INVESTIGATING THE IMPORTANCE OF STEREO DISPLAYS FOR HELICOPTER LANDING SIMULATION

Marc Winterbottom, Charles Lloyd, James Gaska, Logan Williams, Elizabeth Shoda, Charles Bullock, and Steve Hadley

U.S. Air Force medical vision standards currently establish a minimum level of depth perception and ocular alignment. These standards apply not only to pilots but also to aircrew with scanner duty, which applies to aircrew involved in clearing aircraft for landing. This has been especially important since an accident in 1998 involving two H-60 aircraft that collided, where poor depth perception was identified as a contributing factor. However, similar standards do not apply for Army personnel in similar positions, and many other countries do not test for depth perception and ocular alignment even for pilots. Further, although much research has been conducted to examine the importance of stereo acuity and stereo displays in a wide variety of tasks, the results are mixed. The objective of the research presented here was to examine the effect of both stereo acuity and stereo displays on the performance of a helicopter landing task. For this research a representative task was selected in which subjects were required to discriminate the distance between the rear wheel of the aircraft and the top of an object over which the aircraft hovered. The simulation was constructed using the X-Plane software running on a pair of Windows PCs and viewed using a helmet mounted display. A unique aspect of this research is that observer stereo acuity, fusion range, and contrast sensitivity were thoroughly evaluated prior to participation. The results of the first evaluation indicated that observers with good stereo acuity scores performed significantly better on the operational task when stereoscopic video was used relative to monoscopic video. The results of the second evaluation indicated that operational performance could be predicted from a combination of vision test scores, but that stereo acuity test scores did not predict performance when used in isolation.

Stereo vision, stereoeacuity, stereoeacuity test, stereoeacuity assessment, helicopter landing, aircrew vision standards, aircrew vision requirements, helicopter simulation, stereo display, operational performance.
INVESTIGATING THE IMPORTANCE OF STEREO DISPLAYS FOR HELICOPTER LANDING SIMULATION

Marc Winterbottom¹, Charles Lloyd², James Gaska¹, Logan Williams¹, Elizabeth Shoda³, Charles Bullock¹, and Steve Hadley¹

¹Operational Based Vision Assessment (OBVA) Laboratory, Operational Health and Performance Research Division, Aeromedical Research Department, U.S. Air Force School of Aerospace Medicine, 711th Human Performance Wing, Wright-Patterson AFB, OH, USA

²Visual Performance LLC, Ellisville, MO, USA

³Wyle Laboratories, Beavercreek, OH, USA

ABSTRACT

U.S. Air Force medical vision standards currently establish a minimum level of depth perception and ocular alignment. These standards apply not only to pilots but also to aircrew with scanner duty, which applies to aircrew involved in clearing aircraft for landing. This has been especially important since an accident in 1998 involving two H-60 aircraft that collided, where poor depth perception was identified as a contributing factor. However, similar standards do not apply for Army personnel in similar positions, and many other countries do not test for depth perception and ocular alignment even for pilots. Further, although much research has been conducted to examine the importance of stereo acuity and stereo displays in a wide variety of tasks, the results are mixed.

The objective of the research presented here was to examine the effect of both stereo acuity and stereo displays on the performance of a helicopter landing task. For this research a representative task was selected in which subjects were required to discriminate the distance between the rear wheel of the aircraft and the top of an object over which the aircraft hovered. The simulation was constructed using the X-Plane software running on a pair of Windows PCs and viewed using a helmet mounted display. A unique aspect of this research is that observer stereo acuity, fusion range, and contrast sensitivity were thoroughly evaluated prior to participation.

The results of the first evaluation indicated that observers with good stereo acuity scores performed significantly better on the operational task when stereoscopic video was used relative to monoscopic video. The results of the second evaluation indicated that operational performance could be predicted from a combination of vision test scores, but that stereo acuity test scores did not predict performance when used in isolation.

INTRODUCTION

A depth perception standard has been enforced for aviators since the early years of aviation. For example, Wilmer and Berens noted that “the value of stereoscopic vision…is of great value in judging distance and landing…The importance of this qualification seems to grow greater as our experience increases” [1]. Howard developed one of the first tests of depth perception for screening purposes and, on the basis of his research, believed that “to possess normal judgment of distance one’s binocular parallactic angle should not be greater than 8.0” [2]. However, the debate concerning the utility of depth perception has also been ongoing since the early 1900s. Howard (1919) noted “some examiners have questioned the absolute necessity of binocular single vision as a preliminary requirement.[2]” Although a 1996 Delta MD-88 crash at LaGuardia was partly attributed to defective stereopsis, some researchers have concluded that stereopsis is not required for flight safety, owing to the fact that other cues of depth are sufficient [3], [4]. According to U.S. Air Force (USAF) medical policy, good stereo acuity and ocular alignment are both considered to be critical for pilots, but also for non-pilot aircrew (Flying Class III (FCIII) aircrew) involved in certain tasks such as clearing aircraft for landing [5]. An FCIII depth perception standard has been enforced for USAF aircrew since 1998, following a fatal accident involving two H-60 helicopters where defective stereopsis was identified as a contributing factor [6]. However, a similar standard is not maintained for Army personnel in similar aircrew positions, and many other countries do not maintain a depth perception standard, even for pilots.
Research examining the importance of either stereo acuity or the use of stereo displays has had mixed results. In a systematic review of 71 experiments, previous researchers found that although about 67% showed a benefit of three-dimensional (3D) displays, the remaining 33% either did not show a benefit or had very mixed results [7]. Similarly, a review of the importance of depth perception in aviation showed that not only is it difficult to clearly identify the importance of good stereo acuity, traditional methods used to measure stereo acuity may be lacking, which likely contributes to confusion concerning the utility of stereopsis and stereo displays [8]. In simulation and training applications, the use of stereo displays has been very limited. This may be due to several factors. Conventional knowledge has held that stereo is not useful beyond a few meters. Previous studies using electronic displays [9], [10] found stereo acuity thresholds of ~140 arcsec (i.e., many times higher than reported for real objects). Thus, previous experience with inadequate displays may have led to the conclusion that stereo cues would be ineffective for larger distances. Previous implementations of training systems with stereo displays proved difficult to implement [11], and two previous efforts to demonstrate the effectiveness of stereo displays for boom operator training were cancelled. Difficulties with the use of stereoscopic displays are well known and may be attributable to a number of different factors such as vergence-accommodation mismatch, image distortion/ misalignment between the left and right eye images, use of differing filters in the left/right eye (e.g., red/green filtering), conflicting depth cues (e.g., blur vs. disparity, lack of appropriate motion parallax), etc. [12]–[18].

As noted briefly above, a major limitation for many studies examining the utility of depth perception for performance of real-world tasks is that the measures of depth perception are often coarse and suffer from significant floor effects. If stereo acuity or other clinical metrics relevant to binocular health are actually obtained, they are often limited to, for example, a 40- or 60-arcsec minimum threshold, or simply “fly positive,” meaning that subjects could see the 3D fly on a commonly available near stereo acuity test. Thus, part of the confusion concerning the utility of stereopsis may stem from the use of limited measures of binocular health. Although the potential limitations of some commonly used stereo acuity tests have been discussed [19]–[21], these tests are still frequently used. Our own research suggests that a more carefully designed computer-based stereo acuity test, although correlated with a more standard test, differs substantially (Figure 1) in outcome. As shown, there is a substantial floor effect on the standard test, and further, individuals obtaining the best score of 15 arcsec on the standard test may score anywhere from approximately 5 arcsec to 250 arcsec on the adaptive, threshold-based test. These results are consistent with previous research that suggests that the standard stereo acuity tests may actually test something other than stereo acuity. For this reason, we used our computer-based stereo acuity test in the research presented here rather than rely only on the more commonly available chart-based methods.

![Figure 1. Relationship between chart-based AFVT/AO Vectograph stereo acuity test and OBVA Lab, computer-based adaptive stereo acuity test.](image)

The objective of the research presented here was to examine the effect of both stereo acuity and stereo displays on the performance of a helicopter landing task. For this research a representative task was selected in which subjects were required to discriminate the distance between the rear wheel of the aircraft and the top of an object over which the aircraft hovered. To initiate this line of research, we have broken down a very complex call-to-landing task into sub-components, beginning with the hover task presented here. Future research will examine time to contact and height estimation prior to researching performance in a full combat landing simulation. The simulation was constructed using X-Plane software running on a pair of Windows PCs and viewed using a helmet-mounted display (HMD). A unique aspect of this research is that each observer’s stereo acuity, fusion range, and contrast sensitivity were thoroughly evaluated using computer-based vision tests developed in our laboratory prior to participation in the simulated helicopter landing task. It is important to note that depth perception involves much more than binocular disparity. A wide variety of monocular cues, such as optic flow [22], [23], motion parallax [16], [17], [24], relative size, and occlusion [25], all contribute to depth perception. In this research, the use of a head-tracked, wide field of view HMD...

---

**Presented at the IMAGE 2016 Conference**

Dayton, Ohio – 28-29 June 2016

Cleared, 88PA, Case # 2016-2917.
and highly detailed simulated environment preserved many of the cues to depth that would normally be encountered in a natural environment. Thus, the research described here should be relevant for examining the contribution of stereo displays and quality of vision to the performance of a highly complex task such as a helicopter call-to-landing.

METHODS

Three experiments were devised to accomplish different objectives. Experiment 1 was designed to ensure that a difference in performance could be observed for monoscopic versus stereoscopic viewing conditions for a limited number of subjects known to have good stereo acuity. Experiment 2 was designed to examine individual differences in performance and determine whether stereo acuity or fusion range affected performance on this task related to clearing rotary wing aircraft for landing. Experiment 3 was designed to compare stereo acuity thresholds for similar test stimuli displayed on a flat panel display versus in the HMD. A significant effort was made to ensure that binocular disparities displayed in the HMD were accurate. In previous work, we showed that depth thresholds were similar for cross-shaped targets presented in the X-Plane environment in the HMD in comparison to an analogous pair of real-world cross-shaped targets at the same viewing distance [26]. Experiment 3 is a continuation of efforts to confirm that the stereo HMD is configured properly.

Experiment 1

Subjects

Five subjects with far stereo acuity better than 30 arcsec participated in this experiment. All subjects provided informed consent and the experimental protocol was approved by the Wright-Patterson Institutional Review Board (IRB).

Apparatus

The 3D virtual environment was generated using two instances of Laminar Research’s X-Plane® to stereoscopically render a head-tracked, out-the-window visualization. The two instances of X Plane® were implemented using two separate PCs, each incorporating Intel i7 processors and Nvidia Quadro K4200 video cards with Quadro Sync to ensure 60-Hz rendering synchronization between each channel of the stereoscopic visualization. The primary stereoscopic display system consisted of an SA Photonics SA-55 binocular HMD, composed of two 1920x1200 opaque organic light emitting diode displays with 100% overlap, 55° horizontal field-of-view (FOV), 4-meter virtual image distance, and capable of motion blur reduction (variable pixel hold time). This HMD was selected in part because it is one of the highest resolution, largest FOV HMDs currently available (Figure 2). Previous work [27] suggested that a relatively high resolution was needed to adequately display stereo imagery. The FOV limitation of the HMD is operationally similar to performing the call-to-landing task using night vision goggles (NVGs, ~40° FOV) and was thus not considered to be a significant system limitation.

Figure 2. SA-55 HMD configured with 3D printed infrared (IR) reflective rigid body. Photo provided by SA Photonics, used with permission.
The native optics of the SA-55 HMD produce significant pincushion distortion; therefore, the Brown-Conrady model for radially symmetric image warping [28] as implemented to minimize optical distortion while adding negligible latency. The model parameters were determined empirically by subjective evaluation of the final image geometry. However, this reduces the active horizontal FOV to 44° due to the loss/deactivation of pixels near the image borders, as shown in Figure 3.

However, the modified FOV is still within the 40° FOV typical of NVGs. A custom 3D printed rigid body constellation, containing five IR reflectors, was mounted to the HMD for integration with the OptiTrack system. The HMD with rigid body and the structure built by OBVA Lab personnel to configure the head-tracking cameras are shown in Figure 4.

Head tracking was performed using a NaturalPoint OptiTrack (Motive:Tracker) IR tracking system, utilizing seven Flex-13 cameras, with tracking latency of approximately 0.4 ms. Noticeable jitter did require the use of significant smoothing, which introduced a noticeable delay with rapid head movements. However, this artifact is likely inconsequential since subjects tended to hold very still during the distance estimation task. The simulation control host was written in Matlab (MathWorks, 2014) by the OBVA team and operated from a third, separate PC. Communication between the Matlab host and the X-Plane® rendering machines was implemented using UDP multicast packets containing the relevant control parameters (e.g., aircraft position, etc.). The apparatus has also been described in greater detail in a previous publication [26].
Figure 4. Structure and head-tracking cameras constructed by OBVA Lab personnel.

Procedure

Subjects were first fitted with the HMD, ensuring that the binocular imagery was properly aligned. Subjects were instructed to indicate in which trial the tail wheel of the H-60 was most closely aligned vertically with the windsock pole. At the beginning of each trial, the aircraft quickly descended into place from a higher position and then remained stationary for the remainder of the trial. In different blocks of trials, the aircraft height above the tail wheel was varied from 0.1 m to 2 m and stereo imagery was turned on or off. A small degree of random variation was applied to each flight path (between trials) to both enhance realism and eliminate frequent repetition of rendering artifacts that might be used as unintended cues by the observer during the call-to-landing experiment. For the purposes of this experiment, the subjects’ viewing position (the eye-point) was placed such that they could easily view the tail wheel and windsock, as shown in Figure 5, without having to reposition themselves within the eye-tracking structure shown in Figure 4.

For each viewing condition, the $\Psi$ (psi) method [29] was used to estimate the smallest displacement in depth relative to the windsock that each subject could reliably detect.

Results

Figure 6 shows the results of Experiment 1. As shown, smaller thresholds are obtained under stereoscopic viewing conditions in comparison to monoscopic viewing conditions ($t = -6.21, p << 0.001$; averaged over tail wheel height). However, this difference diminishes as the separation between the tail wheel and pole is increased.

Experiment 2

Subjects

Forty subjects volunteered to participate in Experiment 2. All subjects provided informed consent and the experimental protocol was approved by the Wright-Patterson IRB.
**Apparatus**

The apparatus for the helicopter landing simulation used in Experiment 2 was the same as that described for Experiment 1. A Dell Precision T7610 with Nvidia GeForce GTX 680 graphics card was used to administer the stereo acuity and fusion range tests. The tests were displayed on an Asus VG278HE 3D monitor with 1920x1080 pixels that was compatible with Nvidia 3D Vision2 and that used active shutter glasses. At a 1-m viewing distance, the angular pixel size was 1.1 arcmin. A Zotac Z-box mini-PC was used to administer the contrast sensitivity test. The test was displayed using an NEC P232W-BK monitor. A Logitech game controller was used to enter responses.

**Procedure**

All subjects were first administered the stereo acuity and fusion range tests developed by the OBVA Laboratory. Figure 7 shows examples of the test stimuli. For the stereo acuity tests, subjects indicated whether the smaller inner circle appeared popped out or receded in depth relative to the larger reference circle using two buttons on the game controller. The $\Psi$ method was used to estimate stereo acuity. The vertical fusion range test required that participants indicate when a circle viewed at a distance of 1 m on the Asus stereo monitor resulted in double vision (i.e., binocular fusion was broken) using the game controller as the circles (displayed separately to the left and right eyes) moved apart in the vertical direction. The direction of motion then reversed, and the participant next indicated when the circles returned to a single “fused” image using the game controller. This task was repeated several times. The amount of separation between the left and right eye images was recorded at the time the subject pressed the button on the game controller for each trial. A similar procedure was used for horizontal fusion range, except that the two circles moved apart in the horizontal direction, requiring subjects to either cross their eyes or uncross their eyes to maintain binocular fusion.

For the helicopter hover task, the procedure was similar to that described in Experiment 1, except that only two heights above the windsock (0.2 and 2 m) were used.

**Results**

Figure 8 shows the distribution of stereo acuity and horizontal fusion ranges scores. As shown, subjects vary substantially on both dimensions. For stereo acuity, the best observers obtained scores better than 10 arcsec, while the worst were 300 arcsec or greater.
Figure 8. Distribution of far stereo acuity scores (top) and distribution of horizontal fusion range scores (bottom).

Figure 9 shows the distribution of depth thresholds for each viewing condition and aircraft height (0.2 m, top graph; 2 m, bottom). Similar to Experiment 1, the use of a stereoscopic display clearly decreases depth thresholds. Shown in Figure 10 are the average depth thresholds for each viewing condition and aircraft height. The average depth thresholds were 114 cm, 198 cm, 171 cm, and 299 cm for the 0.2 m stereo, 0.2 m no stereo, 2 m stereo, and 2 m no-stereo viewing conditions, respectively. A repeated measures analysis of variance reveals that stereoscopic viewing and height both had significant effects \( F(1, 156) = 25.7, p << 0.001; F(1, 156) = 13.5, p < 0.001 \) but no significant interaction (see Figure 10).

Figure 10. Average depth thresholds in Experiment 2 for each viewing condition and aircraft height.
Shown in Figure 11 are the relationships between individual stereo acuity scores (top) and individual fusion range scores (bottom) and depth thresholds averaged over aircraft height. The correlation between stereo acuity and depth accuracy was not significant. However, the correlation between fusion range and both stereoscopic ($r = 0.5, p < 0.001$) and monoscopic ($r = 0.37, p = 0.02$) depth accuracy was significant. Shown in Figure 12 is the relationship between the combination of stereo acuity and fusion range and depth accuracy.

![Figure 11. Scatterplots showing relationships between individual stereo acuity scores (left) and individual fusion range scores (right) and depth thresholds averaged over aircraft height.](image)

![Figure 12. The relationship between the combination of stereo acuity and fusion range and depth accuracy.](image)

**CONCLUSIONS**

The results of Experiments 1 and 2 clearly show that stereoscopic viewing improves performance when judging the depth of a simulated H-60 tail wheel relative to a landing zone obstacle, in this case a windsock pole. Thus, it could be expected that subjects with better stereo acuity should perform better than subjects with poor stereo acuity on this relative depth task. However, as shown in Figure 11, stereo acuity considered in isolation does not predict performance. As shown in Figure 12, fusion range and stereo acuity test scores may need to be considered together to predict performance on this simulated call-to-landing task. As noted in the introduction, a wide variety of visual cues contribute to depth perception, such as relative size, motion parallax, texture gradient, etc.; thus, observers may be able to use the combination of these cues to perform the task adequately despite weak stereo acuity. Nonetheless, it is still somewhat surprising that observers with superior stereo acuity were not clearly superior in performing this operationally relevant task involving depth estimation.

Shown in Figure 13 is a comparison of the average stereo acuity test results and average hover task depth thresholds. As shown, average hover task depth discrimination thresholds (170 arcsec) are nearly 4x greater than average stereo acuity test thresholds (45 arcsec). While the simulated tail wheel and windsock pole are not ideal stimuli for best stereo acuity [30], it is still worth noting that depth discrimination thresholds for real-world objects at similar viewing distances were found to be approximately 3-8 cm in previous research, which is consistent with stereo acuity thresholds of approximately 5-10 arcsec [31]. Similarly, in research examining display requirements for aerial

Presented at the IMAGE 2016 Conference
Dayton, Ohio – 28-29 June 2016

Cleared, 88PA, Case # 2016-2917.
refueling simulation and training, depth judgments were consistent with stereo acuities of 5-10 arcsec (Lloyd & Nigus, 2012). While we went to some length to validate the stereoscopic simulation for this research, and depth judgments were shown to be equivalent for simulated and real-world targets [26], it is possible that some aspect of the simulation may have limited depth discrimination thresholds. In future research we will continue to examine simulation requirements for stereoscopic displays, as well as the effect of stereo displays and stereo acuity on time to contact and height above terrain judgments, as well as performance in clearing the aircraft for landing in a complex combat landing simulation.

Figure 13. Average stereo acuity test results and average hover task depth thresholds.

ACKNOWLEDGMENTS

The views expressed are those of the authors and do not necessarily reflect the official policy or position of the Air Force, the Department of Defense, or the U.S. Government. This work was supported by U.S. Air Force contract FA8650-12-6280 to Wyle Laboratories and was funded by the 711 HPW/USAFSAM.

REFERENCES


