Quantifying Immersion in Virtual Reality

Randy Pausch[1], Dennis Proffitt, George Williams[2]

University of Virginia

pausch@cmu.edu, drp@virginia.edu, gcw@best.com

ABSTRACT

Virtual Reality (VR) has generated much excitement but little formal proof that it is useful. Because VR interfaces are difficult and expensive to build, the computer graphics community needs to be able to predict which applications will benefit from VR. In this paper, we show that users with a VR interface complete a search task faster than users with a stationary monitor and a hand-based input device. We placed users in the center of the virtual room shown in Figure 1 and told them to look for camouflaged targets. VR users did not do significantly better than desktop users. However, when asked to search the room and conclude *if* a target existed, VR users were substantially better at determining when they had searched the entire room. Desktop users took 41% more time, re-examining areas they had already searched. We also found a positive transfer of training from VR to stationary displays and a negative transfer of training from stationary displays to VR.

INTRODUCTION

In 1968, Ivan Sutherland implemented the first virtual reality system. Using wire-frame graphics and a head-mounted display (HMD), it allowed users to occupy the same space as virtual objects [Sutherland]. In the 1980's, VR captured the imagination of the popular press and government funding agencies [Blanchard, Fisher]. Potential VR applications include architectural walk-through [Brooks], simulation [Bryson], training [Loftin], and entertainment [Pausch 1996]. For the purpose of this paper, we define "virtual reality" to mean any system that allows the user to look in all directions and updates the user's viewpoint by passively tracking head motion. Existing VR technologies include HMDs and CAVEs[tm] [Cruz-Neira].

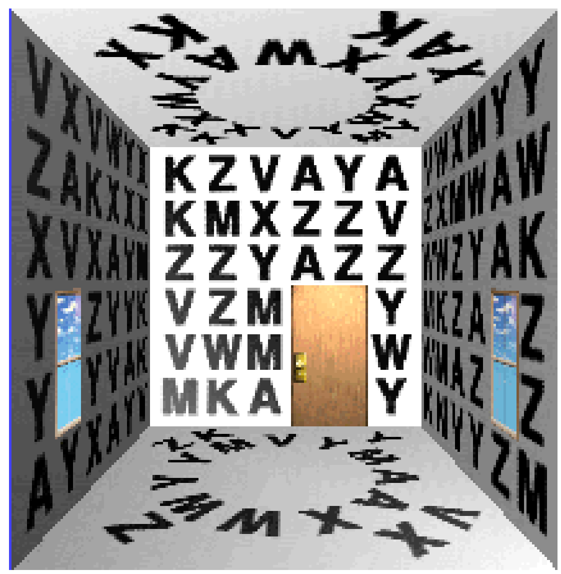


Figure 1: *Users Stood in the Center of This Room and Looked For Target Letters.*

The National Academy of Sciences report on VR [NAS] recommends an agenda to determine when VR systems are better than desktop displays, and states that without scientific grounding many millions of dollars could be wasted. Ultimately, we would like a predictive model of what tasks and applications merit the expense and difficulty of VR interfaces. In this paper, we take a step towards quantifying immersion, or the sense of "being there." We asked users, half using an HMD and half using a stationary monitor, to search for a target in heavily camouflaged scenes. In any given search, there was a 50/50 chance that the target was somewhere in the scene. The user's job was to either find the target or claim no target was present. Our major results are:

1) VR users *did not* find targets in camouflaged scenes faster than traditional users.

2) VR users were substantially faster when no target was present. Traditional users needed to re-search portions of the scene to be confident there was no target.

[1]Currently at Carnegie Mellon University.

[2]Currently at Fakespace, Inc. 241 Polaris Ave. Mountain View, CA 94043

From these two findings, we infer that the VR users built a better mental frame-of-reference for the space. Our second two conclusions are based on search tasks where the users needed to determine that no target existed in the scene:

3) Users who practiced first in VR *positively transferred* that experience and improved their performance when using the traditional display.

4) Users who practiced first with the traditional display *negatively transferred* that experience and performed *worse* when using VR. This negative transfer may be relevant in applications that use desktop 3D graphics to train users for real-world tasks.

In a practical sense, the only way to demonstrate that VR is worthwhile is to build real applications that have VR interfaces, and show that users do better on real application tasks. That can be expensive, and new technologies take time to mature. But the computer graphics community has not even achieved a lower standard: showing, *even for a simple task*, that VR can improve performance. We show improvement in a search task and discuss *why* a VR interface improved user performance.

RELATED WORK

Several researchers have attempted to qualitatively define immersion with taxonomies [Robinett, Zeltzer] or subjective ratings by users [Heeter]. Others have measured "fish tank VR" head-tracked performance [Authur, McKenna, Ware 1993, Ware 1996], or compared variables such as resolution and frame rate in virtual environments [Smets]. We know of no work that formally measures that VR is better than a desktop interface for any search task; the closest is Chung, who compared VR against hand-based manipulation of an object, rather than the viewpoint [Chung].

COMPARING VR AND DESKTOP INTERFACES

To see if VR is useful, one could pick a representative task, such as finding an object in a scene, and compare performance with the best possible VR and desktop interfaces. That introduces many variables, as shown in Table 1. We do not wish to ask if *current* VR interfaces are useful, but rather if VR will *ever* be useful. Simply put, do users perform measurably better when controlling the viewpoint with their head instead of with their hand?

	Desktop	HMD
resolution	1280x1024	240x120
horizontal FOV	40 degrees	93 degrees
vertical FOV	30 degrees	61 degrees
input device	mouse or joystick	6 DOF tracker

Table 1: *Typical Values for Displays*

To hold the variables constant we used the same HMD as both the head-tracked display and the stationary monitor. In both cases, the scenes were rendered in stereo (The use of stereo was probably not significant, as all objects in the scene were at least two meters from the user). Figure 2 shows the stationary condition, where we bolted

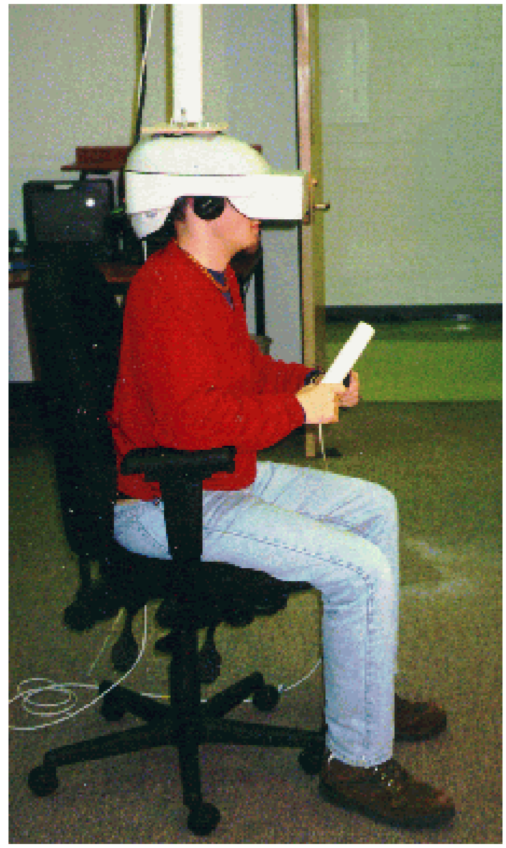


Figure 2: *Using the HMD as a Stationary Monitor.*

the HMD onto a ceiling-mounted post, thus turning the HMD into a stationary monitor. This provided the same resolution, field of view, and image quality in both VR and desktop interfaces. Table 1 gives the values for this particular HMD, the Virtual Research Flight Helmet^{tm} [HMD]. We chose a task where the display resolution was unimportant because the targets were large and easily visible. Using a mouse or joystick as the desktop input device would have introduced variables in lag and sampling rate. Therefore, we used the same magnetic 6DOF electromagnetic tracker [Tracker] from the HMD as our hand input device. All we did to create the desktop interface was to seat the user in a chair and take the 6DOF tracker off the user's head and place it in a comfortable device held in user's hands. By holding all other variables constant, we can claim our results are dependent on head-input versus hand-input. For the remainder of this paper, we refer to these groups as "the VR users" and "the desktop users." While we acknowledge that our desktop users are hardly using a conventional configuration, we claim their setup contains the essential components: a stationary monitor and a hand-input device.

We attempted a pilot experiment [Pausch 1993] where 28 users searched for easy-to-find, uncamouflaged, targets at random locations in a virtual room. VR users found the targets 42% faster than desktop users. We feared we had measured how fast users could move the camera, rather than how immersed they were. For example, finding the red 'Y' in Figure 3 is a *pre-attentive* task, where the

```
AVMNVWXZWXZWXZAVMNVWXZWXZAVMNVWXZWXZAVMNV
AMNVNYVAVMNVWXZMNVAVMNVWXZWXZAVMNVAVMNVWX
ZWXZANVWAVMNVWAVMNVWXZXZXVMAVMNVMWXZNVZAV
MNVWMNZAVMNVAVMNVWXZWXZWXZWXZAVMNVWMAVYNV
AMNVNVAVMNVWXZWXZAVMNVAVMNVWXZAVMNVWXZMXZ
```

Figure 3: *Find the red Y, the Black Y, and the Black K.*

human visual system can find the target without having to consider the camouflage. In a room surrounding the user, the time to find a red Y might be limited only by how fast one could move the camera. But the time to find an object like the black ‘Y’ in Figure 3 is limited by one’s ability to serially examine the items. Searching for a black ‘K’ in Figure 3 is another mentally limited task; there is no ‘K’, and the only way to be certain of that is to systematically search the entire scene. We claim that VR users are much better at systematic searches because they can better remember where they have already looked in the scene that surrounds them.

We placed users in the center of a simple virtual room, 4 meters on each side. The room contained a door and two windows which served as orientation cues. During each search task, the room contained 170 letters arranged on the walls, ceiling and floor. Figure 1 shows a third-person view of the scene, with one wall removed. Letters measured 0.6 meters in length and were easily visible through the display. Users needed to apply some degree of concentration and focused attention to locate the target letter among the similar looking “camouflage” letters. In any given task, we chose target and camouflage letters from either the set “AKMNVWXYZ” (whose primarily visual features are slanted lines), or “EFHILT” (whose primary features are horizontal and vertical lines). We began each search by displaying the target letter in a fixed location over the door, and waiting for the user to say the target letter in order to begin the search. On the user’s cue, we rendered the 170 camouflage letters, placing the target letter at a random location. When they found the target letter, users said “there it is,” which we confirmed by watching an external monitor which displayed what the users were seeing.

48 users participated in the experiment, 24 using VR and 24 using the desktop configuration created by bolting the HMD into a fixed position. Desktop users controlled their viewpoint with the handheld “camera” controller shown in Figure 4, which contained the same 6DOF tracker used to track the VR users’ heads. We did a large number of informal experiments to design a reasonable handheld camera controller. Based on that experience, we also removed the roll component of tracking for the hand input device. The end-to-end system latency in all cases was roughly 100 milliseconds, measured by the technique described by Liang [Liang], and we rendered a constant 60 frames per second on an SGI Onyx Reality Engine2.

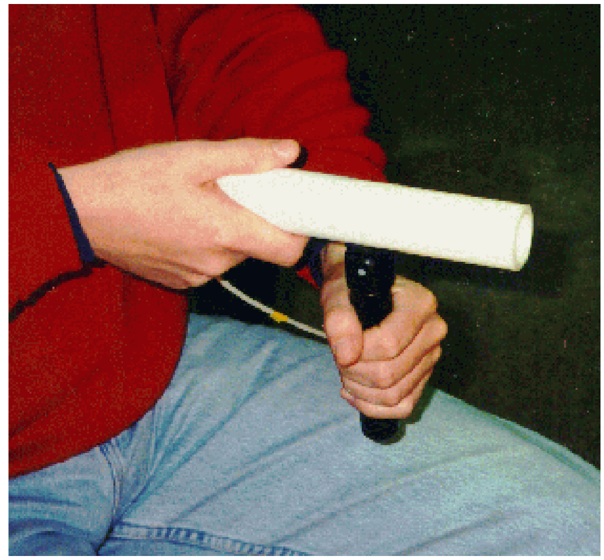
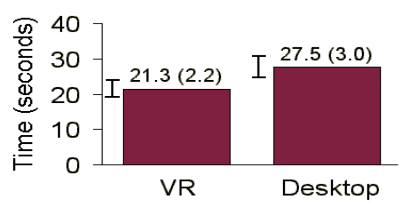


Figure 4: *The Hand Input Device, Containing the 6DOF Tracker.*

RESULTS

Graph 1 shows the average time users needed to locate a target.



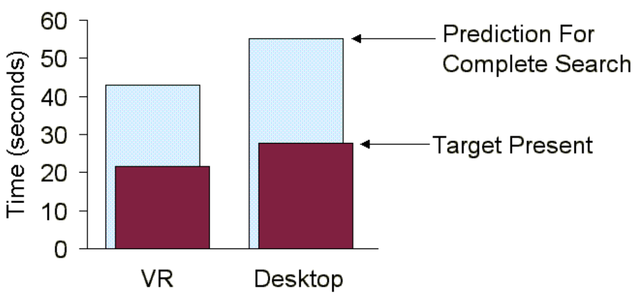
Graph 1: *VR versus Desktop Performance. The difference is not statistically significant.*

Each user performed five searches which we averaged together to form a single data point for that user. The bars in Graph 1 are the average of the 24 VR users and the 24 desktop users. The error bars show the standard error for each data set. The VR and desktop times are very similar, and their difference is not statistically significant. We constructed a cognition-limited task, so it is reasonable that the VR and desktop times are similar. We informally observed that users never physically turned the camera as fast as they could have. The cognitive portion of the search task slowed the users down.

Practice did not appear to be a factor. We asked users to do practice searches until they were comfortable; we required two practice, searches, and some users did three. We did not count practice searches in the results. Users took roughly 15 minutes to perform the searches, and none appeared fatigued. To measure practice and fatigue, we ran separate control groups with eight users each, who ran double the number of trials on both the VR and desktop interfaces. These users showed no statistically significant differences between their earlier and later trials. Users made essentially no errors. All users were between 18 and 25 years old, mostly undergraduates with no VR experience. Both groups were evenly balanced by gender. All users said they could easily see the targets. In addition to the 48 users we report, 3 other users began but did not complete this study. All 3 felt slightly nauseous, and they all reported that they were generally prone to motion sickness.

We now consider searching for a target which is *not* in the scene. Of course, if the user *knows* the target is not there, then the task is pointless. Therefore, we had users perform a sequence of searches, each of which had a 50% likelihood of containing a target. Users were instructed to either locate the target, or claim no target existed. In this way, we measured the time users needed to *confidently* search the entire scene.

If the targets are dense, and the users are efficient in their searching, we can predict how long this will take. Working backwards, consider an efficient user who takes 40 seconds to completely search a scene, with no wasted effort. On average, when a target *is* present, that user should find it in 20 seconds. Random placement may make the letter appear earlier or later in the search process, but on average the user will find the target halfway through the search. We know how long it takes users to find targets when they are present. If the users searched perfectly, it should take twice that long to search the entire room and confidently conclude the target is not there. Any time over that would imply that the users were re-examining portions of the room that they had already searched. This prediction is shown in Graph 2.



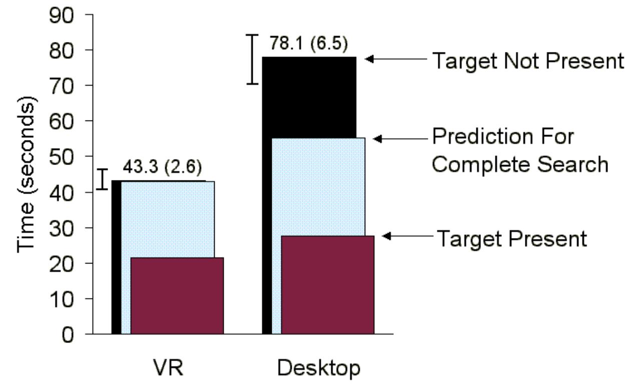
Graph 2: *Predicted Times for A Complete Search.*

Graph 1 showed the results of users who each performed five searches for targets that were in the room. In fact, these users each performed a sequence of ten searches, where on any given search, the target might or might not have existed. For each of the ten searches, the user was told to either find the target, or announce that it was not there. Users did not know beforehand whether a target would be present in any given search. Graph 3 shows the average time users required to locate a target that was in the room (Graph 1 results), the predicted time to search the entire room (Graph 2 results), and the observed time to search the entire room and conclude that no target existed.

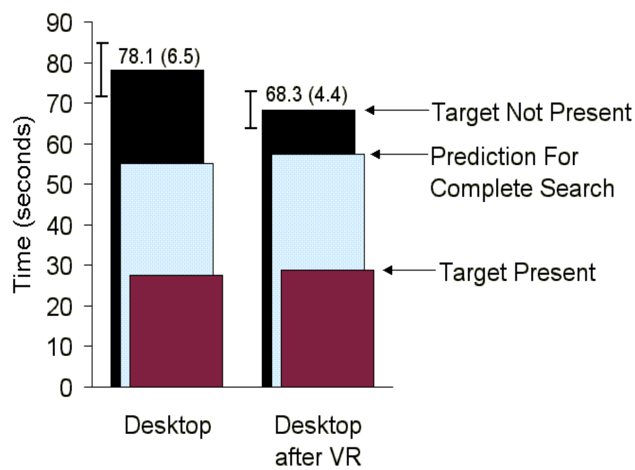
The VR user data is only 1.4% above the prediction for efficient search. This concurs with our personal observations of VR users, who appeared to search the entire room without rescanning. However, desktop users typically examined portions of the room a second time. As shown in Graph 3, the desktop users spent 41% above the time that a perfect search would take.

IMPLICATIONS

The VR community claims that a head-tracked, egocentric camera control provides a stronger sense of immersion, or “being there,” than does a desktop display. Our results indicate that VR can help users remember where they have and have not looked. The ratio shown in the “desktop” performance in Graph 3 implies that back-



Graph 3: *Observed Times To Search the Entire Room and Determine that No Letter is Present.*



Graph 4: *Positive Transfer: Users Who Practice in VR Improve Their Performance on the Desktop*

tracking is occurring [Braddick]. If the desktop users were slower for some biomechanical reason, such as our choice of input device, we assume it would have also slowed them when the target was present.

TRANSFER EFFECTS

We wondered how users would perform the desktop search tasks if they first practiced in VR. If VR allows the user to develop a good frame-of-reference for a space, perhaps that memory would carry over to a desktop interface. We had each of the VR users perform their ten searches, rest for five to ten minutes, and then perform ten more searches using the desktop interface. In this way, we could see if the experience with VR affected later use of the desktop interface. The ten desktop searches, just like the first ten VR searches, contained five with a present target and five without. Graph 4 shows a *positive transfer effect*, where practicing in VR improves performance of the same task when using a desktop interface. This result is statistically significant ($p < 0.0096$). We also performed the reverse experiment — we had the desktop users rest and then perform ten more searches using the VR interface. Graph 5 presents the results, which show a *negative transfer of training*. Practicing on the desktop decreases performance of the same task when using a VR interface. This result is statistically significant ($p < 0.0493$). The implications here are powerful. If we assume that VR and the real world are similar, Graph 5 indicates that training with desktop 3d graphics could potentially degrade real-world performance.

FUTURE WORK

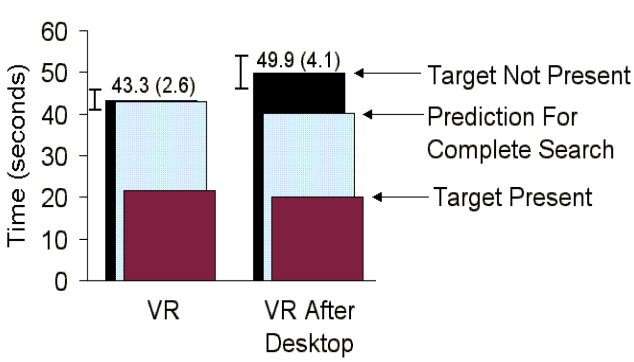
Our claims, particularly about negative transfer of training, rely on VR search performance being similar to real-world search performance. Therefore, we should perform the study using real objects in a real scene. While the absolute times may improve due to improved vision and reduced lag, we expect that "no target exists" searches will take twice as long as searches where the target exists. In a similar vein, we expect that we would see similar results in a CAVE or a BOOM, and we are curious what the results would be for a PUSH device [PUSH]. In general, we believe the computer graphics community should actively pursue this kind of evaluation, which was a primary recommendation of a recent National Academy of Science study on VR [NAS]. Especially given our finding on the potential negative transfer of training, we feel the computer graphics community can benefit from performing this kind of measurement.

CONCLUSIONS

Proponents of virtual reality claim that it can improve user performance via *immersion*, or giving an enhanced sense of "being there." We compared the performance of users searching for targets in heavily camouflaged scenes. Half of the users used VR and the other half used a stationary display with view controlled by a hand input device. In 50% of the searches, we randomly placed a target in the scene. For each search, we asked the user to either find the target, or conclude that no target was in the scene.

1) When targets were present, VR did not improve performance. We believe this is because the task was cognitively limited, and the ability to move the camera quickly and/or naturally was not the bottleneck.

2) When there was no target present, VR users concluded this substantially faster than traditional users. We believe that VR users built a better mental frame-of-reference for the space, and avoided redundant searching.



Graph 5: *Negative Transfer: Users Who Practice with the Desktop Interface Degrade Later Performance with the VR Interface.*

3) Users of traditional displays improved by practicing first with VR. This underscores that something occurred in the user's mental state and could be transferred to using a different interface.

4) VR users who practiced first with traditional displays hurt their performance in VR. This may imply problems with using desktop 3D graphics to train users for real world search tasks.

We believe this is the first formal demonstration that VR can improve search task performance versus a traditional interface. More importantly, the results give us insight into *why* VR is beneficial. This is a step towards our long term goal of being able to predict which real-world tasks will benefit from having a VR interface.

REFERENCES

Arthur, K.W. Booth, Kellogg S., Ware, Colin. 1993. Evaluating 3D Task Performance for Fish Tank Virtual Worlds, ACM Transactions on Information Systems, vol.11, no.3, 239-265.

Blanchard, C., Burgess, S., Harvill, Y., Lanier, J., Lasko, A, Reality Built for Two: A Virtual Reality Tool, ACM SIGGRAPH 1990 Symposium on Interactive 3D Graphics, March 1990.

Braddick, O.J., & Holliday, I.E. (1991). Serial search for targets defined by divergence or deformation of optic flow. Perception, 20_, 345-354.

Brooks, Frederick P, Walkthrough - A Dynamic Graphics System for Simulating Virtual Buildings. In Proceedings of the 1986 ACM Workshop on 3D Graphics, Chapel Hill, NC, October 23-24, 1986, pages 9-21.

Bryson, S., Levit, C., The Virtual Wind Tunnel, IEEE Computer Grahpics and Applications, July 1992, pages 25-34.

Chung, J., A Comparison of Head-tracked and Non-head-tracked SteeringModes in the Targeting of Radiotherapy Treatment Beams, In Proceedings of the 1992 ACM Symposium on Interactive 3D Graphics., pages 193-196.

Cruz-Neira, C., Sandin, T.A., DeFanti R.V. 1993. Surround screen projection-based virtual reality: the design and implementation of the cave. Proceedings of SIGGRAPH 1993. 135-142.

Fisher, S., McGreevy, A, Humphries, J., Robinett, W., Virtual ENvironment Diplay System, Proceedings on the 1986 Workshop on Interactive 3D Graphics, October 23-24, 1986, pages 77-87.

Heeter, Carrie. 1992. Being There: The Subjective Experience of Presence. Presence, vol.1, no.2, 262-271.

HMD: Virtual Research Flight Helmet[tm]. Company information available via www.virtualresearch.com.

Liang, J., Shaw, C., and Green, M., On Temporal-Spatial Realism in the VIrutal Reality Environment, In Proceedings of the ACM SIGGRAPH Symposium on User Interface Software and Technology, 1991, pages 19-25.

Loftin, R., Kenny, Patrick. 1995. Training the Hubble Space Telescope Flight Team. IEEE Computer Graphics and Applications, 31-37.

McKenna, Michael. 1992. Interactive viewpoint control and three-dimensional operations. In Proceedings of the 1992 ACM Symposium on Interactive 3D Graphics., pages 53-56.

NAS: Virtual Reality: Scientific and Technological Challenges, Committe on VIrtual Reality Research and Development, National Research Council, Nathaniel Durlach and Anne Mavor, editors, National Academy of Science Press, 1995, ISBN 0-309-05135-5.

PUSH: Information for the PUSH device can be found at http://www.fakespace.com

Pausch Randy, Shackelford M. Aanne., Proffitt, Dennis., A User Study Comparing Head-Mounted and Stationary Displays, IEEE 1993 Symposium on Research Frontiers in Virtual Reality, October, 1993, pages 41-45.

Pausch, Randy, Snoddy, Jon., Taylor, Robert., Watson, Scott., Haseltine, Eric, Disney's Aladdin: First Steps Toward Storytelling in Virtual Reality, Proceedings of SIGGRAPH, 1996, pages 193-203.

Robinett, Warren. 1992. Synthetic Experience: A Proposed Taxonomy. Presence, vol. 1, no. 2, 229- 247.

Smets, G., and Overbeeke, K., Trade-Off Between Resolution and Interactivity in Spatial Task Performance.

Sutherland, I. 1968. A head-mounted three-dimensional display. In Proceeding of the Fall Joint Computer Conference. AFIPS Conference Proceedings, vol. 33. AFIPS, Arlington, VA., 757-764.

Tracker: Polhemus FASTRACK^{tm}. Company information available via www.polhemus.com.

Ware, C. and Franck, G. (1996) Evaluating and stereo and motion cues for visualizing information nets in 3D. ACM TOG 15(2). 121-139.

Ware, Arthur and Booth. (1993) "FishTank Virtual Reality" ACM CHI'93 Proceedings, 37-42. is also somewhat relevant in that it looks at the phenomenological perception of space. Motion parallax is found to be more important that stereo information.

Zeltzer, D. 1992. Autonomy, interaction, and presence. Presence vol. 1, no. 1, 127-132.