Efficient and Accurate Performance with Unconstrained Mid-air Interaction

Vanessa Vuibert

Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

August 2015

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Master of Engineering.

2015/08/14

Abstract[1]

Mid-air interaction has the potential to manipulate objects in 3D with more natural input mappings. We compared the performance attainable using various unconstrained 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 unconstrained 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.

In order to compare constrained and unconstrained mid-air interactions, we evaluated the typical design of the docking task used for this purpose and improved it through several iterations. In our first design iteration, we used difficulty levels that were set by the experimenter and assumed that the device with the smallest docking time at the most difficult level was the most accurate. In our next design, we compared the fastest device to the most accurate device at the most difficult level and found that they were not the same, even though the difficulty levels were set through a pilot study. These results led us to use only one tolerance level and to evaluate the input conditions by their docking time, and position and orientation errors.

[1]This thesis is an expanded version of a publication by the author [1]

Résumé[2]

L'interaction aérienne permet de manipuler des objets en 3D de façon plus naturelle. Dans un exercice de positionnement des objets présentant 6 degrés de liberté, nous avons comparé la performance, aussi bien en terme de précision que de temps d'exécution, en utilisant différentes méthodes d'interaction aérienne sans contrainte et avec contrainte. Nous avons constaté que les dispositifs d'entrée aériens sans contrainte favorisent les performances de positionnement, tout en présentant une précision proche de celle des dispositifs avec contrainte. L'interaction aérienne à mains nues atteint des performances de temps et de précision similaires par rapport à l'appareil avec contrainte.

Afin de comparer les interactions aériennes avec contrainte et sans contrainte, nous avons réévalué la conception spécifique de l'exercice de positionnement et amélioré celui-ci. Lors de notre premiere analyse, nous avons utilisé des niveaux de difficulté précédemment établis par l'expérimentateur et présumé que le dispositif avec le temps d'accueil le plus court au niveau le plus difficile, a été le plus précis. Nous avons ensuite comparé le dispositif le plus rapide à l'appareil le plus précis au niveau le plus difficile et avons constaté qu'ils diffèrent, bien que les niveaux de difficulté aient été créés au travers d'une étude pilote. Ces résultats nous ont amené à utiliser un seul niveau de tolérance et à évaluer les conditions d'entrée selon leur temps d'accueil, et leurs erreurs de position et d'orientation.

[2]Cette thse est une version tendue d'une publication par l'auteur [1]

Acknowledgments

The author would like to thank Wolfgang Stuerzlinger and Jeff Blum for valuable discussions, Shrey Gupta for help with statistics, Ziad Ewais for the 3D printed chair and Maéva Langouët and Céline Bensoussan for the French abstract. The author would also like to thank her supervisor, Jeremy Cooperstock. This research was funded by Graphics, Animation and New Media (GRAND) NCE.

Contents

List of Figures

List of Tables

List of Acronyms

6 DoF	Six degrees of freedom
3D	Three dimensions
IR	Infra red
RF	Radio frequency
ART	Aligned rank transform
ANOVA	Analysis of variance
GES	Generalized Eta-Squared measure of effect size
GPU	Graphics processing unit
RGB	Red green blue

Chapter 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 a special-purpose constrained desktop device such as the Phantom Omni?

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

[1] www.atheerlabs.com

[2] www.spaceglasses.com

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 from real world object manipulation. 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. The feedback we provided on accuracy also evolved as we improved our evaluation methods and observed participants perform the task. Table 1.1 gives an overview of the experiments we performed to improve the design of the docking task experiment.

Chapter	Tracking device	Evaluation Method	Accuracy Feedback
3	Leap Motion	docking time with difficulty levels	binary visual feedback and continuous sound feedback on orientation
4	OptiTrack Motion Capture System	docking time with difficulty levels	added binary feedback on position
5	OptiTrack Motion Capture System	docking time and accuracy error	continuous visual feedback and multi-modal sound feedback

Table 1.1: Overview of the evaluation methods used and accuracy feedback provided.

In our first experiment, described in Chapter 3, we use the Leap Motion to compare unconstrained mid-air interactions to the Phantom Omni, which is a mechanically constrained desktop device designed for 6 DoF operations, and has demonstrated its superior performance in previous docking tasks studies [5, 6]. Since the ability of the Leap Motion to track fingers was not sufficiently robust for this purpose, we used motion capture technology for tracking in the subsequent experiments.

This reduces the possibility that the results are affected by limitations in tracking robustness. In our second experiment (Chapter 4), we evaluated the traditional way of comparing input conditions in a docking task. We found that it might be better to use the final position and orientation errors to measure accuracy instead of using difficulty levels [7].

Our improved design of the docking task experiment is described in Chapter 5. Our final experiment shows that tangible unconstrained mid-air interactions are faster and achieve similar accuracy to a desktop constrained device. Unconstrained bare hands interactions achieved similar performance and accuracy.

Chapter 2

Related Work

First, we explore different methods to evaluate input conditions used for 3D manipulation. Next, we look at different visualization techniques that allow for 3D perception. Then, we analyze different gestures from previous studies that allow for rotations and translations in 3D.

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 [8], Rockin'Mouse [9], GlobeFish and GlobeMouse [10], multi-touch surfaces used in conjunction with indirect [11] and direct interaction techniques, e.g., DS3 and StickyTools [11, 12, 13, 14], and mid-air interaction techniques such as Go-Go [15], as summarized in Table 2.1. Very few experiments compared constrained desktop-based devices to unconstrained devices that can be manipulated freely in mid-air. A notable exception is early research by Zhai and Milgram [16], which demonstrated that for a docking task, isomorphic manipulation through a 6 DoF unconstrained 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 constrained vs. unconstrained 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

Study	Task	Fastest Technique	Other Compared Techniques
Zhai et al. [16]	docking	mid-air	constrained device
Froehlich et al. [10]	docking	GlobeFish & GlobeMouse	mouse
Berard et al. [7]	placement	mouse	DepthSlider, SpaceNavigator, mid-air
Wang et al. [6]	placement	Phantom	mouse
Kratz et al. [17]	orientation	mid-air	multi-touch screen
Glesser et al. [5]	docking	Phantom, dual multi-touch surfaces	trackpad, mouse

Table 2.1: Past Research on 3D manipulation tasks.

al. [7] compared various devices for a 3D placement 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 [6], consistent with results from a more recent study [5].

Most placement, orientation and docking experiments only measure the time it takes participants to dock the cursor [10, 17, 5], but accuracy is often equally important. A docking task involves gross motion and then fine-tuning once near the target. Zhai et al. [16] 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 since the path taken might be influenced by spatial intelligence of the participant.

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 [3]. 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 [18]. However, they did not improve rotation tasks [19]. In a stereoscopic rendering condition, direct mid-air interactions outperformed multi-touch screen techniques when the target was further away from the screen [20]. 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 [21].

2.3 Gestures

The choice of interaction gestures is a critical factor in usability and performance. Martinet et al. investigated multi-touch screen manipulation techniques that coupled translations and rotations, comparing StickyTools [13] to DS3, an uncoupled technique, and found that the latter significantly improves performance [14]. Kin et al. employed bi-manual operations, where one hand is used for selection and the other for virtual Arcball (2D rotations) [2]. To perform rotations around the third dimension, the second index finger is dragged left or right (Figure 2.1). Similarly, Grossman et al. employ a rotation technique that involves touching the volumetric display with the index finger of the dominant hand and dragging the other index finger in the desired direction of rotation [3] (Figure 2.2). Rotation of the whole hand around the index finger is used to achieve object rotation around the z-axis. Yet another technique enables the user to tap on a point on an object and then tap on another surface to specify where the object should be docked [4] (Figure 2.3). The technique uses collision detection to avoid undesirable object placements. Ideally, only a single degree of freedom has to be adjusted after this form of docking. Cohé et al. performed experiments using a multi-touch screen to compare 3D interaction gestures for translations and rotations [22]. They suggest a grab gesture for such operations, which involves picking a point on a face of the object, and moving the point along a trajectory while the 3D point stays between the finger tips, similar to the Arcball operation.

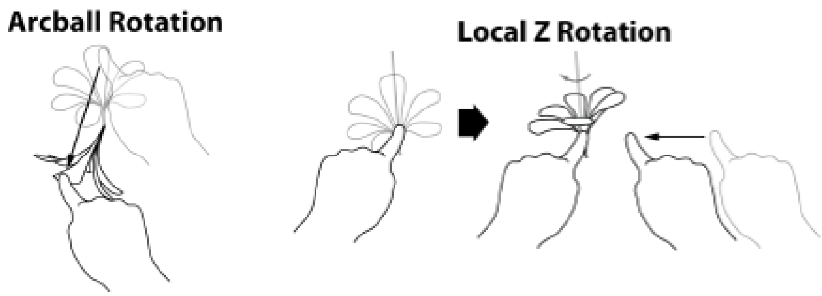


Fig. 2.1: Arcball rotation: the object is touched followed by a drag. Local z rotation: the object is touched followed by a second touch. Then, dragged left or right [2]. ©2011 ACM, Inc. Reproduced with permission of publisher.

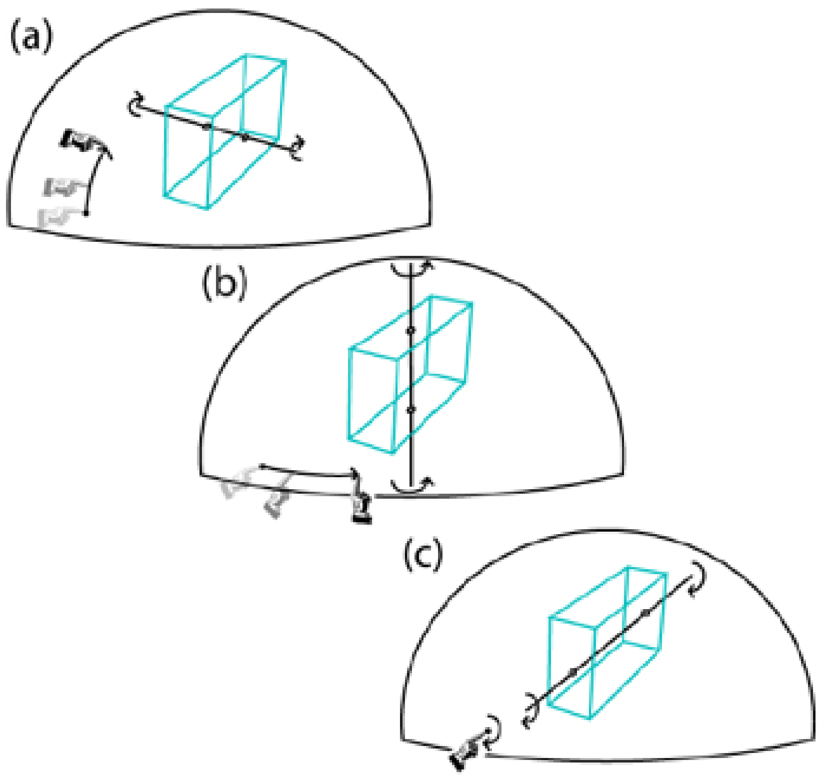


Fig. 2.2: (a) and (b): dragging the index finger in the desired direction of the rotation. (c): twisting the hand around the rotation axis [3]. ©2004 ACM, Inc. Reproduced with permission of publisher.

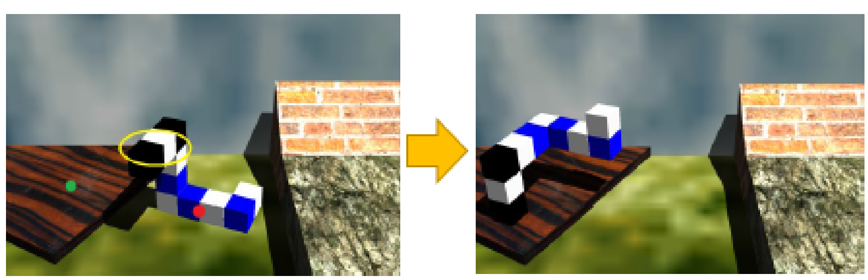


Fig. 2.3: The user first taps at the location of the red dot and then at the green dot. The right image is the result of the operation [4]. ©2013 Scheurich — ACM. Reproduced with permission of author.

Previous studies [23, 24] 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 [25].

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 [26]. 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 [27], similar to the 6 DoF Hand technique described by Mendes et al. [28]. 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 2.4), the object would rotate around the Z world axis instead of the X axis.

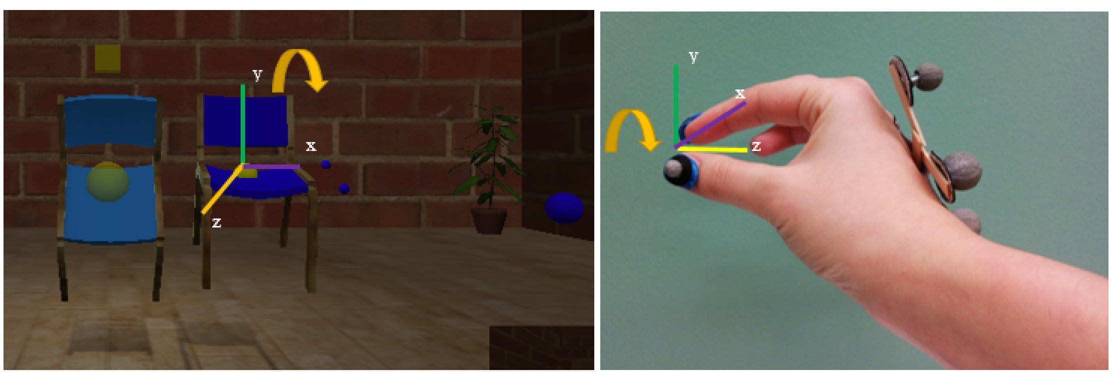


Fig. 2.4: Rotating the dark blue chair around its X axis with the hand.

Most 3D docking tasks used in past research did not evaluate the accuracy of input conditions or only used the accuracy of the trajectory of the object. The mapping described in Section 2.3 is unnatural and does not take advantage of the whole working volume, and thus, does not allow users to manipulate the virtual object from any point in space. We believe that a better approach might be to measure the accuracy of the final position and orientation of the object.

Chapter 3

Initial Docking Task Pilot Study

We were enthusiastic about the possibility of using the low-cost, portable, Leap Motion device as an effective input device for finger-tracking to perform complex 6 DoF manipulation tasks. We therefore designed an initial pilot study to evaluate the feasibility of this approach using a docking task. As a baseline we chose a mechanically constrained desktop device, the Phantom Omni, as it demonstrated its time performance superiority over other desktop devices in a recent docking study [5]. For our study, 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 3.1). Most docking task experiments evaluate devices by comparing their docking time, and use difficulty levels [6, 7] in order to measure accuracy. In this pilot study, we follow the docking task design from previous research where the most accurate device is considered to be the one with the smallest docking time at the most difficult level. The main focus of this chapter is to improve the experimental design of the common docking task and find if it is feasible to use the Leap Motion device for complex mid-air interactions.

3.1 Experimental Task

Previous docking studies used a set of predetermined target positions and orientations to avoid visual ambiguities [16, 5]. 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,

Fig. 3.1: The docking task environment. The virtual representation of the device is visible in the image as an elongated ellipsoid. The second camera on the bottom left offers six different views of the target.

and 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 by a button-press [6, 5, 19]. The participants were instructed to be as fast and accurate as possible [16, 7, 19].

While in other studies the timer started after a loud beep [16] or a key press [10], we initiate timing of the first trial when the participant begins manipulating the input device in order to log the docking time more accurately. For subsequent trials, the timer is started as soon as the new target is displayed. 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 3.1). The previous docking time is also displayed in order to encourage participants to improve their time. The current gesture that is being performed and the difficulty level of the current task (easy, medium, or hard) are displayed at the top-left corner of the screen. Sometimes the target would have a orientation that was difficult to visualize, so we permitted participants to cancel such trials by pressing the 'S' key [29]. This added another trial to the sequence so that all participants completed an equal number of trials.

3.2 Virtual Environment

The scene is rendered in stereo and viewed through NVidia 3D Vision IR 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.

Despite the use of a stereoscopic display, Wang et al. [6] 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 [18]. 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.

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 [21]. To aid in visualizing the target, a second camera shown at the bottom left of the screen offers a choice between six views (top, bottom, left, right, front and back) depending on the orientation of the device (Figure 3.1). For example, if the device is pointing to the left, the second camera will show a view from the left of the room. Initially the second camera showed a dynamic view from the perspective of the device looking at the target, but pilot participants stated that they preferred a more static view since it was difficult for them to orient themselves. Based on this feedback, the current design has several objects positioned around the virtual room to help participants orient themselves while using the second camera.

3.3 Accuracy Feedback

Feedback on accuracy is usually not provided in real-world applications. Yet, such feedback encourages participants to improve their accuracy allowing them to reach the tolerance threshold faster, reducing the docking time. Some 6 DoF docking task experiments use tetrahedra and give color feedback on accuracy if the distance between the vertices of the cursor and target is within a threshold [16, 5, 10]. However, if a participant matches only one vertex, the cursor could still be far from docked.

We believe that a better approach is to increase the volume of background music as the orientation error decreases. The orientation error is the angle between the quaternions of the chairs and the position error is the Euclidean distance between them. This idea was inspired by the game "Zelda Ocarina of Time", where the main character needs to find his way in the woods and the volume of the background music increases as he follows the correct path. For our study, we used the same background music that was used in the aforementioned game.

To give feedback on the combination of the position and orientation errors, the lighting in the scene becomes brighter once the moving chair is within the tolerance level, indicating to participants that they are able to confirm the docking. 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.

3.4 Practice Trials

The participant has to perform five docking tasks at an easy level of difficulty to complete the practice trials. The first practice trial only involves translation and the following trials involve translation and rotation. The first cursor chair looks like the one in Figure 3.2a, and the following three cursors are rotated 45^o around the Y (Figure 3.2b), X and Z axes respectively. Finally, a regular cursor chair, similar to the ones used in the real study, is presented with a random orientation and position, within a predetermined distance range from the target.

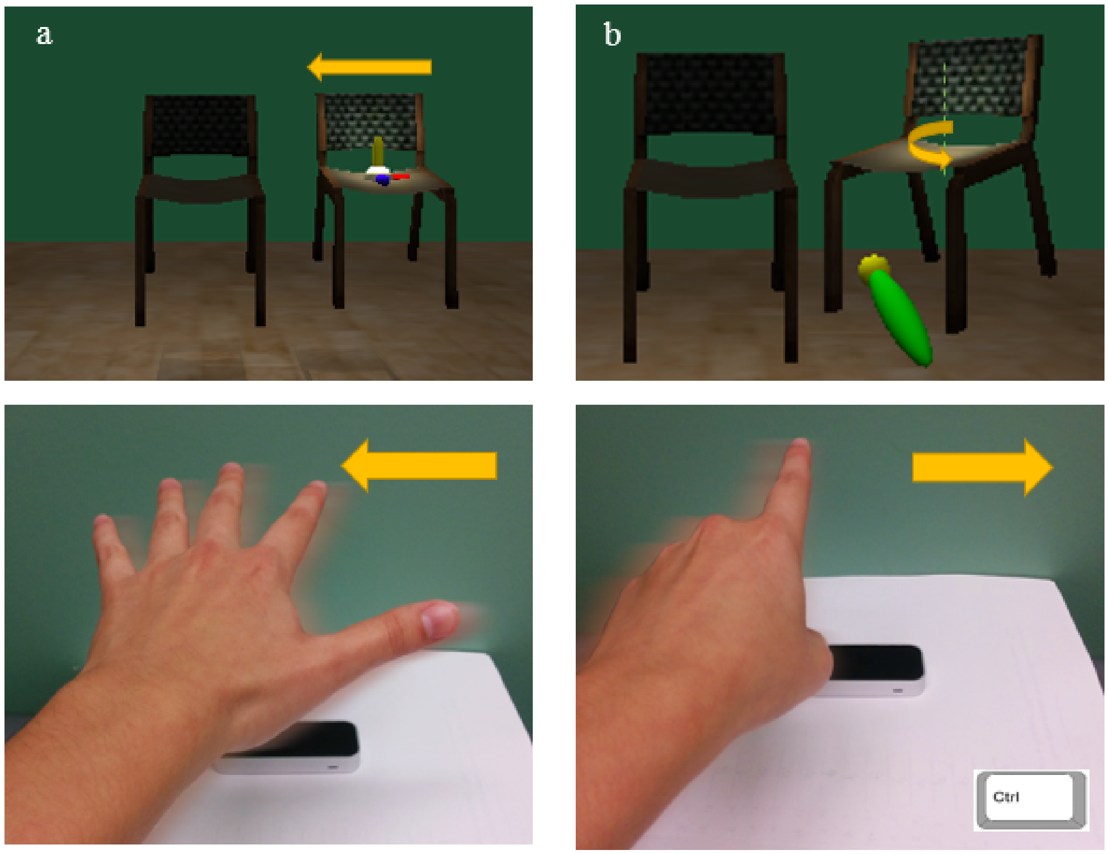


Fig. 3.2: Leap Sphere interaction. The yellow arrow shows the direction of movement. (a) Translation with an open hand. (b) Rotation of the virtual chair around its Y axis while the CTRL key is pressed.

3.5 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 IR shutter glasses. The software environment for the experiment was developed using the Unity3D game engine. Participants manipulated the cursor chair in mid-air using the Leap Motion and the Phantom Omni.

3.6 Input Conditions

To manipulate the cursor chair using the Leap, we developed two sets of gestures for decoupled interaction, separating translations and rotations. This approach was motivated by Martinet et al. [14], who found decoupled interaction to offer improved performance. For maximum robustness, the gestures had to be simple to form and easy for the Leap software to differentiate. In this respect, the Leap was limited in its ability to differentiate certain hand configurations and the orientation of the palm or fingers, and was not able to track a finger when it was pointing directly up or down. For the Phantom, translation and rotation operations were coupled. For the Leap Sphere, described below, and Phantom interactions, the stylus and index finger were represented in the virtual world as an elongated green ellipsoid (Figure 3.2b). The tip of the ellipsoid is green while clutching, and yellow otherwise.

3.6.1 Leap Sphere

For the first set of interaction gestures, participants translate the chair by opening their dominant hand and moving it accordingly in the Leap Motion's tracking volume, as depicted in Figure 3.2a. As feedback, three small axes are rendered at the center of the moving chair when the translation gesture is being performed. A fist gesture is used as a clutching mechanism and to confirm that the chair has been docked.

The cursor chair can be rotated around an axis perpendicular to the direction of movement of the index finger, as shown in Figure 3.2b. Our design of the rotation gesture initially used both index fingers, similar to the gestures described in Section 2.3. The right index finger would point at the cursor chair and the translation of the left index finger would rotate the chair. The rotation axis was the cross product of the pointing vector of the right finger and the translation vector of the left finger (Figure 3.3b), allowing use of the same gesture to perform rotations in the whole volume unlike previous research [2, 3].

However, due to parasitic motion, this approach was found to be insufficiently stable while transitioning between gestures. This prompted us to make use of the CTRL key as a clutching mechanism for the rotation operation, which was reinforced with several visual indicators. A fire trails the path of the virtual finger and a green sphere is rendered on the moving chair when the rotation is being performed.

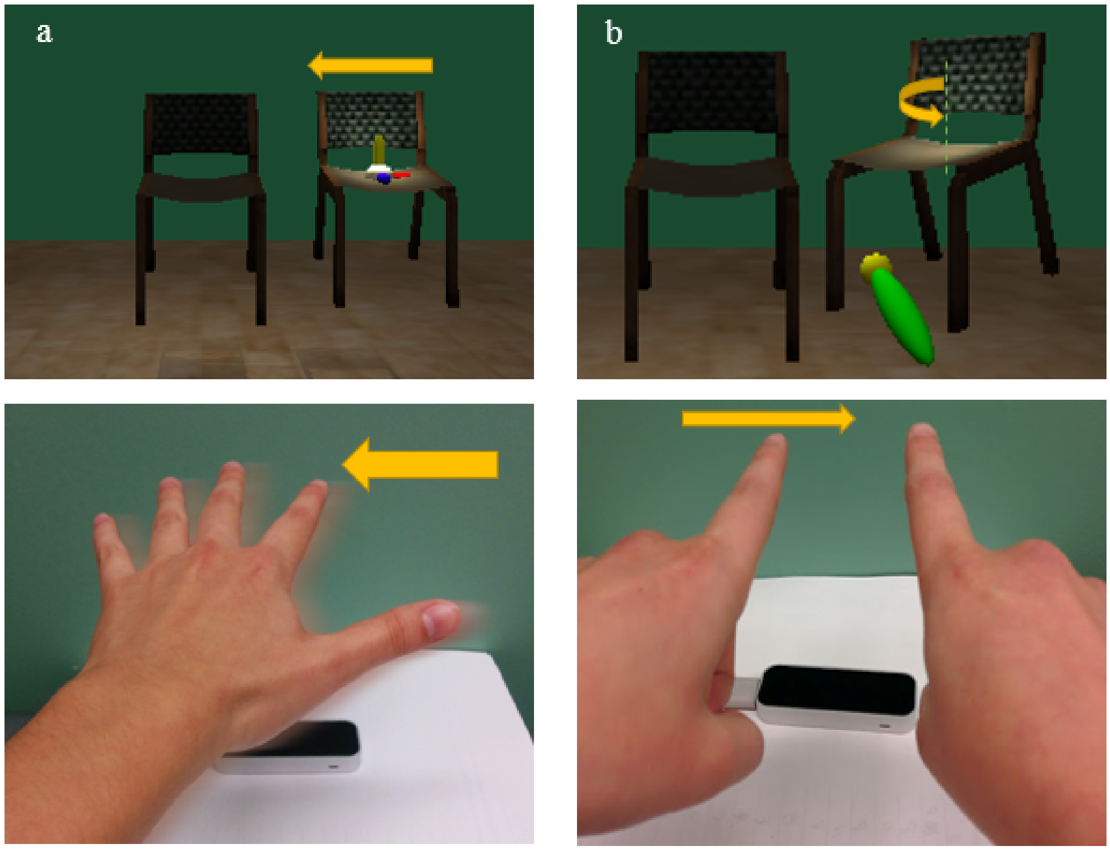


Fig. 3.3: Leap Sphere interaction without the keyboard. (a) Translation with an open hand. (b) Rotation of the virtual chair around its Y axis using the index fingers.

3.6.2 Leap Arcball

For the second set of interaction gestures with the Leap (Figure 3.4), participants translate the chair by closing their fists to make a grabbing gesture. Rotation was achieved following Shoemake's Arcball [8], using a pinching gesture as a clutching mechanism, a gesture commonly used for mid-air selections [30, 31, 32].

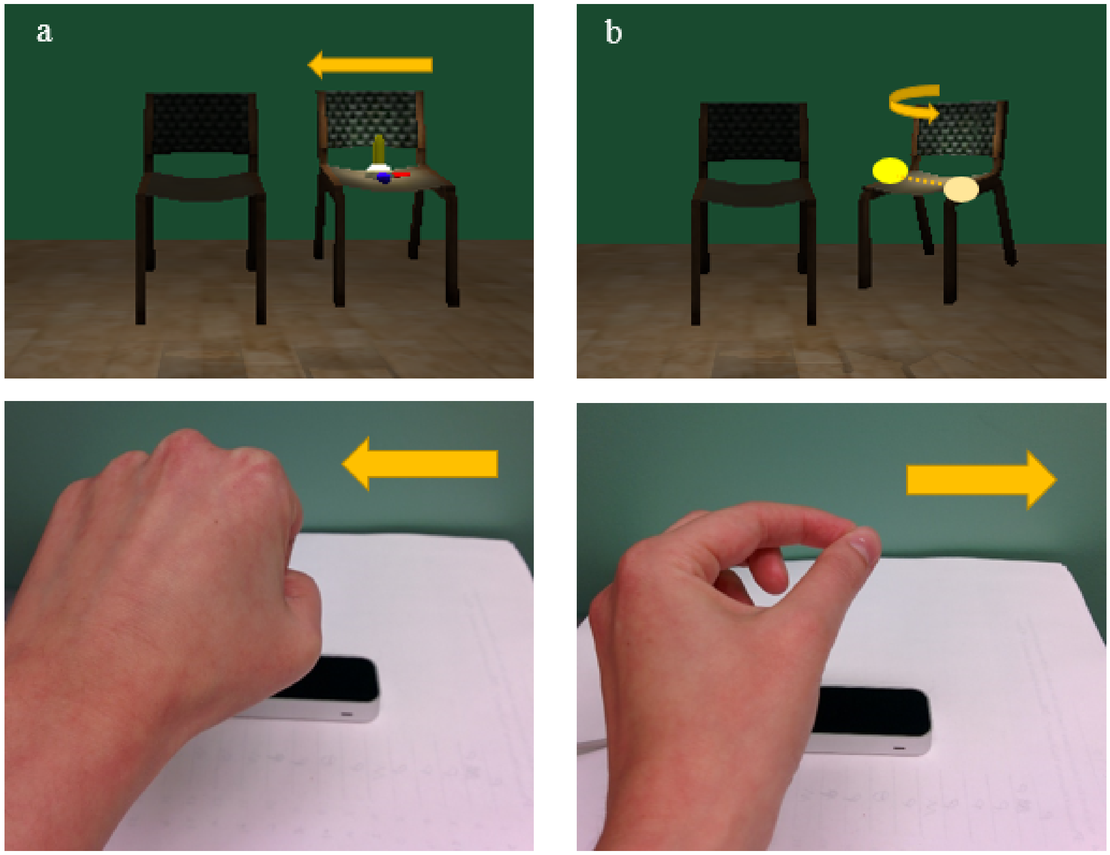


Fig. 3.4: Leap Arcball interaction. (a) Translation with fist. (b) Rotation of the virtual chair around its Y axis by pinching.

3.6.3 Phantom

Translations and rotations performed with the Phantom (Figure 3.5) were applied to the moving chair while pressing the light colored button on the stylus. Participants were able to confirm that the chair was docked by pressing the darker colored button.

Fig. 3.5: Phantom Omni.

3.7 Pilot Study

The pilot study consisted of 6 trials × 3 difficulty levels × 2 input mechanisms for a total of 36 trials per participant and was performed by three unpaid university students between the ages of 22 and 25. In this study the Leap Sphere interaction was compared to the Phantom. The Leap Arcball was not tested because the Leap Motion was not able to track pinching gestures sufficiently reliably, which led to frustration among the pilot test participants. The difficulty levels for the study, shown in Table 3.1, were determined empirically by the experimenter using the Leap Sphere interaction and were used for all the trials with the Phantom and Leap Motion.

3.8 Results and Discussion

The boxplots in Figure 3.6 show that the Phantom was substantially faster than the Leap Sphere for the hard level, but the difference between the interactions was not as big for the easy and medium levels. The participants reported that they preferred the Phantom

Difficulty	Orientation error (degrees)	Position error (cm)
Easy	20.0	3.0
Medium	12.5	2.0
Hard	5.0	1.0

Table 3.1: Difficulty levels for the Leap Sphere study. The level of difficulty of the trial increases as the maximum allowed orientation and position error decrease.

interaction, which makes sense since it allowed for shorter docking times for all difficulty levels.

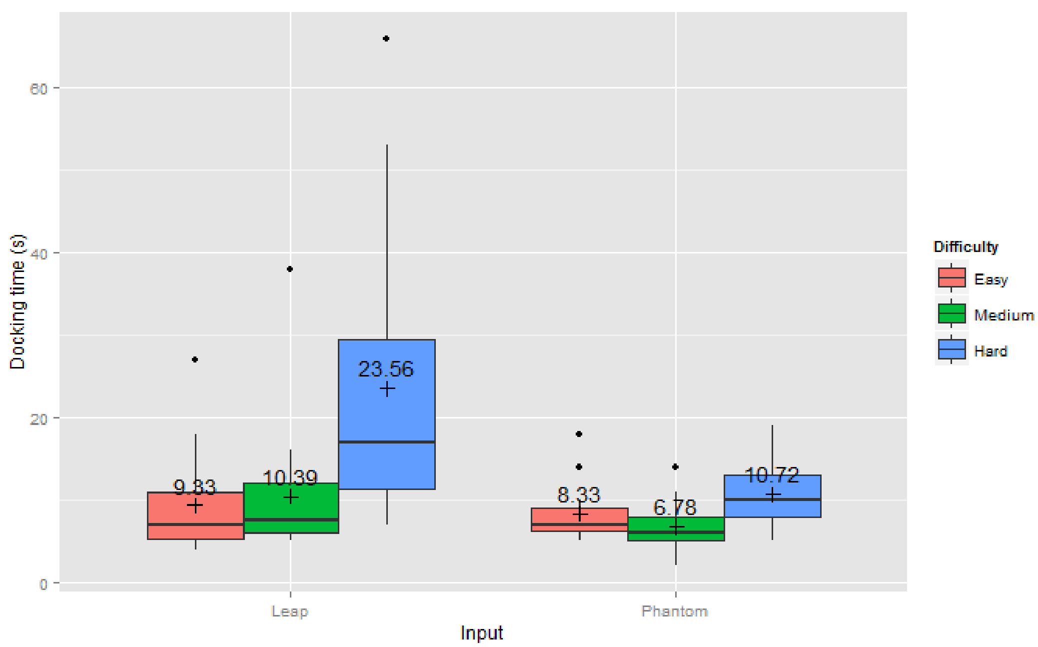


Fig. 3.6: Docking task completion time for Leap Sphere and Phantom grouped by difficulty level, where (+) is the mean docking time.

Following the approach of most previous docking tasks, we only evaluated the time it took to dock the chair. In this design, the most accurate device was assumed to be the one with the lowest docking time at the most difficult level, but we did not log the final orientation and position errors in order to verify if this is true. We suspect that the Phantom had a lower orientation and position error than the Leap Motion, but if true, this difference might be due to detection issues and not because of any inherently superior accuracy of constrained devices vs. unconstrained mid-air interactions. Thus, to ensure a fair comparison between constrained and unconstrained mid-air interaction conditions, we were motivated to adopt a motion capture system due to its more robust tracking technology.

The participants received audio feedback on the orientation error, and additional feedback (room lights turning on) once both the position and orientation were within the required threshold. As a result, participants would sometimes not realize at first that the position of the cursor chair was outside of the required tolerance, due to the difficulty of depth visualization on a 2D display, even with stereo rendering. Thus, we hypothesized that docking time would decrease, on average, if we provided separate feedback on position error. We also suspected that having a texture on the walls of the environment (Figure 3.1) and thicker chairs would improve depth perception, potentially reducing docking time. Participants reported that the sound feedback made the task more enjoyable, but they could not easily hear the change in volume. We could consider creating a more abrupt change in volume, thus making it more salient, by playing the background music only once the cursor chair is properly aligned.

The mapping of the Phantom assumed that participants would always manipulate the chair from the front, similar to the mapping used by Glesser et al. [5]. This turned out to be a problem since we observed multiple participants unsuccessfully attempting to use the Phantom from different sides, possibly because this is how objects are manipulated in the real world. In all following experiments, we implemented a mapping that allowed for manipulations from any side for all devices, including the Phantom.

We observed that changing the side used to manipulate the chair caused some pilot participants to lose their sense of orientation, so we implemented a dynamic second view of the target, that depended on the orientation of the device. Even though participants stated that they found the second camera useful, some reported that they would also like to have the ability to freeze this view for more control. This feature might improve visualization while making fine adjustments to the cursor chair.

In this study, we displayed the docking time of the previous trial to encourage participants to improve their time. However, since successive trials varied in difficulty level, comparisons of docking times between trials would not be meaningful, and could possibly cause frustration. Some participants also reported that they did not look at the timer, which makes sense since the trials did not have a time limit. We assumed that participants would skip targets that were difficult to visualize, but some participants did not, so they were more fatigued than the others. To avoid potential issues from varying fatigue levels across participants, a time limit should be implemented so that the overall experiment takes approximately constant time.

3.9 Conclusion

The analysis above suggests several improvements that should be made to the experimental design. So far, we considered the most accurate device to be the one that supports the fastest performance on the most difficult level. Even though the docking time is a useful variable to compare input devices, we suspect that also evaluating the final orientation and position errors will give us a better comparison.

However, as in previous similar studies, the difficulty levels were set by the experimenter. Since the experimenter is an experienced user, the resulting difficulty levels might be too difficult for the average or novice user. Improvements in the visual and audio feedback may also reduce the cognitive load of docking the chair, which is not a trivial task. In the study described in the present chapter, the Phantom was the fastest device and preferred by participants. However, the comparison may not have been fair, because the Leap Motion was not able to reliably track the gestures performed by the participants.

For our next experimental design, we believed that a better approach would be to determine the difficulty levels through a small study. Setting a time limit should stop participants from trying to solve docking tasks that are more difficult to visualize than the others, reducing fatigue. We also abandoned the Leap Motion as a tracking device and turned, instead, to a more reliable optical-marker-based tracking system for the remaining experiments involving mid-air interaction.

Chapter 4

Accuracy Measured with Difficulty Levels

In the previous chapter we designed a pilot study to compare unconstrained to constrained devices and evaluated them by the time it took participants to dock the chair. Based on previous research, we assumed that the fastest device at the most difficult level was the most accurate. However, we suspect that this might not be the best approach, so for this experiment we will also log the final orientation and position errors to compare the results.

The Leap Motion was not robust enough for our study, so we decided to use motion capture cameras instead. In this study, a docking task is used to compare the performance of three unconstrained mid-air interaction options and a mechanically constrained device, the Phantom Omni. Several changes were made to our first design of the docking task experiment based on our observations and results from the study described in Chapter 3. The first three sections of this chapter describe the changes made to the experimental task, virtual environment and accuracy feedback. The main focus of this chapter is to improve the design of our docking task experiment and explore a more robust system to track unconstrained mid-air interactions.

4.1 Experimental Task

In the pilot study described in Chapter 3, the cursor chair was assigned a random orientation and the orientation of the target chair never changed (Section 3.1). For this experiment, the target chair is assigned a uniformly distributed random orientation. The cursor chair could not take a random orientation because of the particular mapping of one of the mid-air devices, the MiniChair, as will be described in Section 4.6.

In the study from the previous chapter, participants had to confirm that the chair was docked, but we believe this step could be omitted because we want to find which device has the smallest docking time for each difficulty level. Thus, if the participant releases the clutch while the chair is docked within the minimum tolerance, a confirmation sound is played and the trial is completed. We learned from the study with the Leap Motion that a time limit should be implemented to avoid fatigue, so for this experiment each trial needed to be completed within a time limit of 40 s or it is automatically skipped.

4.2 Virtual Environment

Some of the improvements to the virtual environment included textures on the walls, placement of additional objects around the room, and increasing the thickness of the seat and backrest of the chairs in order to help with depth perception (Figure 4.1). The second camera, which shows six different views depending on the position of the device, can now be frozen (or unfrozen) at a desired perspective by pressing the ‘F’ key. This may help perform the more difficult docking tasks by providing a viewpoint that facilitates finer adjustments when necessary.

Fig. 4.1: The docking task environment. The virtual representation of the AirPen is visible in the image as an elongated ellipsoid. Bottom-left corner: the second camera offers six different views of the target. Top-right corner: the number of completed trials with the current input condition and the time elapsed since the trial began. Top-left corner: the difficulty level of the current task.

4.3 Accuracy Feedback

More visual feedback was added to the docking task to give details on orientation and position accuracy by using a sphere and a cube. The color of the cube shown in Figure 4.4 changes from yellow to green once the position is within the tolerance level. The sphere shown in the same figure varies similarly, based on the orientation error. Background music began to play once the orientation error was less than 40^o to increase the saliency of the sound feedback, which is inversely proportional to the orientation error.

4.4 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 (Figure 4.2). Whereas the experiment described in Chapter 3 used IR shutter glasses, for the present experiment, we switched to the NVidia 3D Vision RF shutter glasses to avoid interference with the IR used by the motion capture system. Our OptiTrack motion capture system was composed of seven Flex:V100 cameras. The software environment for the experiment was developed using the Unity3D game engine.

Fig. 4.2: Setup of the experiment while using the Fingers condition.

Participants manipulated the cursor in mid-air using an “AirPen”, a “MiniChair”, or their own hand and fingers, as described below. In each case, retro-reflective markers were attached to the input device and hand, and tracked by an Optitrack Flex:V100 motion capture system. A fourth input device, the Phantom Omni, was also employed. All four input techniques can be seen in Figure 4.3.

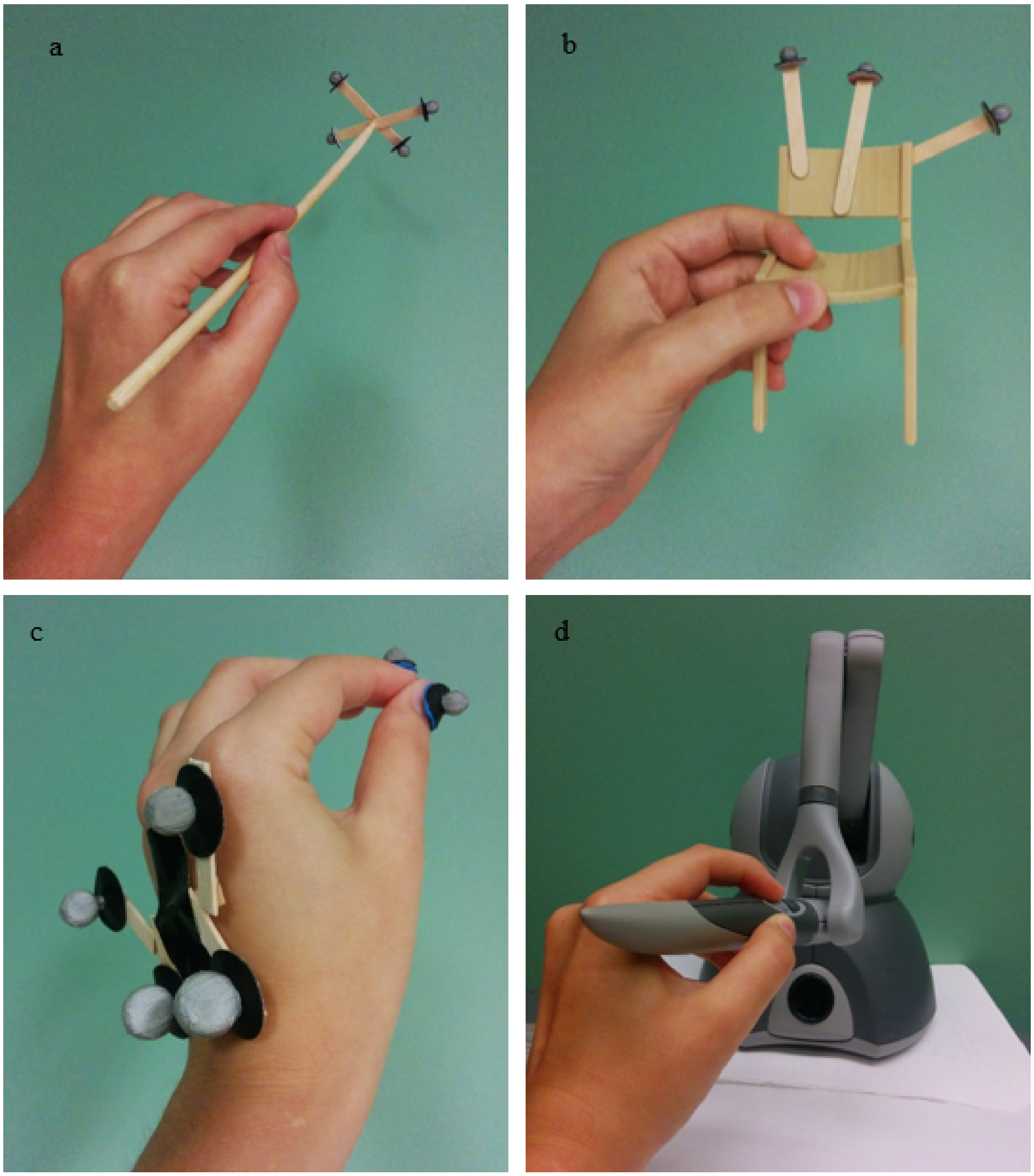


Fig. 4.3: The input conditions used in the experiment were: (a) AirPen (b) MiniChair (c) Fingers and (d) Phantom Omni.

Since latency is known to have a stronger effect than spatial jitter on docking task performance [33], 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 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. 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 [34], 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. [14]. 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 [25]. 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 [35].

4.5 Input Mapping

In contrast to the previous work discussed in Section 2.3 and to improve the input mapping, we did not limit the locations at which the virtual object could be grabbed. Figure 4.4b 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 next section.

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, as illustrated in Figure 4.4) 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.

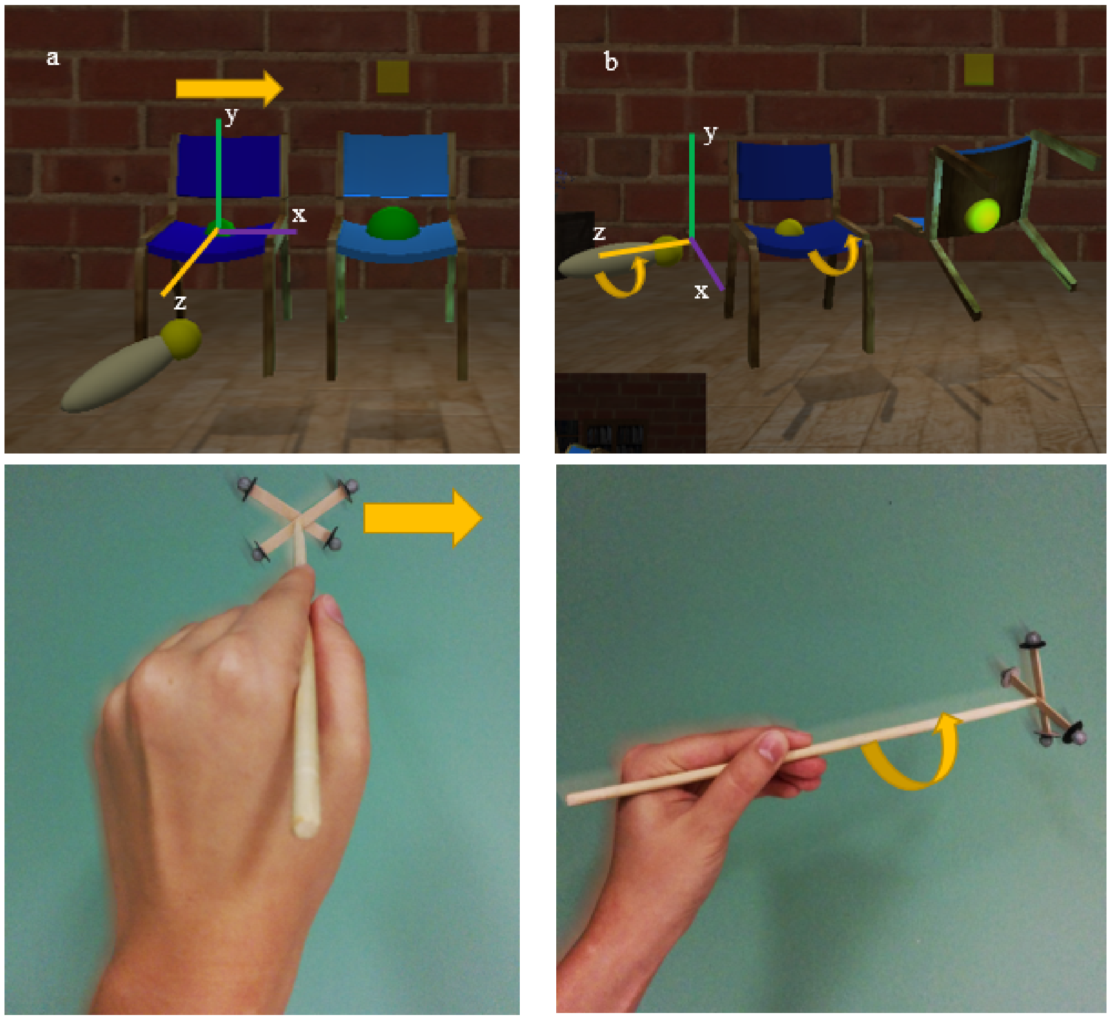


Fig. 4.4: Mapping of the devices. The light blue target chair is under the floating cube. (a) Translating the dark blue virtual chair by dragging the AirPen to the right. (b) Rotating the dark blue virtual chair around the Z axis of the AirPen.

4.6 Input Conditions

4.6.1 AirPen

The AirPen (Figure 4.4a) 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 were 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 by pressing the CTRL key. Clutching indicates when transformations are being performed on the cursor chair by movements of the AirPen.

4.6.2 MiniChair

Inspired by Hinckley's passive real-world interface props [36], the MiniChair (Figure 4.3b) 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 60^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. Clutching by pressing the CTRL key with the non-dominant hand affects translations of the virtual chair.

4.6.3 Fingers

The easiest input device for users to access is, of course, their own hands (Figure 4.4c). 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 manipulated only when the subject is pinching, a gesture commonly used for mid-air selections [30, 31, 32]. 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 used to "clutch" the virtual chair is established through calibration on a per-participant basis.

4.6.4 Phantom

Either button on the stylus is used for clutching (Figure 4.4d). The mapping was improved from the previous study and now allows to manipulate the chair from any point.

4.7 Experiments

4.7.1 Difficulty Levels and Time Limit Study

In the study described in Chapter 3, the experimenter defined the tolerance levels empirically, but we believe that a better approach is to determine the orientation and position tolerances that define easy, medium, and hard tasks through a pilot study. The difficulty levels were established through unconstrained tests performed by four unpaid university students between the ages of 19 and 32. The study consisted of 8 trials $\times$ 2 difficulty levels (easy and hard) $\times$ 4 input mechanisms for a total of 64 trials per participant. The order of the four input devices tested and the difficulty levels were both determined by balanced Latin squares. No feedback on accuracy was provided in this test and there was no minimum tolerance threshold. The participants had to press the ENTER key within a time limit of 60 s to confirm that the moving chair was docked. The initial time limit of 60 s was chosen empirically by the experimenter. For the easy level, the students were instructed to dock the chair as quickly as possible and for the hard level they had to be as accurate as possible within the time limit. Participants were first asked to perform five practice trials.

Results of the Pilot Study

We used the results of the pilot study (Table 4.1) to determine tolerance thresholds for the easy, medium, and difficult levels of the main experiment. We took the largest error over all input conditions in the easy and hard levels from the pilot, i.e., the Fingers and MiniChair, respectively, and used these values as the thresholds for the corresponding levels in the main experiment. The values for the medium level were determined as the average of those of the easy and hard levels. These values were then rounded, as shown in Table 4.2.

Similarly, analysis of the logged docking times led us to select a new time limit of 40 s for trials in the main experiment. This value was sufficient for completion of all trials apart from one outlier (40.10 s).

Difficulty	Orientation error (degrees)	Position error (cm)	Input Condition
Easy	14.82	1.39	Fingers
Hard	4.72	0.60	MiniChair

Table 4.1: Raw largest orientation and position errors for the easy and hard levels.

Difficulty	Orientation error (degrees)	Position error (cm)
Easy	15.0	1.5
Medium	10.0	1.0
Hard	5.0	0.5

Table 4.2: Difficulty levels obtained with the pilot study.

4.7.2 Main Experiment

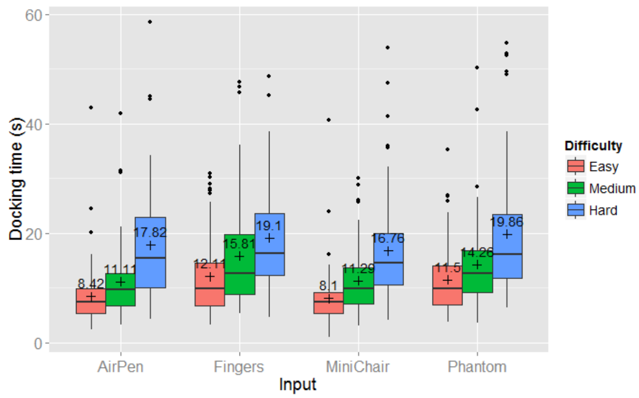
The experiment consisted of 2 blocks × 3 trials × 3 difficulty levels (easy, medium, and hard) × 4 input mechanisms for a total of 72 trials per participant. The order of the four input devices tested and the difficulty levels were both determined by Latin squares. A total of 12 participants took part in the experiment, ranging in age from 20 to 35, drawn from a population of graduate students. For this experiment, we administered a pre-test questionnaire to characterize the participants' experience with use of shutter glasses or gestural interfaces. Two of the participants had previous experience with shutter glasses; none had prior experience with gestural interfaces or had used the Phantom as an input device. We used the difficulty levels and time limit established in the pilot study. If the participant released the clutch while the chair was within the threshold, the orientation and position errors and docking time were logged, and the trial was completed. Participants began by completing a pre-test questionnaire, reading a document with instructions, and watching a short video explaining the visual cues provided in the docking task. They then carried out five practice trials at an easy level before proceeding to the full experiment. The participants were instructed to be as fast and accurate as possible. Following the experiment, participants completed a post-test questionnaire, and were compensated with $10 for their time.

4.8 Results

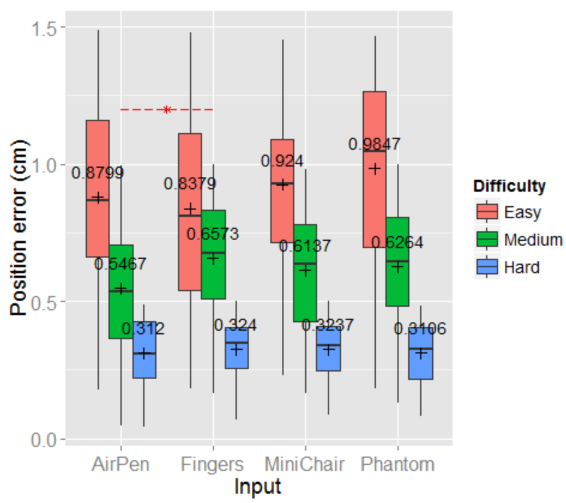
As the data were not normally distributed, we used an aligned rank transform (ART) to conduct a non-parametric ANOVA [37] for the logged docking time, position, and orientation errors. The ANOVA test indicated that the interaction of the device used and the difficulty level had no significant effect on the docking time ($F(6, 66) = ns$). The average docking time for each device, grouped by difficulty level, is shown in Figure 4.5a, but we cannot confidently compare the input conditions since there was not a significant difference between their means. This interaction did have a significant effect on the position error ($F(6, 66) = 3.85$, $p < 0.05$, $GES = 0.11$), so we evaluated the effect of the device used on the position error for each difficulty level. The ANOVA tests found significance between the input conditions for the easy ($F(3, 33) = 3.17$, $p < 0.05$, $GES = 0.12$) and medium ($F(3, 33) = 4.05$, $p < 0.05$, $GES = 0.18$) levels, but not for the hard level ($F(3, 33) = ns$). A pair-wise T test with a Bonferroni correction found no significance between the position error of the devices for the easy level, but there was significance between the position error of the Fingers and the AirPen for the medium level. Figure 4.5b indicates that the AirPen had a smaller position error than the Fingers in the medium level, and this result is significant.

Similarly, the interaction of the device used and the difficulty level had a significant effect on the orientation error ($F(6, 66) = 2.32$, $p < 0.05$, $GES = 0.06$). The ANOVA tests of the device used on the orientation error for each difficulty level found significance only for the medium level ($F(3, 33) = 3.09$, $p < 0.05$, $GES = 0.15$) {easy level: ($F(3, 33) = 1.86$), hard level: ($F(3, 33) = 1.67$)}, and the pair-wise T test with a Bonferroni correction found significance between the the orientation error of the Phantom and the MiniChair for that difficulty level. Figure 4.5c indicates that the Phantom had a smaller orientation error than the MiniChair in the medium level.

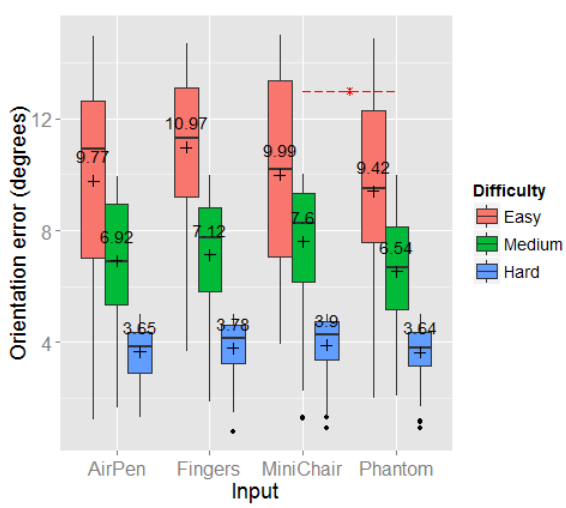
The MiniChair had the smallest docking time for the hard level, but the Phantom had the smallest position and orientation errors for this level, which suggests that using difficulty levels to measure accuracy might not be the best approach. If we ignore the difficulty level variable, the ANOVA tests indicated that the device used had a significant



(a)



(b)



(c)

Fig. 4.5: Average docking time, orientation and position error for each device, grouped by difficulty level, where (+) is the mean and (*) significance.

effect on the docking time ($F(3,33) = 8.45$, $p < 0.05$, $GES = 0.23$), but not on the orientation ($F(3,33) = 2.71$) and position ($F(3,33) = 2.68$) errors. Pair-wise T test with a Bonferroni correction indicated that there was a significant difference between all the input conditions for the docking time except for the MiniChair-AirPen and the Fingers-Phantom. The average docking time, orientation and position errors for each device, are shown in Figure 4.6. We can confidently state that the AirPen and the MiniChair had smaller docking times than the Phantom.

The post-test questionnaire asked participants to rate how they found each interaction, with '5' considered to be a strongly favorable interaction and '1' indicating that it was strongly unfavorable. Participants also rated the level of fatigue they experienced in their wrist and shoulder for each interaction. Results of this questionnaire indicate that subjects preferred the AirPen, while interaction with the fingers proved to be the most fatiguing and least favorable (Figure 4.7).

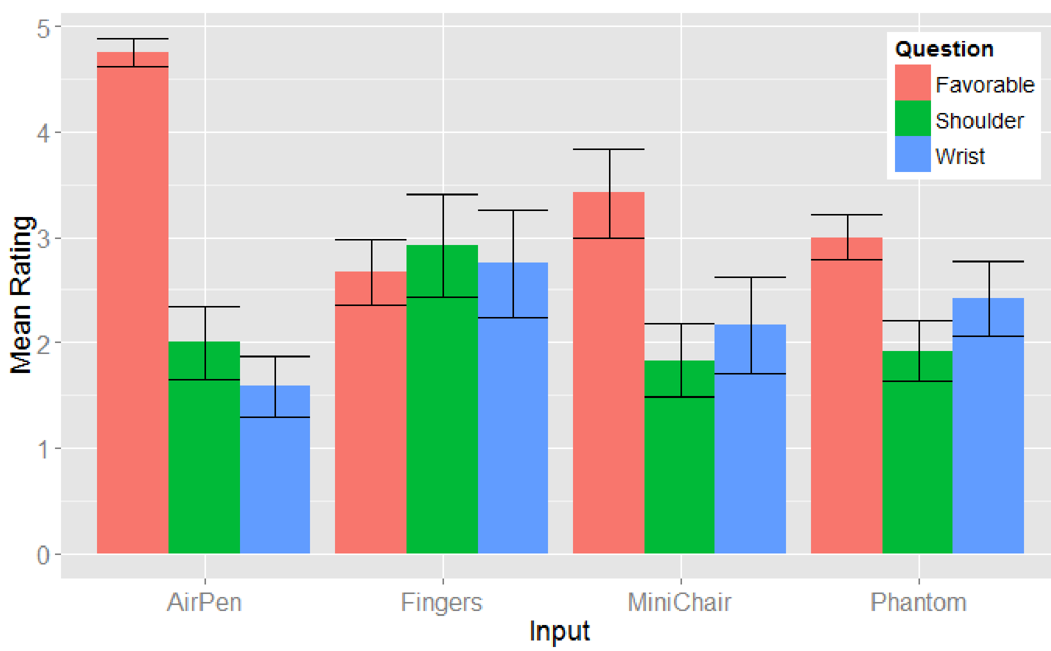
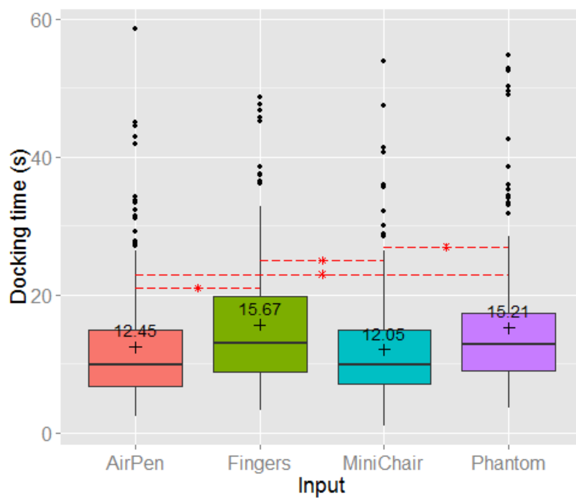
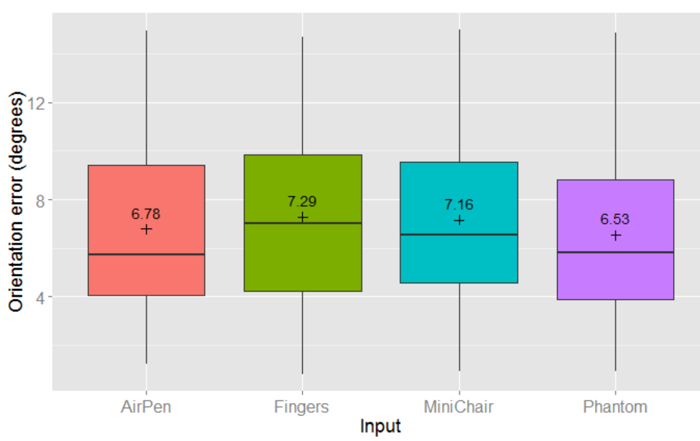


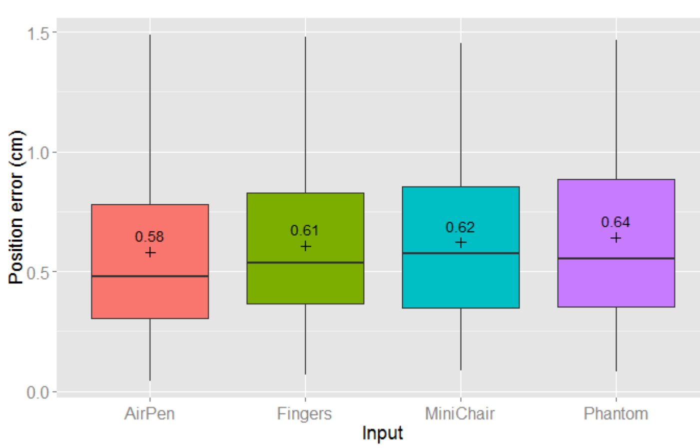
Fig. 4.7: Participants' response to the post-test questionnaire grouped by favorable interaction, shoulder fatigue and wrist fatigue.



(a)



(b)



(c)

Fig. 4.6: Average docking time, orientation and position error for each device, where (+) is the mean and (*) significance.

4.9 Discussion

Our results were consistent with those of Zhai et al. [16] and Bérard et al. [7] in finding that a tangible unconstrained device, such as the AirPen or MiniChair, allowed for completion of docking tasks faster than a constrained device, but we only found a significant effect for this ignoring the difficulty level variable. When we analyzed the effect the interaction of the difficulty level and the input device had on the position and orientation errors, there was only significance for the medium level with one pair of devices. Also, the fastest device at the most difficult level was not the most accurate. These observations might suggest there was an issue with the difficulty levels.

Traditionally, the experimenter chooses the difficulty levels empirically. From our study described in Chapter 3 we thought that a better approach was to determine those levels through a pilot test. 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, so an improved experimental design should not use various difficulty levels to measure accuracy.

A tolerance level is still necessary since it keeps participants from accidentally confirming that the chair has been docked and it also makes the task challenging. However, the tolerance level should not be so high as to prevent participants from achieving success in the absence of additional feedback cues. With only one tolerance level, participants would be able to obtain the lowest orientation and position errors with the most accurate interaction, but for this design there must be a time limit in order to make a fair comparison between trials. Also, participants would have to perform a confirmation gesture once they are within the tolerance and they judge that the chair is docked accurately.

Using a cube to give feedback on position accuracy (Figure 4.4) could split the attention of the participants. A better approach could be to provide multi-modal feedback on position and orientation errors through music. The visual feedback we provided was binary since it only informed participants whether they were within tolerance. It would be better for all feedback to be continuous, similar to the audio feedback we provided for the orientation error.

In Chapter 3 we had to use the keyboard for clutching since the Leap Motion was not robust enough to track clutching gestures. Since we are now using a robust tracking system, we can dispense entirely with the keyboard and use clutching gestures in mid-air instead. The skip feature proved to be unnecessary, as reported by participants. Without that feature, a trial that is overly difficult to visualize would simply cause the time to expire, and the trial to be skipped automatically.

4.10 Conclusion

The analysis above again suggests several improvements that could be made to the design of the experiment. First, we found that using difficulty levels to evaluate the accuracy of the input devices is not the best method since the fastest device in the hard level did not have the smallest position and orientation errors. This led us to move toward the use of a single tolerance level for the experiment described in the next chapter.

However, to make sure that participants are not guessing, they need to confirm once they reach what they believe to be the most accurate orientation. In addition, accuracy feedback should be improved by making it multi-modal and continuous. Finally, since the tracking system is robust enough to track our unconstrained interactions, we will avoid the use of the keyboard and instead adopt a mid-air clutching gesture.

Chapter 5

Accuracy Measured with Position and Orientation Errors

In Chapter 4 we found that for our purposes the docking task experiment should probably just use one tolerance level. In this experiment we use the final orientation and position errors to measure accuracy. The first two sections in this chapter describe the changes made to the experimental design based on our observations and discussion from the previous experiment.

5.1 Experimental Task

Similar to the study described in Chapter 4, a trial was completed once the participant succeeded in aligning the moving chair to the target position and orientation, and confirmed, either by a confirmation gesture or button-press depending on the device. The participants were instructed to be as accurate as possible within the time limit. The easy tolerance threshold (position: 1.5 cm, orientation: 15^o) and time limit(40 s)obtained from the study described in Section 4.7.1 were used. The time limit displayed in the top right corner of the screen turned red when there was 10 s left before the trial ended. This value was chosen empirically by the experimenter.

5.2 Accuracy Feedback

One of our goals was to identify how accurate users could be with the interaction methods under evaluation. Therefore, we provided continuous visual feedback to participants regarding their progress in the docking task, which is an improvement over the binary feedback we provided in our previous experiment, described in Chapter 4. Once the position is within tolerance, drums are heard as audio feedback, and the color of the cube shown in Figure 5.1 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.

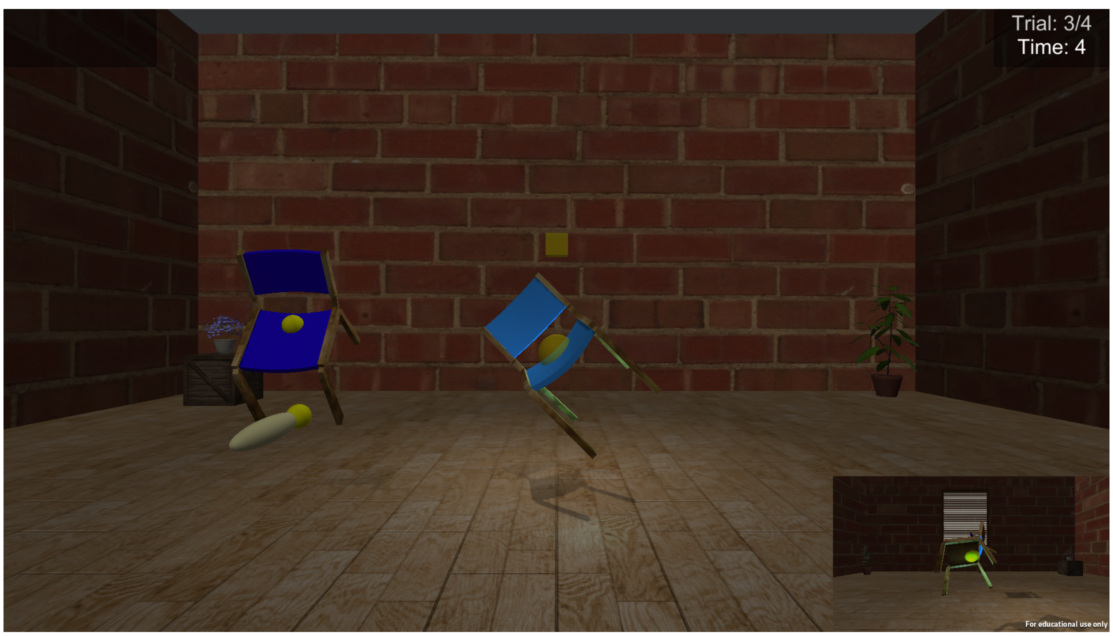


Fig. 5.1: The virtual representation of the AirPen is visible in the image as an elongated ellipsoid. The second camera on the bottom right shows a view from the right side.

5.3 Apparatus

The apparatus and a detailed description of the input conditions are described in Sections 4.4 and 4.6 respectively. In this experiment, a pinching gesture was used for clutching and a fast “tap” gesture, involving contact between the thumb and index finger of less than 0.3 s, confirmed that the chair had been docked. This time threshold for the fast tap was determined empirically by the experimenter.

While the AirPen and MiniChair are held in the dominant hand, the non-dominant hand was used for clutching and confirmation gestures. To avoid occlusions of the markers when the MiniChair is upside down, we limited the orientation between the “up” vector of the target chair and the “down” vector of the virtual world to be always greater than 80^o, a value found empirically to be sufficient. For consistency across conditions, this constraint was applied to all devices. Only the dominant hand was used for manipulation in the Fingers condition. The light colored button on the Phantom was used for clutching and the dark button for confirmation.

5.4 Experiments

Before turning to the main study itself, we first describe a preliminary experiment we conducted to validate the benefits of using an everyday, textured object as the docking cursor and target.

5.4.1 Wireframe Tetrahedron vs. Chair

Some docking tasks in previous work used wireframe tetrahedra as target and cursor [19, 16, 5, 10], with a uniform texture or a checkerboard pattern over the background [6, 7] similar to the environment in Figure 5.2. 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 our 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 5.2) and chairs (Figure 5.1). 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 it was preferred by the participants in Chapter 4. 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.

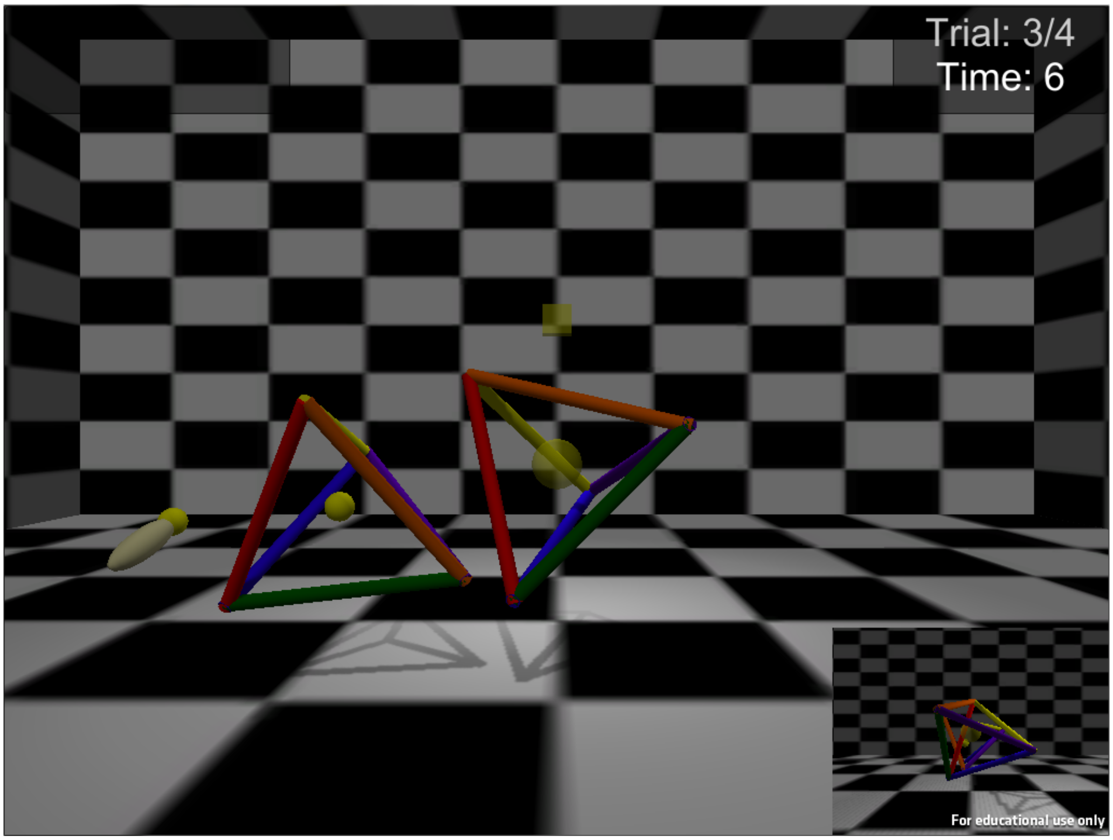


Fig. 5.2: A screenshot from our pilot experiment of a typical docking task experiment using wireframe tetrahedra.

The results in Table 5.1 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 (degrees)**	**Position error (cm)**	**Docking time (s)**
Tetrahedron	9.29	0.71	15.19
Chair	9.30	0.66	11.30

Table 5.1: Average accuracy error and docking time for trials using tetrahedra and chairs.

5.4.2 Main Experiment

The main experiment consisted of 2 blocks × 6 trials × 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.

Results

As the data were not normally distributed, we used ART [37] to conduct a non-parametric ANOVA for the docking time, position and orientation errors. All 19 skipped (timed out) 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 5.3 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 small-

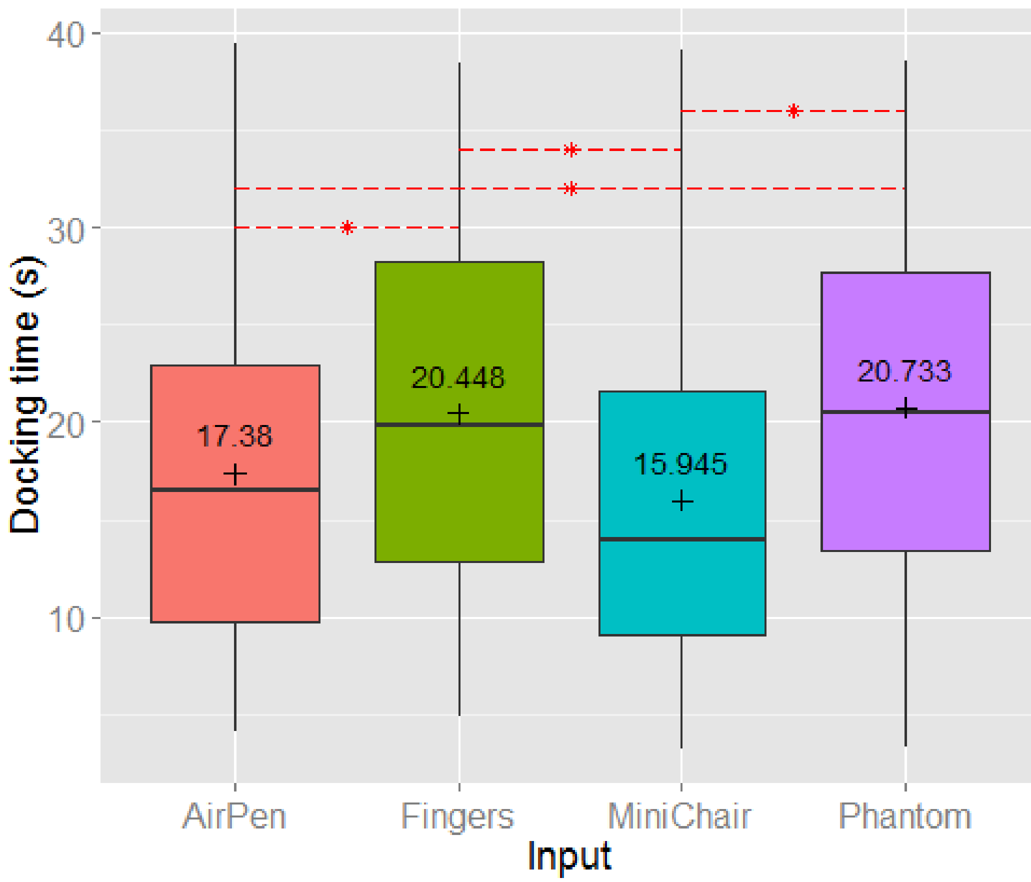


Fig. 5.3: Boxplot of the docking task completion time for each interaction, where (+) is the mean docking time and (*) significance.

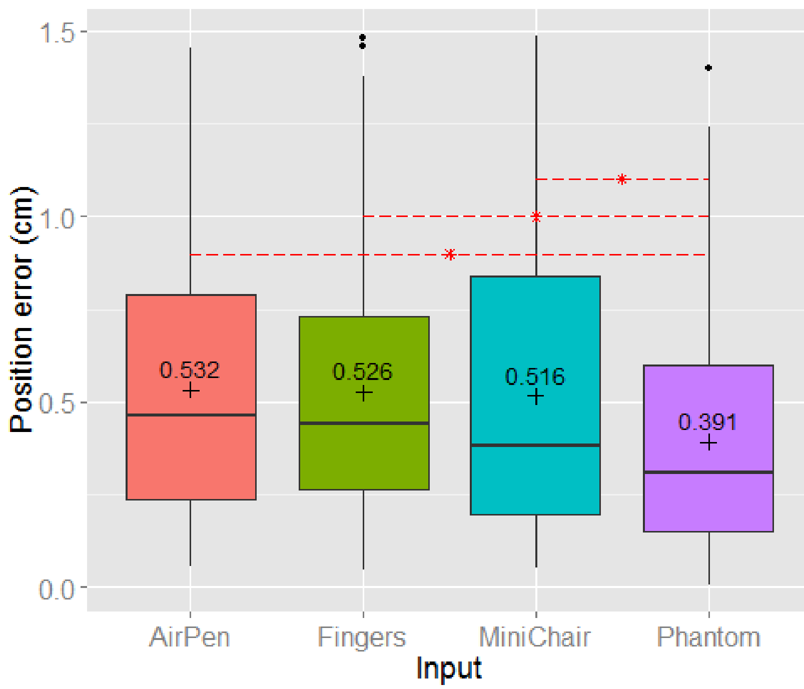
est position error among all input conditions tested (Figure 5.4a). The difference was significant, according to the paired t-tests, although the value of this difference was small: the Phantom was 0.14 cm (26.50%) more accurate than the least accurate interaction for placement, the AirPen. Similarly, the orientation error achieved by participants with the Phantom was the smallest, as shown in Figure 5.4b, 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 5.5.

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.

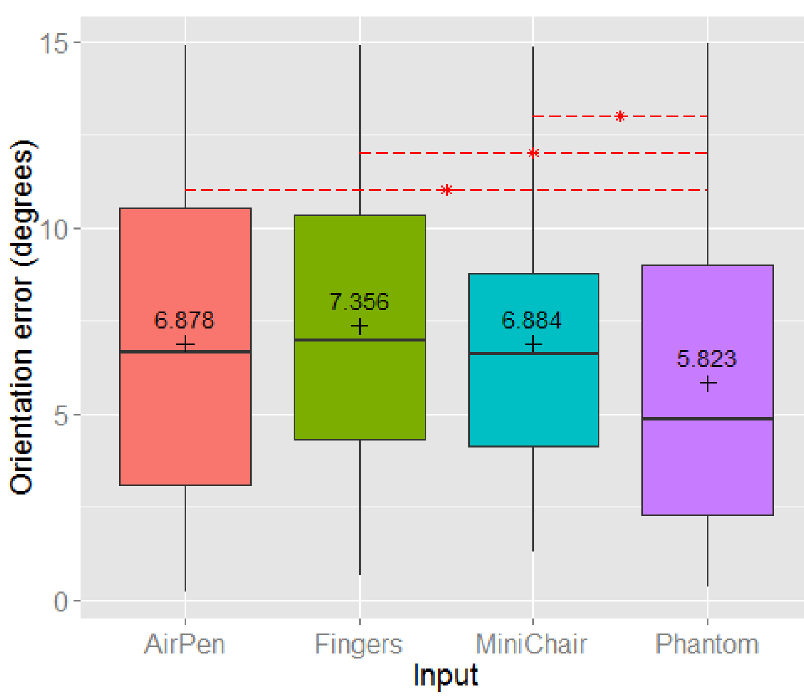
We also found that on average the chair cursor was rotated around its three axes almost equally, but participants preferred rotating the input device around its Z axis while applying rotations to the chair cursor. An ANOVA test indicated that the interaction between the input condition and the rotation axis had a significant effect on the number of rotations performed around each axis ($F(6, 66) = 6.75$, $p < 0.05$, $GES = 0.02$). 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 for 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 5.6). 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.



(a)



(b)

Fig. 5.4: Position and orientation errors for each device, where (+) is the mean and (*) significance.

Fig. 5.5: 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.

5.5 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 [16]. 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

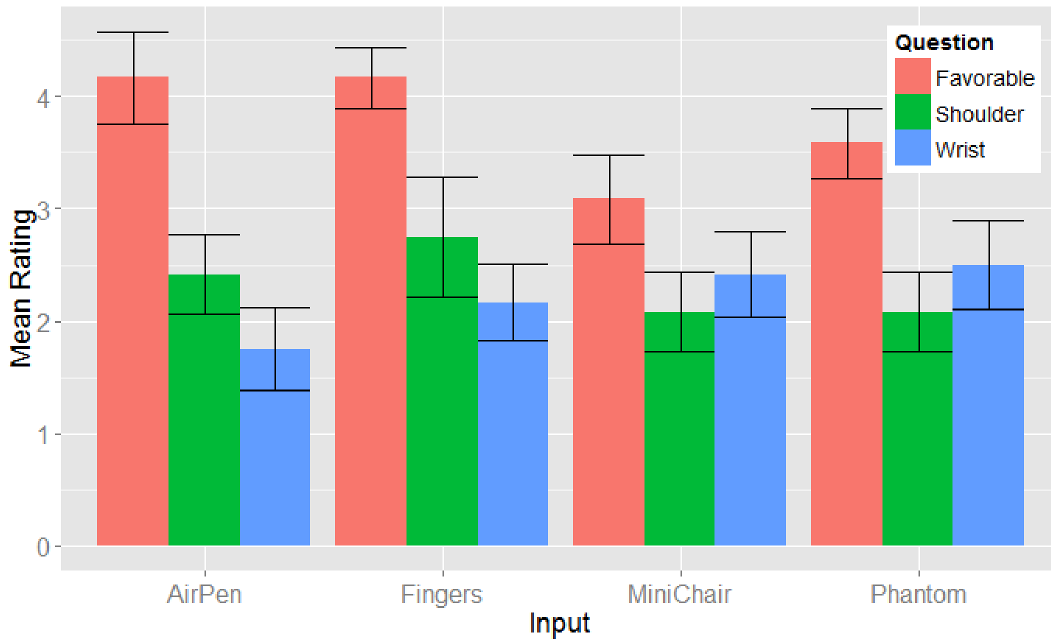


Fig. 5.6: Participants' response to the post-test questionnaire grouped by favorable interaction, shoulder fatigue and wrist fatigue.

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 using a familiar and less symmetrical object improves perception. 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 [38]. Bérard et al. [7] 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 suggested by Hincapie et al. [25], 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 [35] 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.

5.6 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.

Chapter 6

Conclusion

Previous studies have demonstrated that unconstrained mid-air interactions are less accurate but faster than desktop constrained devices that allow for 6 DoF manipulations. We believed that past research was not making a fair comparison between unconstrained and constrained manipulations, so we improved the experimental design of the docking task and explored different input conditions that could be used for unconstrained tracking. First we explored unconstrained input conditions with the Leap Motion, but we found that this was not sufficiently robust to be compared to the Phantom Omni. Thus, for our next experiments we used a passive motion capture system, which offered the necessary robustness.

We developed a mapping that takes advantage of the whole working volume. The mapping that previous studies used assumed that users always pointed the input device towards the screen while manipulating virtual objects, but we observed that users like to manipulate the virtual objects from any point in the working volume. We were also able to reduce the gorilla arm issue by limiting the interaction volume between the hip and shoulders and reducing the need for wide arm extension with a clutching mechanism. With these changes, participants reported the same amount of shoulder and wrist fatigue for the constrained and unconstrained input methods. This is likely because both methods had a small working volume close to the user.

Although previous studies typically docked tetrahedra, we instead used a chair, which is a less symmetrical and more familiar object, with the hypothesis that it would lead to improved performance due to improved perception. We did find improved performance with the chair vs. a tetrahedron, but with the small number of participants, we cannot make broad claims as to the significance of the effect, or attribute it exclusively to improved perception. Depth cues are also important in order to dock the object more easily. To provide additional depth cues, we used textures and lighting effects, and placed other objects around the virtual environment, which was rendered in stereo.

Most docking tasks from the literature evaluated input conditions by comparing only docking time. In our first experiment design, which used three difficulty levels, we assumed that the most accurate device would also exhibit the smallest docking time at the most difficult level. However, in our second experiment, we showed that this was not the case. For our final experiment, we used a single tolerance level and evaluated the input conditions both by their docking time, and by their final position and orientation errors.

We implemented binary visual feedback in our first experiment, but adopted continuous feedback on subsequent experiments after finding that this encouraged participants to achieve greater accuracy. To make the change in feedback more salient, we only provided feedback on accuracy once the chair was inside the tolerance threshold. We also found that multi-modal audio feedback was preferred by participants over visual feedback, probably because it avoided split attention while docking the virtual object, or because it was more salient than the visual feedback.

The mechanically constrained desktop device achieved significantly greater accuracy than the mid-air unconstrained interactions, but the difference was very small. Tangible mid-air interactions proved to be much faster than the constrained interaction probably because of the physical limitations of the Phantom's joints, which constrain the possible movements of the stylus. The fingers interaction had overall the worst performance, but the results for this condition were not very different from the other three conditions. Furthermore, the fingers condition was preferred by participants over the Phantom, the most accurate condition. We also found that participants preferred performing rotations around the Z axis of their stylus.

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.

References

[1] V. Vuibert, W. Stuerzlinger, and J. R. Cooperstock, "Evaluation of docking task performance using mid-air interaction techniques," in *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*, pp. 44–52, ACM, 2015.

[2] K. Kin, T. Miller, B. Bollensdorff, T. DeRose, B. Hartmann, and M. Agrawala, "Eden: a professional multitouch tool for constructing virtual organic environments," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1343–1352, ACM, 2011.

[3] T. Grossman, D. Wigdor, and R. Balakrishnan, "Multi-finger gestural interaction with 3d volumetric displays," in *Proceedings of the 17th annual ACM symposium on User interface software and technology*, pp. 61–70, ACM, 2004.

[4] D. Scheurich and W. Stuerzlinger, "A one-handed multi-touch method for 3d rotations," in *Human-Computer Interaction–INTERACT 2013*, pp. 56–69, Springer, 2013.

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

[6] G. Wang, M. J. McGuffin, F. Bérard, and J. R. Cooperstock, "Pop-up depth views for improving 3d target acquisition," in *Proceedings of Graphics Interface 2011*, pp. 41–48, Canadian Human-Computer Communications Society, 2011.

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

[8] K. Shoemake, "Arcball: a user interface for specifying three-dimensional orientation using a mouse," in *Graphics Interface*, vol. 92, pp. 151–156, 1992.

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

[10] B. Froehlich, J. Hochstrate, V. Skuk, and A. Huckauf, "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*, pp. 191–199, ACM, 2006.

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

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

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

[14] A. Martinet, G. Casiez, and L. Grisoni, "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*, pp. 111–118, ACM, 2010.

[15] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa, "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*, pp. 79–80, ACM, 1996.

[16] S. Zhai and P. Milgram, "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*, pp. 320–327, ACM Press/Addison-Wesley Publishing Co., 1998.

[17] S. Kratz, M. Rohs, D. Guse, J. Muller, G. Bailly, and M. Nischt, "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*, pp. 181–188, ACM, 2012.

[18] G. S. Hubona, G. W. Shirah, and D. K. Jennings, "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*, vol. 34, no. 4, pp. 483–493, 2004.

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

[20] G. Bruder, F. Steinicke, and W. Sturzlinger, "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*, pp. 9–16, ACM, 2013.

[21] J. Schild, J. J. L. Jr, and M. Masuch, "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*, pp. 207–216, ACM, 2014.

[22] A. Cohé and M. Hachet, "Understanding user gestures for manipulating 3d objects from touchscreen inputs," in *Proceedings of the 2012 Graphics Interace Conference*, pp. 157–164, Canadian Information Processing Society, 2012.

[23] D. P. Mapes and J. M. Moshell, "A two-handed interface for object manipulation in virtual environments," *Presence: Teleoperators and Virtual Environments*, vol. 4, no. 4, pp. 403–416, 1995.

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

[25] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani, "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*, pp. 1063–1072, ACM, 2014.

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

[27] 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*, pp. 107–114, ACM, 1997.

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

[29] A. Martinet, G. Casiez, and L. Grisoni, "Integrality and separability of multitouch interaction techniques in 3d manipulation tasks," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 18, no. 3, pp. 369–380, 2012.

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

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

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

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

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

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

[36] 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*, pp. 452–458, ACM, 1994.

[37] J. O. Wobbrock, "Practical statistics for human-computer interaction," in *Annual Workshop of the HCI Consortium*, HCIC, 2011.

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