The Effect of Interaction-Display Offset on User Performance for a 3D-Widget Task in the Cave

Dmitri Lemmerman

May 2006

Abstract

Interaction in Immersive Virtual Environments (VEs) is fundamentally driven by the physical actions of the user's body. The location of rendered virtual objects is, in contrast, the free choice of the VE application designer. The goal of Dmitri Lemmerman's masters degree project is to explore the effect of the position of rendered graphical "widgets" on user performance in the use of those widgets.

The particular task chosen was color-matching: using a 3D color picking widget, users were asked to match a target color. The independent variable was a scalar describing the offset between the interaction frame-of-reference (where the user moves his physical body) and the display frame-of-reference (where the graphical feedback appears). This relationship is not unique to the color picking widget, but generalizes to a larger set of 3D widgets in VEs.

In the experiment, the relative position of the widget with respect to the user is tested in three conditions. While maintaining the same interaction frame-of-reference, the relative position of the display frame-of-reference is varied: to be the same as the interaction frame-of-reference (e.g. collocation), to be displayed at a 3 inch offset, and to be displayed at a 2 feet offset from the interaction frame-of-reference.

Twenty-nine users were tested in a within-subjects design. Each subject was asked to match 15 colors in each of the 3 conditions. A distance scalar was computed based on a perceptual metric. A repeated measures ANOVA test on the data revealed a significant main effect of offset with a confidence of $p < .05$. Paired sample t-tests between the conditions did show a significant difference between the collocated case and each of the offset cases, but did not show a significant difference between the two offset cases.

These results suggest that the presence of the user's body appears to degrade performance when it is collocated with the virtual objects on which the user must focus (as it was in one of the conditions). However, even a short offset of 3 inches seems to alleviate this problem.

Studies in HMD-based VEs have demonstrated significant performance advantages to collocation and the "direct manipulation" of virtual objects. Unlike HMD-based systems, a Cave-based VE platform allows the user to see both his body as well as the virtual objects. The results of this study should inform future VE user interface design for real-world applications and spur further experimental research in human/VE interaction principles.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Immersive virtual reality, as a technological idea, dates back to the work of Ivan Sutherland and his proposal for "The Ultimate Display." *Sutherland* (1965) In this seminal work, Sutherland hypothesized that "the ultimate display would be a room within which the computer can control the existence of matter." Although this 'ultimate' goal is still far from realization, significant progress has been made on technology that simulates the existence of matter in a virtual environment.

Such simulations are achieved through the piecewise stimulation of various aspects of the sensory modalities. In later work, Sutherland conceived and prototyped a system which provided a sensory stimulation characteristic of nearly all modern virtual reality systems: visual stereopsis. *Sutherland* (1968) Stereopsis refers to the sense of depth unconsciously perceived in the world as a result of seeing slightly different images in each eye (see pp. 38-40 *Bowman et al.* (2004) for a discussion of stereopsis in relation to 3D user interfaces; see *Cruz-Neira et al.* (1993) for a discussion of the characteristics of stereo display in Cave-based virtual environments). Sutherland's early system used two separate CRTs to present these distinct images to each eye. Additionally, Sutherland's system implemented another sensory stimulation crucial to most modern virtual reality systems: head-tracked perspective projection. Specifically, he utilized an ultrasonic tracker to determine the user's head position and orientation in 3D space and programmed the graphical feedback to render the 3D objects stored in memory as if they were being viewed from that position (with the additional stereo offset for each eye). This 'kinetic depth effect' combined with stereoscopic display constitutes a suitable baseline level of technology for what are called Immersive Virtual Environments.

The control of matter suggested in Sutherland's 'ultimate display' makes implicit the assumption that the user may interact with this matter just as with its everyday equivalent. This assumption is not valid in current VEs as only an image of matter is displayed. Without additional feedback simulation hardware, this image will only be accessible as such to the user. It cannot be touched or heard or smelt. This lack of material realism necessitates a mediated interaction between the user and the virtual environment. Mediated interaction in VEs can be provided via myriad hardware and software combinations. Overall, the study of such interaction techniques has given rise to the research field of 3D User Interface Design.

With the maturation of the required hardware for effective immersive VEs, 3D User Interface Design emerged as Human-Computer Interaction (HCI) experts and VE application researchers commingled. *Bowman et al.* (2004) Three major thrusts are evident: the development of new interaction techniques for VE applications, theoretical consideration of these techniques (e.g. their classification into a taxonomic structure), and interaction technique evaluation. The focus of this work is on the last of these activities. The literature describing novel interaction technique development for general and specific applications is vast and cannot be summarized in a single source. *Bowman* (1999), *Pierce* (2001), and *Bowman et al.* (2004) each provide thorough discussions of interaction technique development.

No widely accepted paradigm for VE interfaces has been adopted. Various approaches to interface design have nevertheless been proposed. One such approach to interface design is to think in terms of body-centered interaction. Body-centered interaction was first introduced by *Slater and Usoh* (1994). Using a navigation task, they designed an experiment to test the effect of a body-centered travel technique on the user's sense of presence. Presence is a psychological state widely held to be achievable by users of immersive virtual environments. Though the user is not physically located in the scene being simulated, he may nevertheless have a sense of being there. Slater and Usoh describe an immersive environment as a necessary but not sufficient condition for establishing a sense of presence. They hypothesized that using a body-centered navigation technique would have a positive correlation with the degree of presence felt by the subject. In a later experiment to test the effectiveness of several interaction techniques in a selection and manipulation task, *Poupyrev et al.* (1997) introduced a body-centered coordinate system. The current research is focused on testing an intrinsic property of many VE interaction techniques in the context of Poupyrev's coordinate system.

The user necessarily must perform physical actions in some frame-of-reference near his body (referred to as the interaction frame-of-reference in this work). The feedback provided by the VE, however, can be displayed in an arbitrary frame-of-reference at the discretion of the VE application designer (the display frame-of-reference). The relationship between these two frames-of-reference in a body-centered framework shall be explored. Specifically, task performance using a 3D widget will be measured under varying translational offsets of the display frame-of-reference with respect to the interaction frame-of-reference (see *Conner et al.* (1992) for a discussion of UI widgets and specifically 3D widgets in VEs). The results of this study may serve as a guide for VE application designers who need to optimize the display frame-of-reference for a particular 3D interface task.

Chapter 2

Related Work

2.1 Overview

Many approaches have been used to evaluate VE interaction techniques. A broad overview of the motivation and methods of VE interface evaluation can be found in Chapter 11 of *Bowman et al.* (2004). Formal user studies with a sample of the population are a commonly employed method for VE interface evaluation. Typically, during the user study one or more performance measurements will be taken as the user completes a task. Quantitative results garnered by these formal studies provide the approach with a statistical rigor absent in more *ad hoc* or heuristic evaluation schemes.

The purposes of formal VE user studies fall into several increasingly specific categories: to compare different VE hardware configurations (or to compare VE and traditional desktop hardware), to compare different interaction techniques for a VE task, or to test particular parameters of a given interaction technique. This categorization embodies a natural progression of specificity. Since a VE hardware platform is a considerable investment, any particular choice in its acquisition demands justification. User studies that vary attributes of the VE hardware device for a given task or application serve this purpose. Examples include *Swan et al.* (2003) and *Schulze et al.* (2005).

Once a VE hardware platform has been chosen at a site, the VE designer's next goal is to create applications utilizing that platform. Whether these applications are reseach projects, commercial products, or experimental studies, this goal includes the selection of appropriate interaction techniques for every task needed in the application. User studies that vary the interaction technique for a given task in a given VE environment help to inform interaction technique choice. Most often studies are performed to test the effectiveness of an author's new interaction technique against other known techniques that accomplish the same task in a different manner.

As effective interaction techniques become established, the next logical step is to further optimize the performance of these techniques by determining the best choices for all available parameters. Of course, many parameters exist which are unique to a particular interaction technique, but some are common to multiple techniques. User studies that vary these common parameters of a VE interaction technique not only serve to optimize that technique, but also can assist in understanding the effect of that parameter on other interaction techniques. Examples include *Poupyrev et al.* (2000). The current work exists in this final category of user studies. Though a complete history of VE user studies is beyond the scope of this work, a synopsis of notable results is presented.

2.2 Comparisons of Virtual Environments, traditional workstations, and the real world

Although the technological achievement of a VE system dates back to Sutherland, its development remained nascent for several decades while the enabling hardware components matured. Beginning in the early 1990's, VE systems had become common enough to support inquiries into their relationship with traditional workstations, with other VEs, and with the systems they were often times asked to emulate. The following describes a selection of key findings.

Slater et al. (1996) designed an experiment to test egocentric and exocentric immersion on task performance. For Slater, immersion was a factor that varied along with the capabilities of a given VE hardware system. The egocentric versus exocentric distinction is commonly used to describe the user's viewpoint. Egocentric views are defined relative to the user's body; exocentric viewpoints are defined relative to some external frame-of-reference. In his particular experiment, Slater used a head-mounted display (HMD) as an exemplary egocentric immersive VE and a television screen for the exocentric VE. Subjects were asked to observe a series of between seven and nine moves in a 3D chess-like game. They were then asked to repeat the sequence. It was found that those subjects who performed the task in an egocentric manner achieved significantly more accurate repetition of the observed sequence of chess moves than those who performed in the exocentric manner.

The difference between traditional and HMD display for a VE task was further explored by *Pausch et al.* (1997). They developed a search task where subjects were asked to find a camouflaged letter on the walls of a cubical room. Half of the subjects were allowed to freely move with the HMD, their viewpoint updated accordingly; the other half used the HMD for viewing, but were not able to move their heads.

Instead, the viewpoint was updated using a six degree-of-freedom hand tracker. Two situations were presented to the user. In one the target was presented somewhere in the scene and the user was asked to locate the target; the second involved the user determining if in fact a target letter was present in scene or not. No difference was found in the time it took subjects to find a specific target between the two conditions. However, subjects were able to confirm whether or not a target was present in the scene significantly faster using the head-tracked condition. Pausch and his collaborators believed that this increased performance resulted from the superior ability of a subject to remember where he had already looked when he controlled his viewpoint using natural head movement rather than the mapping provided by the hand tracker.

Experiments have also been performed to explore the differences between traditional desktop and Cave displays. *Swan et al.* (2003) utilized a collection of tasks centered around the use of a battlefield planning map. The display device was employed as a factor in an experiment where visual feedback was presented to the subjects using a desktop monitor, a workbench display, a powerwall display, or a four-walled Cave. They discovered a main effect that the workbench always performed worse than the other VE displays.

2.3 Comparisons of interaction techniques

An interaction technique is a way of accomplishing a certain task in a given virtual environment. Different hardware setups for any given deployed virtual environment are possible. After the necessary first step of choosing the components for a particular configuration new choices arise. The next step of analytic granularity leads to the comparison of different interaction techniques to accomplish different virtual environment tasks. This research thrust is exemplified in the testbed evaluation approach put forth by Doug Bowman. *Bowman et al.* (2001) In the testbed approach a task is chosen and the interaction technique itself is used as a factor in a user study. Several different interaction techniques, each capable of accomplishing the task in a different manner, are presented as experimental conditions, and statistics characterizing user performance at completion of the task are recorded. Additionally, Bowman introduced a taxonomy of common virtual environment tasks, establishing a categorical framework towards evaluating a wide range of tasks. *Bowman* (1999)

An example of testbed evaluation performed by Bowman employed the selection and manipulation of a cube as the task. A cube to be selected, identified by color, was placed among eight other cubes in a three-by-three array. A target cube was rendered somewhere else in the environment. The task had two parts. The subject was instructed to select the colored cube and to manipulate it such that it was aligned with the target. To test different interaction techniques, a between subjects experimental design with two factors each of three levels was employed. Within-subjects factors for several parameters of the task were used as well. In this research, Bowman demonstrated that the selection technique known as "Go-go," described in *Poupyrev et al.* (1996), performed significantly better in the selection part of the task than a ray-casting based technique or an occlusion based technique.

2.4 Comparison of parameters for a given interaction technique

User studies that focus on the differences between virtual environment hardware have been examined. Additionally, those comparing different interaction techniques given a constant virtual environment setup have been discussed. Another class of user studies are those which compare different parameter choices for a given interaction technique. These seek to further refine the understanding of the performance characteristics of a given technique. Moreover, knowledge gained from such studies can ideally be applied not just to the particular interaction technique in the user study, but also to a more general class of other VE interaction techniques.

Poupyrev et al. (2000) compared subject performance in a manipulation task requiring rotation. After developing an interaction technique that amplified rotations (i.e. the angular velocity of the virtual object in the display frame-of-reference during rotation exceeded that of the user's hand in the interaction frame-of-reference by a constant amount). They then performed an experiment comparing amplification as a factor. The formal study compared no amplification and an amplified condition for the task of rotating an object to a target location. Additionally, Poupyrev performed an empirical study to arrive at the chosen coefficient of amplification used in the amplified condition. Poupyrev and his colleagues found that the task could be completed significantly faster under the amplified condition, but that the error (measured as distance between the target and chosen location) also increased in the amplified condition.

Examples of user studies exploring interaction technique parameter choices are prevalent in desktop interaction research as well. One such study is presented by Bill Buxton in his "marking menu" studies. *Kurtenbach and Buxton* (1993) A marking menu is a radial menu that is used with a mouse by dragging outward in a particular radial direction to make a desired selection. Hierarchical implementations are possible. Buxton's user study explored the factors of number of selectable items per menu and levels of menu hierarchy in specific tasks to acquire performance data. In this way, the implementation parameters available to any marking menu-based system were analyzed. The current work sits in this final category of analytic approaches, but for virtual environments.

2.5 Body-centered interaction

The previously described works illustrate a progression in virtual environment research from choosing VE hardware deployments to optimizing parameters for particular interaction techniques. Body-centered interaction exists as one framework in which to understand the analysis of virtual environment interaction. Body-centered interaction, as an idea, was first put forth by Mel Slater in his description of a user-study comparing different interaction techniques for a locomotion task. *Slater and Usoh* (1994) Central to adoption of this idea is the belief that for certain tasks user performance in a virtual environment will be best when the feedback provided by the environment closely matches the physical actions undertaken to accomplish the task. Slater felt that those interfaces that maximized the matching between the given feedback and the feedback expected based on the user's proprioceptive sense would maximize the feeling of presence. Slater and his colleagues designed an experiment that questioned whether a user's sense of presence in a virtual environment was affected by whether or not the employed movement technique was "body-centered." They compared a joystick based technique, a walking-in-place technique, and an actual walking (i.e. one-to-one) technique, and found that user's reported a significantly increased sense of presence in the body-centered actual-walking approach.

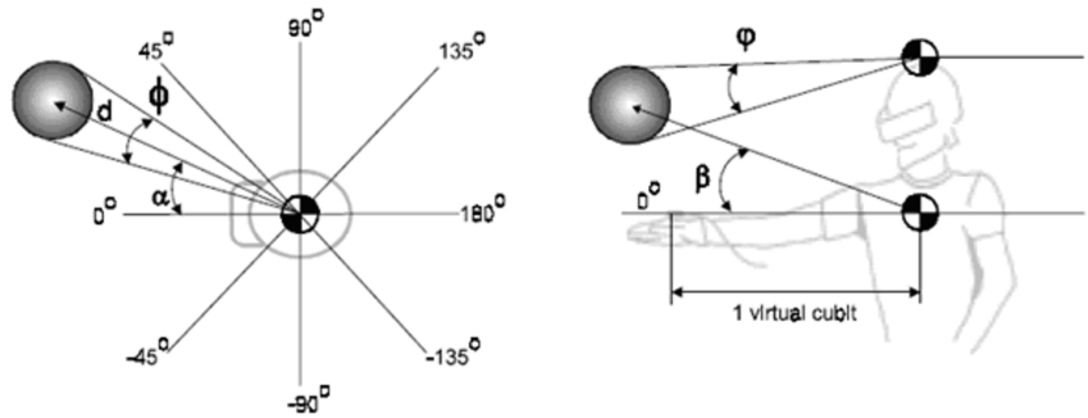


Figure 2.1: Body-centered coordinate system introduced by Poupyrev. *Poupyrev et al.* (1997)

The body-centered approach to analyzing virtual environment interaction was further quantified by Ivan Poupyrev. With his colleagues, Poupyrev established a body-centered coordinate system in terms of which to conduct user studies. *Poupyrev et al.* (1997) The height and reach of the user were employed as intrinsic units. Additionally, the viewing angle with respect to the object being viewed was introduced as an additional parameter within the body-centered framework (see Figure 2.1). This framework was used to establish factors for a user study testing the effect of target distance in a selection task. The current study borrows from Poupyrev's terminology to describe the parameters used as experimental conditions.

Chapter 3

Experimental Design

3.1 Overview

An experiment was designed to test the effect of varying the offset between the display frame-of-reference and a constant interaction frame-of-reference for a 3D widget-based task. In terms of Poupyrev's user-centered coordinates, only the radial distance parameter of the widget's displayed feedback was changed (see Figure 2.1). The angular parameters were held constant. A constant height of $\frac{7}{10}$ of the user's height was maintained as the height of the displayed widget's center, resulting in a constant value for β, the elevation angle relative to the user. Poupyrev's α, the user's azimuth in user-centered coordinates, was set such that the user was always facing forward in the VE system.

Specifically, this system was a four-walled Cave. *Cruz-Neira et al.* (1993) It was driven by a five-node cluster of Linux operating system, Intel Xeon 2.8 Ghz processor machines using the cluster synchronization platform described in *Lemmerman and Forsberg* (2004). Active stereo imagery was provided via Nvidia FX 3000G graphics cards synchronized by infrared with Sterographics CrystalEyes3 LCD shutterglasses. The physical size of each wall was 8 ft. squared with a display resolution of 1024 x 768. Four Marquee Electrohome 9500LC projectors, one per wall, provided images updated at 85 Hz (42.5 Hz for each eye). The brightness of each projector was approximately 200 lumens. Head and hand six degree-of-freedom position and orientation information was acquired using Polhemus Fasttrack magnetic trackers. All distances were specified in terms of this tracking system's intrinsic coordinate system, which was calibrated by the experimenter with assistance from the Brown University Cave support staff prior to the study. The head tracker was attached to the side of the Stereographics glasses. The hand tracker was affixed to a wireless mouse, which provided several buttons and a two degree-of-freedom joystick to support interaction. (see Figures 3.4, 3.2, and 3.3)

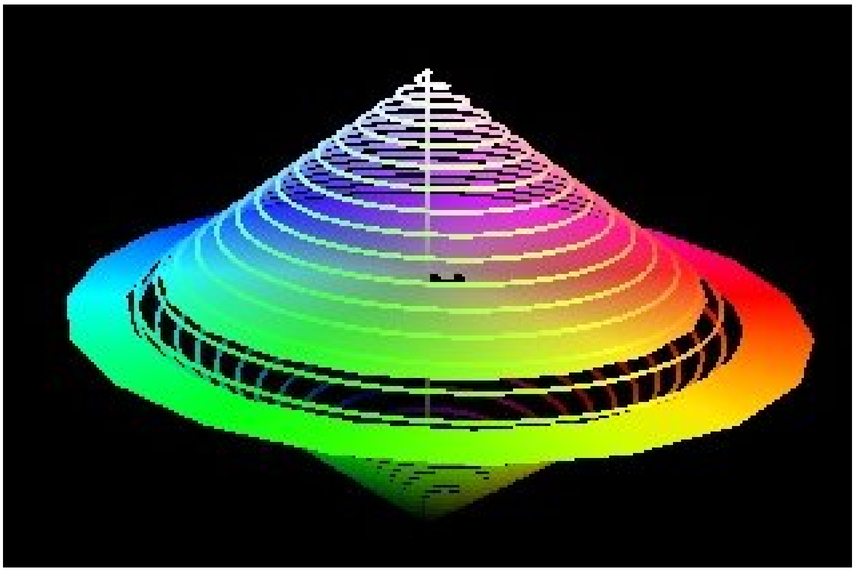


Figure 3.1: Cavepainting color-picking widget.

Given the goal of investigating the relationship between the display and interaction frames-of-reference, an example 3D widget and corresponding task was needed. A 3D color-picking widget based on that from the CavePainting application was chosen. *Keefe et al.* (2001) The widget maps the position of the user's hand to a color. Using cylindrical coordinates in the interaction frame-of-reference, the user's hand position determines a point in HSV color space: the hue of the color is determined by the angular position around the cylinder, the saturation by the radial distance, and the value by the height. In the display frame-of-reference, a double cone is rendered with a spiral wireframe. A spherical cursor colored with the current color and outlined in black is rendered at the current color's position in the widget. Additionally, an isoplane of constant color value (i.e. brightness) equal to the current color's value is rendered. (see Figure 3.1)

Color-matching was chosen as an example task utilizing the color-picking widget. Subjects were presented with two horizontally-adjacent rectangular blocks positioned slightly above the color-picking widget. One of the boxes was colored with the target color and did not change. The other was colored with the currently selected color of the color-picking widget. The user was asked to move his tracked hand to position the cursor in the color widget such that the currently selected color matched the target color, at which point he was instructed to press a button on the wireless mouse. He was asked to perform this task as quickly and accurately as possible for 45 trials.

In order to prevent the location of his hand after a trial from influencing his performance on the following trial, the user was required to complete a hand-centering task prior to each color-matching task. During the centering task the user was presented with two flat discs. One maintained its position and orientation as the target; the other was isomorphically translated and rotated according to the user's hand motion. When the movable disc was within a threshold of angular and translational distance from the target, a green outline appeared around both discs, and the user was instructed to click to start the color-matching trial. Two inches of translational offset and 10 degrees of angular offset (with respect to each axis) were chosen as thresholds for centering. These thresholds were measured in terms of the intrinsic orientation measurements provided by the six degree-of-freedom Polhemus tracker attached to the wireless mouse held by the user.

The independent variable was the offset distance between the display frame-of-reference and the interaction frame-of-reference. Three conditions were tested: no offset, a short offset, and a long offset. The trials were divided into three blocks of 15 trials each. Each block used a different condition. 15 colors were chosen at random. Each color was presented as a target exactly once per block. The ordering of the 15 colors was randomized for each block and each subject, although the same 15 colors were used for all subjects. For each trial, time to completion and the subject's chosen color were recorded.

When the offset between the interaction and display frames of reference is changed, the angular size subtended by the color picking widget must change as well. Decreasing the angular size of a virtual object necessarily decreases the number of pixels used to render that object. In order to avoid this confound, the color picking widget was scaled such that it maintained an approximately constant visual angle of 22.5.

Subjects were presented with each condition twice as practice prior to the recorded experimental trials. A calibration step was performed at the beginning of the experiment to measure the user's height and virtual cubit length, the basis units for Poupyrev's body-centered coordinate system.

3.2 Hypothesis

Prior to engaging in the study, it was predicted that the long offset condition would maximize user performance for the color-matching task. This hypothesis is justified based on a fundamental characteristic of Cave-based VE systems: the user's body cannot be occluded by computer graphics imagery. Borrowing a term from *Mine et al.* (1997), the no offset condition is also referred to as the 'collocated' condition. When there is no offset, the display and interaction frames-of-reference are in the same position. In this case, the physical position of the user's hand will be the same as the displayed position of the color-picking widget. Mine and his colleagues argued in favor of collocation as an effective approach using a head-mounted VR display device. The current work seeks to characterize the performance of collocation for a Cave-based VR task and offers two alternative conditions: a short and long offset between the interaction and display frames-of-reference.

The VE tracking system was calibrated such that in the collocated condition the cursor in the color-picking widget appeared at the very tip of the wireless mouse device. Because of intrinsic error in magnetic tracking, this calibration could not be maintained across the spatial range of the widget. The cursor at some times became occluded by the physical device. This is a possible confound for the collocated condition. In the short offset condition, the user's arm may still occlude part of the widget, although the cursor is always visible. In the long offset condition, the user's arm holding the tracked wireless mouse is generally not in the same visual field as the widget's visual feedback. It is believed that the absence of the user's arm in the same visual field as the widget's visual feedback leads to increased performance in the color matching task.

3.3 Pilot Studies

Two pilot user studies were performed in order to refine the protocol for the formal user study. The results of these studies are presented along with discussion of the potential problems revealed and the modifications made to address each.

3.3.1 First Pilot Study

Seven subjects were asked to complete the color matching task under two conditions: co-located and at a two foot offset. The offset distances were defined in terms of the coordinate system specified by the Polhemus tracking system. Each subject repeated the task 40 times, 20 times per condition. The presentation of conditions was arranged in blocks of 10. Four of the subjects saw a co-located block first; three saw an offset block first. Pilot subjects had a wide range of VR experience.

Time to completion, the target color, and the chosen color were measured. Using a perceptual distance between colors metric (http://www.compuphase.com/cmetric.htm), a scalar value was derived to measure the accuracy of the user's choice. Mean results along with the standard deviations are presented below.

Condition	Time	Distance
Co-located	25.12 $\pm$ 20 seconds	.23 $\pm$ .21
Offset	21.73 $\pm$ 19 seconds	.17 $\pm$ .13

Figure 3.2: The Wanda device (left) and the wireless mouse device (right).

Based on observations of the pilot users and the experimenters, several modifications to the experiment were considered. User's complained that the cursor displayed in the color picker was too obscured by the Wanda device used. To help alleviate this issue, a custom device was fashioned by attaching a Polhemus tracker (identical to that contained inside the Wanda's plastic housing) to a wireless mouse device commonly used for presentations (see Figure 3.2). This reduced the width of the handheld device by half from 2 inches to 1 inch. Furthermore, in order to test whether the visibility of the cursor was the major issue in successful performance of the color matching task (rather than the research hypothesis that the user's hand in his visual field is the major issue), it was decided that a third condition of a short offset should be included.

Subjects also reported problems seeing the darker colors. To address this potential confound it was decided that the lightness range should be limited. A range of 40% to 80% lightness was chosen. This range was chosen based on an informal usability study among several regular users of the Brown University Cave; colors outside this range were deemed too difficult the match.

For each of the 40 color matching trials in the first pilot study, a different random color was selected. The same 40 colors were used in the same order for each subject. However, since the ordering of these 40 colors remained the same for each subject while the presentation order of the conditions changed, not all subjects saw the same colors for the same conditions. Implicit in this decision was a belief that the color matching task would be uniformly difficult across all colors. Based on the feedback from the pilot subjects, this assumption proved incorrect. All subjects reported that colors with lower saturation were harder to match. Additionally, some commented that purple and blue hues were harder as they appeared less brightly. It was concluded that completely random color choices for each trial would lead to a different level of overall task difficulty for each subject. To address this color was introduced as a second experimental factor. A fixed set of random color choices would be chosen, such that each subject would see each color for each condition exactly once (i.e. during the course of the experiment a given color would be seen three times, once per condition).

3.3.2 Second Pilot Study

After completing the aforementioned modifications, the pilot was repeated. Five subjects were asked to complete the color matching task under three experimental conditions: co-located, a short offset of three inches, and a long offset of two feet between the display and interaction frames-of-reference. These offset distances were defined in terms of the Polhemus tracking system's intrinsic measurements. Each subject repeated the task 45 times, 15 times per condition. Trial blocks were of 15 trials; condition block-types were not repeated. 15 colors were randomly chosen (given the lightness constraint introduced). Each color was used as the target exactly once per condition type. The order of the 15 colors for a given condition block was randomized.

The same measurements were made for each trial as in the first pilot study:

Condtion	Time	Distance
Co-located	13.2 $\pm$ 8.9 seconds	.27 $\pm$ .22
3 Inch Offset	14.0 $\pm$ 8.7 seconds	.22 $\pm$ .19
2 Feet Offset	14.0 $\pm$ 10.7 seconds	.21 $\pm$ .17

Additionally, a final phase of the experiment was introduced to gain better insight into the subjects' ideal offset. For 15 trials after the 45 trial previously described, subjects were allowed to specify the display frame-of-reference offset. Trials would begin with a centering task using one of the three offset conditions from the first phase of the experiment (the initial offsets were blocked into groups of five trials). After successfully centering his hand in the interaction frame-of-reference (with visual feedback at one of the three offsets), the user was instructed to adjust the color picking widget to its ideal offset position. The subject performed this interaction using one-dimension of a 2D joystick device on the wireless mouse (see Figure 3.3). Once the subject was satisfied with the position of the widget, he pressed the mouse button. He was then asked to repeat the centering task, this time with visual feedback provided in the newly specified display frame-of-reference. Color matching then proceeded as before.

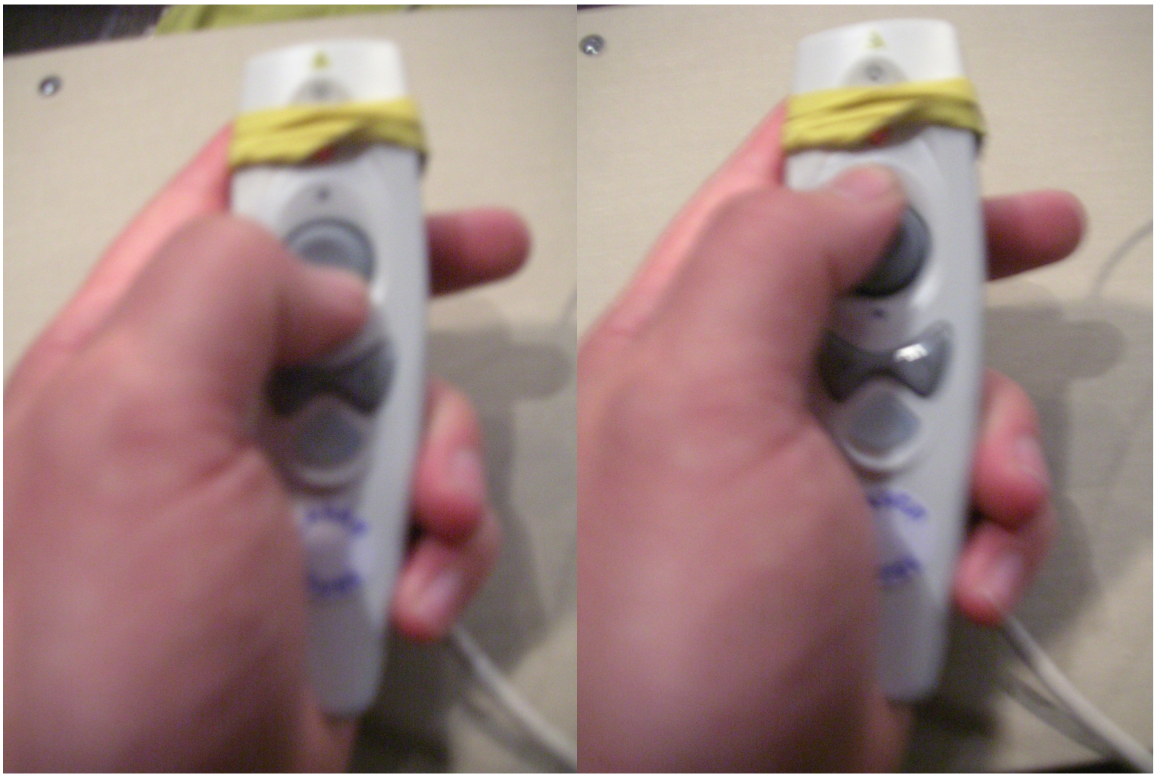


Figure 3.3: The user action needed to decrease and increase the display-interaction offset.

Three days after running the second pilot experiment, the same subjects were re-evaluated using the first phase of the experiment (matching without repositioning the widget). This was done in order to examine the learning effect for the task. For each subject, the condition-type ordering was changed from his initial color matching trials several days earlier. Results follow:

Performance improved during the subjects' second session, both in terms of the absolute time and distance measurements and in consistency (as evinced by the lower standard deviations). The overall variance between condition types decreased as well. This suggests that with practice the subjects were able to overcome some

Condtion	Time	Distance
Co-located	12.2 ± 7.9 seconds	.20 ± .15
3 Inch Offset	11.0 ± 6.9 seconds	.18 ± .14
2 Feet Offset	11.0 ± 8.7 seconds	.19 ± .12

of the detrimental effects initially encountered in the co-located condition.

3.4 Remaining Confounds

Every attempt was made to control or eliminate various factors that might affect user-performance along with the independent variable. To control for the fact that specific colors are intrinsically more or less difficult to match color was introduced as a factor, and the same set of colors were used for each condition. Since the intrinsic error in tracking can cause intermittent occlusion and visibility of the cursor itself, a condition with a small offset is provided, always allowing the cursor to be seen. Nevertheless, some factors remain that could be considered confounding.

Foremost among these is the changing viewing angle that occurs as the color picking widget's position changes relative to the subject. Since the height of the color picker remains constant, as it is further offset, the subject's view of the widget becomes more oblique. An alternative approach is for the offset to be along a constant viewing angle as opposed to along the vector normal to the front wall of the Cave, however this approach was not pursued as it does not prevent the occlusion of the widget by the user's body under any offset condition.

Given the refinements to the experimental approach decided upon over the course of the three pilot studies and keeping in mind the remaining confounds, the description of the procedure and results for the final version of the experiment follow.

3.5 Final Experiment Procedure

The protocol implemented as the final experiment included the refinements garnered from the pilot studies. Its ultimate goal was to gather performance data for a VE task while varying the location of the visual feedback relative to the user's body.

The experiment was performed over ten days, June 16th through June 25th, 2005. Twenty-nine subjects completed the study. Unlike the pilot study where subjects were primarily drawn from the Brown Computer Graphics Research Group, for the final study subjects were drawn from the Brown community at large. Subjects were solicited via signs posted around the Brown University campus. Each subject was paid at least \$10 for approximately an hour of his or her time, though several subjects took slightly longer and were paid an additional \$10 per hour prorated.

Each subject was scheduled to arrive at the Brown University Cave laboratory at a specified time. Upon arrival the subject was allowed to use a limited version of the CavePainting application for several minutes. This provided a brief acclimation period to the overall Cave virtual reality experience. The version of CavePainting used did not allow the subject to see the color-picking widget or any of its other widgets; it simply allowed painting with a preset stroke style.

After the brief CavePainting session, the subject was asked to read several paragraphs introducing the task and to complete a short pre-questionnaire (see Appendix A). Additionally, each subject completed a standard consent form. In the pre-questionnaire, each subject was asked if he or she had color vision deficiency. The questionnaire was followed by the administration of a color vision deficiency screening test. Three respondents to the experiment solicitation were rejected as participants because each was unable to pass this screening.

Following completion of the pre-questionnaire and color vision screening, each subject was led back into the Cave itself. As previously described, the Cave is a VE system which, as deployed at Brown, projects onto three walls and the floor of an 8 foot cube. Two devices were used by the subjects: CrystalEyes Stereo Shutter Glasses and a Kennsington Wireless Mouse. A Polhemus Fasttrack magnetic tracker was affixed onto each of these devices, providing six degree-of-freedom position and orientation data to the system (see Figures 3.4 and 3.5).

Figure 3.4: Views of the wireless mouse device with Polhemus tracker attached.

Figure 3.5: Views of the Stereographics shutter glasses with Polhemus tracker attached.

With the overhead room lights still on, the subject was shown the tracked mouse device, and specifically the button used on the mouse for indicating color selection during the task. The subject was then asked to begin wearing the stereo shutter glasses and complete a practice session of six color matching trials. The experimenter remained in the Cave with the subject to act as a guide during the practice.

Instructional and status messages were displayed on the front wall of the Cave. When the experiment program was launched the first message displayed read: “Click the left button to begin the experiment...” After the subject clicked, the program proceeded.

Figure 3.6: Calibration pose used to measure the subject’s virtual cubit and height.

At the beginning of the practice session, the subject was asked to perform a calibration step to measure his or her height and reach. These basis units of Poupyrev’s body-centered coordinate system were used as parameters for specifying the interaction and display frames-of-reference. The subject was asked to stand straight up in the center of the Cave, reach straight forward, and click the button. The experimenter watched the subject perform this calibration step to ensure that a proper measurement was taken (i.e. the subject approximately assumed the position shown in Figure 3.6). The status message on the front wall read: “Reach the wand straight forward and click the left button.” The height recorded was the height of the tracker attached to the glasses; the reach (i.e. virtual cubit) was the distance of the mouse’s tracker to the line perpendicular to the plane of the floor and through the glasses’ tracker.

Following the calibration step, the subject was asked to complete six practice trials of the color matching task, two per condition. Each practice trial was identical to the trials to be performed during the experiment except that the subject was encouraged to ask any clarification questions of the experimenter, who remained in the Cave. The ordering of the condition blocks was chosen to be the same as it would be during the subject's color matching trials performed on his or her own. A description of the tasks constituting one trial follows. Trials in the practice session were identical to those in the actual experiment except for the presence of the experimenter for coaching during the practice.

Each trial proceeded as follows. For a one second pause the status message read: "Get ready for the next try..." After this pause, the subject was required to perform a centering task. The purpose of the centering task was to ensure that for each trial the subject began in the same position. Centering involved two steps: ensuring that the subject was standing in the center of the Cave and subsequently ensuring that the subject's hand, holding the tracked wireless mouse, was positioned at the predetermined location with respect to his or her body.

Figure 3.7: The first step in the centering task: standing in the center of the Cave.

The requirement for standing in the center of the Cave was defined in terms of the location of the head tracker. If its position projected onto the floor of the Cave was within six inches of the floor's center, the subject was considered to be standing in the center of the Cave. Until this constraint of the centering task was satisfied, a cartoon outline of a pair of feet was displayed as a target on the floor (see Figure 3.7). Additionally a status message reading "Stand on the feet" was displayed on the front wall.

Once centered in the Cave, the subject was required to place his or her hand in the prescribed position and orientation. Two flat discs were displayed: one fixed at

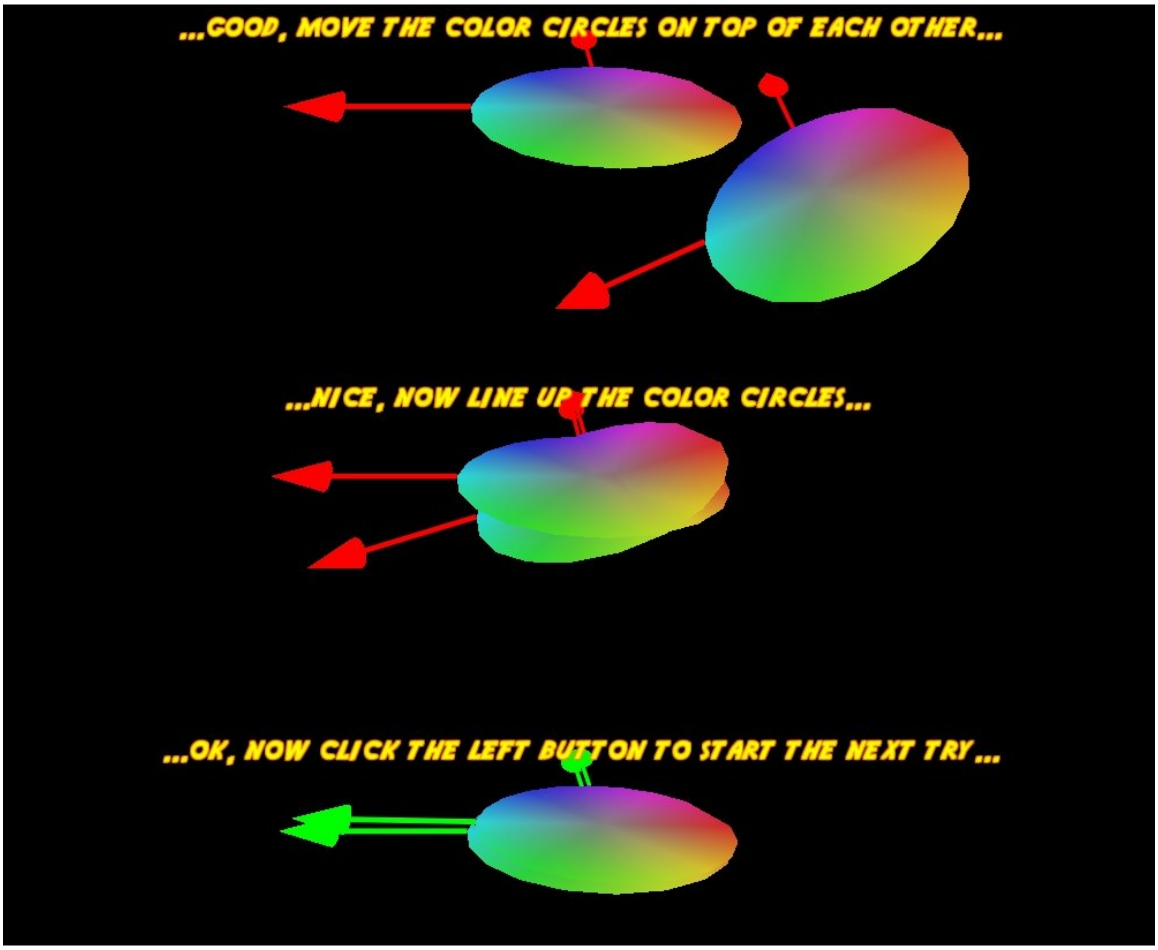


Figure 3.8: Procedure of centering task.

the target position and the other isomorphically mapped to the movement of the six degree-of-freedom tracker attached to the wand device. Additionally, two 3D arrow models were rendered, one at twelve and one at nine o'clock of disc. The arrow models were pointing outward in the direction normal to the disc's edge and in the plane of the disc (see Figure 3.8).

The arrows were colored red until the subject moved his or her hand within the centering thresholds as defined with respect to the interaction frame-of-reference: two inches of translational offset and 10 degrees of rotational offset (measured as rotational Euler angles around the three primary axes). If the subject's hand was not within the translational threshold the status message read: "...good, move the color circles on top of each other..." Once the subject satisfied the translational constraint, the status message changed to: "...nice, now line up the color circles..." Finally, once the angular constraint was also satisfied, the message became: "...ok, now click the left button to start the next try..." With the subject's hand thus in the predetermined position, the red arrows became green. Once he or she clicked the button on the wireless mouse, a single trial of color matching followed.

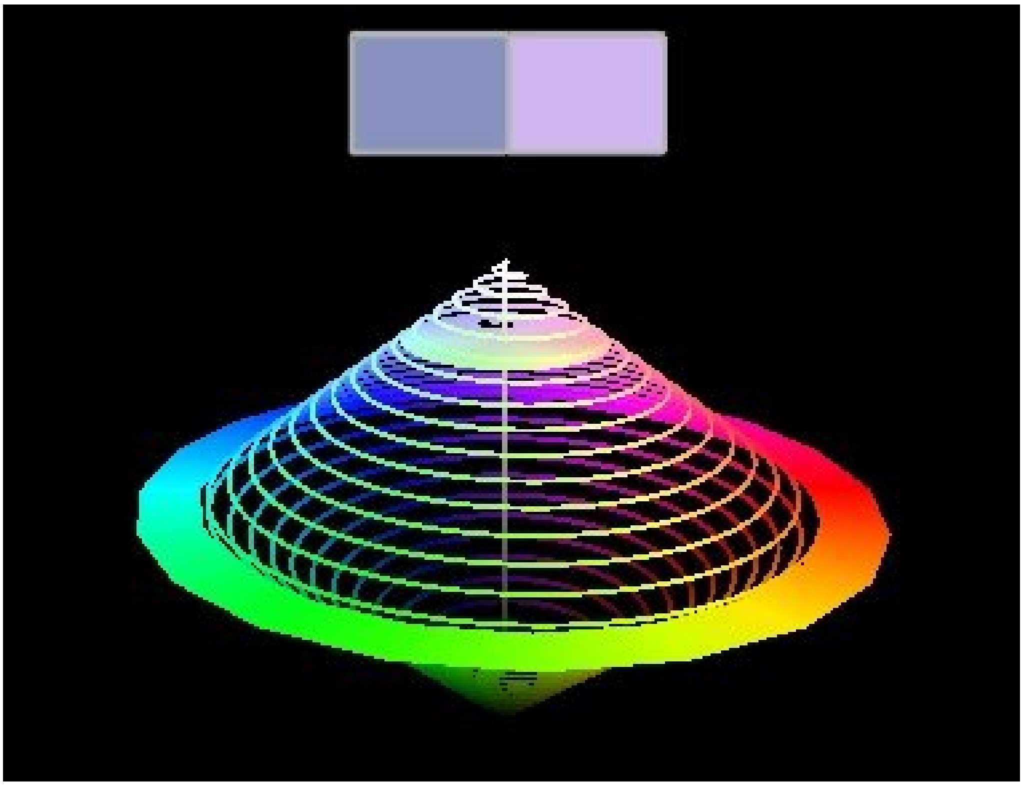


Figure 3.9: Color picking widget with target color and currently chosen color.

Immediately after completion of the centering task, the double-cone color-picking widget appeared. Additionally, the target and currently-selected color swatches appeared. These were rendered just above the top of the color-picking widget. Each was rendered as a square but flat box with a side length equal to the widget's radius and each two inches in depth (see Figure 3.9).

Based on the recommendations derived from the pilots, the target color's brightness was constrained to be between 40% and 80%. When the color-picking widget appeared, the cursor was necessarily at its center: the result of the centering task.

The color widget was rendered as follows. A line-rendered spiral was drawn along the outer edge of the cone, starting at the top tip, increasing in radius to the center plane of the widget, then decreasing in radius to the bottom tip. The color of the line was mapped to the corresponding color that would be selectable in any given location along the line. The same color mapping was displayed with a two inch ring extending outward from the center circumference of the double cones. This showed the 50% lightness (because it was halfway along the height of the cones) and full saturation (as it was along the outer radius) for all the possible hues. Additionally, an isoplane of constant lightness was rendered at the height of the cursor (the cursor's location being isomorphically mapped to that of the tracker held in the subject's hand). An example view of the widget is provided in Figure 3.9. No status message

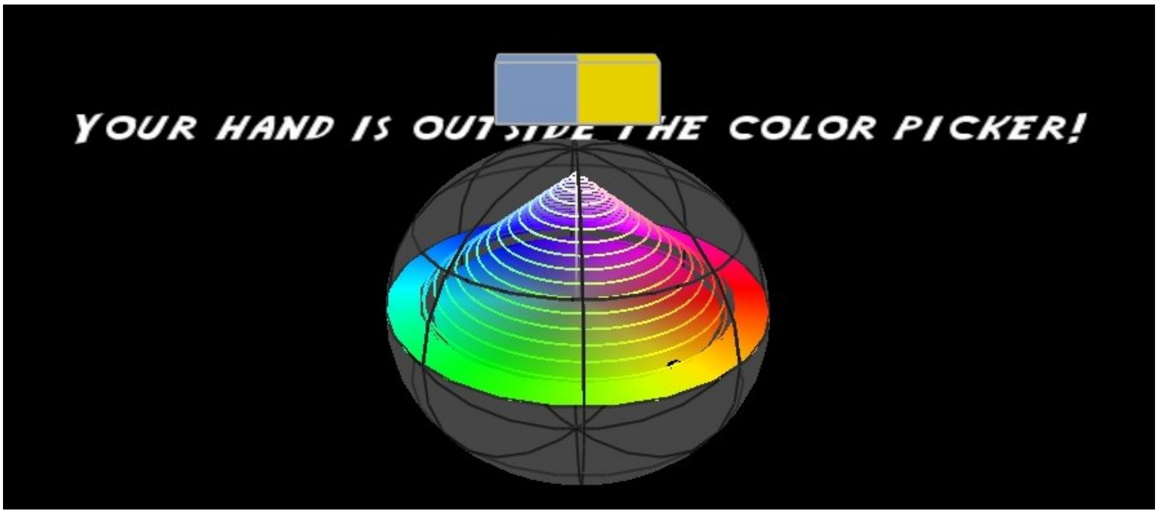


Figure 3.10: Output when the user's hand was outside of the threshold with respect to the interaction frame of reference.

was displayed initially, however if the subject's hand strayed outside of the radius of the color-picking widget, the status message would read: "Your hand is outside the color picker!" (see Figure 3.10)

During the practice trials, the experimenter guided the subject through the matching task, advising which of the hue, lightness, and saturation parameters needed adjustment to achieve a better match and describing the required physical adjustments to the tracker position. As the subject was informed in the instructions, the subject's goal was do his or her best to match each color as quickly as possible, but that no time limit was to be imposed. When the subject clicked the button again to signify that he or she was satisfied with the match, there was a one second pause of blackness. Following the pause, another trial would begin as before. The status message would read "Get ready for the next try..." Then the centering task would proceed. The preceding description of a single trial during the practice serves equally well to describe the trial protocol of the recorded experiment; the practice trials were identical to the actual experiment's (except for the guidance provided by the experimenter during the practice).

The practice session lasted for six trials. After these were completed, the subject was allowed a short break, if he or she chose. Before the experiment trials began, the experimenter turned off the room lights near the Cave and closed a black curtain surrounding the Cave area. This was done to minimize the disturbance caused by ambient light on the perception of the colors provided by the Cave's projectors. The experimenter observed the beginning of the experiment to ensure a smooth start, but left the subject to complete the experiment alone once the first trial began. The experimenter remained in the room on the other side of the black curtain.

Each subject completed 45 trials, 15 per condition. After completion of these 45 trials, the subject was asked to exit the Cave and answer a post-questionnaire (see Appendix A).

Following completion of the post-questionnaire, the subject was asked to complete a short second phase of the experiment. During this part of the experiment, the fundamental task remained color-matching, but the subject was allowed to manipulate the experimental condition of the first phase: the offset distance between the interaction and the display frames-of-reference. The purpose of this second phase was twofold. First, it allowed collection of performance measurements at distances other than the three offset conditions chosen for the first part. The second purpose was to collect data that might suggest the preferred offset.

Just like the first phase of the experiment, the widget placement phase consisted of a short, guided practice session followed by a longer unassisted session of trials. The calibration step was performed again at the beginning of each session. Overall, each trial proceeded as follows: centering, specifying the widget offset, centering, and finally color matching. The three conditions of offset used in the first phase were used in the second phase as initial offsets. The first step was a centering task identical to that completed prior to each trial in the experiment's initial phase. Once the subject centered his or her hand, the color picking widget appeared at the position determined by the current condition. The subject was then asked to use the joystick device on the wireless mouse to adjust the offset to the ideal position. The message on the front wall read: "Now, use the joystick to position the color picker as you like." Pushing up on the joystick caused the offset distance to increase, resulting in the widget moving away from the subject. Pushing down on the joystick caused the color picking widget to move closer to the subject, decreasing the offset distance. Left-right activation of the joystick had no effect on the widget offset; only up-down movements led to corresponding adjustments of the display-interaction offset.

After the subject was satisfied with the widget offset, he or she was instructed to push the button on the wireless mouse. Subsequently, the two discs used for centering—one in the target location, the other isomorphically mapped to the position of the subject's hand—appeared, the target location being that just chosen by the subject. After the subject centered his or her hand in the fixed interaction frame-of-reference, a color matching trial followed, with the color-picking widget displayed at the subject's chosen location.

Each subject completed three practice trials for this second phase. As before, the experimenter guided the subject in the Cave during the practice. Following completion of the practice, the experimenter exited the Cave, and the subject completed 15 unguided trials of color-matching, specifying the precise display-interaction offset prior to each trial. Trials were grouped in blocks of five. During each block of five the initial offset, before repositioning, was set as one of the three conditions of the first phase: collocated, three inch offset, or two feet offset. The ordering of the conditions was chosen to be the same as each subject had already seen during the first part of the experiment. Following the completion of the fifteen unassisted trials, the subject was asked to exit the Cave. After being paid the $10 per hour compensation, the subject's participation concluded and he or she left the Cave building.

Chapter 4

Results

4.1 Overview

Quantitative data were collected for all subjects during each trial. The research hypothesis is that user performance was maximized in the long offset condition. Given this hypothesis, the primary data of interest in the measured results are the chosen color and the time to completion. A metric for user performance is derived from these two parameters. R source code for all descriptive summaries, histograms, and ANOVA analyses is contained in Appendix B.

4.2 Hypothesis

The research hypothesis presented above leads to the following null hypothesis:

H_0: User performance for the color-matching task was not affected by the offset between the display and interaction frames-of-reference.

$$\mu_c = \mu_s = \mu_l$$

where μ_c represents the mean performance of the population from which the subjects were sampled for the collocated condition, μ_s for the mean performance under the short offset condition, and μ_l for the long offset condition.

In the subsequent analysis, rejection of this null hypothesis implies acceptance of the following alternative hypothesis:

H_1: User performance for the color-matching task was correlated with the offset between the display and interaction frames-of-reference.

The following discussion seeks to determine whether the aforementioned null hypothesis should be accepted or rejected.

4.3 User Performance

Two scalar values were collected per color matching trail: time to completion and distance between the chosen and target colors. Two more statistics were derived from these measurements: accuracy and accuracy per time. Descriptives for these measurements and derived statistics follow.

Statistic	Minimum	Maximum	Mean	Standard Deviation
Time	0.534105 s	280.709 s	25.71222 s	24.54542
Distance	0.0115825	1.56416	0.1781673	0.1525316
Accuracy	0	1	0.8927043	0.09824412
Accuracy Per Time	0	0.8400204	0.06198395	0.05422576

Discussion of each statistic along with descriptive breakdowns for each of the experimental conditions follows.

4.3.1 Time

Time was measured as the duration between completion of the centering task and selection of the color. Mean times for each condition type along with the standard deviations are shown below.

Condition	Time	Std. Dev.
Collocated	28.91119 s	30.81845
3 Inch Offset	24.82128 s	22.55296
2 Feet Offset	23.40417 s	18.31405

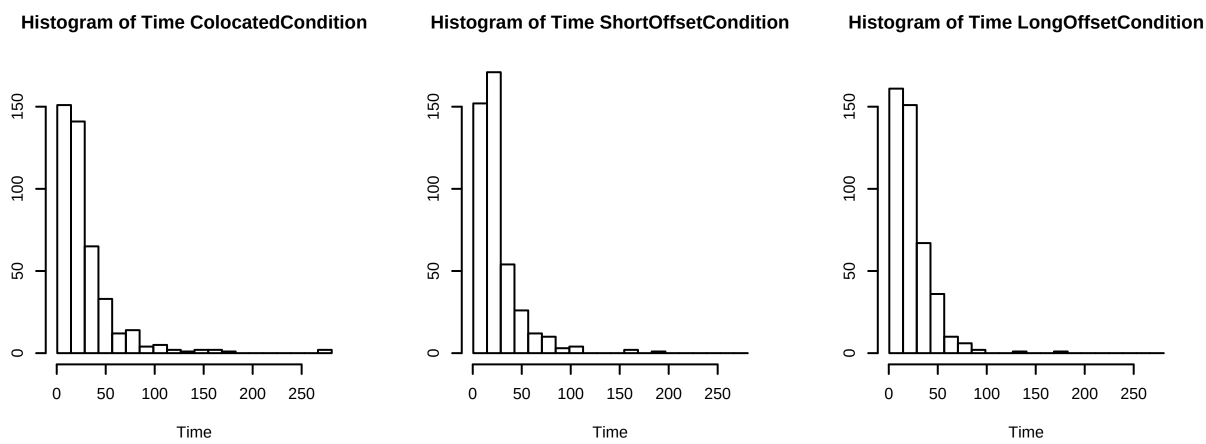


Figure 4.1: Per condition histograms for the time to completion measurement.

From these results, the immediate impression is that both the subjects' time-to-completion and the variability of that measured time decreased as the offset increased.

4.3.2 Distance

The calculation of a single color distance scalar given two colors was employed in order to simplify the assessment of the degree to which the subject successfully completed the task: color matching. As in the pilot studies, a perceptual distance metric was derived given the target and chosen colors. The formula for this metric is:

$$\sqrt{(2+\overline{R})\Delta R^2 + 4\Delta G^2 + (2-(1-\overline{R}))\Delta B^2}$$

The resulting scalar has no unit; the domain of the function is 0 to $2\sqrt{2} \approx 2.82$. Descriptive results of the derived distance scalar follow.

Condition	Color distance	Std. Dev.
Co-located	0.19471	0.1738083
3 Inch Offset	0.1703835	0.1441667
2 Feet Offset	0.1694059	0.1360128

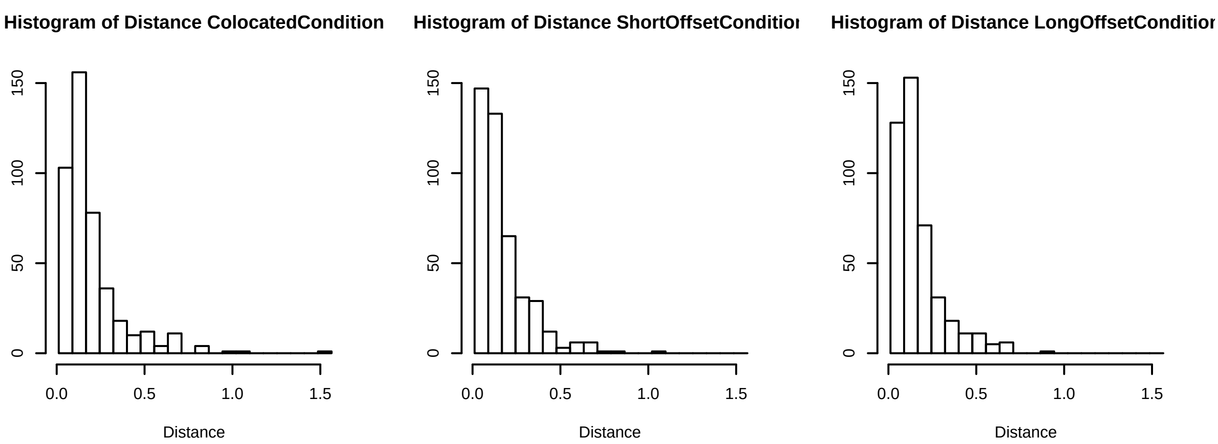


Figure 4.2: Per condition histograms for the perceptual distance statistic.

The mean distance between the chosen color and the target color per condition decreased as the offset between the display and interaction frames-of-reference increased. The variance of the distance metric within each condition also decreased as the offset distance increased.

4.3.3 Accuracy

Given the distance measurement, a normalized and opposite scalar was derived that increased as the subject's performance at the task improved. The calculation of accuracy is:

$$acc_i = \frac{1.0 - (d_i - d_{min})}{d_{range}}$$

Descriptive results of this statistic follow.

Condition	Accuracy	Std. Dev.
Co-located	0.8820477	0.1119482
3 Inch Offset	0.8977178	0.09285637
2 Feet Offset	0.8983475	0.08760454

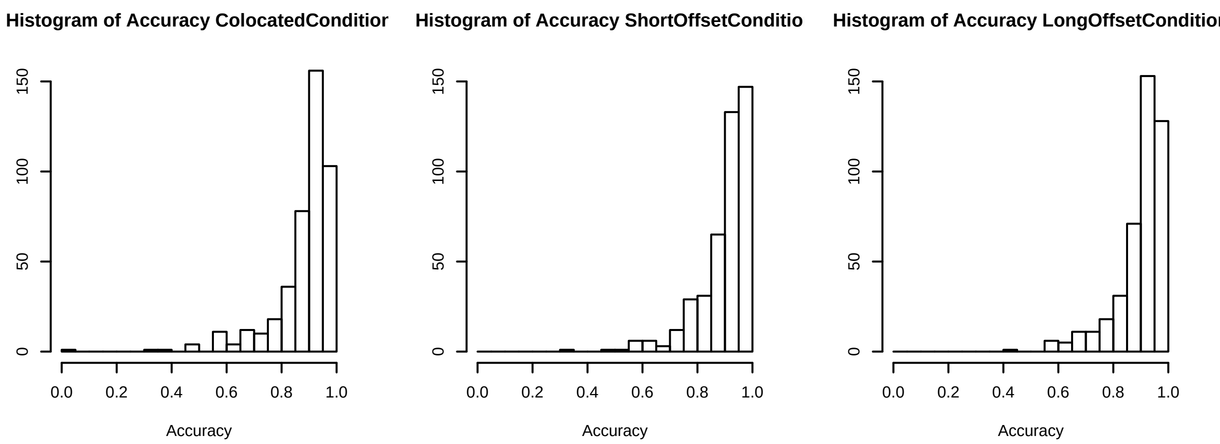


Figure 4.3: Per condition histograms for the derived accuracy statistic.

4.3.4 Accuracy per time

The accuracy scalar was combined with the measured time to derive a single scalar designed to capture the impact of both measured results on overall subject performance. The calculation is:

$$\frac{acc_i}{t_i}$$

Condition	Accuracy per time	Std. Dev.
Co-located	0.05938373	0.05248452
3 Inch Offset	0.06140336	0.04735732
2 Feet Offset	0.06516475	0.0618082

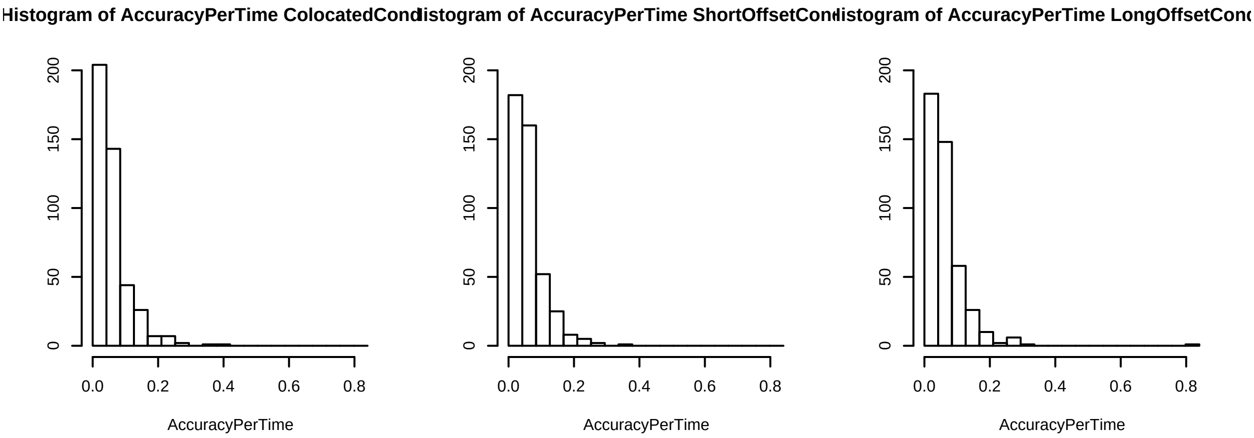


Figure 4.4: Per condition histograms for the derived accuracy per time statistic.

4.4 Experimental Design

The experiment was conducted using a within-subjects design: each subject saw each condition. Given three offset conditions, six ordering types were possible. The preceding descriptive statistics included results for all 29 subjects, and the ordering types were not equally represented. Moreover, it is assumed that the ordering type has an impact on the measured user performance for each condition. A general hypothesis is that performance improves with practice, and thus, for each subject, each successive condition encountered would exhibit better performance. The means over the course of each trial are plotted below along with linear regression lines. Each statistic shows improvement as the subject gains experience across trials.

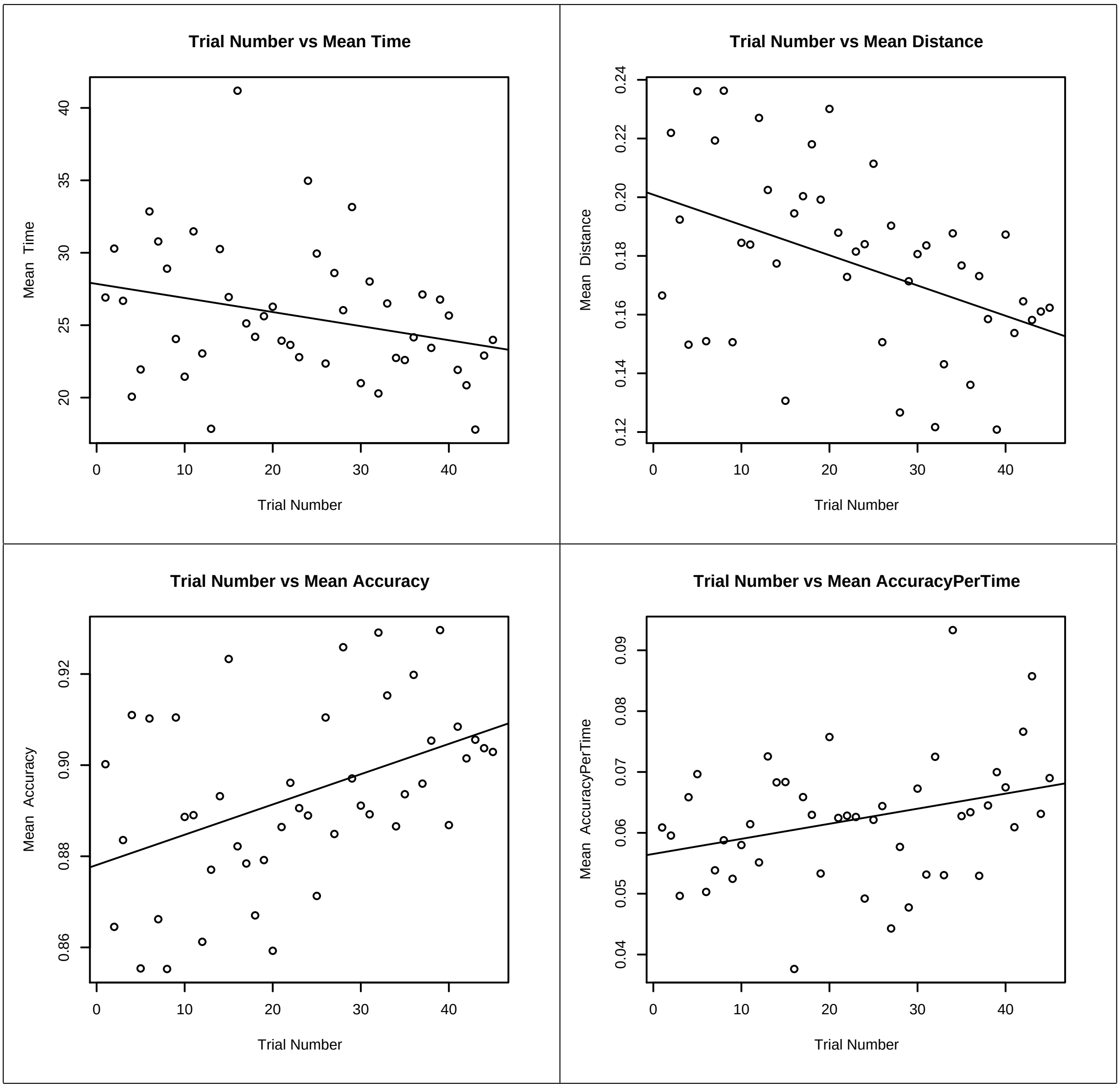


Figure 4.5: Means across subjects are plotted for each trial number. Each statistic demonstrates improved subject performance over the course of the experimental session.

An approach for minimizing the impact of this practice effect during analysis of the measured results is to adopt a balanced design: each ordering type is used an equal number of times. The number of subjects must be a multiple of the number of ordering types. For the experiment being described, this is achieved by discarding data for five subjects and using only twenty-four subjects in the analysis, each ordering type being represented four times. Subject data to discard is chosen not only to balance the design with respect to ordering types, but also with respect to gender. The procedure used to choose the set of twenty-four subjects is described. To achieve balance in ordering types and gender, only certain subject's data may be discarded. Histograms of the ordering types used during the course of the experiment are presented below for each gender. Ordering types are labeled with three letter codes: C stands for the collocated condition, S for the short offset, and L for the long offset condition.

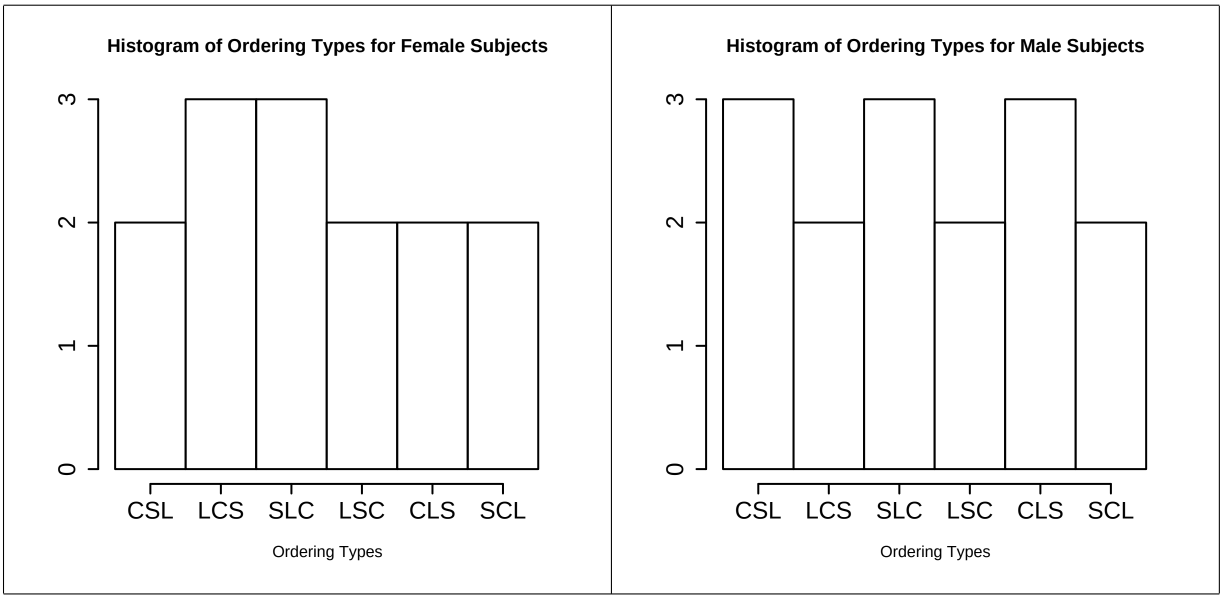


Figure 4.6: Histograms of the ordering types used for female and male subjects.

Balancing the design with respect to ordering type and gender requires that there be exactly two female and two male subjects per ordering type. Thus, one subject's data must be discarded for those ordering type-gender pairs that have three subject measurements. In these cases the subject with the highest overall variance in the accuracy per time statistic was chosen to be discarded. The summary of the discarding process follows.

After discarding data for the subjects noted above, a set of twenty-four subjects balanced with respect to ordering type and gender is achieved. Additionally, a more arbitrary algorithm for discarding subjects is employed to create a different pool of twenty-four subjects: the subject with the lowest subject ID numbers in each

Ordering Type-Gender Pairing	Subject	Standard deviation of Accuracy Per Time	Subject to Discard
LCS-Female	20	0.05138721	20
	24	0.04039673	
	29	0.05066732	
SLC-Female	10	0.13277012	10
	23	0.03906819	
	27	0.04771893	
CSL-Male	1	0.04608350	5
	4	0.04157196	
	5	0.06449871	
SLC-Male	11	0.04783256	28
	12	0.04484750	
	28	0.07896211	
CLS-Male	18	0.06166525	18
	19	0.05868793	
	26	0.03161253	

Table 4.1: The subject with the lowest overall standard deviation of the accuracy per time statistic is chosen to be discarded.

eligible gender-ordering type pair is removed. This favors discarding those data acquired earliest in the study. Descriptive statistics for both balanced subject pools are presented.

Results show lowest overall accuracy-per-time standard deviation discarding approach, lowest subject ID discarding approach.

Statistic	Condition	Mean	Standard Deviation
Distance	Colocated	0.186827, 0.199924	0.169902, 0.181035
	Short Offset	0.169213, 0.175868	0.147747, 0.152034
	Long Offset	0.166166, 0.171215	0.131519, 0.135404
Time	Colocated	29.58258 s, 28.32647 s	30.74937, 30.45468
	Short Offset	25.64511 s, 25.00983 s	23.65118, 23.62795
	Long Offset	24.38133 s, 23.72446 s	18.78458, 18.80541
Accuracy	Colocated	0.887126, 0.878690	0.109432, 0.116603
	Short Offset	0.898471, 0.894185	0.095162, 0.097923
	Long Offset	0.900433, 0.897182	0.084710, 0.087212
Accuracy Per Time	Colocated	0.054473, 0.058766	0.044331, 0.050128
	Short Offset	0.058937, 0.061618	0.043884, 0.047794
	Long Offset	0.059651, 0.062047	0.045216, 0.047480

Table 4.2: Descriptive statistics for the two sets of subjects balanced with respect to ordering type.

4.5 Comparison of means

The null hypothesis under investigation states that the means of each condition are equal. Paired sample t-tests are performed between each offset condition pair for both balanced subject pools. Results are summarized:

Conditions	Statistic	p-value	
Short vs. Long offset	Time	0.4010	0.3900
	Distance	0.7591	0.6486
	Accuracy Per Time	0.8313	0.904
Collocated vs. Long offset	Time	0.005048	0.01191
	Distance	0.05876	0.01223
	Accuracy Per Time	0.1124	0.3470
Collocated vs. Short Offset	Time	0.03722	0.0744
	Distance	0.08922	0.02795
	Accuracy Per Time	0.1611	0.397

Table 4.3: Paired sample t-test results between each condition.

Lower p-values imply the increasing likelihood that a difference exists between the means of the given statistics that is not due to chance alone. The differences between the collocated and the long offset condition are most pronounced, particularly in the time and distance measurements. This result suggests that the null hypothesis may be rejected: there is a difference in means between offset conditions. Using the balanced set of 24 subjects, analysis of variance is used to determine if any of the above differences in means can be attributed to the experimental condition or if the differences must be attributed to random error (in this case: subject).

In the experiment each subject tried to match each color exactly once per condition. This design allows the color to be included as a factor in the analysis. A two-way within-subjects ANOVA is performed for the time, distance, and accuracy per time statistics. The factors are the condition and the target color, the former having three levels and the latter fifteen levels. Subject is the known random error. Breakdowns of the analyses follow.

4.5.1 Time

Two-way ANOVA is executed for the time statistic. ANOVA summaries are given for both balanced sets of twenty-four subjects.

Largest overall standard deviation of accuracy per time discarded group:

Lowest subject ID discarded group:

```
Error: SubjectID
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 23  88513    3848

Error: SubjectID:Condition
          Df Sum Sq Mean Sq F value  Pr(>F)
Condition  2   5298    2649   3.846 0.02854 *
Residuals 46  31686     689
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: SubjectID:ColorName
           Df Sum Sq Mean Sq F value    Pr(>F)
ColorName  14  31907    2279  3.5784 1.641e-05 ***
Residuals 322 205083     637
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: SubjectID:Condition:ColorName
                     Df Sum Sq Mean Sq F value Pr(>F)
Condition:ColorName  28  12136     433  0.9379  0.559
Residuals           644 297612     462
```

Figure 4.7: ANOVA summary for the time measurement in the largest overall accuracy-per-time standard deviation discard group.

```
Error: SubjectID
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 23  99704    4335

Error: SubjectID:Condition
          Df  Sum Sq Mean Sq F value  Pr(>F)
Condition  2  4059.7  2029.9   2.973 0.06106 .
Residuals 46 31407.1   682.8
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: SubjectID:ColorName
           Df Sum Sq Mean Sq F value    Pr(>F)
ColorName  14  31237    2231  3.6621 1.108e-05 ***
Residuals 322 196187     609
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: SubjectID:Condition:ColorName
                     Df Sum Sq Mean Sq F value Pr(>F)
Condition:ColorName  28  14063     502   1.124 0.3020
Residuals           644 287751     447
```

Figure 4.8: ANOVA summary for the time measurement in the first subject discard group.

The null hypothesis that the variability among the time means between conditions and within each subject is due to chance can be rejected if a $p < .05$ is acceptable. The chosen value of p represents the risk one is willing to accept of making a Type I error: rejecting the null hypothesis when it is in fact true. The p-value is the probability that the observed differences in means was in fact due to chance and not to the experimental condition—the lower the p-value, the less likely is such an incorrect identification of a real difference in means. Conversely, a lower p-value increases the risk of a Type II error: accepting the null hypothesis when it is not true. In this case there is a real difference in means, but the difference is not identified by the researcher. The relative importance of avoiding each error type depends on the context of the experiment and the conclusions to be drawn. The effect on time due to the target color is more pronounced. A belief that this would be the case was garnered from feedback after the first pilot and implicit in the decision to use a fixed set of colors per condition per subject in the final design. The extremely low p-value strongly suggests that this belief is indeed true.

4.5.2 Distance

Two-way ANOVA is executed for the perceptual distance metric statistic. Largest overall standard deviation of accuracy per time discarded group:

Lowest subject ID discarded group:

```
Error: SubjectID
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 23 4.9890  0.2169

Error: SubjectID:Condition
          Df  Sum Sq Mean Sq F value Pr(>F)
Condition  2 0.08957 0.04478  1.7149 0.1913
Residuals 46 1.20128 0.02611

Error: SubjectID:ColorName
           Df Sum Sq Mean Sq F value    Pr(>F)
ColorName  14 3.7398  0.2671  13.676 < 2.2e-16 ***
Residuals 322 6.2895  0.0195
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: SubjectID:Condition:ColorName
                     Df Sum Sq Mean Sq F value  Pr(>F)
Condition:ColorName  28 0.5625  0.0201  1.6962 0.01459 *
Residuals           644 7.6275  0.0118
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4.9: ANOVA summary for the distance statistic in the largest overall accuracy-per-time standard deviation discard group.

```
Error: SubjectID
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals 23 5.2429  0.2280

Error: SubjectID:Condition
          Df  Sum Sq Mean Sq F value  Pr(>F)
Condition  2 0.17095 0.08548   3.176 0.05105 .
Residuals 46 1.23799 0.02691
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: SubjectID:ColorName
           Df Sum Sq Mean Sq F value    Pr(>F)
ColorName  14 4.5346  0.3239  15.415 < 2.2e-16 ***
Residuals 322 6.7659  0.0210
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: SubjectID:Condition:ColorName
                     Df Sum Sq Mean Sq F value  Pr(>F)
Condition:ColorName  28 0.5538  0.0198  1.5327 0.03994 *
Residuals           644 8.3108  0.0129
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4.10: ANOVA summary for the time measurement in the first subject discard group.

In the first subject pool ,the results for distance do not afford a high enough F-value (the ratio of the variability between to that within conditions) to reject the null hypothesis. The obtained F-value ($F = 3$) for the second subject pool is more suggestive of a significant effect, but still does not allow rejection of the null hypothesis given $p < .05$. Even more extremely pronounced than with time, the target color has a huge effect on the distance statistic.

4.5.3 Accuracy per Time

Finally, accuracy per time, the derived statistic that is designed to measure a subject's effiency at a task, is analyzed with two-way ANOVA.

Largest overall standard deviation of accuracy per time discarded group:

```
Error: SubjectID
          Df   Sum Sq  Mean Sq F value Pr(>F)
Residuals 23 0.154035 0.006697

Error: SubjectID:Condition
          Df   Sum Sq  Mean Sq F value Pr(>F)
Condition  2 0.005672 0.002836  1.1744 0.3181
Residuals 46 0.111073 0.002415

Error: SubjectID:ColorName
           Df  Sum Sq Mean Sq F value    Pr(>F)
ColorName  14 0.20893 0.01492  7.7027 4.885e-14 ***
Residuals 322 0.62384 0.00194
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: SubjectID:Condition:ColorName
                     Df  Sum Sq Mean Sq F value Pr(>F)
Condition:ColorName  28 0.04211 0.00150  0.9775 0.4997
Residuals           644 0.99091 0.00154
```

Figure 4.11: ANOVA summary for the accuracy-per-time statistic in the largest overall accuracy-per-time standard deviation discard group.

Lowest subject ID discarded group:

```
Error: SubjectID
          Df   Sum Sq  Mean Sq F value Pr(>F)
Residuals 23 0.287994 0.012521

Error: SubjectID:Condition
          Df   Sum Sq  Mean Sq F value Pr(>F)
Condition  2 0.002291 0.001145  0.3973 0.6744
Residuals 46 0.132622 0.002883

Error: SubjectID:ColorName
           Df  Sum Sq Mean Sq F value    Pr(>F)
ColorName  14 0.23616 0.01687  7.9031 1.917e-14 ***
Residuals 322 0.68730 0.00213
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: SubjectID:Condition:ColorName
                     Df  Sum Sq Mean Sq F value Pr(>F)
Condition:ColorName  28 0.04774 0.00171  0.9635 0.5205
Residuals           644 1.13969 0.00177
```

Figure 4.12: ANOVA summary for the accuracy per time statistic in the first subject discard group.

Similar to the distance alone, the combined metric does not allow rejection of the null hypothesis. Again, as with distance, the target color shows a very strong effect on the subject's performance.

4.6 Second experiment phase: Widget placement

The second part of the experiment allowed the subject to specify the offset between the interaction frame-of-reference and the display frame-of-reference prior to each color matching attempt. The purpose of this phase of the experiment was to determine the subjects' preferences for the offset. Descriptive statistics for the offset choice follow. Though the three offset conditions from the first phase were used as initial offsets for each trial (i.e. the offset prior to any subject adjustment), all 29 subjects are represented as there was no *a priori* hypothesis regarding an effect this initial offset would have on the subject's chosen offset.

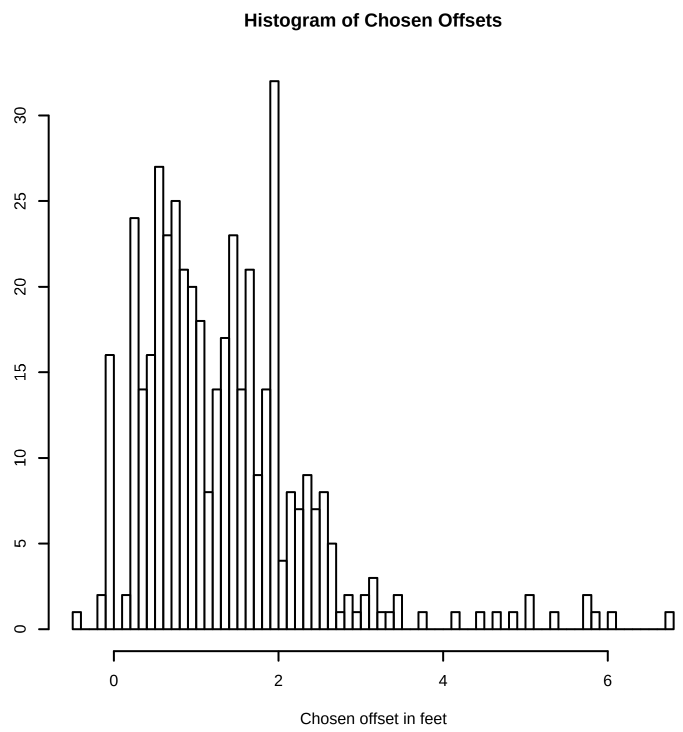


Figure 4.13: Chosen offsets in the widget placement task.

Mean	1.348944 feet
Standard deviation	1.026073

It is apparent from the above histogram that many trials proceeded with the subject making no adjustment to the initial offset. Many reasons are possible for this outcome. One possibility that should be excluded from the analysis is the case where the subject accidentally activated the color matching part of the trial without choosing the desired offset. Using a subset of the trials where the subject displaced the offset by at least .001 inches, a set of trials that likely excludes this possibility is used for further analysis.

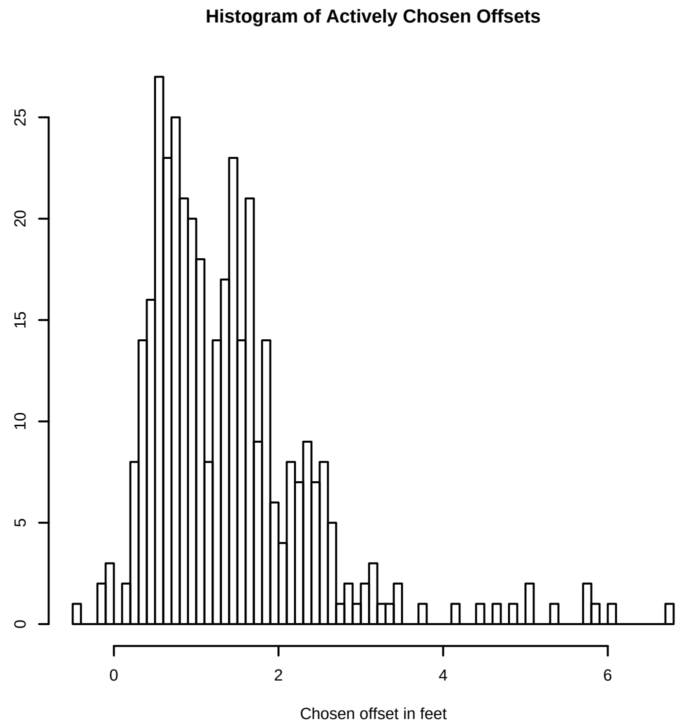


Figure 4.14: Chosen offsets in the widget placement task for only those trials where the subject changed the offset at least .001 feet.

Mean	1.396817 feet
Standard deviation	1.029969

4.7 Discussion

The experiment was performed to test the differences in user performance at a color matching task, specifically to determine if the null hypothesis—that no difference in user performance exists between the conditions—could be rejected. Two performance metrics were measured: time to completion and perceptual distance between the target and the chosen color. A third metric, accuracy per time, was derived as a combination of the two fundamental scalar measurements. Given a $p < .05$, a significant main effect was found for the offset factor in the time to completion metric, but not in the distance or the combined metric. For the distance metric, the condition factor was nearly able to lead to a rejection of the null hypothesis if a $p < .05$ could be deemed acceptable. The factor of target color showed a highly significant effect in each metric, with a very high certainty on the distance metric ($F = 13$). Though less pronounced, the effect of color on the time measurement was also highly significant ($F = 3$). The inherent difference between colors in matching difficultly was a result predicted by subject feedback in the first pilot study. The second phase of the experiment was performed to determine the optimal offset distance between the display and interaction frames of reference from the subject's point of view. The mean chosen offset distance for the actively chosen trials was between the short and long offset conditions.

Chapter 5

Conclusion

A hypothesis was proposed stating that user performance in a color matching task would be better given an offset between the display and interaction frames-of-reference than when there was no offset, and the results from an experiment suggest that this hypothesis can be accepted. The terms interaction frame-of-reference and display frame-of-reference were introduced to draw attention to a key distinction common among all virtual environment interfaces: the location of the user's physical actions versus the location of the rendered feedback. In the spirit of a body-centered approach to virtual environment user interface design, previous research asserted that the collocation of these frames-of-reference is optimal insofar as this affords the most direct form of manipulating the virtual world through intuitive physical action. These conclusions were based on studies using a head-mounted display for the virtual environment hardware platform. The current work serves as a contrast to exemplify an alternative conclusion that may be drawn given a hardware platform that allows the user to see his body and the virtual imagery simultaneously (e.g. Cave-like environments).

Given a Cave-based virtual environment, it has been demonstrated that no offset between the interaction and display frames-of-reference (i.e. collocation) leads to decreased user performance. A recommendation stemming from this study for designers of applications to be deployed in Cave-based VE platforms is that a non-zero offset between the display and interaction frames-of-reference should be considered for all interaction techniques. The current study cannot recommend a specific optimal offset for the color picking widget. Two offset conditions were presented to the subjects, but the data suggest that the preferred offset was between the two provided conditions. A more refined inquiry could be undertaken to determine the optimal offset distance and to examine if any tradeoffs exist between subject performance and preference. Further research could examine more offset conditions at a more regular spacing to refine the estimate of an optimal offset for the color picking widget.

Just as no optimal offset between the display and interaction frames-of-reference for the color picking widget emerge from these results without further experimentation, any conclusion as to the general optimal offset for a larger class of virtual environment widgets cannot as yet be drawn. Nevertheless, given that this study is situated in the most specific category of VE user studies—those that compare the parameter choices within a given interaction technique—it can serve as a model for the specific individual studies that could constitute the basis for the more general question: what is the optimal display-interaction offset given any interaction technique in a Cave-based VE?

For instance, an inquiry could be pursued to compare results of similar user studies. In such an inquiry, the interaction technique would be a variable and the user performance among a group of interaction techniques would be explored. Possible interaction techniques could be chosen based on a taxonomic structure (e.g. *Bowman* (1999)) in order to facilitate generalization across interaction technique types. For example, two branches of Bowman's taxonomy could be compared: the results of a user study in a system control task (e.g. color matching using the Cavepainting color-picking widget as in the current study) and of a user study for a manipulation task (e.g. the rotation of an object to a target as in *Poupyrev et al.* (2000). If both individual studies employed a varying offset between the display and interaction frames-of-reference as a common experimental factor, the comparison of these studies could yield a more complete understanding of the diplay-interaction offset's effect on general user performance. Given a method of finding the optimal offset for each individual widget (e.g. the individual user studies), it may be the case that all interactions in Cave-based environments are optimized at nearly the same offset distance. Alternatively, such a research pursuit may not yield a general suggestion for offset distance, but rather may emphasize the specific considerations that seem to impact performance for the examined interactions techniques and example tasks.

Ultimately, this is not a vein that can be pursued in and of itself, but can only emerge as more exemplary interaction techniques that afford a choice in the display-interaction offset are developed. Those interaction techniques chosen for implementation by virtual environment application designers are motivated by the requirements of the application itself. The choices themselves are influenced by those techniques the designer has read of or experienced. The goal of the current work is to participate in this dialectic between new technique development on one hand and developed technique analysis on the other, so as to serve a role in pushing forward the overall progress of virtual environment research and application design.

References

Bowman, D., Interaction techniques for common tasks in immersive virtual environments: Design, evaluation and application, Ph.D. thesis, Georgia Institute of Technology, 1999.

Bowman, D., D. Johnson, and L. Hodges, Testbed Evaluation of Virtual Environment Interaction Techniques, *Presence*, *10*, 75–95, 2001.

Bowman, D., E. Kruijff, J. LaViola, and I. Poupyrev, *3D User Interfaces: Theory and Practice*, Addison-Wesley, 2004.

Conner, B., S. Snibbe, K. Herndon, D. Robbins, R. Zeleznik, and A. van Dam, Three-dimensional Widgets, in *Proceedings of Symposium on Interactive 3D Graphics '92*, pp. 183–188, 1992.

Cruz-Neira, C., D. Sandin, and T. DeFanti, Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE, in *Proceedings of SIGGRAPH '93*, 1993.

Keefe, D., D. Acevedo, T. Moscovich, D. H. Laidlaw, and J. LaViola, Cavepainting: A fully immersive 3D artistic medium and interactive experience, in *Proceedings of ACM Symposium on Interactive 3D Graphics 2001*, pp. 85–93, 2001.

Kurtenbach, G., and W. Buxton, The limits of expert performance using hierarchic marking menus, in *Proceedings of InterCHI '93*, pp. 482–487, 1993.

Lemmerman, D., and A. Forsberg, Adapting event-based applications for synchronization in VR clusters, in *Proceedings of the 8th International Immersive Projection Technology Workshop*, Iowa State University, Ames, IA, 2004.

Mine, M., F. Brooks, and C. Sequin, Moving objects in space: Exploiting proprioception in virtual-environment interaction, in *Proceedings of SIGGRAPH '97*, pp. 482–487, 1997.

Pausch, R., D. Proffitt, and G. Williams, Quantifying immersion in virtual reality, in *Proceedings of SIGGRAPH '97*, pp. 13–18, 1997.

Pierce, J., Expanding the interaction lexicon for 3d graphics, Ph.D. thesis, Carnegie Mellon, 2001.

Poupyrev, I., M. Billinghurst, S. Weghurst, and T. Ichikawa, The go-go interaction technique: Non-linear mapping for direct manipulation in vr, in *Proceedings of UIST '96*, pp. 79–80, 1996.

Poupyrev, I., S. Weghurst, M. Billinghurst, and T. Ichikawa, A framework and testbed for studying manipulation techniques for lmmersive vr, in *Proceedings of VRST '97*, pp. 21–28, 1997.

Poupyrev, I., S. Weghurst, and S. Fels, Non-isomorphic 3d rotational techniques, in *Proceedings of CHI '00*, pp. 546–547, 2000.

Schulze, J., A. Forsberg, A. Kleppe, R. Zeleznik, and D. H. Laidlaw, Characterizing the effect of level of immersion on a 3D marking task, in *Proceedings of HCI International '05*, 2005.

Slater, M., and M. Usoh, *Artificial Life and Virtual Reality*, chap. Body Centred Interaction in Immersive Virtual Environments, pp. 125–148, John Wiley and Sons, 1994.

Slater, M., V. Linakis, M. Usoh, and R. Kooper, Immersion, presence, and performance in virtual environments: An experiment with tri-dimensional chess, in *Proceedings of VRST '96*, pp. 163–172, 1996.

Sutherland, I., The ultimate display, in *Proceedings of IFIP '65*, vol. 2, pp. 582–583, 1965.

Sutherland, I., A head-mounted three-dimensional display, in *Proceeding of the Fall Joint Computer Conference*, vol. 33, pp. 757–764, 1968.

Swan, J., J. Gabbard, D. Hix, R. Schulman, and K. Kim, Comparative study of user performance in a map-based virtual environment, in *Proceedings of IEEE Virtual Reality '03*, pp. 259–266, 2003.

Appendix A

Pre- and post-questionnaires

The experiment's written instructions, the demographic pre-questionnaire, and post-questionnaire are presented, along with summaries of the collected data.

Pre Questionnaire

In this experiment you will be working in the Cave environment, wearing stereo glasses, and using an interaction wand to match colors. In each try you will see a 3D color picker and a pair of colored boxes above it. The color in the left box will show you the target color you are trying to select; the color in the right box indicates the current color you are picking with the 3D color picker. Additionally, a small cursor will appear inside the 3D color picker to show the location of the current color. A new target color will be randomly chosen for each try.

The 3D color picker is displayed as two back-to-back cones. Each point in the double-cone specifies a color as follows. The hue of the color (the color family) is changed around the circumference of the cone. The brightness of the color is least at the bottom of the picker and most at the top (i.e. the bottom tip is pure black; the top tip is pure white). The saturation (which can also be thought of as related to a color's 'grayness') is most at the outer circumference of the cone and least at the center (i.e. along the outer circumference is pure color; at the center is pure gray).

Before each matching try, you will be asked to center yourself with respect to the color picker. You will have to line up a flat disc attached to your controller with one located at the center of the color picker.

There is no time limit for each matching task. Do your best to match each color as quickly as you can.

First you will be given 6 practice tries. After that the actual experiment will begin as soon as you are ready. At the beginning of both the practice and and the actual experiment, you will be asked to reach forward and click the controller to calibrate the system for your arm length.

Everyone responds to VR in different ways. Though it's unlikely, if for any reason during the experiment you feel too disoriented or uncomfortable to continue, let me know and I'll end the experiment.

After you complete the experiment, you'll be given a set of questions to answer about your experience.

First, please tell me a little about yourself by answering the following questions. This experiment is anonymous, and your name will not be used anywhere. This information will only be used to help analyze the data.

1. Age:
2. Gender?

   ( ) Female ( ) Male

3. Do you have any color vision deficiency?

   ( ) Yes ( ) No

4. Do you wear corrective lenses?

   ( ) Yes ( ) No

5. Which is your dominant hand?

   ( ) Left ( ) Right

6. I have extensive video game experience.

   ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

7. I have extensive virtual reality experience.

   ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

8. I have extensive experience with computer graphics programs involving color selection (e.g. Adobe Photoshop).

   ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

Post Questionnaire

Thank you for completing the color matching tries. Please complete the following statements regarding your experience.

1. I was able to remember locations of color families (e.g. reds, blues, etc.) relative to the color picker's center.

   ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

2. I was physically fatigued after completing all of the coloring matching tries.

   ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

3. I felt that my accuracy in color matching improved over the course of the color matching tries.

   ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

4. I felt that my speed in color matching improved over the course of the color matching tries.

   ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

When the picker appeared closest to my body...

5. I enjoyed using the color picker.

   ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

6. I found matching to be difficult.

   ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

7. The distance I moved my hand was the same distance that the color picker cursor moved.

   ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

8. The distance between my hand and the color picker cursor remained constant when I moved my hand.

   ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

9. I was able to see the locations of colors easily.

   ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

10. I felt the color picker was within my reach.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

11. It was easy to line up the color discs before each matching try.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

12. I felt that color matching was fatiguing.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

13. The volume in which I had to move my hand was a comfortable space to reach.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

14. I was mainly looking at the color picker while trying to match colors.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

15. I saw my own hand while trying to match colors.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

16. Seeing my own hand was distracting for matching colors with the color picker.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

17. I felt that I could move my hand into the color picker.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

18. Using the color picker felt natural.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

When the picker appeared slightly away from my body...

19. I enjoyed using the color picker.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

20. I found matching to be difficult.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

21. The distance I moved my hand was the same distance that the color picker cursor moved.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

22. The distance between my hand and the color picker cursor remained constant when I moved my hand.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

23. I was able to see the locations of colors easily.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

24. I felt the color picker was within my reach.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

25. It was easy to line up the color discs before each matching try.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

26. I felt that color matching was fatiguing.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

27. The volume in which I had to move my hand was a comfortable space to reach.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

28. I was mainly looking at the color picker while trying to match colors.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

29. I saw my own hand while trying to match colors.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

30. Seeing my own hand was distracting for matching colors with the color picker.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

31. I felt that I could move my hand into the color picker.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

32. Using the color picker felt natural.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

When the picker appeared furthest away from my body...

33. I enjoyed using the color picker.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

34. I found matching to be difficult.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

35. The distance I moved my hand was the same distance that the color picker cursor moved.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

36. The distance between my hand and the color picker cursor remained constant when I moved my hand.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

37. I was able to see the locations of colors easily.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

38. I felt the color picker was within my reach.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

39. It was easy to line up the color discs before each matching try.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

40. I felt that color matching was fatiguing.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

41. The volume in which I had to move my hand was a comfortable space to reach.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

42. I was mainly looking at the color picker while trying to match colors.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

43. I saw my own hand while trying to match colors.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

44. Seeing my own hand was distracting for matching colors with the color picker.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

45. I felt that I could move my hand into the color picker.

    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

46. Using the color picker felt natural.

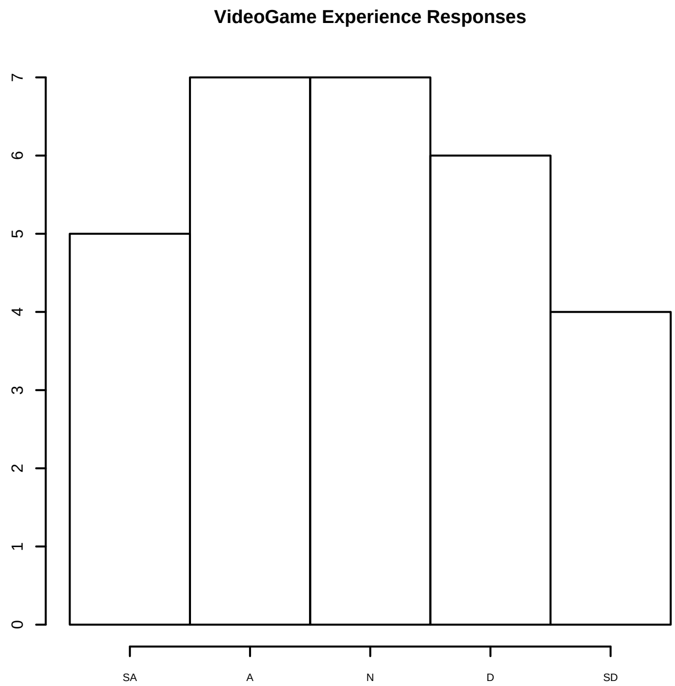
    ( ) Strongly agree ( ) Agree ( ) Neither agree nor disagree ( ) Disagree ( ) Strongly disagree

47. Describe your overall strategy for completing the color matches:

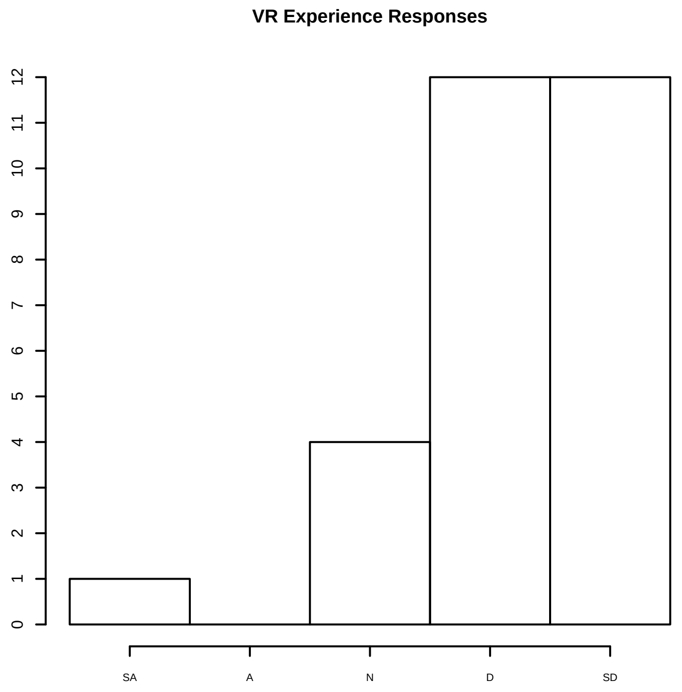
A.1 Pre-questionnaire responses

Demographic data gathered from each subject prior to the study is presented below. Responses to the Likert scale questions are displayed as histograms. The abbreviations SA, A, N, D, and SD are used for Strongly Agree, Agree, Neither Agree nor Disagree, Disagree, and Strongly Disagree—the answers in the questionnaire.

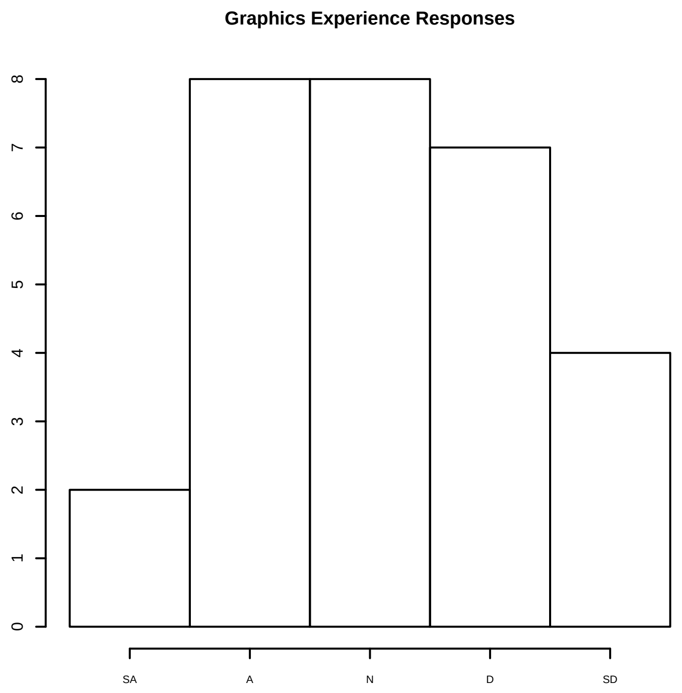
Mean Age	24.33
Gender	14 Female, 15 Male
Normal Color Vision	29 Yes, 0 No
Wears glasses	17 Yes, 12 No
Handed	27 Right, 2 Left



I have extensive video game experience.

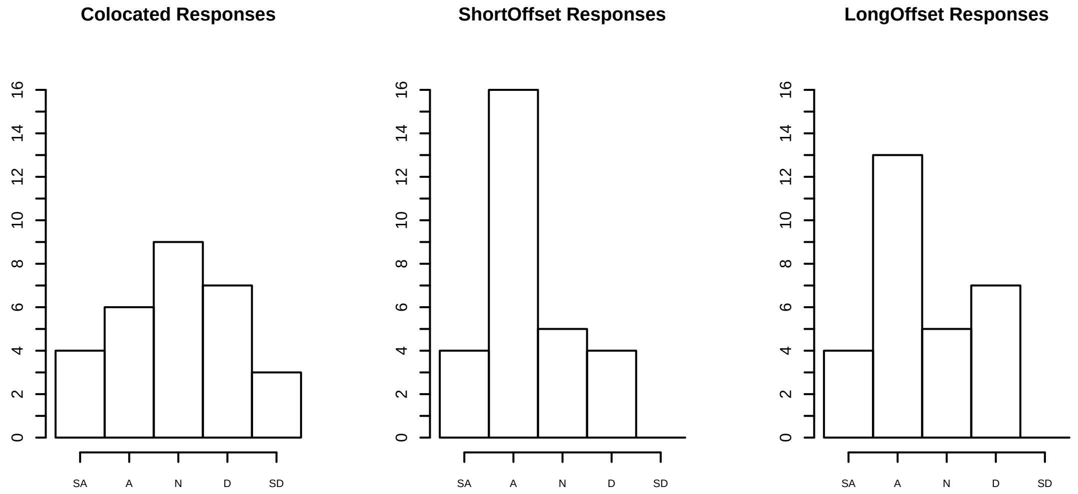


I have extensive virtual reality experience.

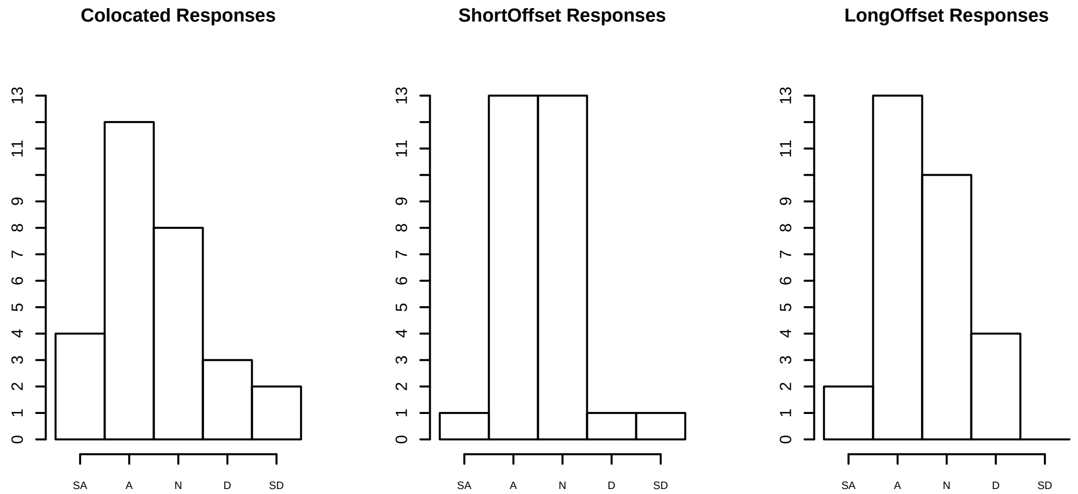


I have extensive experience with computer graphics programs involving color selection (e.g. Adobe Photoshop).

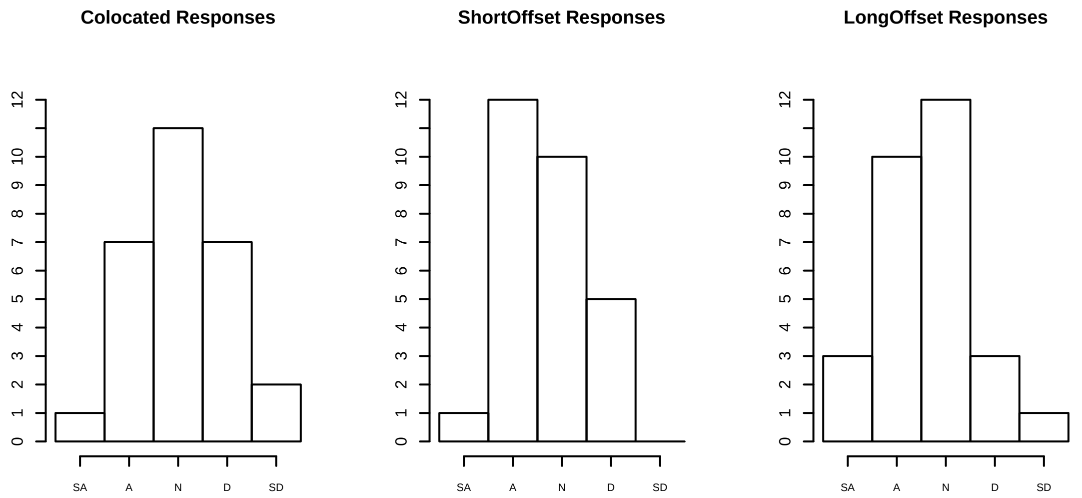
A.2 Post-questionnaires responses



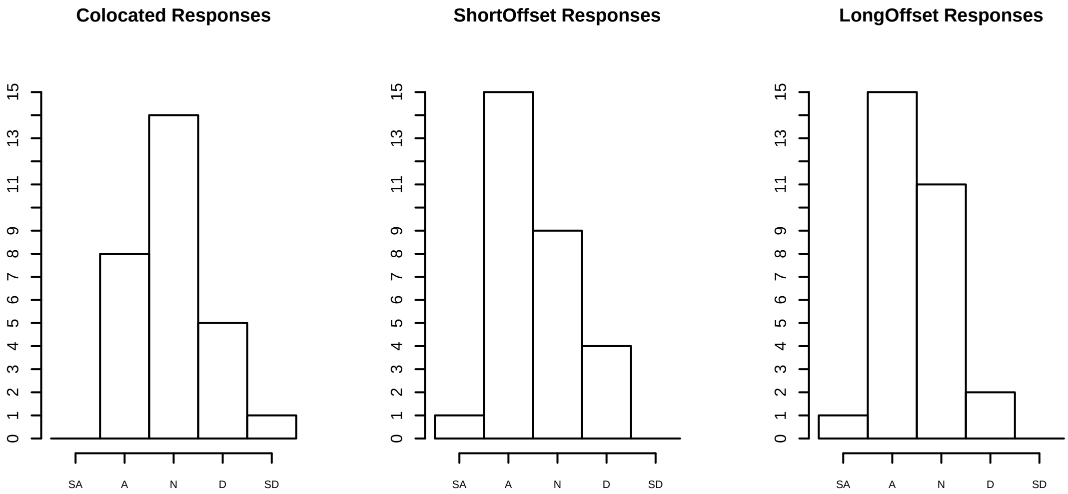
I enjoyed using the color picker.



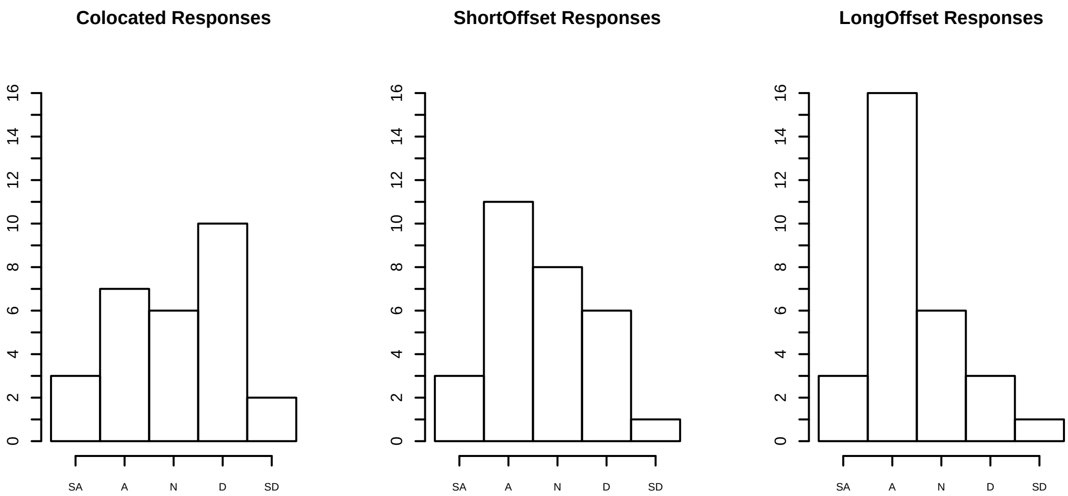
I found matching to be difficult.



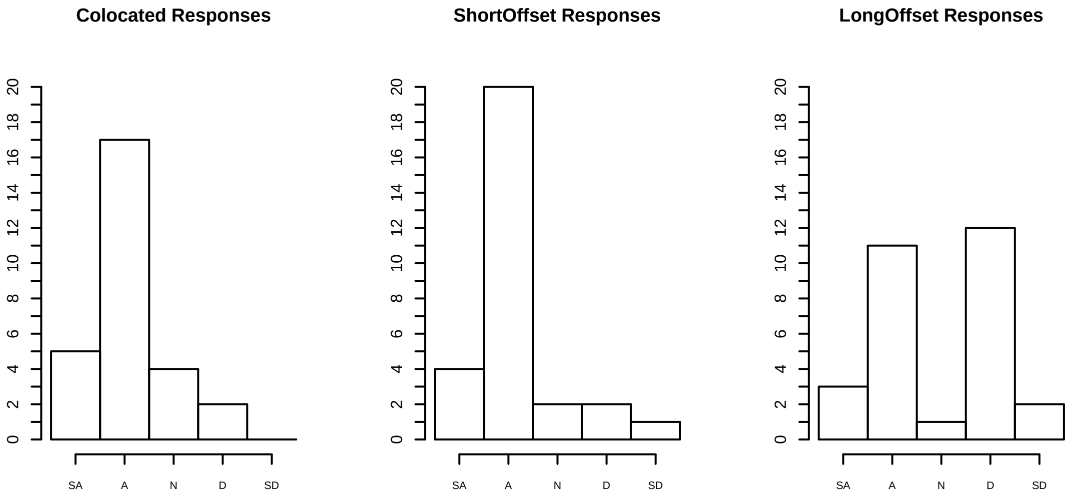
The distance I moved my hand was the same distance that the color picker cursor moved.



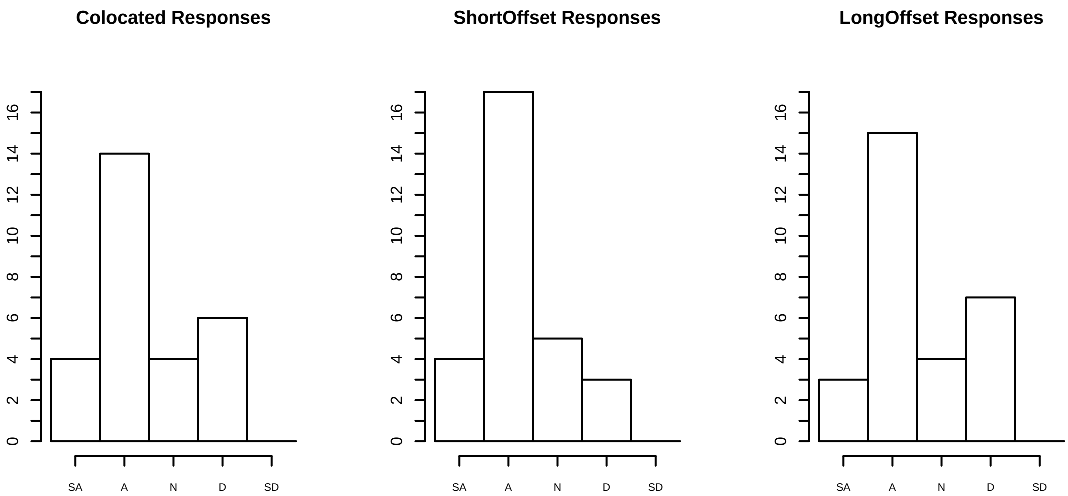
The distance between my hand and the color picker cursor remained constant when I moved my hand.



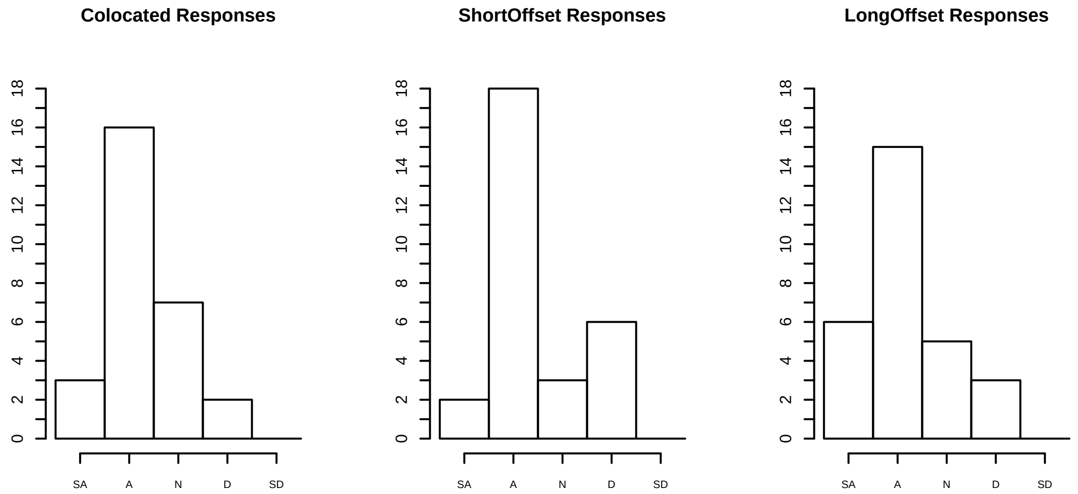
I was able to see the locations of colors easily.



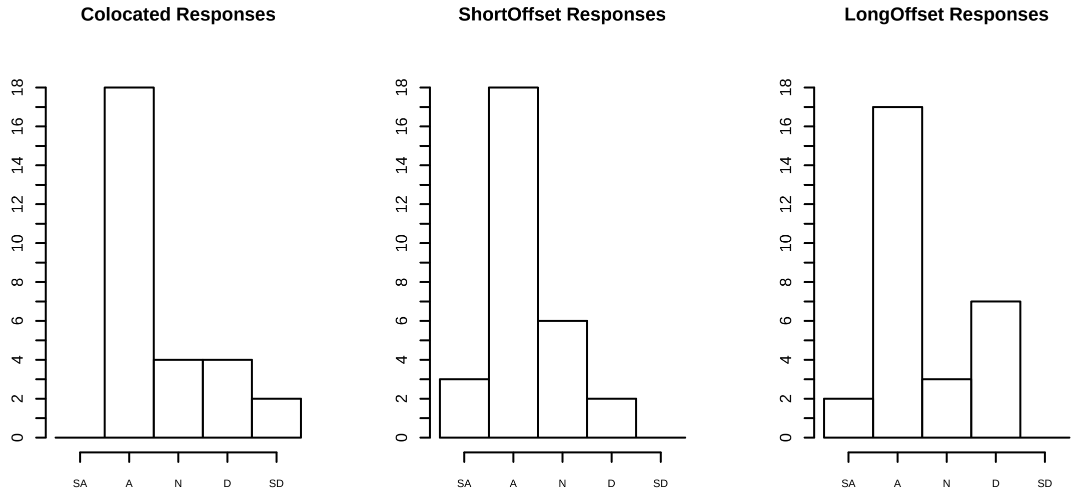
I felt the color picker was within my reach.



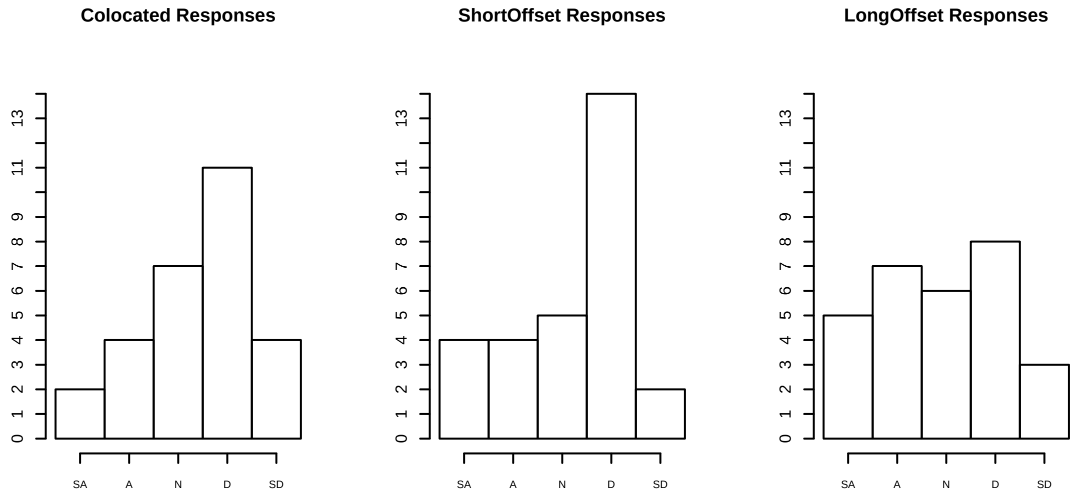
It was easy to line up the color discs before each matching try.



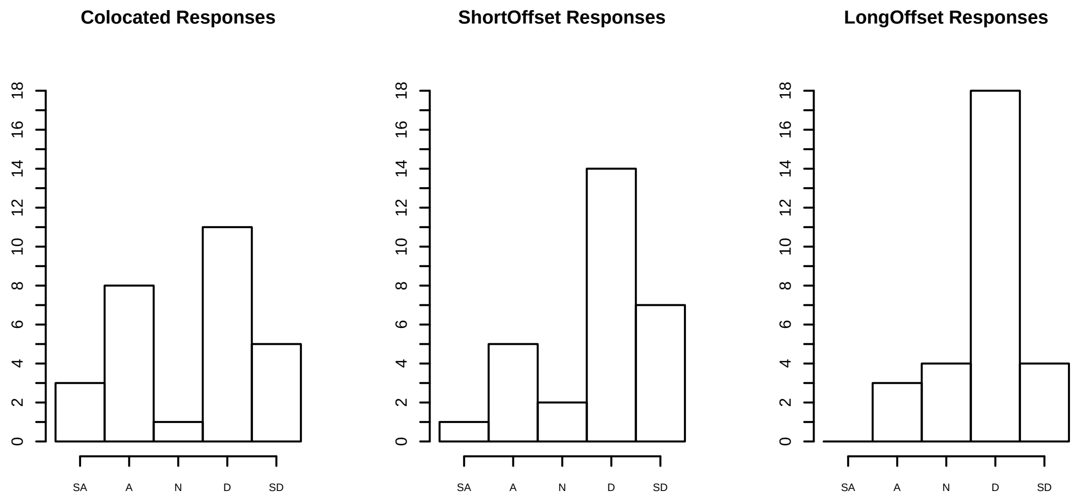
I felt that color matching was fatiguing.



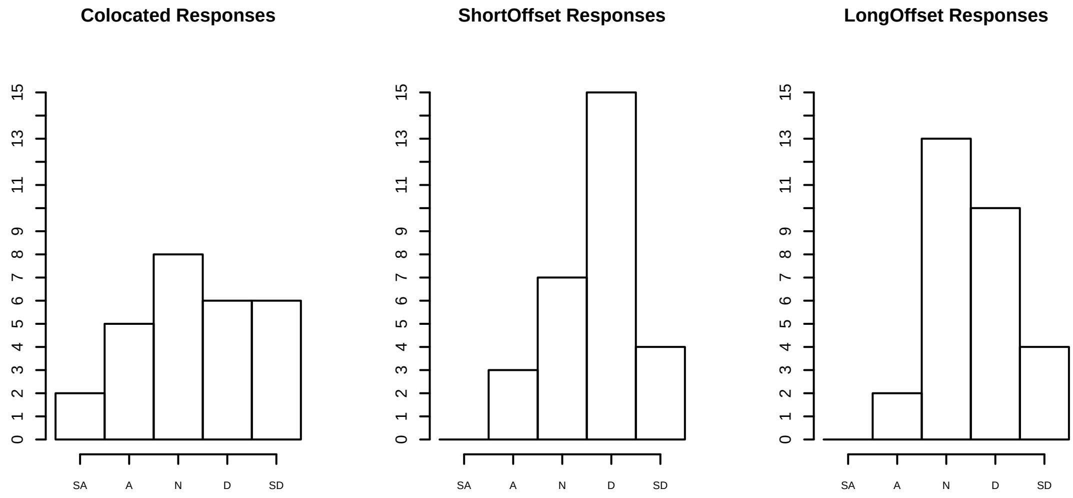
The volume in which I had to move my hand was a comfortable space to reach.



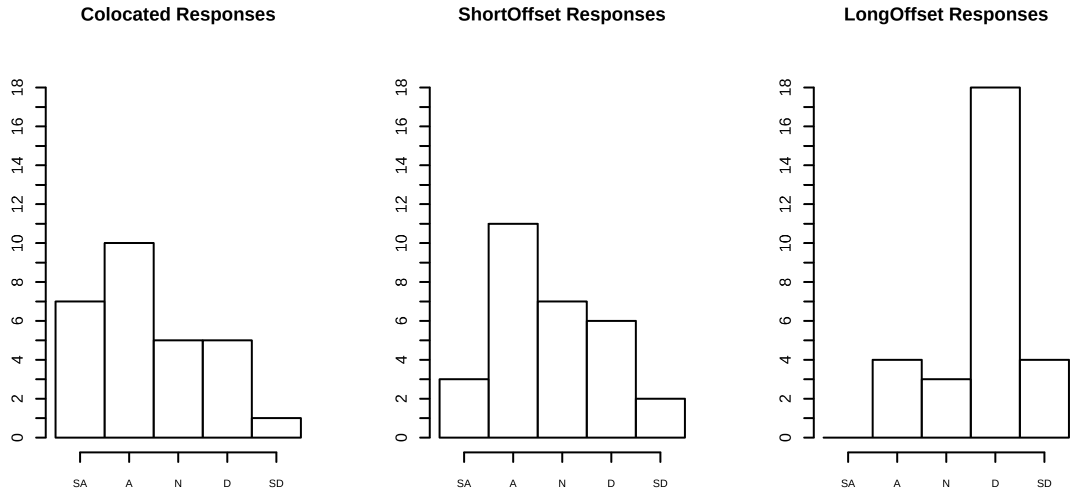
I was mainly looking at the color picker while trying to match colors.



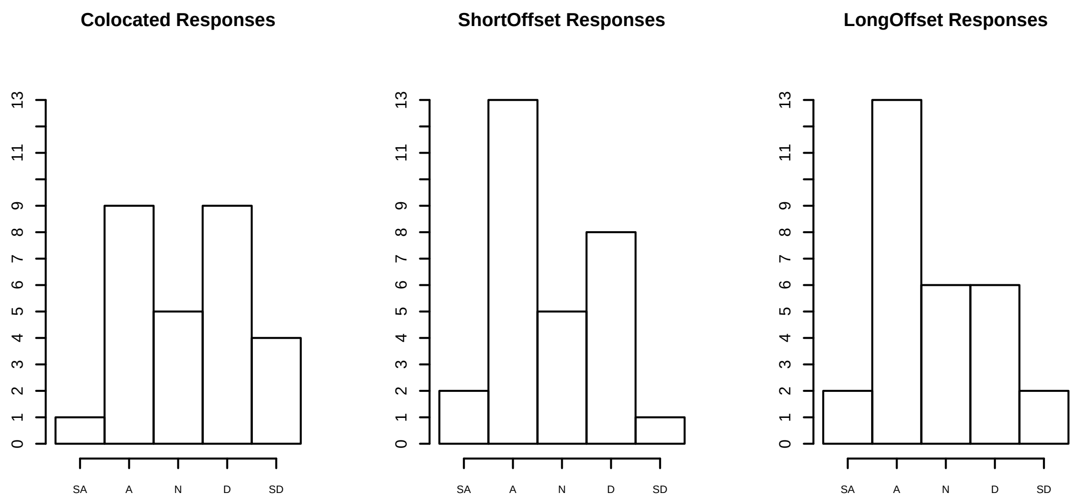
I saw my own hand while trying to match colors.



Seeing my own hand was distracting for matching colors with the color picker.



I felt that I could move my hand into the color picker.



Using the color picker felt natural.

Appendix B

R source code for statistical analyses and graphics

The program below was used to create all the figures and descriptive data contained in this work. The program was written and tested in the R 2.2.1 environment for Windows XP. See http://www.r-project.org for information about R.

```
# Get the data.
dataFilename <- "C:/thesisdata/allofit_tabbed.txt"
myData <- read.table( dataFilename, header=T)

# Derive variables.
rawDist <- myData[,"Distance"]
distPerTime <- rawDist / myData[,"Time"]
normDist <- (rawDist - min(rawDist)) / (max(rawDist) - min(rawDist))
accuracy <- 1 - normDist
accuracyPerTime <- accuracy / myData[,"Time"]
myData$Accuracy <- accuracy
myData$AccuracyPerTime <- accuracyPerTime

# Coerce factors. Condition is already a factor.
# Since it is a string, R makes it so.
myData$ColorName <- as.factor(myData$ColorNum)
myData$SubjectID <- as.factor(myData$Subject)

# Enumerate conditions.
conditions <- c("ColocatedCondition"
                "ShortOffsetCondition"
                "LongOffsetCondition")
```

```r
############################
# Helper functions.
############################

# Output means and standard deviations for a variable
# overall and per condition.
outputDescriptives <- function(theData, theColumn, theConditions)
{
  print("Overall")
  print(paste("Mean: ", mean(theData[theColumn])), digits=4)
  print(paste("StdDev: ", sd(theData[theColumn])), digits=4)

  for ( c in theConditions )
    {
      s <- subset(theData, Condition == c)
      print(c)
      print(paste("Mean: ", mean(s[theColumn])), digits=4)
      print(paste("StdDev: ", sd(s[theColumn])), digits=4)
    }
}

# Output descriptives for all variables of interest in a dataset.
outputAllDescriptivesForData <- function(theData)
{
  variablesOfInterset <- c("Distance","Time","Accuracy","AccuracyPerTime")
  conditions <- c("ColocatedCondition",
                  "ShortOffsetCondition",
                  "LongOffsetCondition")
  for ( interestingVar in variablesOfInterset )
    {
      print(interestingVar)
      outputDescriptives(theData, interestingVar, conditions)
    }
}


# Get the means for conditions in a data frame.
getMeans <- function(theData, theColumn, theConditions)
{
```

```
  m <- c()
  for ( c in theConditions )
    {
      s <- subset(theData, Condition == c)
      m <- c(m, mean(s[theColumn]))
    }
  data.frame(theConditions, m)
}

# Make the whole histogram of acc/time.
createPerConditionHistograms <- function(colName)
{
  maxDatum <- max(myData[colName])
  minDatum <- min(myData[colName])
  numBins <- 20
  bins <- c(minDatum
            minDatum + (1:numBins / numBins) * (maxDatum - minDatum))
  maxCnt <- 0
  accRange <- c(0,max(myData[colName]))

  # Get the range of possible counts so that the set of three histograms
  # will have the same scales.
  for ( cond in conditions )
    {
      accData <- subset(myData, Condition == cond, select = colName)
      accHist <- hist(accData[,1], breaks = bins, plot=F)
      maxCnt <- max(maxCnt, accHist$counts)
    }
  cntRange <- c(0,maxCnt)

        # Prepare a plot for 3 histograms.
  pdfFilename <- paste("C:/thesisdata/", colName,
                       "_per_condition_histograms.pdf", sep="")
  pdf(file = pdfFilename, width = 6.5, height = 2.5)
  layout(matrix(c(1,2,3),1,3))

        # Cycle over the conditions and output a histogram.
  for ( cond in conditions )
    {
      accData <- subset(myData, Condition == cond, select = colName)
```

```
      accHist <- hist(accData[,1], breaks = bins, plot=F)
      par(cex=.5)
      plot(accHist,
           main = paste("Histogram of", colName, cond),
           xlab = colName, ylab = "",
           xlim = accRange, ylim = cntRange)
    }

        # Write the file.
  dev.off()
}

############################
# Section 4, Figures 1-4.
############################

# Create the per conditions histograms.
variablesOfInterset <-
  c("Distance","Time","Accuracy","AccuracyPerTime")
for ( interestingVar in variablesOfInterset )
{
        createPerConditionHistograms( interestingVar )
}

############################
# Section 4, Figures 5.
############################

# Make a graph of statistics per trial number.
# To see if overall it improves with time.
variablesOfInterset <-
  c("Distance","Time","Accuracy","AccuracyPerTime")
trialNums <- 0:44
w <- 3
h <- 3

for ( interestingVar in variablesOfInterset )
{
  trialMeans <- c()
  for ( t in trialNums )
```

```
    {
      m <- mean(subset(myData, Trial == t,
                        select = interestingVar))
      trialMeans <- c(trialMeans, m)
    }
  theTitle <- paste("Trial Number vs Mean", interestingVar)
  xTitle <- "Trial Number"
  yTitle <- paste("Mean ", interestingVar)
  pdf(file = paste("C:/thesisdata/TrialNumvMean",
        interestingVar,".pdf", sep=""),
      width = w, height = h)
  par(cex=.5)
  plot(trialMeans, main = theTitle, xlab = xTitle, ylab = yTitle)
  fit<-lm(trialMeans~trialNums)
  abline(fit)
  dev.off()
}

###########################
# Section 4, Figure 6.
###########################

# Make a histogram of the ordering types used.
w <- 3 # inches.
h <- 3 # inches.
maleSubjectsOrdtypes <-
  subset(myData, Trial == 1 & Gender == "M",
         select = c(Subject, OrdType))
maleSubjectOrdHist <-
  hist(maleSubjectsOrdtypes[,"OrdType"],
       breaks=(.5 + (0:6)), plot=F)
femaleSubjectsOrdtypes <-
  subset(myData, Trial == 1 & Gender == "F",
         select = c(Subject, OrdType))
femaleSubjectOrdHist <-
  hist(femaleSubjectsOrdtypes[,"OrdType"],
       breaks=(.5 + (0:6)), plot=F)
orderingTypes <- c("CSL","LCS","SLC","LSC","CLS","SCL")
pdf(file = paste("C:/thesisdata/orderingtypes_males.pdf"),
    width = w, height = h)
```

```
par(cex=.5)
plot(maleSubjectOrdHist,
     main = "Histogram of Ordering Types for Male Subjects",
     xlab = "Ordering Types", axes=FALSE, ylim=c(0,3), ylab = "")
par(cex=.75)
axis(side=1, labels=orderingTypes, at=1:6)
axis(side=2, at=0:3)
dev.off()
pdf(file = paste("C:/thesisdata/orderingtypes_females.pdf"),
    width = w, height = h)
par(cex=.5)
plot(femaleSubjectOrdHist,
     main = "Histogram of Ordering Types for Female Subjects",
     xlab = "Ordering Types", axes=FALSE, ylim=c(0,3), ylab = "")
par(cex=.75)
axis(side=1, labels=orderingTypes, at=1:6)
axis(side=2, at=0:3)
dev.off()

# Find the golden 24 subjects. Discard subjects with the
# highest variance in these gender/ordering type pairs:
# F2, F3, M1, M3, M5.
for ( genderType in c("F","M") )
{
  for ( orderingType in 1:6 )
    {
      # Get the subjects for this ordering type/gender pair.
      genderOrdTypeSubjects <-
        subset(myData,
               Gender == genderType
               & OrdType == orderingType
               & Trial == 1, select=c(Subject))
      # If there was 3 subjects, it is one of
      # the pairings for which we need to discard one subject.
      if ((dim(genderOrdTypeSubjects)[1]) == 3) {
        print(c(genderType, orderingType))
        print(genderOrdTypeSubjects[1])
      # In the discard-first approach, we remove the
      # subject with the lowest id.
        discardFirstSubj <- c(discardFirstSubj, min(genderOrdTypeSubjects[1]))
```

```
      }
    }
}


##################################################
# Section 4.6 Comparison of means with ANOVA.
##################################################

compareMeansForBalancedPool <- function( subjectsToKeep )
{
        # Make the data.
  balData <- subset(myData, Subject %in% subjectsToKeep)

        # Output the descriptives
  outputAllDescriptivesForData(balData)

        # Output the t-test
        # Define 'without' from the help on in.

  "%w/o%" <- function(x,y) x[!x %in% y]

  for ( condition in conditions )
    {
      print(conditions %w/o% condition)
      print( t.test(Time ~ Condition, balData,
                    Condition != condition, paired=TRUE) )
      print( t.test(Distance ~ Condition, balData,
                    Condition != condition, paired=TRUE) )
      print( t.test(AccuracyPerTime ~ Condition, balData,
                    Condition != condition, paired=TRUE) )
    }

        # Output the anova results.
  taov <- aov(Time~(Condition*ColorName)
              +Error(SubjectID/(Condition*ColorName)), balData)
  aaov <- aov(AccuracyPerTime~(Condition*ColorName)
              +Error(SubjectID/(Condition*ColorName)), balData)
  daov <- aov(Distance~(Condition*ColorName)
              +Error(SubjectID/(Condition*ColorName)), balData)
```

```
    print("ANOVA summary for Time")
    print(summary(taov))
    print("ANOVA summary for Distance")
    print(summary(daov))
    print("ANOVA summary for Accuracy per Time")
    print(summary(aaov))
}

# Make the list of subjects to discard.
discardHighestAccPerTimeSD <- c(20, 5, 28, 18, 10)
discardFirstSubj <- c(20, 10,  1, 11, 18)

# Define 'without' from the help on in.
"%w/o%" <- function(x,y) x[!x %in% y]

compareMeansForBalancedPool( 1:29 %w/o% discardHighestAccPerTimeSD )
compareMeansForBalancedPool( 1:29 %w/o% discardFirstSubj )


##############################
# Section 4, Figure 7-8.
##############################

# Widget placement.
dataFilename <- "C:/thesisdata/widgetplacement.txt"
placementData <- read.table( dataFilename, header=T)

# Chosen offset histogram.
mean(-placementData$Offset)
sd(placementData$Offset)
offsetHist <- hist(-placementData$Offset, breaks=80, plot=F)
pdf(file = paste("C:/thesisdata/chosen_offset_hist.pdf")
    width = 4, height = 4)
par(cex=.5)
plot(offsetHist, main = "Histogram of Chosen Offsets",
     xlab = "Chosen offset in feet", ylab = "")
dev.off()

# Actively chosen offset histogram.
activeOffset <-
```

```
  subset(placementData, abs(InitOff - Offset) > .001,
         select=c("Offset"))
mean(-activeOffset$Offset)
sd(activeOffset$Offset)
activeOffsetHist <- hist(-activeOffset$Offset, breaks=80, plot=F)
pdf(file = paste("C:/thesisdata/chosen_active_offset_hist.pdf"),
    width = 4, height = 4)
par(cex=.5)
plot(activeOffsetHist, main = "Histogram of Actively Chosen Offsets",
     xlab = "Chosen offset in feet", ylab = "")
dev.off()

# Chosen offset vs. variables.
# Time
pdf(file = paste("C:/thesisdata/chosen_offset_v_time.pdf"),
    width = 4, height = 4)
par(cex=.5)
myfm <- placementData$Time ~ placementData$Offset
plot(myfm, main = paste("Chosen offset v Time"),
     xlab = "Chosen offset in feet", ylab = "Time")
abline(lm(myfm))
dev.off()

# Distance
pdf(file = paste("C:/thesisdata/chosen_offset_v_distance.pdf"),
    width = 4, height = 4)
par(cex=.5)
myfm <- placementData$Distance ~ placementData$Offset
plot(myfm, main = paste("Chosen offset v Distance"),
     xlab = "Chosen offset in feet", ylab = "Distance")
abline(lm(myfm))
dev.off()


#####################################
# Appendix A
#####################################

# Demographics.
# Get the data.
```

```
demographicDataFilename <- "C:/thesisdata/demographics.csv"
demoData <- read.csv( demographicDataFilename )

summary(demoData)

# Make pdfs for the likert questions.
likertAnswers <- c("SA","A","N","D","SD")

for ( q in c("VideoGame","VR","Graphics") )
{
  pdf(file = paste("C:/thesisdata/experience_",q,"_responses.pdf",
        sep=""),
      width = 4, height = 4)
  qHist <- hist(demoData[,q], breaks=(.5 + (0:5)), plot=F)
  par(cex=.5)
  plot(qHist, main = paste(q, "Experience Responses"), xlab = "",
       axes=F, ylab = "")
  axis(side=2, at=0:max(qHist$counts))
  par(cex=.3)
  axis(side=1, labels=likertAnswers, at=1:5)
  dev.off()
}

# Per condition post questions.
postDataFilename <- "C:/thesisdata/perconditionpost.csv"
postData <- read.csv( postDataFilename )

# Make histograms for the post questionnaire responses.
likertAnswers <- c("SA","A","N","D","SD")
conditions <- c("Colocated","ShortOffset","LongOffset")
questions <- sapply(1:14, function (x) { paste("Q",x, sep="") })

for ( q in questions )
{
        # Find the max count.
  maxCnt <- 0
  for ( c in conditions )
    {
      s <- subset(postData, Condition == c)
      qHist <- hist(s[,q], breaks=(.5 + (0:5)), plot=F)
```

```
      maxCnt <- max(maxCnt, max(qHist$counts))
    }

  pdf(file = paste("C:/thesisdata/", q, "responses.pdf", sep=""),
      width = 6, height = 3)
  layout(matrix(c(1,2,3),1,3))
  for ( c in conditions )
    {
      s <- subset(postData, Condition == c)
      qHist <- hist(s[,q], breaks=(.5 + (0:5)), plot=F)
      par(cex=.5)
      plot(qHist, main = paste(c, "Responses"), xlab = "", axes=F,
            ylab = "", ylim = c(0,maxCnt+1))
      axis(side=2, at=0:maxCnt)
      par(cex=.3)
      axis(side=1, labels=likertAnswers, at=1:5)
    }
  dev.off()
}
```