Comparing CAVE, Wall, and Desktop Displays for Navigation and Wayfinding in Complex 3D Models

Colin Swindells[1], Barry A. Po[1], Ima Hajshirmohammadi[2], Brian Corrie[3]
John C. Dill[2], Brian D. Fisher[1], and Kellogg S. Booth[1]

[1] *Computer Science*
University of British Columbia
{swindell, po, fisher, ksbooth}@cs.ubc.ca

[2] *Engineering Science*
Simon Fraser University
{dill, ihajshir}@cs.sfu.ca

[3] *Computer Science*
University of Victoria
bdcorrie@uvic.ca

Abstract

Computer-aided design (CAD) and 3D visualization techniques are at the heart of many engineering processes such as aircraft, ship, and automobile design. These visualization tasks require users to navigate or wayfind through complex 3D geometric models consisting of millions of parts. Despite numerous studies, it remains unclear whether large-screen displays improve user performance for such activities. We present a user study comparing standard desktop, immersive room (i.e., CAVE), and wall displays with 3D stereo / head-tracking, and mono / no head-tracking. We observed individual differences between users and found that the presence of contextual structure greatly impacted performance, suggesting that providing structure and developing interaction techniques accommodating a wide range of users yields better performance than focusing on display characteristics alone.

1. Introduction

Structured 3D geometry is commonly used in manufacturing areas such as aircraft, automobile, and ship design. Manufacturing companies are interested in visualizing 3D models to support tasks such as vehicle maintenance and safety training. For example, at the Boeing Company, thousands of aircraft mechanics around the world must inspect, maintain, and repair aircraft on a daily basis. In collaboration with management and development teams at Boeing, we have identified several practical applications of complex 3D visualization. How does a mechanic find a single part in a model 747 aircraft containing over six million parts? How does one part relate to another part? What are good access routes, or paths, to specific parts? Because aircraft are so large and complex, the answers to these questions are critical, but they may not necessarily be so obvious.

We evaluate the importance of the physical display environment in the visualization of complex 3D models. While it is often assumed that large-screen and immersive display environments offer advantages that are unavailable with typical desktop configurations, little empirical evidence supports this claim [4]. In collaboration with the Boeing Company, we compare user performance in navigation and wayfinding tasks in subset 3D models of Boeing 737 and 777 aircraft across different display environments. Task performance is compared across a standard desktop monitor, a tiled wall display, and an immersive room (i.e. CAVE-style) environment [3]. Since virtual reality (VR) rendering capabilities are among the potential advantages of immersive displays over smaller desktop displays, we also examine binocular (stereo)/head-tracked rendering versus monocular (standard)/non-head-tracked rendering on task performance in the tiled wall and CAVE-style configurations.

Our results indicate the presence of large individual performance differences and similarly large differences between the (unstructured) navigation and (structured) wayfinding tasks. This suggests that considering display characteristics, while important, is not the only factor determining user performance. This implies that task structure and interaction style are more likely to benefit from the usability of complex 3D visualization than a singular emphasis on display and rendering type.

1.1. Background and Related Work

Our dialogue with the Boeing Company has revealed that at least five major atomic visualization tasks occur in the design and review activities of aircraft design (personal communication, October, 2002):

1. *Finding* objects in a scene.

2. *Inspecting* objects for discrepancies, conformity, overlaps and interference.

3. *Visually scanning* scenes to optimize for misplaced objects, forgotten objects, and redundancy.

4. *Tracing paths* in scenes, typically through animation, to detect dynamic interference conditions, signal paths through systems, and distortion.

5. *Comparing* objects from multiple release states to understand subjective preference and similarity.

With these tasks in mind, Kasik *et al.* [6] explored the navigation and wayfinding of structured 3D geometry on three small and medium displays: a 20-inch CRT monitor, a 50-inch widescreen plasma display, and a 5-foot hemispherical Elumens VisionStation display. They studied navigation and wayfinding times in various sections of Boeing aircraft models. Their results found that model navigation was fastest for the monitor and slowest for the hemispherical display.

Numerous studies compare the performance of immersive displays and VR input devices (see [11, 13, 16, 17, 18] for a review). Arthur *et al.* [1] found that the adding head-coupled stereoscopic visualization contributed to significant performance and preference improvements for particular perceptual discrimination tasks. Participants greatly preferred the immersion provided with head-tracking, and, to a lesser degree, stereo. Participants viewing a head-coupled stereo display were faster in their experiments and made significantly fewer errors than participants who used a standard display without head-coupling or stereo. Such a performance difference due to increased immersion is consistent with a study reported by Czerwinski *et al.* [5]. They found women took significantly longer to find a target cube in a virtual 3D world using smaller displays when compared to larger displays.

Pausch *et al.* [7] compared user performance when searching for letters (i.e., 'K', 'V', 'Z') in desktop and head-mounted virtual reality (HMD) displays. They found no significant improvements in user performance between those who engaged in target search tasks on a traditional desktop display versus those who engaged in target search tasks under immersive conditions. In their study, there were statistically significant transfer effects between desktop displays and HMDs. They found a "positive transfer" training effect from HMD to desktop display: participants who practiced with an HMD performed better on subsequent search tasks on a desktop display. They also found a reciprocal "negative transfer" training effect from desktop to HMDs. In an extension to their work, Robertson *et al.* [10] found that HMD performance characteristics did not transfer to desktop VR, and they found navigation cues significantly affected performance.

Thus, it is still relatively unclear whether immersive display environments improve user performance in 3D visualization tasks. Often, the performance effects of different displays appear to be overshadowed by other issues such as task type, input device, and previous experience. Since the advantages attributed to large, immersive displays are often abstract, we evaluated their apparent value in the context of an industrially-relevant scenario (i.e., navigation and wayfinding through computer-aided design models of aircraft).

1.2. Objectives

Our collaboration with Boeing generated research interest into whether the physical characteristics of large, immersive-scale displays would confer user performance advantages for 3D navigation and wayfinding tasks. We were given access to subset models of a Boeing 737 and 777 aircraft, consisting of over 500,000 related parts between both models. Akin to the study by Kasik *et al.* [6], we used these datasets to empirically evaluate large-screen and immersive display environments in a manner consistent with the atomic visualization tasks that occur frequently during aircraft design review sessions at Boeing.

We proposed the following experimental hypotheses when comparing navigation and wayfinding performance:

1. Immersive environments should be significantly faster and more accurate compared to desktop environments.

2. Stereo/head-tracked environments should be significantly faster and more accurate compared to monocular/ non-head-tracked rendering.

Combining these two hypotheses, we expected stereo/head-tracked immersive environments to outperform monocular/non-head-tracked immersive environments. At the very least, we expected the larger display configurations, stereo/head-tracked or otherwise, to outperform the standard desktop configuration.

Like Kasik *et al.* [6], we use the terms "Where's Waldo?" and "Hansel & Gretel" to refer to the navigation and wayfinding tasks, respectively. The Where's Waldo task is similar to the books of the same name, where individual scenes have a large, dense population of relatively similar cartoon characters and viewers must find the Waldo character. Our Where's Waldo task required participants to find a specific part by navigating through a large collection of 3D geometric parts. The Hansel & Gretel task is similar to the fairy tale of the same name, where two children leave a trail of shiny pebbles behind them as they travel into the forest. Using these pebbles, they find their way back home. Our Hansel & Gretel task guided participants through a large collection of 3D geometric parts toward a specific part. Participants were subsequently returned to their original location and were asked to find their way back to that part.

1.3. METHODS

To test our formal hypotheses, we refined and extended Kasik *et al.'s* [6] experimental framework. Unlike their study, which focused on small-to-medium size displays, we focused on evaluating very large, immersive environments, such as tiled wall displays and CAVE-style configurations. Also novel, we evaluated the potential advantages of VR-style immersion on these large screens through a performance comparison across stereo/head-tracked and monocular/non-head-tracked visualizations.

As shown in Figure 1, our user study employed a mixed-design involving one between-participants factor (task type) and two within-participants factors (display type and rendering type). Participants were equally divided into two groups: one group was assigned to a Where's Waldo (unstructured navigation) task and the other group was assigned to a Hansel & Gretel (structured wayfinding) task. Within each group, participants were asked to perform their designated task once in each of five different display conditions.

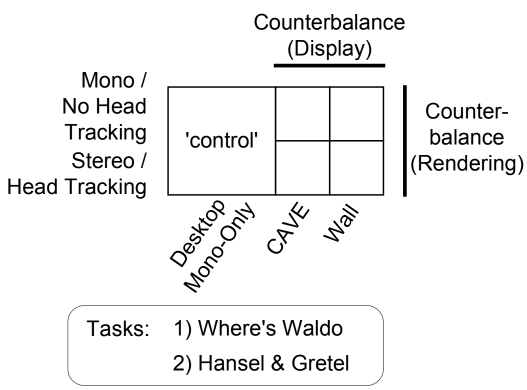

Figure 1. Experimental design. To maintain a manageable study size (24 participants for each task), and focus on effects related to large displays, two within-subjects factors (display type and rendering type) were counterbalanced. All participants started with the desktop condition. Additionally, one between-subjects factor (task) was performed.

Every participant began with the desktop display condition, which served as a control to compare and contrast performance differences between the larger displays. The remaining four wall and CAVE-style conditions were fully counterbalanced such that each participant in a given task group was presented with the larger displays in a unique order. The lack of counterbalancing for the desktop display was deemed necessary, but acceptable for two reasons. First, the addition of a fifth display to counterbalance would have yielded an unmanageable group size (5! = 120 participants per task group, as compared to 4! = 24 participants per task group in the current design). Second, it was clear that all participants would have far more familiarity with a desktop display than with the immersive displays, so having the desktop displays presented first was appropriate given the applied nature of the task and the context of usage where these displays would most likely be employed. An alternate Latin Squares-style design could have been used, but this would have assumed no interaction effects in a statistical analysis between the immersive displays (i.e., learning effects), which could not be assumed given the results of previous work in this area [7, 8, 12].

To prevent tainting our user study with experimental confounds, we carefully controlled for possible system performance artifacts, such as display resolution, refresh rate, and model complexity, which are often ignored in empirical evaluations of display environments. We also carefully controlled for possible participant performance artifacts, such as depth perception using a clinically validated diagnostic screening test known as the Titmus Stereo Fly test [14]. This ensured participant performance across different display configurations would not merely be the result of rendering and physiological artifacts, but could in fact be attributed to differences in physical display characteristics.

The Titmus Stereo Fly test required participants to wear polaroid glasses, then look at images on vectograph plates. If an image on the vectograph appeared to 'pop out' towards the participant (i.e., a 3D effect), then the participant passed that portion of the test. Each participant in our user studies performed both gross and fine depth acuity components of the Titmus Stereo Fly test. The gross acuity component required participants to observe an image of a fly containing a disparity of 1 degree of arc. The fine depth acuity component required participants to look at Writ rings ranging from 800 – 40 seconds of arc (i.e., look at a cluster of 4 rings, and test if the participant could discern which ring 'popped out' of the plate).

1.4. Participants

A total of 48 participants, divided into two groups, took part in this user study. 24 participants (14 male, 10 female) with a mean age of 24.8 (SD = 8.3) years took part in a Where's Waldo task. The other 24 participants (14 male, 10 female) with a mean age of 26.2 (SD = 7.2) years took part in a Hansel & Gretel task. All participants had normal or corrected-to-normal vision. All participants were pre-screened to have binocular depth acuity of at least 200 arc seconds. The Writ ring component of the Stereo Fly test evaluated participants to have a group mean stereo acuity of 44.8 arc seconds (*SD* = 16.6).

1.5. Apparatus

Figures 2 and 3 illustrate the display apparatus used during the user study. Our accompanying video gives additional apparatus details (contact an author for a copy).

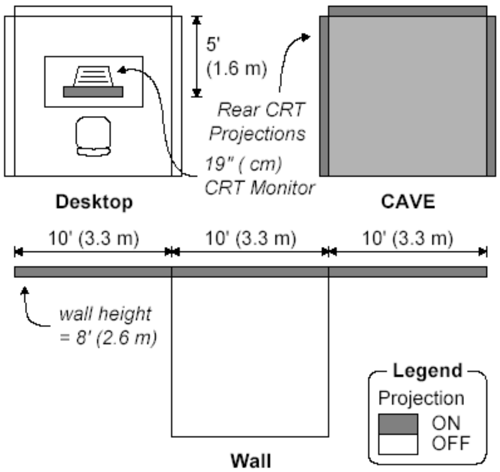

Figure 2. Plan view of the user study apparatus. Clockwise from the top-left, the desktop, CAVE-style room, and tiled wall configurations.

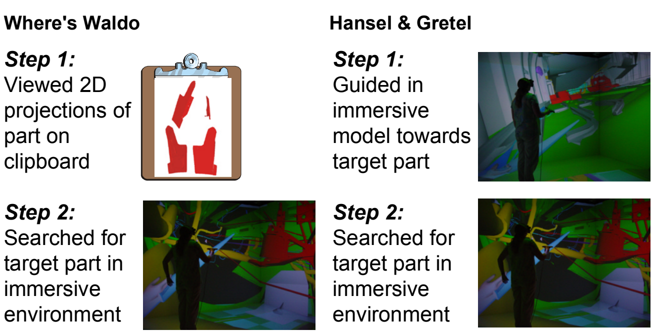

Figure 3. Summary of Where's Waldo and Hansel & Gretel tasks.

The user study was conducted inside a reconfigurable immersive display environment controlled by an SGI Onyx 3200/InfiniteReality 3 computer with 4 MIPS R12000 processors and two rendering pipes. The immersive display was configurable into a panoramic, three-screen tiled wall display (bottom of Figure 2) or into a closed, four-screen CAVE-style immersive room (upper right-hand of Figure 2). In the closed configuration, three screens surrounded the participant, and one screen was oriented to project onto the floor. In the open, panoramic configuration, the three wall screens were oriented to create a wide, flat surface (floor projection was off).

The three walls and floor were rear and front projected, respectively, with four Electrohome 9500 CRT projectors. The desktop monitor used in this study was a 19" CRT monitor and was always situated in the centre of the immersive room during the desktop display condition. Incidentally, this allowed us to maintain similar visual angles across all viewing conditions, despite changes in the display configuration Even though participants were free to move around in the CAVE and wall configurations, the experimenters observed that participants generally stayed near the center of the immersive display. Thus, similar viewing angles were maintained. The resolution of all displays was 1024x768 pixels at a constant refresh rate of 96 Hz at all times. Frame rates were locked at a consistent 10 frames per second, and ambient illumination was dimmed to near-darkness for all tasks. For desktop conditions, participants sat in front of the desktop monitor. For wall and CAVE-style conditions, participants stood, but could roam freely inside the "floor" area of the immersive environment.

A six degree-of-freedom Intersense IS900 ultrasonic tracker provided head-tracking and wand (i.e., "bat mouse") input. Participants wore a head tracker and CrystalEyes LCD shutter glasses at all times (i.e., even when graphics were rendered in monocular format and were not head-tracked). Participants always (including desktop conditions) navigated and interacted with 3D models using the wand. To navigate, participants used the wand's integrated joystick to move forward and backward, and to rotate left and right. Pointing the wand in a specific direction while moving the joystick forward moved participants in the pointed direction.

Model rendering and navigation software to display the Boeing aircraft datasets were written in C++ using OpenGL Performer and CAVELib. The software displayed the Boeing aircraft models and recorded participant performance data. To maintain a viable interactive frame rate, the Boeing datasets were preprocessed to a 90% level of detail (LOD) using commercial data reduction software [9]. This LOD was maintained at all rendering distances, ensuring that parts maintained the same geometry regardless of virtual distance from viewer to parts.

1.6. Procedure

Participants took part in single 60-90 minute sessions with either Where's Waldo or Hansel & Gretel tasks (see Figure 3) in each of the five possible display conditions. Participants were initially pre-screened through an application of the Titmus Stereo Fly test. For each display condition, participants were presented with a partial model from the Boeing dataset, through which they were asked to navigate and find a prescribed part (for Where's Waldo) or find their way back to a prescribed part (for Hansel & Gretel). Five different model subsets (see Figure 4) and prescribed part pairs were chosen, each consisting of approximately 500 similarly sized parts. All prescribed

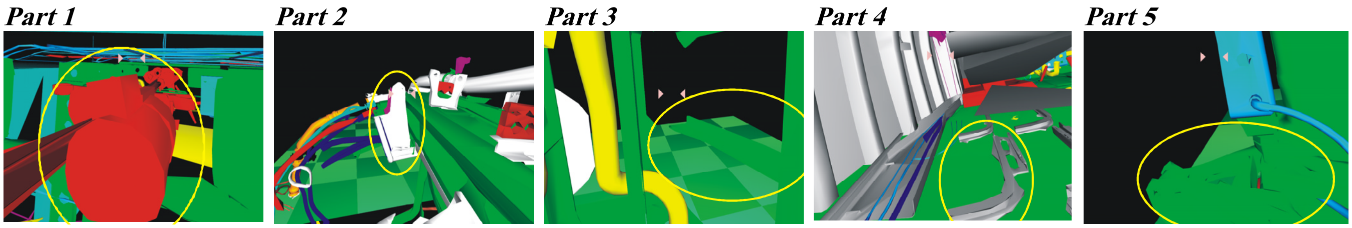

Figure 4. Five parts used for both tasks: Where's Waldo and Hansel & Gretel. Part 1 was always used for the desktop display condition, and parts 2 - 5 were used for the wall and CAVE conditions. Yellow ellipses highlight target parts in the Figure.

parts were of roughly the same size. Every display condition was presented with a unique model/part pair such that no pair was used twice in a given participant session. Participants were given a chance to take a short break between display conditions while the experimenter set up the display apparatus for the next presentation. After the five display conditions, participants were asked to fill out a computer experience and display preference questionnaire. An experimenter was always present to answer questions and to facilitate progress during the session.

1.6.1. Training Prior to completing each display condition, participants were presented with a small set of parts from the Boeing dataset (designated as "practice" parts), which were not used during any of the five experimental display conditions. Participants were given two to three minutes to become familiar with navigating through the display configuration using the tracking wand. In the stereo/head-tracked conditions, this also gave participants an opportunity to become comfortable with dynamically changing viewpoints based on head position.

1.6.2. Where's Waldo? Participants were placed at a common start location in the 3D model, and were centered physically in each of the five display conditions (see Figure 2). They were given an illuminated clipboard holding an 8" x 11" piece of paper with a target 3D part in four different views. Participants were instructed to look for the target 3D part in the 3D virtual model presented to them. Participants were informed they would have ten minutes to successfully locate the target part. Once participants viewed the target part on the clipboard and were ready to start, they gave a verbal cue to the experimenter to begin. The display switched from a solid black background into a 3D virtual model and a ten minute countdown timer began.

Participants were then allowed to navigate through the model in an attempt to find the target part. If they were able to successfully find their target part, the participants gave a second verbal cue to the experimenter indicating they had done so. At this point, the experimenter would stop the timer and record the final completion time. If a participant failed to find the target part within ten minutes, the experimenter told the participant that the task was over for that particular display condition and recorded a completion time of 10 minutes. If the participant found an incorrect part, the completion time was still recorded, but an error was recorded. Participants were not told whether they had correctly identified the part or not.

1.6.3. Hansel and Gretel Participants followed the same procedure as the Where's Waldo task described above, with the following exception. Instead of receiving an illuminated clipboard with a piece of paper indicating a target part, participants were initially guided from a nominal start position to a given part using an indirect, but continuous, pre-recorded motion path. The viewing position and orientation in the model would change until a target part was clearly visible. Participants could then see the target part and its surrounding context until they were ready to search for the part. It is important to note that the path from the starting point to the target part did not follow a direct, shortest path. Thus, participants would be unable to retrace their path simply by employing some simple, linear movement. Instead, they would have to rely on remembered navigation cues to find their way back to the target part. Wayfinding and experimenter recording was accomplished in the same fashion as the Where's Waldo task described above.

2. Results

Figure 5 presents aggregate performance across the different display conditions in terms of task completion (i.e., how many participants were able to successfully find prescribed parts) and completion time (i.e., of those who successfully found prescribed parts, how long it took them to do so). We generally observed that individual participants varied widely in their ability to complete their respective tasks, and that display condition appeared to have little influence on task completion and overall completion time. Noticeable differences in task completion rates and completion times between the unstructured Where's Waldo task and the structured Hansel & Gretel task were also observed.

To formally test our experimental hypotheses and these observations, we used a variety of parametric and non-parametric techniques to analyze collected participant data. These analyses are consistent with alternative methods of statistical inference, which emphasize the importance of characterizing variance patterns, individual differences, and other inferential measures that do not depend on a null hypothesis [2, 15]. Chi-Square and Cochran's Q tests were used to compare task completion performance within and between the different display conditions. For the Chi-Square tests, we tested the hypothesis that completion rates would be significantly above or below a completion threshold of 50%. Chi-Square tests were also used to compare each of the immersive display conditions against the completion rate of the desktop condition (i.e., for each immersive display condition, the hypothesis was that each condition performed significantly better or worse than the desktop condition). Two-way within-subjects ANOVAs compared completion times between independent factors of display type (tiled wall and CAVE) and rendering type (stereo/head-tracked and mono/non-head-tracked).

Individual participants were ranked based on relative performance, and performance concordance was analyzed using Kendall's W test. Relationships between sex, age, and completion rates were tested using a non-parametric Kendall's Tau-B correlation test. Correlations for sex, age, and completion times were tested using a Pearson correlation analysis. Preference data from the participant questionnaires were statistically evaluated using a Friedman test of ranked preference. Performance differences between task type (Where's Waldo versus Hansel & Gretel) were analyzed using an independent samples *t*-test.

2.1. Task Completion

The top row of Figure 5 illustrates how participants successfully completed an average of two of the five display conditions in the Where's Waldo task and four of the five display conditions in the Hansel & Gretel task, suggesting a factor of two difference in performance between the unstructured Where's Waldo navigation task and the structured Hansel & Gretel wayfinding task. A non-parametric Cochran's Q analysis test for statistical differences in task completion rates (i.e., number of people who successfully found their target parts) failed to find any differences between the five conditions in both the Where's Waldo [Q(4, N = 24) = 2.246, p = 0.691] and Hansel & Gretel task types [Q(4, N = 24) = 5.826, p = 0.213].

Chi-Square tests for differences in task completion within the display conditions found the tiled wall conditions (both stereo/head-tracked and mono/non-head-tracked) performed significantly worse than the "expected" 50% completion threshold [$\chi^2(1, N = 24) = 4.167$, p = 0.041]) in the Where's Waldo task. An identical Chi-Square test for the Hansel & Gretel task found the desktop display condition, tiled wall (mono/non-head-tracked), and immersive CAVE-style room (stereo/head-tracked) had better participant performance than the "expected" 50% completion threshold [Desktop: $\chi^2(1, N{=}24) = 4.167$, p = 0.041; Wall (mono/non-head-tracked): χ^2[(1, N=24) = 10.667, p = 0.001; CAVE: $\chi^2(1, N{=}24) = 13.500$, $p < 0.001$].

We attribute the inconsistencies in statistical significance to characteristic individual differences between participants, which would mean that performance was largely dictated by personal human factors instead of just display factors. To examine individual differences, we used a statistical measurement independent of rejection of a null hypothesis. The presence of large individual differences between participants was observed by a Kendall's W test, which generates a normalized coefficient of concordance between 0 and 1. This coefficient is typically interpreted as the degree to which participants agreed (or were consistent) with one another. On our ranked participant data between display conditions, the Kendall's W coefficient value was extremely low for both the Where's Waldo (W = 0.008) and the Hansel & Gretel (W = 0.033) tasks. Thus, participants varied largely in the specific display conditions that they successfully completed – suggesting that personal factors affected completion time more than display factors.

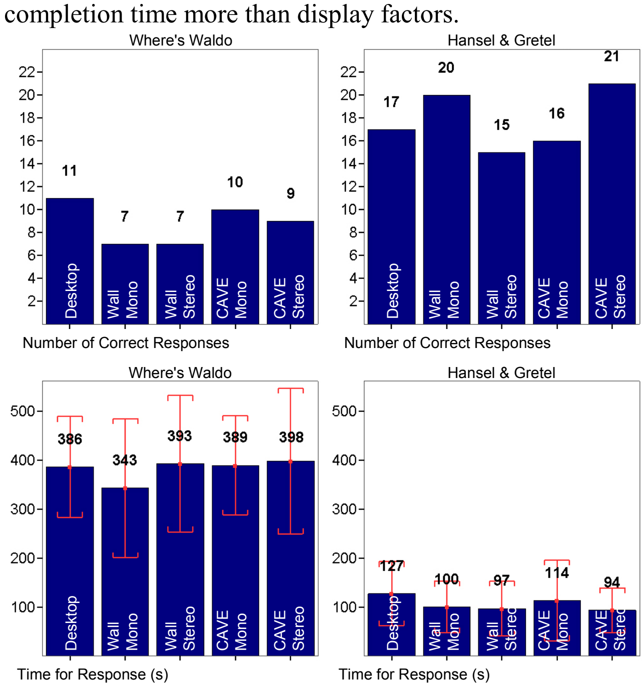

Figure 5. Task completion rates by number of participants (top row) and completion times in seconds (bottom row) for participants who successfully found target parts in each display condition. Notice the relative consistency in completion rates and times across displays, but the large differences between the unstructured Where's Waldo navigation task and the structured Hansel & Gretel wayfinding task.

2.2. Completion Time

Figure 5 indicates a large participant completion time variation within individual display conditions, although it was clear the Hansel & Gretel participant group was consistently faster than the Where's Waldo group. The two-way within-subjects ANOVAs testing completion time against display type (wall or immersive room) and rendering type (stereo/head-tracked or mono/non-head-tracked) for each task type yielded no significant main effects or interactions [all models had $F(1, 23) < 2.308$, $p > 0.142$]. A subsequent analysis of accounted variance using η^2 values was done to determine whether this lack of significance was due to small effect size or because there really was no difference.

This is consistent with a discussion of statistical relevance by Cohen [2], who suggests using effect size as a metric to infer the underlying presence or non-presence of differences. Cohen suggests categorizing .01, .059, and .138 to be small, medium, and large effect sizes respectively. Because the η^2 values were greater than 0.076 for the independent factor of display type (between medium and large effect size) in both the Where's Waldo and Hansel & Gretel tasks, we have some measure of certainty that if a statistically significant difference did exist, our user study would have detected it, especially given the number of participants who completed our study. Thus, since we found no such differences between display conditions with relatively high effect sizes, display type appears to have had little effect on participant performance. This is also consistent with the observed individual differences in the task completion analyses described above.

2.3. Between Tasks and Other Effects

An independent samples *t*-test confirmed that there was a statistically significant difference in relative performance between the Where's Waldo and Hansel & Gretel tasks [$t(46) = 6.339$, $p < 0.001$]. An evaluation of the post-session preference questionnaires indicated that a majority of the participants from both task groups preferred using the more immersive displays compared to the desktop displays, but largely disagreed on how easy they were to use to complete their assigned tasks. A Friedman test of ranked preference failed to show that any of the display conditions was easier than any other [Where's Waldo: $\chi^2(2, N = 24) = 0.250$ $p = 0.882$; Hansel & Gretel: [$\chi^2(2, N = 24) = 5.333$, $p = 0.069$]. No statistically significant correlations were observed between sex, age, and task performance.

As an additional measure of assurance that our results were not confounded by experimental factors, no order effects were observed and each of the five subset model/prescribed part pairs were successfully completed near-equal numbers of times, so it was not the case that some display conditions and some models were more difficult than others.

Qualitative observations by the experimenters support the hypothesis that display type had little effect on user performance. Participants tended to navigate then focus on a particular area of the large displays. For example, in the immersive CAVE-style configuration, several participants would initially use the centre of the left screen as their "working area." After some period of time, they would switch "working areas" to another area of the display such as the lower-left corner of the centre screen. Thus, it was rarely the case that participants would scan large areas of the display. This may also help explain why participant performance on the desktop display was more consistent than the more immersive displays: the desktop display did not require participants to distribute their time over a very wide area of visual space.

3. DISCUSSION AND FUTURE WORK

Our results are largely consistent with the results collected by Kasik *et al.* [6], who failed to find any statistically significant relationships between their tested display conditions. Since our study provided greater rigor in terms of limiting possible confounds and controlling experimental factors, we lend further credence to the possibility that other influences aside from display characteristics and rendering type may be important in improving user performance for 3D visualization tasks involving complex, structured geometric models. As the marked difference between the Where's Waldo and Hansel & Gretel task shows, task structure and navigational cues may have a far greater influence on performance than any increase in immersion and display size.

However, this does not imply that large and immersive display characteristics are entirely unimportant. Other benefits, such as a feeling of presence in the larger displays, may be important but not lead towards short term performance benefits. In collaborative scenarios, where several people may be working together, the scale of a large-screen display allows each person to more easily share the same display space, which simply cannot be accomplished on a small desktop monitor. Moreover, such displays can be useful in situations where graphical data are presented in showrooms and exhibits for public display.

Choice of user interface may be a greater issue in wayfinding and navigation performance than display size or rending type. For example, simply looking at the back of a target object can be difficult using a device such as a bat mouse because the world would need to pivot about the target object instead of the positionally tracked user. During common, everyday interactions with physical objects, such mode changes are so natural that we rarely think about them. An excellent interaction methodology for navigation and wayfinding in 3D models has yet to be developed (either for a desktop or large immersive display). Future work could explore effects of different user interfaces with experiments similar to ours. For example, comparing a bat mouse to a person wearing tracked gloves and walking on a multi-dimensional treadmill might yield interesting results.

The observed individual performance differences between participants also merits future work. Identifying those characteristics of participants who were most successful in completing the navigation and wayfinding tasks would be useful in further improving 3D visualization for tasks such as the ones outlined by Boeing. Moreover, those individual differences could be important for future visualization applications where it may be possible to accommodate for these kinds of differences by creating personalized interfaces tailored specifically for an individual's physiology and behavior.

4. CONCLUSION

We presented a user study that investigated the role of large screen and immersive display characteristics in improving task performance for 3D visualization tasks involving navigation and wayfinding. Using a comparative framework, results suggest that factors such as display type and rendering style were less important than factors such as task structure, navigational context, and individual differences. Thus, more powerful improvements in the usability of 3D visualization tools may come in accommodating the widest range of users possible and structuring complex 3D environments to make scene context as evident as possible.

5. ACKNOWLEDGEMENTS

We thank Dave Kasik and The Boeing Company for their comments and 3D CAD models, and Dr. Hong-Yee Wong, Todd Zimmerman, and the New Media Innovation Centre for use of their virtual reality facilities and technical support. The research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC).

6. REFERENCES

[1] Arthur, K. W., Booth, K. S., & Ware, C. "3D task performance in fish tank virtual worlds." *ACM Transactions on Information Systems 11*(3), 1993, pp. 239-265.

[2] Cohen, J. "Eta-squared and partial eta-squared in communication science." *Human Communication Research, 28,* Oxford Journals, 1973, pp. 473-490.

[3] Cruz-Neira, C., Sandin, D., DeFanti, T., Kenyon, R., & Hart, J. "The CAVE: Audio Visual Experience Automatic Virtual Environment." *Communications of the ACM 35*(6), 1992, pp. 65-72.

[4] Czwerinski, M., Smith, G., Regan, T., Meyers, B., Robertston, G., & Starkweather, G. "Toward Characterizing the Productivity Benefits of Very Large Displays." In Proceedings of *International Conference on Human Computer Interaction* (*INTERACT),* IFIP Press, 2003, pp. 252-255.

[5] Czerwinski, M., Tan, D. S., & Robertson, G. G. "Women take a wider view." In Proceedings of *Conference on Human Factors and Computing Systems (CHI), CHI Letters 4*(1), ACM Press, 2002, pp. 195-202.

[6] Kasik, D. J., Troy, J. J., Amorosi, S. R., Murray, M. O., & Swamy, S. N. "Evaluating Graphics Displays for Complex 3D Models." *Computer Graphics and Applications, 22*(3), IEEE Press, 2002, pp. 56-64.

[7] Pausch, R., Proffitt, D., & Williams, G. "Quantifying immersion in virtual reality." In Proceedings of *International Conference on Computer Graphics and Interactive Techniques* (*SIGGRAPH)*, ACM Press, 1997, pp. 13-18.

[8] Pausch, R., Shackelford, M. A., & Proffitt, D. "A User Study Comparing Head-Mounted and Stationary Displays." In Proceedings of *Symposium on Research Frontiers in Virtual Reality*, IEEE Press, 1993, pp. 41-45.

[9] Rational Reducer "Systems in Motion AS". *http://www.rational-reducer.com*, Accessed October, 2003.

[10] Robertson, G. G., Czerwinski, M., & van Dantzich, M. "Immersion in desktop virtual reality." In Proceedings of *Symposium on User Interface Software and Technology* (*UIST)*, ACM Press, 1997, pp. 11-19.

[11] Slay, H., Thomas, B., & Vernik, R. "Tangible User Interaction Using Augmented Reality." In Proceedings of *Australasian Conference on User Interfaces (AUIC)*, ACM Press, 2002, pp. 13-20.

[12] Tan, D. S., Gergle, D., Scupelli, P. G., & Pausch, R. "With Similar Visual Angles, Larger Displays Improve Spatial Performance." In Proceedings of *Conference on Human Factors in Computing Systems (CHI), CHI Letters 5*(1), 2003, pp. 217-224.

[13] Tan, D. S., Robertson, G. G., & Czerwinski, M. "Exploring 3D Navigation: Combining Speed-coupled Flying with Orbiting." In Proceedings of *Conference on Human Factors and Computing System (CHI), CHI Letters 3*(1), ACM Press, 2001, pp. 418-425.

[14] Somers, W. W. & Hamilton, M. J. "Estimation of the stereoscopic threshold utilizing perceived depth." *Ophthalmic and Physiological Optics. 4*(3), 1984, pp. 245-250.

[15] Vicente, K. J., and Torenvliet, G. L. "The Earth is spherical ($p < 0.05$): alternative methods of statistical inference." *Theoretical Issues in Ergonomic Science 1*(3), 2000, pp. 248-271.

[16] Wanger, L. R., Ferwerda, J. A., & Greenberg, D. P. "Perceiving Spatial Relationships in Computer-Generated Images." *Computer Graphics and Applications, 12*(3), IEEE Press, 1992, pp. 44-58.

[17] Ware, C., & Fleet, D. "Context sensitive flying interface." In Proceedings of *Symposium on Interactive 3D Graphics.(I3D)*, ACM Press, pp. 127-130.

[18] Zhai, S., & Milgram, P. "Human Performance Evaluation of Manipulation Schemes in Virtual Reality." In Proceedings of *Virtual Reality Annual International Symposium,* IEEE Press, 1993, pp. 155-161.