ELSEVIER

Displays 12 (2000) 17–24

DISPLAYS

www.elsevier.nl/locate/displa

Visuospatial task performance as a function of two- and three-dimensional display presentation techniques

K.F. Van Orden[a,b,*], J.W. Broyles[a]

[a]*Space and Naval Warfare Systems Center, 5425 Patterson Road, San Diego, CA 92152-7150, USA*
[b]*Naval Health Research Center, San Diego, CA, USA*

Received 24 May 1999; received in revised form 18 October 1999; accepted 1 November 1999

Abstract

Three-dimensional (3D) displays are becoming more widely available and are being applied to a greater variety of human–computer interface domains. Previous research has shown that 3D display of objects and information often is more appealing to the users, but for many tasks it is less useful than the two-dimensional (2D) displays. New display techniques must be assessed for their ability to improve human operator performance. The purpose of this research was to compare human performance on several 2D and 3D display formats across four visuospatial tasks. Qualified military and civilian air traffic controllers completed altitude and speed judgement tasks, a vectoring task, and a collision avoidance task on 2D top-down (plan-view), 3D perspective, 3D stereo, and laser-based 3D volumetric display systems. Each subject's speed and accuracy were measured on each task. Results indicated that the 2D plan or side-view displays yielded performance as good or better than any other display system for speed and altitude judgment tasks. Data presentation on the 3D volumetric display was superior to 3D perspective, 3D stereoscopic, and 2D displays only for the collision avoidance task. These results support previous research suggesting that 3D displays are useful in very specific tasks. The results from the collision avoidance experiment suggest that tasks requiring operators to view and predict future locations of multiple display symbols traversing a confined space (such as relationships between aircraft within the airspace around an airport) appear to be well suited for 3D rendering. Compared to 3D stereoscopic and perspective displays, the veridical display of localized spatial information within a volumetric display may provide high fidelity stereoscopic and parallax cues, improving human performance for some tasks. © Published by Elsevier Science B.V.

Keywords: 2D displays; 3D displays; Stereoscopic displays; Volumetric displays; Human performance

1. Introduction

Human operators within civilian air traffic control and military command centers are required to develop three-dimensional (3D) mental models of their respective environments by relying on information presented upon two-dimensional (2D) display surfaces. 3D displays are more readily available and might possibly improve the ability of operators to understand the actions and activities of real-world entities. 3D displays are often more expensive (in terms of cost and processing overhead) than 2D displays, and depending upon an operator's task, may not offer specific advantages over 2D displays. Furthermore, a growing number of 3D display techniques are available, including 3D perspective (on 2D media), stereoscopic glasses, and virtual reality, each claiming better "real world" representation and implying better human performance. While 3D displays are often more appealing to users, performance data is not necessarily correlated with opinion data [1]. For example, Steiner and Dotson [2] compared 2D and 3D aircraft cockpit tactical displays during simulated air combat scenarios. They found that while pilots preferred the 3D displays, performance was significantly better when the 2D display was used. Command and control as well as air traffic control system users may benefit from 3D presentation of data, although the results of human performance studies comparing different 3D display technologies (e.g. 3D perspective and stereoscopic) have been inconclusive.

Although previous comparative studies of 3D rendering technologies have been limited in scope, and may not have kept pace with rapid technological development, some consistent findings and themes have emerged. Several reviews of effects of 3D displays on performance [3,4] indicate that there is limited evidence that 3D perspective displays facilitate tasks such as air traffic direction any

* Corresponding author. Space and Naval Warfare Systems Center, 5425 Patterson Road, San Diego, CA 92152-7150, USA.
E-mail address: vanorden@spawar.navy.mil (K.F. Van Orden).

0141-9382/00/$ - see front matter © Published by Elsevier Science B.V.
PII: S0141-9382(00)00024-X

Fig. 1. 3D Volumetric Display. During operation room lights were dimmed and subject sat or stood in front of display. Stimuli were presented as shown in Fig. 2.

better than well-designed versions of 2D displays that contain equivalent information. Tham and Wickens [5] investigated the performance of air traffic controllers, pilots, and students on 2D top down (plan view), 3D-perspective, and stereoscopic radar displays. They found no performance differences for the majority of task judgments among 2D, 3D perspective, and 3D stereoscopic displays. Stokes et al. [6] summarized the difficulty one faces when choosing to display information in a 3D versus a 2D display format for air traffic control. They conclude that "3D representation is clearly a compatible and natural representation of airspace, while the depicted objects and their location in the space are integrated as single display entitles, not separated as different views across two (or three) displays. Yet on the other hand, the 3D representation carries with it a number of less obvious disadvantages." Stokes et al. point out that the benefit of an integrated perspective of a space and its constituent parts engendered by 3D displays comes at the cost of making precise estimates of position along one of the dimensions. Thus while relative locations are well represented, absolute locations can be ambiguous.

New 3D display technologies require systematic investigation of human performance over a variety of tasks in order to determine their relative value for real-world applications compared to more traditional 2D display techniques. Stereoscopic and volumetric 3D rendering techniques provide additional visual cues (e.g. parallax, stereo cues) over 3D perspective displays [7]. The present study represents an initial human performance evaluation of a 3D volumetric display (3DVD). The 3DVD allows the display of true 3D images by projecting laser light onto a rapidly rotating helix. The timing of laser light pulses and the position of the helix renders stationary "voxels" that can be used to construct images of objects [8–11]. This display scheme permits a more natural method of determining intra- and inter-object relationships—through the parallax created by the head and body movement of the observer, as well as true stereoscopic vision. At issue is whether the veridical display of 3D data, as presented by the 3D volumetric display, is advantageous for human performance relative to more traditional 3D rendering technologies. The present study represents a first step in the process of evaluating 2D and 3D displays for use as military tactical and air traffic control displays.

The present experiment examined four visuospatial tasks presented on as many as seven different display formats (2D-plan view, 2D stereo plan view, 2D side view, 2D stereo side view, 3D-perspective, 3D-stereoscopic, and 3DVD). The dependent variables were accuracy (percent correct judgments made in each task) and response time. In addition, several derived calculations were made on specific task data (e.g. distance deltas, time before collision, 3D display adjustment times, etc.) to better understand usability issues in viewing and manipulating objects in a 3D space.

2. Method

2.1. Subjects

Twenty military and civilian air traffic controllers (15 males and five females) with a mean age of 28.9 years and ranging in age from 19 to 52 years, participated in the experiment. The subjects had approximately eight years of Naval service and 5.8 years of experience performing air traffic control (ATC) duties. Two subjects failed to complete the study (one left the Navy and the other failed to complete all assigned tasks) leaving 18 subjects. Subjects were tested for stereo or fine depth discrimination using a Stereo Optical Company's RANDOT stereoscopic vision test.

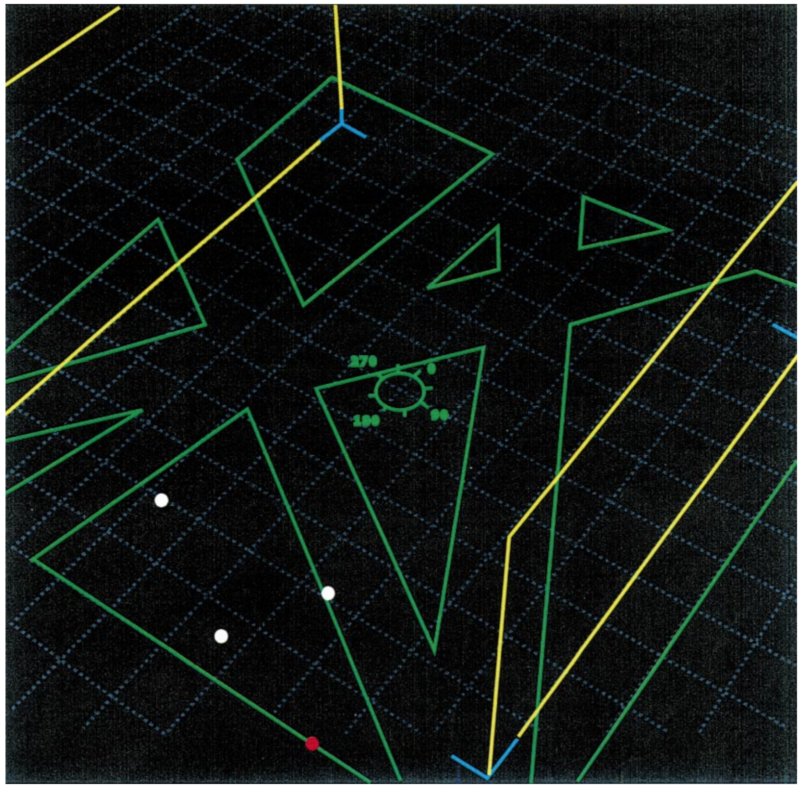

Fig. 2. 3D Perspective view of experimental display on SGI computer. Spheres represent aircraft, red sphere indicates aircraft selected by subject.

2.2. Apparatus

The ATC tasks were presented on our laser-based (36-in. diameter 18-in. high double helix) 3D Volumetric Display System with an INTEL 486DX2 50 MHz CPU and V3D controller cards (see McDonald and Arney [8]; Soltan, Trias, Dahlke, Lasher, and McDonald, [9,10]). The 3DVD is shown in Fig. 1. The ATC tasks were also presented on a Silicon Graphics (SGI) Indigo 2 High Impact Graphics computer system with a 200 MHz MIPS 4400 CPU, Sony 20 inch diagonal Stereo-ready color monitor, and Stereographics CrystalEYES 2 glasses (for stereo display format only).

Display formats. Display formats consisted of a 2D-“plan” view (vertical top down view of airspace), 2D-stereoscopic plan view, 3D-perspective, 3D-stereoscopic, and 3D volumetric display (3DVD). The SGI Indigo 2 High Impact Graphics computer system presented all display formats with the exception of the 3D Volumetric display view, which was shown on the 3D Volumetric Display. Each format presented the same scene; sphere shaped aircraft flying above a grid-lined airfield with three runways and a small compass rose in the middle of the airfield (see Fig. 2). The displays of each system were constructed to mimic the resolution of the 3DVD system (lowest resolution). The scene was optimized for the respective depth cues afforded by the presentation on each display format. The 2D stereoscopic view offered the same views, plus stereo depth effects.

The 3D Perspective and 3D Stereoscopic views offered subjects free rotational viewing capability. Subjects could rotate their viewing angle from 0 to 90° in the horizontal axis and $+60$ or $-60°$ off the center in the vertical axis.

The 3D Volumetric display offered 3D viewing limited only by the 120°, frontal viewing space. This limitation was caused by the solid (non-translucent) rotating helix. The 3D Volumetric display used a Logitech remote control, hand-held, 3-button mouse. The SGI High Impact computer used a programmable 3-button mouse. The input devices were programmed to interact with the 2D and 3D environments in an identical manner.

2.3. Tasks

Software from a study by Tham and Wickens’ [5] air traffic control study was modified to present the following specific tasks:

- *Altitude judgment tasks.* Subjects viewed 4–5 aircrafts in flight on each trial and determined the aircraft with the highest or lowest altitude. In this and all other tasks, the subjects depressed a keyboard or mouse button to cycle through the displayed targets, highlighted by a cube surrounding the aircraft, until they selected the aircraft perceived to be of highest or lowest altitude. The subject used another button to enter the aircraft as an answer for the trial. Aircraft velocity was equivalent across all display platforms.

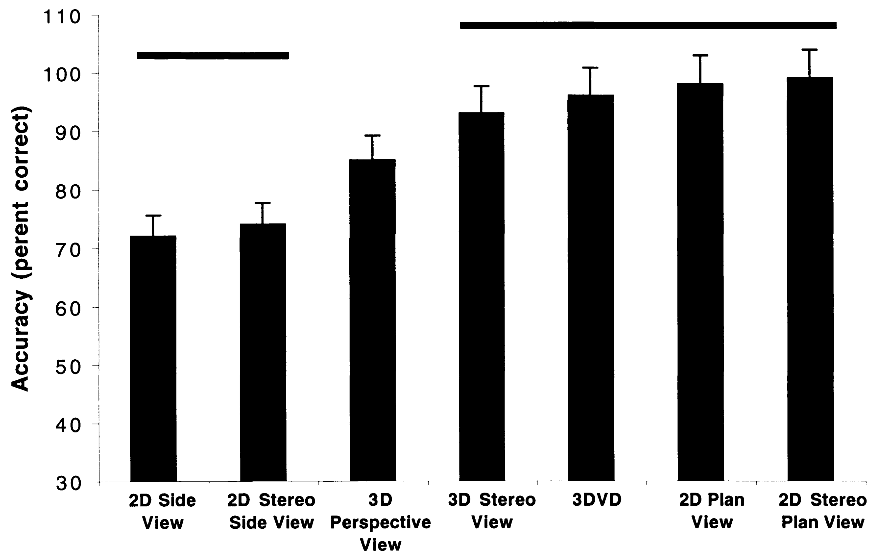

Fig. 3. High-altitude judgment accuracy means as a function of display condition. Error bars represent one standard deviation from the mean. Horizontal bars represent homogeneous sub-groups in post-hoc testing for differences among the means.

- *Speed judgment tasks.* In different blocks of trials, subjects were required to determine which of three aircraft was flying the fastest or the slowest.
- *Aircraft vectoring task.* The subject was required to vector an aircraft to a position marker on the display using as few turn commands as possible. Subjects viewed the display with a single aircraft present, and called out course changes in absolute degrees. The experimenter entered the new course into the computer to direct the aircraft. The subject would say "stop" when the aircraft was as close to the marker as possible, and the experimenter would terminate the trial.
- *Collision avoidance task.* Subjects would first view four aircraft in motion on the display, then freeze the display with a button press. With another button press the subject would respond "yes" or "no" that a collision was imminent. In the case of a yes response, the subject would select the two aircraft they believed would collide.

2.4. Procedure

At the start of each session, the subjects received an orientation on the scope of the experiment, trial descriptions, salient safety and display features, and 3D mouse input device operation. For each ATC task, subjects received 10 practice trials followed by the experimental trials (20–37 trials, depending on the specific task). The subjects completed all five modified ATC tasks on each display format. Display format trials were counterbalanced across the subjects for 2D and 3D formats (half of the subjects saw the 2D plan view followed by 3D perspective, while the other half saw first the 2D stereo plan view followed by 3D stereo). In the altitude judgment task only, subjects also were shown a 2D side view or the 2D stereo side view before the start of the 2D and 3D formats described above. Upon completion of each display format session, subjects filled out a usability rating form on the input device and answered a post-experimental questionnaire. The subjects completed each display format in approximately one hour. A total of 5.5 h was required to complete the experiment. The experiment was conducted in both our 3DVD laboratory and at an on-site location at a local air traffic control center.

3. Results

3.1. Altitude judgment task

The subjects performed two types of altitude judgments. For the first block of trials they selected the highest flying aircraft among four to five planes, and in the following block of trials they selected the lowest aircraft among those in flight. Seven different display formats were used. A 2×7 two-way (altitude judgment type by display format) repeated measures analysis of variance (ANOVA) was performed for each of two dependent measures. The subjects mean performance for accuracy and response time were compared across display formats to determine if there were reliable performance differences.

The analysis yielded a significant interaction effect for accuracy between display format and type of altitude (high vs. low) judgment, $F(6, 119) = 35.20$, $p < 0.0001$, as well as significant main effects for display and type of judgment. A significant interaction effect for mean response time to make a correct judgment between display format and type of altitude (high vs. low) judgment, $F(6, 119) = 6.87$, $p < 0.0001$, as well as significant main effects for both display and type of judgment were also observed.

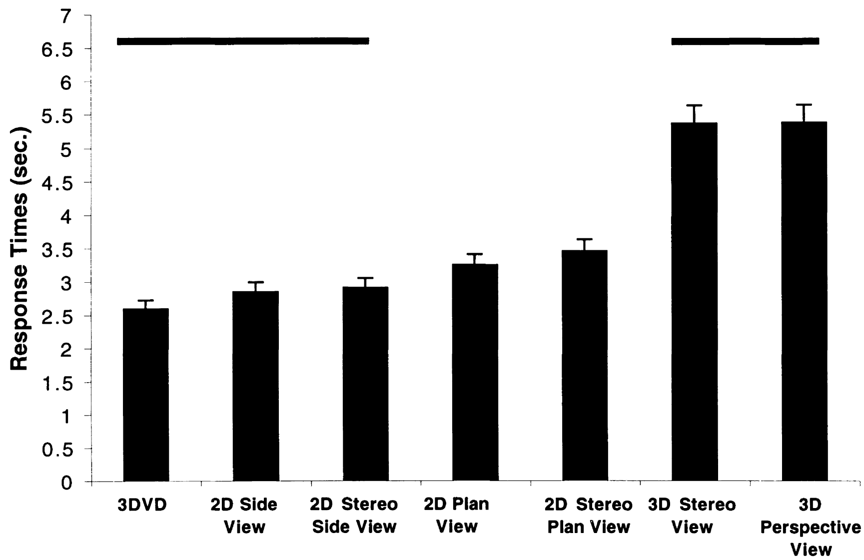

Fig. 4. High altitude judgment response times for each display format. For details see Fig. 5 legend.

The high and low altitude judgment data were analyzed separately to examine these interaction effects.

High altitude judgment accuracy. A one-way (Display Format x Subjects) repeated measures ANOVA was performed to analyze the mean percent correct accuracy scores across all seven display formats, yielding a significant main effect of display type, $F\,(6, 18) = 68.44$, $p < 0.0001$. Post-hoc pairwise comparisons using Scheffe's method (in this and all subsequent post hoc comparisons) yielded significant differences of performance between several displays. These results are shown in Fig. 3. The horizontal bars near the top of the figure represent homogeneous groups; means falling outside of these bars were significantly different from each other ($p < 0.01$). Lower accuracy scores were found for the 2D side view and 2D stereo side view compared to the subjects' performance on the other display formats.

High altitude judgment response times. Means and standard deviations for the high-altitude judgment response time are shown in Fig. 4. A one-way (Display Format $\times$ Subjects) repeated measures ANOVA for the high altitude condition on mean response time scores for correct judgments across all seven display formats yielded a significant main effect for display, $F\,(6, 18) = 75.65$, $p < 0.0001$. Post-hoc Scheffe tests indicated significant performance differences between several displays as indicated by the horizontal bars in Fig. 4.

Low altitude judgment accuracy. In the low altitude judgment task, subjects were instructed to select the lowest flying aircraft from among 4–5 aircrafts presented. Mean

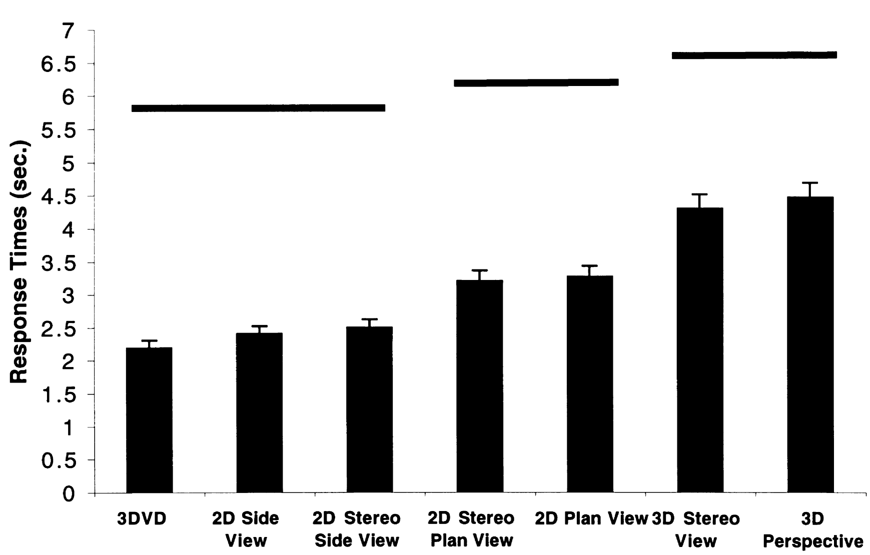

Fig. 5. Low altitude judgment response times as a function of display format.

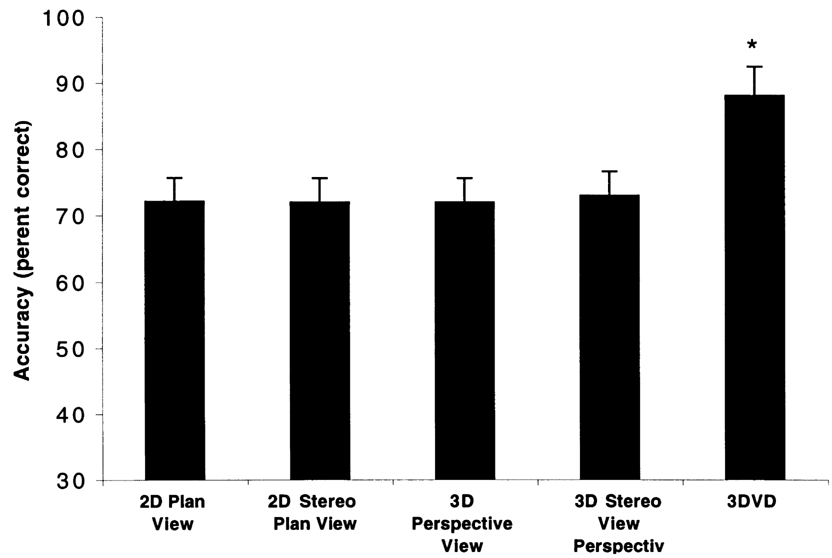

Fig. 6. Collision avoidance task accuracy as a function of display type. Asterisk indicates significant difference, $p < 0.01$.

accuracy for this task was 98% or better across all display formats.

Low altitude judgment response times. A one-way (Display Format × Subjects) repeated measures ANOVA was performed to analyze the mean response time scores across all seven display formats yielding a significant main effect for displays, $F\,(6, 18) = 127.38$, $p < 0.0001$. Post-hoc comparisons indicated significant response time differences between several display formats, as shown in Fig. 5.

3.2. Speed judgment task

In this task the subjects performed two types of speed judgments (i.e. first selecting the fastest flying aircraft among three planes and then on a following session selecting the slowest aircraft among the three planes in flight) across five different display formats. A 2 × 5 two-way (Speed Judgment Type × Display Format) repeated measures ANOVA was performed for each of the two dependent measures.

For the analysis on response accuracy, a significant main effect for type of speed judgment (fast vs. slow) was observed; $F\,(1, 85) = 14.76$, $p < 0.0002$ with the main effects for display formats barely approaching significance, $F(4, 85) = 2.33$, $p < 0.0627$ and no interaction effects. The effect for judgment type on the accuracy measure was the result of a difference between mean judgment scores, collapsed across all display types, of 70 and 74.4% for the slow and fast conditions, respectively.

For response times, significant main effects for display format and type of speed judgment made were indicated ($F\,(4, 85) = 5.33$, $p < 0.0007$; $F\,(1, 85) = 16.88$, $p < 0.0001$, respectively) and no interaction effects. The effect of judgment type was due to more rapid responses on the slow than the fast judgment condition (e.g. 4.58, 5.18 s, respectively). The speed judgment data was collapsed over type of speed judgment in order to examine the pooled display format means, $F\,(4, 85) = 5.29$, $p < 0.001$. Post-hoc Scheffe comparisons showed significantly faster response times ($p < 0.01$) when using the 2D plan view as compared to all other displays.

3.3. Aircraft vectoring task

The subjects were instructed to maneuver or vector a single aircraft to a specific location on the airfield (marked by a red field marker) as rapidly and with the fewest number of turns as possible. The number of turns the subjects made in vectoring the aircraft to the desired location and the times for completing the vectoring task were recorded for each of the five following displays: 2D Plan View, 2D Stereo plan View, 3D Perspective View, 3D Stereoscopic View, and 3D Volumetric.

A one-way repeated measures ANOVA was performed to analyze the mean number of turns made to vector an aircraft for the five display formats. A significant main effect for display, $F\,(4, 17) = 27.09$, $p < 0.0001$, followed by post-hoc tests for differences among the means, indicated that the subjects used significantly fewer turns on the 3DVD display (1.8) compared to all other display formats (range: 2.67–2.90).

An ANOVA performed to analyze the mean task completion time scores across all five display formats yielded a significant main effect for display, $F\,(4, 17) = 18.98$, $p < 0.0001$. Post-hoc statistical tests for differences among the means indicated that completion times were significantly longer for the 3D Volumetric Display (19.5 s) compared to all other means (range: 14.2–15.2 s).

The data reveal a tradeoff between the completion times and the number of turns required to vector the aircraft. Subjects spent an average of 10.7 s on each leg of the

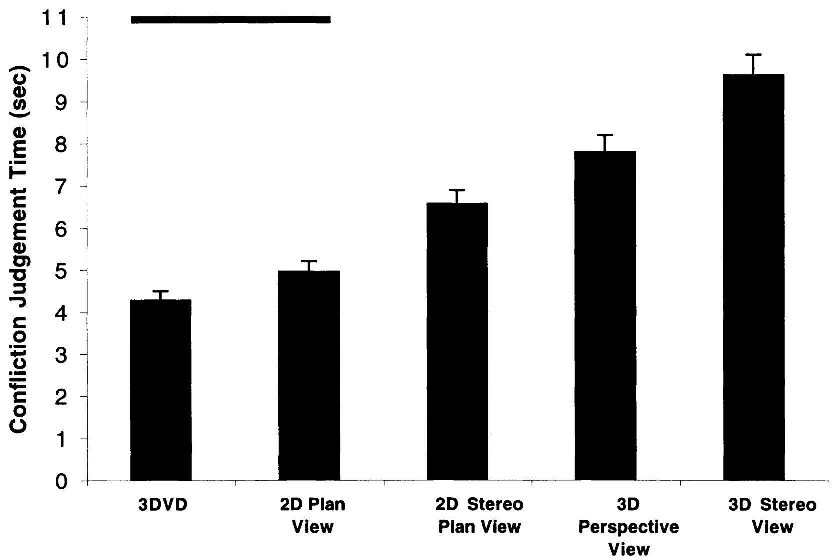

Fig. 7. Collision avoidance response times.

aircraft route on the 3DVD display, compared to 5.1–5.4 s per leg observed on all other display formats. Some subjects demonstrated a strategy of waiting until an aircraft was approaching a major axis (e.g. 0, 45, 90°) relative to the field marker and then ordered a turn while others pursued a more aggressive approach in vectoring their aircraft directly at the field marker. The data indicate that this latter strategy was more evident on the 3DVD display. The possible cause for this difference is unknown. As discussed in Section 2 (Methods), aircraft traversed the airspace on each display at the same rate, although perceived velocity differences brought about by different viewing distances may have influenced strategy on the 3DVD.

3.4. Collision avoidance judgment task

On each trial the subject was presented with a four aircraft scenario in which half of the time a collision between two aircraft was imminent and the other half of the time a collision situation did not exist. The subjects had to determine if a collision situation existed or not and report a judgment to the experimenter. A correct judgment was scored when a collision situation existed and the subjects identified it as one (e.g. picked the two correct colliding aircraft) or if there was not a collision situation and they reported the same. An incorrect judgment consisted of not reporting an imminent collision, falsely reporting a collision, or correctly reporting a collision situation, but incorrectly identifying one or both of the colliding aircraft (a rare occurrence).

A one-way repeated measures ANOVA was performed on the percent correct scores (see means and standard deviations presented in Fig. 6). The analysis yielded a significant effect for displays, $F(4,17) = 17.87$, $p < 0.0001$. Scheffe's post-hoc comparisons showed the 3D Volumetric Display produced higher accuracy compared to all other display formats ($p < 0.01$).

An ANOVA on the mean response time scores indicated a significant effect for displays, $F(4,17) = 21.10$, $p < 0.0001$, and subsequent post-hoc tests showed that subject responded fastest with 3D Volumetric and 2D Plan View displays compared to all other display formats (see Fig. 7). The 3D Stereoscopic display was associated with the longest response times ($p < 0.01$). Compared to the other 2D and 3D display formats the volumetric display was associated with significantly higher accuracy and shorter response times.

4. Conclusions

Results from the present study are congruent with findings by Tham and Wickens [5] and others [3,4], that for many tasks, there are no differences in performance between 2D and 3D displays. Many tasks used in previous studies (and most of those used in this experiment) are similar in that they typically rely upon perception along one dimension. Air traffic control, for the most part, involves routing traffic in the *x* and *y*-dimension, and then checking to make sure vertical *z*-dimension parameters are nominal. A similar situation exists for military command and control: The general altitude of an unknown aircraft approaching friendly forces from 100 km away is of little concern, and rendering such a situation in 3D is of little use. An operator may wish to review an altitude history of the unknown contact, but such information is easily rendered upon a properly formatted 2D display. Within air traffic control and military command and control domains 3D may be useful for disambiguating dynamic relationships between multiple aircraft maneuvering within a local geographical area. For example, airport or aircraft carrier approach and takeoff control, and naval amphibious operations often involve numerous aircraft operating within a constrained airspace. The conflict/collision task of the present experiment demonstrated that 3D data display results in greater accuracy and faster response times when subjects were required to anticipate whether simulated radar contacts traversing a relatively limited airspace would collide. These findings may be useful to current efforts underway to develop 3D air traffic control simulators (e.g. the NASAs ATC Test Facility [12]). Further research is required to determine how best to configure 3D displays to best convey dynamic 3D data.

Of particular interest is the finding that the performance on the collision task was best with the 3DVD display. Subjects were both less accurate *and* produced slower response times when using 3D Stereoscopic and 3D Perspective displays. The veridical presentation of the data may provide highly accurate disparity, convergence, stereoscopic and parallax cues. Further research will be necessary to identify those conditions and depth cues giving rise to the performance improvements observed with the 3DVD display over other 3D display techniques in the present experiment. Display resolution, data selection, and data manipulation methods remain significant technological challenges for volumetric display technology.

5. Summary

In summary, the present study found that for all except one task (collision avoidance), the performance of 2D displays was as good as or better than performance on 3D displays. As evidenced by higher accuracy and faster response times, the 3D Volumetric Display (3DVD) was superior and well suited for perceiving complex and dynamic data relationships in a confined 3D space in comparison to other 3D and 2D displays. The research results supported the notion that air traffic control and command and control system users may benefit from 3D representations of data for tasks requiring integration and prediction of moving display elements within limited spatial areas. 2D displays appear to be well suited for most air traffic control and command and control tasks. Given the present findings, future application of 3D display technology might be best suited for use in tasks such as local approach/takeoff air traffic control, the planning and control of military operations occurring in a confined region (e.g. complex naval amphibious), planning of invasive medical treatment, product prototyping, and some graphical data mining tasks. Despite reports on the appeal of 3D displays to system users [2], further research will be required to determine the extent to which 3D displays result in performance improvements, or decrements, within each application or for specific types of applications.

Acknowledgements

This research was supported by a grant from the Office of Naval Research Office of Naval Research under Program Element 0602233N. The views expressed in this article are those of the authors and do not reflect official policy or position of the Department of the Navy, Department of Defense, or the US Government. The authors wish to thank Neil Acantilado and Alvin Yue for their assistance with the study. Correspondence concerning this article should be addressed to Karl F. Van Orden, Space and Naval Warfare Systems Center, Code D44209, 53560 Hull Street, San Diego, CA 92152-5001. Electronic mail may be sent to vanorden@spawar.navy.mil

References

[1] A.D. Andre, C.D. Wickens, When users want what's not best for them, Ergonomics in Design, October (1995) 10–14.
[2] B.A. Steiner, D.A. Dotson, The use of 3-D stereo display of tactical information, Proceedings of the Human Factors Society's 34 Annual Meeting. Human Factors and Ergonomics Society, Santa Monica, 1990.
[3] S.R. Ellis, M.K. Kaiser, A.J. Grunwald, Pictorial Communication in Virtual and Real Environments, Taylor and Francis, New York, 1991.
[4] C.D. Wickens, S. Todd, K. Seidler, Three-dimensional displays: perception, implementation, applications, Crew System Ergonomics Information Analysis Center's State of the Art Report (CSERIAC SOAR-89-01), 1989.
[5] M. Tham, C.D. Wickens, Evaluation of perspective and stereoscopic displays as alternatives to plan view displays in air traffic control, Aviation Research Laboratory, Institute of Aviation Technical Report. (ARL-93-4/FAA-93-1), University of Illinois at Urbana-Champaign, June 1993.
[6] A. Stokes, C. Wickens, K. Kite, Display technology: human factors concepts, Society of Automotive Engineers, Inc., Warrendale, PA, 1990.
[7] W.N. Dember, J.S. Warm, Psychology of Perception, 2nd ed, Holt, Rinehart & Winston, New York, 1979, p. 303 (p. 303).
[8] M.C. McDonald, D.V Arney, The mechanical design of the helix-based 3-D volumetric display system, NCCOSC RDTE Division, San Diego, NCCOSC RDTE Technical Note 1738, January 1995.
[9] P. Soltan, J. Trias, W. Dahlke, M. Lasher, M. McDonald, Laser-based 3-D volumetric display system (second generation), Naval Engineers Journal 107 (3) (1995) 233–243.
[10] p. Soltan, J. Trias, W. Dahlke, M. Lasher, M. McDonald, Laser-based 3-D volumetric display system: second generation, in: R.M. Satava, K. Morgan, H.B. Sieburg, R. Mattheus, J.P. Christensen (Eds.), Interactive Technology and the New Paradigm for Healthcare, IOS Press, Washington, DC, 1995.
[11] R.D. Williams, Volumetric three-dimensional display technology, in: D. McAllister (Ed.), Stereo Computer Graphics and other True 3D Technologies, Princeton University Press, Princeton, NJ, 1993.
[12] M. Mecham, Airport planners gain 3D tool for ATC simulation, Aviation Week and Space Technology, May (1998) 82–84.