In Press: *The Proceedings of the 11th International Conference on Human Computer Interaction*

Development of a Benchmarking Scenario for Testing 3D User Interface Devices and Interaction Methods

Albert A. Rizzo[1], Gerard J. Kim[2], Shih-Ching Yeh[1], Marcus Thiebaux[1], Jayne Hwang[2] & J. Galen Buckwalter[1]

[1]University of Southern California Institute for Creative Technologies, 13274 Fiji Way, Marina del Rey, CA. 90292
arizzo@usc.edu

[2]Pohang University of Science & Technology,Virtual Reality Lab, San 31, Hyoja-Dong, Pohang, Kyungbuk, Korea, POSTECH CSE. gkim@postech.ac.kr

Abstract

To address a part of the challenge of testing and comparing various 3D user interface devices and methods, we are currently developing and testing a VR 3D User Interface benchmarking scenario. The approach outlined in this paper focuses on the capture of human interaction performance on object selection and manipulation tasks using standardized and scalable block configurations that allow for measurement of speed and efficiency with any interaction device or method. The block configurations that we are using as benchmarking stimuli are accompanied by a pure mental rotation visuospatial assessment test. This feature will allow researchers to test users' existing spatial abilities and statistically parcel out the variability due to innate ability, from the actual hands-on performance metrics. This statistical approach could lead to a more pure analysis of the ergonomic features of interaction devices and methods separate from existing user abilities. An initial test was conducted at two sites using this benchmarking system to make comparisons between 3D/gesture-based and 2D/mouse-based interactions for 3D selection and manipulation. Our preliminary results demonstrated, as expected, that the 3D/gesture based method in general outperformed the 2D/mouse interface. As well there were statistically significant performance differences between different user groups when categorized by their sex, visuospatial ability and educational background.

1 Introduction

Virtual Reality (VR) has now emerged as a promising tool in many domains of human experience (Stanney, 2002). Continuing advances in VR technology along with concomitant system cost reductions have supported the development of more usable, useful, and accessible VR systems that can uniquely target a wide range of application areas and research questions. What makes VR application development so distinctively important is that it represents more than a simple linear extension of existing computer technology for human use by way of its immersive and interactive features. However, before this vision can be fully reached, conceptual and technological advances need to occur in the area of 3D User Interface (3DUI) devices and interaction methods. From a human computer interaction perspective, a primary concern involves how to design more effective, efficient and easily learnable methods for human interaction with complex VR systems. Current methods are still limited in the degree to which they allow users to naturalistically interact with the challenges presented in a virtual environment. A case can be made that short of fostering truly realistic naturalistic interaction, perhaps interaction methods based on "magic" that give the user "suprahuman" interaction abilities are a viable option (Bowman, Druijff, Laviola & Poupyrev, 2001). However, these sometimes less than intuitive methods may not be easily learnable and often require extensive usability testing even before the efficacy of the scenario can begin to be evaluated apart from the interaction design. Even if novice users are capable of using a less natural interaction method at a basic level, the extra non-automatic cognitive effort required to interact/navigate could serve as a distraction and limit the ultimate value of a VR application for accomplishing a specified set of goals. In this regard, Psotka (1995) hypothesizes that facilitation of a "single egocenter" found in highly immersive interfaces would serve to reduce "cognitive overhead" and thereby enhance information access and learning. While many modes of VR interaction (i.e., wands, joy sticks, 3D mice, etc.) can be easily mastered by expert users, the learning curve for use of such methods by novices is often poorly documented. Problems with the learnability of 3D user interfaces and interaction methods could limit the widespread usage of VR by potential "everyday" users who do not have prior experience or training with such interface technology.

Despite some progress in interaction modeling (Bowman et al., 2001), the search for the best 3DUI and interaction method for a given application usually requires costly and time-consuming usability testing. Other major challenges for designing usable interfaces include the rapid changes in hardware capabilities, device availability, cost and the lack of a mature methodology in interaction design. The situation is often compared to the case of 2D desktop interfaces where a mature interface methodology has emerged over the last 30 years using devices that are standardized/fixed (mouse, keyboard, monitor) (Preece, Rogers & Sharp, 2002). With the absence of such an established design methodology, we are still limited to a trial-and-error exploratory approach to VR interaction design. This is an area that needs the most attention in the current state of affairs for VR and better multidisciplinary collaboration in application development may be a key element for future success.

To address a part of the challenge of testing and comparing various 3DUI devices and interaction methods, we are currently developing a VR benchmarking scenario. This application focuses on the capture of human interaction performance on object selection and manipulation tasks using standardized and scalable block configurations that allow for measurement of speed and efficiency with any interaction device or method. It will be possible with this system to store performance data across different defined user groups (age/sex/years of computer experience/etc.) and build performance norms for future comparison purposes. As well, the block configurations that we are using as benchmarking stimuli are accompanied by a pure mental rotation visuospatial assessment test. This feature will allow researchers to test users' existing spatial abilities and statistically parcel out the variability due to innate ability, from the actual hands-on performance metrics. This statistical approach could lead to a more pure analysis of the ergonomic features of 3DUIs and methods with the variability of preexisting user abilities removed.

An initial test was conducted using this benchmarking system to make comparisons between gesture and mouse interaction. The block configuration methodology was derived from some of our previous work (Rizzo et al., 2001) with Shepard and Metzler (1971) mental rotation stimuli (Figure 1). These block stimuli have many features that make them attractive as 3D interaction benchmarking stimuli. There is a rich history of research examining human performance with these stimuli, particularly in the area of spatial ability investigations across age and sex (Voyer, Voyer & Bryden, 1995). The stimuli are easily capable of being hierarchically presented from very simple renderings to much more complex configurations and in the types of rotations required to perform tasks with them. The influence of prior learning and experience on interaction performance measures with these stimuli is lessened when compared with using realistic objects. And the influence of a person's innate level of visuospatial ability can be statistically parceled out of the user's hands-on interaction performance with the blocks, by factoring out results from user performance on the highly reliable Vandenberg and Kuse (Vandenberg & Kuse, 1978; Peters et al., 1995) Mental Rotations Test (Figure 2).

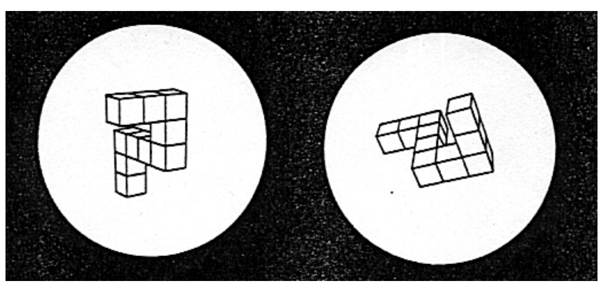

Figure 1: Shepard & Metzler Block Configuration Stimuli. (Shepard & Metzler, 1971)

Now Look at this object:

Two of these four drawings show the same object. Can you find those two? Put a big X across them

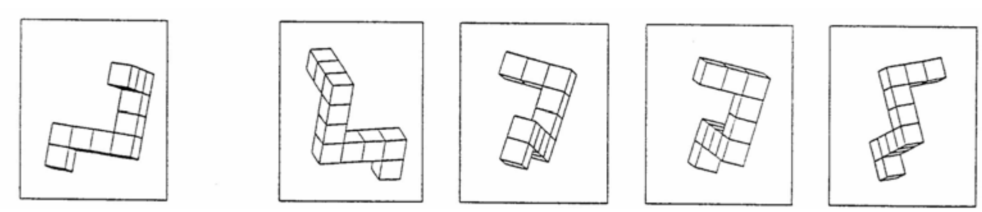

Figure 2: Vandenberg & Kuse Mental Rotations Test. (Vandenberg & Kuse, 1978)

We have now designed a series of progressively more complex block configurations that can be rendered in either monoscopic or stereoscopic mode on a common PC CRT monitor (shutter glasses are required for stereo mode). The task presents pairs of identical block stimuli in orientations (that are readily controllable by the experimenter) and the user is instructed to select and manipulate one of the block configurations and to superimpose it upon the target configuration using a designated input device. When the user completes the trial, data is automatically collected on the speed of successful superimposition and the efficiency of the actual movement as measured by the ratio of the shortest path of travel to the actual path. We are now able to configure the system such that the blocks can be actuated by any input device in order to compare speed and efficiency across interaction devices and methods. As well, by administering the Mental Rotations Test (MRT) (Vandenberg & Kuse, 1978; Peters et al., 1995) prior to interaction testing, we are able to make comparisons within our samples based on this visuospatial variable, along with an analysis of sex and educational background influences on interaction performance. By the time of the conference we intend to present results from a more sophisticated process whereby we can statistically parcel out the variability due pre-existing "mental" rotation performance from the actual hands-on task and thereby get more definitive measures of 3DUI effectiveness and efficiency that could perhaps better inform interaction design.

2 Initial Experimental Tests

2.1 General Procedure

It has been reported that users with varying backgrounds and spatial abilities exhibit different behaviors when interacting with computers (Gomez et al., 1983; Snyder & Harris, 1993). Spatial ability in our experiments was assessed using the MRT. The 24-Item MRT (Vandenberg & Kuse, 1978; Peters et al., 1995) was administered and scored using standard procedures, as a measure of the subjects' visuospatial ability. The MRT was delivered on a computer monitor and subjects used a mouse to click on the correct answers in a multiple choice format. In this test, subjects are required to match 2D perspective drawings of block configurations to identical target drawings. As has been consistently reported in the literature, sex differences are often found on large-sample studies of spatial ability (Voyer et al., 1995). In the initial experiment conducted at the University of Southern California, an equal number (n=10) of male and female university student subjects with diverse educational backgrounds were tested. In the second experiment conducted at Pohang University in Korea, a similar number of male and female university student subjects were tested (n=12 for each sex), however, their educational background was exclusively in the area of engineering.

After collecting background information and administering the MRT, the subjects were then asked to carry out 3D object selection and manipulation tasks using different interaction devices and methods, administered in counterbalanced form. The selection and manipulation tasks used the same block stimuli as are found in the mental rotation problems in the MRT. In this "hands-on" task subjects used the interface devices to rotate and translate fully controllable blocks into superimposition with an identical block configuration target stimulus (Figure 3).

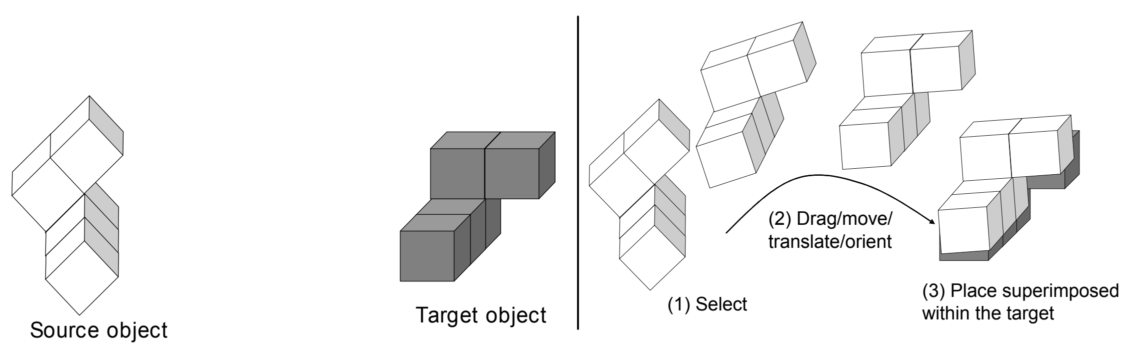

Figure 3: Selection and manipulation task for testing interaction methods.

Subjects were tested on all interaction methods using a within subjects design for both experiments, with the order of interaction tests being counterbalanced across the experiments in an effort to control for carry-over effects on learning from the previous interaction tests. Prior to each interaction test, the subject was briefed about the purpose of the experiment, given instructions on how to use the interface, and were administered a short 10-item pre-test training session to get familiarized with the interface. This pre-test used very basic block configurations in very simple orientation and depth rotations for interface familiarity training purposes. The familiarity training was then followed by a more complex 24-item task. In our initial experiment (USA), users were presented with block selection and manipulation tests using both a 3-button mouse 2D interaction method and a simple gesture-based system, in counterbalanced fashion. The second experiment (Korea) replicated the original mouse and gesture interaction method tests, with the addition of the Arcball method as a second form of mouse-based interaction.

2.2 The Three Interaction Methods

Three interaction methods for 3D object selection and manipulation were tested for their usability across the two experiments presented in this paper. Two methods were mouse-based and the third used a six degree of freedom (DOF) tracked glove. The first mouse-based interaction method (hence called the "Standard Mouse") controlled the 6 DOF by combinations of various mouse buttons and hand movement. Table 1 summarizes the functionalities and the required actions for the standard mouse method.

Functionality		Action	Prior action needed
Selection		Press left, middle, or right mouse button	N/A
Release		Release left, middle, or right mouse button	N/A
Translate	Horizontal	Drag left/right	Object must be selected by left button first
	Vertical	Drag up/down	Object must be selected by left button first
	Perpendicular to screen	Drag up/down	Object must be selected by left button first
Rotate Around	Vertical axis	Drag left/right	Object must be selected by right button first
	Horizontal axis	Drag up/down	Object must be selected by right button first
	Axis perpendicular to screen	Drag left/right	Object must be selected by right button first

Table 1: The "Standard Mouse" interaction method for object selection and manipulation.

The second mouse-based interface was an implementation of the Arcball interface originally proposed by Chen, Mountford & Sellen (1988) and also implemented and improved by Shoemake (1992). The Arcball interface uses a sphere metaphor: that is, a (visible or translucent) sphere is overlaid on the object to be rotated, and by "dragging" on the surface of the projection of the object (or the encompassing sphere), the object is rotated accordingly (like in using the trackball mouse). The amount of the rotation is controlled by the arc length between the starting point and the end point of the "drag." The starting point is captured upon a button press and the end point at the button release. Thus, the main difference between the Arcball and the "Standard Mouse" is that the Arcball does not use the middle button in rotation and the "dragging" allows rotation around an arbitrary axis. The detailed functionality is summarized in Table 2.

Functionality		Action
Selection		Press left, middle, or right mouse button
Release		Release left, middle, or right mouse button
	Horizontal (sel. by left button first)	Drag left/right
	Vertical (sel. by left button first)	Drag up/down
	Perpendicular to screen (sel. by middle button first)	Drag up/down
Rotate (sel. by right button first)		Drag on 2D projection surface of the control sphere

Table 2: The "Arcball" interaction method for object selection and manipulation.

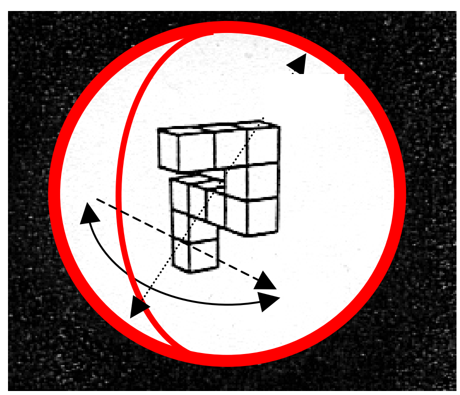

Figure 4: The Arcball interface interaction method (Shoemake, 1992). The dotted line is the dragging action on the 2D projection display surface using the mouse and the solid line represents the rotation direction around the axis of the red sphere overlaid around the object.

The third interaction method employed a 6 DOF magnetic tracked glove for gesture interaction. The grasping (and opening) gesture, inferred from the sensors on the glove, was used to select (and deselect) the object. Once the object was selected (or "grabbed"), the object could be moved and rotated according to the motion of the hand captured by the tracker (See Figure 5). Subjects could release the object by opening their hand and then "regrap" and actuate the object from another position until it was superimposed within the target stimulus.

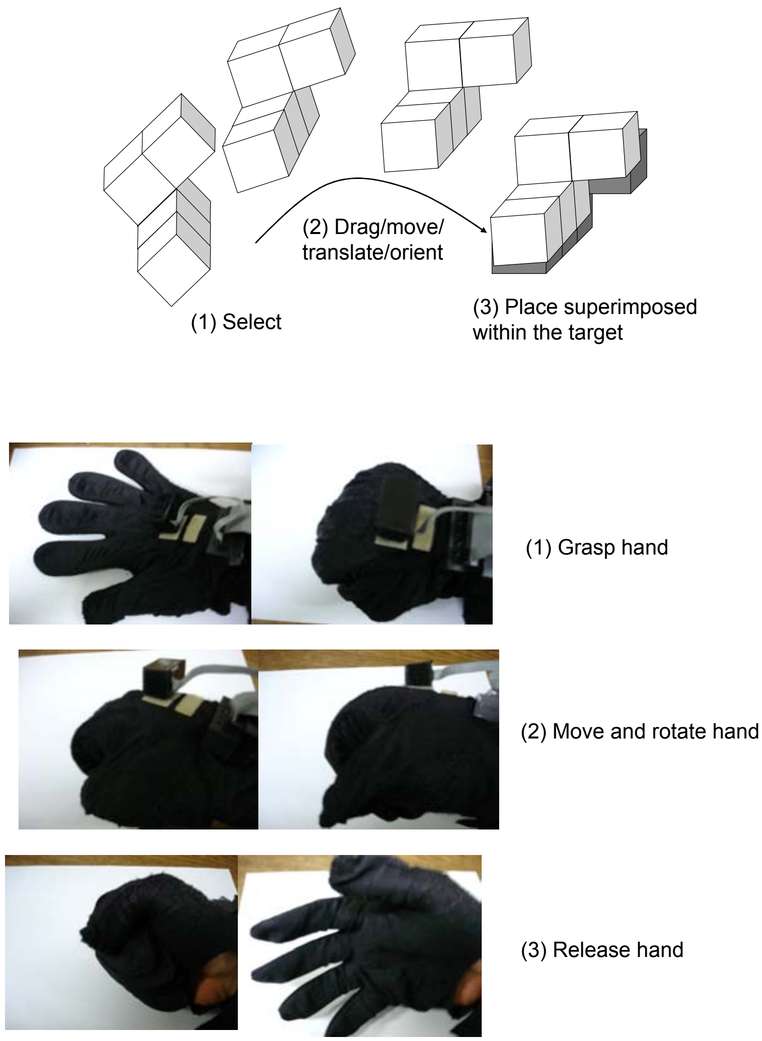

Figure 5: Task and the Tracker/Glove Interaction Method

2.3 Experiments 1 and 2

Our initial experiment was conducted in the USA and compared two interaction methods - the Standard Mouse and Tracked/Glove. We carried out an "almost" identical second experiment in Korea for three purposes: (1) to increase the number of subjects and enhance the statistical power and generalizability of the experimental results, (2) to demonstrate the feasibility of standardizing the process of interface evaluation and comparison and (3) to include another interface, the Arcball, in the comparison. The experimental set-up and conditions in Korea were identical to that of the one used in the USA, except that a third additional interaction method was evaluated and the experiment was conducted in Korean. Twenty subjects were tested in Experiment 1 (USA) (10 females and males, average age 29, 65% with prior 3D UI experience) and were recruited from the general student population resulting in a sample with diverse (>6) educational backgrounds. In Experiment 2 (Korea), 24 subjects were tested (12 females and males, avg. age 26, 63 % with prior 3D UI experience) and all of the subjects were engineering students. All participants received a compensation for their efforts (5 US Dollars at USC and 10,000 Korean Won (about 8 US Dollars) at POSTECH).

2.4 Equipment

A 20-inch CRT monitor was used as a display and Stereographics shutter glasses were used to produce the stereo effect (see Figure 6). A commonplace Logitech 3-button mouse was used for the Standard Mouse and Arcball interaction methods. For the Tracker/Glove interaction method, the Ascension Flock of Birds 6 DOF magnetic tracker and the 5DT Ultra 14 Dataglove were used.

Figure 6: Using the Standard Mouse to carry out the block selection and manipulation task.

2.5 Evaluation

Both quantitative and qualitative data were collected for analysis purpose. The quantitative data included the number of correct answers on the MRT test and for the interaction tests, the average task completion time was measured for each task (subjects carried out 20 manipulation tasks for each interaction method). After carrying out all the tasks, a questionnaire was administered to subjects that evaluated and compared the usability of the interaction methods in terms of ease of learning, cumbersomeness, fatigue and comfort, and general preference.

3 Results

3.1 Task Performance Across Interaction Methods

Since no systematic carry-over effects were found due to the order of test administration, and the experimental procedures were essentially the same, we combined the results from Experiment 1 (USA sample) and Experiment 2 (Korean sample) to compare performance across interaction methods. Overall, the glove interaction method outperformed the standard mouse and the Arcball as illustrated in Figure 7 ($p< .03$). This result was also found when the samples were compared separately. Note that this result does not consider any other possible factors such as sex, spatial ability, user background, etc.

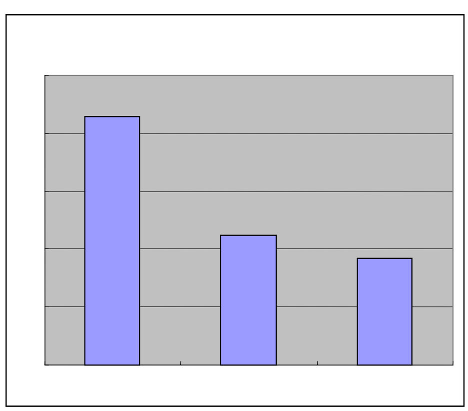

Figure 7: Average task performances across all interaction methods tested.

3.2 Sex and Visuospatial Ability Related to Interaction Method Task Performance

3.2.1 Sex and Visuospatial Ability (MRT)

In accordance with previous studies, there was a significant difference between males and females on the MRT, in favor of higher performance in males. However, while this result was found for the USA sample alone and the combined sample ($p < .01$, n=44), no significant difference was found in the Korean sample ($p < .20$, n=24) that was comprised exclusively of engineering students (see Figure 8, right graph). This suggests that the educational and/or cultural background of subjects may be a factor in this measure of spatial ability and in turn, these subject variables may exert some influence on interface performance tests.

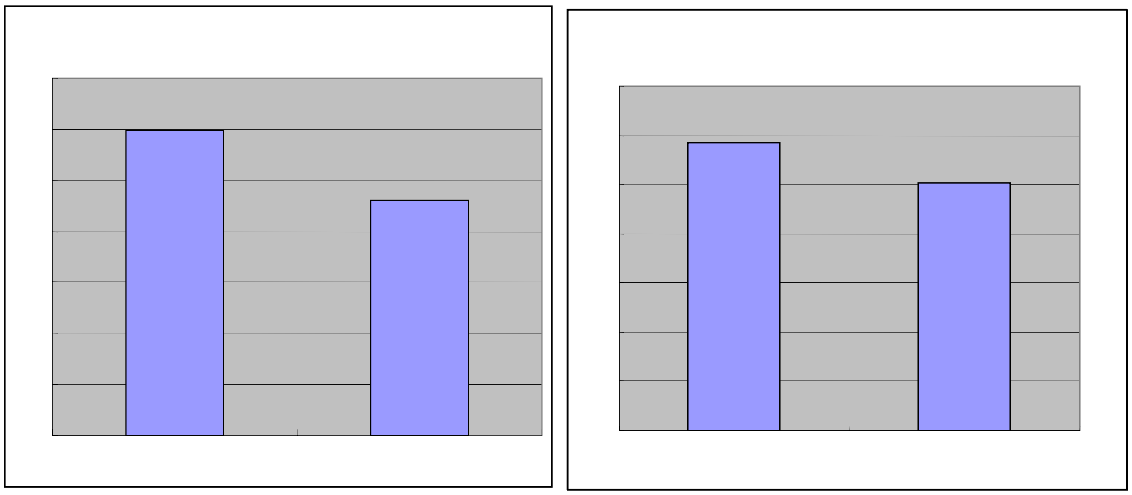

Figure 8: Difference in spatial ability and sex in two different subject pools: Left graph shows the overall MRT results from USA and Korean sample combined. Right graph shows the MRT results from the Korean sample only.

3.2.2 Sex and Task Performance

For the combined USA and Korean sample, there was no significant performance difference between males and females for the standard mouse test ($p < .12$), while the glove method did produce a difference between the sexes (p.< .008), in favor of the males. In the Korean sample of engineering students only, where no differences were found between males and females on the MRT, no sex differences were found on performance using the standard mouse or glove interaction method.

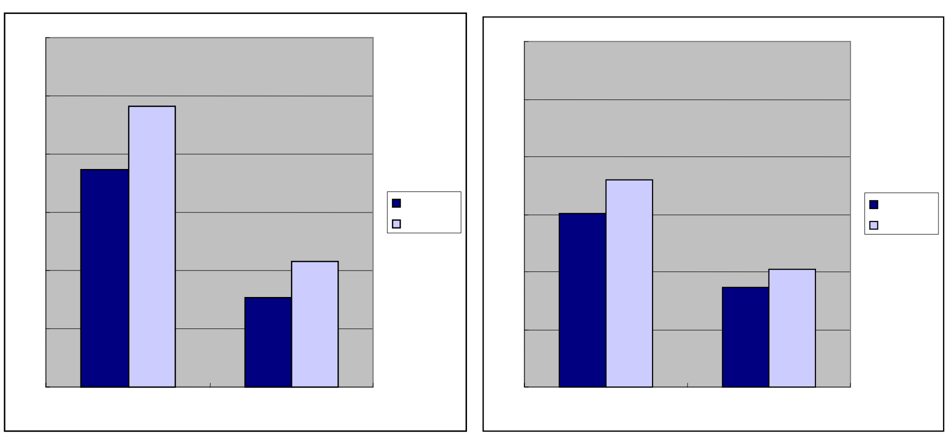

Figure 9: Task completion time: Left graph shows results from combined subjects in Experiments 1 and 2 (n=44), with no significant sex difference in performance using the mouse ($p < .12$) and a significant sex difference using the glove ($p < .008$). Right graph shows results from subjects in Experiment 2 only (n=24), with no significant sex differences in performance with the mouse ($p < .12$) and with the glove ($p < .28$).

3.2.3 Visuospatial ability and Task Performance

Visuospatial ability appears to be an important moderating factor in evaluating interface task performance. To examine this, we broke the total sample into three groups based on the MRT scores (Low, Medium, High MRT performance) and compared the time to select, manipulate and successfully superimpose the block stimuli, across the standard mouse and glove conditions. As we see in Figure 10, the group that had the lowest MRT scores (Group 1) exhibited relatively large differences in task completion time between the mouse and glove. As the visuospatial ability of the subjects gets better (group1 < group2 < group3 on MRT scores), the difference gets smaller and mouse performance gets better. This was supported by a significant correlation between visuospatial ability and the task completion time on the mouse interface ($r = -0.454$, $p = .002$). This suggests that if the interaction method is less "natural" (i.e., mouse), the users' entry-level of visuospatial ability becomes more of a factor in actual performance (i.e. higher visuospatial ability=shorter performance times). In more natural interaction methods, as seen with the glove method, visuospatial ability appears to exert little influence on performance. Also note that despite the differences in visuospatial ability, the task completion time for the glove interface is still always smaller and significantly faster with than the standard mouse. From this, we can infer that the glove interaction method was more efficient overall, and that entry-level visuospatial ability as measured by the MRT, becomes less of a limitation when using a more efficient interaction method.

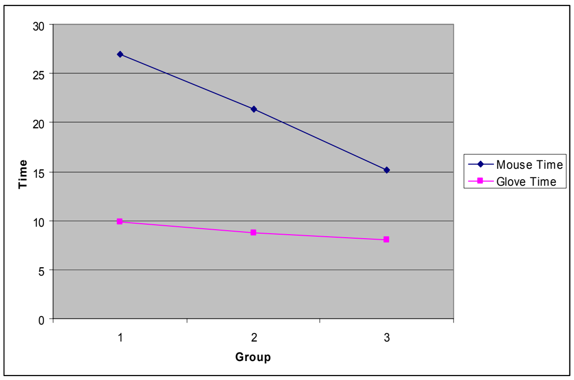

Figure 10: Visuospatial ability and task performance. Different subject groups exhibit different performance gaps between the mouse and the glove.

3.2.3 Previous 3DUI experience and task performance

Surprisingly, no statistically significant effect was found for interaction method task performance based on self-reported prior experience with 3DUI's (see Figure 11). This absence of effect was seen in all combinations of

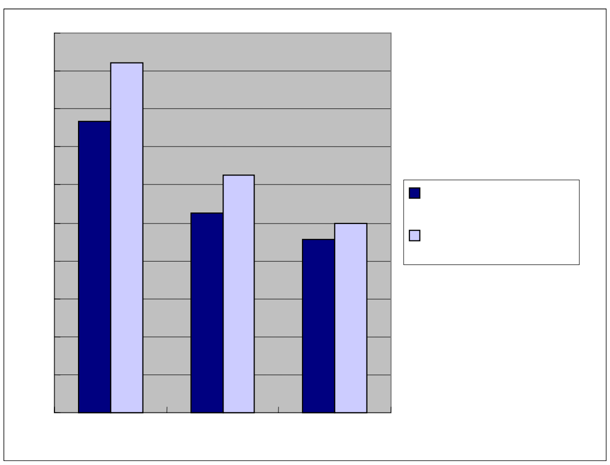

Figure 11: Prior 3D UI experience and task performance (no significant difference among task performance).

samples (independent USA and Korea and combined) and this may suggest that spatial ability, sex and educational background may exert more of an influence on interaction method performance comparisons than actual previous use of such methods. This may due to the novelty of the task compared to the 3DUI's that these subjects may have used previously, but specific data on this question was not collected.

4 Conclusions and Future Work

In this paper, we have presented initial pilot data on a benchmarking scenario for evaluating performance across three types of 3DUI devices and methods. Our results indicate that while a glove-based gesture interaction method produced faster performance than the mouse interaction methods tested, this effect was moderated by statistically significant performance differences across different user groups. For example, sex differences and its interaction with visuospatial ability appeared to be a significant factor in our results. Results from the Korean sample of engineering students revealed no differences between the sexes on entry level visuospatial ability as well as no sex differences on both mouse and glove interaction methods. Conversely, with the USA sample alone and in the combined sample, where significant sex differences occurred in visuospatial ability (as well as the additive effect of introducing more educational diversity into the sample), males performed better on the glove interaction method, but not on the less intuitive mouse method. In a previous report on usability analysis for 3D rotation techniques, a significant male performance advantage was found in terms of accuracy (Hinckley, Tullio, Pausch, Proffitt & Kassell, 1997). In their study, subjects rotated a 3D model of a house using the Arcball. Although it is possible that the model of a house, due to its familiarity, could affect the user in terms of preferred orientation and variance in the task performance, there are other reports of consistent sex differences in human visuospatial performance across a wide range of stimulus types (Geary, Gilger & Elliot-Miller, 1992; Snyder & Harris, 1993; Voyer et al., 1995). However, there is evidence to suggest that this robust sex difference may be less pronounced when comparing individuals with higher overall spatial skills. For example, Turner (1998) tested male and female engineering students on a mental rotation task and found the typical male superiority effect, but found it to be significantly smaller between males and females within a subset of higher spatial performers. Thus, visuospatial ability may vary depending on the subject's sex, educational background and on other related spatial ability skills and needs to be considered in the 3DUI evaluation equation.

Our results also suggest that visuospatial ability may also moderate 3DUI performance regardless of sex and background when the interaction method is less intuitive or "natural". Across all subjects in our samples, those with higher visuospatial abilities performed significantly better on the standard mouse test, while no impact from this factor was seen on the glove interaction performance. This suggests that if the interaction method is less "natural", the users' entry-level of visuospatial ability becomes more of a factor in actual performance. When using more natural interaction methods, as seen with the glove method, increasing visuospatial ability appeared to exert little influence on performance. Yet visuospatial ability still differentiated group performance better than self-reported previous use of 3DUI's. Our choice of the MRT as a standard assessment of visuospatial ability was based on its consistent history of producing reliable sex differences and as a pure "mental" measure of the actual hands-on block selection and manipulation benchmarking task that we have modeled. As well, the use of such generic block configuration test stimuli may control for the bias that could be introduced into the assessment if one was to use more familiar stimuli based on "real-life" objects. However, spatial ability is a complex process that can be assessed using other measures and our future research will explore other facets of spatial ability that could provide added value for understanding the complex relationship between the user and the 3DUI.

Although these preliminary results suffer from a relatively small sample size and potentially confounding (or illuminating!) cross-cultural influences, they are encouraging for future research in this area. These results provide evidence that one's sex, entry-level visuospatial ability and educational/cultural background could exert an influence on performance comparisons across interaction methods, perhaps even more than self-reported previous experience with 3DUI's. Our continuing research in this area involves applying our benchmarking technique to other 3DUI's and display configurations (e.g. auto-stereoscopic display comparisons) (Alpaslan, Yeh, Rizzo & Sawchuk, 2005). We are also exploring the addition of other types of user background information that could serve to produce a more precise accounting of the underlying dynamics of 3DUI performance and to help guide the process for determining more optimal ways for users to interact with 3D content. It is our view that systematic investigation of such variables on actual task performance via the use of standardized benchmarking tests needs to occur and that these factors should be accounted for in the 3DUI evaluation and analysis process.

References:

Alpaslan, Z.Y., Yeh, S.C., Rizzo, A.A. & Sawchuk, A.A. (2005). Quantitative Comparison of Interaction with Shutter Glasses and Autostereoscopic Displays. In: *The Proceedings of the 16th Annual SPIE International Symposium on Electronic Imaging: Science and Technology*, San Jose, CA.

Bowman, D., Kruijff, E., LaViola, J. J. & Poupyrev, I. (2001). An introduction to 3-D user interface design. *Presence: Teleoperators and Virtual Environments*, 10, 96-108.

Chen, M., Mountford, S. J., & Sellen, A. (1988). A study in interactive 3-D rotation using 2-D control devices. *SIGGRAPH '88,* 121-129.

Geary, D. C., Gilger, J. W., Elliot-Miller, B. (1992). Gender differences in three-dimensional mental rotation: A replication. *Journal of Genetic Psychology*, 153, 115-117.

Hinckley, K., Tullio, J., Pausch, R., Proffitt, D. & Kassell, N. (1997). Usability of 3D rotation techniques, *Proceedings of the 10th annual ACM symposium on User interface software and technology*, 1-10.

Shoemake, K. (1992) ARCBALL: a user interface for specifying three-dimensional orientation using a mouse. *Proceedings of the conference on Graphics interface* '92, 151 – 156.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R. & Richardson, C. (1995). A redrawn Vandenberg and Kuse Mental Rotations Test: different versions and factors that affect performance. *Brain and Cognition, 28,* 39-58.

Preece, J., Rogers, Y. & Sharp, H. (2002). *Interaction Design*. (1st Ed.) Hoboken, NJ: Wiley Text Books.

Psotka, J. (1995). Immersive training systems: Virtual reality and education and training. *Instructional Science,* 23, 405-431.

Rizzo, A.A., Buckwalter, J.G., Bowerly, T., McGee, J., van Rooyen, A., van der Zaag, C., Neumann, U., Thiebaux, M., Kim, L., Pair, J. & Chua, C. (2001). Virtual Environments for Assessing and Rehabilitating Cognitive/Functional Performance: A Review of Project's at the USC Integrated Media Systems Center. *Presence: Teleoperators and Virtual Environments.* Vol. 10 (4), 359-374.

Shepard, R. N. & Metzler J. (1971). Mental rotation of three-dimensional objects, *Science*, 171, 701-703.

Snyder, P. J. & Harris, L. J. (1993). Handedness, sex, and family sinistrality effects on spatial tasks, *Cortex*, 29, 115-134.

Stanney, K. (2002). Handbook of Virtual Environments, New York: L.A. Earlbaum.

Turner, G.F.W. (1998). The effects of stimulus complexity, training, and gender on mental rotation performance: A model-based approach. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 58, 6865.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47, 599-604.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, 117, 250-270.