The Effects of Display Fidelity, Visual Complexity, and Task Scope on Spatial Understanding of 3D Graphs

Felipe Bacim*, Eric Ragan, Siroberto Scerbo, Nicholas F. Polys
Center for Human-Computer Interaction
Virginia Tech

Mehdi Setareh
School of Architecture and Design
Virginia Tech

Brett D. Jones
Department of Learning Sciences and Technologies
Virginia Tech

ABSTRACT

Immersive display features can improve performance for tasks involving 3D, but determining which types of spatial analysis tasks are affected by immersive display features for different applications is not simple. This research adds to the knowledge of how the level of display fidelity (i.e., the realism provided by the display output) affects task performance for a variety of 3D spatial understanding tasks. In this study, we control visual display fidelity with the combination of stereoscopy, head-based rendering, and display area and study performance analysis of 3D graphs. Through a controlled study, we evaluated the relationship among display fidelity, visual complexity, task scope, and a user's personal spatial ability. Over a variety of task types, our results show significantly better overall task performance with higher display fidelity. We also found that visual complexity and task scope affect speed, with higher levels of either type of complexity leading to slower performance. These results show the importance of considering multiple factors when calculating the overall difficulty and complexity of a spatial task, and they suggest that visual clutter makes a greater impact on speed than correctness. Further, the study of different task types suggest enhanced virtual reality displays offer more benefits for spatial search and fine-grained component distinction, but may provide little gain than for sense of scale or size comparison.

Keywords: Spatial understanding, benefits of immersion, display fidelity, virtual reality.

Index Terms: I.3.7 [Three-Dimensional Graphics & Realism]: Virtual Reality

1 INTRODUCTION

Virtual reality (VR) and immersive technologies provide display features to support improved spatial perception. For example, stereoscopy provides binocular disparity to aid depth perception, head tracking makes it possible to use familiar head movements to observe motion parallax, and large display areas make it possible to use natural physical head and body movements (rather than virtual camera adjustments) to analyze 3D content from different vantage points. Because of this, immersive displays are often used to help understand 3D visualizations for a variety of purposes, such as scientific visualization (e.g., [10]), social networks analysis (e.g., [8]), engineering (e.g., [21]), and architectural design (e.g., [2]).

In addition, many studies have shown that the addition of immersive display features can improve performance for tasks involving 3D visualization and spatial understanding (e.g., [5, 15, 19]). However, immersive technology is generally more expensive than common computer displays. VR systems also often require large physical spaces and lack the portability of simpler displays. To help researchers and practitioners of visualization and VR to balance the tradeoffs between the costs and benefits associated with immersive technology, our research adds to the knowledge of how the level of *display fidelity* (that is, the realism provided by a display's output [11]) affects task performance for a variety of spatial investigation tasks. The display fidelity of a VR system can be seen as a combination of all its display properties, which include visual components such as field of view (FOV), screen resolution, frame rate, and stereoscopy [4, 11].

However, the evaluation of how display fidelity affects spatial understanding tasks does not depend only on these display characteristics. Different domains and applications require different types of investigation and spatial judgments, and the difficulty of the task depends on a combination of multiple characteristics of the visualization. For example, the amount of visual clutter, lighting and rendering choices, the number of components that need to be considered, the contrast among visual components, and the scale of the structural components of interest all affect the overall complexity of the task and visualization.

While other studies have investigated the effects of display fidelity on spatial analysis tasks (e.g., [1, 12, 15, 19, 20]), our work is novel in that it isolates and controls several additional factors that could impact these effects. To help make our findings applicable to a variety of spatial understanding tasks in different applications, we conducted our study using abstract mathematical graphs (undirected node graphs; see Figure 1) as the testbed for the experimental tasks. Using these graphs, we performed a controlled study in a CAVE-like display to evaluate how the level of display fidelity (specifically, the combination of display size, stereo, and head tracking) affects task performance.

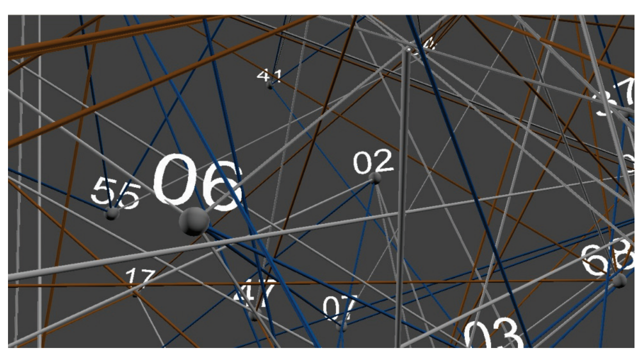


Figure 1: View within a 3D graph used for the experiment.

* e-mail: fbacim@vt.edu

Additionally—and perhaps more importantly—we also controlled for two factors contributing to the difficulty of the spatial task: visual complexity and scope of the task. We define *visual complexity* as the amount of visual clutter within the environment. We were able to differentiate this visual complexity from the *task scope* for each graph and task. In the spatial context, the task scope relates to the amount of content being analyzed and the distribution of the spatial structures. In our testbed, task scope was determined by the number of nodes, edges, and connections that must be considered for the task, as well as their spatial and mathematical proximity to each other. Though a few previous studies have partially controlled for mathematically complexity while studying spatial understanding [9, 19, 20], these studies had difficulty separating visual complexity and task scope.

By individually controlling display fidelity, visual complexity, and task scope for four types of tasks (intersection search, path following, connection identification, and edge length comparison), our experiment was able to not only test for effects of the individual variables, but also to detect interactions among them. Further, our design accounted for the spatial abilities of the individual users, which made it possible to study if these personal differences influenced any observed effects of the primary factors.

Finally, our main contributions can be summarized as follows.

- We present a user study that separates visual complexity and task scope from the overall task complexity, and demonstrates that these factors individually affect performance for spatial analysis tasks.
- We add to the body of evidence demonstrating that using a higher level of visual display fidelity can result in performance improvements. The results also suggest that these benefits persist over a range of spatial and visual types of complexity.
- We present results for abstract tasks that can inform the design of different applications.

2 Related Work

This work builds upon many previous studies on how display features of immersive VR systems affect task performance for spatial analysis tasks.

2.1 Display Fidelity and Spatial Understanding

In evaluating how immersive display components affect task performance, this work follows the immersion framework presented by Bowman and McMahan [4]. By this model, rather than categorizing a display as purely *immersive* or *non-immersive*, the system's level of display fidelity can fall anywhere along a multi-dimensional continuum of display components and the level of realism each supports [4, 11]. McMahan et al. [11] define *display fidelity* as "the objective degree of exactness with which real-world sensory stimuli are reproduced." Therefore, display fidelity is determined by the display's physical characteristics, such as FOV, resolution, latency, and stereoscopy. The closer the system simulates the real world, the higher the display fidelity. For example, a system that has a higher FOV and stereoscopy, like a CAVE or head-mounted display (HMD), has higher display fidelity than a system with lower FOV and no stereoscopy, like a typical monitor or large screen display. We note that previous research used the term *immersion* in place of *display fidelity* (e.g., [4, 16, 17]). However, because *immersion* is often confused with the concepts of engagement or presence (i.e., the feeling of being in the virtual world, rather than simply using a computer system [16]), recent publications have opted for *display fidelity* for clarity in the relation to the realism provided by the display [11, 12].

Many studies have investigated how varying levels of display fidelity can affect user performance for a variety of spatial understanding and analysis tasks. For example, separate research by Ruddle et al. [14] and by Chance et al. [5] involving egocentric spatial orientation found that the addition of head tracking improved user orientation in first-person navigation tasks. In another example, Arns et al. [1] compared desktop workstation and a CAVE display (with four screens and stereo) for a visual statistical analysis task. They found that the CAVE display allowed for better understanding of the statistical data when identifying structural features. Studying a task involving path-planning for oil-wells, Gruchalla [7] required participants to edit paths within complex branching 3D structures. In their comparison between a stereo-desktop display and a CAVE-like display, the researchers found that increasing display area and the addition of head tracking significantly reduced task time.

Other researchers have studied how display fidelity affects performance on variety of tasks within the same environment. For example, Schuchardt and Bowman [15] compared high-fidelity (four screens with stereo and head tracking) to low-fidelity (one screen with no stereo and no head tracking) to determine how display fidelity affects spatial understanding of complex underground cave layouts. This study evaluated performance on a number of tasks such as: structural feature search, identification search, 3D structure-to-surface projection, angle measurement, and size comparison. This work is relevant to our research because the study found significant performance improvements for some of the tasks (intersection search, structural feature search, and size comparison), but not for others (3D projection and angle measurement). The findings clearly suggest that the effects of display fidelity on task performance depend on the specifics of the spatial investigation task. For this reason, and to support generalization of any findings to other application domains, we also chose a variety of task types for our study of the effects of display fidelity and spatial complexity.

2.2 Display Fidelity and Graph Analysis

A number of prior studies have not only studied the effects of display components on spatial understanding tasks, but have also focused specifically on graph analysis tasks. For example, Ware et al. [18] found that the addition of stereo and head tracking significantly improved user accuracy for tracing paths in 3D tree graphs. In an experiment using network graphs, Henry and Polys [9] studied how display fidelity and navigation modes affect different types of spatial knowledge. They compared a four-wall CAVE to a single-wall display, both with stereo and head tracking, and compared egocentric to exocentric navigation. Four task types were used: counting the number of nodes connected to a given node, searching for node with highest number of connections, counting the number of nodes in a path connecting two nodes, and counting the number of nodes of a specific type. Overall, display fidelity did not significantly affect speed. However, participants tended to be slower when using a high-fidelity display for counting the number of nodes between two nodes. While this study is similar to our own, it did not take visual complexity or task scope into consideration. It is also interesting to note that the negative effect of additional display area on spatial task performance, which is particularly surprising, since other studies found additional display area to have the greatest positive effect involving spatial understanding [12, 13].

Ware and Franck [19] also conducted studies with graph analysis. They evaluated the effects of stereo and head tracking on path tracing using graphs with varying numbers of nodes and edges. Overall, they found that higher levels of display fidelity improved accuracy, but not speed. In addition, the authors also found that using a larger graph size resulted in reduced speed. In a similar study, Ware and Mitchel [20] explored the effects of stereo, 2D and 3D graph layout, and 2D and 3D rendering with different graph sizes. While the task was the same as in the previous study (path tracing), participants only had five seconds of viewing time, and the display was blacked out when they had to answer. The results showed a positive effect of stereo on both time and accuracy. While both studies are similar to our own, visual complexity was not varied independently of the mathematical complexity of the graphs. For example, while visual complexity increased by overlaying graphs with similar amounts of nodes and connections, the number of connections of the nodes used for the tasks varied. Our study removes the confound between visual complexity and task scope by controlling them independently, and also evaluates a wider range of tasks.

3 EXPERIMENT

We conducted a controlled experiment to investigate the relationship among immersive display fidelity, visual complexity, and spatial ability with regard to spatial investigation of 3D node graphs.

3.1 Hypotheses and Goals

The goals of this research were to study the effects of visual display fidelity on task performance for 3D spatial understanding tasks. We also sought to test whether these effects are influenced by task scope and the visual complexity. Further, we aimed to investigate whether the user's spatial ability had any impact on these potential effects.

Based on previous studies of the effects of immersive display features on spatial understanding tasks, (e.g., [6, 12, 15, 19]), we hypothesized that higher display fidelity (specifically, increased display area and the addition of stereo and head tracking) would lead to improved performance on 3D spatial inspections. We also expected an interaction effect between display fidelity and the level of visual complexity, where the higher display fidelity would yield greater advantages for visualizations with greater visual complexity. Similarly, we expected to see greater benefits of the enhanced spatial cues of higher display fidelity for more advanced task scope.

Furthermore, we hypothesized that these effects and interactions would persist regardless of users' levels of spatial ability, though we did predict that participants with better spatial abilities would perform better on the spatial tasks.

3.2 Task

To test our hypotheses, we chose several spatial investigation tasks using 3D graphs. Figure 1 shows a screen shot of the graph visualization. The graphs edges were blue, orange, and gray to make it easier to differentiate between edges, and lighting was used to provide depth cues. These colors were assigned randomly. Graph nodes were represented with gray spheres with white billboarded numeric labels. The graphs were shown against a darker gray background (preliminary tests found that stronger contrasts strained the subjects' eyes).

Four separate task types were chosen for the experiment: intersection search, path following, connection identification, and length comparison. For the *intersection search* task, participants were asked to find two edges that were clipping through each other, and to then name those edges using the node numbers. For the *path following* task, participants were given the numbers for two nodes in the graph. Participants then had to find a path that would connect the two given nodes and list the intermediate nodes along the path. In the next task, *connection identification*, participants were given the name of a single node and shown its location. Participants were asked to give the names of all the nodes that were directly connected to the original node. Finally, in the *length comparison* task, the colors of two edges within the graph were changed, so that one was bright red and the other was bright yellow. Participants had to determine which edge was longer. Participants reported their answers for all tasks verbally.

To allow us to study effects due to task, we designed simple and advanced versions of each task type. Participants always completed both versions sequentially, starting with the advanced version of the task in each graph. This way, the advanced-scope version was always done without prior exposure of the graph. For intersection search, participants were required to find two intersections. We considered the search for the first intersection to be more advanced due to a larger search area and a lack of familiarity with the graph. For the path following task, the shortest path between the two given nodes was longer (i.e., contained more nodes) in the advanced versions than in the simple versions of the task. For connection identification, the more advanced versions had greater numbers of connected nodes. Lastly, for the length comparisons, the advanced versions of the tasks had the two target edges further away from each other and in different orientations, which made them more difficult to compare to each other and required participants to navigate more.

Participants were instructed to prioritize completing the tasks successfully, but were also asked to try to give their answers quickly. A time limit of five minutes was enforced for all tasks.

3.3 Apparatus

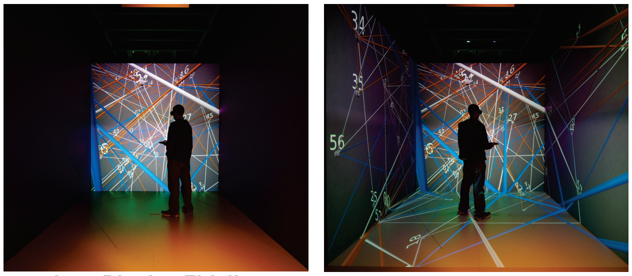
The display system used for the experiment was a VisBox VisCube, a CAVE-like display with four screens (three rear-projected display walls and a front-projected floor). Each screen measured 10x10 feet with graphics shown with 1920x1920 pixel resolution. The display supported passive stereoscopy through Infitec stereo glasses. Wireless head tracking was supported with an Intersense IS-900 motion tracking system. For interaction with the virtual content, participants used a wireless handheld wand controller, which was tracked in six degrees of freedom. Participants could move (translate) through the visualizations by physically pointing the wand in the intended direction of travel, and pushing forward or backward on the wand's joystick. Participants could also move the joystick to the left or right to rotate about the vertical axis of the display.

All participants used the same VisCube display, but the display properties were adjusted based on the controlled level of display fidelity. Participants in the high-fidelity condition used all four screens, view the imagery in stereo, and had head-tracking enabled. Participants in the low-fidelity condition did not have stereo or head tracking enabled, and only had one screen (the front) available to complete the tasks. Figure 2 illustrates both setups. To control the field of view in both levels of display fidelity, participants in the low-fidelity conditions wore the same Infitec glasses as those in the high-fidelity conditions.

The experiment's graph visualizations were rendered from X3D using the InstantReality Simple Avalon Player. Frame rate varied between 45 and 60 frames per second.

3.4 Experimental Design

We ran a 2x3x2x4x2 mixed experimental design for *display fidelity* (high and low), *visual complexity* (low, medium, and high), participant *spatial ability* (low and high), *task type* (intersection search, path following, connection identification, and length comparison), and *task scope* (simple and advanced), respectively. Display fidelity and spatial ability were studied between subjects, while visual complexity, task type, and task scope were varied within subjects.



Low Display Fidelity **High Display Fidelity**

Figure 2: Two display-fidelity conditions used in the experiment.

The level of display fidelity determined the display features and interaction methods used to complete the tasks. Two levels of display fidelity were used: high and low. As explained in the Apparatus section, the high-fidelity condition used all four screens of the VisCube with stereo and head tracking, while the low-fidelity condition used one screen without stereo or head tracking. Each participant used either the high- or low-fidelity setup for their tasks.

The level of visual complexity indicated how much visual clutter was present in the visualization. Visual complexity was controlled based on the number of individual graphs present within the same space. Note that the *visual complexity* was independent of mathematical *graph complexity*. All individual graphs used in the different conditions of the experiment were isomorphic. That is, all graphs had the same number of nodes and the corresponding nodes of all graphs were connected in the same ways. Visually, different graphs had their nodes in different positions, but their mathematical structures were identical. For example, Figure 3 shows isomorphic red and blue graphs in the same space.

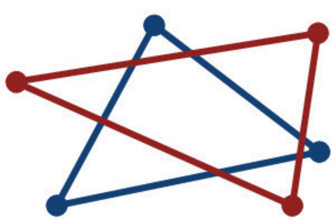


Figure 3: Simple example of two separate isomorphic graphs within the same space.

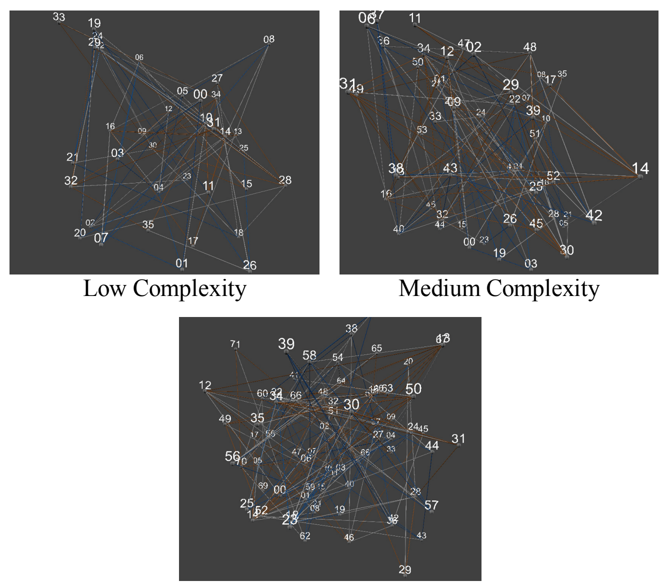
By controlling graphical isomorphism, we were able to control the task scope for the graph inspection tasks (i.e., identifying connected nodes and path tracing). At the same time, we were able to control visual complexity by the number of separate graphs within the same space, which determined the amount of visual clutter in the space.

The experiment used three levels of visual complexity (shown in Figure 4): low, medium, and high. Conditions with low complexity had two identical and separate graphs in the same space, medium-complexity conditions had three graphs within the same space, and high-complexity conditions had four graphs. Each separate graph within the space was mathematically identical, having a total of 18 nodes and 47 edges. Visual complexity was controlled within subjects, so that every participant completed the tasks with all graph visualizations. The orders for complexity levels were balanced across other conditions using a Latin-square design.

Additionally, we considered participants' spatial ability as another between-subjects variable. All participants completed a standard visual spatial ability test (the Purdue Visualization of Rotations Test [3]) prior to participating in the study. Based on their scores on this test, participants were assigned to a high-spatial-ability group or a low-spatial-ability group. Participants in these groups were balanced across conditions.

All participants completed all four task types, which were described in the Task section. We also controlled task scope as a within-subjects variable by selecting a simple task and an advanced task for each combination of complexity level and task type.

For performance metrics, we recorded completion time and correctness for each task attempt. Participants were given a time limit of five minutes for each attempt. If the attempt was not completed before reaching the time limit, the experimenter notified the participant and moved on to the next attempt. Because raw completion time does not take correctness into account, a time score was calculated for each task. If the participant provided a correct answer for the task, the raw time value was used as the value of the time score. If the participant answered incorrectly, the time score was calculated as the time limit plus the standard deviation of all raw times for that task. This way, the time score gives a measure for speed while accounting for errors.



Low Complexity Medium Complexity

High Complexity

Figure 4: Three different levels of visual complexity used in our experiment.

3.5 Participants

Twenty-eight undergraduate university students participated in the study, with equal numbers of males and females. Ages ranged from 19 to 25, and the median age was 20. All participants were architecture majors enrolled in a junior-level *Building Structures* architecture class. Through this class, all participants had experience working with visualization software to inspect 3D structures.

3.6 Procedure

Prior to the day of the experiment, all participants had completed a visual spatial ability test as a homework assignment. The scores from this test were used to balance participants across conditions by spatial ability.

On the day of the experiment, participants first completed a background survey to provide demographic information and estimations about weekly video game playing. Participants were then introduced to the VisCube display and travel techniques. Participants were required to practice until they could demonstrate proficient control with the wand device. Next, using a graph similar to those used to the experimental trials, the experimenter explained the tasks. The experimenter walked the participants through each task. Participants were then asked to complete each task themselves, while the experimenter provided feedback and answered any questions.

After this familiarization, the participants completed the primary trials. For each of the three levels of complexity, the participant completed two tasks of each of the four types (intersection search, path following, connection identification, and length comparison). Participants had the option of taking breaks throughout the study, and the experimenter asked participants if they would like to rest after completing each level of complexity.

Next, participants were briefly interviewed about the trials and their experiences with the display. Finally, participants completed a visual spatial ability test (the same spatial ability test that they had previously taken prior to the study). The entire procedure took approximately 60 to 75 minutes.

3.7 Results

We tested for effects of the experiment's independent variables on the task correctness and time scores. We analyzed the results with mixed-design factorial ANOVAs. We ran the tests considering display fidelity and spatial ability as between-subjects factors, and considering visual complexity, task type, and task scope as within-subjects factors.

3.7.1 Time Score Results

The time score metric provided an overall performance measure that accounted for both speed and correctness (as described in the Task section). The ANOVA for time score found a significant effect of display fidelity, with $F(1, 24) = 5.733$ and $p = 0.025$. Performance was faster with high fidelity ($M = 91.055$, $SD = 719.43$) than with low fidelity ($M = 128.89$, $SD = 779.43$). This confirms our hypothesis that a higher level of display fidelity would affect positively 3D inspection tasks. The combination of increased display area and the addition of stereo and head tracking made it easier to quickly complete the tasks. This overall result is consistent with the results found in several different reports in the literature that show how spatial understanding of 3D visualizations can be positively affected by higher levels of display fidelity (e.g., [15, 19]).

But the more important results from this experiment are related to visual complexity, task scope, and task types. The ANOVA detected an overall significant effect of visual complexity on time scores, with $F(2, 48) = 5.728$ and $p = 0.006$. As expected, the mean time scores got worse as visual complexity increased (low complexity: $M = 94.17$, $SD = 301.99$; medium complexity: $M = 112.78$, $SD = 336.44$; and high complexity: $M = 122.96$, $SD = 419.13$), as can be seen in Figure 5. This confirms our main hypothesis that visual complexity does affect the overall speed in completing 3D inspection tasks. Figure 5 also shows that the higher display fidelity yields better overall performance with all the different levels of visual complexity. A Bonferroni pairwise post-hoc comparison showed significant differences between low and high levels of visual complexity, with $p = 0.013$. No other significant effects were found between different visual complexity conditions.

Contrary to our hypothesis that higher display fidelity would cause greater performance benefits with increased visual complexity, no interaction was detected between display fidelity and visual complexity, so we reject this hypothesis. Figure 5 suggest that the performance benefits of using higher level of display fidelity are the same for all levels of visual complexity.

Overall, there was also a significant effect of task type on time scores, with $F(3, 72) = 35.187$ and $p < 0.001$. Figure 6 shows that time scores were better with higher display fidelity for all task types. There was no interaction detected between task type and display fidelity, with $F(3, 72) = 1.48$ and $p = 0.227$. To test the effects of display fidelity for each task type, we ran independent t-tests as planned contrasts. A test found significantly better time scores with higher fidelity for the intersection search task, with $t(26) = 2.594$ and $p = 0.015$. Participants also tended to be faster with higher display fidelity for the path tracing task, with $t(26) = 1.79$ and $p = 0.085$, and for the connection identification task, with $t(26) = 1.902$ and $p = 0.068$, though these effects are not significant at the 0.05 level. The planned contrast failed to detect evidence of an effect of display fidelity for the length comparison task, with $t(26) = 0.627$ and $p = 0.536$.

The two levels of task scope were significantly different for time scores, with $F(1, 24) = 59.484$ and $p < 0.001$. As can be seen in Figure 7, the results confirm our hypothesis that the advanced-scope versions of the tasks were more challenging than the simple versions (simple: $M = 79.02$, $SD = 458.79$; advanced: $M = 140.92$, $SD = 550.55$).

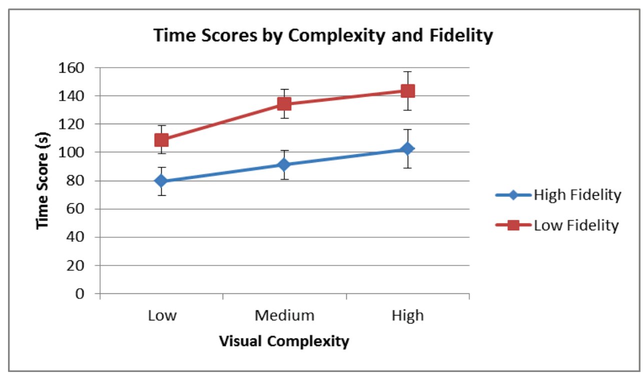


Figure 5: Overall average time score for different visual complexity conditions for high and low display fidelity. The error bars indicate standard error. Lower time scores are better.

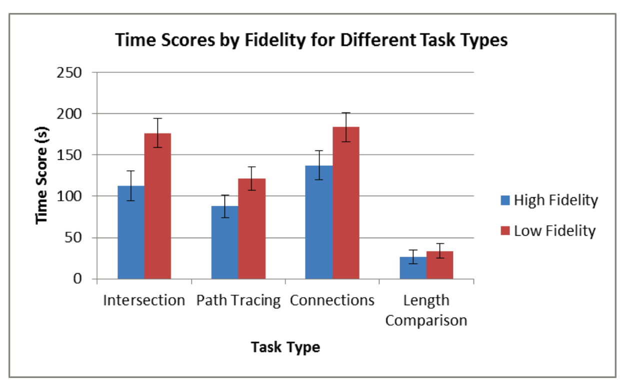


Figure 6: Overall average time score for different task types for high and low display fidelity. The error bars indicate standard error. Lower time scores are better.

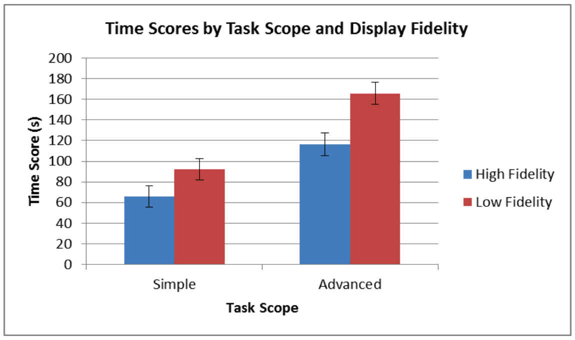


Figure 7: Average time score for different task-scope conditions for high and low display fidelity. The error bars indicate standard error. Lower time scores are better.

The ANOVA for time scores also detected a significant interaction between visual complexity and task scope, with F(2, 48) = 7.113 and p = 0.002. A Bonferroni post-hoc pairwise comparison test showed significant differences between simple and advanced task scopes for low complexity (simple: M = 34.7; advanced: M = 49.1; p < 0.001), medium complexity (simple: M = 35.1; advanced: M = 77.7; p < 0.001), and high complexity (simple: M = 46.6; advanced: M = 59.1; p = 0.014).

There was no significant interaction between display fidelity and task type. On the other hand, the interaction between task scope and task type was significant, with F(3, 72) = 15.474 and p < 0.001. A Bonferroni post-hoc pairwise comparison test showed significant differences between the simple and advanced task scopes for the path tracing task (p < 0.001) and the connection identification task (p < 0.001). As shown in Figure 8, speed on the intersection and length comparison tasks were not significantly affected by task scope. No other significant interactions were detected.

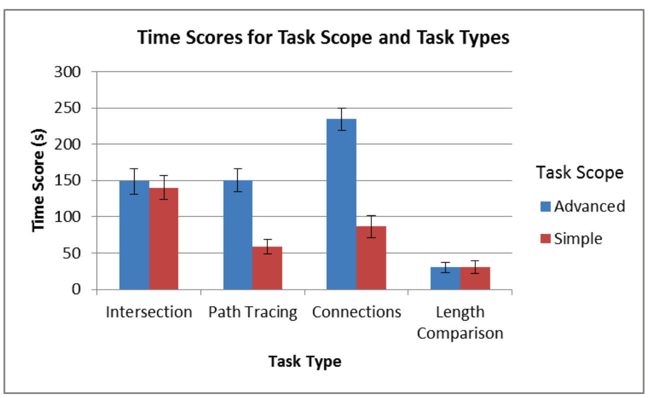


Figure 8: Average time score for different task types and simple and advanced task scope. The error bars indicate standard error.

The test for time scores failed to detect effects due to spatial ability, with F(1, 24) = 0.003. Moreover, spatial ability did not have any significant interactions with display fidelity, visual complexity, or task scope. These results suggest that participant spatial ability was not a significant factor for performance on our graph analysis tasks.

3.7.2 Correctness Results

The correctness results provided a measure for successful task completion without regard for speed. Our ANOVA for correctness did not detect an overall effect due to display fidelity, with F(1, 24) = 2.347 and p = 0.139. Though the time score results do support the hypothesis that higher fidelity conditions would cause performance improvements, this is not supported by the correctness results alone. This discrepancy suggests that the more advanced display features made a bigger difference for task efficiency, rather than for pure accuracy.

The simple scope conditions were significantly better than the advanced task scope conditions, with F(1, 24) = 5.559 and p = 0.027. The test also found significant differences due to task type, with F(3, 72) = 13.617 and p < 0.001. Figure 9 shows the average correctness of the answers for each task type with simple and advanced task scope. It highlights that task scope affected the path tracing and connections tasks more than the other tasks, which is consistent with the results for time score. The ANOVA failed to detect significant effects of the interaction between display fidelity and task type. In planned contrast tests for the effects of display fidelity for each individual task type, independent t-tests found no effects of display fidelity for any of the tasks.

The ANOVA for correctness failed to detect significant effects for visual complexity, with F(2, 48) = 1.997 and p = 0.147. Combined with the results for the time scores, the lack of effects for correctness suggests that visual clutter makes a greater impact on speed than correctness.

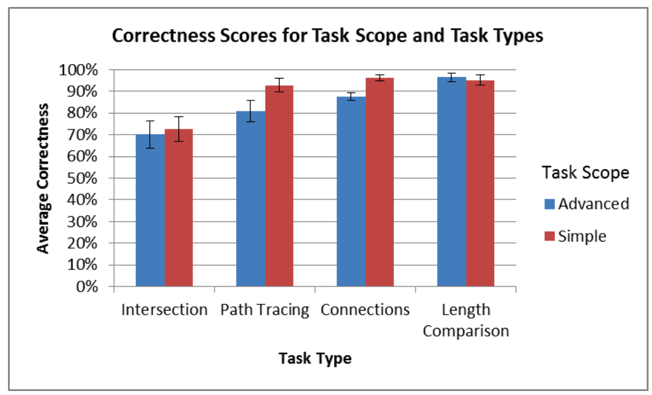


Figure 9: Average correctness for the different task types and simple and advanced task scope. The error bars indicate standard error.

Just as with time scores, no significant effects were found due to spatial ability for correctness, with F(1, 24) = 0.014 and p = 0.908, and the test found no significant interactions with variables for correctness.

3.7.3 Correlations

We also tested whether time scores were correlated with reported hours of video game playing. Spearman's correlations were performed because reported gaming hour were not normally distributed. However, no significant correlations were found for different levels of visual complexity, task scope, task type, and overall average time score and accuracy.

Though level of spatial ability was treated as a binary variable for the ANOVA effect tests, we also tested for correlations between time scores and the exact spatial ability test scores. The test found significant correlations for time score between the spatial ability score with the low visual complexity (ρ = -0.528, p = 0.004), high visual complexity (ρ = -0.462, p = 0.013), path tracing (ρ = -0.392, p = 0.039), connection identification (ρ = -0.471, p = 0.011), advanced task scope (ρ = -0.560, p = 0.002), and average time score considering all conditions (ρ = -0.499, p = 0.007). For accuracy, only one significant correlation was found, between spatial ability and more advanced task scope (ρ = 0.391, p = 0.04). This means that higher scores in the spatial visualization test led to higher speed for a variety of conditions, and higher accuracy when the task was more demanding.

4 Discussion

Based on the time score results, which account for both speed and accuracy of responses, the high-fidelity display condition had better overall performance for the experiment's graph analysis and spatial understanding tasks. Display fidelity did not make a significant difference for correctness scores. These findings suggest that higher levels of display fidelity may be important for improving the speed for spatial analysis tasks to be successfully completed, but our results do not provide evidence of differences for accuracy without consideration for time. This is contrary to the findings reported Ware and Franck [19], which showed that both stereo and head tracking improve accuracy for a path tracing task, and Ragan et al. [12], which showed that a larger display area and head tracking improve accuracy for an intersection detection task. However, we believe this is due to a number of differences between the environments and tasks (e.g., path tracing in [19] required users to determine if there was a path with a specific number of nodes, instead of finding any possible path between two nodes), as well as the lack of control for visual complexity and task scope in those studies.

It is also important to note that the lack of significant interactions between display fidelity and other conditions suggests that the positive effect was independent of visual complexity and task scope. Of course, it is possible that the reason we did not see interactions can be attributed to a failed detection by the statistical analysis, but we do not believe this to be the case. As Figure 4 shows, the graphs had a considerable range of visual complexity for the three levels. More convincingly, Figure 5 shows the relationship between display fidelity and visual complexity for time scores, and the trend is consistent for all levels of visual complexity. Figure 7 presents a similar relationship for task scope and display fidelity. These results are similar to those reported by Ware and Mitchell [20], which showed persistent advantages of stereo for graphs with varying numbers of nodes. The advantage to our approach was the ability to add to this finding by separating the visual complexity and task scope. With the addition of our results, we have evidence that the significant effects of the combination of additional stereo, larger display area, and head tracking are not confounded by the relationship between structural variance and the amount of visual clutter.

Our results also show that the level of visual complexity and task scope each individually affect the overall task performance. This illustrates the importance of considering both factors when calculating the overall difficulty and complexity of a spatial task, and both elements should be taken into account when considering display options for a 3D application. Our results support the claim that even for relatively easy tasks (as simulated with conditions having simple scope and/or low visual complexity), higher display fidelity can still provide performance improvements.

With this in mind, we can consider how the differences between the display conditions affected performance in the tasks used in our experiment. By separately analyzing the effects of display fidelity on four separate tasks types, it is possible to speculate why aspects of the task were more strongly affected by the display variable. Though performance was better for all tasks with higher fidelity, the statistical analysis shows significant performance gains in the intersection search task and near-significant improvements for connection identification and path tracing. These results suggest that the high-fidelity features of virtual reality displays offer benefits for spatial search and fine-grained component distinction, but may provide less noticeable gains than for tasks involving sense of scale or size comparison. We suspect that the addition of stereo, head tracking, and display area could benefit each of these tasks, however each display component could be more or less useful for the different tasks.

Stereo is useful for depth perception, and we hypothesize that it helped participants to more easily distinguish between edges and nodes at different depths. This would certainly assist in identifying collisions between edges in the *intersection search* task, and in distinguishing between edges while following connections the *connection identification* and *path following* tasks. Better depth perception could also make it easier to gauge the length of edges for *length comparisons*.

As shown in previous studies (e.g., [12, 19]), head tracking works especially well in combination with stereo. Requiring participants to distinguish between small-scale intersections, which was similar to the *intersection search* task of our experiment, Ragan et al. [12] found that adding head-tracking significantly improved accuracy, and improved speed when combined with stereo. The ability to use natural head and body movement to adjust the view point and control motion parallax would be especially useful for identifying collisions between spatial structures. But the same effect would obviously be useful for adjusting the view and physically looking around geometry for the other tasks in our study, as well. While participants in both display conditions could still use joystick-control to achieve motion parallax, we hypothesize that head tracking simply made it faster and easier to adjust the view with familiar body movements.

The difference in display area between the low- and high-fidelity conditions provides similar advantages of physical view control. Unlike the single display screen of the low-fidelity condition, the three walls and floor of the high-fidelity display made it possible to use physical head and body rotation to view more virtual content. VR researchers sometimes refer to the total angular range accessible with physical rotation as the *field of regard* (FOR), and many studies have controlled this factor separately from other display components (e.g. [11-13]). These previous studies have shown a strong effect of FOR on both accuracy and speed when completing spatial inspection tasks. We hypothesize this is due to the fact that physical rotation allows for faster turning and movements than commonly used in virtual navigation alternatives, therefore providing more contextual information, which could increase accuracy, in a shorter amount of time, which affects speed. While increased FOR does not necessarily mean better performance for all tasks, the collective findings from previous studies and our own support the hypothesis that higher FOR improves performance of spatial understanding and inspection tasks.

It is important to note here that the results of the study performed by Henry and Polys [9] contradict this hypothesis, in which FOR surprisingly tended to have an overall negative effect on speed. While the authors believe such results were caused by the additional walls being distracting, we hypothesize that two other factors affected their results. First, participants always performed tasks in both low- and high-fidelity conditions, which could have biased the strategy used for all conditions. Second, this was an overall effect considering both exocentric and egocentric navigation modes, while most other studies (including our own) use egocentric navigation. This could mean that participants simply performed so much worse with one of the navigation modes and high-display fidelity that the overall effects changed. However, that paper does not provide enough details to support these hypotheses.

Interestingly, this issue shows another contribution of our work, which is related to the validation of previous studies. By creating a more controlled environment (separating visual complexity and task scope) and testing a diverse set of spatial tasks, we not only provide evidence to support existing findings and derive new ones, but we are able to understand these findings better.

5 Conclusion

In this paper, we controlled for and evaluated the effects of visual complexity, task scope, display fidelity, and a user's personal spatial ability on performance of spatial inspection tasks. We used an abstract environment with a 3D undirected graph, and tested four different types of tasks: intersection search, path following, connection identification, and length comparison. Our results show a strong positive effect of the level of display fidelity on speed for performing tasks, as well as a clear effect of visual complexity and task scope, with higher levels of either type of complexity leading to slower performance.

Since we use a generic 3D graph and tasks, our results can be extended for a variety of applications. For example, in structural engineering applications [2], users have to closely inspect complicated structures to perform comparisons between structural components. In oil-well path planning [7], it is important that users search for intersections between different wells. In network visualization [8], for example, users are interested in tracing a path between different nodes or determining connectivity. However, while we provide general guidelines about different factors, it is also important for future work to investigate these effects in realistic applications.

Acknowledgements

We would like to thank the Virginia Tech Visionarium for the hardware and facilities used to run this study. This study was supported by the National Science Foundation under Grant No. CCLI/TUES 0817106. This support is gratefully acknowledged. Any opinions, findings, and conclusions expressed in this paper are those of the writers and do not necessarily reflect the views of the National Science Foundation.

References

[1] Arns, L., D. Cook, and C. Cruz-Neira, The benefits of statistical visualization in an immersive environment, in *Proceedings of IEEE Virtual Reality*, 1999. p. 88-95.

[2] Bacim, F., N. Polys, J. Chen, M. Setareh, J. Li, and L. Ma, Cognitive scaffolding in Web3D learning systems: a case study for form and structure, in *Proceedings of the 15th International Conference on Web 3D Technology* 2010, ACM: Los Angeles, California. p. 93-100.

[3] Bodner, G.M. and R.B. Guay, The Purdue visualization of rotations test. The Chemical Educator, 1997. **2**(4): p. 1-17.

[4] Bowman, D.A. and R.P. McMahan, Virtual reality: how much immersion is enough? *IEEE Computer*, 2007. **40**(7): p. 36-43.

[5] Chance, S.S., F. Gaunet, A.C. Beall, and J.M. Loomis, Locomotion Mode Affects the Updating of Objects Encountered During Travel: The Contribution of Vestibular and Proprioceptive Inputs to Path Integration. *Presence: Teleoperators and Virtual Environments*, 1998. **7**(2): p. 168-178.

[6] Chen, J., H. Cai, A.P. Auchus, and D.H. Laidlaw, Effects of Stereo and Screen Size on the Legibility of Three-Dimensional Streamtube Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 2012. **18**(12): p. 2130-2139.

[7] Gruchalla, K., Immersive well-path editing: investigating the added value of immersion, in *Proceedings of IEEE Virtual Reality, 2004*. p. 157-164.

[8] Halpin, H., D. Zielinski, R. Brady, and G. Kelly, Exploring semantic social networks using virtual reality. The Semantic Web-ISWC 2008, p. 599-614.

[9] Henry, J.A.G. and N.F. Polys, The effects of immersion and navigation on the acquisition of spatial knowledge of abstract data networks. Procedia Computer Science, 2010. **1**(1): p. 1737-1746.

[10] Laha, B., K. Sensharma, J.D. Schiffbauer, and D.A. Bowman, Effects of Immersion on Visual Analysis of Volume Data. *IEEE Transactions on Visualization and Computer Graphics*, 2012. **18**(4): p. 597-606.

[11] McMahan, R.P., D.A. Bowman, D.J. Zielinski, and R.B. Brady, Evaluating Display Fidelity and Interaction Fidelity in a Virtual Reality Game. *IEEE Transactions on Visualization and Computer Graphics*, 2012. **18**(4): p. 626-633.

[12] Ragan, E., R. Kopper, P. Schuchardt, and D. Bowman, Studying the Effects of Stereo, Head Tracking, and Field of Regard on a Small-Scale Spatial Judgment Task. *IEEE Transactions on Visualization and Computer Graphics*, 2012. doi: 10.1109/TVCG.2012.163

[13] Ragan, E.D., A. Sowndararajan, R. Kopper, and D.A. Bowman, The effects of higher levels of immersion on procedure memorization performance and implications for educational virtual environments. *Presence: Teleoperators and Virtual Environments*, 2010. **19**(6): p. 527-543.

[14] Ruddle, R.A., S.J. Payne, and D.M. Jones, Navigating large-scale virtual environments: what differences occur between helmet-mounted and desk-top displays? *Presence: Teleoperators & Virtual Environments*, 1999. **8**(2): p. 157-168.

[15] Schuchardt, P. and D.A. Bowman, The benefits of immersion for spatial understanding of complex underground cave systems, in *Virtual Reality Software and Technology: Proceedings of the 2007 ACM symposium on Virtual reality software and technology,* 2007. p. 121-124.

[16] Slater, M., A note on presence terminology. *Presence Connect*, 2003. **3**(3).

[17] Slater, M., Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2009. **364**(1535): p. 3549-3557.

[18] Ware, C., K. Arthur, and K.S. Booth, Fish tank virtual reality, in *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems* 1993. p. 37-42.

[19] Ware, C. and G. Franck, Evaluating stereo and motion cues for visualizing information nets in three dimensions. *ACM Transactions on Graphics* (TOG), 1996. **15**(2): p. 121-140.

[20] Ware, C. and P. Mitchell, Reevaluating stereo and motion cues for visualizing graphs in three dimensions, in *Proceedings of the 2nd symposium on Applied perception in graphics and visualization*2005. p. 51-58.

[21] Watson, A.S. and C.J. Anumba, The need for an integrated 2D/3D cad system in structural engineering. *Computers & Structures*, 1991. **41**(6): p. 1175-1182.