Specimen Box: A Tangible Interaction Technique for World-Fixed Virtual Reality Displays

David J. Zielinski*
Duke Immersive Virtual Environment
Duke University, USA

Derek Nankivil †
Department of Biomedical Engineering
Duke University, USA
Johnson & Johnson Vision Care Inc.

Regis Kopper‡
Department of Mechanical Engineering
and Materials Science
Duke University, USA

Abstract

Object selection and manipulation in world-fixed displays such as CAVE-type systems are typically achieved with tracked input devices, which lack the tangibility of real-world interactions. Conversely, due to the visual blockage of the real world, head-mounted displays allow the use of many types of real world objects that can convey realistic haptic feedback. To bridge this gap, we propose *Specimen Box*, an interaction technique that allows users to naturally hold a plausible physical object while manipulating virtual content inside it. This virtual content is rendered based on the tracked position of the box in relation to the user's point of view. Specimen Box provides the weight and tactile feel of an actual object and does not occlude rendered objects in the scene. The end result is that the user sees the virtual content as if it exists inside the clear physical box. We hypothesize that the effect of holding a physical box, which is a valid part of the overall scenario, would improve user performance and experience. To verify this hypothesis, we conducted a user study which involved a cognitively loaded inspection task requiring extensive manipulation of the box. We compared Specimen Box to Grab-and-Twirl, a naturalistic bimanual manipulation technique that closely mimics the mechanics of our proposed technique. Results show that in our specific task, performance was significantly faster and rotation rate was significantly lower with Specimen Box. Further, performance of the control technique was positively affected by experience with Specimen Box.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities

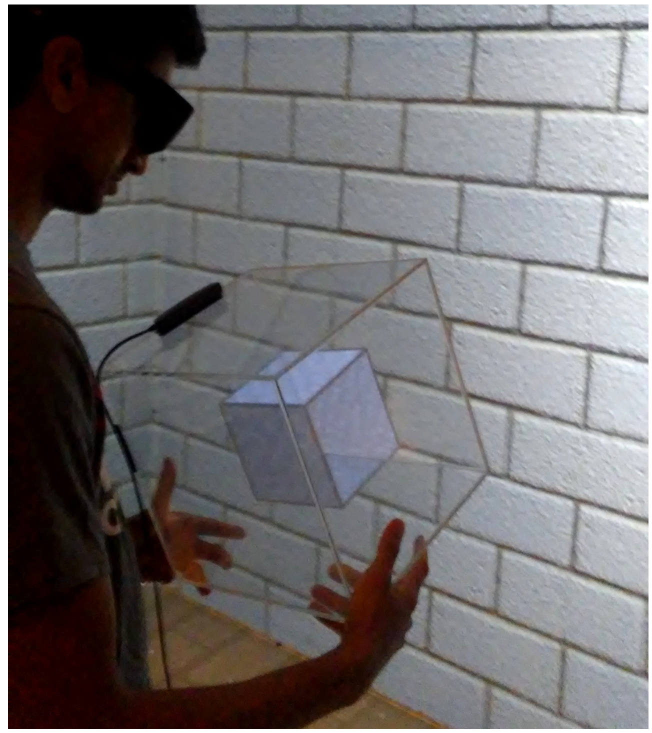

Figure 1: A user utilizing the proposed Specimen Box interaction technique

1 Introduction

Selection and manipulation of objects in world-fixed displays[1] such as CAVE-type virtual reality (VR) systems are limited by the narrow avenue of interaction possibilities provided by wands and game controllers. Issues such as occlusion and the lack of realistic haptic feedback caused by these input devices hinder the ability to perform realistic interactions. This paper proposes Specimen Box[2], a novel tangible interaction technique that uses passive haptics to overcome limitations of traditional world-fixed display input devices.

1.1 World-Fixed Displays

Options to display and experience virtual content can generally be classified as either world-fixed, such as CAVE-type systems, or user-fixed, such as HMDs. Even though user-fixed displays offer convenient advantages such as mobility and affordability, world-fixed displays show advantages which invite further research on usable and efficient interaction techniques.

Researchers have found that CAVEs lead to increased presence [22] and reduced simulator sickness [24] when compared to HMDs. These advantages may come from multiple factors, originating from one or more of the following: superior field of view [16], seams in the display which can act as rest frames [10], comfortable accommodation distances (the walls of the CAVE are at a pleasant distance from the user's eyes) [18], scene continuity (available even during quick head turns), and personal awareness (the ability to see one's own real body) [4]. Additionally, world-fixed displays allow users to have a social group experience (although in most world-fixed systems there is still only one truly correct view point).

Further, world-fixed displays are able to offer very high res-

*e-mail:djzielin@duke.edu

†e-mail:dnankivi@its.jnj.com – This work was carried out entirely at Duke University without consult or influence by Johnson & Johnson Vision Care Inc. As of June 2016, Derek Nankivil is employed by Johnson & Johnson Vision Care Inc.

‡e-mail:regis.kopper@duke.edu

[1]The term world-fixed display was proposed by Jerald [21] in contrast with head-mounted displays (HMDs). These include CAVE-type systems, wall-sized displays and desktop VR displays. Our proposed technique applies to any display defined as world-fixed.

[2] This technique is presented in short form in [46].

olution and frame rates which are not foreseeable for user-fixed displays. For example, research on extremely high frame rates (>1000 fps) is only currently available via DLP projectors [27] and there are possibilities with projector-based systems to achieve autostereoscopy (permitting completely unencumbered operation) [34]. Collectively, world-fixed displays present a number of potential benefits that should be explored through new techniques, such as Specimen Box, proposed and demonstrated herein.

1.2 Passive Haptics in Virtual Reality

Passive haptics offer a compelling way to increase user experience beyond indirect input through game controllers and wands. Studies conducted by Insko [20] used a VR pit simulator with passive haptics cues by raising up the walkway off the floor slightly so that the edge of the walkway could be sensed by the participants [30]. This passive haptics ledge led to a significant increase in heart rate and skin conductivity responses. In a second experiment, participants were trained to navigate a maze-like environment. One group trained in VR and had styrofoam walls co-located to the walls in the virtual environment, while the other group trained in VR with no haptics. Results showed that participants who trained with passive haptics had significantly faster completion time and less collisions when assessed on the maze while blindfolded [20].

Another successful use of passive haptics devices is in the medical domain. Early VR work showed that passive props representing the head and cutting planes helped neurosurgeons with visualization [17]. Laparoscopic surgery simulators were developed with force feedback probes [37]. Results showed that trainees utilizing VR were faster, made less errors, and had more economy of motion on the follow up assessment.

Researchers have realized that clear tangible objects may be employed in world-fixed displays. Early work by Encarnação *et al.* on the "Translucent Sketchpad" [13] and work by Coquillart *et al.* on the "Virtual Palette" [6] also utilized the idea that the user could hold a clear prop (in their case the sketchpad or pallete) for the user to interact on. By carefully co-locating the rendered image (from their VR workbench / single wall setup) they were able to allow the user to virtually interact with or write on the prop. While the use of a transparent prop is similar to our technique, "Translucent Sketchpad" pursues a pen and paper metaphor for interaction and "Virtual Palette" pursues a menu metaphor. Specimen box is different as it pursues a 2 handed 3D box grabbing metaphor.

Work by Kawakami *et al.* [23] developed the concept of "Object-Oriented Displays", where the faces of the objects are used as display screens. Taking that concept further Inami *et al.* [19] proposed a head-mounted projector that was able to project onto objects. More recently, researchers proposed autostereoscopic displays on the faces of a cube [28]. These techniques are all conceptually related to Specimen Box, but with key differences. Displaying objects on the faces of a cube restricts the display of content behind the cube. Part of the appeal of the Specimen Box technique is that we see through the box (the user can see both the virtual scene and their hands), thus making the translucent box more plausible in an encompassing VR scene. In addition, Specimen Box requires only a see-through tracked cube, while previously proposed techniques require actual screens or displays on the cube.

It is clear that there are advantages to incorporating passive haptics into simulations and training. However, passive haptics have been historically hard to incorporate into world-fixed displays because they are visible to the user. Our proposed interaction technique overcomes this limitation by using a clear box and displaying its virtual content directly inside it through the world-fixed display.

1.3 Historical Motivation

The inspiration for the proposed technique drew from a study of history and a desire to conduct exposure studies of stimuli eliciting disgust and fear. We refer to the 1800s "specimen jars" that were used to house preserved specimens of various animals and human organs. Some specimen jars date back to the 1600s, and were stored together (along with other oddities: shrunken heads, sculptures, and even fantastic fabrications of mythical creatures) in "curiosity cabinets" by aristocrats, collectors, and what would come to be called natural scientists. These objects were available for seeing, and even smelling and tasting [5]. Such jars became more common place after the discovery of formaldehyde (1867) and the formalization of the field of Natural History. We thus propose our virtual reality interaction technique and dub it "Specimen Box" (see figure 1) as a reference to the historical practice of specimen jars.

1.4 Specimen Box Technique and Implementation

Specimen Box embodies an object manipulation technique, in a clear acrylic box, where users can pickup, touch and feel the box, but can't quite physically reach its contents (figure 1). By using this interaction metaphor, Specimen Box affords natural manipulation, while maintaining plausibility of the inner object, through consistent tracking and realism of its perspective rendering.

To implement this technique, We commissioned a box consisting of a cube with outer dimensions of 26.35 cm. While this number is somewhat arbitrary, we wanted an object that would be comfortable to grasp with two hands. Holding our hands out in front of us, shows we can avoid flexing outward if the object is smaller than about 30cm. In looking at some of the balls utilized in sports, we saw that football (soccer) balls are 22 cm, American basketballs are 24 cm in diameter. So, we estimate that a box in the range of 20 cm to 30 cm may be comfortable to handle with two hands. Also, the size of the box was selected to balance two competing factors. If the box is too small, the size of the virtual object that can be contained within it will be overly restricted, but if the box is too large, handling and manipulation will be overly cumbersome.

We were concerned that the user might be able to squeeze the box, and thus create curvature that would distort the image. So, it was important to choose a wall thickness that could withstand the user pressures. From some simple calculations we determined that deflections would be negligible with a 3/16 inch (0.476 cm) thick transparent acrylic. The box could theoretically be constructed from materials available at most local hardware stores; however, in order to have an aesthetically pleasing box, we had it professionally made [1] for approximately 100 USD. The manufacturer utilized SC-94 brand acrylic cement for gluing the panels together. These dimensions and thickness led to a box box that weighed 2,105 g. Adding the affixed tracking sensor (36 g), led to a total weight of 2,141 g. To preserve the box integrity that could be compromised from the glue off-gassing, a small hole of 5.86 mm diameter was drilled near one of the edges.

In order to use the box within a VR context, it needs to be furnished with 6 degrees of freedom (6-DOF) tracking. We attached an IS-900 wireless head tracking sensor to the top of the box. A cable ran from the sensor on the box to a power pack/transmitter clipped onto the belt/pocket of the user. The tracking sensor can be seen on Figure 1. By referencing the tracking sensor values through software, a virtual object can be rendered at an offset from the incoming tracker position, such that it appears inside the transparent acrylic box.

2 User Study

In order to evaluate the possible advantages of the Specimen Box interaction technique, we conducted a user study. Our goal was to understand the differences between Specimen Box and a usable virtual object manipulation technique that closely mimicked the biomechanics of Specimen Box in terms of user performance (time and correctness), physical motions taken (translation and rotation), and subjective feedback from the participants.

2.1 Choice of Virtual Manipulation Technique

We propose to compare Specimen Box to an existing bimanual interaction technique, already known and developed by the research community. Prior work was previously undertaken to create a series of classifications of the way 2 hands can cooperate [15]. Guiard studied everyday actions to create classifications such as unimanual (one hand used), bimanual symmetric (2 hands used, and each hand either performs the same action or an action in phase with the other hand), and finally bimanual asymmetric (a more complicated action/interaction between the two hands).

Cutler *et al.* [8] used Guiard's framework to develop and explore a series of bimanual interaction techniques. From the list, Cutler proposes that the Grab-and-Carry method (defined as bimanual symmetric) matches how users handle objects in the real world. Conceptually, a line is created between the two hands. We can imagine that, if the object we want to manipulate is forced to stay on that line, by moving the hands, we can achieve a 5-DOF system. A slight addition is made in Cutlers Grab-and-Twirl method, in which a rotation from one of the hands (Cutler suggests using only the right hand) provides roll around the line connecting the two objects. With the *Grab-and-Twirl* method we have found a 6-DOF bimanual interface that closely mimics how objects are manipulated in the real world. We have, thus, decided to compare Specimen Box to Grab-and-Twirl.

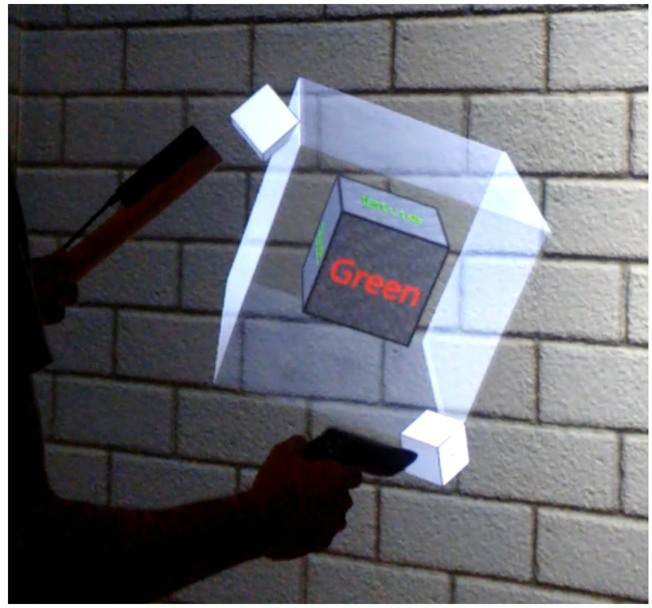

Figure 2: Example of user performing an inspection task with the Grab-and-Twirl manipulation technique. Note that a virtual acrylic box is rendered.

2.2 Display Apparatus and Software

We used the Duke immersive Virtual Environment (DiVE), a six-sided CAVE-type [7] system to perform the experiment. Tracking was provided via an Intersense IS-900 tracking system. For the Grab-and-Twirl evaluation, we utilized a tracking sensor attached to a short wood stick measuring approximately 30cm. The dominant hand used the standard Intersense wireless wand, which was furnished with 6 buttons and 2 joystick axes. In our experiment, only the wand trigger was used for Grab-and-Twirl manipulation.

Each wall of the DiVE has two Christie Digital WU7K-M projectors running at 120 Hz. These projectors are overlapped and blended, and give a total resolution of 1920x1920 pixels per wall. Active stereo is achieved via Volfoni Edge RF liquid crystal shutter glasses. Our simulation was visually simple enough that we were able to run at our full system performance, which was 120 Hz active stereo - effectively 60 fps.

For the simulation software, we used Unity 5.3 with the MiddleVR plugin [26] to support clustered CAVE-type renderings. The scripts to control various aspects of the simulation inside Unity were written in C#. Logs were written out to disk in the form of CSV files, and later combined via a helper program into one large CSV file (containing all of the users) that was loaded into the IBM SPSS statistical software.

The virtual environment was a small room, coinciding with the walls of the DiVE. A Stone texture was shown on the walls, and a wood floor texture on the floor. The room scenario was created in order to give a sense of presence and to allow some objects to be displayed behind the box. In the case of the Grab-and-Twirl condition, we added a virtual representation of the physical box to eliminate confounds (see figure 2) and make the task technique as similar to Specimen Box as possible.

2.3 Participants

Participants were screened for family history of photosensitive epileptic seizure and asked to have normal or corrected-to-normal vision (we conducted in person screening for color blindness). We recruited 22 volunteers, but due to technical difficulties of the equipment, the data from 2 participants were not analyzed, leading to 20 participants included in the data analysis. The ages ranged from 18-47 years (M=28.2, SD=7.6). Seven participants (35%) were female. All participants (in the Grab-and-Twirl condition) held the wand with the same hand they used for a mouse (the right hand for every participant in our cohort). Nine participants (45%) had previous experience with CAVE-type systems. Twelve participants (60%) had prior HMD experience. Every participant that reported CAVE experience, had also experienced an HMD before, so our overall level of prior VR experience on any type was 60%. Participants were compensated with a $10 gift card.

2.4 Word Counting Stroop Task

In an effort to create a task with varying levels of cognitive load, we utilized the well known Stroop Effect [41]. John Ridley Stroop released one of the most widely cited papers in Psychology in 1935 on what came to be known as the Stroop Effect. In this effect, when a user reads a word, for example the word "Red" written in ink colored red - there is a congruent condition between the word and the color. However, interesting things happen when the word itself "Red", is written in blue ink. This creates a mismatch in the brain and is referred to as an "incongruent condition". The user will be asked to read the word or asked about the color of the ink and the time for users to respond is taken. The incongruent conditions (with the word not the same as the ink) generally shows slower response times. This has been traditionally a pen and paper test, but modern researchers have used it with success in recent VR studies [33, 24]. One important thing to consider is that since the Stroop Effect utilizes words in a particular language, results may differ when the participant is bilingual [35]. Specifically, lower fluency in English corresponds to a reduction in performance on the Stroop test (leading to more time to answer the question) [42]. This could be an issue in our experiment, as all of our participants claimed English fluency but 7 participants (35%) identified as non-native speakers. However, we anticipated some protection against this potential confounding factor as all participants were their own control.

In our study, we used 3 levels of increasing task difficulty of the Stroop task–congruent with blank faces (a cube in this configuration could have 4 faces with the word green written in green, and 2 blank faces), congruent with distractors (the same example could have 4 faces with the word green written in green, and 2 faces with the word blue written in blue), and finally incongruent (2 faces with

with the word red written in green ink, 2 faces with with the word blue written in green ink, and the two remaining faces could be inked red with non red words (blue, green)). See figure 3 for visual examples of difficulty conditions 2 and 3.

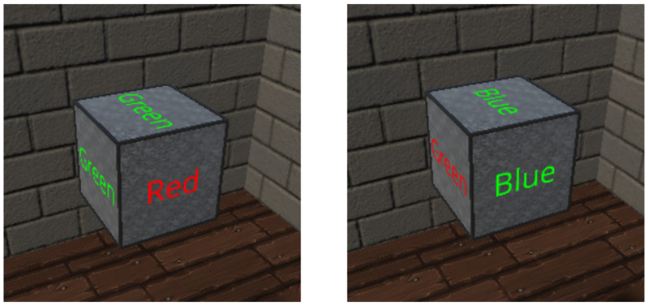

Figure 3: Two different Stroop conditions. Left: congruent (ink color matches word); right: incongruent (ink color does not match word)

2.5 Procedure

Upon arrival, participants signed an informed consent form, were evaluated for color blindness, conducted a spatial aptitude test (Cube Comparison Test S-2 [11]), and completed a brief background survey containing demographic items as well as video game and VR experience. A tutorial followed where participants were shown how to manipulate the box. The user was asked to stand in the center of the system and face the front wall. Participants were asked to remain standing in the same place, but were allowed free movement of their head and hands. As initial practice, they were allowed to manipulate the box on their own for 1 minute. Following that, 2 examples of each of the 3 difficulty levels were presented (for a total of 6 training trials). During the tutorial, if the participant proposed the wrong answer, the experimenter corrected the participant, and allowed them to reinspect the target cube. During the actual trials, no feedback on correctness was provided.

Each trial began with the computer proposing a question visually to the user, for example "How many faces have words colored blue?", and also spoken using the built-in Microsoft text-to-speech voice. Then a ding sound was played to alert the user that the trial was beginning. The participants could then freely rotate/manipulate the physical/virtual box however they wanted. When they were sure of the answer to the proposed question they stated "ready!". This allowed the experimenter to press the space bar, which stopped the timer. Then, the participant spoke the answer, which was also keyed in by the experimenter. Thus, the answering of the proposed question could be thought of as utilizing a "wizard of oz" voice control system [9]. The system then asked the user to align the box to a common starting orientation. On completion of the alignment, the next trial began.

As providing an exhaustive list of all possible cube configurations would be impractical, we presented 12 randomly selected cube configurations. Here we summarize the randomization algorithm. Each 12 trials per difficulty block were split into 2 groups of 6. This provided 6 trials for each type of question - looking for the word or looking for the ink color. In each sub-grouping of 6, we had one each of the number of faces 1,2,3,4,5,6. Within each subgroup of 6 we would randomly assign 2 red, 2 blue, 2 green. Thus, while not covering all theoretically possible combinations, we were ensuring that the answer colors and number of faces were evenly represented. Once all 12 trials were configured, specific configurations for the cube (that could represent 4 faces of green for example), for each cell were generated. Finally the blocks of 12 trials were randomly shuffled, thus interspersing questions about the word or about the ink. The algorithm would then move up to the next difficulty level (12 more, congruent with distractors), and finally 12 in the hardest (incongruent) for a total of 36 trials per interaction technique. This schedule was computed and written out by a helper program to a CSV file (and then loaded by the Unity-based Virtual Environment) before each launch of the program, ensuring that each run was randomized, and thus minimizing any learning effects that could occur.

After the completion of all 3 difficulty blocks from easiest to most difficult (i.e., 36 trials), participants completed a survey about their experience with the interaction technique. At this point, the participants were offered a break for bathroom or water. When ready, the tutorial was started again, this time with a different input device. On completion of the tutorial the 36 trials were run, the device evaluation survey was completed, and then a final overall preference survey was completed. The order of interaction technique was counterbalanced between subjects.

2.6 Design

The study was conducted with three within-subjects independent variables – interaction technique (IT–Specimen Box, Grab-and-Twirl), difficulty level (congruent, congruent+distractors, incongruent) and trial number (12). We counterbalanced the order of IT presentation and considered it as a between-subjects factor to test for ordering effects. The study thus had a mixed design with 3 within-subjects factors and 1 between-subjects factor. Within an IT, each difficulty block consisted of 12 trials. The tasks were presented in order of difficulty from easiest to hardest and each participant performed a total of 72 trials (2 ITs $\times$ 3 difficulty $\times$ 12 trials). No user took longer than one hour to complete the whole experimental session.

3 RESULTS

3.1 Data Collection

Participants performed a standard spatial abilities test and background data on age, gender, native language, video gaming experience and VR experience was collected.

We collected objective data as well as subjective responses from participants. The objective measures were accuracy, time to complete a trial (s), rotation rate of the real or virtual box (deg/s) and translation rate of the real or virtual box (m/s). Subjective responses comparing the ITs were collected for a subset of standard presence questions[39] ("Please rate your sense of being in the virtual environment."; "To what extent were there times during the experience when the virtual environment was the reality for you?"; "When you think back about your experience, do you think of the virtual environment more as images that you saw, or more as somewhere that you visited?"), ease of use, learnability and cumbersomeness. We also collected overall ratings for each IT as well as the overall preferred method.

3.2 Data Analysis

Data was analyzed with IBM SPSS 24. For time, rotation rate and translation rate, data was visually inspected for normality and skewness and kurtosis were within approximate normal ranges. Data was tested for sphericity, using the Maulchly's test, and when the assumption was violated, the Greenhouse-Geisser correction on the degrees of freedom was used. For pairwise multiple comparisons, the Bonferroni correction was used.

A 4-way mixed-design factorial analysis of variance (ANOVA) with three repeated-measures factors and one between-subjects factor was performed for time, rotation rate and translation rate while subjective ratings were analyzed with the Wilcoxon signed rank test.

Pearson's correlation tests were conducted to relate spatial ability test score, VR experience, video game frequency of use and age to the outcome variable time.

For clarity, units are not displayed along with the means. Time units are seconds, rotation rate units are degrees per second and translation rate units are meters per second. Also for clarity, we refer to Specimen Box as *SB* and to Grab-and-Twirl as *GT*, while difficulty is represented from 1 to 3 as easiest (congruent without distractors) to most difficult (incongruent).

3.2.1 Missing Data

On a few occasions, tracking was lost, or participants waited too long before they performed the first trial due to misunderstanding. In those situations, data was missing for that particular trial. In order to avoid the list-wise removal of a participant from a single missing cell, we replaced the missing data with the average for the trial block of the respective participant for each instance [25]. A total of 15 (1.04%) missing data points were replaced.

3.2.2 Accuracy Measure

Participants were asked to perform as fast as possible, but always respond with confidence. For that reason, the vast majority of the trials were accurate ($1421/1425 = 98.5\%$ of valid trials). For that reason, we did not perform statistical analysis on accuracy, but focused on time, rotation rate and translation rate.

3.3 Correlations

Time was significantly related to the age of participants ($r = .214, p = .019$), indicating that the older participants were, the longer the trials took. There were significant inverse relationships of time with spatial ability ($r = -.412, p < .0001$)–participants with higher spatial ability completed trials faster. Similarly, there was an inverse relationship of time with frequency of playing real time strategy ($r = -.266, p = .003$) and sports/action ($r = -.309, p = .001$) video games–the more frequently participants played video games the faster they were. There was no significant relationship of time with amount of VR experience.

3.4 Main Effects

Looking at the order factor of IT, there were no main effects in the study for time ($\mu_{SB\_first} = 5.34$, $\mu_{GT\_first} = 5.70$, $F_{1,18} = .37, p = .55$), rotation rate ($\mu_{SB\_first} = 75.16$, $\mu_{GT\_first} = 78.97$, $F_{1,18} = .12, p = .733$) or translation rate ($\mu_{SB\_first} = .197$, $\mu_{GT\_first} = .170$, $F_{1,18} = 1.69, p = .21$). There were, however, significant interaction effects with ordering (see section 3.5).

For the IT factor (see figure 4), time was significantly lower with SB ($\mu_{SB} = 5.28$, $\mu_{GT} = 5.77$, $F_{1,18} = 5.53, p < .05$) and rotation rate was significantly lower with SB ($\mu_{SB} = 64.79$, $\mu_{GT} = 88.34$, $F_{1,18} = 22.16, p < .0001$), while translation rate did not show a significant difference ($\mu_{SB} = .183$, $\mu_{GT} = .184$, $F_{1,18} = 0, p = .984$).

Difficulty also showed a main effect for time (figure 5; $\mu_1 = 5.71$, $\mu_2 = 5.53$, $\mu_3 = 5.32$, $F_{2,36} = 5.42, p < .01$), where pairwise comparisons showed that trials completed with difficulty 1 took significantly longer than those completed with difficulty 3 ($p < .05$). Difficulty levels were not significantly different among each other for rotation rate ($F_{1.42,25.6} = .439, p = .648$) or translation rate ($F_{2,36} = .262, p = .771$).

The trial variable had significant differences for time ($F_{5.58,100.5} = 4.71, p < .0001$). Pairwise comparisons on time show that the only significant differences occurred between the first and second trials, indicating that there was a steep but quick learning curve, and time was consistent after trial 1. There were no main effects of trial for rotation rate ($F_{5.20,93.5} = 4.71, p = .062$) or translation rate ($F_{3.25,58.5} = 1.27, p = .293$).

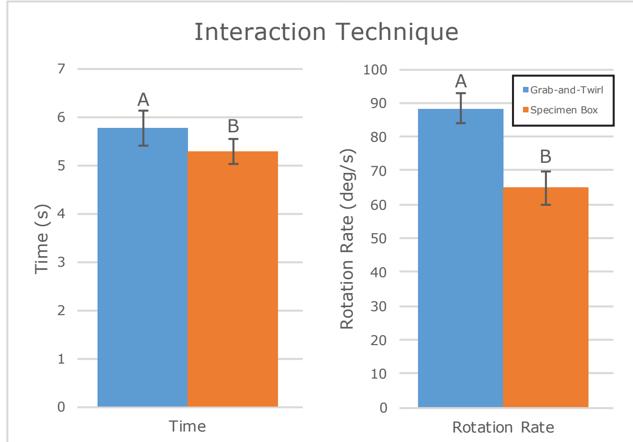

Figure 4: Time and rotation rate for each interaction technique. Error bars show standard error. Levels not connected by the same letter are significantly different.

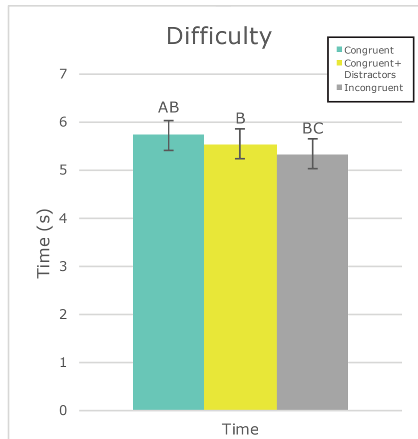

Figure 5: Time for each Difficulty. Error bars show standard error. Levels not connected by the same letter are significantly different.

3.5 Interaction Effects

Although there was no main effect of ordering, we observed some interaction effects which involved ordering. There was a significant interaction of ordering and IT for time ($F_{1,18} = 16.177, p < .005$), as shown in Figure 6. Pairwise comparisons show that when GT was performed first, it took significantly more time than SB ($p < .0001$). However, there were no significant differences between GT and SB when SB was performed first ($p = .253$). This is an interesting result, and we discuss its implications in section 4.2.

There was also a 3-way interaction effect across ordering, difficulty and IT for time ($F_{2,36} = 6.394, p < .005$). Pairwise comparisons provide insight into this interaction effect. Looking at GT with difficulty 1, it was significantly slower when it was performed first than when it was performed last ($p < .05$), while there were no significant differences for difficulties 2 ($p = .108$) or 3 ($p = .324$). Conversely, when looking at the ordering where SB was performed first, GT was significantly faster than SB in difficulty 1 ($p < .05$), with no significant differences with difficulty 2 ($p = .340$) or 3 ($p = .777$). Finally, GT was significantly slower in difficulty 1 as compared to difficulty 3 when it was performed first ($p < .005$), and SB was significantly slower in difficulty 1 as compared to difficulty 3 when it was performed first ($p < .05$).

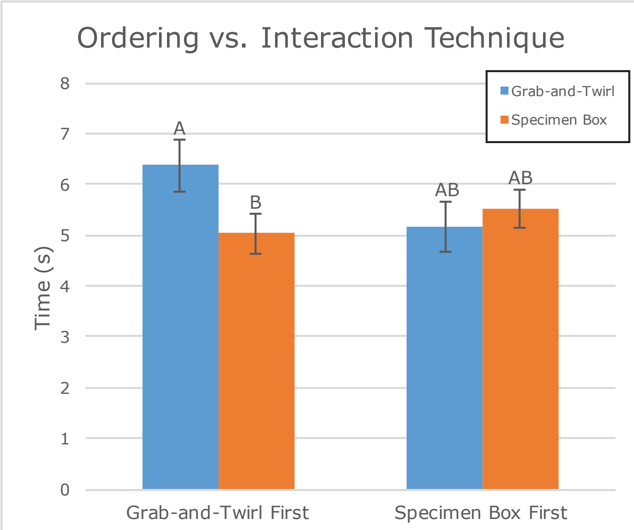

Figure 6: Interaction effect of ordering with interaction technique for time. Error bars show standard error. Levels not connected by the same letter are significantly different.

The last significant interaction effect observed in the study was a four-way interaction between all factors (ordering, difficulty, trial and IT) for *rotation rate*. An interaction effect with so many factors is difficult to interpret, and pairwise comparisons didn't elicit any immediate insights.

3.6 Subjective Feedback

The analysis of survey questions showed that presence was significantly perceived as higher for GT for the three questions asked ("sense of being there" – $Mdn_{GT} = 5.5$, $Mdn_{SB} = 4$, $N = 17$, $z = -2.53$, $p < .05$; "VR experience was reality" – $Mdn_{GT} = 5$, $Mdn_{SB} = 3$, $N = 15$, $z = -1.99$, $p < .05$; "images vs. places"– $Mdn_{GT} = 4$, $Mdn_{SB} = 2$, $N = 15$, $z = -2.78$, $p < .01$).

Overall, 14 participants (70%) preferred GT over SB, although there was not a significant difference in the overall rating of each IT ($Mdn_{GT} = 4$, $Mdn_{SB} = 3$, $N = 17$, $W = 40.5$, $insignificant$). Participants found SB significantly more cumbersome than GT ($Mdn_{GT} = 2$, $Mdn_{SB} = 3$, $N = 18$, $z = -2.20$, $p < .01$). However, ease of use was not significantly different across ITs ($Mdn_{GT} = 4$, $Mdn_{SB} = 5$, $N = 14$, $W = 44.5$, $insignificant$), nor was perceived learning speed ($Mdn_{GT} = 4$, $Mdn_{SB} = 5$, $N = 8$, $W = 10.5$, $insignificant$).

Free form comments were collected and coded. Fourteen participants commented that the box was too heavy. A small number of participants commented on visual distortions of the physical box as distracting (1 participant), on reflections of the physical box as distracting (3 participants), and on how tethering of physical box tracker interfered with motions (3 participants). We discuss these comments in section 4.

4 Discussion

4.1 Time Advantage of Specimen Box

The Specimen Box technique was significantly faster than Grab-and-Twirl, even with the heavy weight of the box (see discussion in section 4.6). The tactile nature of Specimen Box could have provided accuracy in the movements of the user. Thus the user would not overshoot or have to redo intended operations. However, it would be hard (in the context of our current experiment) to gain insight into this factor (accuracy/intention), but future studies could probe deeper in this direction.

Another hypothesis is that the tactile nature of the box allowed the user to intuitively (perhaps through proprioception [32]), know what face of the inner target box they were looking at. This could result in less time needed to render judgment if they had indeed covered all faces for inspection.

A further factor could be in terms of the action-response cycle, or closed-loop actions, of the user. We could conceive that with Specimen Box, the user could operate the physical box at the unrestricted frame rate of reality. In the Grab-and-Twirl condition, however, all the stimulus is rendered virtually, and with added latency. Although less likely than the previous hypotheses, the combination of a lower-than-reality 60 Hz frame rate with latency in the order of 50 ms could have caused artifacts that affected time performance [12, 44].

4.2 Learning Effects

Perhaps the most interesting insight gained from this study is the ordering effect of interaction technique. When Grab-and-Twirl was performed first, it was significantly slower than Specimen Box; however, there were no significant differences between Grab-and-Twirl and Specimen Box when Specimen Box was performed first. This indicates that performing the tasks initially with a more natural technique caused participants to improve their performance with the less intuitive Grab-and-Twirl. One interpretation for this result is that participants were able to exercise an optimal strategy from the beginning with the physical box, whose learning readily transferred to the virtual box manipulation.

Only in one condition, when looking at the interaction effect among ordering, difficulty and interaction technique, was Grab-and-Twirl faster than Specimen Box. That was when Specimen Box was performed first with the easiest difficulty. Here we are looking at the very first block of trials performed with Specimen Box, and at the fourth block of trials, performed with Grab-and-Twirl (because it was performed second). Again, this provides evidence that Specimen Box could be learned rather quickly, such that the most difficult blocks did not suffer from an ordering effect as compared to Grab-and-Twirl.

4.3 Rotational Minimization of Specimen Box

The Specimen Box technique showed significantly less rotation per second as compared to Grab-and-Twirl. This seems to show a certain economy of motion in completing the task. Our main hypothesis is that when faced with physical actions, the brain automatically optimizes for economy of motion. This is a natural survival strategy, as the brain wants to avoid spending additional energy [2]. We see similar economy of motions (faster time and more accuracy) in previous VR research, which compares naturalistic interfaces (real walking [36]) or similar biomechanical motions (walking in place [14]). We believe that this exposes an interesting and important issue for the design of training systems in VR. It seems that users may adopt different strategies when physical muscle motions are involved. Thus, to maximize efficiency of training transfer for real physical tasks, it would be preferable to favor the use of tangible interaction devices and to avoid interaction techniques that offer artificially weightless manipulation.

4.4 Negative Presence Effects of Specimen Box

When comparing the reduced SUS Presence questionnaire between Specimen Box and Grab-and-Twirl, the results show that presence was significantly higher when participants were using Grab-and-Twirl. This was initially not expected, as previous studies have shown that passive haptics [20] and real world biomechanics (real walking) [43] increase presence. In the Specimen Box case, however, since the physical box exists in relation to low-realism visual renderings (our virtual environment is not photo realistic), it may have caused a break in plausibility illusion (Psi) [38]. In other words, when the rendered target cube was shown with the rendered box and the rendered walls (i.e., everything rendered) it may have allowed greater suspension of disbelief, increasing the Psi of the experience. When a real object–the acrylic box–was added to the scene, it may have had the opposite effect. It may also be that the issues with the box (discussed is section 4.6), especially its weight, added a load or distraction to the users that prevented a higher feeling of presence. Ultimately, our study aimed mainly at performance and usability testing, and thus the virtual environment was not built with the goal of presence in mind. Future studies could aim to examine the Specimen Box technique in terms of presence. It would also be valuable to focus on user subjective reaction to the object inside the box, rather than the overall environment/experience, through feelings of "social" presence to the box.

4.5 Favorability of Specimen Box vs. Grab-and-Twirl

Although participants generally preferred Grab-and-Twirl over Specimen Box, our proposed technique outperformed the bimanual virtual manipulation technique in terms of time and rotation rate. We believe the participants negative subjective ratings were largely due to the size and weight of the box, which was not very comfortable. In the following section, we discuss some potential ways to overcome current limitations of Specimen Box that could lead to improved performance, experience and subjective feelings.

4.6 Specimen Box Limitations

4.6.1 Weight of the Box

The most common comment (14 participants), was that the box was too heavy and became fatiguing. The most direct way to overcome this limitation is to make the box wall thinner. Scaling the box down would also reduce its weight. However, scaling down the box limits the size of the object that can be virtually placed inside it. A more creative option would be to fill the box with a lighter-than-air gas, such as helium. We also considered reducing the density of the air inside the box to approach a vacuum (first proposed by jesuit priest Francesco Lana de Terzi in 1670, and recently considered for military use [31]), although this would only reduce the weight of the box by 1 or 2 g.

4.6.2 Visual Distortions

We noticed in our own observations of the specimen box (also reported by one participant) that minimal distortions are present when looking through the box. This could be due to issues with our tracker. We hypothesize that comparing the virtual object to the physical box could show variations, revealing a sort of misregistration in the mapping between real world positions and tracker-reported values. The distortions could also be due to the fact that optical imaging systems such as the human eye are subject to the effects of refraction when light travels from one medium to another medium with a different refractive index.

Even though the distortions are minimal and were mentioned by a single experiment participant, it is possible to correct the refractive distortions and provide a more seamless user experience. With knowledge of the refractive index, shape and size of the transparent structure and of the viewer's gaze direction and the orientation and relative position of the box, the images displayed can be pre-warped to effectively correct the refractive distortions from the perspective of the viewer. More importantly, applying corrections to refractive distortions would not be limited to structures composed of parallel plates, but to any three dimensional transparent structure, solid or hollow, (such as a cuboid, cylinder, cone, pyramid, torus, etc...). Three dimensional refraction correction based on work by Zhao *et al.* [45] could be implemented to correct for refractive distortions arising from any of these structural forms. Future work should consider applying such corrections, especially for physical objects that have convex or concave shapes. In a future manifestation of the transparent Specimen Box, we anticipate the use of thinner walled acrylic material which will reduce some of observed optical distortions.

4.6.3 Reflections on the Box Walls

In addition to positional visual distortions, we also noticed that the images rendered to appear inside the box would sometimes show as reflections on different inside surfaces of the box. In looking at the free form comments, three users mentioned this artifact. We have considered the idea of anti-reflection coatings for the plastic walls. The limitation of this approach is that the anti-reflection coatings are designed for a particular angle of incidence and for a particular wavelength. They can be broadened to cover the full visible spectrum, and the performance can be flattened somewhat over a range of angles, but it is limited. The typical reflection is about 4%. With anti-reflection coatings, this can be reduced to $< 1\%$, but as a drawback, the reflection at high angles of incidence will take on a colored appearance. Although the cost of producing a clear box with anti-reflective coating would be higher, future work could explore the benefits of such an improvement.

4.6.4 Tethering of the Tracker

The tethering of the box tracker was also seen as an encumbrance for 3 participants of the study. The tracking sensor was mounted to the box, and a cable came down to a power pack on the participant's belt. Our other option was to mount the power pack to the box. This would make it untethered, but would increase the area of occlusion and weight of the box. In principle we prefer no devices on the box to minimize occlusions. We see this limitation as a feature of current technology and as tracking sensors become miniaturized, the issue would be eliminated or significantly reduced.

4.6.5 Occlusion From User's Hands

One potential limitation of the Specimen Box technique is that the user's hands could potentially occlude the content meant to be displayed inside the box. This would happen if the user was holding the box, and instead of gripping the box on either side, gripped the box on what would be far side of the box. Another, perhaps mechanically more realistic gripping motion would be to support the box from underneath. However, in our study no users mentioned this occluding issue. We hypothesize that users accept the fact that occlusion exists and suspend enough disbelief to see it as a feature of the technology, not of the experience. We also note that this same issue occurs in CAVE-type systems in general–if one imagines outstretching one's arm, content that was meant to appear between the user's head and the user's blocking arm will be occluded.

4.7 Extensions to the Paradigm

4.7.1 Multiple Boxes

To improve ecological realism, a range of boxes could be made to match the weight of the box to the assumed weight of the virtual object inside it. This could be useful for an ecologically valid skills training session, where the user has to manipulate different objects with varying weights. Multiple boxes with varying weights would be available to the user, and the appropriate box would be selected to match the weight of the virtual object. The ability to design the weight of box to mimic the object seems one of the most alluring potentialities for the interaction paradigm introduced by Specimen Box.

4.7.2 Possibilities of a Spherical Specimen Container

One idea we have entertained is to make the container a sphere, or some other shape. A sphere could be used in complementary situations to the box. However, to make a optically high quality sphere is much more difficult than making a box. Considering standard subtractive manufacturing techniques, compared to a box, a sphere requires a larger volume of material for manufacturing. It would also require more machining time and polishing time, causing it to be significantly more expensive than the box. If made, the refractive distortions would be significant. The distortions of a hollow transparent sphere–essentially a hollow ball lens–could be corrected. Optically, this would behave very similarly to a pair of meniscus lenses in series, which would add defocus, spherical aberration and significant barrel distortion. The latter of which, being the most significant distortion, could be corrected by simply warping the displayed images to remap the distorted field points properly. These distortions would be very noticeable if uncorrected, even if the wall thickness were very thin.

4.8 Application of Specimen Box

The Specimen Box in its current form and with extensions proposed in this paper has great potential for a myriad of application areas. Our results pointed in the direction that the use of a physical box in virtual object manipulation benefits not only time and movement efficiency, but it also may directly impact training transfer, as it affords more natural biomechanical symmetry [29] of a VR-trained task and its real-world counterpart. We believe that using Specimen Box for assembly and manufacturing training is perhaps the most direct use of its potential. While we don't believe that Specimen Box can replace in-situ training, we strongly believe that it has the potential to provide training that is more physically consistent with the real-world task.

Another potential application area for Specimen Box is in learning and exploration. We can conceive smaller scale versions of Specimen Box as dynamic museum exhibits, where visitors can pick up a box, contained in front of a small world-fixed display, and explore many different virtual artifacts. The ability to manipulate the box and see the object inside it from different perspectives poses a great potential benefit to the learning experience.

5 Conclusion and Future Work

We have presented Specimen Box, a tangible interaction technique for object manipulation in world-fixed VR displays. Specimen Box opens up exciting future possibilities for bringing a tangible interface into reach for those utilizing world-fixed display systems. We have shown in our initial user study that, for our specific task, the Specimen Box technique allowed user's to have faster performance in a cognitively loaded task compared to an existing bi-manual technique (Grab-and-Twirl). Results from subjective feedback informed us of several changes that need to be made to the current design, especially making the box lighter (perhaps by making the walls of the box thinner). We believe that with additional work on the physical box form factor, the Specimen Box technique could gain greater favorability ratings.

We propose to compare our world-fixed technique to user-fixed displays using a passive haptics box. We hypothesize there may be benefits from interacting with passive haptics devices in a world-fixed display, as recent research showed that, in cognitively loaded tasks, seeing one's hands (even if just as an avatar) may have positive effects [40]. We would like to explore whether seeing one's real hands, as possible in a CAVE-type system, is different from an avatar representation of the hands, which are often slightly misregistered spatially in an HMD system.

In the future, a systematic exploration of the various form factors of the box should be conducted. Endeavoring to discover what works best in terms of box shape and box size would help those wishing to better utilize the Specimen Box technique, and also potentially give insight at what was driving the advantages seen in the technique.

While the goal of this work was the description and initial evaluation of a new technique, it should ideally be considered against the broad array of interaction techniques available to users of world-fixed VR displays. Specifically the Specimen Box technique should be compared to the simple virtual hand interaction technique[3].

Finally, this technique needs to be explored on a task with more ecological validity. Instead of the abstract counting faces task, inspection of a real artifact, looking for cracks or flaws, could be an interesting use of Specimen Box. Another use that we did not explore yet, but believe has strong potential is to use Specimen Box for manipulation of virtual objects beyond rotation. For example, the box could be used to pickup, translate and drop objects in the virtual environment.

References

[1] Clear solutions displays - manufacturer of clear acrylic objects. `http://cleardisplays.com/`. Accessed: 2016-09-14.

[2] H.-R. Berthoud. Multiple neural systems controlling food intake and body weight. *Neuroscience & Biobehavioral Reviews*, 26(4):393–428, 2002.

[3] D. A. Bowman and L. F. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics*, pages 35–ff. ACM, 1997.

[4] G. Bruder, F. Steinicke, K. Rothaus, and K. Hinrichs. Enhancing presence in head-mounted display environments by visual body feedback using head-mounted cameras. In *CyberWorlds, 2009. CW'09. International Conference on*, pages 43–50. IEEE, 2009.

[5] C. Classen. Museum manners: The sensory life of the early museum. *Journal of Social History*, 40(4):895–914, 2007.

[6] S. Coquillart and G. Wesche. The virtual palette and the virtual remote control panel: a device and an interaction paradigm for the responsive workbench/sup tm. In *Virtual Reality, 1999. Proceedings., IEEE*, pages 213–216. IEEE, 1999.

[7] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti. Surround-screen projection-based virtual reality: the design and implementation of the CAVE. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pages 135–142. ACM, 1993.

[8] L. D. Cutler, B. Fröhlich, and P. Hanrahan. Two-handed direct manipulation on the responsive workbench. In *Proceedings of the 1997 symposium on Interactive 3D graphics*, pages 107–114. ACM, 1997.

[9] N. Dahlbäck, A. Jönsson, and L. Ahrenberg. Wizard of oz studies: why and how. In *Proceedings of the 1st international conference on Intelligent user interfaces*, pages 193–200. ACM, 1993.

[10] H. B.-L. Duh, D. E. Parker, and T. A. Furness. An independent visual background reduced simulator sickness in a driving simulator. *Presence: Teleoperators and Virtual Environments*, 13(5):578–588, 2004.

[11] R. B. Ekstrom, J. W. French, H. H. Harman, and D. Dermen. Manual for kit of factor-referenced cognitive tests. *Princeton, NJ: Educational testing service*, 1976.

[12] S. R. Ellis, F. Bréant, B. Manges, R. Jacoby, and B. D. Adelstein. Factors influencing operator interaction with virtual objects viewed via head-mounted see-through displays: viewing conditions and rendering latency. In *Virtual Reality Annual International Symposium*, pages 138–145. IEEE, 1997.

[13] L. Encarnacão, O. Bimber, D. Schmalstieg, and S. Chandler. A translucent sketchpad for the virtual table exploring motion-based gesture recognition. In *Computer Graphics Forum*, volume 18, pages 277–286. Wiley Online Library, 1999.

[14] S. C. Grant and L. E. Magee. Navigation in a virtual environment using a walking interface. Technical report, DTIC Document, 2000.

[15] Y. Guiard. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Journal of motor behavior*, 19(4):486–517, 1987.

[16] C. Hendrix and W. Barfield. Presence within virtual environments as a function of visual display parameters. *Presence: Teleoperators & Virtual Environments*, 5(3):274–289, 1996.

[17] K. Hinckley, R. Pausch, J. C. Goble, and N. F. Kassell. Passive real-world interface props for neurosurgical visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 452–458. ACM, 1994.

[18] W. A. IJsselsteijn, H. de Ridder, and J. Vliegen. Effects of stereoscopic filming parameters and display duration on the subjective assessment of eye strain. In *Electronic Imaging*, pages 12–22. International Society for Optics and Photonics, 2000.

[19] M. Inami, N. Kawakami, D. Sekiguchi, Y. Yanagida, T. Maeda, and S. Tachi. Visuo-haptic display using head-mounted projector. In *Virtual Reality, 2000. Proceedings. IEEE*, pages 233–240. IEEE, 2000.

[20] B. E. Insko. *Passive haptics significantly enhances virtual environments*. PhD thesis, University of North Carolina at Chapel Hill, 2001.

[21] J. Jerald. *The VR Book: Human-Centered Design for Virtual Reality*. Morgan & Claypool, 2015.

[22] M. C. Juan and D. Pérez. Comparison of the levels of presence and anxiety in an acrophobic environment viewed via hmd or cave. *Presence*, 18(3):232–248, 2009.

[23] N. Kawakami, M. Inami, D. Sekiguchi, Y. Yanagida, T. Maeda, and S. Tachi. Object-oriented displays: a new type of display systems-from immersive display to object-oriented displays. In *Systems, Man, and Cybernetics, 1999. IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on*, volume 5, pages 1066–1069. IEEE, 1999.

[24] K. Kim, M. Z. Rosenthal, D. J. Zielinski, and R. Brady. Effects of virtual environment platforms on emotional responses. *Computer methods and programs in biomedicine*, 113(3):882–893, 2014.

[25] R. Kopper, D. A. Bowman, M. G. Silva, and R. P. McMahan. A human motor behavior model for distal pointing tasks. *International journal of human-computer studies*, 68(10):603–615, 2010.

[26] S. Kuntz. MiddleVR a generic VR toolbox. In *2015 IEEE Virtual Reality (VR)*, pages 391–392. IEEE, 2015.

[27] P. Lincoln, A. Blate, M. Singh, T. Whitted, A. State, A. Lastra, and H. Fuchs. From motion to photons in 80 microseconds: Towards minimal latency for virtual and augmented reality. *IEEE transactions on visualization and computer graphics*, 22(4):1367–1376, 2016.

[28] R. Lopez-Gulliver, S. Yoshida, S. Yano, and N. Inoue. Poster: Toward an interactive box-shaped 3d display: Study of the requirements for wide field of view. In *3D User Interfaces, 2008. 3DUI 2008. IEEE Symposium on*, pages 157–158. IEEE, 2008.

[29] R. P. McMahan. Exploring the effects of higher-fidelity display and interaction for virtual reality games. 2011.

[30] M. Meehan, S. Razzaque, M. C. Whitton, and F. P. Brooks. Effect of latency on presence in stressful virtual environments. In *Virtual Reality, 2003. Proceedings. IEEE*, pages 141–148, March 2003.

[31] T. T. Metlen. Design of a lighter than air vehicle that achieves positive buoyancy in air using a vacuum. Technical report, DTIC Document, 2012.

[32] M. R. Mine, F. P. Brooks Jr, and C. H. Sequin. Moving objects in space: exploiting proprioception in virtual-environment interaction. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 19–26. ACM Press/Addison-Wesley Publishing Co., 1997.

[33] T. Parsons and C. Courtney. Interactions between threat and executive control in a virtual reality stroop task.

[34] T. Peterka, R. L. Kooima, J. I. Girado, J. Ge, D. J. Sandin, A. Johnson, J. Leigh, J. Schulze, and T. A. DeFanti. Dynallax: solid state dynamic parallax barrier autostereoscopic vr display. In *2007 IEEE Virtual Reality Conference*, pages 155–162. IEEE, 2007.

[35] M. Rosselli, A. Ardila, M. N. Santisi, M. D. R. Arecco, J. Salvatierra, A. Conde, , and S. Bonie Leni. Stroop effect in spanish–english bilinguals. *Journal of the International Neuropsychological Society*, 8(06):819–827, Sept. 2002.

[36] R. A. Ruddle and S. Lessels. The benefits of using a walking interface to navigate virtual environments. *ACM Trans. Comput.-Hum. Interact.*, 16(1):5:1–5:18, Apr. 2009.

[37] N. E. Seymour, A. G. Gallagher, S. A. Roman, M. K. OBrien, V. K. Bansal, D. K. Andersen, and R. M. Satava. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Annals of surgery*, 236(4):458, 2002.

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

[39] M. Slater, M. Usoh, and A. Steed. Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 3(2):130–144, 1994.

[40] A. Steed, Y. Pan, F. Zisch, and W. Steptoe. The impact of a self-avatar on cognitive load in immersive virtual reality. In *2016 IEEE Virtual Reality (VR)*, pages 67–76. IEEE, 2016.

[41] J. R. Stroop. Studies of interference in serial verbal reactions. *Journal of experimental psychology*, 18(6):643, 1935.

[42] P. A. Suarez, T. H. Gollan, R. Heaton, I. Grant, and M. Cherner. Second-language fluency predicts native language stroop effects: Evidence from spanish–english bilinguals. *Journal of the International Neuropsychological Society*, 20(03):342–348, 2014.

[43] M. Usoh, K. Arthur, M. C. Whitton, R. Bastos, A. Steed, M. Slater, and F. P. Brooks Jr. Walking $>$ walking-in-place $>$ flying, in virtual environments. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 359–364. ACM Press/Addison-Wesley Publishing Co., 1999.

[44] C. Ware and R. Balakrishnan. Reaching for objects in vr displays: lag and frame rate. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 1(4):331–356, 1994.

[45] M. Zhao, A. N. Kuo, and J. A. Izatt. 3D refraction correction and extraction of clinical parameters from spectral domain optical coherence tomography of the cornea. *Optics express*, 18(9):8923–8936, 2010.

[46] D. J. Zielinski, D. Nankivil, and R. Kopper. 6 degrees-of-freedom manipulation with a transparent, tangible object in world-fixed virtual reality displays. In *2017 IEEE Virtual Reality (VR)*. IEEE, 2017.