Evaluation of Docking Task Performance Using Mid-air Interaction Techniques

Vanessa Vuibert
Centre for Intelligent Machines
McGill University
Montréal QC, H3A 0E9
vvuibert@cim.mcgill.ca

Wolfgang Stuerzlinger
School of Interactive Arts + Technology
Simon Fraser University
Surrey, BC V3T 0A3
w.s@sfu.ca

Jeremy R. Cooperstock
Centre for Intelligent Machines
McGill University
Montréal QC, H3A 0E9
jer@cim.mcgill.ca

ABSTRACT

Mid-air interaction has the potential to manipulate objects in 3D with more natural input mappings. We compared the performance attainable using various mid-air interaction methods with a mechanically constrained input device in a 6 degrees-of-freedom (DoF) docking task in both accuracy and completion time. We found that tangible mid-air input devices supported faster docking performance, while exhibiting accuracy close to that of constrained devices. Interaction with bare hands in mid-air achieved similar time performance and accuracy compared to the constrained device.

Keywords

3D interaction, 3D docking task, unconstrained mid-air

1. INTRODUCTION

Computer-vision-based tracking systems, exemplified by products such as the Kinect One and the Leap Motion, are now easily accessible on the mass-market. Simultaneously stereoscopic displays for gaming and entertainment have also become increasingly popular. These trends support and encourage the possibility of unconstrained mid-air interaction with a virtual 3D world, in a manner that approximates how we interact with the physical world. This vision is also promoted by augmented reality products, such as the Atheers One[1] and Meta [2], which render stereoscopic 3D content and track the users' hand gestures with built-in depth-sensing cameras. But can we, in fact, manipulate virtual 3D content quickly and accurately without the benefit of special-purpose constrained desktop devices?

This question motivated the studies described here. Our intent was to determine how mid-air interaction compares to existing alternatives for a non-trivial task in virtual environments. Specifically, we wanted to evaluate the possibility that efficient and accurate manipulation of 3D content may be supported without the need for a constrained desktop input device. If so, we would also like to determine whether a hand-held input device is even necessary, or if tracking of the user's hands can potentially suffice.

We chose to study 3D docking as our main task, which requires both orientation and positioning of an object with respect to a target. Our contribution is an exploration of docking performance using various mid-air interaction techniques, and the comparison of this performance to that attained with a desktop device that is considered to be ideally suited for 6 DoF manipulations. The focus of our study was not the docking strategy itself, but rather the performance attainable with various input devices.

Several earlier studies investigated docking tasks using traditional wireframe graphics. However, for our experiment, we chose a richer graphical environment, offering improved depth cues with lighting and shadow effects, as this permits users to reuse existing skills. We also propose a mapping that allows users greater flexibility in the manner in which they manipulate the virtual object. Moreover, we discuss the need to evaluate not only the docking time, but also the accuracy of the final position and orientation of the object.

2. RELATED WORK

2.1 Evaluation of Input Conditions

There has been significant prior research investigating 3D manipulation using desktop devices, in particular, investigating 3D position and/or orientation tasks. These include the virtual trackball [26], Rockin'Mouse [1], GlobeFish and GlobeMouse [7], multitouch surfaces used in conjunction with indirect [6] and direct interaction techniques, e.g., DS3 and StickyTools [6, 24, 10, 18], and mid-air interaction techniques such as Go-Go [23], as summarized in Table 1. Very few experiments compared constrained desktop-based devices to unsupported devices that can be manipulated freely in mid-air. A notable exception is early research by Zhai and Milgram [33], which demonstrated that for a docking task, isomorphic manipulation through a 6 DoF unsupported device was faster but less accurate than non-isomorphic manipulation with a 6 DoF elastic-rate-controlled device. However, it was unclear whether the time-accuracy tradeoff was more a result of the differences between isomorphic and rate-controlled input, or supported vs. mid-air interaction.

Placement (3 DoF), orientation (3 DoF) and docking (6 DoF) are fundamental tasks for manipulation of 3D content. However, comparisons of performance between input devices on such tasks are often frustrated by the lack of a standard experimental design. Bérard et al. [2] compared various devices for a 3D placement

[1] www.atheerlabs.com
[2] www.spaceglasses.com

SUI 2015, August 8–9, 2015, Los Angeles, CA, USA.

ACM 978-1-4503-3703-8/15/08 ...$15.00.
DOI: http://dx.doi.org/10.1145/2788940.2788950

Study	Task	Fastest Technique	Other Compared Techniques
Zhai et al. [33]	docking	mid-air	constrained device
Froehlich et al. [7]	docking	GlobeFish & GlobeMouse	mouse
Berard et al. [2]	placement	mouse	DepthSlider, SpaceNavigator, mid-air
Wang et al. [30]	placement	Phantom	mouse
Kratz et al. [14]	orientation	mid-air	multi-touch screen
Glesser et al. [8]	docking	Phantom, dual multi-touch surfaces	trackpad, mouse

Table 1: Past Research on 3D manipulation tasks.

task and found that the mouse, used in conjunction with orthographically projected views, was the fastest. However, computer-generated scenes often lack some of the depth cues that we rely on in the physical world to discriminate depth. This factor may account for at least some of the difference in human performance observed for tasks in the virtual compared to the physical world. With the addition of an improved visualization technique to compensate for limited depth cues, Wang et al. demonstrated that the Phantom could achieve higher performance on the same task [30], consistent with results from a more recent study [8].

Most placement, orientation and docking experiments only measure the time it takes participants to dock the cursor [7, 14, 8], but accuracy is often equally important. A docking task involves gross motion and then fine-tuning once near the target. Zhai et al. [33] measured how much the cursor's actual path differed from the shortest path to the target, both in terms of position and orientation. While there is interesting information in such trajectories, we are more interested in the accuracy of the final position and orientation, i.e., the docking result, as the evaluation criterion.

2.2 Visualization

Grossman et al. used motion capture cameras to track hand gestures as the fingers interact on the transparent spherical enclosure of a 3D volumetric display [9]. Although such volumetric displays offer the benefit of a true 3D display, consumer-level stereoscopic 3D, as used in many virtual and augmented reality displays, is a considerably more affordable and easily obtainable technology.

Stereoscopic 3D rendering and shadow-casting were found to improve accuracy in positioning tasks and permitted subjects to perform 3D placement tasks faster [13]. However, they did not improve rotation tasks [3]. In a stereoscopic rendering condition, direct mid-air interactions outperformed multi-touch screen techniques when the target was further away from the screen [4]. This may be due, in part, to the fact that while focusing on the finger that touches the multi-touch screen, the stereo image rendered above the screen appears blurred. Although stereoscopic rendering is often associated with simulator sickness, this is not a problem for docking tasks because the scene is static and the user focuses on a single object [25].

2.3 Gestures

The choice of interaction gestures is a critical factor in usability and performance. Previous studies [16, 27] used a handle bar metaphor to perform mid-air translations and rotations, where the virtual object being manipulated is imagined to be between the fists of the user. The main limitation of this technique is that the handle bar pose becomes fatiguing when users need to keep their arms extended to manipulate the handle bar for longer periods of time. A study by Hincapie et al. recommended to keep motions between the hip and the shoulder, and to minimize arm extension [11].

Tracking the translation and rotation of one hand is less fatiguing than using the handle bar technique. Levesque et al. proposed using the left hand for selection and the right hand for translation and rotation operations [15]. Cutler et al. proposed a more natural approach by using a pinch gesture to grab the virtual object and performing the 6 DoF operations with the same hand [5], similar to the 6 DoF Hand technique described by Mendes et al. [19]. Although not specified in their published descriptions, we suspect that these techniques require the users to always start their operations with their hand oriented so that it is pointing at the display. For example, if users were to grab the virtual object from the right side with their right hand and twist their wrist around its Z axis (local frame) (Figure 1), the object would rotate around the Z world axis instead of the X axis.

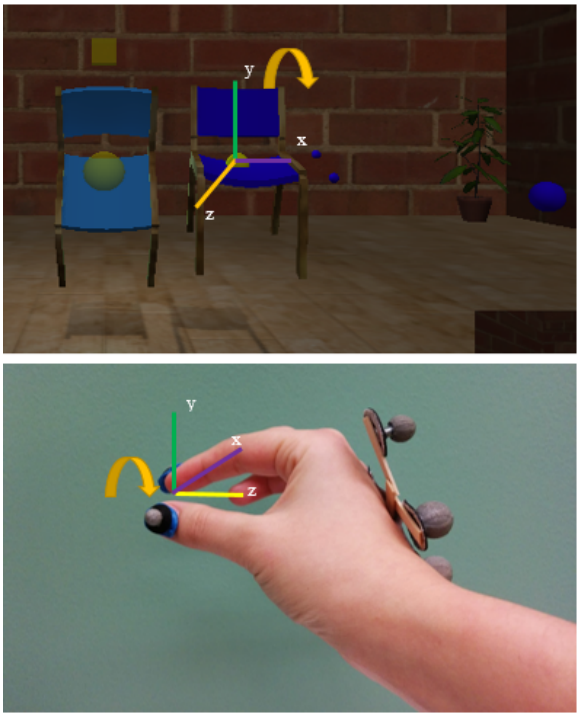

Figure 1: Rotating the dark blue chair around its X axis with the hand.

3. METHODOLOGY

A docking task was used to compare performance of three mid-air interaction options, using either a physical replica of the virtual object, a wand-like device or the user's hands. As baseline we chose a mechanically constrained input device, the Phantom Omni,

which can be used to provide 6 DoF input. This device demonstrated its superiority in terms of time performance in relation to other desktop devices in a recent docking study [8]. For our experiment, participants were asked to dock a moving "cursor" chair using a combination of translation and orientation operations, with a similar lighter colored target that remains fixed in the middle of the screen throughout each trial (Figure 2).

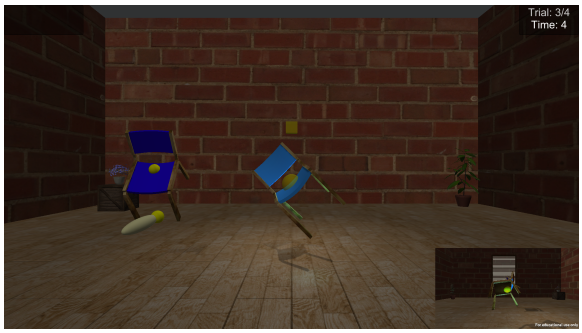

Figure 2: The virtual representation of the AirPen is visible in the image as an ellipsoid. The second camera on the bottom right shows a view from the right side.

3.1 Experimental Task

Previous docking studies used a set of predetermined target positions and orientations to avoid visual ambiguities [33, 8]. Since we offer a richer virtual environment for the docking task and thus are less affected by the visual ambiguity issue, the moving chair is placed randomly at the start of each trial, within a predetermined distance range from the target, which is assigned a uniformly distributed random orientation. A trial is completed once the participant succeeds in aligning the moving chair to the target position and orientation within a tolerance level, and confirms, either by a confirmation gesture or button-press [30, 8, 3].

While in other studies the timer started after a loud beep [33] or a key press [7], we initiate timing of the first trial when the participant begins manipulating the input device. For subsequent trials, the timer is started as soon as the new target is displayed. Each trial needs to be completed within a given time limit or it is automatically skipped. In this case, a new trial is added to the sequence, ensuring that all participants complete an equal number of trials. The number of completed trials for the current device, as well as the time elapsed since the trial began, is displayed in the top-right corner of the screen. (Figure 2).

3.2 Visual Environment

The scene is rendered in stereo and viewed through NVidia 3D Vision RF shutter glasses, thereby providing the participants with stereoscopic depth cues. We used the default stereo settings of the NVidia drivers, because these were picked to be appropriate for a large variety of viewers at desktop viewing distances. We did not attempt to perform any individual calibration of stereo viewing parameters for each participant. Our objective was simply to attain a quality of depth perception commensurate with what one achieves with "out-of-the-box" commodity 3D hardware.

Although stereoscopic rendering is often associated with simulator sickness, this is not a problem for docking tasks because the scene is static and the user focuses on a single object [25]. To assist in visualization of the target orientation, a second camera window, shown at the bottom right of the screen, offers a view of the target from the right side (Figure 2).

Despite the use of a stereoscopic display, Wang et al. [30] raised the concern that depth discrimination may be affected by impoverished depth cues, thus increasing task complexity. To minimize the potential impact of this factor, we designed a more graphically rich virtual environment, in which depth cues are also conveyed by the textures of the floor and walls. Lighting effects and shadows cast by the chairs further improve 3D perception and aid positioning [13]. However, we did not evaluate the improvement in task performance resulting from these factors. Instead, the objective of our experiment was to evaluate human performance with different input devices on the docking task. Theoretically, users can perform such tasks even without the benefit of stereo rendering, as there are enough depth cues available in our virtual environment.

3.3 Accuracy Feedback and Error Measures

We use orientation and position errors to measure accuracy. The orientation error is the angle between the quaternions of the chairs and the position error is the Euclidean distance between them. Once the orientation and position errors of the moving chair are within a threshold, a confirmation message appears in the top left panel. If the participant confirms the position and orientation while the chair is docked within the tolerance level, a confirmation sound is played and the trial completes. The tolerance level for orientation and position was determined in a pilot study, described in the following section.

Similar to previous experimental docking studies [3, 7, 8, 33], we provide color feedback as a means of informing participants that they have docked the chair within the required tolerance and can complete the trial. However, we have two additional objectives. First, we wish to determine the limits of accuracy that participants can achieve with the interaction methods under evaluation. Second, we wish to explore the use of auditory feedback to avoid the problem of split visual attention between the docking task itself and visually verifying accuracy feedback.

To encourage participants to achieve the highest accuracy possible, we provide continuous visual and audio feedback regarding their progress in the docking task. Once the position is within tolerance, drums are heard as audio feedback, and the color of the cube shown in Figure 2 changes from yellow to green. The cube remains fixed in position above the target at all times. Similarly, once the orientation is within tolerance, a bass track is heard as audio feedback, and the color of the spheres also changes from yellow to green. Both the volume of the audio tracks and brightness of the visual cues increase as position and orientation improve further.

3.4 Apparatus

The experiment was conducted on a computer equipped with a Nvidia Quadro FX 3800 GPU that drove a 1920x1080 120 Hz 53 cm wide display, viewed by participants through NVidia 3D Vision RF shutter glasses. The software environment for the experiment was developed using the Unity3D game engine. Participants manipulated the cursor in mid-air using an "AirPen", a "MiniChair", or the participants' own hand and fingers, as described below. A fourth input device, the Phantom Omni, was employed as control condition. All four input techniques can be seen in Figure 3. For all mid-air conditions, retro-reflective markers were attached to the input device and hand, and tracked by an Optitrack Flex:V100 motion capture system.

Since latency is known to have a stronger effect than spatial jitter on docking task performance [28], another set of measurements was performed to determine whether end-to-end latency might be a factor in our experiment. For these measurements, the scene consisted of a gray circle, which the experimenter translated back and

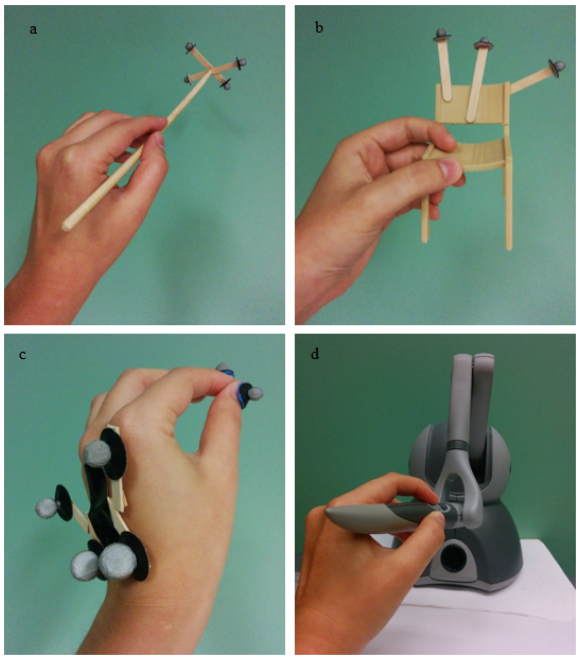

Figure 3: The input conditions used in the experiment were: (a) AirPen (b) MiniChair (c) Fingers and (d) Phantom Omni.

forth using the Phanton and the AirPen. The circle was overlaid on a 2D black-and-white checkerboard pattern, chosen to facilitate detection of movement by a high-speed black-and-white video camera, which captured the scene at 250 Hz. This procedure was repeated five times for each device and the recorded frames were reviewed to find the offset between movement of the physical device and the corresponding movement of the virtual device on screen. The results indicated a mean latency for the Phantom of 76.8 ms versus 72.0 ms for our Optitrack motion capture system.

We then sought to also confirm that the comparison of device performance was not affected by the sampling rate of the motion capture hardware or the Phantom. Towards this, we recorded the position and orientation reported through logging for each device over a 1 s interval, during which the experimenter used the device to translate and rotate the virtual chair. This procedure was repeated five times for each device. From inspection of the data, the sampling rate of the Phantom was determined to be approximately 73 Hz, versus 61 Hz for the devices tracked by the motion capture cameras.

In other words, both sampling rates were above 60 Hz, and the absolute difference between their mean latencies was approximately 5 ms. From these measurements, which are consistent with previous work [22], we are confident that neither sampling rate nor latency was substantially different between the Phantom and the other input conditions.

For all devices, translations and rotations are coupled, allowing both operations to be carried out simultaneously. This choice was preferred by all participants in a pilot, contradicting the findings of Martinet et al. [18]. To reduce shoulder fatigue, the width of the tracking volume was designed to reside between the hip and the shoulder of the participants. The need for arm extension and un-ergonomically large hand rotations was minimized through a clutch mechanism [11]. The participants sat approximately 75 cm from the screen and were allowed to rest their elbows on their lap or the armrests of the chair. All interaction involved indirect manipulation, which was found to be considerably faster than direct manipulation [20].

3.5 Input Mapping

To improve the input mapping, and in contrast with the previous work discussed in the Related Works section, we did not limit the locations at which the virtual object could be manipulated. Figure 4b shows how one can rotate the dark blue virtual chair around the Z axis of the AirPen (along the stick) in order to match the orientation of the target. The same mapping was used for all of the devices in our experiment, except for the rotation operation of the MiniChair, on which we elaborate in the following subsection.

The translation of each input device was applied to the virtual chair. Similarly, the virtual chair was rotated based on the change in Euler angles of the orientation of the input device. The virtual device had the same orientation as the real device except for its rotation around the Z axis (local frame), which was always set to 0^o, consistent with the assumptions of our pilot participants. The "up" vector of the virtual device (the Y direction) was transformed from local space to world space. The rotation operation of the virtual chair was performed around the previously obtained axis passing through its center. The same was done with the X and Z axes.

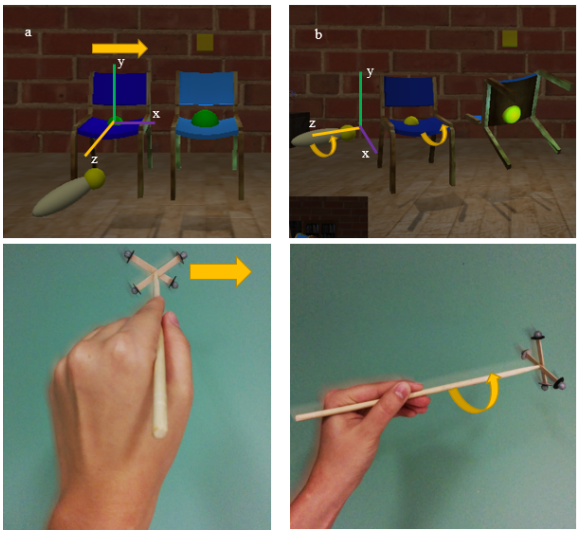

Figure 4: Mapping of the devices. The light blue target chair is under the floating cube. (a) Translating the dark blue virtual chair by dragging the AirPen to the right. (b) Rotating the dark blue virtual chair around the Z axis of the AirPen.

3.6 Input Conditions

3.6.1 AirPen

The AirPen (Figure 3a) was designed to be functionally similar to an unconstrained version of the Phantom stylus. It serves as an example of a familiar object that could plausibly be tracked as an input device by a virtual or augmented reality system, since the Leap Motion is already capable of tracking a stylus. The AirPen consists of a chopstick, to which a set of short sticks affixed with retro-reflective markers was attached perpendicularly to both track the third degree of rotation and to avoid occlusions.

While the AirPen is held in the dominant hand, the non-dominant hand is used for clutching and confirmation gestures. We use a fast tap of the index finger and the thumb of the non-dominant hand as confirmation gesture and a (longer) pinch for clutching. While

the user is holding the clutch gesture, movements of the AirPen are applied to the chair cursor, as in previous studies [17, 21, 31]. The confirmation gesture indicates completion of a trial. The pinch gesture is detected by observing the proximity of two spherical retro-reflective motion-capture markers, placed on the index finger and thumb using putty. The threshold distance for the "clutch" was established through calibration on a per-participant basis. A fast "tap" gesture, involving contact between the thumb and index finger of less than 0.3 s, is used to confirm the final docking position and orientation. The experimenter empirically determined the time threshold for the fast tap.

3.6.2 MiniChair

Inspired by Hinckley's passive real-world interface props [12], the MiniChair (Figure 3b) is a 3D printed chair, to which we attached sticks with retro-reflective markers. To avoid marker occlusions when the chair is upside down, we constrained the angle between the "up" vector of the target chair and the "down" vector of the virtual world to be greater than 80^o, a value found empirically to be sufficient. For consistency across conditions, this constraint was applied to all devices. Because of the one-to-one mapping between the orientation of the physical MiniChair and its virtual representation, clutching was unnecessary and inappropriate for performing rotations. In theory, this represents a docking time advantage for the MiniChair for rotation operations. As with the AirPen, clutching by pinching with the non-dominant hand affects translations of the virtual chair. A fast tap was again used for confirmation.

3.6.3 Fingers

The easiest input device for users to access is, of course, their own hands (Figure 3c). This is especially true in the mobile context, for which other input devices would need to be carried or worn. As with the stylus, tracking of hands and fingers is available through existing RGB and depth cameras, although doing so robustly often remains a challenge. To avoid this potential confound, retro-reflective motion capture markers, configured as a trackable object, are taped to the back of the dominant hand for our experiment, while single markers are placed on the index finger and thumb. The virtual chair is then manipulated only while the subject is pinching. Since no object needs to be grasped in this condition, a fast tap of the thumb and index finger of the dominant hand is used to confirm docking.

3.6.4 Phantom

The Phantom Omni (Figure 3d) is a mechanically tracked, constrained device designed for 6 DoF operations that has demonstrated its superior performance in previous studies [8, 30]. We used the light colored button on the Phantom for clutching and the dark button for confirmation.

4. EXPERIMENTS

Before turning to the main study itself, we first describe several preliminary experiments we conducted to establish docking thresholds and the trial time limit, as well as to validate the benefits of using an everyday, textured object as the docking cursor and target.

4.1 Tolerance Level and Time Limit

Prior to running the main experiment, we needed to determine an appropriate tolerance level for both position and orientation errors. This was established through a pilot with four unpaid university students, without giving the participants feedback regarding their accuracy.

The pilot began with practice trials, where each device was tested on a series of five targets. Presentation order of the four input devices tested was determined by a four-level Latin square. The first target was presented in a standard orientation, Figure 4a, and the following three targets were rotated 45^o around the Y, X and Z axes respectively. The final target was assigned a random orientation, subject to the constraints explained above in the "MiniChair" section.

After completing the practice trials, each participant performed eight trials with each of the four interactions for a total of 32 trials per participant. We arranged the mean orientation and position error by input condition, and chose the biggest errors. The largest errors in both position (1.5 cm) and orientation (15^o) were observed in the Fingers condition; these values were then used as the respective tolerances for the position and orientation errors for all conditions in the following experiments. Similarly, analysis of the logged docking times during the practice trials led us to select a time limit of 40 s for trials in the following experiments. This value was sufficient for completion of all trials apart from one outlier (40.1 s).

4.2 Wireframe Tetrahedron vs. Chair Pilot

Some docking tasks in previous work used wireframe tetrahedra as target and cursor [3, 33, 8, 7], with a uniform texture or a checkerboard pattern over the background [30, 2] similar to the environment in Figure 5. However, anecdotal reports suggest that the use of a everyday, more familiar, and less symmetrical object, such as a chair, could reduce the perceptual complexity of the docking task. Since the goal of a docking task experiment is to evaluate input methods and not the spatial intelligence of the participants, we conducted a pilot test with three unpaid participants to compare their performance using tetrahedra (Figure 5) and chairs (Figure 2). For the former condition, each edge of the tetrahedron was assigned a different color to avoid ambiguity in perception of orientation, and a checkerboard texture was used as a background.

We chose the AirPen device for this test, since the participants in our pilot studies preferred it. The pilot test consisted of 2 blocks $\times$ 6 trials $\times$ 2 docking environments for a total of 24 trials per participant. Before starting the trials, the participants completed four practice runs in each docking environment. The participants were instructed to be as accurate as possible within the time limit. The diameters of the bounding spheres of the virtual chair and tetrahedron were 9 cm and 10 cm, respectively.

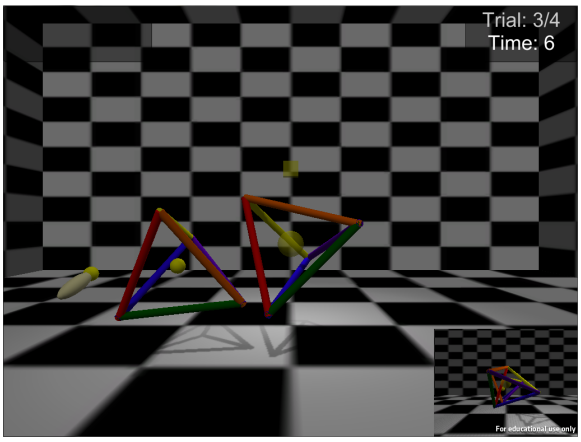

Figure 5: A screenshot from our pilot experiment of a typical docking task experiment using wireframe tetrahedra.

The results in Table 2 show similar accuracy in both environments. However, participants docked the chair noticeably faster than the tetrahedron, and reported greater difficulty docking the tetrahedron, consistent with our hypothesis.

Environment	Orientation error (degree)	Position error (cm)	Docking time (s)
Tetrahedron	9.29	0.71	15.19
Chair	9.30	0.66	11.30

Table 2: Average accuracy error and docking time for trials using tetrahedra and chairs.

4.3 Main Study

The main experiment consisted of 2 blocks $\times$ 6 trials $\times$ 4 input conditions for a total of 48 trials per participant. The order of the four input conditions tested was determined by Latin squares. A total of 12 participants took part in the experiment, ages ranging from 19 to 27 (median 22), drawn from a population of students. Half of the participants performed 3D virtual tasks at least two to five times per week and the other half less often. Participants began by completing a pre-test questionnaire, reading a document with instructions, and watching a short video explaining the visual and sound feedback provided in the docking task. They then carried out four practice trials before proceeding to the full experiment for each interaction. Following the experiment, participants completed a post-test questionnaire, and were compensated $10 for their time. We used the tolerance threshold found in our pilot study (position: 1.5 cm, orientation: 15^o) and limited the task time to 40 s. The participants were instructed to be as accurate as possible within the time limit.

4.3.1 Results

As the data was not normally distributed, we used ART [32] to conduct a non-parametric ANOVA for the docking time, position and orientation errors. All 19 skipped trials were discarded, and we analyzed only the 48×12 successful trials. The ANOVA test indicated that the interaction method used had a significant effect on the docking time ($F(3, 33) = 6.95$, $p < 0.05$, $GES = 0.09$), position error ($F(3, 33) = 4.21$, $p < 0.05$, $GES = 0.07$), and orientation error ($F(3, 33) = 3.36$, $p < 0.05$, $GES = 0.04$). Pairwise comparison using paired t-tests with a Bonferroni correction was then used to analyze individual effects within these measures.

For docking time, there was a significant difference between all the interaction methods except for the MiniChair-AirPen and Phantom-Fingers pairs. Figure 6 shows that all the tangible mid-air interactions were faster than with the Phantom, a constrained device. The slowest mid-air method, the fingers, was 0.29 s faster (1.37%), on average, than the Phantom. The fastest device, the MiniChair, was 4.79 s faster (23.09%) than the Phantom. Although the MiniChair had the smallest mean docking time, the difference between it and the next fastest device, the AirPen, was not significant.

The Phantom was the most accurate device, allowing participants to achieve the smallest position error among all input conditions tested (Figure 7). The difference was significant, according to the paired t-tests, although the value of this difference was small: the Phantom was 0.14 cm (26.50%) more accurate than the least accurate interaction for placement, the AirPen. Similarly, the orientation error achieved by participants with the Phantom was the smallest, as shown in Figure 8, which was again significantly different from all the mid-air interactions according to the t-tests. Even though the Phantom was 20.84% more accurate for rotation operations than the worst mid-air interaction, the fingers, the absolute difference in degrees was minor, at only 1.53^o. Thus, the Phantom was the most accurate device for both position and orientation, but not by a large margin. There was no significant difference in terms of either accuracy measures between the mid-air interaction conditions. A representative illustration of the average accuracy error of Fingers (position: 0.53 cm, orientation: 7.36^o), overall the least accurate interaction condition, is shown in Figure 9.

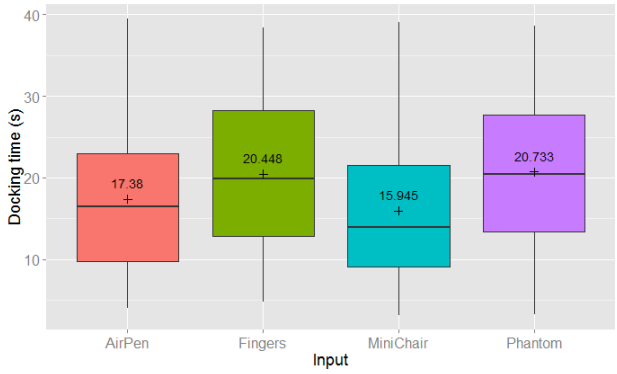

Figure 6: Boxplot of the docking task completion time for each interaction, where (+) is the mean docking time.

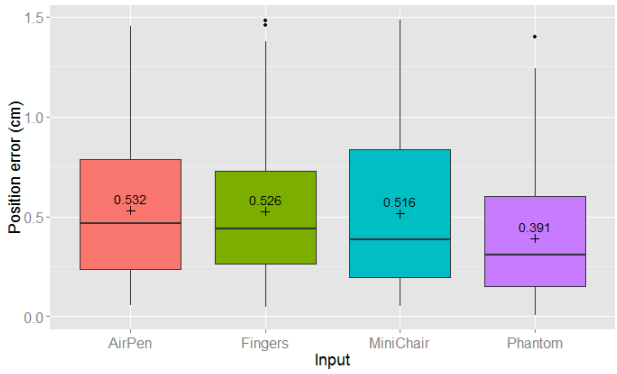

Figure 7: Boxplot of the position error for each interaction, where (+) is the mean position error.

On average, participants, applied transformations to the virtual chair (clutched) during 76% of the total time for each trial. An ANOVA test indicated that input condition had a significant effect on the clutching time ($F(3, 33) = 4.36$, $p < 0.05$, $GES = 0.07$). T-tests with a Bonferroni correction identify significance between all pairs of input conditions except for the fingers-AirPen and fingers-Phantom pairs. The average clutching time for the AirPen, fingers, MiniChair and Phantom were 13.62, 15.20, 12.24 and 15.61 seconds, respectively.

We also found that on average the chair cursor was rotated around its three axes almost equally, but participants preferred rotating the input device around its Z axis while applying rotations to the chair cursor. An ANOVA test indicated that the interaction between the input condition and the rotation axis had a significant effect on the number of rotations performed around each axis ($F(6, 66) = 6.75$,

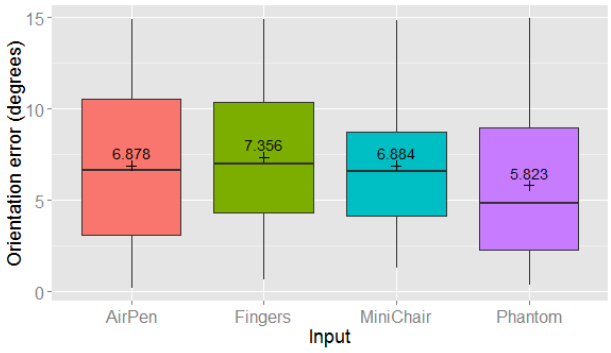

Figure 8: Boxplot of the orientation error for each interaction, where (+) is the mean orientation error.

Figure 9: Visual representation of the average accuracy error for the Fingers interaction, which was the least accurate, overall. The dark blue chair cursor had an orientation error of 7.36^o and a position error of 0.53 cm.

$p < 0.05$, $GES = 0.02$). After separating the data by input condition, the ANOVA tests found that the rotation axes had a significant effect on the number of rotations performed around each axis for the AirPen ($F(2, 22) = 36.56$, $p < 0.05$, $GES = 0.19$), MiniChair ($F(2, 22) = 7.00$, $p < 0.05$, $GES = 0.04$), Fingers ($F(2, 22) = 4.19$, $p < 0.05$, $GES = 0.02$) and Phantom ($F(2, 22) = 31.10$, $p < 0.05$, $GES = 0.16$). T-tests with a Bonferroni correction identify significance between the Z axis and the X and Y axes for all input conditions. Rotations were performed around the Z axis 42.8% of the time for the AirPen, 45.5% for the Phantom, and 37.5% for the Fingers and 38.1% for the MiniChair. The AirPen (X:29.8%, Y:27.3%) and the MiniChair (X:29.1%, Y:32.8%) also had significance between their X and Y axes.

The post-test questionnaire asked participants to rate how favorably they found each interaction, with '5' considered to be strongly favored and '1' strongly unfavorable. Participants also rated the level of fatigue they experienced in their wrist and shoulder for each interaction and were asked their opinion about the auditory and color feedback. Results of this questionnaire indicate that subjects preferred the AirPen and Fingers, while interaction with the MiniChair was the least favored (Figure 10). The level of fatigue reported by the participants was similar across devices. The participants gave the auditory feedback an average rating of 4.42 and the color feedback an average rating of 3.42 out of 5.

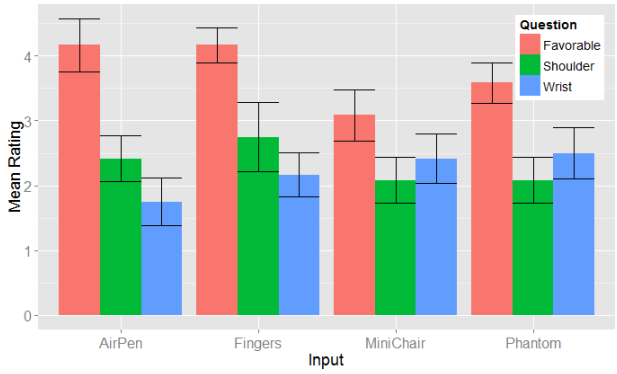

Figure 10: Participants' response to the post-test questionnaire grouped by favorable interaction, shoulder fatigue and wrist fatigue.

5. DISCUSSION

Overall, we found that the Phantom, a mechanically tracked and constrained device, was the most accurate device for position and orientation, whereas the tangible mid-air interactions (AirPen and MiniChair) were the fastest. This is consistent with previous research [33]. Interestingly, the Phantom interaction exhibited the highest completion time, and the highest clutching time, on average. These observations may be due to the physical limitations of the Phantom's joints, which constrain the possible movements of the stylus, thereby making it more complicated to perform the required manipulations. However, we also found that the tested mid-air conditions achieved an average accuracy that is close to that of the constrained device, which was the most accurate. Our results also highlight that time, orientation error, and position error are all important factors in evaluating docking tasks, since these measures offer insights into suitable applications for the device.

The results of our pilot study indicate that participants were able to dock an everyday virtual object, such as a chair, faster than the traditional wireframe tetrahedron, which has been used in the past for docking tasks. One possible explanation is that a more familiar and less symmetrical object is easier to perceive. As confirmed by the post-test questionnaire data for our main experiment, audio feedback offers the benefit that each musical instrument can provide information regarding a different variable. Even though visual feedback was provided by the color of two dedicated objects (square and sphere), participants preferred the audio feedback. This may have been due to the audio feedback not requiring split attention, or because it was more salient than the visual feedback.

We observed that some participants were more accurate than others, although at the cost of longer trial completion times. This speed-accuracy trade-off is known from Fitts' law research in human-computer interaction and has been observed in 3D selection tasks [29]. Bérard et al. [2] also found a trade-off between time and accuracy, further motivating the imposition of a time limit on trials. Such a limit should be determined through a pilot test, during which one can simultaneously determine an appropriate tolerance level. If the time limit is too high, some participants will become tired trying to achieve the maximum possible accuracy. If it is too low, some participants will not be able to complete the trials successfully.

Since the participants reported similar fatigue for the desktop device and the mid-air interactions, our experiment does not seem to suffer from the "gorilla arm problem". The likely reason is that users kept their movements between the hip and shoulders, as sug-

gested by Hincapie et al. [11], and manipulated the chair during 76% of the trial time, limiting arm extension with the clutching mechanism. For maximum flexibility, we deliberately enabled a larger working volume for the unconstrained interactions than that provided in the Phantom condition. We observed that many participants would initially perform large gestures to avoid clutching, but soon switched to small gestures after realizing that these are less fatiguing, much like what one can observe with typical mouse usage. After the practice trials, most participants used approximately the same volume for all interactions.

Our analysis did not indicate any user preference for rotating the virtual chair around its X, Y or Z axis. This might suggest that participants deliberately select different orientations of the input device around the volume in order to manipulate the virtual object more comfortably, a behavior enabled by our mapping. The target was always assigned a random orientation and the rotations applied to the chair cursor from its reference frame also seemed random. Yet, the log data indicates that participants applied the transformations with their input device in a non-uniform manner, preferring rotations of the AirPen and Phantom device around its Z axis, which they did for 42.8% and 45.5% of all rotations respectively. We believe this to be due to the fact that it is easier to roll the stylus around its longitudinal axis between the fingers, relative to other rotations, which involve moving the wrist.

The MiniChair was the fastest option, likely because it was a replica of the virtual target and did not require clutching for rotation operations. However, participants rated the MiniChair as the least favorable condition, which we speculate was due to its more complex shape, which made it difficult to manipulate. In fact, some participants used both hands to rotate the MiniChair, possibly due in part to their small hand size. While it is not practical to have a replica of every virtual object we want to manipulate, such replicas may still be convenient for some applications, such as action figures in an augmented reality game.

Based on our results, we believe that the AirPen can serve as a multi-purpose device due to its ergonomic shape, speed and the high acceptance from the participants. The Flystick [20] behaves in a similar manner, but is held with a power grip, which precludes rolling around its Z axis, a feature of the stylus preferred by our participants. The user's fingers are a convenient input condition, since there is no need for an extra device. However, this requires accurate and reliable finger tracking in the presence of potentially large hand rotations. While it would have been possible to use the same gestures for clutching across all conditions, the AirPen and MiniChair needed the second hand for clutching, while the Phantom and the fingers conditions were manipulated with the dominant hand. We acknowledge that this might have increased fatigue for the bi-manual conditions, but the participants reported similar levels of fatigue across all conditions.

As described earlier, we determined the maximum position and orientation errors acceptable for the docking task through a pilot experiment. Traditionally, the experimenter chooses such values empirically. Yet, in pilot tests we observed that if the threshold is too high, participants repeatedly make small adjustments until they receive feedback of being within the required tolerance. In that case, the results may be more a reflection of luck than the performance achievable with any given input device. Given these factors, we attempted to set a tolerance threshold that is sufficiently difficult to make the task challenging, but not so difficult that success becomes tedious and overly fatiguing.

6. CONCLUSION

We conducted a study to compare the completion time and accuracy achievable on a docking task, performed with a 6 DoF mechanically constrained desktop device, to three alternatives employing mid-air interactions. We found that the constrained desktop device achieved greater accuracy than mid-air unconstrained interactions, as expected. Interestingly, however, the performance difference was very small, and possibly overshadowed by the faster speed of the tangible mid-air interaction methods. Even though the fingers did not outperform the Phantom in accuracy or speed, the difference between these two conditions was small. Thus, fingers may serve as a reasonably accurate and efficient input method, especially for mobile environments. We also found that participants prefer performing rotations around the Z axis of a stylus, and preferred multi-modal audio feedback to visual feedback for accuracy.

Given these results, we believe that rich mid-air interaction with virtual 3D content is not only plausible, but also reasonably fast. Future work should address the challenge of accurately tracking input devices with RGB and depth cameras.

7. ACKNOWLEDGEMENTS

We thank Jeff Blum for valuable discussions, Shrey Gupta for help with statistics and Ziad Ewais for the 3D printed chair.

8. REFERENCES

[1] Balakrishnan, R., Baudel, T., Kurtenbach, G., and Fitzmaurice, G. The rockin'mouse: integral 3d manipulation on a plane. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, ACM (1997), 311–318.

[2] Bérard, F., Ip, J., Benovoy, M., El-Shimy, D., Blum, J. R., and Cooperstock, J. R. *Did Minority ReportÂİ get it wrong? Superiority of the mouse over 3D input devices in a 3D placement task*. Human-Computer Interaction-INTERACT 2009. Springer, 2009, 400–414.

[3] Boritz, J., and Booth, K. S. A study of interactive 6 dof docking in a computerised virtual environment. In *Virtual Reality Annual International Symposium, 1998. Proceedings., IEEE 1998*, IEEE (1998), 139–146.

[4] Bruder, G., Steinicke, F., and Sturzlinger, W. To touch or not to touch?: comparing 2d touch and 3d mid-air interaction on stereoscopic tabletop surfaces. In *Proceedings of the 1st symposium on Spatial user interaction*, ACM (2013), 9–16.

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

[6] de la Rivière, J.-B., Kervégant, C., Orvain, E., and Dittlo, N. Cubtile: a multi-touch cubic interface. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, ACM (2008), 69–72.

[7] Froehlich, B., Hochstrate, J., Skuk, V., and Huckauf, A. The globefish and the globemouse: two new six degree of freedom input devices for graphics applications. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, ACM (2006), 191–199.

[8] Glesser, D., Bérard, F., and Cooperstock, J. R. Overcoming limitations of the trackpad for 3d docking operations. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, ACM (2013), 1239–1244.

[9] Grossman, T., Wigdor, D., and Balakrishnan, R. Multi-finger gestural interaction with 3d volumetric displays. In

Proceedings of the 17th annual ACM symposium on User interface software and technology, ACM (2004), 61–70.

[10] Hancock, M., Ten Cate, T., and Carpendale, S. Sticky tools: full 6dof force-based interaction for multi-touch tables. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ACM (2009), 133–140.

[11] Hincapié-Ramos, J. D., Guo, X., Moghadasian, P., and Irani, P. Consumed endurance: A metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, ACM (2014), 1063–1072.

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

[13] Hubona, G. S., Shirah, G. W., and Jennings, D. K. The effects of cast shadows and stereopsis on performing computer-generated spatial tasks. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on 34*, 4 (2004), 483–493.

[14] Kratz, S., Rohs, M., Guse, D., Muller, J., Bailly, G., and Nischt, M. Palmspace: continuous around-device gestures vs. multitouch for 3d rotation tasks on mobile devices. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, ACM (2012), 181–188.

[15] Lévesque, J.-C., Laurendeau, D., and Mokhtari, M. An asymmetric bimanual gestural interface for immersive virtual environments. In *Virtual Augmented and Mixed Reality. Designing and Developing Augmented and Virtual Environments*. Springer, 2013, 192–201.

[16] Mapes, D. P., and Moshell, J. M. A two-handed interface for object manipulation in virtual environments. *Presence: Teleoperators and Virtual Environments 4*, 4 (1995), 403–416.

[17] Markussen, A., Jakobsen, M. R., and Hornbaek, K. Vulture: a mid-air word-gesture keyboard. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, ACM (2014), 1073–1082.

[18] Martinet, A., Casiez, G., and Grisoni, L. The effect of dof separation in 3d manipulation tasks with multi-touch displays. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*, ACM (2010), 111–118.

[19] Mendes, D., Fonseca, F., Araujo, B., Ferreira, A., and Jorge, J. Mid-air interactions above stereoscopic interactive tables. In *3D User Interfaces (3DUI), 2014 IEEE Symposium on*, IEEE (2014), 3–10.

[20] Moehring, M., and Froehlich, B. Effective manipulation of virtual objects within arm's reach. In *Virtual Reality Conference (VR), 2011 IEEE*, IEEE (2011), 131–138.

[21] Muller, J., Geier, M., Dicke, C., and Spors, S. The boomroom: mid-air direct interaction with virtual sound sources. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, ACM (2014), 247–256.

[22] Pavlovych, A., and Stuerzlinger, W. The tradeoff between spatial jitter and latency in pointing tasks. In *Proceedings of the 1st ACM SIGCHI symposium on Engineering interactive computing systems*, ACM (2009), 187–196.

[23] Poupyrev, I., Billinghurst, M., Weghorst, S., and Ichikawa, T. The go-go interaction technique: non-linear mapping for direct manipulation in vr. In *Proceedings of the 9th annual ACM symposium on User interface software and technology*, ACM (1996), 79–80.

[24] Reisman, J. L., Davidson, P. L., and Han, J. Y. A screen-space formulation for 2d and 3d direct manipulation. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, ACM (2009), 69–78.

[25] Schild, J., Jr, J. J. L., and Masuch, M. Altering gameplay behavior using stereoscopic 3d vision-based video game design. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, ACM (2014), 207–216.

[26] Shoemake, K. Arcball: a user interface for specifying three-dimensional orientation using a mouse. In *Graphics Interface*, vol. 92 (1992), 151–156.

[27] Song, P., Goh, W. B., Hutama, W., Fu, C.-W., and Liu, X. A handle bar metaphor for virtual object manipulation with mid-air interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2012), 1297–1306.

[28] Teather, R. J., Pavlovych, A., Stuerzlinger, W., and MacKenzie, I. S. Effects of tracking technology, latency, and spatial jitter on object movement. In *3D User Interfaces, 2009. 3DUI 2009. IEEE Symposium on*, IEEE (2009), 43–50.

[29] Teather, R. J., and Stuerzlinger, W. Pointing at 3d targets in a stereo head-tracked virtual environment. In *3D User Interfaces (3DUI), 2011 IEEE Symposium on*, IEEE (2011), 87–94.

[30] Wang, G., McGuffin, M. J., BÃČÂľrard, F., and Cooperstock, J. R. Pop-up depth views for improving 3d target acquisition. In *Proceedings of Graphics Interface 2011*, Canadian Human-Computer Communications Society (2011), 41–48.

[31] Wang, R., Paris, S., and Popović, J. 6d hands: markerless hand-tracking for computer aided design. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, ACM (2011), 549–558.

[32] Wobbrock, J. O. Practical statistics for human-computer interaction. In *Annual Workshop of the HCI Consortium*, HCIC (2011).

[33] Zhai, S., and Milgram, P. Quantifying coordination in multiple dof movement and its application to evaluating 6 dof input devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press/Addison-Wesley Publishing Co. (1998), 320–327.