To the Graduate Council:

I am submitting herewith a thesis written by Alan Joseph Wigdahl entitled "A Limited Evaluation of Pilot Technical Performance, Situational Awareness, and Workload when Flying with an Airframe-Referenced 3D Audio Display." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of requirements for the degree of Master of Science, with a major in Aviation Systems.

Richard Ranaudo, Major Professor

We have read this thesis
and recommend its acceptance:

Dr. Stephen Corda

John Muratore

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of
the Graduate School

(Original signatures are on file with official student records.)

A LIMITED EVALUATION OF PILOT TECHNICAL PERFORMANCE, SITUATIONAL AWARENESS, AND WORKLOAD WHEN FLYING WITH AN AIRFRAME-REFERENCED 3D AUDIO DISPLAY

A Thesis
Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Alan Joseph Wigdahl
May 2008

ACKNOWLEDGMENTS

I wish to thank professor Richard Ranaudo for his mentorship and advice, for serving as a safety pilot, and for his tireless efforts in seeing this test effort to completion. I would also like to thank Dr Stephen Corda for his guidance in planning the various modifications to the test aircraft and for serving as a safety pilot for the program. Dr Bruce Whitehead, Mr. Michael Leigh, and Mr. Steven Amyot deserve special recognition for their critical roles in the design and development of the experimental three-dimensional (3D) audio system and aircraft modifications related to the 3D system. Serving as flight test engineers, Mr. Leigh and Mr. Amyot also ensured proper system operation and data collection during the test flights.

DEDICATION

I wish to dedicate this paper to my wife Clara, who is my best friend and true companion, the most beautiful person I know. Clara supported and strengthened me through the significant efforts and time requirements of this thesis. And more importantly, she occasionally made me get out and have fun, especially when I needed it! I love you Clara.

ABSTRACT

The visual workload of general aviation pilots operating alone in adverse conditions is often high. Airframe-referenced 3D audio systems may improve safety and performance by transferring a portion of this workload to the audio modality. These systems use the aircraft's axes as a reference to present audio directional cues to the pilot via a headset.

This experiment compared the technical performance, workload, and situational awareness of pilots as they performed instrument flight procedures with and without an airframe-referenced 3D audio system. Five pilot participants flew six tasks of each audio condition in a Piper Navajo aircraft modified with a 3D audio system.

With the 3D audio system, the pilots showed slightly better technical performance and lower workload, as well as improved situational awareness. However, the improvements from 3D audio were not significant when expanded to the entire pilot population. The lack of statistical significance appeared to be the results of a small pilot sample and better than expected pilot performance in both audio conditions.

The most important results of the experiment were the modes of the responses to a situational awareness questionnaire consisting of 26 Likert-type questions. Pilots clearly indicated they preferred flying with the 3D system to flying without it, suggesting the airframe-referenced 3D audio system may have been beneficial in completing flight tasks. Further research should be performed to document the system's impact on pilot performance, especially at higher levels of workload than were evaluated in this study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table **Page**

LIST OF ABBREVIATIONS

3D	three-dimensional
AGL	above ground level
AHRS	attitude heading reference system
ASCII	American Standard Code for Information Interchange
DVD	digital versatile disk
FAA	Federal Aviation Administration
FTE	flight test engineer
GA	general aviation
GPS	Global Positioning System
HRTF	head-related transfer function
HSI	horizontal situation indicator
IFR	instrument flight rules
IID	interaural intensity difference
IMC	instrument meteorological conditions
ITD	interaural time difference
MSL	mean sea level
MTSU	Middle Tennessee State University
NASA	National Aeronautics and Space Administration
PCMCIA	Personal Computer Memory Card International Association
RMS	root mean square
RNAV	area navigation
TLX	task load index
UTSI	University of Tennessee Space Institute
VMC	visual meteorological conditions
WAAS	wide area augmentation system

1. INTRODUCTION

General Aviation (GA) aircraft are frequently operated by a single pilot who flies complex navigational procedures such as departures and approaches in Instrument Meteorological Conditions (IMC) while adhering to Instrument Flight Rules (IFR). Under these conditions, pilot visual workload is often high, and can be further increased if an aircraft emergency or malfunction occurs that requires immediate pilot attention. Any means for reducing the burden on the pilot's visual channel in these circumstances should improve performance, situational awareness, and safety. One such reduction involves the use of a 3D audio system to transfer navigation and situational awareness cues from the visual channel to the audio channel.

This study investigated the utility of an aircraft-referenced 3D audio system by evaluating the performance, situational awareness, and workload of five instrument rated pilots as they flew a series of controlled flight experiments in a Piper PA 31 Navajo test bed aircraft. Results when flying with an airframe-referenced 3D audio system were compared to results when flying without a 3D audio system. The intent of this experiment was to evaluate, in a realistic high workload setting, an airframe-referenced 3D audio technology that could potentially be implemented in a system affordable for most GA pilots.

BACKGROUND

To appreciate the benefits and applications of 3D audio technology, it is useful to understand the basis of the technology. In general aviation, sounds, such as radio communications and warning tones, have typically been presented to aircrew in a monophonic format via a fixed speaker mounted in the cockpit or through headphones. When presented through the fixed speaker, the crew recognizes all sounds as emanating from the single point where the speaker is located (see figure A-1. Note: All figures and tables are located in the Appendices.). When monophonic audio signals are presented through headsets, they appear to originate from a point on a line halfway between the speakers, which means the sound appears to originate from within the listener's own head. To improve upon the basic monophonic presentation, a number of technologies have been developed. These include stereo, quadraphonic, surround sound, and 3D systems.[1]

Stereo presentation allows sounds to seem as if they originate from any point on a line between two speakers. This is achieved by varying the volumes of the speakers to shift the sound toward one speaker or the other. When using headphones, stereo sounds again appear to originate from a point on a line between the headphone speakers (figure A-1). Thus the sounds appear to originate from within the listener's head. Despite this unusual origin, the listener can clearly discern the directional nature (left or right) of the cue.[2]

Quadraphonic and surround sound systems provide even higher fidelity directional presentation of audio information. Quadraphonic systems utilize four external speakers arranged in a square around the listener or special headsets with two speakers in each earpiece arranged front and back. With this system, sounds can appear to originate from any point on the perimeter of the square (figure A-1). Surround sound systems use four speakers in a quadraphonic arrangement plus an additional speaker located near the listener. Surround sound cues can appear to come from any point within the square bounded by the speakers (figure A-1). The addition of speakers above and below the listener can provide basic vertical cueing.[1]

CONCEPT OF 3D AUDIO

The preceding systems use audio volume and the location of speakers to present audio directional information. 3D systems achieve directional cueing by different means. Spatiality of sounds in a 3D audio system is achieved by altering the sounds to mimic changes that occur to natural sounds as they travel from their sources to each of the listener's ears. The most critical changes are Interaural Time Differences (ITD) and Interaural Intensity Differences (IID). In a situation where a sound originates 45 degrees to a listener's left (see figure A-2), it will first arrive at the listener's left ear and then at his right ear. This time delay is its ITD. Additionally, the listener's head will create an acoustic shadow that attenuates the sound arriving at the listener's right ear. This difference in intensity is its IID.[3] The amount of intensity reduction caused by the listener's head is based, in part, on the frequencies composing the sound. Lower frequency sounds bend more easily around objects and consequently arrive at the right ear with less volume reduction than higher frequency sounds.[2]

Beyond perceiving ITD and IID, the human body improves its ability to pinpoint sounds through the use of its pinnae, or outer ears. The pinnae tend to reflect or suppress various midrange and high frequency sounds based on arrival direction. These pinnae cues are more precise for higher frequencies. Because higher frequency sounds do not easily bend around the listener's head, the ear that is angularly best aligned to the sound source experiences the greatest localization benefit from the pinnae. Even though each listener experiences different sound characteristics based on the unique shape of his or her pinnae, the listener has learned and reinforced his individual "sound map" since birth.[2]

Putting these characteristics together, a sound originating directly in front of a listener has small interaural time and intensity differences, as well as pinnae effects that match the brain's expectation for a sound originating in front of the head versus behind it. As the sound is moved to the left or right of the listener, IDT and IID increase. Pinnae cues also change based on the frequencies and angle of the sound, with the closer ear benefiting from more precise, detailed cues.[4] Using this information, localization accuracies of five degrees in azimuth and fifteen degrees in elevation can be achieved in the forward hemisphere under optimal conditions. The average localization error in the rear hemisphere is approximately twice as large.[3]

In order for a 3D audio system to be most effective, it must replicate all of this information, but especially the ITD and IID. The more accurately it replicates the sounds, the more precisely the listener will be able to spatially locate them. These "sound models", used by 3D audio systems to replicate sounds, are Head-Related

Transfer Functions (HRTFs). Basic HRTFs do little more than apply generalized IIDs and ITDs to sound. More accurate HRTFs are listener-specific and take into consideration the unique shape of the listener's pinnae.[4]

To obtain an individualized HRTF, small microphones are placed in a person's ear canals near the eardrum. The person then keeps their head immobile as multiple sound frequencies and intensities are played at several hundred locations in a purpose-built sound booth.[5] This process precisely maps IIDs, ITDs, and pinna effects for the user. The main drawback to individualized HRTFs is that the development process is lengthy, expensive, and generally impractical if large groups of people are to use the 3D system.

A practical solution is to use generic HRTFs. Generic HRTFs map the acoustic characteristics of one or more common head/pinna shape combinations, and are subsequently used by various users. If multiple generic HRTFs are available, users perform basic localization tests to determine which HRTF provides the best performance.

Once an appropriate HRTF is selected, 3D audio sounds are created by using the HRTF and spatial information to filter and adjust a monaural sound stream. The output has appropriate ITD, IID, and frequency characteristics for a sound originating from the desired direction and distance.

PREVIOUS 3D AUDIO RESEARCH AND TESTS

Research has been conducted in the field of 3D audio for several decades. Initially, research focused on defining and developing this audio technology. As the technology matured, research increasingly focused on the effectiveness of 3D audio in various practical applications. Two notable applications of 3D audio technologies have been in home entertainment (video gaming and some specialized Digital Versatile Disk (DVD) movies) and aviation, especially military aviation. Aviation applications have varied from fielded systems that improve the situational awareness of ground based air traffic controllers at Nellis Air Force Base, Nevada to high priced developmental systems that aim to increase the survivability and effectiveness of the new F-35 Lightning II fighter. One common theme among these aviation applications is that they have been expensive relative to general aviation budgets and have been based upon complex technologies such as helmet mounted visual displays. Due to their price and complexity, these technologies remain out of the reach of the general aviation community.

Nonetheless, the applicability of 3D audio technology to general aviation flight tasks remains apparent. With a focus on GA, a number of research efforts have been conducted to assess less complex, more affordable 3D audio displays for use in aviation situations. Most of this aviation-related 3D audio research has been conducted using ground-based flight simulations. The projects have assessed the effectiveness of 3D audio displays in tasks such as traffic collision avoidance, target acquisition, threat identification, waypoint navigation, taxiway navigation, and improving communications.[1,6-17]

A common finding of these projects is that the use of a 3D audio display resulted in statistically significant performance improvements, regardless of the specific details of 3D audio system implementation.[7-12] Johnson and Dell[1] reported on one notable exception. Their experiment found no significant performance difference between a system employing a stereo audio display and a system employing a 3D audio display. However, the researchers felt their results were impacted by limitations in both the experimental design and the audio display as directional cues were only displayed on the horizon at 45-degree increments. Despite these limitations, this experiment's results seemed to support the conclusions of other researchers including Simpson et al[11] who postulated that even simple non-directional or stereo audio alerting cues had major positive impacts upon which higher fidelity directional cues could build.

The results of the ground-based simulations showed great promise. However, a number of researchers noted the inability of their ground-based tests and simulations to accurately replicate the tasks and workload of the flight environment, and expressed the need for higher fidelity evaluations, whether through improved flight simulation or actual flight test.[8,13,18,19] Yet, little high fidelity evaluation was completed.

Recently, several efforts have collected quantitative flight test data for different implementations of 3D audio. A common variation among these studies was the reference source for their audio cues. The systems used either a head-referenced system or an airframe-referenced system. Head-referenced 3D audio signals are based on the position of the listener's head. For example, if a sound is coming from the right side of the aircraft and the pilot looks in the direction of the sound, the sound will appear to be directly in front of the pilot's head. Conversely, airframe-referenced systems are based on the aircraft's position. With this system, if a sound is coming from the right side of the aircraft and the listener looks in the direction of the sound, the sound will still seem to be coming from the listener's right since the cue is based on the orientation of the aircraft axes.

One recent flight experiment that employed these reference schemes used a Beech King Air to compare performance while flying under three conditions: without an audio directional cueing system, with a head-tracked 3D audio system, and with an airframe-referenced 3D audio system.[18,20] A later study evaluated the performances of a head-referenced system and an airframe-referenced system in completing a navigation task in a modified Cirrus aircraft.[19] Though these experiments did not employ operationally-representative flight tasks, both demonstrated the usefulness of 3D audio display systems for aircraft navigation.

A surprising result emerged from these tests. In both experiments, pilots achieved better performance while flying with an airframe-referenced system than when flying with a head-referenced system. These results contradicted the large pool of data from ground simulations, which consistently indicated that airframe-referenced systems were less intuitive and more likely to lead to localization errors. The researchers attributed the unexpected result to two factors. First, if a pilot was tasked to capture an exact heading while using a head-referenced system, the pilot's head had to be precisely aligned with the aircraft fuselage to avoid introducing angular errors. Second, the effectiveness of any particular reference system depended upon the task at hand. Most ground-tests used a target capture task, which appears to be best accomplished with a head-referenced 3D audio display. The airborne tests evaluated pilots' performance of navigation tasks, for which an airframe-reference system appears to deliver superior results.

The present study expanded upon previous tests by applying an airframe-referenced 3D audio display to a realistic, high workload, operational flight environment. An airframe-referenced system was chosen for two reasons. First, airframe-referenced systems are more practical and less expensive than head-referenced systems, and thus are more accessible for the average pilot.[19] Second, previous results indicated that an airframe-referenced system was better suited for the navigation task of this experiment. It is important to note that, while capable of providing vertical cues, the airframe-referenced 3D audio system only provided horizontal directional cueing during the present research.

The operational flight environment that was an unique aspect of this experiment was achieved by tasking multiengine, instrument rated pilots to simultaneously fly actual instrument approaches, perform normal checklist procedures, and deal with realistic, but simulated, emergency situations in a multiengine airplane. The goal of these measures was to increase pilot workload, especially in the visual channel.

OBJECTIVES

The purpose of the present research was to evaluate the effects of an aircraft-referenced 3D audio system on pilot workload, situational awareness, and technical performance while performing high-workload flight tasks. The flight test was structured in a manner allowing comparison between flight with an airframe-referenced 3D audio system and flight without the system.

Test Objectives

1. Identify differences in pilot technical performance when flying with and without the 3D audio system by evaluating track and cross-track errors during point-to-point and holding pattern navigation tasks.
2. Identify differences in pilot technical performance when flying with and without the 3D audio system by evaluating time elapsed from waypoint passage to initiation of a turn toward the next waypoint in the route of flight.
3. Determine differences in pilot workload when flying with and without the 3D audio system by using the National Aeronautics and Space Administration (NASA) Task Load Index (TLX).
4. Highlight trends in the pilots' perception and opinion of the 3D audio system, their technical performance, workload, and situational awareness.

2. SYSTEM DESCRIPTION

AIRCRAFT DESCRIPTION

Flight tests were accomplished in the University of Tennessee Space Institute (UTSI) Piper PA-31 Navajo research aircraft, tail number N11UT (see figures A-3 and A-4). The Navajo is a small 6-8 passenger twin-engine aircraft that has a maximum takeoff weight of 6500 lb, with a useful load of approximately 2700 lb. The aircraft is powered by two 310 hp Lycoming reciprocating engines connected to Hartzell adjustable pitch propellers. The aircraft has a maximum cruise speed of 188 knots, a never exceed speed of 236 knots, a stall speed of 70 knots, and a service ceiling of 24,000 feet. At typical flight conditions the aircraft has an endurance of more than five hours. During test flights, the pilots maintained approximately 120 knots at 4000 ft above mean sea level (MSL), with an approximate sortie duration of 1.7 hours.

3D AUDIO SYSTEM DESCRIPTION

Aircraft N11UT was modified with a 3D Audio System consisting of two laptop computers, a global positioning system (GPS) unit, a stereo intercom system, and a video system as depicted in figure A-5.

GPS

A panel-mount Garmin 530W GPS was installed in the aircraft. The 530W was Wide Area Augmentation System (WAAS) equipped, providing the capability to fly precision Area Navigation (RNAV) approaches.

During testing, the pilots' primary visual source of navigation information was the map display mode of the GPS (see figure A-6), which had a typical position accuracy of 3-5 meters and a position update rate of 5 Hz. The pilot zoomed in or out of the GPS map as desired to precisely track the desired flight path. Although the evaluation pilot utilized the GPS map display as his primary source of navigational information, aircraft control was maintained by referencing the aircraft's conventional flight instruments (airspeed indicator, altimeter, attitude direction indicator, and horizontal situation indicator).

For data collection purposes, GPS positional data was streamed through a serial output port to the data acquisition computer at a 1 Hz rate. Data, presented in standard engineering units, was transmitted in American Standard Code for Information Interchange (ASCII) format. The GPS data pertinent to this test were present position, actual and desired ground tracks, cross-track error (lateral error from the desired ground track), altitude, time, selected waypoint, and distance from waypoint.

Pressure Transducers

Two Honeywell pressure transducers were installed. An absolute pressure unit (model number HPA200W2DA-B) measured ambient pressure at the static port of the aircraft. A differential pressure unit (PPT0001DWW2VA-B) measured dynamic pressure between the pitot and static ports. The output of each of these units was transmitted electronically to the data acquisition computer, where the signals were converted to altitude and airspeed using standard pitot-static calculations and aircraft-specific corrections. Of the two data streams, only altitude was utilized during data reduction to gain a general sense of aircraft stability during testing.

Attitude Heading Reference System

The aircraft had also been modified with a Crossbow Attitude Heading Reference System (AHRS), model number AHRS400CD-100, that contained angular rate gyros, accelerometers, and magnetic flux detectors on each of three axes (X, Y, Z). The system utilized this information to derive stabilized roll, pitch, and yaw solutions. These parameters were output to the data acquisition computer. Though the AHRS portion of the 3D audio system will likely be important to future 3D audio research, AHRS data were not pertinent to the present test effort.

Data Acquisition Computer

A laptop computer performed the role of a data acquisition computer. The laptop received GPS, AHRS, and pressure transducers data via four RS-232 signals, which were fed into a four port Personal Computer Memory Card International Association (PCMCIA) adaptor card. A purpose-designed LabVIEW computer program, created by the UTSI Flight Research Facility, combined these data streams into composite text files, which were later imported into Microsoft Excel for reduction and analysis. During testing, a flight test engineer interacted with the data acquisition computer to start and stop data recording during each test sequence.

In addition to recording data inputs, the data acquisition computer provided positional commands to a second laptop, which served as an audio rendering computer. These commands were provided in spherical coordinates (horizontal angle, vertical angle, distance) at a 1 Hz rate.

Audio Rendering Computer

The second of the two laptops served as an audio rendering computer. The purpose of this computer was to output a three dimensional audio cue (only two-dimensional for the present research) that presented the location of a selected waypoint. Though the system had the capability to provide vertical audio directional cues, they were not provided or utilized during this research.

The audio renderer received spherical coordinates describing the direction and distance of the waypoint via an Ethernet connection to the data acquisition computer. The renderer used these inputs in conjunction with a generic HRTF, provided by the Air Force Research Lab's Human Effectiveness Directorate at Wright-Patterson Air Force Base, OH, to adjust the amplitude, tone, frequency, and delay attributes of a carrier recording. The HRTF, along with similarly constructed HRTFs, had previously been used in target location and navigation experiments.[11,18,19]

The carrier recording was a pulse of three tones of white noise that repeated at five-second intervals. Broadband white noise was chosen due to its broadband audio frequency characteristics, which promotes accurate spatial localization.[21]

Intercom System

A Flight Com intercom system was designed to combine into one audio presentation the 3D audio cues from the audio renderer, radio communications, and other audio from the aircraft audio panel (intra-cabin "hot mike", Morse code, etc.). However, to minimize distractions to the evaluation pilot during this test effort, external radio communications were deselected via intercom system pushbuttons. Thus, the intercom combined all planned inputs, except external radio communications, into one audio stream, which was played to both pilots and to the primary flight test engineer (FTE) over Bose noise canceling headsets. A fourth audio signal was provided to the video system for recording.

Video System

A video system recorded flight test activity by using a video camera located in the cockpit and an audio input line from the intercom system. The audio input captured 3D audio tones presented to the pilot and crew voice annotations. The voice annotations supplemented the flight test engineer's handwritten notes and data from the acquisition computer. Data was recorded on a HI-8 digital tape, and was later transferred to DVD for archiving.

OTHER TEST INSTALLATIONS

FTE station

A workstation for a flight test engineer was installed in the left side of the Navajo's cabin immediately behind the pilot seats. The station included a custom designed table that held the data acquisition and audio rendering computers. The FTE monitored the audio renderer for proper function, interacted with the data acquisition computer to record the experiment's results, and kept hand-written notes of pertinent aspects of the flight.

Vision restricting device

During flight, the evaluation pilots wore specially designed eyeglasses, commonly used in flight training, that restrict the pilots' forward and peripheral vision. The goal of this restriction was to replicate flight in IMC in a manner that was also consistent with safety.

Installations from Previous Programs

Due to its primary role as a performance, stability, and control research aircraft, the UTSI Piper Navajo is equipped with additional instrumentation including control position, pilot force, and inertial sensors. Except for information from the pressure transducers described in the 3D audio system description, the aircraft's pre-existing flight test capabilities were not applicable to this test program.

EVALUATION PILOTS AND OTHER AIRCREW

Five pilots from the aviation-training program at Middle Tennessee State University (MTSU) were chosen to serve as evaluation pilots in this investigation. Each pilot was 21 or 22 years old and had multi-engine land and instrument ratings. Total flight time (pilot in command, second in command, student) ranged from 210 to 680 hours. Multi-engine and IFR time for each pilot was less than 100 hrs. Only one evaluation pilot had previous flight experience in a Piper Navajo (2.5 hours flight time). All pilots were familiar with the operation of the Garmin 530W GPS. None had previous experience with 3D audio systems.

In addition to five evaluation pilots, two safety pilots and two flight test engineers were involved in this program. Each safety pilot was both an experienced test pilot and a professor at UTSI. The flight test engineers were on staff at UTSI and had extensive experience in flight test.

3. FLIGHT TEST

After the 3D audio system was installed in the aircraft, training was provided to each evaluation pilot before evaluating performance. This training consisted of classroom training, aircraft familiarization, and 3D audio system familiarization. Additionally, each evaluation pilot flew one orientation flight before flying the evaluation profiles.

PREPARATION AND TRAINING

Classroom Training

Classroom training provided necessary background on the nature of the test, the theory underlying 3D audio cueing, the implementation of the audio system in the research aircraft, and the “ground rules” and constraints of the test program. Evaluation pilots were instructed on the flight characteristics, limitations, systems, and normal and emergency procedures of the PA 31aircraft. The procedures were reviewed for each instrument approach to be flown during testing. Finally, guidance was provided concerning the flight scenario, and the NASA TLX and post-flight questionnaires. Training lasted approximately two hours and was conducted by the safety pilots involved with this program.

Cockpit Familiarization

After classroom training, each evaluation pilot received one hour of cockpit familiarization training. The purpose of this training was to ensure that the pilot could efficiently locate instruments and controls, and could perform emergency checklist items for planned simulations of aircraft malfunctions. As an ungraded "graduation" exercise, evaluation pilots located important cockpit controls while blindfolded.

3D Audio System Familiarization

Each evaluation pilot also received ground familiarization with the 3D audio system. During this training, power was applied to applicable aircraft systems to allow operation of the GPS, AHRS, 3D audio system, and aircraft audio system. The output of the 3D audio system was presented to the evaluation pilot for various waypoints so that he could determine how the audio cues related to the GPS displays of the waypoints. During this training, the characteristics of the airframe-referenced system were emphasized. This event lasted 15-30 minutes per pilot.

Orientation Flight

Each pilot was given approximately one hour of flight training, which included normal procedures, simulated emergency procedures, and flight with and without 3D audio cueing. During the preflight briefing, the safety pilot spelled out the roles and responsibilities of each crewmember, priorities for the evaluation pilot, weather requirements, planned flight events, and transfer of aircraft control criteria. During flight, Visual Meteorological Conditions (VMC) were maintained. The purposes of the orientation flights were to ensure a baseline level of competence in instrument flight procedures, to practice executing checklist procedures for simulated aircraft malfunctions, to familiarize the pilots with the local flying area (see figure A-7), and to develop proficiency in the use of the GPS and 3D audio displays in a flight environment.

EVALUATION FLIGHTS

Each evaluation pilot participated in one evaluation flight of approximately 1.7 hours duration. The minimum crew for each evaluation flight was an evaluation pilot, a safety pilot, and an FTE. Occasionally a second FTE was also on board. During test runs, the evaluation pilot completed each planned task to the best of his ability while wearing a vision-restricting device.

The safety pilot was responsible for the overall safe conduct of the mission. The safety pilot took control of the aircraft between test runs to perform normal in-flight checks. When required by the test cards (see example card in figure A-8), the safety pilot operated the GPS and conducted the flight scenario, which required him to introduce simulated emergency procedures and act as an air traffic controller. The FTE operated the data collection and 3D audio systems. The FTE also kept a time-based log of the sortie's events.

When completing test events, the evaluation pilots flew the aircraft at an altitude of at least 2000 ft above ground level (AGL) in daytime VMC. Furthermore, the evaluation pilots were generally limited to yoke, rudder pedal, and throttle inputs. All actions were simulated when performing simulated emergency or abnormal event checklist procedures.

Aircraft control was transferred between the evaluation and safety pilot at various points during the flight, as indicated on the test cards. The safety pilot was responsible for taking control of the aircraft if the evaluation pilot exceeded test limits, approached aircraft limits, or if safety of the flight was jeopardized. However, during testing the safety pilot did not have to take control of the aircraft, except as noted in the test cards.

The evaluation flight consisted of a round robin profile consisting of six test cards. Each test card described the procedures for two test runs: a short navigation route over three waypoints, and a holding task that was anchored upon the third waypoint. These flight tasks were based on the ground tracks of standard RNAV (GPS) approach procedures serving the Tullahoma Regional Airport in Tennessee (see figure A-9). The altitudes indicated on the approach procedures were disregarded. Instead, the evaluation pilot sought to maintain a consistent VFR hemispheric altitude assigned by the safety pilot.

In selecting an altitude, the safety pilot chose one altitude that minimized turbulence, avoided clouds, and minimized wind. The maximum acceptable wind was 15 knots. Winds of greater than 15 knots were expected to introduce significant crosswind corrections, which were beyond the scope of this evaluation. To minimize learning bias, the order of testing, which included 3D cue availability, approach procedures, and simulated emergency procedures, was alternated.

To start each test run, the evaluation pilot began flying when the aircraft was over an initial waypoint and headed toward the intermediate waypoint indicated on the card. A simulated clearance was provided via the safety pilot, which the evaluation pilot acknowledged. The evaluation pilot then selected the appropriate RNAV/GPS instrument approach and then performed a "direct to" the next waypoint in the GPS procedure. Selection of "direct to" was analogous to snapping a notional chalk line between the present position of the aircraft and the next waypoint. This effectively zeroed cross-track error, calculated a more accurate desired track, and also provided an updated display of ground track to the pilots at the start of each run.

The evaluation pilot completed an instrument approach checklist while flying his "cleared" route. As the evaluation pilot neared the intermediate waypoint, the approach checklist was interrupted by the introduction of a simulated emergency, which was called out by the safety pilot. The goal of introducing the simulated emergency was to increase pilot workload, especially visual workload. The corrective actions used to address this emergency were listed on the top portion of test card A (see figure A-10). The evaluation pilot located each object that was indicated on the checklist and simulated or described the action that applied to the object. He did not actually move or adjust any aircraft or subsystem controls with the exception of the flight controls and throttles, as noted above.

While completing this emergency procedure checklist, the evaluation pilot maintained responsibility for navigating precisely along the navigation route at the altitude assigned by the safety pilot. This included performing a lead turn toward the final waypoint as soon as the evaluation pilot recognized that the GPS was commanding the turn.

Throughout the turn and during the subsequent navigation leg, the evaluation pilot continued with the emergency procedure checklist until it was complete. Also, it was important to perform a second “direct to” when the intermediate turn was completed. This established a new desired ground track that linked the aircraft’s current position and the next steerpoint, against which the pilot’s actual performance could be compared. After completing both the initial simulated emergency and the intermediate turn, the evaluation pilot was subjected to a second simulated emergency. The procedure for this emergency is listed on the bottom portion of test card A (see figure A-10).

The evaluation pilot continued to fly, prioritizing his actions between the navigation task, the normal instrument and before landing checklists, and the procedures to address the simulated emergency. Upon reaching the holding fix, the evaluation pilot was to immediately perform either a direct entry or parallel outbound holding entry, as appropriate for the situation. The evaluation pilot did not have the option of performing a teardrop entry.

Upon commencing a turn to enter holding, the safety pilot took control of the aircraft, completed appropriate in-flight checks, and maneuvered the aircraft to the back of the holding pattern. The evaluation pilot completed a NASA TLX questionnaire (see figures C-1 and C-2) during this period.

Once the evaluation pilot completed the TLX questionnaire, the safety pilot positioned the aircraft on the inbound leg of the holding pattern, issued a clearance back to a holding fix, and then transferred control of the aircraft back to the evaluation pilot. The holding task began when the evaluation pilot took control of the aircraft. The evaluation pilot performed a “direct to” the holding fix on the GPS and executed the holding pattern as depicted on the approach plate. Shortly after the run commenced, the safety pilot introduced a third simulated emergency. In response to this simulated emergency, the evaluation pilot accomplished the checklist procedures found on test card B (see figure A-11). The test run was complete when the evaluation pilot reached the holding fix and began the initial turn of the second holding pattern.

At this point, the safety pilot took control of the aircraft, completed appropriate in-flight checks, and maneuvered the aircraft to the starting point of the next test card. The evaluation pilot completed another NASA TLX questionnaire.

After the flight was complete, the evaluation pilot completed a post-flight questionnaire (see figures D-1 through D-6) that consisted of 26 Likert (or Likert-type) questions. Additionally, the FTE performed a cursory review of the flight’s recordings to determine that the required data streams were present and complete.

4. DATA REDUCTION AND ANALYSIS METHODOLOGY

After each evaluation flight, recordings were downloaded from the data acquisition laptop and distributed for analysis. The data recordings contained the following data streams: GPS data (latitude, longitude, altitude, ground speed, distance to waypoint, desired and actual track, cross-track error, waypoint identifier, bearing to waypoint, and magnetic variation) and aircraft instrumentation data (pressure altitude, indicated airspeed, attitude [roll, pitch, heading], rates and accelerations [roll, pitch, yaw], flight control displacements, and time). An example of the data recordings is provided in table A-1.

Of the available data streams, the following were used for this experiment: distance to waypoint, desired and actual track, cross-track error, waypoint identifier, pressure altitude, and time. A visual depiction of the raw cross-track and track error data for a sample test run is presented in figure A-12. The figure shows errors typical of the test effort and demonstrates the relationship wherein track errors result in cross-track errors.

Additionally, the video/audio recording was transferred to digital format for use as needed. Finally, the hardcopy TLX and post-flight questionnaires were scanned into electronic format and distributed for analysis.

PILOT TECHNICAL PERFORMANCE

Cross-Track Error

During flight, the GPS provided cross-track error calculations to the data acquisition computer at a 1 Hz rate. The errors were calculated internally by the GPS by measuring the length of a line normal to the desired ground track that passed through the aircraft's current position. These data were presented in hundredths of a nautical mile, which was the standard output format of the GPS.

During post-flight data analysis, these data were converted to a nautical miles unit of measure. Time references and FTE handwritten notes were added to the spreadsheet for improved insight into the data. The beginning and end of each lead turn and holding turn were identified. Also, the beginning and end of each flight task was identified.

In order to consider the data for any particular task to be valid, the aircraft had to be stable and tracking along course with errors zeroed or minimized. This occurred when three constraints were simultaneously satisfied: cross track error less than 0.05 nm, track error less than 10 deg, and stable aircraft parameters (a subjective judgment). Generally, these conditions were satisfied when the evaluation pilot commanded a "direct to" on the aircraft GPS. However, a "direct to" was not required for the data to be valid. Once the initial conditions were satisfied, data continued to be valid until the aircraft began a turn.

Referencing the procedures described in the preceding Flight Test section, valid data normally began shortly after the evaluation pilot established the aircraft on course to the intermediate waypoint. Data stopped being valid as soon as the aircraft's navigation system commanded a lead turn at the intermediate waypoint. When the aircraft completed the intermediate turn and flew toward the hold point, valid data recommenced as soon as the three conditions mentioned above were again satisfied. Normally, this occurred shortly after the pilot rolled out of the turn. Upon reaching the lead turn or entry turn for holding, data were considered invalid. These first two segments of data corresponded with the top portion of the test card and were designated the "A" run of the card.

The "B" run of the test card corresponded to the inbound leg of the holding task. Data was considered valid as soon as the aircraft was established inbound in the holding pattern with all three conditions met. Data became invalid as soon as the pilot reached the holding point and commenced his second turn in holding.

After the good data were stripped out, they were organized into "cue on" and "cue off" groupings for each pilot. Two variations of root mean square (RMS) analyses were performed on the data. The first was performed about the data's mean, and the second about the course centerline (zero error line). The results are presented as the data's mean, standard deviation, and standard error. In addition to presenting results on a pilot-by-pilot basis, results are presented that have been averaged across all pilots. Statistical significance was established only if a null hypothesis that "3D audio cueing does not impact pilot performance" was rejected. The student T-test (2 tail, p = 0.95) and Mann-Whitney U-test (2 tail, p = 0.95) were used, as appropriate.

Track Error

Overall, track error data were handled in roughly the same manner as cross-track error. There were a few differences worth noting though. First, the GPS provided desired track and actual track data to the data acquisition computer. Both were raw directional data versus error calculations. Desired track and actual track were output in tenths of a degree and degrees, respectively.

Track error was determined post-flight by calculating the difference between actual track and the desired track. As described in the previous section on cross-track error, times and explanatory notes were added to the data. Periods of valid data were determined. A similar analysis of the data was performed for each evaluation pilot and for all pilots combined.

It is important to note the relationship between cross-track error and track error. The primary measure of pilot performance during this test was cross-track error. Evaluation pilots were briefed to fly using the techniques, procedures, task prioritization, and mindset they normally used when flying in IMC. This included minimizing cross-track error due to the fact that Federal Aviation Administration (FAA) procedures required a pilot to maintain a defined "black line" route and not simply a heading (unless being vectored, which was not the case in this experiment). Track error was evaluated as a method of correlating pilot performance.

Time Elapsed Until Turn Initiation

To evaluate the second objective, a time error was calculated for each turn by subtracting the time of day that the GPS commanded a turn from the time of day that the pilot achieved 20 degrees of turn. Additionally, turns were evaluated to determine if they were performed in the procedurally correct direction (left or right).

NASA TLX DATA

The responses from the NASA TLX questionnaires, which were completed while airborne, were input into a computer program developed by the NASA Human Performance Research Group called NASA-TLX v1.0.[22] The program automatically calculated the scores and weightings of each of the six dimensions, as well as the overall workload. The overall workloads from the test runs were listed by individual subjects and by all subjects combined. A Mann-Whitney U Test (2 tail, p = 0.95) was performed for each pilot and for the combined pilot listing. Noteworthy trends were highlighted.

Also, each individual component of the TLX assessment was evaluated. The six components were mental demand, physical demand, temporal demand, performance, effort, and frustration level. Though not a part of a formal TLX assessment, each dimension was treated as a simple, straightforward workload-related question. The ratings for each dimension were grouped by individual pilot and by all pilots combined. A Mann-Whitney U Test (2 tail, $p = 0.95$) was performed to discern the significance of the results. In a manner similar to the TLX evaluation, results of the various workload dimensions were evaluated for trends.

POST-FLIGHT QUESTIONNAIRE DATA

The post-flight questionnaire (see figures D-1 through D-6) consisted of 26 questions arranged into 6 main groups: system implementation, navigation performance, the airframe-reference system, situational awareness, workload, and overall utility. All questions were Likert or Likert-type questions. Basic demographic and flight experience data for the evaluation pilots were also gathered at the end of the questionnaire. An analytical psychologist, a human factors engineer, a flight test engineer, and four pilots reviewed the questionnaire to improve understandability and minimize bias. After the evaluation pilots completed the questionnaires, the mode for each question was highlighted and trends were qualitatively assessed.

5. RESULTS AND DISCUSSION

PILOT TECHNICAL PERFORMANCE

Three aspects of pilot technical performance were evaluated both with and without 3D audio cueing: cross-track error, track error, and time elapsed until turn initiation.

Cross-Track Error Results

An evaluation of cross-track error did not reveal a significant difference in performance between runs flown with 3D audio cueing and runs flown without cueing.

Figure A-13 depicts mean cross-track errors in hundredths of nautical mile when flying with and without the 3D audio system. These means describe the central tendencies of the data, which is to say they showed whether the pilot flew on course or to the left or right. The error bars represent one standard deviation, a measure of the data's compactness. Note that standard deviation information is presented more clearly in figure A-14.

Both the mean errors and their corresponding standard deviations are very small in magnitude considering the flight task that was performed. The magnitudes of the means range from 0.004 NM to 0.044 NM. Stated differently, the mean errors range from 24 to 267 feet, which is approximately 0.5 to 6.5 aircraft wingspans. These are remarkably accurate values given that the pilot was performing complex procedures while flying at 120 knots. Even when accounting for the variability in the flight path of the "least precise" pilot, 95% (2σ) of the data still was less than 0.4 NM from the desired course. This is well within the FAA's expectation for a pilot to remain within two nautical miles of route laterally when navigating under instrument flight rules.

Though four of five pilots flew to the left of course when using the 3D audio system and four of five flew to the right when not using the system, the differences in the means of each condition are not significant (T-test, 2 tail, 95% confidence). Based on the good accuracy experienced in each condition, it seems likely that any apparent trends in mean cross-track error are merely chance occurrences.

Figure A-14 depicts standard deviation of the cross-track error data in hundredths of nautical mile between the no audio condition (baseline aircraft) and the 3D audio condition. As previously stated, standard deviation is a measure of precision or compactness of the data about the mean. Relating this to the experiment, a large standard deviation implies that the evaluation pilot had large lateral deviations in his flight path. Conversely, a small deviation indicates precise navigation with small lateral deviations about the mean value.

During this experiment, pilot 4 (3D audio condition) flew most precisely with a standard deviation of 0.03 NM, or approximately 180 ft. Pilot 1 (baseline condition) flew least precisely with a standard deviation of 0.16 NM or 970 ft. Of note, four of five pilots flew more precisely when using the 3D audio system than without the system. No apparent reasons existed for pilot 3's less precise performance with 3D cueing. The average difference between the baseline and 3D audio conditions was 170 ft. However, these differences are not significant (U-test, 2 tail, $p = 0.95$).

Figure A-15 presents standard error in both the baseline and 3D audio conditions on a pilot-by-pilot basis. Standard error takes into consideration the effects of both the data's mean and standard deviation. However, the individual effects of each component are not readily discerned in this measure. Whereas standard deviation measures the compactness of data about its mean, standard error measures the compactness of the data about a selected reference (in this case the desired flight path).

As is evident in the standard error figure, the results are essentially identical to those of the standard deviation figure. This is due to the small mean errors observed during the experiment. It is notable that four of the five pilots had better overall performance, defined by a smaller standard error, when flying with the 3D audio display. Due to the study's small sample of pilots, it is also no surprise that the differences are not statistically significant (U-test, 2-tail, 95% confidence).

Track Error Results

In general, the pilots' track errors were in line with their cross-track errors. The track errors do not reveal a significant difference in performance between runs flown with 3D audio cueing and runs flown without cueing.

Figure A-16 depicts the mean track error in degrees between the no audio condition (baseline aircraft) and the 3D audio condition. The error bars represent one standard deviation in the data.

Again, the mean errors were small with a worst case of 1.9 degrees. The differences in the data are not statistically significant (T-test, 2-tail, 95% confidence). Track performance was better than expected and is in line with the results of the cross-track error analysis.

Figure A-17 depicts the standard deviations of cross-track error in degrees between the no audio condition (baseline aircraft) and the 3D audio condition. The standard deviations range from 3.0 degrees (pilot 4, 3D audio condition) to 6.1 degrees (pilot 1, baseline condition). Similar to cross-track error performance, all pilots displayed greater heading precision in the 3D audio configuration except pilot 3. These differences are not significant (U-test, 2-tail, 95% confidence).

Figure A-18 presents standard track error in degrees for both the baseline and 3D audio conditions. Again due to the small values of mean track error, there is very little difference between the values of the standard errors and the standard deviations. Thus data on figures A-17 and A-18 are similar. Four of five pilots flew with more precisely in the 3D condition, but without statistically significant differences (U-test, 2-tail, 95% confidence).

Time Elapsed Until Turn Initiation

The second objective of this study, which involved an evaluation of elapsed times until turns were initiated, was not successfully completed. During classroom training, the evaluation pilots were briefed that they should turn as soon as the GPS commanded a turn, but not before the command. The goal was to note the time that elapsed from the turn command until 20 degrees of turn had been completed. During flight however, the pilots generally did not comply with this rule and turned early instead. This was not a willful violation of the mission rules; rather, it was simply a demonstration of solidly ingrained pilot technique.

Though no statistical analysis was performed due to the unexpected behavior, important observations remain. The fact that the pilots tended to turn early seems to imply a good awareness of approaching turn points. This is supported by the fact that the pilots always turned in the correct direction, including during the more complex entry turns into the holding patterns (except for two cases where turns were not completed due to overriding circumstances). It is further supported by the pilots' midrange ratings of workload and by their responses to the post-flight questionnaires, both of which are discussed in subsequent portions of this paper. Thus, pilots seemed to have good situational awareness of approaching turn points and the turns associated with them, regardless of whether or not 3D audio cues were present.

Discussion

A known limitation of this experiment was the sample size of five pilots. This small sample required large, clear-cut performance differences between the baseline and 3D audio conditions for significance to be established. Though it was likely that sufficiently large differences would not be observed, they were possible nonetheless. In day-to-day operations, a high potential exists for pilots to exhibit substantial cross-track and track errors when flying light twin aircraft in high workload IFR situations.

Sizeable errors were not observed during this test, however. In fact, the errors tended to be exceptionally small for both the baseline and 3D audio conditions. The most likely reasons for this appears to be the very effective presentation of positional information by the GPS map display, a high level of performance by the evaluation pilots (in specific response to the challenge of this experiment), and the failure of the experimental procedure to adequately raise pilot workload (especially visual workload) to the point at which performance became degraded. Note that workload is discussed in more detail in the NASA TLX data results subsection.

The exceptional performance of the pilots had a second consequence as well. The small errors associated with the baseline condition provided very little room for improvement in the 3D audio condition. Borrowing from a popular phrase, it is difficult to improve on (near) perfection. Thus, it is likely that any benefit of the 3D audio system on pilot performance is not evident in the data, since there was essentially no room to improve upon the baseline data.

In an opposite sense, it is noteworthy that the 3D audio system did not show a statistically significant negative impact on pilot technical performance. Having a primary goal of evaluating an airframe-referenced audio system as opposed to more common head-referenced systems, the observation that the data did not show a negative impact on pilot performance was especially important. Given that the baseline condition was characterized by very good performance, the experiment offered every opportunity to highlight negative impacts of the 3D system on performance (it is usually easy to perform worse than perfect). However, this did not occur. During testing four out of five pilots performed better with the system than without it. Statistically, the null hypothesis could not be disproved, meaning that the results could not be broadened beyond the test runs to the entire pilot population with meaningful significance.

In order to improve the meaningfulness of the results, follow-on experiments could utilize a visual display that is more difficult to interpret than the GPS map display. For example, a horizontal situation indicator (HSI) display could be used instead of the more intuitive GPS map display. However, the use of a less intuitive display should be tempered against the goal to assess 3D audio performance in a representative, realistic flight environment. With the prevalence of GPS in today's modern GA aircraft, it may be most appropriate to continue using the map display.

More importantly, workload should be increased such that pilot performance degrades from the excellent levels of performance observed in this experiment. Though there are many ways to accomplish this, the use of asymmetric thrust should be considered. Asymmetric thrust would bring the aircraft out of trim, require constant pilot interaction with the flight controls, and greatly increase the overall challenge of flying the aircraft. Additionally, the use of asymmetric thrust would improve the realism of the simulated engine failure procedures used during each test run (see figures A-10 and A-11).

NASA TLX DATA

Table A-2 shows the results of the NASA TLX surveys for the baseline and 3D audio conditions. Three of the five evaluation pilots experienced a slightly lower workload when using 3D audio. The other two experienced slightly higher workload. In no case are the results significant (U-test, 2 tail, $p = 0.95$). Thus, in the third objective of this evaluation, no significant overall differences in NASA TLX results exist between the baseline and 3D audio conditions.

In tables C-2 through C-7, each aspect of the NASA TLX survey is presented separately. In contrast to the formal evaluation that weighed the components based on the respondent's pairwise comparisons, here each component is treated as a simple workload question.

The responses of three of the five pilots do not have statistically significant differences in any of the six components of the NASA TLX. The responses of the other two pilots are significant for several of the components (U-test, 2 tail, $p = 0.95$). Pilot 2 perceived statistically significant improvements in mental demand, performance, and frustration when using the 3D audio system. Pilot 5 indicated statistically significant improvements in performance and effort with the system. These results indicate that the 3D audio system lowered some aspects of workload for two of the pilots, but perhaps not enough to lower their overall workload as measured by the formal NASA TLX procedure. In no case did the results indicate a statistically significant negative impact on workload.

Overall, the TLX assessments and most of the subcomponent assessments do not have significant differences. Based on these statistical analyses alone, the impact of the 3D audio system on pilot workload cannot be determined. However, when also considering the results of the pilots' technical performance and their responses to the post-flight questionnaire, it seems likely that workload was not adequately elevated during the experiment to allow differences between the two audio conditions to become manifest.

The pilots provided low or midrange responses for mental demand, physical demand, performance, and frustration. Temporal demand and effort levels were moderately high, but not greatly so. Across both conditions, the average NASA TLX score was 62, on a scale of zero to 100. If workload is elevated to the point where technical performance becomes significantly degraded, the 3D audio cues might exhibit a larger impact, resulting in greater variation between the baseline and 3D audio workload responses. Again, there are a number of methods to increase workload. One method is to incorporate a realistic asymmetric thrust condition, which will increase the difficulty of the flight task.

POST-FLIGHT QUESTIONNAIRE DATA

While the results of pilot technical performance and the NASA TLX generally lack significance, pilot responses to the post-flight questionnaire clearly favor the use of the 3D audio display during flight. Percentages listed in the following paragraphs refer to percent of pilots associated with the responses and are extracts of the complete questionnaire. The complete questionnaire and associated responses are presented in figures D-1 through D-6.

Overall, pilots felt that the addition of the 3D audio system to the aircraft was somewhat (80%) or very (20%) useful. The pilots (100%) did not find the audio cues to be annoying or distracting.

Four of the five pilots felt that the cue correctly presented directional information to the pilot and that they could localize the cue somewhat precisely. Interestingly, the remaining pilot (pilot 2) strongly disagreed with the statement that the 3D audio cues correctly conveyed directional information to the pilot. He also felt that he could only locate the direction of the cue very imprecisely. His responses are not supported by his technical performance or TLX results, both of which were typical of the pilots' average results. Furthermore, pilot 2 indicated elsewhere in the questionnaire that his navigational performance and turn timeliness were much better with the 3D audio cues. His negative responses might have been a simple misunderstanding of the questions. Alternatively, the generic HRTF used by the experiment might not have been a good fit for pilot 2. A poorly matched HRTF could have notably decreased the pilot's ability to precisely locate cues, a shortcoming for which he could have compensated to achieve better performance than in the baseline condition.

Concerning their ability to maintain the desired ground track and make timely turns, the pilots felt their performance with directional audio to be the same or better than the baseline condition (40% same, 40% somewhat better, 20% much better). Furthermore, all pilots felt their navigational performance was improved when performing emergency procedures while using 3D audio. These replies were consistent with the pilots' workload assessments and technical performance. If the navigation errors in the baseline condition had been large enough to allow significant performance differences between the test conditions, it seems reasonable that the responses to these questions would have shifted more strongly in favor of the 3D audio system. However, the technical results do not show much improvement between the conditions, and the pilot responses to these questions seem to indicate their awareness of this fact.

Similarly, when using 3D audio cues the pilots perceived their workload to be the same or lower than the baseline condition (40% same, 40% somewhat lower, 20% much lower). Only one of the pilots felt he had to concentrate on the audio cues to determine the location of the waypoint, whereas the other four did not have to concentrate to determine the location. Also, 80% of the pilots felt that less mental effort was required to complete flight tasks with the 3D cues. These same pilots stated that they could devote more attention to completing emergency procedures when they relied upon 3D audio cues to alert them of approaching turnpoints.

While it may seem surprising that the pilot who indicated higher workload with the 3D audio display (pilot 4) is the pilot who exhibited the best technical performance of the group, this may be evidence that his workload under the 3D condition was more optimally elevated with 3D audio than without it, allowing superior performance. If workload is too low, performance is normally low due to complacency or boredom. If workload is too high, a performance can suffer due to task saturation and overload. Performance tends to be maximized when workload is elevated, but not overly so.

In this specific case, pilot 4 indicated that somewhat more mental effort was required when using 3D audio cues. He did not feel that the cues allowed him more time to focus on other tasks, such as simulated emergency procedures. To this pilot the need to attend to the audio cues was yet another task added to an already busy flight. The NASA TLX workload results for this pilot are higher in the 3D audio condition, but not significantly so. As previously mentioned, his technical performance was the best of the group. These results may indicate that pilot 4 performed better precisely because he focused more consciously on the audio cues than the other pilots. While this is an interesting possibility to consider, the data ultimately are unable to support a firm conclusion on this issue.

It is noteworthy that all evaluation pilots felt they were able to adapt to and use the airframe-referenced system of 3D audio cues. The cues were somewhat (40%) or very (60%) intuitive when looking forward, in which case the pilot's head was aligned with the airframe-reference system. Even when the pilot's head wasn't aligned with the airframe, a majority (60%) of the pilots felt the cues to be somewhat intuitive.

It seems logical that an airframe-referenced system would work well during tasks where the pilot is looking forward, such as when flying in the weather using instrument flight procedures. Pilot responses concerning the airframe-referenced 3D audio cues strongly support this assumption. The responses also indicate that an airframe-referenced system of cues might be useful even when the pilot is looking to the sides. It is possible that the pilots developed a mental construct for audio cues that were airframe-referenced and visual cues that were head-referenced. Further research into this topic should be performed before conclusions are reached.

Finally, each pilot felt that the 3D audio display improved his situational awareness. 80% of the pilots felt that the audio cues improved their sense of direction to the next waypoint (60% somewhat improved, 20% much improved). However, three of the five pilots felt that the cues did not allow them to precisely determine heading errors. Top-performing pilot 4 was in this group. This may have been another result of using a generic HRTF, as these generic functions tend to be less precise than individualized HRTFs. Considering that the present experiment used a generic HRTF and presented basic unaugmented directional information, results might have been improved through the use of more precise HRTFs or changes to how the audio display presented information.

Overall, the questionnaire results are decidedly in favor of the 3D audio display. These responses are the most important results of the experiment and indicate that further research is merited. Based on the strength of the responses, it seems likely that significant performance improvements would be observed between the 3D audio and baseline conditions if a larger group of pilots was sampled and the pilots' baseline performance had greater variability and lower accuracy.

6. CONCLUSIONS AND RECOMMENDATIONS

Building upon research that demonstrated strong potential for an airframe-referenced 3D audio display to positively impact pilot performance, this experiment compared pilot technical performance, workload, and situational awareness when flying with and without an airframe-referenced 3D audio display.

CONCLUSIONS

Most pilots participating in this study showed improved technical performance when using an airframe-referenced 3D audio display. However, their results were not statistically significant. This lack of significance seems to be a result of very good pilot performance in the baseline audio condition that left little room for improvement when flying in the 3D audio condition. Though no significant positive impacts to technical performance were observed, it is equally important to note that significant negative impacts were also not observed.

The results of the NASA TLX analyses do not show a significant difference in pilot workload between the baseline and 3D audio conditions. Across both audio conditions, the NASA TLX responses revealed midrange, but not high, pilot workloads. If workload had been elevated to the point where technical performance was significantly degraded, the 3D audio cues might have had a larger impact, resulting in greater variation between the baseline and 3D audio workload responses.

Despite a lack of significant differences in the pilot technical performance and workload data, the pilots clearly favored using an airframe-referenced 3D audio display in their responses to the post-flight questionnaire. The pilots unanimously felt the 3D system improved their situational awareness and their navigational performance when performing emergency procedures. Furthermore, the pilots found the airframe reference system of the display to be at least somewhat intuitive, especially when looking straight ahead. The pilots' distinct preference for flying with the airframe-referenced 3D audio system to flying without the system is the most important result of this experiment.

RECOMMENDATIONS

The overall results of the present research, and especially the pilot's preference for flying with the system under test, indicate that further research is merited. Follow-on research efforts should continue to evaluate differences in pilot performance with and without the use of an airframe-referenced 3D audio system.

Where possible, these research efforts should utilize a larger pilot sample to better evaluate quantitative performance differences associated with the 3D system. Even more critical, these tests should ensure that the pilot's workload, especially in the visual modality, is sufficiently elevated to allow researchers to ascertain improvements resulting from the auditory localization system. The use of asymmetric thrust, established by setting idle thrust on one engine, should prove a simple and effective method of raising pilot workload.

LIST OF REFERENCES

List of References

[1]Johnson, C. W., and Dell, W., "Limitations of 3D Audio to Improve Auditory Cues in Aircraft Cockpits," *Proceedings of 2003 International Systems Safety Conference*, Ottawa, 2003, pp. 990-999.

[2]Kapralos, B., Jenkin, M. R. M., and Milios, E., "Auditory Perception and Spatial (3D) Auditory Systems," York University Technical Report CS-2003-07, North York, Ontario, 2003.

[3]Jarrett, D. N., *Cockpit Engineering*, Ashgate, Hampshire, England, UK, 2005, p. 252.

[4]Gardner, W. G., "3D Audio and Acoustic Environment Modeling." Wave Arts, Inc., Arlington, MA, 1999.

[5]Schutte, J., "Aviation Safety Thru 3-D Audio," *Federal Laboratory Consortium for Technology Transfer News Link* [online journal], October/November 2006, URL: http://www.federallabs.org/pdf/OctoberNovember2006-NewsLink.pdf [cited 25 February 2008].

[6]Begault, D. R., "Challenges to the Successful Implementation of 3-D Sound," *Journal of the Audio Engineering Society*, Vol. 39, No. 11, 1991, pp. 864-870.

[7]Begault, D. R., "Head-Up Auditory Displays for Traffic Collision Avoidance System Advisories: A Preliminary Investigation," *Human Factors*, Vol. 35, No. 4, 1993, pp. 707-717.

[8]Begault, D. R., and Pittman, M. T., "Three-Dimensional Audio Versus Head-Down Traffic Alert and Collision Avoidance System Displays," *The International Journal of Aviation Psychology*, Vol. 6, No. 1, 1996, pp. 79-93.

[9]Perrott, D. R., Cisneros, J., McKinley, R. L., and D'Angelo, W. R., "Aurally Aided Visual Search Under Virtual and Free-Field Listening Conditions," *Human Factors*, Vol. 38, No. 4, 1996, pp. 702-715.

[10]Bolia, R. S., D'Angelo, W. R., and McKinley, R. L., "Aurally Aided Visual Search in Three-Dimensional Space," *Human Factors*, Vol. 41, No. 4, 1991, pp. 664-669.

[11]Simpson, B. D., Brungart, D. S., Gilkey, R. H., Cowgill, J. L., Dallman, R. C., Green, R. F., Youngblood, K. L., and Moore, T. J., "3D Audio Cueing for Target Identification in a Simulated Flight Task," *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*, New Orleans, 2004, pp. 1836-1840.

[12]Sandberg, S., Håkansson, C., Elmqvist, N., Tsigas, P., and Fang, C., "Using 3D Audio Guidance to Locate Indoor Static Objects," *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, San Francisco, 2006, pp. 1581-1584.

[13]Bronkhorst, A. W., Veltman, J. A., and Van Vreda, L., "Application of a Three-Dimensional Auditory Display in a Flight Task," *Human Factors*, Vol. 38, No. 1, 1996, pp. 23-33.

[14]Foyle, D. C., Andre, A. D., McCann, R. S., Wenzel, E. M., Begault, D. R., and Battiste, V., "Taxiway Navigation and Situational 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*, Vol. 105, 1996, pp. 1411-1418.

[15]Begault, D. R., Wenzel, E. M., and Lathrop, W. B., "Augmented TCAS Advisories Using a 3-D Audio Guidance System," *Proceedings of the Ninth International*

Symposium on Aviation Psychology, The Ohio State University, Columbus, OH, 1997, pp. 353-357.

[16]Nelson, W. T., Hettinger, L. J., Cunningham, J. A., Brickman, B. J., Haas, M. W., and McKinley, R. L., "Effects of Localized Auditory Information on Visual Target Detection Performance Using a Helmet-Mounted Display," *Human Factors*, Vol. 40, No. 3, 1998, pp. 452-460.

[17]Veltman, J. A., Oving, A. B., and Bronkhorst, A. W., "3-D Audio in the Fighter Cockpit Improves Task Performance," *International Journal of Aviation Psychology*, Vol. 14, No. 3, 2004, pp. 239-256.

[18]Simpson, B. D., Brungart, D. S., Dallman, R. C., Joffrion, J., Presnar, M. D., and Gilkey, R., "Spatial Audio as Navigation Aid and Attitude Indicator," *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*, Orlando, 2005, pp. 1602-1606.

[19]Brungart, D. S., Simpson, B. D., Dallman, R. C., Romigh, G., Yasky, R., and Raquet, J., "A Comparison of Head-Tracked and Vehicle-Tracked Virtual Audio Cues in an Aircraft Navigation Task," *Proceedings of the 13th International Conference on Auditory Display*, ICAD, Montréal, 2007, pp. 32-37.

[20]Joffrion, J. M., Delsing, F. A., Hamilton, C. E., Presnar, M. D., Shinoskie, J. R., and Reed, S. M., "Inertial Head Tracking for 3D Audio," AFFTC TIM 04-06, Edwards AFB, CA, 2004.

[21]King, R. B., and Oldfield, S. R., "The Impact of Signal Bandwidth on Auditory Localization: Implications for the Design of Three-Dimensional Audio Displays," *Human Factors*, Vol. 39, No. 2, 1997, pp. 287-295.

[22]NASA Human Performance Research Group, Task Load Index (NASA-TLX) Computerized Version, Ver. 1.0, NASA Ames Research Center, Moffett Field, CA, 1987.

[23]Fanshier, B., *Garmin 500W Series Pilot's Guide and Reference,* Garmin International, Olathe, KS, 2006, p. 33.

[24]National Aeronautical Charting Office of the Federal Aviation Administration, *Atlanta Sectional Aeronautical Chart* [online chart], 15 March 2007, URL: http://naco.faa.gov/index.asp?xml=naco/catalog/charts/vfr/sectional [cited 30 March 2007].

[25]National Aeronautical Charting Office of the Federal Aviation Administration, *Tullahoma Regional Airport RNAV (GPS) 18 Instrument Approach Plate* [online plate], 15 March 2007, URL: http://204.108.4.16/d-tpp/0703/00508R18.PDF [cited 30 March 2007].

APPENDICES

APPENDIX A – FIGURES AND TABLES

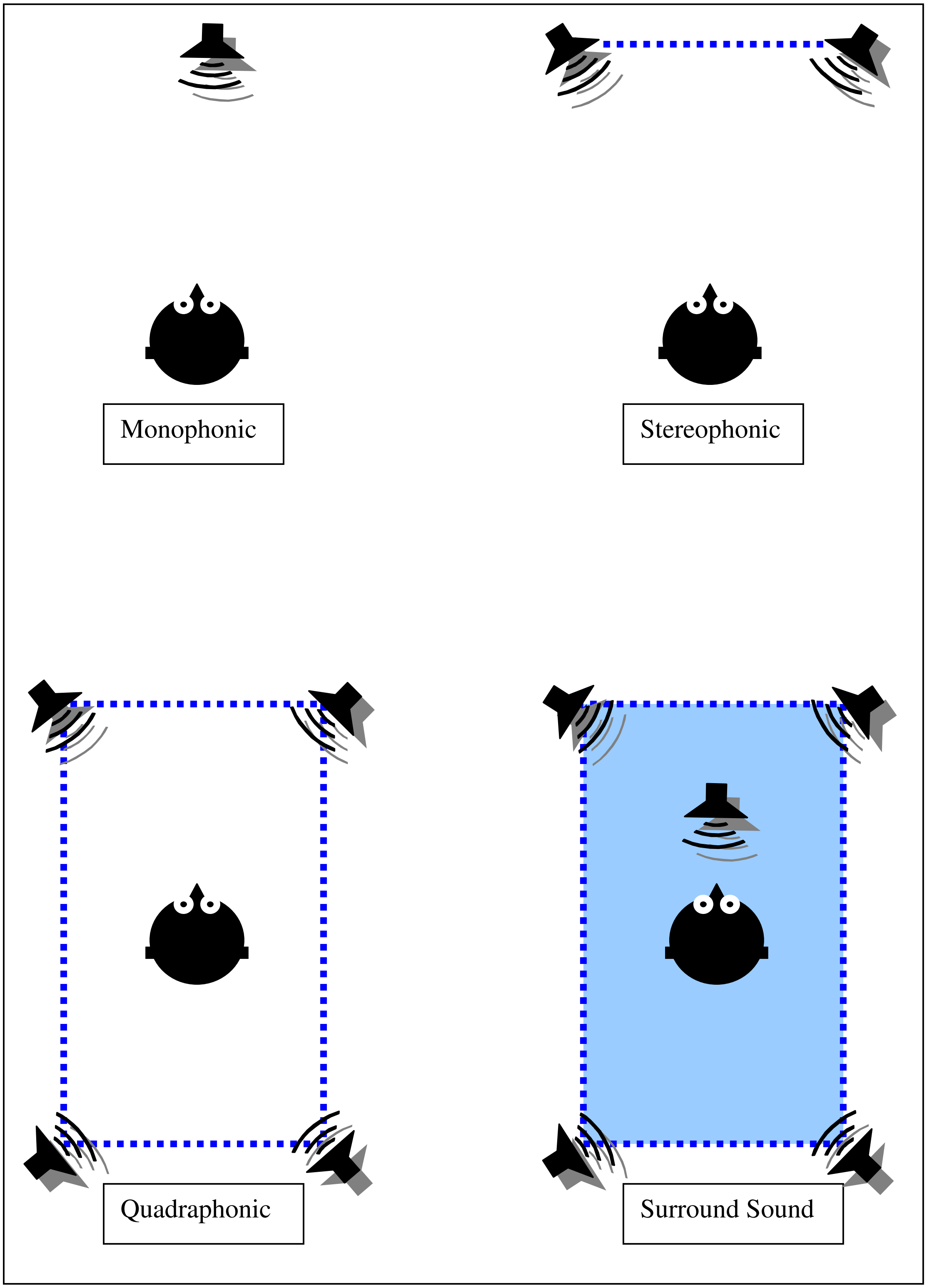

Figure A-1: Basic Audio Technologies

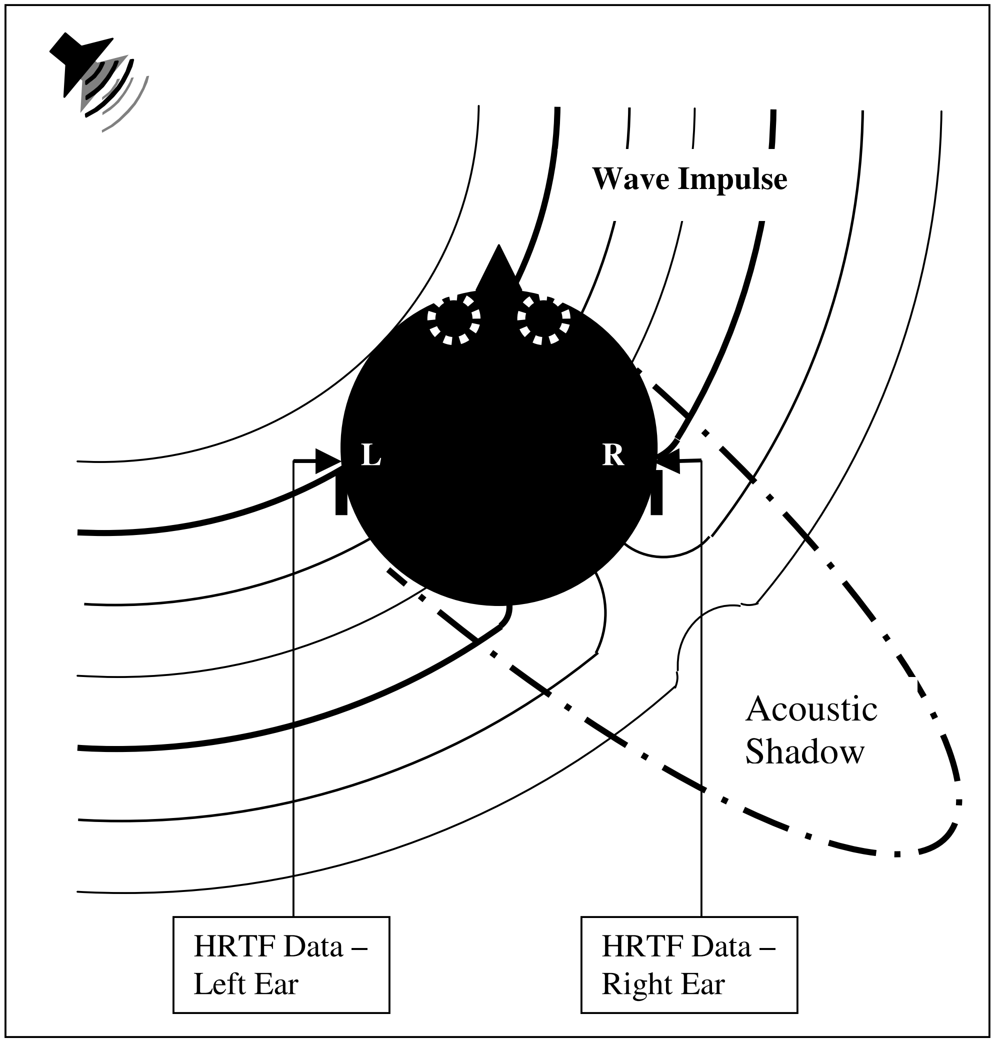

Figure A-2: Data Collection for HRTF Development[3]

Figure A-3: UTSI Piper PA 31 Navajo Research Aircraft

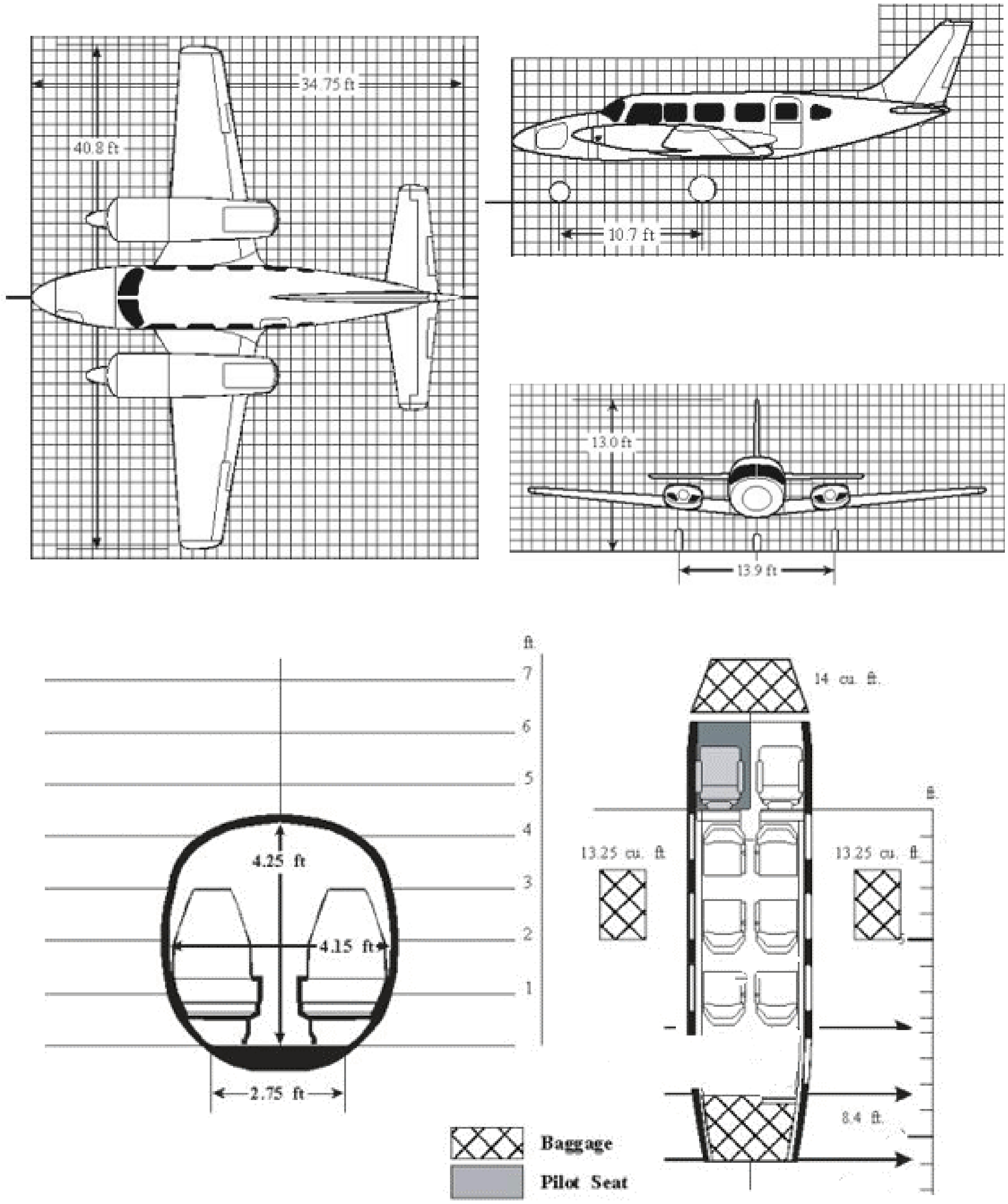

Figure A-4: Piper PA 31 Navajo 3-View Drawing and Typical Interior Dimensions

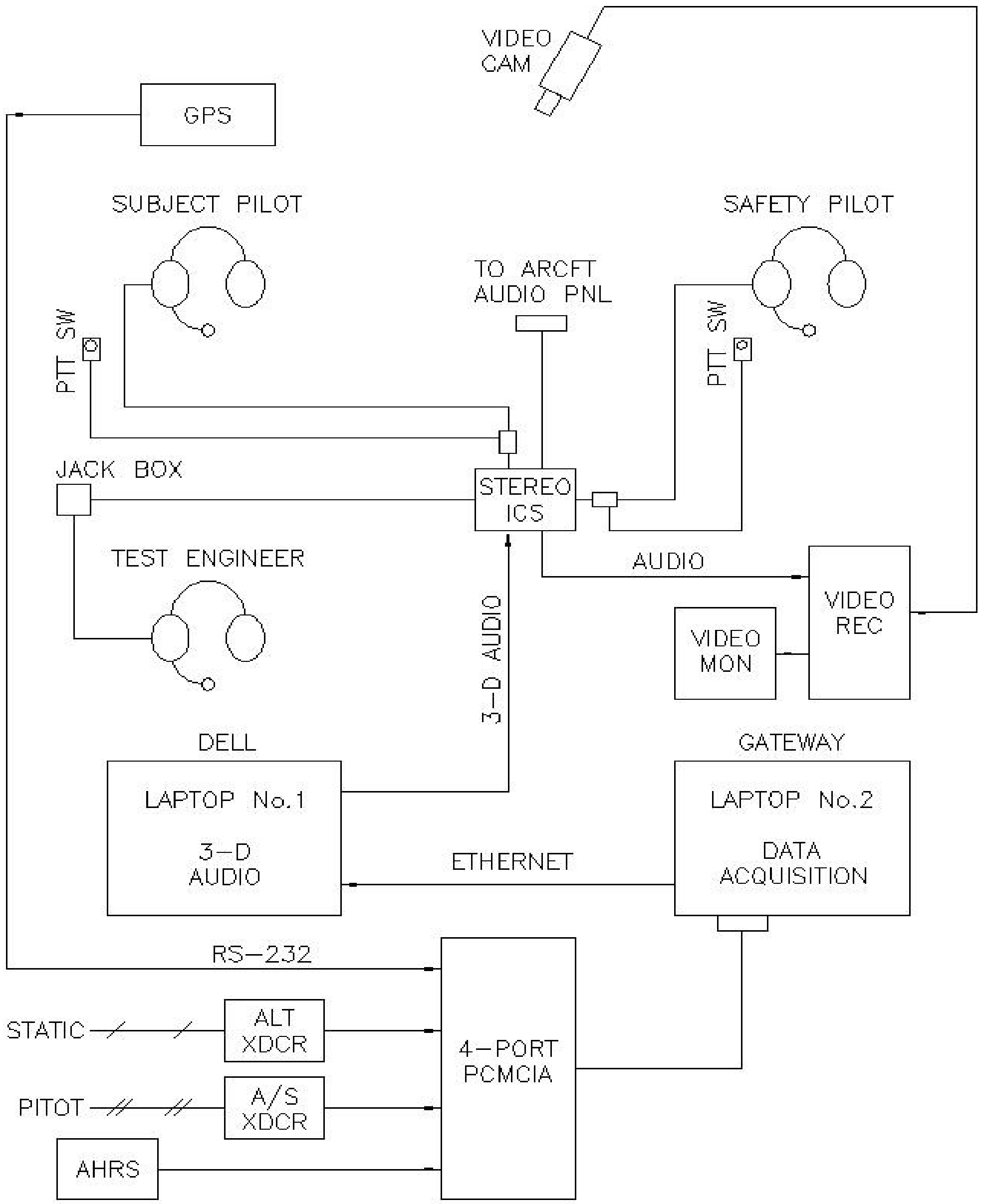

Figure A-5: Functional Depiction of 3D Audio System

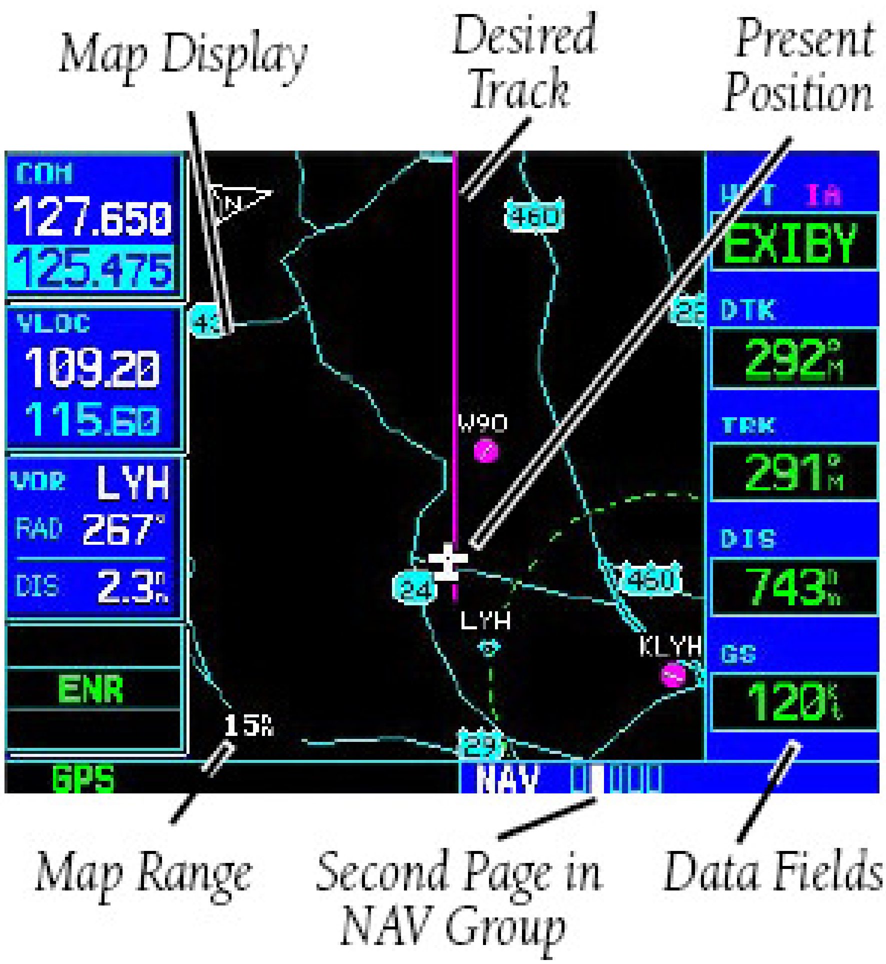

Figure A-6: GPS Map Display Format[23]

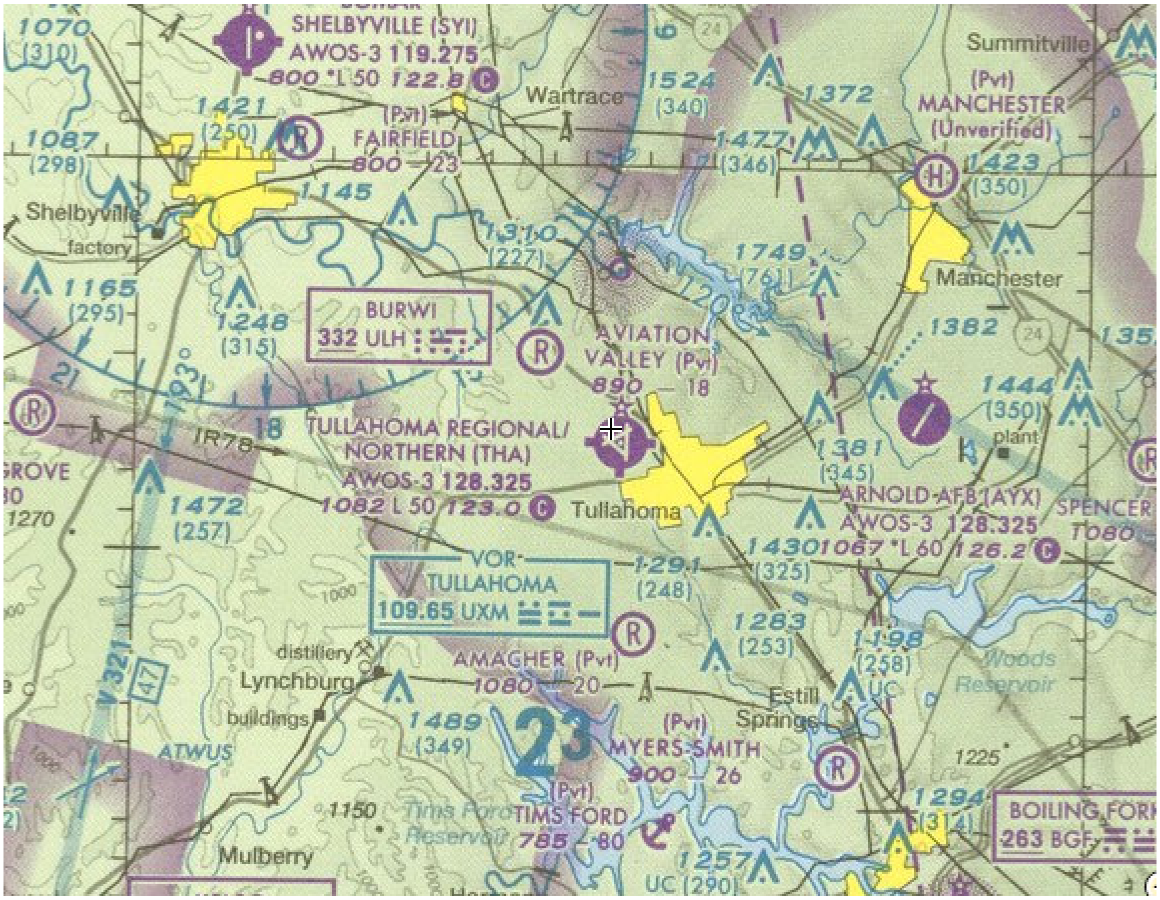

Figure A-7: Approximate Area of Flight Test Operations[24]

1	TETCO→KEYSE→LOYSI (RNAV 18)

Setup:	**Limitations:**
Alt – 4000' MSL (Local Altimeter) Trim for 120 KIAS **(FTE) 3D Audio = OFF**	Alt: 3000' MSL minimum Bank: 45 deg max Airspeed: 100 KIAS minimum

TETCO→KEYSE→LOYSI

1. (PIC) Cross TETCO heading 031°
2. (PIC) Transfer aircraft control to subject pilot
3. (PIC) **Read:** "11UT you are cleared for the RNAV 18 Approach to Tullahoma via direct KEYSE"
4. (SUB) Select RNAV 18 (KEYSE), "direct to" KEYSE
5. (PIC) **RECORDS = ON**
6. (PIC) ~6 nm from KEYSE, **State:** "Simulated left engine roughness"
7. (SUB) Completes top portion of **Card A**

KEYSE

8. (PIC) When subject initiates lead turn, **say:** "The left engine has now failed"
9. (SUB) Completes bottom portion of **Card A**
10. (PIC) When within 20 deg of direct LOYSI, **State:** "11UT, the runway is temporarily closed, expect two turns in holding, report LOYSI"
11. (SUB) Select RNAV 18 (LOYSI w/Holding), "direct to" LOYSI
12. (SUB) After EP, start **WHOLD** check (Remind to do WHOLD, if req)
    (**Weather** / **Holding** / **Obtain Clearance** / **Letdown plate** / **Approach check)**

LOYSI

13. (SUB) Announce "LOYSI, established in holding"
14. (PIC) After 30 deg of turn, **RECORDS =OFF** and take control of aircraft
15. **(SUB) TLX questionnaire**

HOLDING

16. (PIC) Return aircraft control to subject
17. (SUB) When inbound, select RNAV18 (LOYSI w/holding), "direct to" LOYSI
18. (PIC) **RECORDS = ON**
19. (PIC) **State:** "11UT report LOYSI"
20. (PIC) ~3 nm from LOYSI, **State:** "Simulated left engine failure"
21. (SUB) Completes **Card B**
22. (SUB) Over LOYSI, announce "LOYSI outbound"
23. (PIC) After 30 deg of turn, **RECORDS =OFF** and take control of aircraft
24. **(SUB) TLX questionnaire**

TASK #2 ON DECK: **HUSKU→KEYSE→IDEYA (RNAV 24)**

Figure A-8: Example Test Card

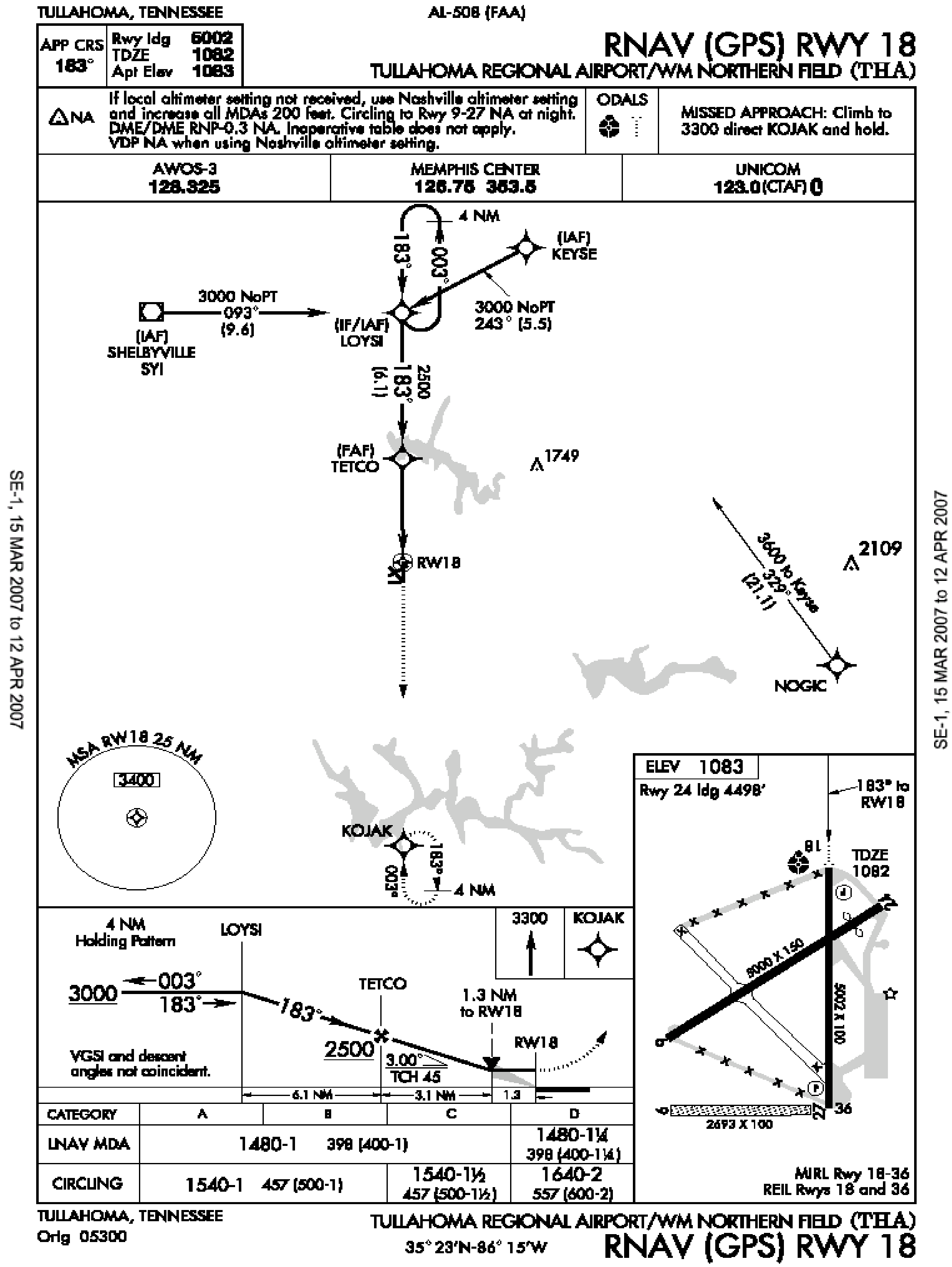

Figure A-9: Example Instrument Approach Procedure[25]

A	Emergency Procedure #1

(PIC) State: "Your left/right engine is running roughly"

Engine Roughness

A1. Emer Fuel Pumps	On
A2. Engine Instrument	Scan for cause
A3. Mixture	Adjust as required
A4. Alternate Air	OPEN
A5. Cowl Flaps	Adjust for proper CHT
A6. Fuel	Switch tanks if fuel in second tank
A7. Magnetos	Check

(PIC) State: "Your left/right engine has now failed"

Engine Failure During Flight

A8. Inop Eng	Identify
A9. Operative Eng	Adjust as req'd
A10. Airspeed	Attain and maintain at least 94 KIAS
Before securing inop engine	
A11. Fuel Flow	Check (if deficient, emer fuel pump on)
A12. Fuel Quantity	Check
A13. Fuel Selector (inop eng)	Switch to other tank containing fuel
A14. Oil Pressure and Temp	Check
A15. Magneto Switches	Check
A16. Air Start	Attempt

Airstart

A17. Fuel Selector	On
A18. Magnetos	On
A19. Throttle	Open ½ inch
A20. Propeller	½ forward
A21. Mixture	Forward
A22. Starter	Engage until prop unfeathers
A23. Propeller	Pull back to low RPM position as propeller speed accelerates through 1000 RPM
A24. Throttle	Reduce power until warm; 2000 RPM max
A25. Alternator	On
A26. Air Conditioner	(As desired) On
A27. Propeller	Manual sync with operating engine
A28. Throttle	Set as desired
A29. Prop Sync	(As desired) On

(PIC) State: "The left engine has restarted and is running satisfactorily"

Figure A-10: Emergency Procedures, Card A

B	Emergency Procedure #2

(PIC) State: "Your left/right engine has failed once again"

Engine Failure During Flight

B1. Inop Eng	Identify
B2. Operative Eng	Adjust as req'd
B3. Airspeed	Attain and maintain at least 94 KIAS
Before securing inop engine	
B4. Fuel Flow	Check (if deficient, emer fuel pump on)
B5. Fuel Quantity	Check
B6. Fuel Selector (inop eng)	Switch to other tank containing fuel
B7. Oil Pressure and Temp	Check
B8. Magneto Switches	Check
B9. Air Start	Attempt

Airstart

B10. Fuel Selector	On
B11. Magnetos	On
B12. Throttle	Open ½ inch
B13. Propeller	½ forward
B14. Mixture	Forward
B15. Starter	Engage until prop unfeathers
B16. Propeller	Pull back to low RPM position as prop speed accelerates through 1000 RPM

(Note: Engine doesn't restart. Subject should perform Eng Securing Procedure)

Engine Securing Procedure (Feathering Procedure)

B17. Throttle	Close
B18. Propeller	Feather (1000 RPM min)
B19. Mixture	Idle cut-off
B20. Cowl Flap (inop eng)	Close
B21. Air Conditioner	Off
B22. Magneto Switch	Off
B23. Emergency Fuel Pump	Off
B24. Fuel Selector	Off
B25. Fuel Boost Pump CB	Pull
B26. Alternator CB Switch	Off
B27. Prop Sync	Off
B28. Electrical Load	Reduce
B29. Crossfeed	If Req'd

Engine Failure During Flight (remainder of procedure started earlier)

B30. Power (operative eng)	As required
B31. Mixture (operative eng)	Full rich
B32. Fuel Quantity (oper eng tank)	Sufficient
B33. Emerg Fuel Pump (oper eng)	As required
B34. Cowl Flap (oper eng)	As required
B35. Trim	Adjusted (5° bank into good eng)
B36. Electrical Load	Decrease to min required

B37. Land as soon as practical at nearest suitable airport

Figure A-11: Emergency Procedures, Card B

Table A-1: Example of Output from Data Acquisition Computer

Galt (ft)	Lat (deg)	mins	Long (deg)	mins	Trc (deg)	G/S (kgs)	Dist (1/10 nm)	X/Ter (1/100 nm)	D/T_ (1/10 deg)	WP/ID	Brg (1/10 nm)	MgVr (1/10 deg)	DesDst (1/10 nm)	PAlt__ (ft)	IAS_ (KIAS)	RollAtt (deg)	PitchAtt (deg)
3964	35	3253	86	1097	19	151	43	-3	270	KEYSE	270	W032	320	3692	127.33	0.665	1.829
3961	35	3258	86	1095	20	151	42	-4	270	KEYSE	270	W032	319	3687	126.87	0.934	2.379
3959	35	3262	86	1094	20	151	42	-4	270	KEYSE	270	W032	319	3683	129.73	-0.341	2.565
3960	35	3266	86	1092	20	151	41	-5	270	KEYSE	270	W032	318	3681	128.94	0.906	2.609
3966	35	3270	86	1091	20	151	41	-5	270	KEYSE	280	W032	318	3688	127.66	2.483	2.444
3972	35	3274	86	1089	20	151	41	-6	270	KEYSE	280	W032	318	3693	128.63	2.038	2.697
3979	35	3278	86	1088	20	151	40	-7	270	KEYSE	280	W032	317	3700	128.26	1.329	2.983
3986	35	3282	86	1086	20	151	40	-7	270	KEYSE	280	W032	317	3710	128.6	0.5	2.763
3990	35	3286	86	1085	20	151	39	-7	270	KEYSE	280	W032	316	3713	127.16	-0.055	2.571

HeadAtt (deg)	RollRat (deg/sec)	PitchRat (deg/sec)	YawRat (deg/sec)	X-Accel (deg/sec^2)	Y-Accel (deg/sec^2)	Z-Accel (deg/sec^2)	X-Mag_	Y-Mag_	Z-Mag_	E/T__	Date/Time____________	Loops
20.198	1.181	3.397	-0.233	-0.029	-0.055	0.937	0.196	-0.071	0.498	1	12/11/2007 14:17	3396
18.902	-0.371	-2.051	0.769	-0.042	0.065	1.031	0.203	-0.068	0.501	2	12/11/2007 14:17	3397
19.825	0.801	-1.964	-0.041	-0.039	-0.01	1.031	0.199	-0.082	0.499	3	12/11/2007 14:17	3398
18.809	3.168	-1.666	-3.589	-0.051	0.014	1.172	0.187	-0.065	0.502	4	12/11/2007 14:17	3399
18.479	0.453	0.385	2.403	-0.041	0.058	0.984	0.197	-0.051	0.502	5	12/11/2007 14:17	3400
20.539	1.167	-1.186	0.865	-0.03	-0.002	0.937	0.186	-0.061	0.503	6	12/11/2007 14:17	3401
21.275	0.179	4.628	-0.897	-0.031	-0.034	1.031	0.195	-0.071	0.502	7	12/11/2007 14:17	3402
19.764	-0.691	1.881	-2.477	-0.024	0.035	0.844	0.197	-0.072	0.502	8	12/11/2007 14:17	3403
20.347	-0.998	2.206	1.918	-0.032	0.006	0.937	0.185	-0.079	0.501	9	12/11/2007 14:17	3404

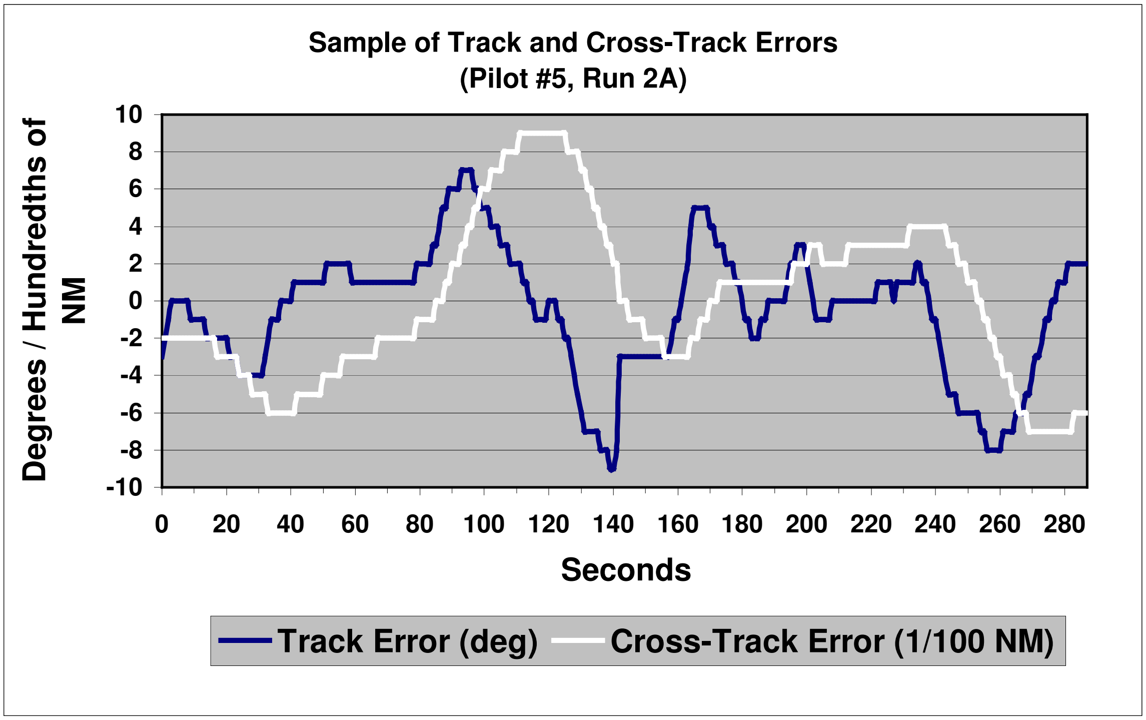

Figure A-12: Track and Cross-Track Errors During Sample Test Run

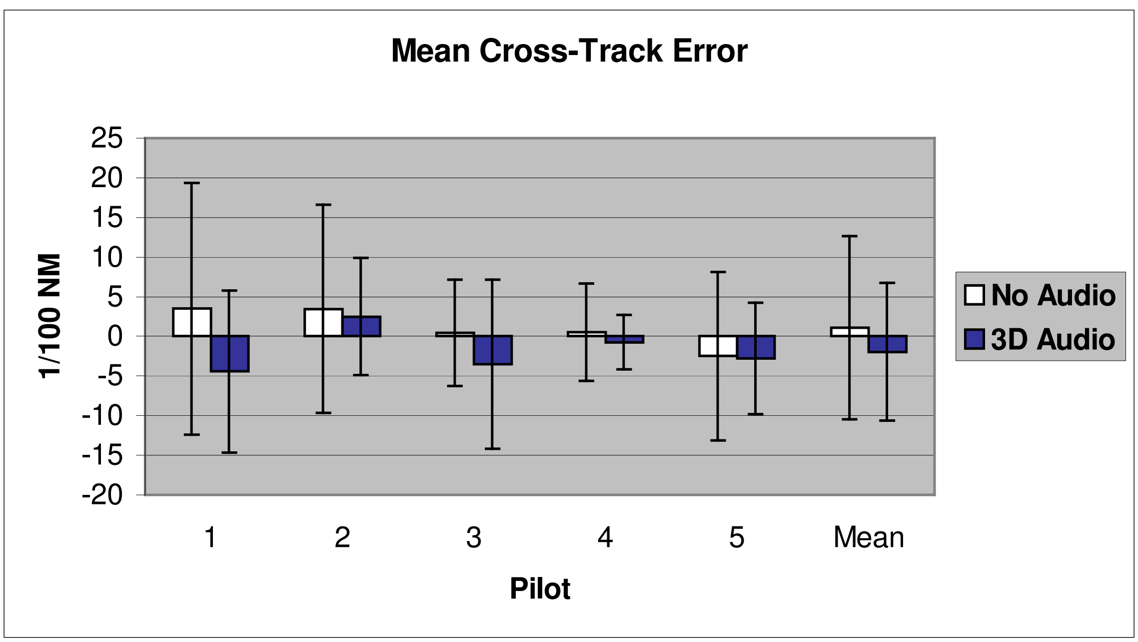

Figure A-13: Mean Cross-Track Error

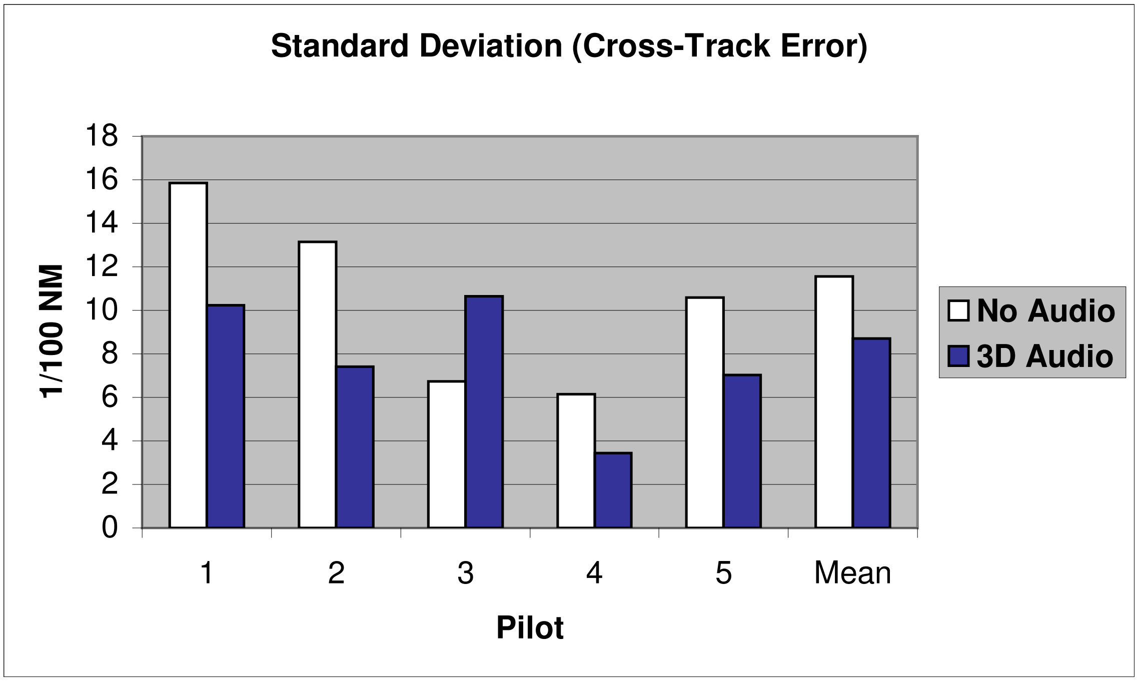

Figure A-14: Standard Deviation of Cross-Track Error

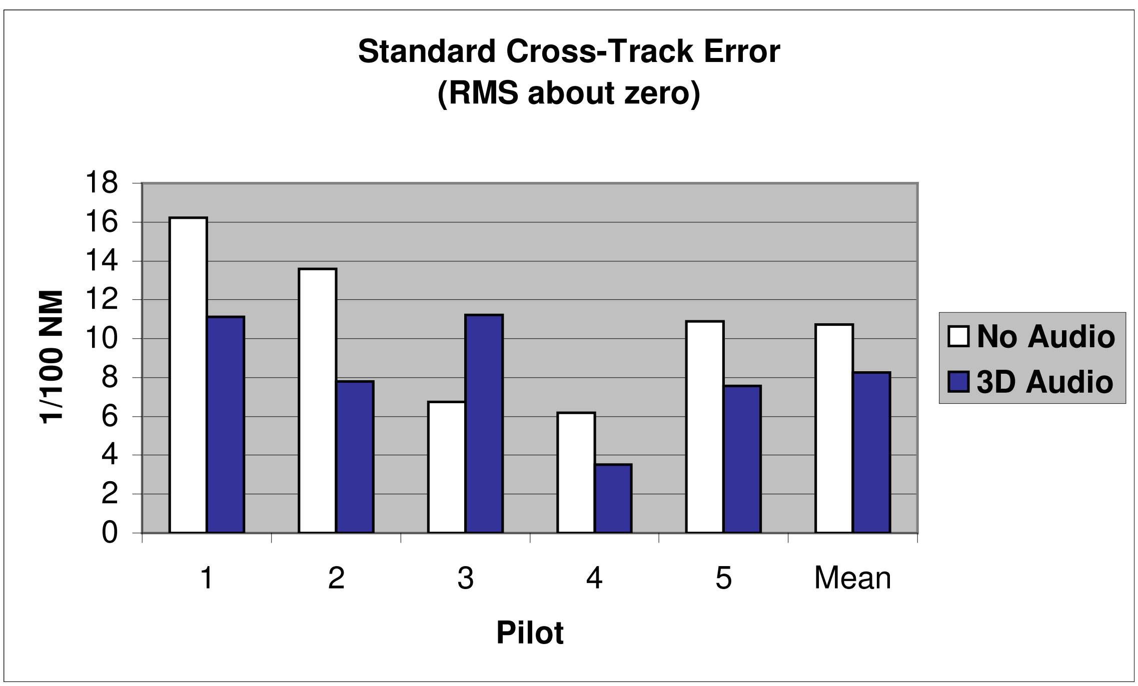

Figure A-15: Standard Cross-Track Error

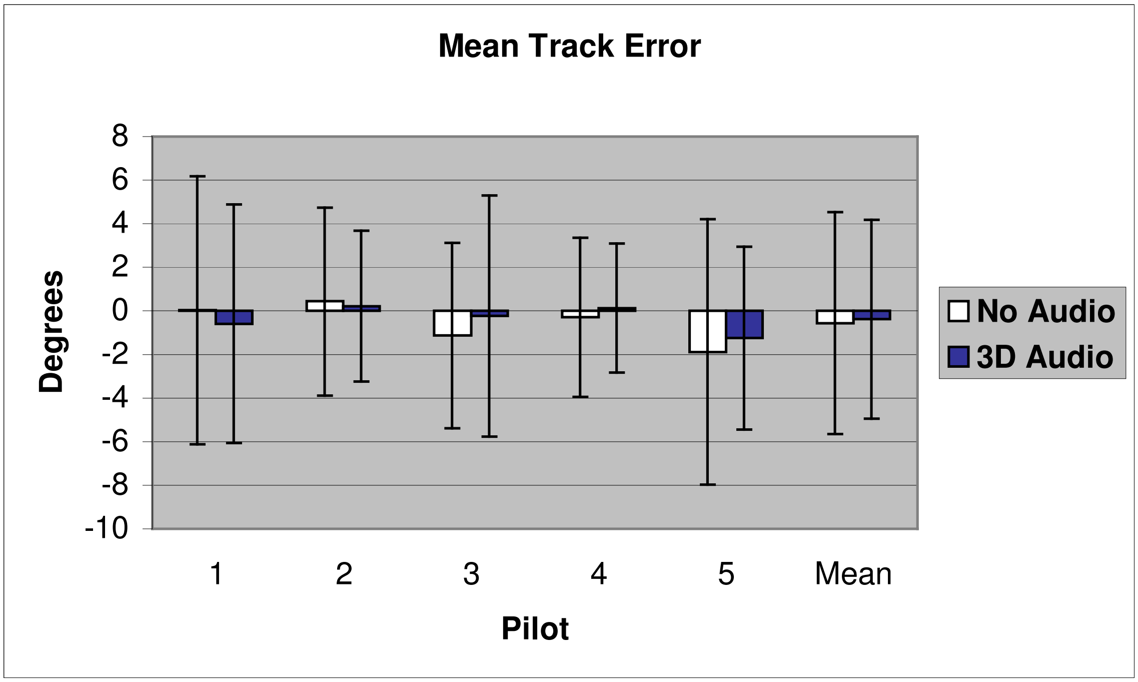

Figure A-16: Mean Track Error

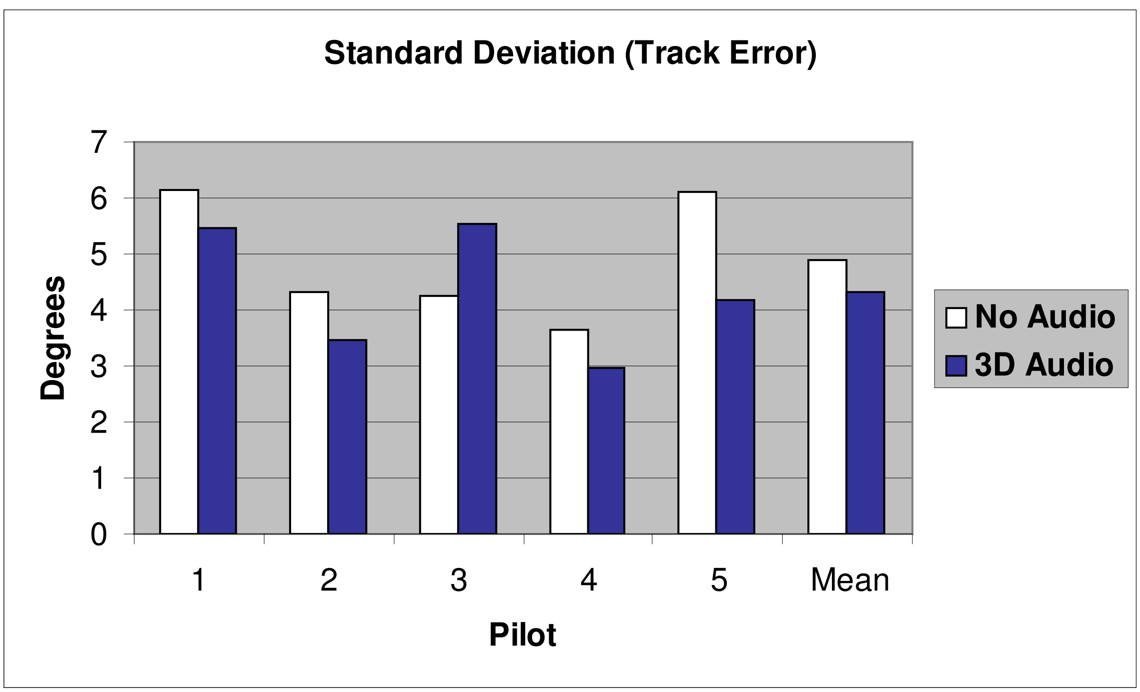

Figure A-17: Standard Deviation of Track Error

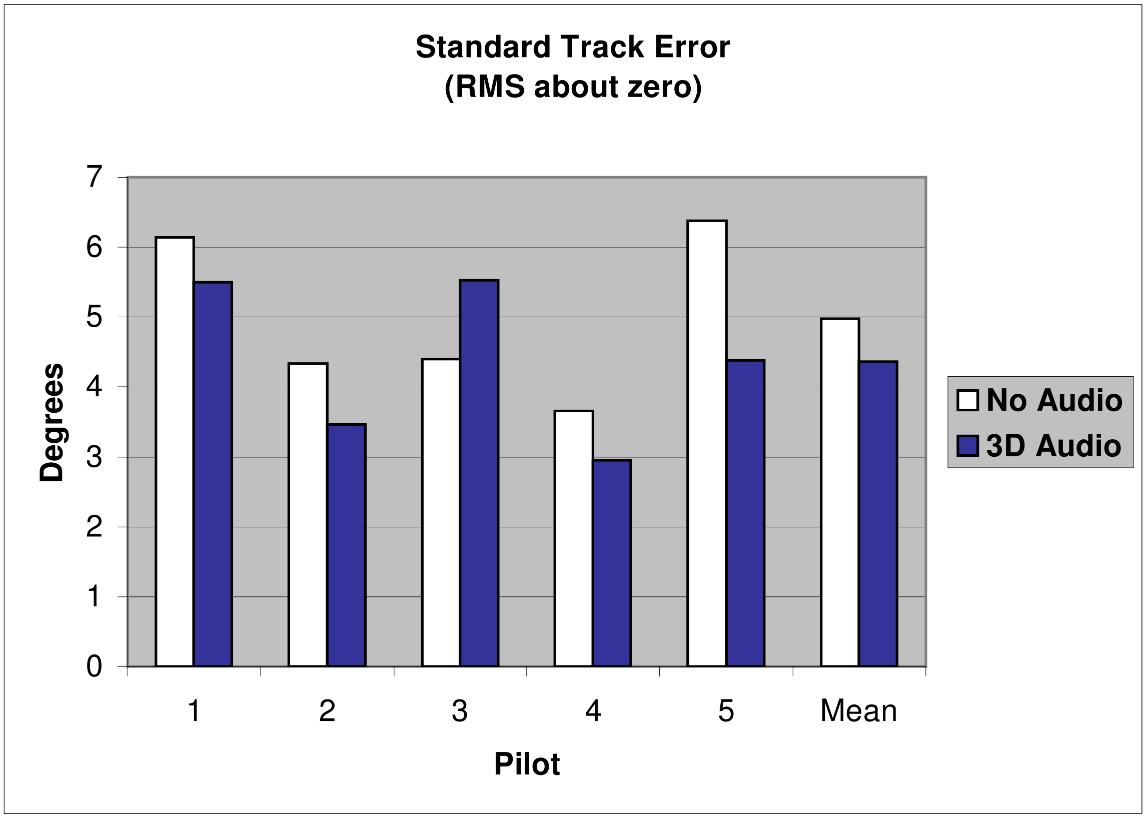

Figure A-18: Standard Track Error

Table A-2: Analysis of NASA TLX Data

	All Pilots	Pilot 1	Pilot 2	Pilot 3	Pilot 4	Pilot 5
Mean (Cue OFF)	62.14	54.67	63.67	63.11	58.00	71.28
Mean (Cue ON)	61.71	50.22	54.50	65.50	70.11	68.22
U1	13.00	23.00	28.00	13.00	9.00	25.00
U2	12.00	13.00	8.00	23.00	27.00	11.00
U_{req} for $\alpha = 0.05$	2.00	5.00	5.00	5.00	5.00	5.00
Significance?	**NO**	**NO**	**NO**	**NO**	**NO**	**NO**

APPENDIX B -- PILOT TECHNICAL DATA / RESULTS

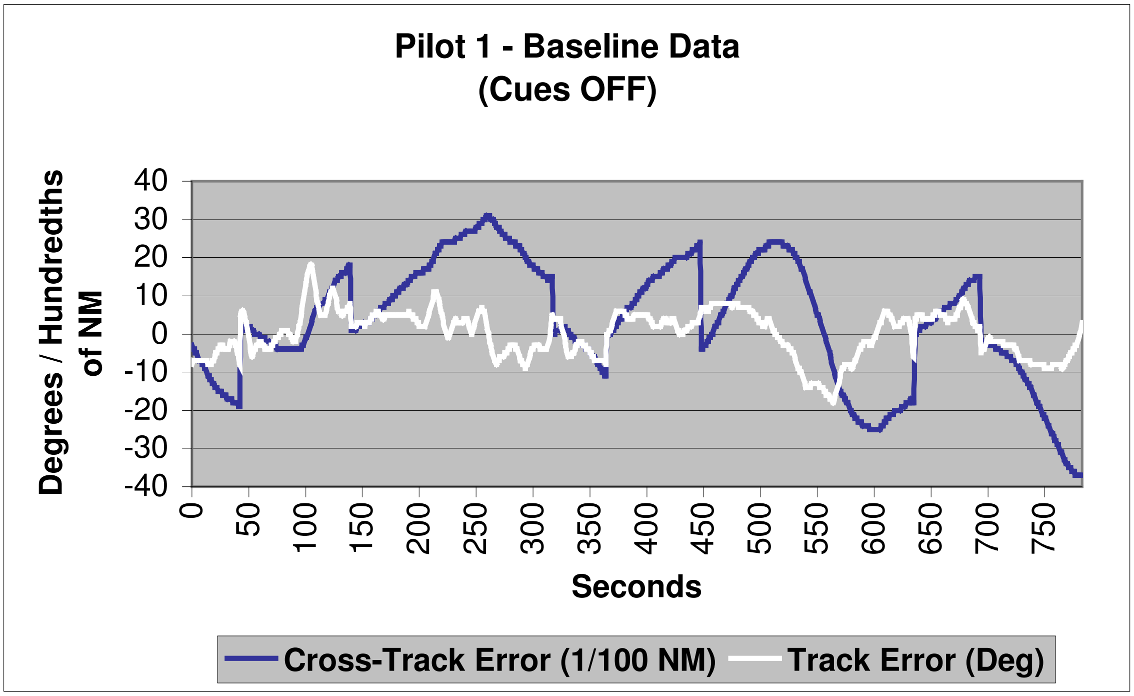

Figure B-1: Performance Errors (Pilot 1, Baseline Condition)

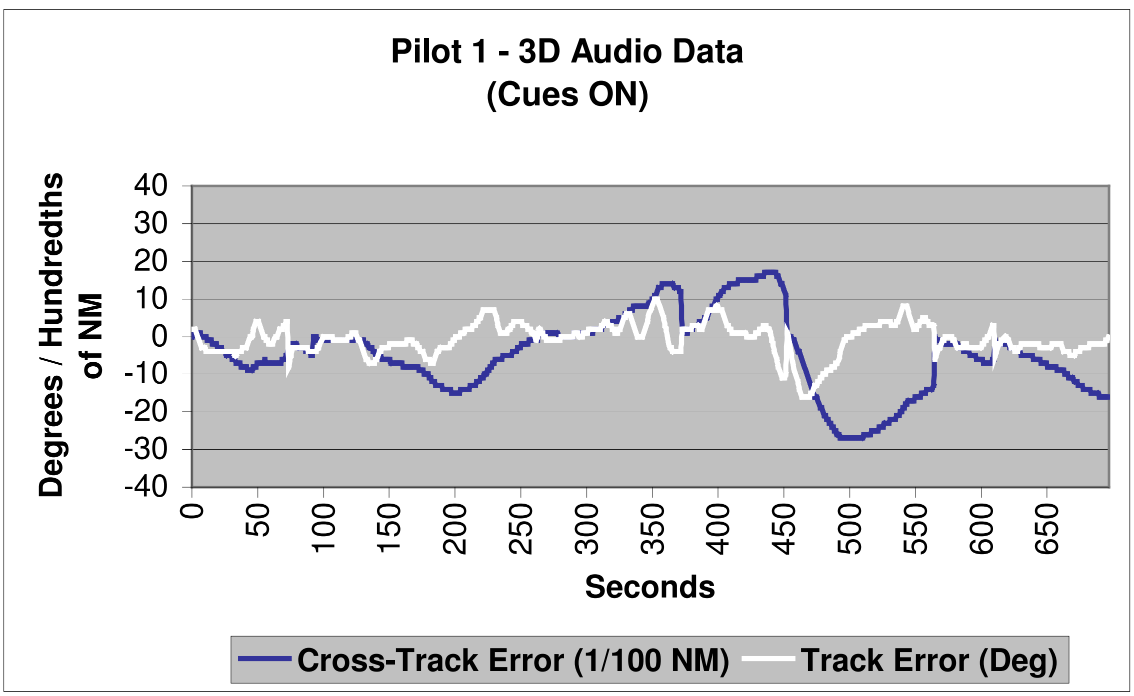

Figure B-2: Performance Errors (Pilot 1, 3D Audio Condition)

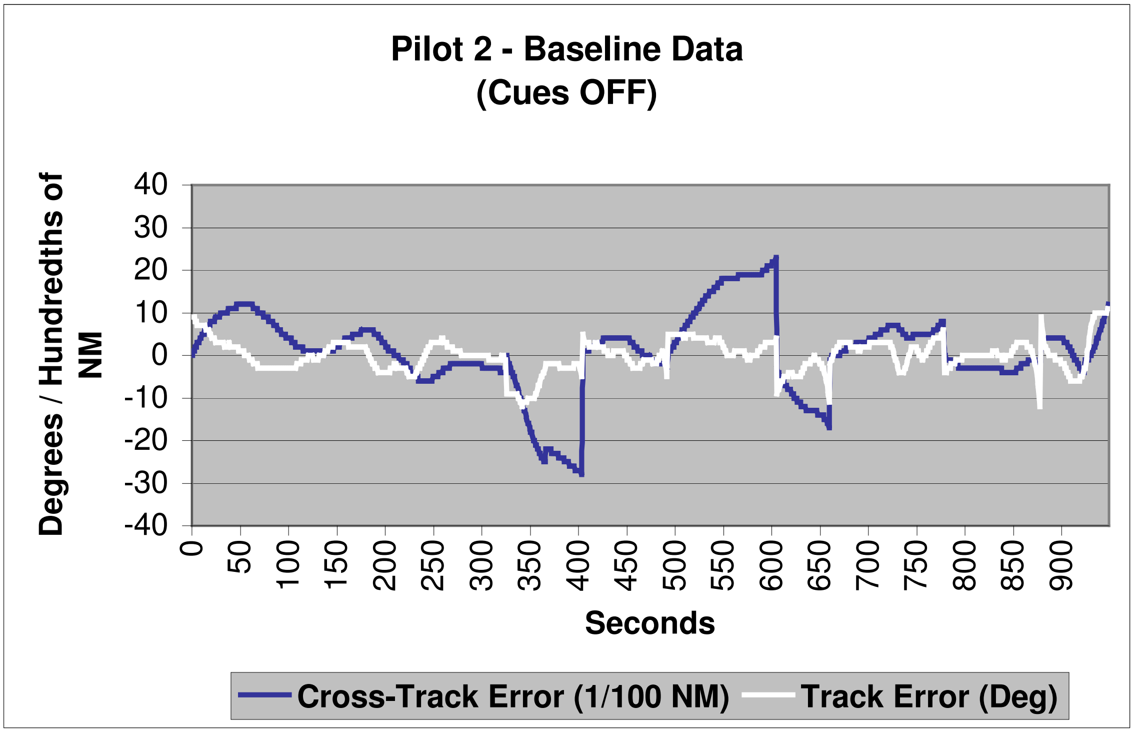

Figure B-3: Performance Errors (Pilot 2, Baseline Condition)

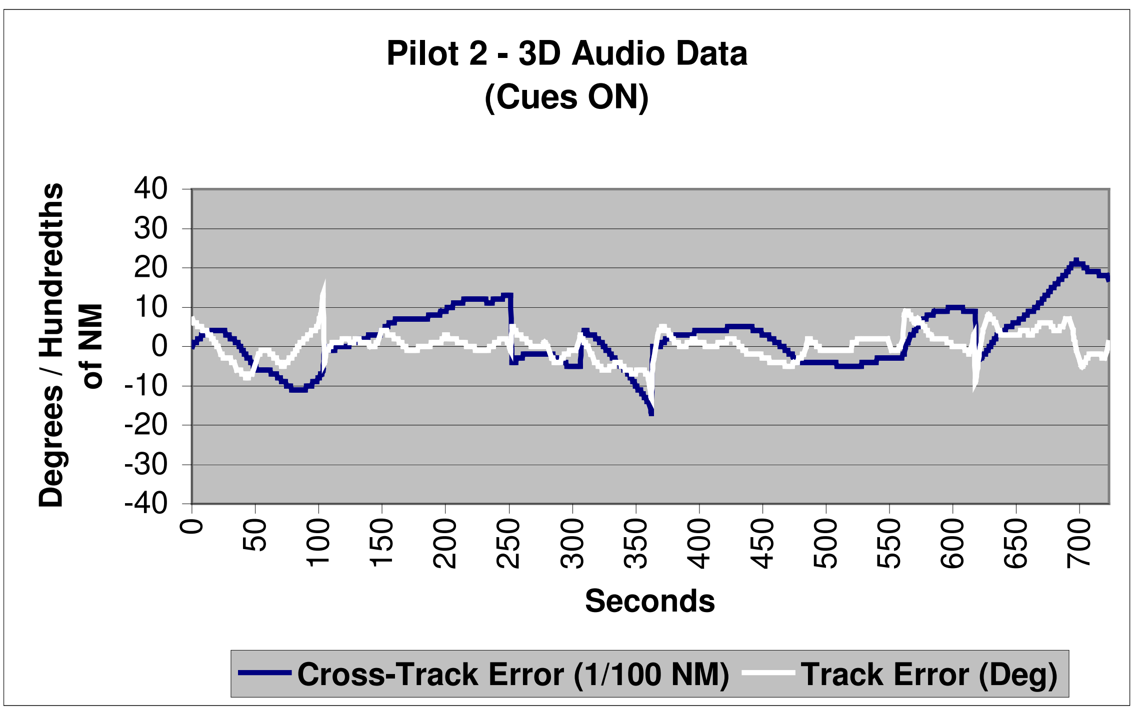

Figure B-4: Performance Errors (Pilot 2, 3D Audio Condition)

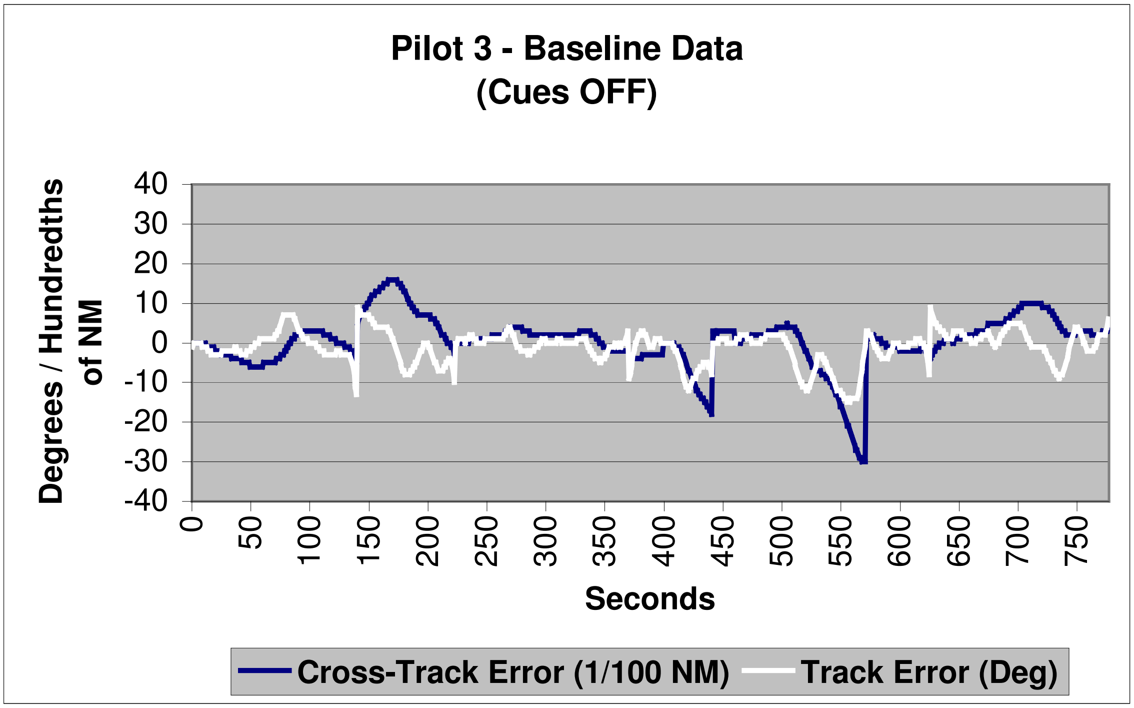

Figure B-5: Performance Errors (Pilot 3, Baseline Condition)

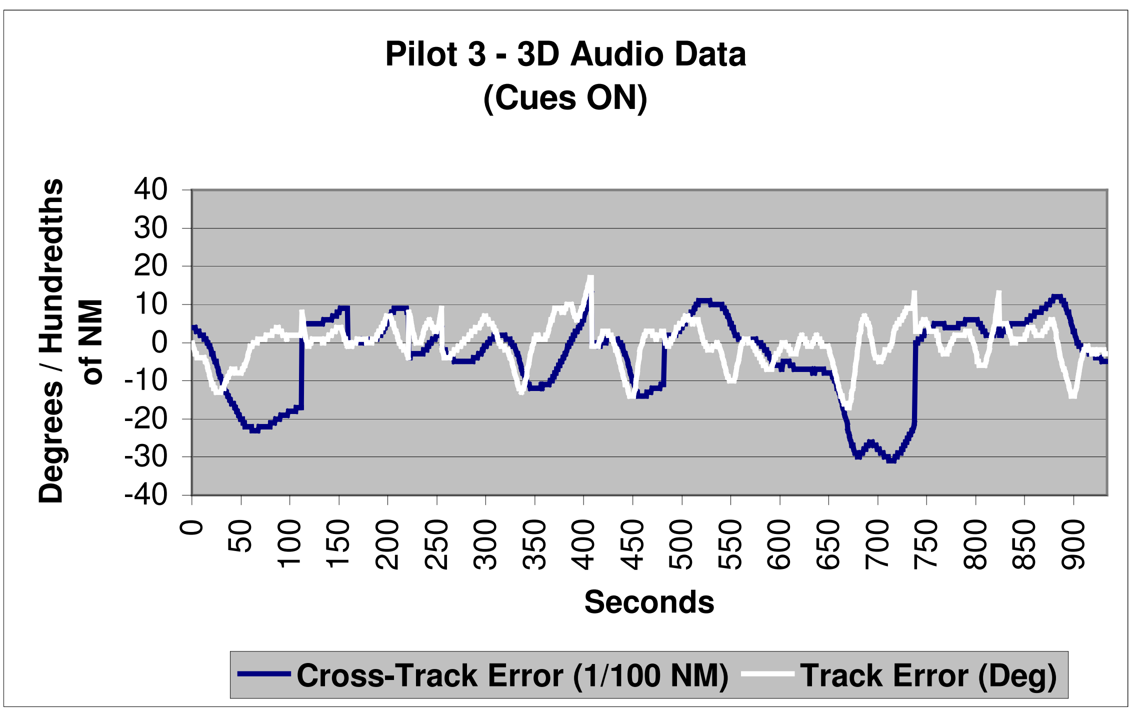

Figure B-6: Performance Errors (Pilot 3, 3D Audio Condition)

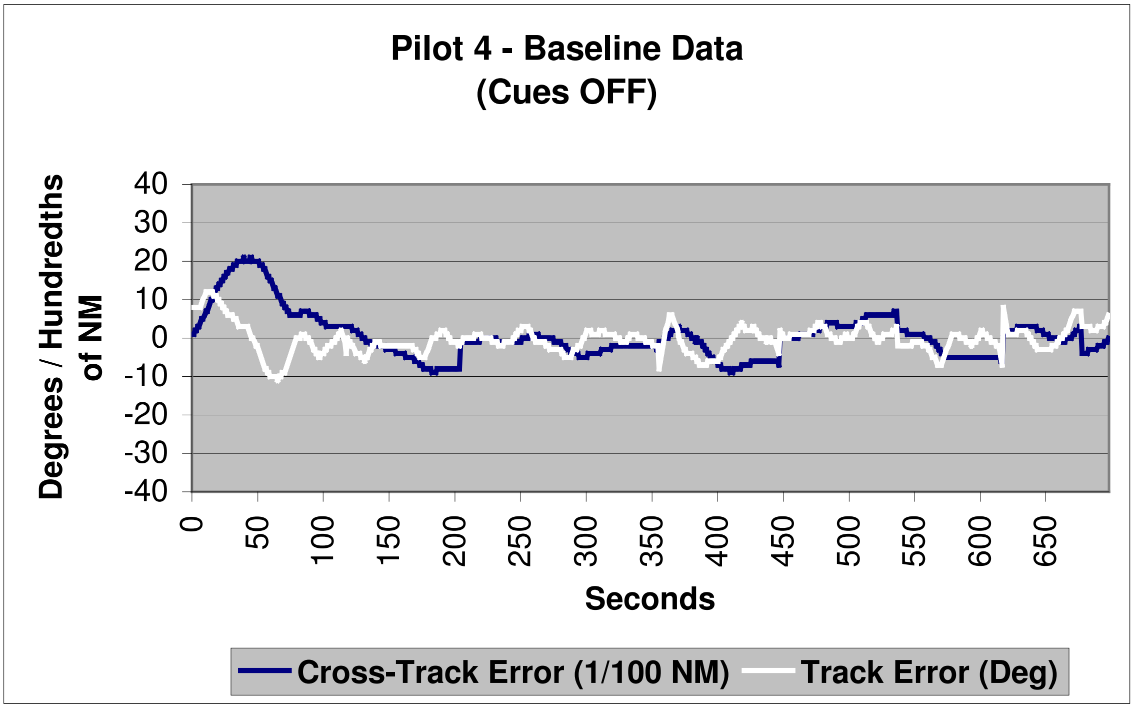

Figure B-7: Performance Errors (Pilot 4, Baseline Condition)

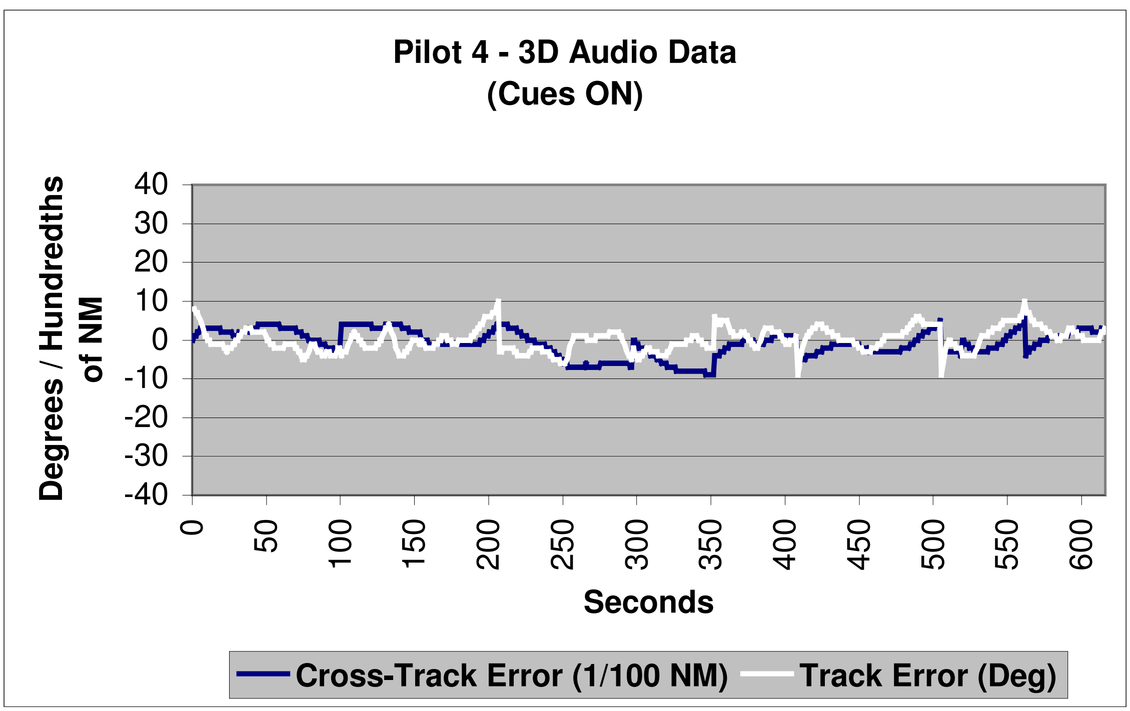

Figure B-8: Performance Errors (Pilot 4, 3D Audio Condition)

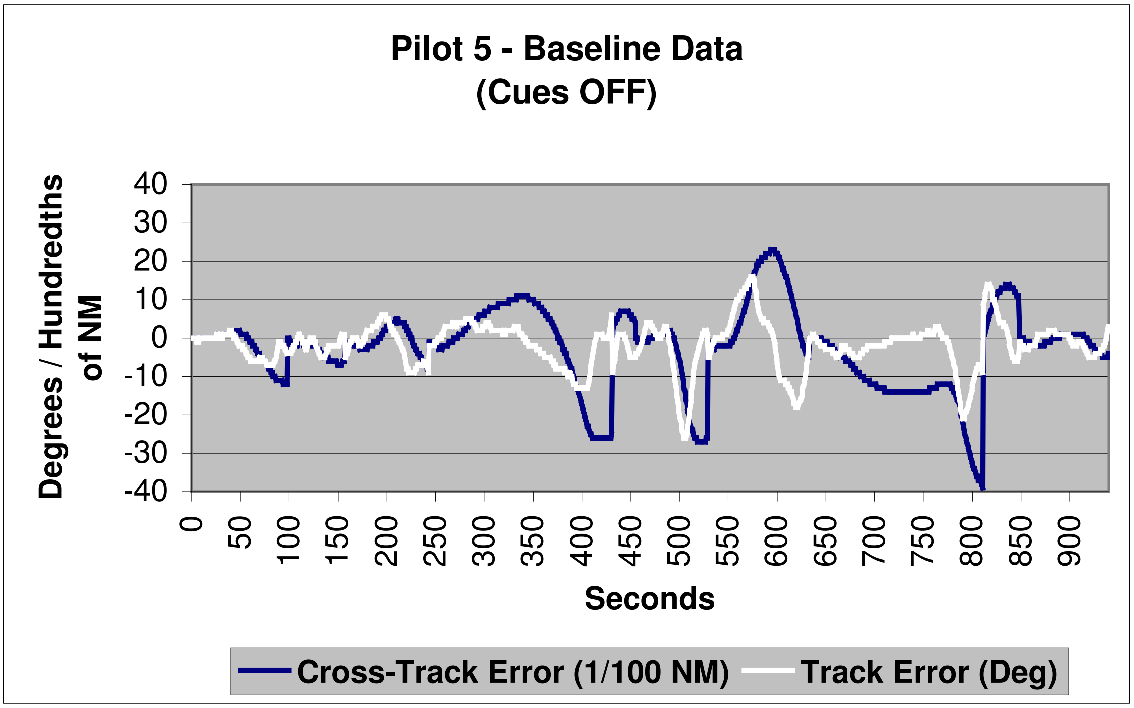

Figure B-9: Performance Errors (Pilot 5, Baseline Condition)

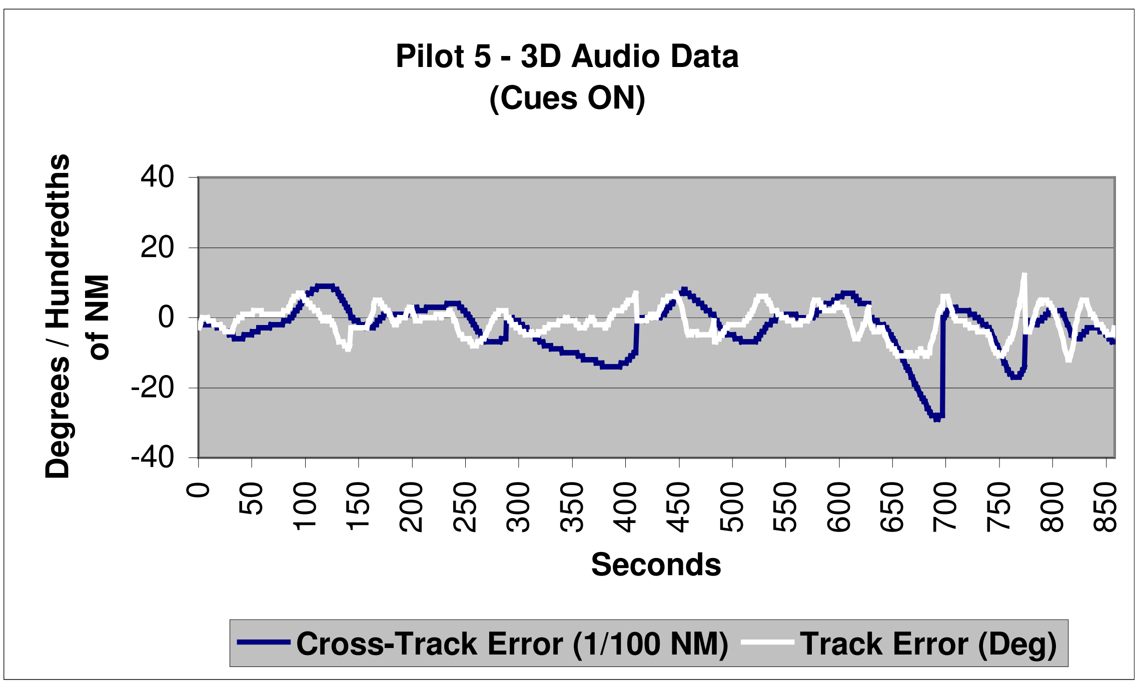

Figure B-10: Performance Errors (Pilot 5, 3D Audio Condition)

Table B-1: Cross-Track Error Results

	Mean (NM)		Std Deviation (NM)	
	Cue OFF	Cue ON	Cue OFF	Cue ON
All Pilots	**0.010933**	**-0.01948798**	**0.115491721**	**0.087049**
Pilot 1	0.034758	-0.04440459	0.158498518	0.102219
Pilot 2	0.034594	0.024861878	0.13136498	0.074097
Pilot 3	0.004389	-0.03522484	0.067359386	0.106478
Pilot 4	0.005229	-0.00756494	0.06155579	0.034319
Pilot 5	-0.02504	-0.02815851	0.105957264	0.070151

Table B-2: Track Error Results

	Mean (Deg)		Std Deviation (Deg)	
	Cue OFF	Cue ON	Cue OFF	Cue ON
All Pilots	**-0.56111**	**-0.38505747**	**5.100639247**	**4.554304**
Pilot 1	0.038265	-0.58249641	6.142294391	5.463766
Pilot 2	0.434018	0.208852006	4.310059251	3.454385
Pilot 3	-1.11969	-0.24089936	4.248387918	5.521351
Pilot 4	-0.29656	0.121753247	3.642490309	2.952157
Pilot 5	-1.87872	-1.24592075	6.089339702	4.195779

Table B-3: Turn Procedure Results

DID PILOT TURN IN CORRECT DIRECTION?					
	Pilot 1	**Pilot 2**	**Pilot 3**	**Pilot 4**	**Pilot 5**
1A	YES	YES	YES	YES	YES
1B	YES	YES	YES	YES	YES
2A	YES	YES	YES	YES	YES
2B	YES	YES	YES	YES	YES
3A	YES	YES	YES	YES	YES
3B	YES	YES	YES	YES	YES
4A	YES	YES	YES	NO TURN	YES
4B	YES	YES	YES	YES	YES
5A	YES	YES	YES	NO TURN	YES
5B	YES	YES	YES	YES	YES
6A	YES	YES	YES	YES	YES
6B	YES	YES	YES	YES	YES

APPENDIX C – NASA TLX QUESTIONNAIRE / RESULTS

1.1	NASA TLX RATING SCALE
TASK ID:______________________	SUBJECT ID:___________________

Place a mark on each scale that represents the magnitude of each factor.

Mental Demand

How much mental and perceptual activity was required?
(thinking, deciding, calculating, remembering, looking)

LOW **HIGH**

Physical Demand

How much physical activity was required?
(pushing, pulling, turning, controlling, activating, etc)

LOW **HIGH**

Temporal Demand

How much time pressure did you feel ?
Was the pace slow and leisurely or rapid and frantic?

LOW **HIGH**

Performance

How successful do you think you were?
How satisfied were you with your performance?

EXCEL **POOR**

Effort

How hard did you have to work mentally or physically to accomplish your level of performance?

LOW **HIGH**

Frustration Level

Low = secure, gratified, content, relaxed and complacent
High = insecure, discouraged, irritated, and stressed

LOW **HIGH**

Figure C-1: NASA TLX Sheet 1

1.2	NASA TLX RATING SCALE
Circle the factor of each pair that was more significant source of workload variation.	

1. Mental Demand / Physical Demand
2. Mental Demand / Temporal Demand
3. Mental Demand / Performance
4. Mental Demand / Effort
5. Mental Demand / Frustration

6. Physical Demand / Temporal Demand
7. Physical Demand / Performance
8. Physical Demand / Effort
9. Physical Demand / Frustration

10. Temporal Demand / Performance
11. Temporal Demand / Effort
12. Temporal Demand / Frustration

13. Performance / Effort
14. Performance / Frustration

15. Effort / Frustration

Figure C-2: NASA TLX Sheet 2

Table C-1: NASA TLX Results (Overall)

	ALL PILOTS		PILOT 1		PILOT 2		PILOT 3		PILOT 4		PILOT 5	
	Cue OFF	Cue ON	Cue OFF	Cue ON	Cue OFF	Cue ON	Cue OFF	Cue ON	Cue OFF	Cue ON	Cue OFF	Cue ON
RUN 1	62.67	36.33	62.67	36.33	58.33	43.67	55.33	59.33	65.00	82.33	68.67	65.00
RUN 2	52.00	53.33	52.00	53.33	58.67	50.00	63.67	69.00	55.00	62.33	63.67	72.67
RUN 3	54.33	50.33	54.33	50.33	53.67	51.67	59.67	56.67	40.33	64.33	75.00	71.67
RUN 4	56.00	57.67	56.00	57.67	64.00	59.00	66.00	67.00	50.67	60.00	77.67	72.00
RUN 5	51.67	47.33	51.67	47.33	72.00	66.33	70.33	69.00	68.00	80.33	74.00	65.67
RUN 6	51.33	56.33	51.33	56.33	75.33	56.33	63.67	72.00	69.00	71.33	68.67	62.33
RUN 7	58.33	43.67										
RUN 8	58.67	50.00										
RUN 9	53.67	51.67										
RUN 10	64.00	59.00										
RUN 11	72.00	66.33										
RUN 12	75.33	56.33										
RUN 13	55.33	59.33										
RUN 14	63.67	69.00										
RUN 15	59.67	56.67										
RUN 16	66.00	67.00										
RUN 17	70.33	69.00										
RUN 18	63.67	72.00										
RUN 19	65.00	82.33										
RUN 20	55.00	62.33										
RUN 21	40.33	64.33										
RUN 22	50.67	60.00										
RUN 23	68.00	80.33										
RUN 24	69.00	71.33										
RUN 25	68.67	65.00										
RUN 26	63.67	72.67										
RUN 27	75.00	71.67										
RUN 28	77.67	72.00										
RUN 29	74.00	65.67										
RUN 30	68.67	62.33										

Analysis of Results

	All Pilots		Pilot 1	Pilot 2	Pilot 3	Pilot 4	Pilot 5
Mean (Cue OFF)	62.14		54.67	63.67	63.11	58.00	71.28
Mean (Cue ON)	61.71		50.22	54.50	65.50	70.11	68.22
U1	452.00	13.00	23.00	28.00	13.00	9.00	25.00
U2	448.00	12.00	13.00	8.00	23.00	27.00	11.00
U for $\alpha = 0.05$	317.00	2.00	5.00	5.00	5.00	5.00	5.00
Significance?	NO		NO	NO	NO	NO	NO

Table C-2: NASA TLX Results (Mental Demand)

Mental Demand																							
Rating	**Low (-10)**	**-9**	**-8**	**-7**	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**High (+10)**	**U-Test (2-tail, α=.05)**	
Pilot 1																						Significance?	**NO**
Cue Off										1	4	1										U1	23.5
Cue On							1			2	2	1										U2	12.5
Pilot 2																						Significance?	**YES**
Cue Off										1			2	1	2							U1	36
Cue On					1		4		1													U2	0
Pilot 3																						Significance?	**NO**
Cue Off				1						1	1	3										U1	9
Cue On										1		3		2								U2	27
Pilot 4																						Significance?	**NO**
Cue Off									1				2	1	2							U1	10
Cue On												1		2		1	1	1				U2	26
Pilot 5																						Significance?	**NO**
Cue Off															1	1	2	2				U1	20.5
Cue On															1		5					U2	15.5

Table C-3: NASA TLX Results (Physical Demand)

Physical Demand																							
Rating	**Low (-10)**	**-9**	**-8**	**-7**	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**High (+10)**	**U-Test (2-tail, α=.05)**	
Pilot 1																						Significance?	**NO**
Cue Off									1	1	2		2									U1	25.5
Cue On								1	1	2	1	1										U2	10.5
Pilot 2																						Significance?	**NO**
Cue Off													2		2	1	1					U1	18.5
Cue On											1		1	1	1		1	1				U2	17.5
Pilot 3																						Significance?	**NO**
Cue Off												2	4									U1	19
Cue On											1	1	4									U2	17
Pilot 4																						Significance?	**NO**
Cue Off						2	2	1	1													U1	9
Cue On						1		2	2			1										U2	27
Pilot 5																						Significance?	**NO**
Cue Off													1	1		3		1				U1	13
Cue On															1	3	1	1				U2	23

Table C-4: NASA TLX Results (Temporal Demand)

Temporal Demand																							
Rating	**Low (-10)**	**-9**	**-8**	**-7**	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**High (+10)**	**U-Test (2-tail, α=.05)**	
Pilot 1																						Significance?	**NO**
Cue Off									2	1		2	1									U1	22.5
Cue On						1		2				2	1									U2	13.5
Pilot 2																						Significance?	**NO**
Cue Off									1				2		2		1					U1	13.5
Cue On											1		1		2		1	1				U2	22.5
Pilot 3																						Significance?	**NO**
Cue Off													2		1	3						U1	23
Cue On											1		1		3	1						U2	13
90Pilot 4																						Significance?	**NO**
Cue Off								1	1			1	1		2							U1	7
Cue On														3	1		1		1			U2	29
Pilot 5																						Significance?	**NO**
Cue Off															1	1	4					U1	6
Cue On																	3	3				U2	30

Table C-5: NASA TLX Results (Performance)

Performance																							
Rating	**Excel (-10)**	**-9**	**-8**	**-7**	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Poor (+10)**	**U-Test (2-tail, α=.05)**	
Pilot 1																						Significance?	**NO**
Cue Off									1	3			1	1								U1	24.5
Cue On								1	2	1	1	1										U2	11.5
Pilot 2																						Significance?	**YES**
Cue Off					1	1	1		1	1					1							U1	35
Cue On			2	2	2																	U2	1
Pilot 3																						Significance?	**NO**
Cue Off						3	1		1	1												U1	17.5
Cue On			1			1	1	2		1												U2	18.5
Pilot 4																						Significance?	**NO**
Cue Off								1	2	1	1		1									U1	9
Cue On									1	1	1		2	1								U2	27
Pilot 5																						Significance?	**YES**
Cue Off							3		2			1										U1	31.5
Cue On					1	3	1	1														U2	4.5

Table C-6: NASA TLX Results (Effort)

Effort																							
Rating	**Low (-10)**	**-9**	**-8**	**-7**	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**High (+10)**	**U-Test (2-tail, α=.05)**	
Pilot 1																						Significance?	**NO**
Cue Off									1	3	1	1										U1	18.5
Cue On							1			3	1	1										U2	17.5
Pilot 2																						Significance?	**NO**
Cue Off											2		2	1			1					U1	20.5
Cue On					1				1				2	1	1							U2	15.5
Pilot 3																						Significance?	**NO**
Cue Off												2	2	1		1						U1	9.5
Cue On													2	1	1	2						U2	26.5
Pilot 4																						Significance?	**NO**
Cue Off									1				2			2	1					U1	9.5
Cue On														3			1	1	1			U2	26.5
Pilot 5																						Significance?	**YES**
Cue Off															1	1	3	1				U1	34
Cue On												1	1	2	1	1						U2	2

Table C-7: NASA TLX Results (Frustration)

Frustration																							
Rating	**Low (-10)**	**-9**	**-8**	**-7**	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**High (+10)**	**U-Test (2-tail, α=.05)**	
Pilot 1																						Significance?	**NO**
Cue Off					1		1	1	3													U1	12.5
Cue On				1				1	2	2												U2	23.5
Pilot 2																						Significance?	**YES**
Cue Off							1		1		3		1									U1	32
Cue On					3		1		1	1												U2	4
Pilot 3																						Significance?	**NO**
Cue Off				1	1		1				1	2										U1	19
Cue On							2		3	1												U2	17
Pilot 4																						Significance?	**NO**
Cue Off									3				2		1							U1	11
Cue On									1		1	1			1	1	1					U2	25
Pilot 5																						Significance?	**NO**
Cue Off									1			2	2	1								U1	20.5
Cue On								1	1	1			2		1							U2	15.5

APPENDIX D – POST-FLIGHT QUESTIONNAIRE / RESULTS

	Implementation of the 3D audio system					
Question 1	If 3D audio cues were the only directional cues available, how precisely could you locate the direction of the "go to" waypoint?	Very imprecisely	Somewhat imprecisely	Neither precisely nor imprecisely	Somewhat precisely	Very precisely
		1			4	
Question 2	Indicate your level of agreement/disagreement with the following statement: "The use of white noise by the 3D audio cue allowed directional information to be correctly conveyed to the pilot."	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
		1			3	1
Question 3	Indicate your level of agreement/disagreement with the following statement: "The 3D audio cue was distracting or annoying."	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
		1	4			
	3D audio system and navigational performance					
Question 4	When performing emergency procedures <u>with</u> 3D audio cueing (vs. <u>without</u> cueing), my navigational performance was:	Much worse	Somewhat worse	About the same	Somewhat better	Much better
					4	1
Question 5	How much better or worse do you believe you maintained the desired ground track when <u>using</u> 3D audio cueing (vs. <u>without</u> cueing)?	Much worse	Somewhat worse	About the same	Somewhat better	Much better
				2	2	1

Figure D-1: Post-Flight Questionnaire (Sheet 1 of 6)

Question 6	Upon arriving at turnpoints, how much more or less timely were your turns when <u>using</u> 3D audio cueing (vs. <u>without</u> cueing)?	Much less timely	Somewhat less timely	Neither more nor less timely	Somewhat more timely	Much more timely
				2	2	1
	Use of an airframe-referenced system					
Question 7	Indicate your level of agreement/disagreement with the following statement: "After a reasonable period of use, I was able to adapt to and use an airframe-referenced system of 3D audio cues."	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
					3	2
Question 8	When looking forward at the instrument panel and using the 3D audio cue, how intuitive or counterintuitive was it for you to determine the direction of the selected waypoint?	Very counterintuitive	Somewhat counterintuitive	Neither intuitive nor counterintuitive	Somewhat intuitive	Very intuitive
					2	3
Question 9	When looking to the sides (head turned more than 30 degrees from forward) and using the 3D audio cue, how intuitive or counterintuitive was it for you to determine the direction of the selected waypoint?	Very counterintuitive	Somewhat counterintuitive	Neither intuitive nor counterintuitive	Somewhat intuitive	Very intuitive
			1	1	3	

Figure D-2: Post-Flight Questionnaire (Sheet 2 of 6)

Question 10	When looking forward, how often did the direction of the 3D audio cue seem to match the direction shown on the GPS display	Never	Infrequently	Some of the time	Frequently	Always
					4	1
Question 11	How often did you find yourself rotating your head left and right in an attempt to better locate the 3D audio cue?	Never	Infrequently	Some of the time	Frequently	Always
		3	2			
	3D audio system and workload					
Question 12	Overall, did the use of 3D audio cueing raise or lower your workload? With 3D audio cueing, my workload was:	Much higher	Somewhat higher	Neither higher nor lower	Somewhat lower	Much lower
				2	2	1
Question 13	Indicate your level of agreement with the following statement: "I had to concentrate (focus) on the 3D audio cues to determine the location of the selected waypoint."	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
			4		1	
Question 14	Indicate your level of agreement/disagreement with the following statement: "When 3D audio cues were present as I was performing required navigation tasks, I felt I had more time available for other matters such as monitoring aircraft systems and performing emergency checklists."	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
			1	1	1	2

Figure D-3: Post-Flight Questionnaire (Sheet 3 of 6)

Question 15	Was more or less mental effort required to complete flight tasks when <u>using</u> 3D audio cueing (vs <u>without</u> cueing)?	Much less mental effort	Somewhat less mental effort	Neither more nore less mental effort	Somewhat more mental effort	Much more mental effort
			4		1	
Question 16	Was it easier or more difficult to perform emergency procedures when <u>using</u> 3D audio cueing (vs <u>without</u> cueing)?	Much more difficult	Somewhat more difficult	Neither easier nor more difficult	Somewhat easier	Much easier
				2	2	1
Question 17	Indicate your level of agreement/disagreement with the following statement: "Attending to the 3D audio cues was yet one more task added to an already busy flight."	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
			4		1	
Question 18	Indicate your level of agreement/disagreement with the following statement: "By relying on 3D audio cues to alert me of arrival at the next turnpoint, I could devote more attention to completing emergency procedures (than I could without 3D audio cues)."	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
			1		4	

Figure D-4: Post-Flight Questionnaire (Sheet 4 of 6)

	3D audio and situational awareness	(Situational awareness is the ability to identify, comprehend, and maintain awareness of the various elements of a person's environment. More simply stated, it's knowing what is going on around you.)				
Question 19	Overall, how did your situational awareness (S. A.) with 3D audio cueing compare to your situational awareness without 3D audio cueing?	Much worse	Somewhat worse	Neither better nor worse	Somewhat better	Much better
					4	1
Question 20	How did the use of 3D audio cueing affect your sense of direction to the "go to" waypoint?	Greatly impaired	Somewhat impaired	Neither improved nor impaired	Somewhat improved	Much improved
				1	3	1
Question 21	How did the use of 3D audio cueing affect your sense of waypoint passage?	Greatly impaired	Somewhat impaired	Neither improved nor impaired	Somewhat improved	Much improved
				1	2	2
Question 22	Upon arriving at turnpoints, how much more or less accurately did 3D audio cueing allow you to determine the heading to the next waypoint (vs without audio cueing)?	Much less accurately	Somewhat less accurately	Neither more nor less accurately	Somewhat more accurately	Much more accurately
				2	3	

Figure D-5: Post-Flight Questionnaire (Sheet 5 of 6)

Question 23	Using 3D audio cueing only, how precisely could you detect heading errors?	Very imprecisely	Somewhat imprecisely	Neither precisely nor imprecisely	Somewhat precisely	Very precisely
			2	1	1	1
Question 24	Indicate your level of agreement or disagreement with the following statement: "The 3D audio cues consistently reinforced my mental image of where I was in relation to the instrument approach."	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
			1	1	2	1
Question 25	Indicate your level of agreement or disagreement with the following statement: "When the 3D audio cue stepped from one waypoint to the next, the change in direction of the cue allowed me to more easily determine in which direction to turn (exclude those points at which you entered holding)."	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
					4	1
	Overall impression of utility					
Question 26	Overall, how useful/useless was the addition of 3D audio cues to the aircraft?	Very useless	Somewhat useless	Neither useful nor useless	Somewhat useful	Very useful
					4	1

Figure D-6: Post-Flight Questionnaire (Sheet 6 of 6)

VITA

Alan Joseph Wigdahl was born in Kettering, OH in 1973. He was raised in Casstown, OH, where he graduated from Miami East Local Schools in 1991. From there he attended Edison State Community College, Piqua, Ohio for a year before being appointed to the United States Air Force Academy in Colorado Springs, CO. At the Academy he received a B.S. in Civil Engineering (French language minor) in 1996 and was recognized as the top civil engineering graduate and third overall graduate. During 1997-1998 he attended Euro-NATO Joint Jet Pilot Training in Wichita Falls, TX and the F-16 Replacement Training Unit Course in Phoenix, AZ before settling at Cannon Air Force Base, NM for his first operational assignment in 1999. Alan upgraded to Instructor Pilot and Mission Commander while in New Mexico. Following this assignment, he served at Kunsan Air Base, Republic of Korea for a year before attending United States Air Force Test Pilot School (TPS) in 2003-2004. After graduating from TPS he was assigned to the 416th Flight Test Squadron, Edwards AFB, CA as an Experimental Test Pilot and Instructor Pilot. Subsequently he served as the Chief of Test System Safety at the Air Force Flight Test Center. He was rated as a senior pilot with more than 2000 flight hours in 32 different aircraft types. He received his M.S. in Aviation Systems at the University of Tennessee Space Institute, Tullahoma, TN in 2008.