pilot	Mean	Std Error	Lower Bound	Upper Bound
1	Captain	2.182	1.625	-1.395	5.758
	First Officer	1.870	1.342	-1.084	4.824
2	Captain	1.870	3.036	-4.813	8.553
	First Officer	1.247	2.362	-3.953	6.446
3	Captain	-4.675	2.441	-10.049	.699
	First Officer	.312	4.378	-9.325	9.949
Total	Captain	-.623	6.326	-14.547	13.300
	First Officer	.609	4.094	-8.401	9.619

5.6 Post-Test Questionnaire Results

A MANOVA statistical procedure was performed on four paired comparison scales administered to the captain and first officer of each flight crew. The paired comparison scales asked the EP to factorially evaluate each of the four display conditions in comparison to one another on four constructs: Situation Awareness (SA-SWORD), Mental Workload (SWORD), Taxi Efficiency, and Taxi Safety. The analyses were conducted separately for captain and first officer responses. Significant results reported are at the $p < 0.01$ significance level.

5.6.1 Situation Awareness

For the captain ratings, there was a significant main effect found for situation awareness, $F(3, 10) = 26.621$. A post-hoc planned comparison revealed an effect "approaching significance" ($p = 0.06$) for the comparison between Advanced HUD and Advanced HWD. EPs rated the Advanced HWD higher in terms of SA than the Advanced HUD although not statistically significant at the 0.05 alpha level. Because the Advanced EMM and Advanced HUD were found not to be significantly different from each other, but in contrast the Advanced HWD was found to be rated significantly higher. Therefore, the HWD versus the HUD results for SA-SWORD are likely an artifact of the power of the experimental design and do provide trend evidence to posit an effect between display concepts.

Table 10. Significance Table for Situation Awareness (SA-SWORD) Paired Comparison Ratings

Display Comparisons	CAPTAIN		FIRST OFFICER	
	Significant	Non-Significant	Significant	Non-Significant
Paper - Adv EMM	X		X	
Paper - Adv HUD	X		X	
Paper - Adv HWD	X		X	
Adv EMM - Adv HUD		X		X
Adv EMM - Adv HWD	X			X
Adv HUD - Adv HWD		X		X

For the first officer ratings, there was a significant main effect for situation awareness, $F(3, 30) = 17.9$. A post-hoc test (Table 10) revealed that the Paper Chart condition was rated significantly lower for situation awareness than the other three display concepts. No other effects were found to be significant.

5.6.2 Mental Workload

For the captain ratings, there was a significant main effect found for mental workload, $F(3, 30) = 366.69$. A post-hoc test (Table 11) revealed that the Paper Chart condition was rated significantly higher for mental workload than the other three display concepts. There were no differences between the other three display conditions for mental workload.

Table 11. Significance Table for Mental Workload (SWORD) Paired Comparison Ratings

Display Comparisons	CAPTAIN		FIRST OFFICER	
	Significant	Non-Significant	Significant	Non-Significant
Paper - Adv EMM	X		X	
Paper - Adv HUD	X		X	
Paper - Adv HWD	X		X	
Adv EMM - Adv HUD		X		X
Adv EMM - Adv HWD		X		X
Adv HUD - Adv HWD		X		X

For the first officer ratings, there was a significant main effect found for mental workload, $F(3, 30) = 91.33$. A post-hoc test (Table 11) revealed that the Paper Chart condition was rated significantly higher for mental workload than the other three display conditions and there were no differences between these three conditions.

5.6.3 Taxi Efficiency

For the captain ratings, there was a significant main effect found for taxi efficiency, $F(3, 30) = 25.76$. A post-hoc test (Table 12) revealed that (a) the Paper Chart condition was rated significantly lower for taxi efficiency than the other three display conditions; (b) the Advanced EMM was rated significantly lower for taxi efficiency than both the Advanced HUD and the Advanced HWD; and (c) there were no significant differences between the Advanced HUD and the Advanced HWD conditions.

Table 12. Significance Table for Taxi Efficiency Paired Comparison Ratings

Display Comparisons	CAPTAIN		FIRST OFFICER	
	Significant	Non-Significant	Significant	Non-Significant
Paper - Adv EMM	X		X	
Paper - Adv HUD	X		X	
Paper - Adv HWD	X		X	
Adv EMM - Adv HUD	X			X
Adv EMM - Adv HWD	X			X
Adv HUD - Adv HWD		X		X

For the first officer ratings, there was a significant main effect found for taxi efficiency, $F(3, 30) = 32.96$. A post-hoc test (Table 12) revealed that the Paper Chart condition was rated significantly lower for taxi efficiency than the other three display concepts. No other effects were found to be significant.

5.6.4 Surface Operations and Taxi Safety

For the captain ratings, there was a significant main effect found for taxi safety, $F(3, 30) = 4.9$. However, subsequent post-hoc (Table 13) pair-wise comparisons (Bonferroni) failed to find any mean difference significant at the $\alpha = 0.05$ level.

Table 13. Significance Table for Taxi Safety Paired Comparison Ratings

Display Comparisons	CAPTAIN		FIRST OFFICER	
	Significant	Non-Significant	Significant	Non-Significant
Paper - Adv EMM		X	X	
Paper - Adv HUD		X	X	
Paper - Adv HWD		X	X	
Adv EMM - Adv HUD		X		X
Adv EMM - Adv HWD		X		X
Adv HUD - Adv HWD		X		X

For the first officer ratings, there was a significant main effect found for taxi safety, $F(3, 30) = 14.74$. A post-hoc test (Table 13) revealed that the Paper Chart condition was rated significantly lower for taxi safety than the other three display conditions, of which there were no significant differences between them for taxi safety.

5.6.5 HWD Usability

To get a general appreciation of the HWD usability for surface operations, a relatively simple, but broad-based usability questionnaire was used [62] (Appendix C, Fig. C11). After the completion of the experiment, the captains completed a questionnaire addressing a variety of technology usability issues. By using this usability questionnaire, a large range of issues were addressed from complexity to usefulness. A comparative evaluation with the head-down or head-up displays was not conducted. The 10 statements of the HWD usability were:

1. I think that I would like to use this system frequently
2. I found the system unnecessarily complex
3. I thought the system was easy to use
4. I think that I would need the support of a technical person to be able to use this system
5. I found the various functions in this system were well integrated
6. I thought there was too much inconsistency in this system
7. I would imagine that most people would learn to use this system very quickly
8. I found the system very cumbersome to use
9. I felt very confident using the system
10. I needed to learn a lot of things before I could get going with this system

The captains rated their agreement with each of the usability statements on a scale of 1 (strongly disagree) to 5 (strongly agree). In order to compare the 10 usability questions, each statement was scored with an equal weighting. The five positive statements (the odd numbered statements, Fig. 47) were weighted by a factor of two. Essentially, this made each statement score between two and 10. However, for the five negative statements (the even numbered statements, Fig. 48), they were reversed scored (i.e., a rating of 1 was a score of 5, etc.) and then weighted by a factor of two. The 10 weighted scores were then added together to give an overall rating between 0 and 100 (Fig. 49). Figures 47 and 48 are boxplots [63] that present the smallest and largest values, the lower and upper quartiles and the median for each usability statement. For the experiment, the average score for the HWD concept was a 75 for 12 EPs. The scores ranged from a maximum of 95 to a low of 52.5 with a 10 point standard deviation around the mean. The rationale for these grades can be determined from responses to the individual questions (Figs. 47 and 48).

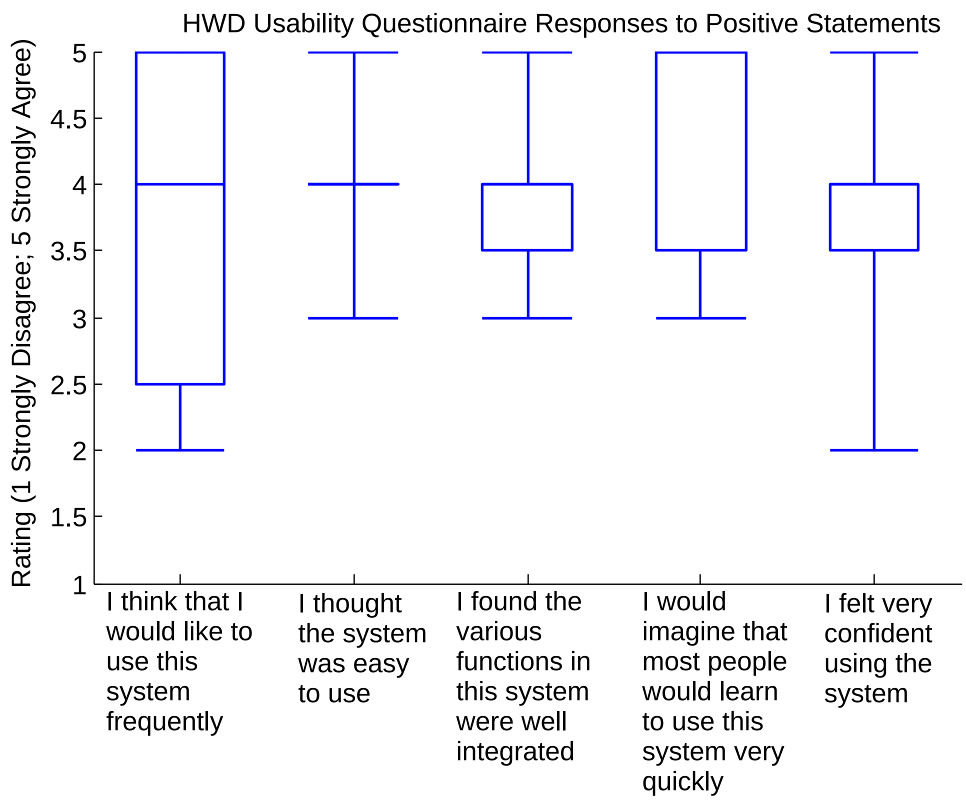


Figure 47. Captain responses to positive Usability Questionnaire.

Overall, the HWD system was given high marks by almost all the crews for being easy to use, not overly complex, and well integrated. The poor marks were due primarily to some strong negative opinions by a few users. For these questions, a "bi-polar" response was given to whether the system operation could be easily learned (7 EPs strongly agreed that it could, but 3 EPs were neutral to this question) and whether the EPs thought the system was cumbersome to use (3 strongly disagreed with this statement, but 4 were neutral to in moderate agreement.) Clearly, if the weight and encumbrance of the system were improved, more positive statements to this question might have been obtained.

Another source of disagreement and negative ratings was in response to whether the EPs would "use the system frequently." Four EPs strongly agreed with this statement but three EPs moderately disagreed. This question should have been better posed since it could be interpreted several ways. For instance,

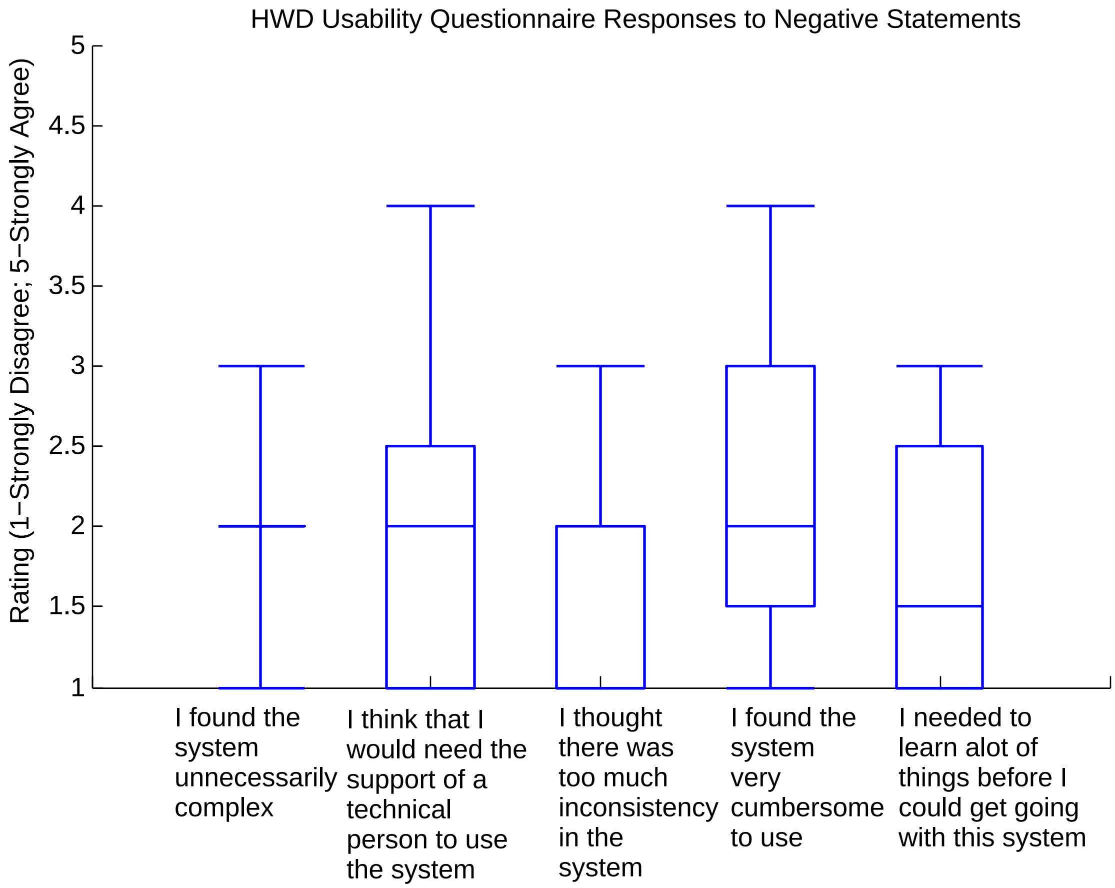


Figure 48. Captain responses to negative Usability Questionnaire.

if taken in the context of everyday operation, the EPs might have been rating how often poor weather and limited visibility necessitate the need for HWD taxi assistance. Or, they might have interpreted the question as asking whether they felt this system improved their ability to safely and efficiently conduct surface operations in general.

The good scores are encouraging but the negative scores point to needed areas for improvement. The captains felt the HWD showed high potential but that refinement is clearly needed.

5.7 Discussion of Experiment II Results

Experiment II was designed to be a follow-on to Experiment I to improve the color of the displayed traffic and to evaluate additional display concepts to the baseline condition of an airport paper map in order to fully complete the matrix of possible display comparisons. From Experiment I, even though traffic was displayed, some crews missed the displayed traffic and ended up in a "nose-to-nose" situation for the rare event scenario. For Experiment II,

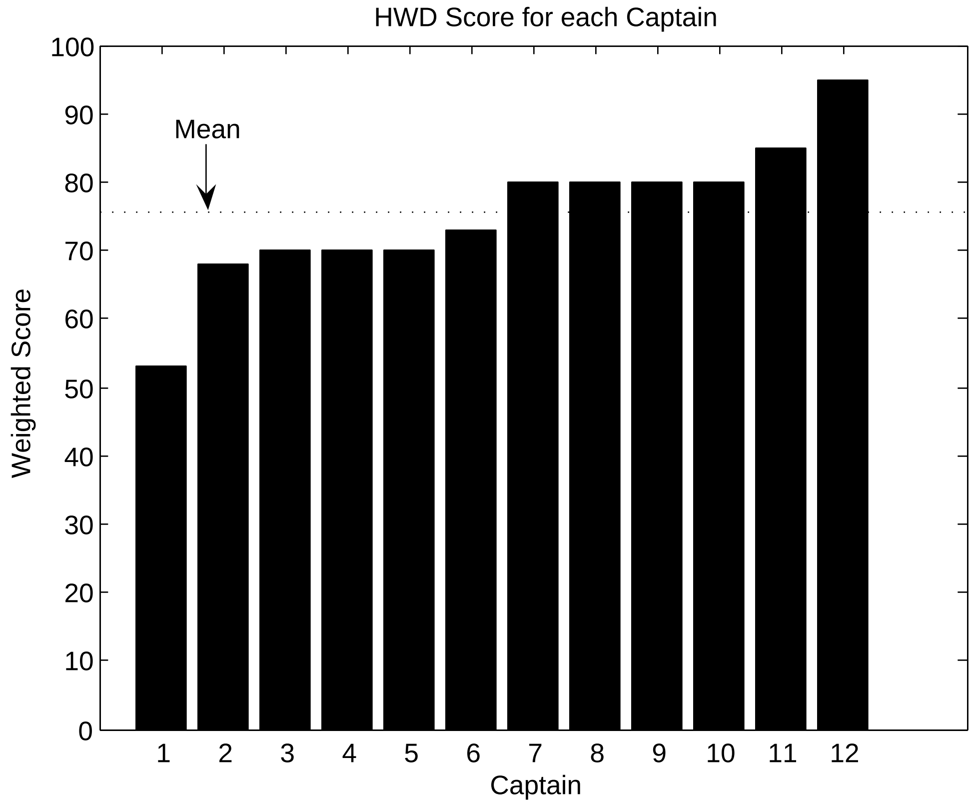


Figure 49. Captain scores for the HWD Usability Questionnaire.

none of the crews who had traffic displayed got into a nose-to-nose situation and, therefore, it appears likely that the color and size of the traffic icon was the main factor for the differing results from Experiment I.

The performance data showed that all of the advanced concepts provided better route accuracy and faster taxi speeds compared to paper charts alone. On average, the EPs were able to complete the taxi route 15% faster with the advanced concepts compared to paper charts. Previous T-NASA research reported taxi speed increases in the range of 16% to 26% [47].

One surprising result was that there were 2 taxi conflict events in Experiment II during the non-rare event data trials. These events occurred with the Advanced EMM and Advanced HUD conditions, respectively. Each of these events occurred despite the traffic being clearly represented on the head-down moving map display (Advanced EMM). These "near misses" were avoided through traffic detection made out-the-window. No such taxi conflict events occurred with the Paper Chart or the Advanced HWD condition.

6 General Discussion of Results

6.1 Quantitative Results

From both experiments, the results with the Advanced HUD shown in this paper are similar to the results from previous surface operations research conducted by NASA Ames and NASA Langley. The performance data showed no significant differences between the Advanced EMM, Advanced HUD, and Advanced HWD display concepts for the dependent variables measured, but the EPs taxied at significantly faster speeds and more accurately with these displays than when taxiing with just paper charts alone. In general, the Advanced display concepts provided information (e.g., cleared route, ownship position, taxi guidance cues) enabling fast and efficient taxi. However, no quantitative performance differences differentiating head-up versus head-down display concepts (when displaying essentially the same or similar information) were found in this study. Additionally, the crews made significantly more navigation errors with the paper charts than with any of the other three advanced display concepts.

From Experiments I and II, none of the display concepts were within surface RNP requirements as proposed by Cassell et al [51]. Even when trials with navigational anomalies were removed, RNP was still not obtained and there were no significant differences among the various concepts. The taxi routes used in this experiment were very challenging by design, especially with the given visibility conditions. Further, for concepts where a path was not displayed, a significant amount of navigational errors were made. Making an incorrect turn will cause the deviation from the intended route centerline to grow large quickly. The result was that every data run was a challenge in that some type of mistake was made by almost every crew. There were no "nominal" type runs to offset the experimental trials. Further, the crews were instructed to get to the gate quickly as this is the nominal taxi instruction at Chicago O'Hare. Therefore, for these experiments, the RNP measurements were used to compare the EP's performance between the display concepts and were not used to evaluate surface RNP for the Chicago O'Hare airport. When these navigational errors were removed from the RNP analyses, there were no significant differences among the concepts, which was expected.

Comparing the Advanced HWD and Advanced HUD concepts across Experiments I and II, a MANOVA on RMS path error, delta taxi speed and delta time-to-taxi with display condition and visibility as the main factors was performed. There were no display effects, visibility effects, or interaction effects of these two main factors for the three measures listed above. Therefore, in terms of taxi performance, the Advanced HWD and the Advanced HUD were statistically the same.

There were significantly fewer taxi conflicts with Experiment II as compared with Experiment I. There were several contributing factors leading to this result. First, Experiment II took half the time per crew than Experiment I, because the 12 longer, more complex nominal runs were replaced with 12 short, simple routes. Also, the displayed traffic visibility was improved. Given that the two conflict events occurred with the improved traffic readability displays and not with paper charts, the dominant factors between the two experiments are the shortened time (less fatigue) and simpler routes for half the data runs.

6.2 Qualitative Results

The results of the paired comparisons showed that the addition of a head-up or head-worn display subjectively increased taxi efficiency as compared with having an advanced EMM alone. These results agree with past research conducted at NASA Ames Research Center, which demonstrated that the combination of head-up display and head-down display taxi concepts provides superior taxi performance.

From Experiment I, the 2-D head-down Baseline EMM and the 3-D head-up Intermediate HWD concepts displayed the same information, but at different virtual camera perspectives. From the results, there were no significant differences between these two concepts except that the Intermediate HWD was rated higher in SA than the Baseline EMM. The EPs commented that the Intermediate HWD presented the taxi signage in a format that they were already familiar, thus, it was easy to interpret.

The NASA-TLX results showed no significant differences for mental workload, suggesting that the Advanced HWD display does not increase, nor reduce, mental workload demands compared to current navigation methods (i.e., using paper charts). Moreover, there were no differences found between the advanced display concepts further lending evidence that the introduction of an Advanced HWD concept would not increase cognitive and attentional demands for the flight crew.

The mental workload results are mirrored by the situation awareness post-run questionnaire and taxi situation awareness paired comparison results that also failed to show significant differences in perceived situation awareness between the advanced display concepts. These three display concepts were all rated significantly higher for situation awareness than taxiing with paper charts alone. When the captains were asked to rate their overall impressions of situation awareness, however, the SA-SWORD results did reveal that both the Advanced HUD and Advanced HWD provided significantly higher SA than the Advanced EMM concept. Although no statistically significant results were found for all ratings between the Advanced HUD and the Advanced HWD, statistical evidence for a trend toward significance for situation awareness ratings was observed (i.e., $p = 0.06$). Because of the limited number of flight crews tested and the reported limitations of the prototype HWD, these results highly support the conclusion of a likely effect in favor of the HWD particularly if the shortcomings of the current HWD concept are addressed (see Sec. 6.6 below).

6.3 Taxi Conflicts

During data collection for Experiments I and II, there were a total of 19 taxi conflict events. Several factors contributed to these conflicts. First, the visual acuity of the out-the-window scene was degraded because of technology limitations of the scene image generator. The measured visual acuity was 20/80. As a result, the scene would appear blurry in the far field making it difficult to read taxiway signs. Also, since ATC was a simple automated program, simulated controllers did not provide feedback for ground traffic, which is normally done. In addition, the party line chatter was not correlated to the actions of ground traffic. These conditions, combined with the complexity of the Chicago airport under the experiment visibility conditions, created very challenging taxiing scenarios.

Collapsing the data across Experiment I and II, an ANOVA was performed on the number of taxi conflicts committed by the crew for the display concept (Advanced HUD, Advanced HWD), and visibility condition (Night VMC, Day 700 RVR) as the main factors. There were no significant differences ($p > .05$) among the main factors or their interactions for this measure. On average, there were more taxi conflicts with the Advanced HUD than with the Advanced HWD, but these differences were not statistically significant.

6.4 Rare Event Results

The rare event provided another measure of situation awareness for unexpected events. The experiments were designed to create mild fatigue by the final run to create line operation conditions that increase the likelihood of runway incursion situations. For the ATC error in which the controller gave a verbal instruction to turn the wrong way, it was clear that crews with a displayed route could detect the error before making the wrong turn. All crews that did not have a displayed route (i.e., Baseline EMM and Intermediate HWD) for this rare event made the wrong turn and did not realize the mistake until cross referencing with paper charts. Further, the flight crews commented that the displayed route and EMM display provided significant situation awareness over paper charts.

The rare event involving the non-transponding aircraft was designed to examine cognitive capture effects. By providing routing, clearance information, and traffic, it was thought that such information might keep the crews head-down rather than eyes out. This rare event occurred late in the trials so that the crew was familiar with the display and the information presented on the display. The results showed that all crews were still mainly “eyes out” for traffic surveillance, which is consistent with their current training. Most crews were able to detect that the non-transponding traffic was not represented on the display(s).

The nose-to-nose rare event was designed to highlight traffic awareness by the crew. From Experiment I, two crews with the Advanced HWD display concept did not see the traffic in conflict even though it was displayed both head up and head down. Both crews commented that the brown color of the displayed traffic was difficult to distinguish without concentrating on the displays. They commented that it was desired that traffic be detectable with a quick glance; however the traffic icons should not adversely clutter the displays. For Experiment II, the traffic color and size was changed to improve readability. For Experiment II, all crews using display concepts that contained traffic information (Advanced EMM, Advanced HUD, and Advanced HWD) were able to avoid the nose-to-nose situation. As with the ATC rare event, the crews had information available within the cockpit that contradicted the controller's clearance. In both situations, the crews contacted the ground controller to resolve the discrepancy to avoid mistakes. Further, the crews commented that the information presented on the HUD provided no additional benefit in detecting this rare event in contrast to the information available on the HWD (e.g., traffic). In other words, the Advanced HWD presented another source for displaying traffic information that supplemented information being presented on the Advanced EMM head-down display.

6.5 Simulator and HWD-Induced Sickness

In all experiments, some form of simulator or cyber (virtual reality) sickness was observed with multiple crews. It has been shown that though the symptoms are similar, cyber and simulator sickness are different [57]. These experiments have the potential to cause severe cases of sickness as the experiments are a combination of virtual reality (the HWD) in a simulator. This confound causes difficulty in determining the cyber/simulator effects of the HWD. From all the cases of cyber or simulator sickness in both Experiments I and II, 41% of the cases were with the first officers who never wore the HWD. Even the best simulators (in terms of causing simulation sickness) can have as high as 20% of users experiencing significant simulator sickness effects [57]. Therefore, flight trials in real aircraft would be necessary to isolate the cyber sickness effects of the HWD. Future simulator studies should require the administering of the SSQ and real time monitoring of sickness to avoid severe EP discomfort.

6.6 Future Research Issues

There are many research questions that were not evaluated in this effort but will be considered in the future, including:

- SV database accuracy and integrity (evaluated with fused / integrated S/EVS evaluations).
- Real-Time Obstacle detection.
- Criticality of traffic reporting (i.e., ADS-B accuracy and integrity).
- SV design considerations for surface operations in HWD applications.
- Introduction of EV for HWD and HUD applications; optimal fusion of EV and SV for surface operations.
- Mixed-fleet equipage, particularly aircraft with ADS-B out, ADS-B out and in, and some non-equipped or failed ADS-B equipment.
- HWD color considerations.
- Does a HWD affect a pilot's head-movement compared to present-day operations.

The experiments revealed numerous directions of future research to better optimize and develop these concepts. One future direction involves the integration of enhanced vision sensor technology with the optimized HWD concept. Further, for these experiments, the routing and clearance information was relayed to the aircraft displays via a simulated controller datalink. Currently, the IIFD/Crew-Vehicle Interface team is conducting research employing voice recognition technology to quickly and accurately enter routing information during read-back. The potential also exists for conducting analysis of the speech and airport information for route awareness and route / track analysis.

A significant body of research has shown that runway incursions can be mitigated or even prevented via flight deck alerting. For this experiment, however, the crew's situation awareness in the absence of alerting was of most interest. Alerting, in conjunction with these displays, would clearly add significantly to enhancing further the safety of surface operations. Future research will evaluate the additive effects of including such alerting algorithms, derived from the NASA RIPS research, to determine whether further safety enhancements to airport surface operations are possible.

Two issues that influenced this work were the head-tracker size and its alignment/accuracy. For these experiments, the HWD was installed on a helmet to provide a stable mounting location for the head-tracker. This configuration resulted in significant pilot encumbrance and head-borne weight. Also, the HWD was aligned with the scene by displaying a grid pattern in the HWD and the same pattern in the out-the-window visuals. For actual operations, the alignment process must be quick, reliable and with a pre-determined degree of integrity and assurance. Further, the HWD image stability and alignment must be maintained during operation. (With a HUD, this boresighting procedure is done once and “hard-mounted” into the aircraft.) Current research efforts are exploring the use of optical head tracking techniques that would minimize or eliminate these HWD “costs.” Otherwise, any dollar savings, derived by weight reductions for HWD-equipage, would be out-weighed by the cost in developing a robust procedure for HWD alignment, image correlation, and pilot “encumbrance.”

For both experiments, the first officers had a repeater display of the head-up device (either the HUD or HWD). All first officers commented that the repeat of the HUD did not provide any significant situation awareness. The HUD repeater did not have the out-the-window image; therefore the first officer was unable to easily correlate the symbology to the scene. For the HWD repeater, the first officers commented that the repeater was a distraction. The HWD repeater displayed all the captain’s head movement, thus the image on the repeater was constantly changing. This highly dynamic image would tend to unnecessarily capture the first officer’s attention. Essentially, early in the trials, the first officers ignored the repeater and commented that the Advanced EMM with routing, clearance and traffic information provided the essential information for surface operations in these experiments.

Another issue that NASA will be addressing in the future is obscuration of the outside world view by the pilot, in this “augmented reality” created by the HWD. For instance, semi-conformal display concepts, binocular/monocular displays, and clutter countermeasures will be explored in this area.

7 Conclusions

The results suggest that the Advanced HUD and Advanced HWD are comparable to each other with regard to situation awareness, mental workload, taxi efficiency, taxi performance, and perceived taxi safety. There were a few limitations of the implementation of the HWD concept that may have reduced its full potential to demonstrate marked differences between the capabilities of the HUD and HWD concepts. The EPs commented that the Advanced EMM concept was a “quantum leap” for situation awareness. At the start of a data collection day, without experiencing any of the concepts, most crews commented that the head down Advanced EMM was the only display needed. However, as experienced was gained with the head-up devices (HUD and HWD), the EPs concluded that a head-up device enhanced their situation awareness compared to just a head-down display.

In addition to the efficiency and safety advantages of the advanced HWD, there are other considerations that argue for a HWD solution. The HWD provides potential weight savings that would have significant cost advantages to commercial airlines. Further, the typical viewing area of a HUD is 30°H by 24°V, which is sufficient for flight, but not necessarily for surface operations because of its limited field-of-regard. On the ground, one of the main tasks of the crew is to survey all around the aircraft to avoid collisions with other airplanes or objects on the airport surface. The limited field-of-regard of the HUD was especially evident in the present experiment when the flight crew attempted turns but the path was only displayed as virtual turn flags in the HUD due to required over-steering.

The research showed significant potential for the HWD for commercial surface operations. However, because the technology is new for Part 121 operations, there are significant impediments that exist which currently limit its potential application. The conclusions drawn from the experimental data demonstrate that the HWD has similar performance and situational awareness benefits compared to a HUD. However, the HUD has substantially greater operational time and pilot interaction with the display device. Because the HUD has significantly more operation time, it is more optimized compared to the HWD. Therefore, evidence for the replacement of the HUD with a HWD must clearly demonstrate a cost-to-benefit value due to the initial difficulties that would come with certification of a new display device in the commercial cockpit and associated training requirements. Despite these hurdles, the data collected to date substantiate the value of the technology for commercial aircraft operations if implementation barriers are addressed.

Consequently, as a result of being an emerging technology, the HWD has significant user encumbrance issues associated with it that are being addressed by the manufacturers of the technology. These issues include: 1) latency issues for head-tracking systems; 2) ergonomic issues such as weight, balance and comfort needs; 3) transition issues for other phases of flight; 4) imagery issues such as resolution, brightness, contrast, and use of color; 5) alignment issues and the accuracy requirements needed; and 6) HWD operating procedures. NASA is currently researching these issues through in-house experiments, SBIR contracts and NASA Research Announcement (NRA) agreements.

References

1. Joint Planning and Development Office. Next-generation air transportation system integrated plan. Technical Report Version 1, U.S. Department of Transportation, December 2004.

2. National Transportation Safety Board. Most wanted transportation safety improvements. List Brochure, November 2007.

3. FAA. Runway incursion trends and initiatives at towered airports in the United States, FY 2001 through FY 2004. FAA runway safety report, US Department of Transportation, Federal Aviation Administration, Washington DC, USA, August 2005.

4. D. C. Foyle, A. D. Andre, R. S. McCann, E. Wenzel, D. Begault, and V. Battiste. Taxiway navigation and situation awareness (T-NASA) system: Problem, design philosophy and description of an integrated display suite for low-visibility airport surface operations. *SAE Transactions: Journal of Aerospace*, 105:1411–1418, 1996.

5. B. L. Hooey and D. C. Foyle. A post-hoc analysis of navigation errors during surface operations: Identification of contributing factors and mitigating solutions. In *Proceedings of the Eleventh International Symposium on Aviation Psychology*, Columbus, OH: The Ohio Sate University, 2001.

6. R. S. McCann, B. L. Hooey, B. Parke, D. C. Foyle, A. D. Andre, and B. Kanki. An evaluation of the taxiway navigation and situation awareness (T-NASA) system in high-fidelity simulation. *SAE Transactions: Journal of Aerospace*, 107:1612–1625, 1998.

7. M. D. Byrne and A. Kirlik. Using computational cognitive modeling to diagnose possible sources of aviation error. *International Journal of Aviation Psychology*, 15(2):135–155, 2005.

8. D. R. Jones, C. C. Quach, and S. D. Young. Runway incursion prevention system - demonstration and testing at the Dallas/Fort Worth international airport. In *20th Digital Avionics Systems Conference*, October 2001.

9. D. R. Jones and J. M. Rankin. A system for preventing runway incursions. *Journal of Air Traffic Control*, 44(3), July – September 2002.

10. D. R. Jones. Runway incursion prevention system simulation evaluation. In *21th Digital Avionics Systems Conference*, October 2002.

11. D. R. Jones. Runway incursion prevention system testing at the Wallops flight facility. In Jacques G. Verly, editor, *Enhanced and Synthetic Vision Proceedings of SPIE*, volume 5802, pages 47–58, Bellingham, WA, 2005. SPIE.

12. D. R. Jones and L. J. Prinzel III. Runway incursion prevention for general aviation operations. In *25th Digital Avionics Systems Conference*, October 2006.

13. J. L. Williams, B. L. Hooey, and D. C. Foyle. 4-D taxi clearances: pilots' usage of time- and speed-based formats. In *Proceedings of the AIAA Modeling and Simulation Technologies Conference*, number AIAA-2006-6611, 2006.

14. J. J. Arthur III, L. J. Prinzel III, L. J. Kramer, R. V. Parrish, and R. E. Bailey. Flight simulator evaluation of synthetic vision display concepts to prevent controlled flight into terrain (CFIT). Technical Report NASA/TP-2004-213008, NASA Langley Research Center, Hampton, VA, April 2004.

15. L. J. Kramer, J. J. Arthur III, R. E. Bailey, and L. J. Prinzel III. Flight testing an integrated synthetic vision system. In Jacques G. Verly, editor, *Enhanced and Synthetic Vision Proceedings of SPIE*, volume 5802, Bellingham, WA, 2005. SPIE.

16. J. R. Kerr and S. P. Way. New infrared and systems technology for enhanced vision systems. In *RTO SET Workshop on Enhanced and Synthetic Vision System*, number NATO RTO-MP-107, held in Ottawa, Canada, September 2002. NATO.

17. R. V. Parrish, A. M. Busquets, S. P. Williams, and D. E. Nold. Evaluation of alternate concepts for synthetic vision flight displays with weather-penetrating sensor image inserts during simulated landing approaches. Technical Report NASA TP-2003-212643, NASA Langley Research Center, Hampton, VA, October 2003.

18. J. J. Arthur III, L. J. Kramer, R. E. Bailey, and L. J. Prinzel III. Flight test comparison between enhanced vision (FLIR) and synthetic vision systems. In Jacques G. Verly, editor, *Enhanced and Synthetic Vision 2005*, volume 5802, Bellingham, WA, April 2005. SPIE.

19. S. D. Harrah, W. R. Jones, C. W. Erickson, and J. H. White. The NASA approach to realize a sensor enhanced-synthetic vision system (SE-SVS). In *21st IEEE / AIAA Digital Avionics Conference*, volume 2, pages 11A4–1 – 11A4–11. IEEE/AIAA, October 2002.

20. M. Uijt de Haag, J. Sayre, J. Campbell, S. D. Young, and R. A. Gray. Flight test results of a synthetic-vision elevation database integrity monitor. In J. G. Verly, editor, *Enhanced and Synthetic Vision 2001*, volume 4363 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 124–133. SPIE, August 2001.

21. G. Craig, S. Jennings, K. Link, and R. Kruk. Flight test of a helmet-mounted, enhanced and synthetic vision system for rotorcraft operations. In *RTO SET Workshop on Enhanced and Synthetic Vision System*, number NATO RTO-MP-107, held in Ottawa, Canada, September 2002. NATO.

22. R. E. Bailey, L. J. Kramer, and L. J. Prinzel III. Fusion of synthetic and enhanced vision for all-weather commercial aviation operations. In *NATO Human Factors and Medicine Symposium on Human Factors and Medical Aspects of Day/Night All Weather Operations: Current Issues and Future Challenges*, number NATO RTO-HFM-141, pages 11–1–18. NATO, 2007.

23. B. L. Hooey and D. C. Foyle. Pilot navigation errors on the airport surface: Identifying contributing factors and mitigating solutions. *The International Journal of Aviation Psychology*, 16(1):51–76, 2006.

24. S. D. Young and D. R. Jones. Flight testing of an airport surface guidance, navigation, and control system. In *Institute of Navigation Conference*, January 1998.

25. T. A. Andre. Information requirements for low-visibility taxi operations: What pilots say. In R.S. Jensen and L.A. Rakovan, editors, *Eighth International Symposium on Aviation Psychology*, pages 484–488, Columbus, Ohio, 1995. Ohio State University.

26. J. J. Arthur III, L. J. Prinzel III, S. P. Williams, and L. J. Kramer. Synthetic vision enhanced surface operations and flight procedures rehearsal tool. In Jacques G. Verly and Jeff J. Guell, editors, *Enhanced and Synthetic Vision Proceedings of SPIE*, volume 6226, page 62260I, Bellingham, WA, 2006. SPIE.

27. J. J. Arthur III, L. J. Prinzel III, K. J. Shelton, R. E. Bailey, L. J. Kramer, S. P. Williams, and R. M. Norman. Design and testing of an unlimited field-of-regard synthetic vision head-worn display for commercial aircraft surface operations. In Jacques G. Verly and Jeff J. Guell, editors, *Enhanced and Synthetic Vision 2007*, volume 6559, Bellingham, WA, April 2007. SPIE.

28. J. J. Arthur III, S. P. Williams, L. P. Prinzel III, L. J. Kramer, and R. E. Bailey. Flight simulator evaluation of display media devices for synthetic vision concepts. In Clarence E. Rash and Colin E. Reese, editors, *Helmet- and Head-Mounted Displays IX: Technologies and Applications*, volume 5442, pages 213–224, Bellingham, WA, 2004. SPIE.

29. H. A. Rediess. An augmented reality pilot display for airport operations under low and zero visibility conditions. In *Collection of Technical Papers. Pt. 2*, number AIAA-97-3680, pages 912–929, New Orleans, LA, August 1997. AIAA Guidance, Navigation, and Control Conference.

30. J. W. Ruffner, J. E. Fulbrook, and M. Foglia. Near-to-eye display concepts for air traffic controllers. In Clarence E. Rash and Colin E. Reese, editors, *Helmet- and Head-Mounted Displays IX: Technologies and Applications*, volume 5442, pages 120–131, Bellingham, WA, April 2004. SPIE.

31. R. E. Bailey, J. J. Arthur III, and S. P. Williams. Latency requirements for head-worn display S/EVS applications. In Jacques G. Verly, editor, *Enhanced and Synthetic Vision 2004*, volume 5424, pages 98–109, Bellingham, WA, April 2004. SPIE.

32. S. D. Young and L. Quon. Integrated Intelligent Flight Deck: Technical Plan, May 2006.

33. J. J. Arthur III, S. P. Williams, L. P. Prinzel III, L. J. Kramer, and R. E. Bailey. US patent pending: Multi-modal cockpit interface for improved airport surface operation, April 2007. NASA LAR-17290.

34. M. Velger. *Helmet-Mounted Displays and Sights*. Artech House, Inc., 1998.

35. C. E. Rash. Helmet-mounted displays: Design issues for rotary-wing aircraft. Technical report, US Army Aeromedical Research Laboratory, Fort Rucker, AL, 1999.

36. V. Klymenko, T. H. Harding, H. H. Beasley, J. S. Martin, and C. E. Rash. The effect of helmet mounted display field-of-view configurations on target acquisition. Technical Report USAARL Report No. 99-19, US Army Aeromedical Research Laboratory, Fort Rucker, AL, September 1999.

37. R. V. Parrish and S. P. Williams. Trade-offs arising from mixture of color cueing and monocular, binoptic, and stereoscopic cueing information for simulated rotorcraft flight. Technical Report NASA Technical Paper 3268, NASA Langley Research Center, Hampton, VA, January 1993.

38. E. Peli. Visual issues in the use of head-mounted monocular display. *Optical Engineering*, 29, Issue 8:883–892, August 1990.

39. M. J. Wells and R. K. Osgood. The effects of head and sensor movement on flight profiles during simulated dive bombing. In *Human Factors and Ergonomics Society, 35th annual meeting*, pages 22–26, 1991.

40. M. J. Wells and M. Venturino. Performance and head movements using a helmet-mounted display with different sized field-of-views. *Optical Engineering*, 29, Issue 8:870–877, August 1990.

41. M. J. Wells, M. Venturino, and M. A. Gresty. Coordination of head and eye movements to fixate continuous and intermittent targets. *Vision Research*, 14:395–403, 1974.

42. Boeing. Boeing commercial aircraft - design groups/codes (FAA/ICAO). Design code table, Airport Technology, Boeing Commercial Airplanes, Seattle, WA, USA, July 2007.

43. M. A. Vidulich and E. R. Hughes. Testing a subjective metric of situation awareness. In *Human Factors Society 35th Annual Meeting*, pages 1307–1311, Santa Monica, CA, 1991. Human Factors Society.

44. L. J. Kramer, L. J. Prinzel III, R. E. Bailey, J. J. Arthur III, and R. V. Parrish. Flight test evaluation of synthetic vision concepts at a terrain challenged airport. Technical Report NASA/TP-2004-212997, NASA Langley Research Center, Hampton, VA, February 2004.

45. D. C. Foyle and B. L. Hooey. Improving evaluation and system design through the use of off-nominal testing: A methodology for scenario development. In *12th International Symposium on Aviation Psychology*, pages 397–402, Dayton, OH, 2003.

46. M. L. Atkins. Head-up display symbology for surface operations: Comparisons among scene-linked symbology sets for optimum turn navigation. In *10th International Symposium on Aviation Psychology*, Columbus, OH, 1999. Ohio State University.

47. B. L. Hooey, D. C. Foyle, and A. D. Andre. The design of aircraft cockpit displays for low-visibility taxi operations. In A.G. Gale, editor, *Vision in Vehicles IX*, Holland, 2001. Elsevier Science Publishers.

48. NASA Ames Research Center. Enhanced descriptions of off-route navigation errors. Working Paper, 2001.

49. M. Yeh. Human factors considerations in the design and evaluation of moving map displays of ownship on the airport surface. Technical Report DOT/FAA/AR-04/39, US Department of Transportation, Federal Aviation Administration, Washington DC, USA, September 2004.

50. D. A. Nakamura. Required navigation performance. Technical Report D780-10251-1, The Boeing Company, Seattle, WA, November 2003.

51. R. Cassell, A. Smith, and D. Hicok. Development of airport surface required navigation performance (RNP). Technical Report NASA/CR-1999-209109, NASA Langley Research Center, Hampton, VA, June 1999.

52. R. E. Bailey, J. J. Arthur III, S. P. Williams, and L. J. Kramer. Latency in visionic systems: Test methods and requirements. In *RTO HFM Workshop on "Toward Recommended Methods for Testing and Evaluation of EV and E/SV Based Visionic Devices"*, volume RTO-MP-HFM-125, held in Williamsburg, Virginia, April 2004. NATO.

53. R. E. McFarland. CGI delay compensation. Technical Report NASA/TM-86703, NASA Ames Research Center, Moffett Field, CA, 1986.

54. S. G. Hart and L. E. Staveland. Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati, editors, *Human mental workload*, pages 139–183, Amsterdam, The Netherlands, 1988. Elsevier.

55. R. M. Taylor. Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In *AGARD Conference Proceedings No 478, Situational Awareness in Aerospace Operations*, number AGARD-CP-478, pages 3–1 – 3–37, Aerospace Medical Panel Symposium, Copenhagen, October 1990.

56. M. R. Endsley and D. J. Garland. *Situation Awareness Analysis and Measurement*. Lawrence Erlbaum, 2000.

57. K. Stanney, R. S. Kennedy, and J. M. Drexler. Cybersickness is not simulator sickness. In *Human Factors Society and Ergonomics Society 41st Annual Meeting*, pages 1138–1142, 1997.

58. W. T. Nelson, R. S. Bolia, M. M. Roe, and R. M. Morley. Assessing simulator sickness in a see-through HMD: Effects of time delay, time on task, and task complexity. In *IMAGE 2000 Conference*, Scottsdale, AR, July 2000.

59. R. T. Hennessy, T. J. Sharkey, J. A. Matsumoto, and J. W. Vorrhees. Simulator induced alteration of head movements (SIAHM). In *AIAA/AHS Flight Simulation Technologies Conference*, number AIAA-1992-4134, pages 29–36, Hilton Head Island, SC, August 1992.

60. R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology*, pages 203–220, 1993.

61. M. A. Vidulich. The use of judgment matrices in subjective workload assessment - the Subjective WORkload Dominance (SWORD) technique. In *Human Factors and Ergonomics Society, 33rd annual meeting*, pages 1406–1410, Denver, CO, October 1989.

62. J. Brooke. SUS: A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, and I. L. McClelland, editors, *Usability Evaluation in Industry*, pages 189–194, London, UK, 1996. Taylor and Francis.

63. J. W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, 1977.

64. J. E. Melzer and K. Moffitt. *Head Mounted Displays: Designing for the User*. McGraw - Hill, 1997.

65. FAA. Standards for airport markings. Advisory Circular 150/5340-1J, US Department of Transportation, Federal Aviation Administration, Washington DC, USA, April 2005.

Appendix A

The Four Display Modes of HWD Usability Study

Four display modes were developed for the Head-Worn Display (HWD) usability study. In some of the data trials, the Evaluation Pilots (EPs) were allowed to choose the display mode via the Voice Recognition System (VRS). Using the VRS, the EPs would choose the desired mode by pressing a button on the joystick to activate the VRS, then speak one of six commands:

1. *Clearance* - a text only representation of the cleared route.
2. *Bread Crumb* - showed the 2-Dimensional (2-D) precision guidance display.
3. *Map* - showed the overhead map. In this mode, pilots would say *Range Up* or *Range Down* to scale the display range up or down.
4. *Perspective* - this rendered the virtual airport as viewed from a virtual camera at the pilot eye point looking out the window.
5. *Range Up* - increased the map range scale in *Map* mode (zoom out).
6. *Range Down* - decreased the map range scale in *Map* mode (zoom in).

In all modes, the ground speed in knots was displayed to the EP at the top center of the display. Figure A1 shows an example of the *Clearance* mode. In this mode, the cleared route was displayed in chronological order starting at the top. As each stage of the clearance was completed, the color changed from white to green. The white box denoted the current instruction being executed. For these trials, the input of the clearance information was automatic and required no interaction on the EP to enter the route.

Figure A2 shows an example of the *Bread Crumb* display. The *Bread Crumb* display was at a fixed scale of 0.1 nautical miles. This mode provided precision guidance to the EP. The display included an ownship symbol with nose and main gear and taxiway edge lines and markings, all drawn to scale. This display was designed to provide the EP with a sense of the relationship of the gear to the taxiway. The cleared route was displayed as “bread crumbs” (the magenta triangles). A trend vector (not shown) was displayed to further help in turning. The trend vector, similar to Navigation Display (ND) type trend vectors, was a 5 second ahead predictor and was anchored at the guidance point. The guidance point, depicted as an oval on the ownship symbol, was a NASA developed point which provided a reference for judgmental over steering. Placing the guidance point symbology on the path symbology during a turn would keep the aircraft gear on the pavement. This point was located 20 feet behind the nose gear.

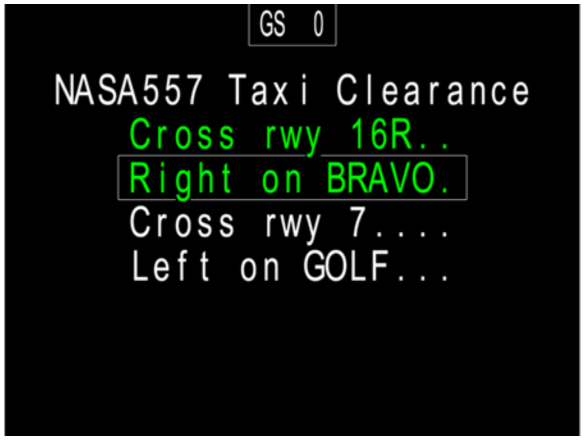


Figure A1. Clearance mode showing cleared taxi route information.

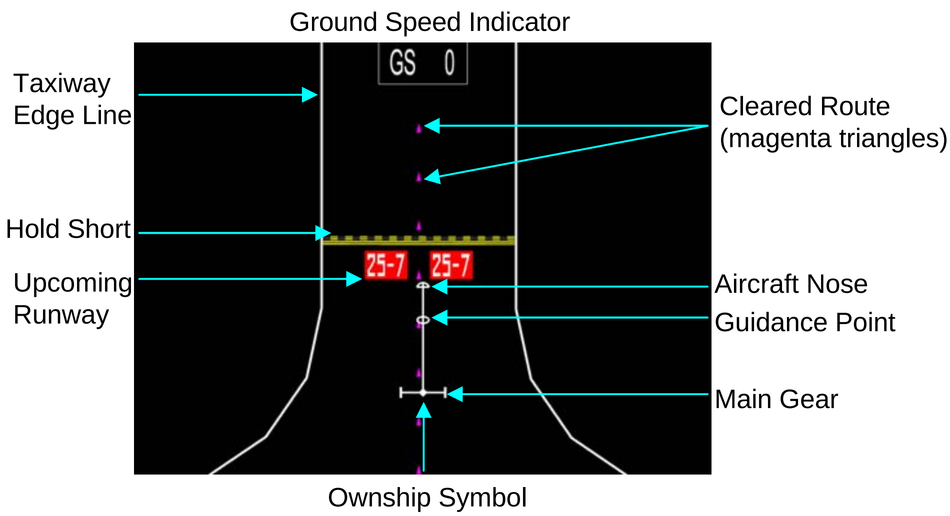


Figure A2. *Bread Crumb* mode showing precision guidance.

Also included in the *Bread Crumb* mode was the typical airport signage as if painted on the airport surface [65]. Further, virtual signs were displayed to further enhance situation awareness. The term virtual sign is used to denote symbology for markings that would not exist on the actual airport surface. For example, typical road side stop signs where used to denote hold shorts. Once cleared, the stop sign would disappear. Turn signs were also displayed to show upcoming turns along the cleared route.

The overhead *Map* mode (Fig. A3) was a overhead view of the airport. It was scalable (from 0.5 nmi to 5.0 nmi) and was a larger, more strategic view of the airport surface. The mode is like a typical ND with a detailed model of the airport. The map was oriented track up and displayed the cleared route.

Figure A4 shows an example of the *Perspective* mode. This mode was rendered as if a virtual camera was placed at the EP's eye. Using ownship

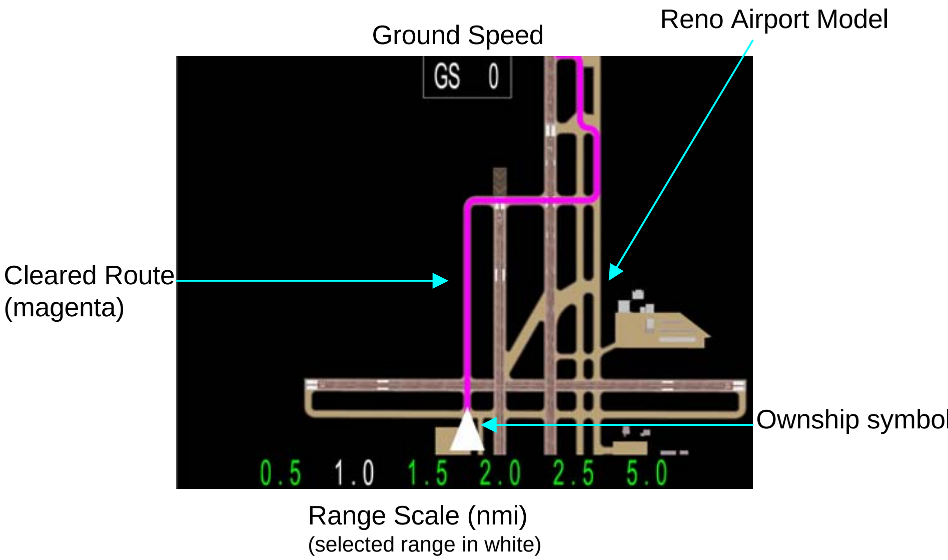


Figure A3. Map mode showing the strategic overhead display (only mode in Single-Mode).

position and real-time head tracking data, a virtual scene can be drawn as if seen by the EP's unaided vision. Elements in the *Perspective* mode would overlay objects in the real world scene. Magenta taxiway edge cones were used to denote the cleared route. As with the *Bread Crumb* mode, virtual road signs were displayed to enhance Situation Awareness (SA) for turns and holding positions.

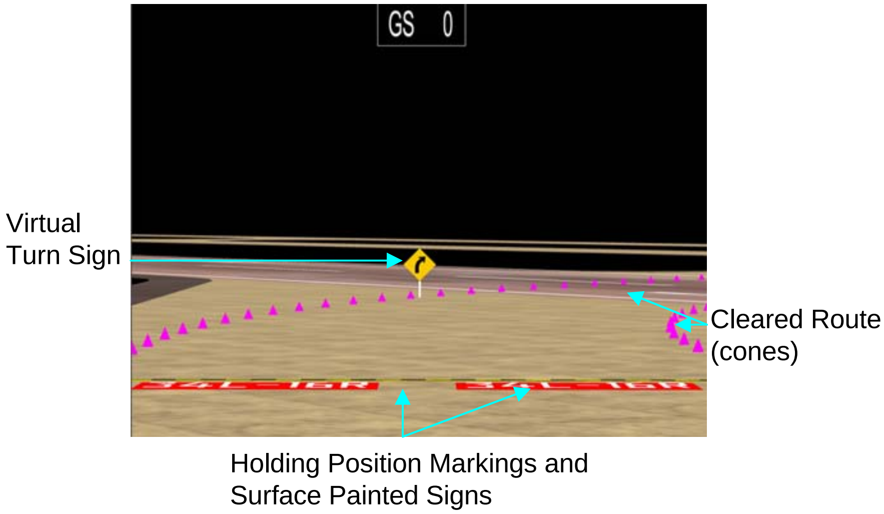


Figure A4. Perspective mode showing the virtual airport.

Appendix B

Usability Study Run Sheet

Table B1 shows the run order for the usability study. After training, run 1 was a runway to gate route using a paper chart of the Reno airport. Ground speed and heading were displayed head down. Run 2 was a runway to gate route using an Electronic Flight Bag (EFB) format on the head down display.

Runs 3 through 8 were HWD runs for the Multi-Mode, Clear weather condition. Note that for runs 3, 5 and 7, the Field-of-View (FOV) was changed during the run. At the hold short position, the EPs were notified that the FOV was changed from Conformal to a typical Head-Up Display (HUD) FOV (30°x 24°) for the remainder of the run. Changing the FOV during a run was necessary due to time constraints. After run 8, the EPs were asked to rank order their preferred display. From the rank order, runs 9 through 11 used the three EP-preferred display concept with the Clear weather condition.

Runs 12 through 17 were HWD runs for Multi-Mode, 1000 Runway Visual Range (RVR) weather condition. At the end of run 17, the EPs were asked to rank order their preferred display for the Multi-Mode 1000 RVR condition. Runs 18 through 20 were runs using the EP preferred display concept with the 1000 RVR weather condition.

Runs 21 through 26 were Single-Mode data runs. Since Single-Mode consisted of a 2-D moving map, head tracking and FOV were not applicable to these runs. At the end of run 23, the EPs were asked to rank order their preferred display concept for the Single-Mode, Clear weather condition. At the end of run 26, the EPs were asked to rank order their preferred display concept for the Single-Mode, 1000 RVR weather condition. Post-run questionnaires were then administered to the EPs after the completion of run 26.

Table B1. Usability Study Run Sheet

Run	Display	Mode	See-Through	FOV	Head Tracker	Weather	Route
1	Paper Chart					Clear	Rwy to Gate
2	EFB					Clear	Rwy to Gate
3	640x480	Multi	No	Conformal	Off	Clear	Rwy to Gate
	640x480	Multi	No	HUD	Off	Clear	Rwy to Gate
4	640x480	Multi	No	Conformal	ON	Clear	Rwy to Gate
5	800x600	Multi	Yes	Conformal	Off	Clear	Rwy to Gate
	800x600	Multi	Yes	HUD	Off	Clear	Rwy to Gate
6	800x600	Multi	Yes	Conformal	ON	Clear	Rwy to Gate
7	800x600	Multi	No	Conformal	Off	Clear	Rwy to Gate
	800x600	Multi	No	HUD	Off	Clear	Rwy to Gate
8	800x600	Multi	No	Conformal	ON	Clear	Rwy to Gate
9	1st Preference	Multi				Clear	Gate to Rwy
10	2nd Preference	Multi				Clear	Gate to Rwy
11	3rd Preference	Multi				Clear	Gate to Rwy
12	640x480	Multi	No	Conformal	Off	1000 RVR	Rwy to Gate
	640x480	Multi	No	HUD	Off	1000 RVR	Rwy to Gate
13	640x480	Multi	No	Conformal	ON	1000 RVR	Rwy to Gate
14	800x600	Multi	Yes	Conformal	Off	1000 RVR	Rwy to Gate
	800x600	Multi	Yes	HUD	Off	1000 RVR	Rwy to Gate
15	800x600	Multi	Yes	Conformal	ON	1000 RVR	Rwy to Gate
16	800x600	Multi	No	Conformal	Off	1000 RVR	Rwy to Gate
	800x600	Multi	No	HUD	Off	1000 RVR	Rwy to Gate
17	800x600	Multi	No	Conformal	ON	1000 RVR	Rwy to Gate
18	1st Preference	Multi				1000 RVR	Gate to Rwy
19	2nd Preference	Multi				1000 RVR	Gate to Rwy
20	3rd Preference	Multi				1000 RVR	Gate to Rwy
21	640x480	Single	No			Clear	Gate to Rwy
22	800x600	Single	Yes			Clear	Gate to Rwy
23	800x600	Single	No			Clear	Gate to Rwy
24	640x480	Single	No			1000 RVR	Gate to Rwy
25	800x600	Single	Yes			1000 RVR	Gate to Rwy
26	800x600	Single	No			1000 RVR	Gate to Rwy

Appendix C

Questionnaires

RANK ORDERING

<u>Block 1: Multi-Mode, Unlimited Visibility</u>

_______Micro, non-conformal, FOV1
_______Micro, non-conformal, FOV2
_______Micro, conformal
_______Le500 (open), non-conformal, FOV1
_______Le500 (open), non-conformal, FOV2
_______Le500 (open), conformal
_______Le500 (closed), non-conformal, FOV1
_______Le500 (closed), non-conformal, FOV2
_______Le500 (closed), conformal

Top Choice (unlimited visibility) : ________________________________

<u>Block 2: Multi-Mode, Restricted Visibility</u>

_______Micro, non-conformal, FOV1
_______Micro, non-conformal, FOV2
_______Micro, conformal
_______Le500 (open), non-conformal, FOV1
_______Le500 (open), non-conformal, FOV2
_______Le500 (open), conformal
_______Le500 (closed), non-conformal, FOV1
_______Le500 (closed), non-conformal, FOV2
_______Le500 (closed), conformal

Top Choice (restricted visibility) : ________________________________

Top Choice (Multi-Mode Overall): ________________________________

<u>Block 3: Single-Mode, Unlimited Visibility</u>

_______Micro, non-conformal
_______Le500 (open), non-conformal
_______Le500 (closed), non-conformal

Top Choice (unlimited visibility) : ________________________________

<u>Block 4: Single-Mode, Restricted Visibility</u>

_______Micro, non-conformal
_______Le500 (open), non-conformal
_______Le500 (closed), non-conformal

Top Choice (restricted visibility) : ________________________________

Top Choice (Single-Mode Overall): ________________________________

TOP CHOICE (Both Modes): ____________________________________

Figure C1. Usability Study Rank Order post-block questionnaire. Micro refers to the 600x480 pixel HWD and Le500 refers to the 800x600 pixel HWD.

Final Questionnaire

1. Please indicate your level of agreement with the following statement for all display concepts.

Low | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | ---- | High
1 2 3 4 5 6 7 8 9 10

<u>Where Am I?</u> The display concept provides sufficient awareness of my ownship position with respect to runways, taxiways, and stationary objects.

_________ Paper Charts only
_________ Electronic Flight Bag Taxi Map
_________ Micro, non-conformal, multi-mode
_________ Micro, conformal, multi-mode
_________ Micro, non-conformal, single-mode
_________ Le500 (open), non-conformal, multi-mode
_________ Le500 (open), conformal, multi-mode
_________ Le500 (open), non-conformal, single-mode
_________ Le500 (closed), non-conformal, multi-mode
_________ Le500 (closed), conformal, multi-mode
_________ Le500 (closed), non-conformal, single-mode

If unequal rating, please indicate reason ____
____
____

2. Please rank order the following for runway taxi / surface situation awareness:

_____ Paper Charts only
_____ Electronic Flight Bag Taxi Map
_____ Top Choice Multi-Mode (_____________________)
_____ Top Choice Single-Mode (_____________________)

Figure C2. Usability Study post-test questionnaire (page 1 of 4).

3. Please provide comparisons each of the following in terms of runway taxi / surface situation awareness:

Minimal | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | Substantially
1 2 3 4 5 6 7 8 9 10

Check Left or Right Column for choice that provides the greater runway taxi/surface situation awareness

Provide 0 - 10 rating in center for how much greater runway taxi/surface situation awareness in provided (0 = equal)

______ Paper Charts Only	_____	EFB Taxi Map ______
______ Paper Charts Only	_____	Top Choice Multi-Mode _____
______ Paper Charts Only	_____	Top Choice Single-Mode _____
______ EFB Taxi Map	_____	Top Choice Multi-Mode _____
______ EFB Taxi Map	_____	Top Choice Single-Mode _____
______ Top Choice Multi-Mode	_____	Top Choice Single-Mode _____

4. If Paper Charts and/or EFB Taxi Map were rated higher than top choice multi- or single-mode, please provide the reason for the rating:

____
____
____
____

If applicable, what can be done, if anything, to improve the efficacy of the multi-mode and/or single-mode display concepts?

____
____
____
____

Figure C3. Usability Study post-test questionnaire (page 2 of 4).

5. Please provide comparisons each of the following in terms of workload:

Minimal | —- | — | — | —- | — | — | — | — | — | — | Substantially
1 2 3 4 5 6 7 8 9 10

Check Left or Right Column for choice that you feel may provide greater or increase workload (lower workload = better)

Provide 0 - 10 rating in center for how much greater the workload that may be / was experienced (0 = equal)

______ Paper Charts Only	_____	EFB Taxi Map ______
______ Paper Charts Only	_____	Top Choice Multi-Mode _____
______ Paper Charts Only	_____	Top Choice Single-Mode _____
______ EFB Taxi Map	_____	Top Choice Multi-Mode _____
______ EFB Taxi Map	_____	Top Choice Single-Mode _____
______ Top Choice Multi-Mode	_____	Top Choice Single-Mode _____

Please indicate the reason for your ratings and any improvements that may be made to reduce workload (if applicable)

____
____
____
____
____
____

5. Please rank order the following multi-mode concepts in terms of which you consider most useful for surface operations:

_____ Taxi Instructions / Clearances
_____ Taxi Director "Bread Crumb"
_____ "2D Top-Down Map Display"
_____ Perspective Map Display

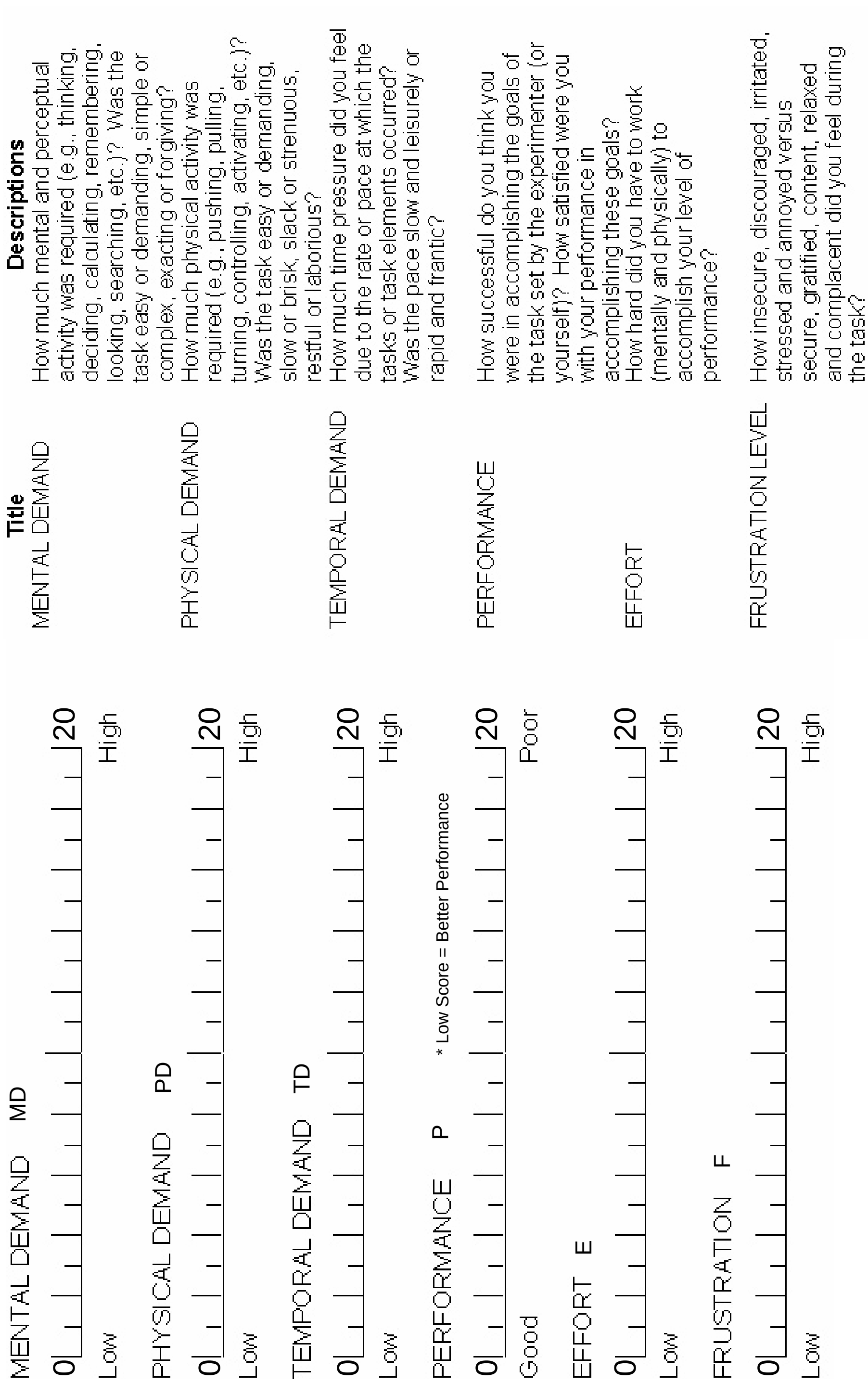
____
____
____
____

Figure C4. Usability Study post-test questionnaire (page 3 of 4).

6. Are there any other display modes you would like available for surface operations as part of multi-mode display concepts?

7. The display concepts you evaluated today are initial designs that will be modified and optimized in later studies. Please provide your thoughts, comments, etc. on how we can best improve upon these display concepts to improve their efficacy for surface operations

Figure C5. Usability Study post-test questionnaire (page 4 of 4).



Title	Descriptions
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?
EFFORT	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Figure C6. NASA TLX post-run questionnaire used in Experiments I and II.

Please rate your how much overall situation awareness you had while taxiing the scenario. One could define situation awareness as "...the pilot has an integrated understanding of the factors that will contribute to the safe taxiing of the aircraft under normal or non-normal conditions."

S1 100 |———————— 50 ————————| 0

High Situation Awareness — Low Situation Awareness

Please rate your overall impression of the scenario in terms of how much attention and effort was required to perform the scenario successfully. Things to consider are the degree of instability, complexity, and variability that you perceived while taxiing the scenario.

S2 100 |———————— 50 ————————| 0

Low Demand — High Demand

Please rate your overall impression of the scenario in terms of the amount of "spare" attention to give to other tasks. Was 100% of your attention directed towards successfully completing the scenario? Or, could you have completed other sub-tasks while taxiing the scenario? Things to consider include your level of arousal, level of concentration, and if you attention was divided across many sub-tasks.

S3 100 |———————— 50 ————————| 0

High Supply — Low Supply

Please rate your overall understanding of what was was happening with the aircraft. Mark on the line below the degree to which you felt confident that you were aware of the elements in your environment. Things to consider include the level of information quantity and quality as well as familiarity that you felt you had with what was taking place during the scenario.

S4 100 |———————— 50 ————————| 0

High Understanding — Low Understanding

Figure C7. SART post-run questionnaire used in Experiments I and II.

1. Please rate the contribution of the display concept to overall taxi efficiency*

R1 |------|------|------|------|
1 2 3 4 5
Not At All ... Very Much

2. Please rate the display concept for navigational awareness for the following five dimensions*:

2a. Overall Navigation Awareness

R2a |------|------|------|------|
1 2 3 4 5
Not At All ... Very Much

2b. Taxi Route Awareness During Final /Roll-Out

R2b |------|------|------|------|
1 2 3 4 5
Not At All ... Very Much

2c. Taxi Route Awareness During Taxiing

R2c |------|------|------|------|
1 2 3 4 5
Not At All ... Very Much

2d. Awareness of Other Aircraft

R2d |------|------|------|------|
1 2 3 4 5
Not At All ... Very Much

2e. Awareness of Direction of Travel

R2e |------|------|------|------|
1 2 3 4 5
Not At All ... Very Much

3. Please rate how beneficial each display concept was to your ability to safety taxi the aircraft*

R3 |------|------|------|------|
1 2 3 4 5
Not At All ... Very Much

*McCann et al., 1998

Figure C8. Taxi Awareness post-run questionnaire used in Experiments I and II.

Instructions: Please circle the severity of any symptoms that apply to you <u>right now</u>.

	(0)	(1)	(2)	(3)
1. General Discomfort	None	Slight	Moderate	Severe
2. Fatigue	None	Slight	Moderate	Severe
3. Headache	None	Slight	Moderate	Severe
4. Eye Strain	None	Slight	Moderate	Severe
5. Difficulty Focusing	None	Slight	Moderate	Severe
6. Increased Salivation	None	Slight	Moderate	Severe
7. Sweating	None	Slight	Moderate	Severe
8. Nausea	None	Slight	Moderate	Severe
9. Difficulty Concentrating	None	Slight	Moderate	Severe
10. Fullness of Head	None	Slight	Moderate	Severe
11. Blurred Vision	None	Slight	Moderate	Severe
12. Dizzy (Eyes Open)	None	Slight	Moderate	Severe
13. Dizzy (Eyes Closed)	None	Slight	Moderate	Severe
14. Vertigo*	None	Slight	Moderate	Severe
15. Stomach Awareness**	None	Slight	Moderate	Severe
16. Burping	None	Slight	Moderate	Severe

*Vertigo refers to a loss of orientation with respect to upright (i.e., you don't know "which way is up")

**Stomach awareness is usually used to indicate a feeling of discomfort which is just short of nausea

Are there any other symptoms that you are experiencing <u>right now</u>? If so, please describe the symptom(s) and rate their severity below.

Figure C9. Simulator Sickness post-run questionnaire used in Experiments I and II.

1. What was your present heading? ___________

2. What was your present ground speed? _________

3. Which taxiway is the aircraft presently located? Taxiway __________

4. Is the aircraft currently ____ left ____ center _____ right of taxiway centerline?

5. What was your full taxi clearance? ____________________________________

6. Which taxiway did the ground controller next expect the aircraft to taxi to from present location? ____

7. Which direction were you to next turn the aircraft onto taxiway? _________

8. Estimate the time of taxi route completion from present location to concourse in seconds ______.

9. How many aircraft were you aware of (perceived out-the-window and/or display concept(s)) _________________

10. Where was the location of the nearest aircraft to your aircraft? _______________

11. What was the heading and ground speed of nearest aircraft to your ownship?
    Heading __________ Ground Speed _____________

12. Please provide a description of the objective of the nearest aircraft to your ownship ____
    ____

Figure C10. SAGAT questionnaire used in Experiment I.

Head-Worn Display Usability

		Strongly Disagree 1	2	3	4	Strongly Agree 5
1	I think that I would like to use this system frequently					
2	I found the system unnecessarily complex					
3	I thought the system was easy to use					
4	I think that I would need the support of a technical person to be able to use this system					
5	I found the various functions in this system were well integrated					
6	I thought there was too much inconsistency in this system					
7	I would imagine that most people would learn to use this system very quickly					
8	I found the system very cumbersome to use					
9	I felt very confident using the system					
10	I needed to learn a lot of things before I could get going with this system					

Comments:

Figure C11. HWD Usability post-test questionnaire used in Experiment II.

Appendix D

Taxi Routes for Experiments I and II

The routes used for Experiment I are shown in Tables D1 - D3 (refer to Fig. E2). There were a total of 27 data runs (Table D4) per crew: 12 nominal routes in Table D1, 12 "error" routes in Table D2 and the three rare event scenarios in Table D3. The error routes were recreated from NASA Ames Taxiway-Navigation And Situation Awareness (T-NASA) experiments in which crews committed an error during the route. The display configurations were experimentally blocked. After training, the first block consisted of evaluations using "conventional" displays (head-down and head-up display concepts), followed by a second block consisting of evaluations using the HWD concepts. This pattern was repeated twice more. The routes were randomly assigned with the exception of the rare event routes in Table D3. The Air Traffic Control (ATC) error rare event was always the 6th run of the day. The non-transponding traffic rare event was always the 25th run. The nose-to-nose rare event scenario was always the last run.

For Experiment II, the routes in Table D5 replaced the routes in Table D1 and the only rare event was the nose-to-nose (scenario 26 in Table D3). There were a total 25 data runs (Table D6) per crew. The 12 error routes from Experiment I were used in Experiment II. As with Experiment I, the nose-to-nose rare event scenario was the last run of the day.

Table D1. Nominal Taxi Routes for Experiment I

Scenario	**Clearance**	**SAGAT**
1	P4 > P > A > Concourse K	
2	P4 > P > A > Concourse E	
3	P4 > P > H > B > A4 > Concourse C	YES
4	T7 > A7 > A > Concourse B	
5	T7 > T > A10 > B > A11 > Concourse F	YES
6	T7 > T > A10 > B > F > Concourse H	YES
7	T > A10 > A > Concourse C	
8	T > M > M2 > B > F > Concourse H	
9	T > M > M2 > A12 > A > Concourse F	
10	C > B > A10 > A > Concourse E	
11	C > B > A10 > A > Concourse B	
12	C > B > R > A > H > P > A > Concourse L	YES

Table D2. Taxi Routes Used in Previous Experiments in Which an Error was Made

Scenario	Clearance
13	M7 > D > A17 > Concourse K
14	T6 > T > M > M2 > B > A15 > A > Concourse H
15	C > B > A9 > A > Concourse E
16	T6 > T > M > M2 > B > F > A > Concourse E
17	M6 > M > F > B > A18 > Concourse K
18	Q > M > M5 > D > A16 > A > Concourse G
19	T6 > T > M > M2 > B > F > A > Concourse K
20	M7 > M > D > A17 > A > Concourse H
21	M7 > M > M5 > D > B > F > A > Concourse F
22	C > B > A10 > A > Concourse E
23	M7 > M > F > B > A17 > Concourse H
24	M7 > M > F > B > D > D1 > A19 > A > Concourse L

Table D3. Rare Event Taxi Routes

Scenario	Clearance	Description
25	T > M > K > T10 > A10 > Concourse F	ATC Error
26	M7 > D > M6 > M > D > A17 > Concourse K	Nose to Nose
27	M7 > D > A17 > Concourse K	Non-transponding Traffic

Table D4. Typical Run Sheet for Experiment I

Run	**Display**	**Weather**	**Scenario**
1	Advanced HUD	Day 700 RVR	15
2	Baseline EMM	Night VMC	5
3	Baseline EMM	Day 700 RVR	13
4	Advanced HWD	Night VMC	4
5	Intermediate HWD	Day 700 RVR	14
6	Advanced HWD	Night VMC	25
7	Intermediate HWD	Day 700 RVR	18
8	Advanced HWD	Night VMC	8
9	Advanced HUD	Day 700 RVR	23
10	Advanced HUD	Night VMC	3
11	Baseline EMM	Day 700 RVR	17
12	Baseline EMM	Night VMC	9
13	Advanced HWD	Day 700 RVR	16
14	Intermediate HWD	Night VMC	10
15	Advanced HWD	Day 700 RVR	20
16	Intermediate HWD	Day 700 RVR	22
17	Advanced HWD	Night VMC	12
18	Intermediate HWD	Night VMC	2
19	Baseline EMM	Night VMC	1
20	Advanced HUD	Night VMC	7
21	Baseline EMM	Night VMC	21
22	Advanced HUD	Night VMC	11
23	Advanced HUD	Day 700 RVR	19
24	Advanced HUD	Day 700 RVR	27
25	Intermediate HWD	Night VMC	6
26	Advanced HWD	Day 700 RVR	24
27	Baseline EMM	Day 500 RVR	26

Table D5. Nominal Taxi Routes for Experiment II

Scenario	**Clearance**
1	T3 > H > E > Concourse C
2	T7 > A7 > Concourse C
3	T10> A10 > Concourse E
4	M4 > F > Concourse H
5	M4 > M > M2 > A12 > Concourse F
6	M3 > F > A > Concourse G
7	M3 > F > B > A15 > Concourse H
8	M5 > D1 > B > A19 > Concourse K
9	M6 > D4 > Concourse M
10	C > B > E > Concourse C
11	C > B > R > Concourse C
12	C > H > Concourse B

Table D6. Typical Run Sheet for Experiment II

Run	Display	Weather	Scenario
1	Advanced HUD	Day 700 RVR	10
2	Advanced EMM	Day 700 RVR	8
3	Advanced EMM	Day 700 RVR	20
4	Advanced HWD	Day 700 RVR	11
5	Advanced HWD	Day 700 RVR	23
6	Paper Chart	Day 700 RVR	9
7	Paper Chart	Night VMC	17
8	Advanced HUD	Day 700 RVR	22
9	Advanced HUD	Night VMC	2
10	Advanced EMM	Night VMC	16
11	Advanced EMM	Day 700 RVR	24
12	Advanced HWD	Night VMC	3
13	Advanced HWD	Night VMC	15
14	Advanced HWD	Day 700 RVR	19
15	Paper Chart	Night VMC	1
16	Paper Chart	Night VMC	5
17	Paper Chart	Night VMC	13
18	Advanced EMM	Night VMC	4
19	Advanced HUD	Night VMC	14
20	Advanced EMM	Day 700 RVR	12
21	Advanced HUD	Night VMC	18
22	Advanced HUD	Night VMC	6
23	Paper Chart	Day 700 RVR	21
24	Advanced HWD	Day 700 RVR	7
25	Paper Chart	Day 500 RVR	26

Appendix E

Airport Charts

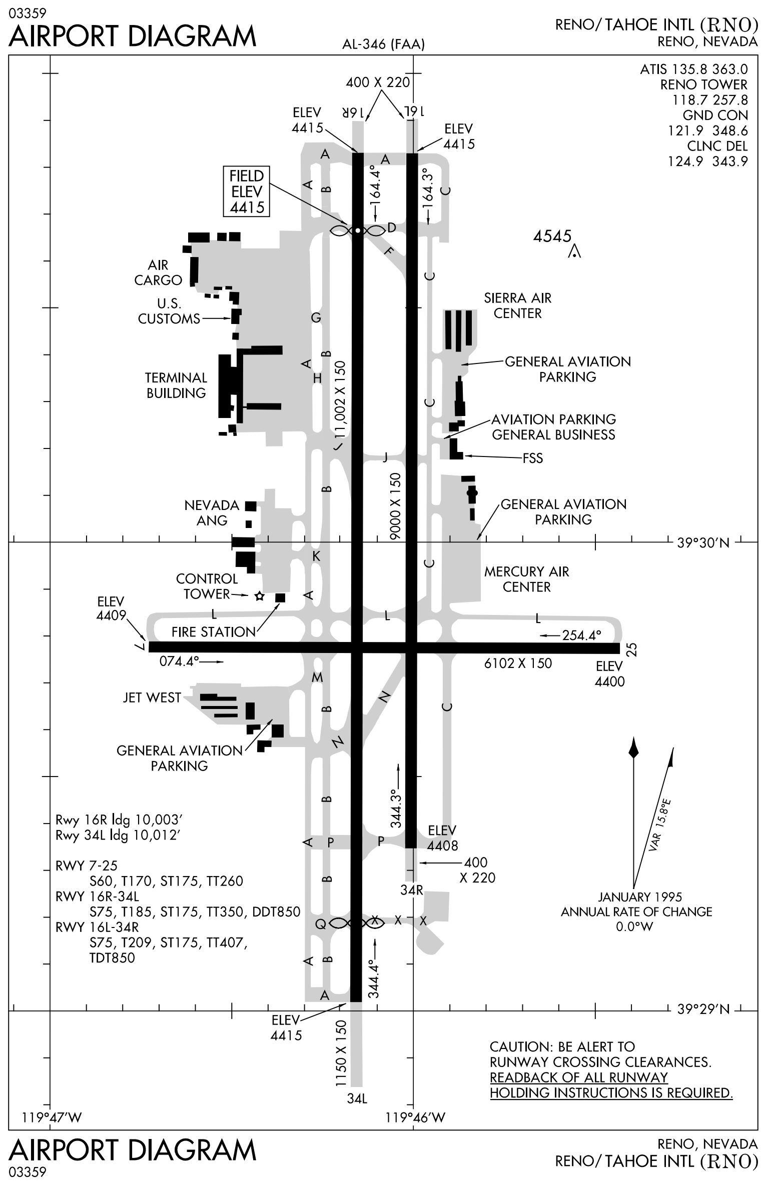


Figure E1. Reno/Tahoe International airport chart. (Not for navigation).

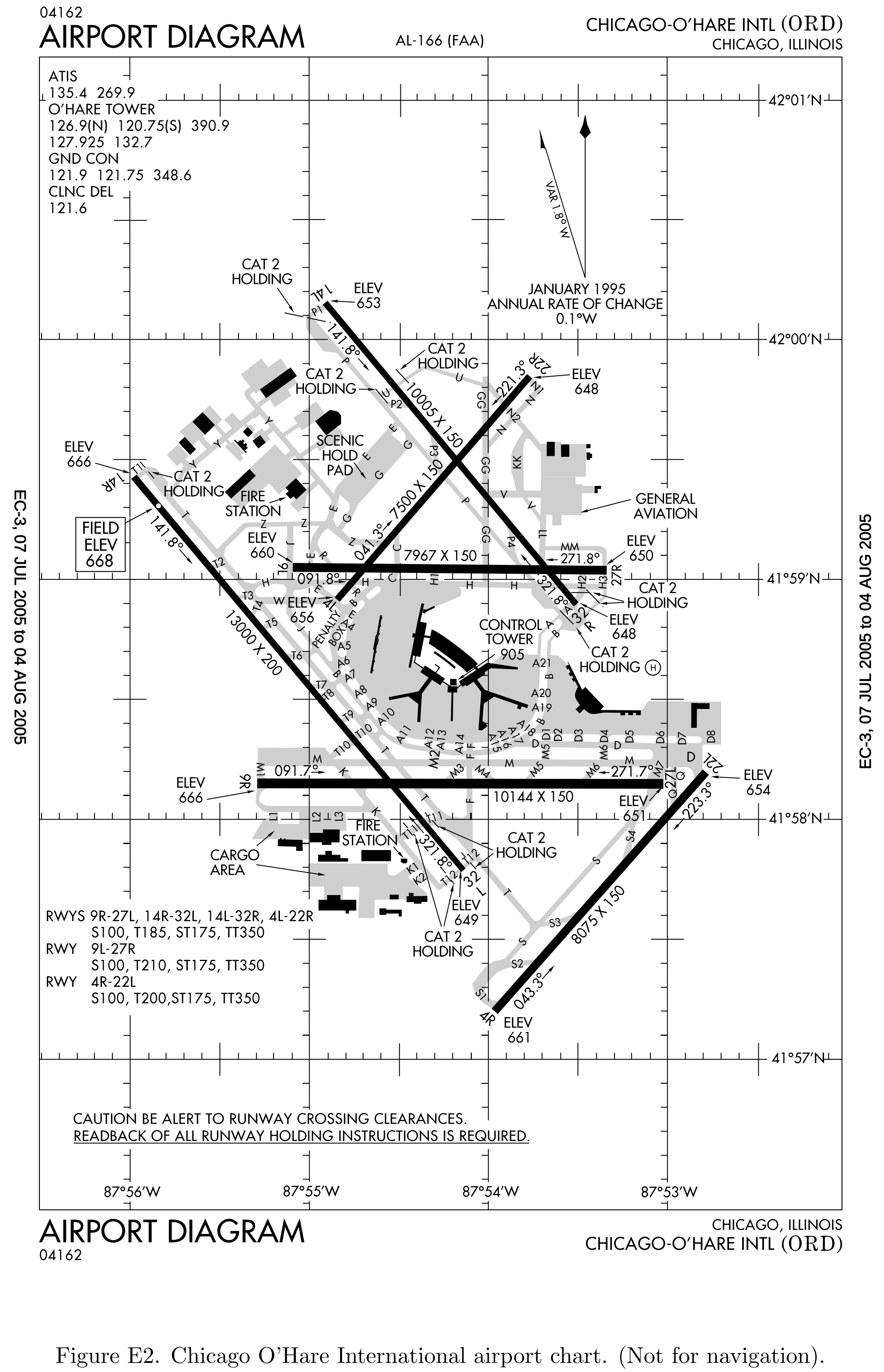


Figure E2. Chicago O'Hare International airport chart. (Not for navigation).

Appendix F

Symbology for Experiment I

The symbology for Experiment I was redesigned based on comments from the usability study. The elements of different modes from the usability study were combined. The Head-Down Display (HDD) was designed around the *Overhead* mode as this was the top-rated mode, therefore, it was always displayed head down. The clearance data were added as two lines of text at the bottom of the display. Further, the virtual camera angle was moved from directly overhead to a "tethered" view. The tethered view provided intuitive perspective depth cues (items farther away appear smaller). Table F1 shows the placement of the virtual camera as a function of the range scale.

Table F1. Camera Placement for Viewpoint of EMM. The camera angle (Angle column) is measured from the vertical axis (airport surface to the sky).

Range Scale	**Angle**	**Distance from Ownship**	**Height Above Ground**
1X	25°	2000 ft	150 ft
2X	25°	4000 ft	200 ft
3X	25°	6000 ft	310 ft
4X	25°	8000 ft	400 ft

The Baseline Electronic Moving Map (EMM) (Fig. F1) consisted of a model of the Chicago Airport (including taxiway markings), taxiway/runway labels, ownship symbol, range scale, ground speed and heading indicator. Taxiway centerlines were included on all EMM display configurations. The T-NASA concept did not include taxiway centerlines on the EMM because it was not intended to be a precision guidance display. For the experiments in this paper, the EMM was also not intended to be a precision guidance display and the EPs were instructed not to use the EMM for precision guidance. The taxiway centerlines were displayed on the EMM as the centerlines were part of the Chicago airport model. The EMM display was EP selectable to be track-up or north-up. The north-up mode showed the entire airport and was not scalable (Fig. F2). The taxiway/runway labels were "halo"-ed in order to improve readability.

The Advanced EMM (Fig. F3) contained all of the elements of the Baseline EMM with the addition of a displayed route, clearance information and traffic icons. The traffic icons were a fixed size and did not scale as the range scale changed. The traffic icons provided location and heading information but there was no indication of size. In other words, a small and large aircraft were depicted by the same size traffic icon. When the ownship came within 300 feet of another airplane, a range ring would be displayed around the traf-

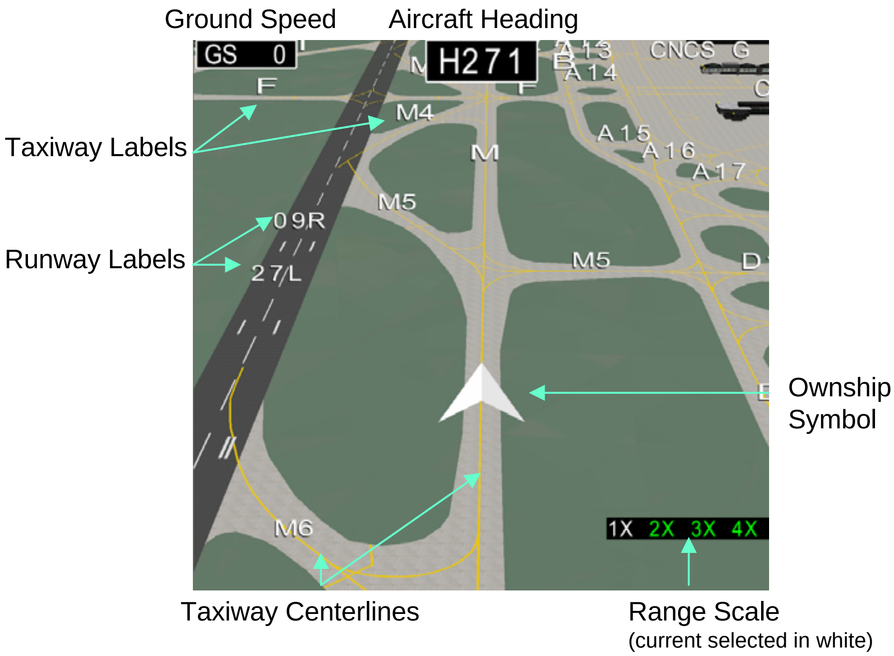


Figure F1. Baseline EMM display.

fic icon. This range ring represented (to scale) the longest dimension (plus an additional 20 ft as a safety margin) of the aircraft. In addition, the aircraft call sign and type were displayed as text. The traffic icons were colored brown as Automatic Dependent Surveillance-Broadcast (ADS-B) targets are colored brown for aircraft on the surface with a white "halo" outline to improve contrast readability. When the traffic ring was visible, the traffic icon halo color changed from white to yellow.

Three different types of head-up displays were used in Experiment I. The Intermediate HWD (Fig. F4) was analogous to the Baseline EMM as the display contained static, non-transponding airport objects. The difference was the virtual camera angle of the Intermediate HWD was placed at the pilot's eye reference point (Table F2) rather than the "tethered" camera placement (Table F1).

Table F2. Pilot Eye Vector for the Captain from CG.

Distance	Direction
24.0 meters	+ out nose
-0.5334 meters	+ starboard
1.0851 meters	+ up

The HUD concept was developed based on the T-NASA and Runway Incursion Prevention System (RIPS) concepts (Fig. F5). The cleared path was displayed as scene linked symbology and an overhead display (known as the

Figure F2. Advanced EMM in north-up mode.

taxi director). The scene-linked symbology consisted of cones that were placed at the edge of the route. The centerline scene-linked symbology was depicted as regularly spaced squares on the route centerline. For turns, flags were placed where the pennant of the flag pointed toward the turn, as was done in the T-NASA system. The flags added awareness of turns because the scene-linked cones would go out of the FOV of the HUD. In addition, virtual turn signs, similar to road side turn signs, enhanced turn awareness. The RIPS-developed taxi director is a non-conformal, top down view of the airplane similar to the "bread crumb" display (Fig. A2).

The Advanced HWD display (Fig. F6) consisted of three major parts: 1) the perspective view of the airport, 2) a small "bread crumb" type display and 3) the route clearance text. The perspective view was the same as the Intermediate HWD with the addition of route symbology and 3-Dimensional (3-D) traffic icons. The route was displayed as a magenta line, 80 feet wide along the centerline of the cleared route. The 3-D traffic icons were a generic airplane shape and size and were a brown color to match the color of the traffic icons on the Advanced EMM. The insert display was a small top down view of the airport surface with the addition of the airplane outline. The insert display could be removed by the EP with a press of the declutter button. The clearance text display was the same as that displayed on the Advanced EMM.

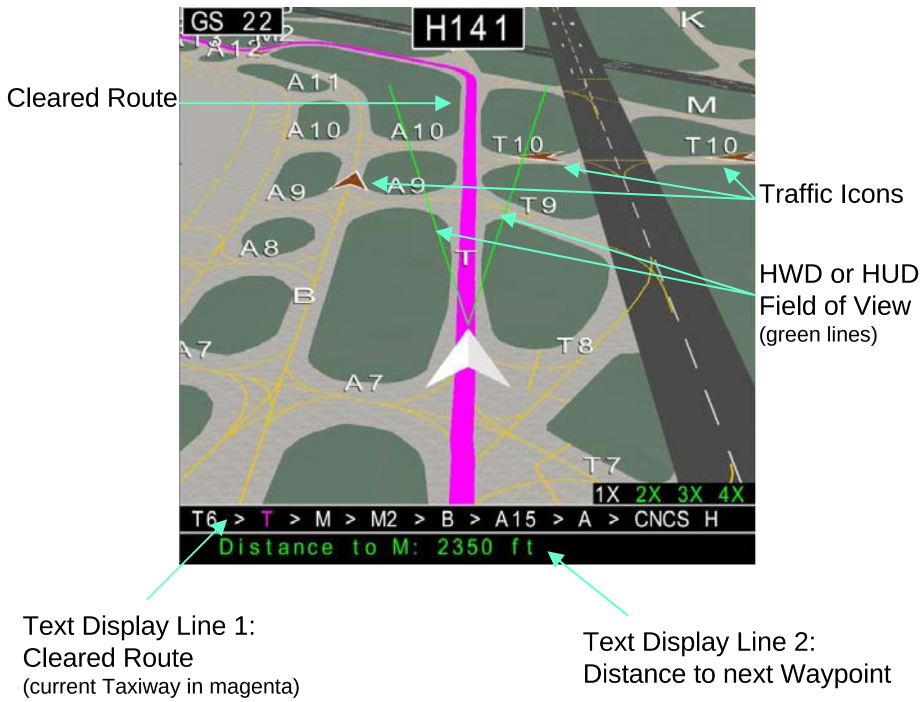


Figure F3. Advanced EMM display.

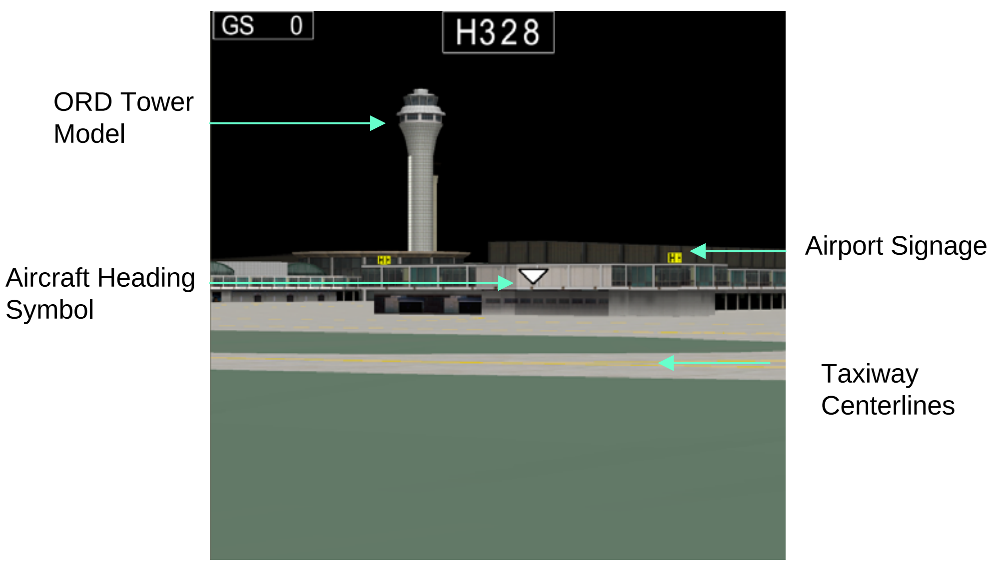


Figure F4. Intermediate HWD.

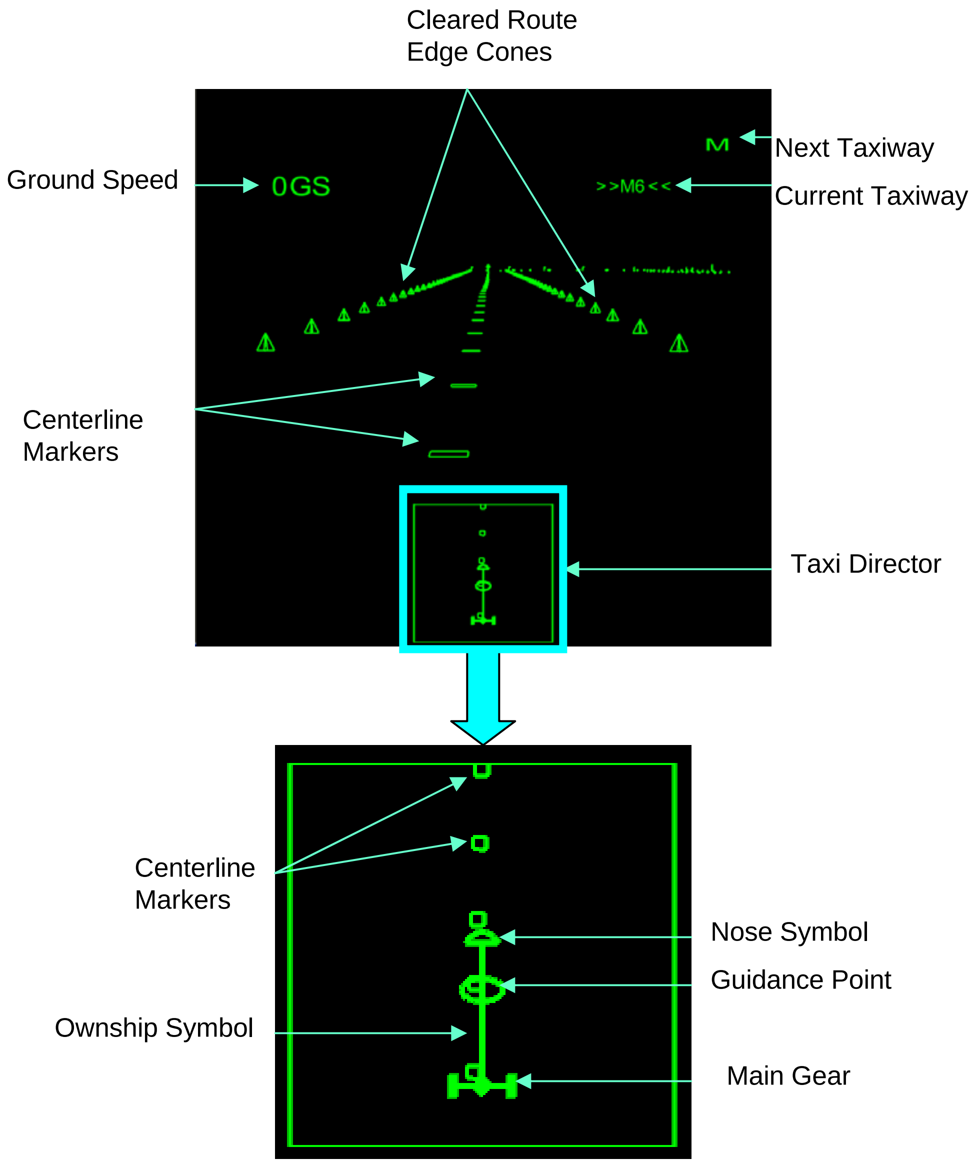


Figure F5. Advanced HUD.

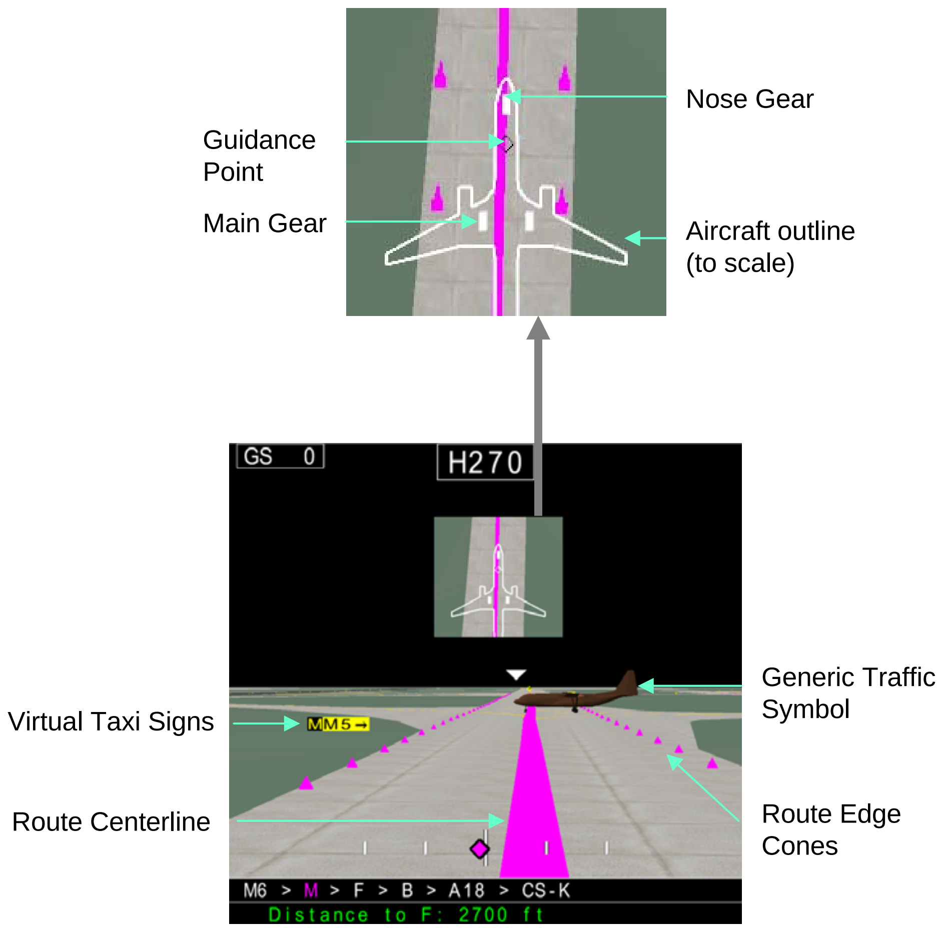


Figure F6. Advanced HWD.

Appendix G

Symbology for Experiment II

For Experiment II, the Baseline EMM concept was replaced with a Paper Chart concept. For the Paper Chart concept, the EPs were given a paper surface map (Fig. E2) and the HDD were traditional cockpit displays (i.e., there was no EMM displayed). The Intermediate HWD concept from Experiment I was replaced with an Advanced EMM display (Fig. G1). There was no head-up display with the Advanced EMM concept. Also, the traffic icon color was changed from brown to cyan to provide for a higher contrast. The size of the traffic icons was also changed as a function of range (Table G1). The size of the traffic in Table G1 represents the wingspan and the fuselage length of the traffic icon.

Table G1. Traffic Size as a Function of Pilot Selected Range

Range	Traffic Size
1X	130 ft
2X	130 ft
3X	170 ft
4X	200 ft

The Advanced HUD for Experiment II was the same as Experiment I. The Advanced HWD was modified slightly for Experiment II. The 3-D traffic icons were colored cyan to match the traffic icons on the Advanced EMM of Experiment II. Also, the 3-D traffic icons had directional and non-directional strobe lights on the models. The insert display was modified to improve readability. In Experiment I, the insert display contained a full virtual airport model. For Experiment II, this insert airport model was reduced to taxiway and runway edge lines. The clearance text was unchanged from Experiment I. Figures G1 through G4 show the symbologies that were changed from Experiment I to Experiment II .

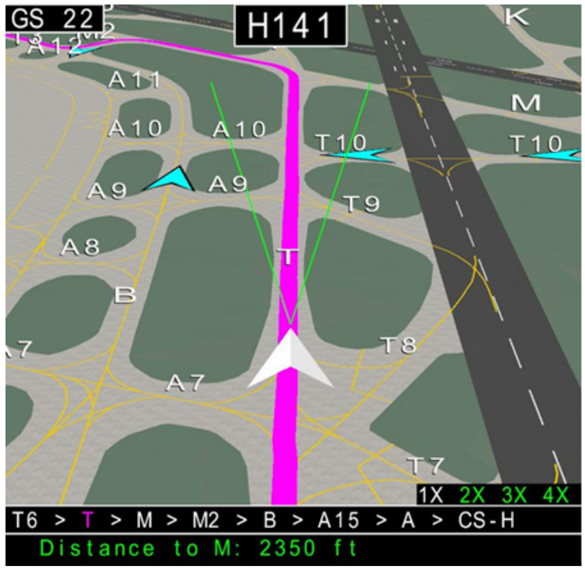


Figure G1. The Advanced EMM with traffic icons (blue chevrons), routing and clearance text.

Figure G2. Advanced EMM in north-up mode for Experiment II.

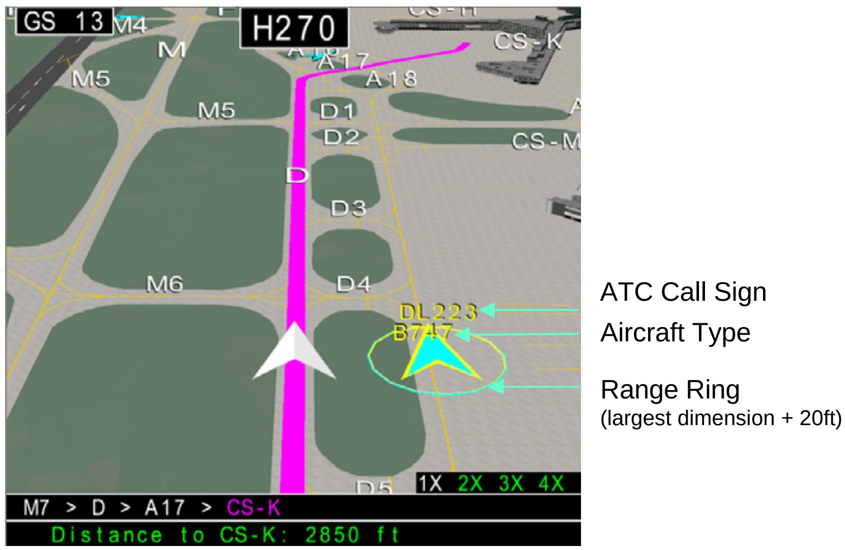


Figure G3. Advanced EMM showing traffic within 300 feet of ownship.

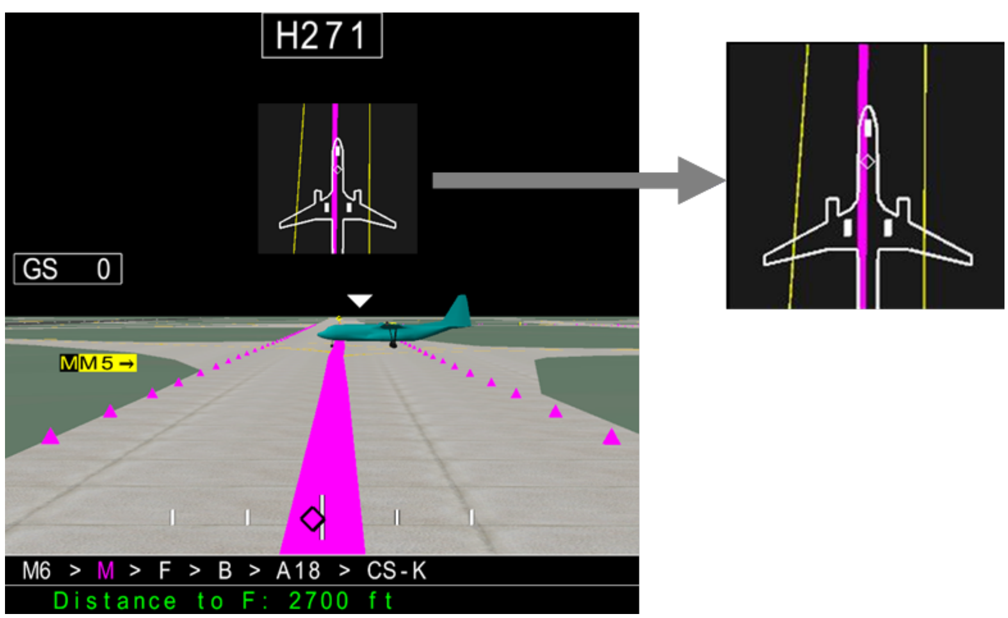


Figure G4. Advanced HWD in Experiment II showing cyan traffic model and improved Insert Display. The higher contrast of the Insert Display improved readability.

REPORT DOCUMENTATION PAGE					*Form Approved* *OMB No. 0704–0188*

1. REPORT DATE *(DD-MM-YYYY)*
01-06-2008

2. REPORT TYPE
Technical Publication

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
Head-Worn Display Concepts for Surface Operations for Commercial Aircraft

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Jarvis (Trey) J. Arthur III, Lawrence J. Prinzel III, Randall E. Bailey, Kevin J. Shelton, Steven P. Williams, Lynda J. Kramer and Robert M. Norman

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
609866.02.07.07.02

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NASA Langley Research Center
Hampton, Virginia 23681-2199

8. PERFORMING ORGANIZATION REPORT NUMBER
L–19449

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSOR/MONITOR'S ACRONYM(S)
NASA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
NASA/TP–2008–215321

12. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified-Unlimited
Subject Category 03
Availability: NASA CASI (301) 621-0390

13. SUPPLEMENTARY NOTES
An electronic version can be found at http://ntrs.nasa.gov.

14. ABSTRACT
Experiments and flight tests have shown that a Head-Up Display (HUD) and a head-down electronic moving map (EMM) can be enhanced with Synthetic Vision for airport surface operations. While great success in ground operations was demonstrated with a HUD, the research noted that two major HUD limitations during ground operations were its monochrome form and limited, fixed field-of-regard. A potential solution to these limitations found with HUDs may be emerging with Head Worn Displays (HWDs). HWDs are small display devices that may be worn without significant encumbrance to the user. By coupling the HWD with a head tracker, unlimited field-of-regard may be realized. The results of three ground simulation experiments conducted at NASA Langley Research Center are summarized. The experiments evaluated the efficacy of head-worn display applications of Synthetic Vision and Enhanced Vision technology to improve transport aircraft surface operations. The results of the experiments showed that the fully integrated HWD provided greater pilot performance with respect to staying on the path compared to using paper charts alone. Further, when comparing the HWD with the HUD concept, there were no differences in path performance. In addition, the HWD and HUD concepts were rated via paired-comparisons the same in terms of situation awareness and workload.

15. SUBJECT TERMS
head worn display, synthetic vision, surface operations

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			STI Help Desk (email: help@sti.nasa.gov)
					19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	135	(301) 621-0390