Benoît Bossavit*
Asier Marzo
Oscar Ardaiz
Luis Diaz De Cerio
Alfredo Pina
Public University of Navarre
Campus Arrosadia, s/n
Edificio de las Encinas
31006 Pamplona, Spain

Design Choices and Their Implications for 3D Mid-Air Manipulation Techniques

Abstract

Manipulation is one of the most important tasks required in virtual environments and thus it has been thoroughly studied for widespread input devices such as mice or multi-touch screens. Nowadays, the Kinect sensor has turned mid-air interaction into another affordable and popular way of interacting. Mid-air interaction enables the possibility of interacting remotely without any physical contact and in a more natural manner. Nonetheless, although some scattered manipulation techniques have been proposed for mid-air interaction, there is a lack of evaluations and comparisons that hinders the selection and development of these techniques. To solve this issue, we gathered four design choices that can be used to classify mid-air manipulation techniques. Namely, choices are based on the required number of hands, separation of translation–rotation, decomposition of rotation, and interaction metaphors. Furthermore, we developed, adapted, and compared three manipulation techniques selected for studying the implications of the design choices. These implications are useful to select among already existing techniques as well as to inform technique developers.

1 Introduction

Nowadays, virtual reality (VR) is extensively used in computer-aided design (CAD), simulations, training, and entertainment. In order to deliver the visual information to the user, different output devices such as monitors, glasses, and wall displays are employed. Additionally, different input devices like mice, keyboards, joysticks, and gloves can be used to affect and explore the virtual environments. Interaction with VR environments can be classified into four categories: manipulation, selection, navigation, and system control (Bowman, Kruijff, LaViola, & Poupyrev, 2001).

Manipulation tasks comprise setting the position and orientation of virtual objects, although other properties such as color, scale, or texture can be manipulated (Poupyrev & Ichikawa, 1999). Every singular value of an object that can be changed is called a degree of freedom (DOF); thus, in 3D environments both position and rotation represent 3 DOF each. Similarly, input devices support a different number of DOF. For example, the traditional mouse has 2 DOF (x and y position). The purpose of a manipulation technique is to define how the DOF of the input affects the DOF of the virtual object. Normally,

Presence, *Vol. 23, No. 4, Fall 2014*, 377–392
doi:10.1162/PRES_a_00207

*Correspondence to benoit.bossavit@unavarra.es.

position and rotation are some of the most challenging types of manipulation as 3 DOF have to be manipulated in a coordinated way. Consequently, they have been thoroughly studied for mainstream devices such as mice and multi-touch surfaces.

Recently, the Microsoft Kinect sensor has enabled new interaction possibilities. This sensor can compute in real time the 3D skeleton pose of a person using a combination of a color and a depth camera. This presents several advantages; for instance, it enables mid-air interaction, which permits users to give commands to a computer without the necessity of having any physical contact with it. Additionally, the interaction is more intuitive and efficient as users utilize their body movements and have a perception of its position and orientation (Mine, Brooks, & Sequin, 1997). Moreover, the Kinect provides up to 3 DOF per skeleton point; therefore, it acts as a 6-DOF input by using only the input of both hands. Finally, it is affordable, available for everyone, and requires neither calibration nor wearing additional devices.

Due to the advantages of the Kinect sensor and the importance of manipulation interactions, some manipulation techniques have been developed for the Kinect and similar devices. However, previous research has focused only on creating manipulation techniques. Furthermore, there are no empirical comparisons between existing techniques and the implications of the design choices that were taken to create the techniques are not discussed. As a result, the literature on manipulation techniques for Kinect consists of just isolated techniques and no indications are provided about which technique should be used in each case. Moreover, the lack of information regarding the effects of the design choices also hampers the development of new techniques.

In order to clarify and guide the selection or development of manipulation techniques for mid-air interaction, we collected four design choices from the existing literature for other interaction modalities. Afterwards, we selected three techniques to cover the most important combinations of design choices. Finally, the three techniques were evaluated in two empirical studies employing the docking task. As the techniques covered the most important combinations of design choices, the consequences of these choices were revealed in the results. These design implications can be used to guide the development and selection of manipulation techniques using the Kinect sensor.

2 Existing Techniques

Existing uni-manual and bimanual interactions for 3-DOF (3D rotations) and 6-DOF (3D translation + 3D rotation) mid-air manipulations are presented subsequently. Some techniques were implemented for hardware different from the Kinect. Nonetheless, their findings are also relevant for our study.

2.1 Uni-Manual Interactions

Segen and Kumar (1998) developed GestureVR, a system that recognizes hand gestures using a video-processing algorithm. It was improved by O'Hagan, Zelinsky, and Rougeaux (2002) with a more robust algorithm. The authors proposed a set of hand gestures that allows the user to manipulate translation and rotation. The rotation was performed by twisting the wrist, whereas the translation was done by closing and moving the hand. It was detected that light conditions could negatively affect the tracking and that this can be solved using infrared cameras.

Another issue derived from previous hand-tracking recognition is the short distance that must be kept between the interaction space and the camera. Lu, Shark, Hall, and Zeshan (2009) proposed a system to rotate and translate a 3D object in an immersive environment. They used a data glove to allow the user to interact from a wider range of distances.

Across all one-hand manipulation techniques, translations are usually made by a drag gesture with the hand closed, whereas rotations are made by twisting, tilting, or swiping the wrist. Raj, Creem-Regehr, Rand, Stefanucci, and Thompson (2012) compared wrist gestures to control rotation, namely wrist tilt with wrist swipe. Results showed that participants used the swipe gesture more.

2.2 Bimanual Interactions

Schlattmann and Klein (2009) proposed a bimanual manipulation technique based on video recognition. They employed the metaphor of grasping an object with both hands. Specifically, a virtual object can be grasped by moving both hands closer. Then, the user could translate the object, displacing both hands concurrently. The rotation of the virtual object followed the averaged orientation between the hands and the middle point between them. Finally, to release the object the users had to separate their hands. The study showed that for precise movements, this technique was faster than a 3D mouse.

Wang, Paris, and Popovic (2011) suggested separating translation from rotation. Translation was made by dragging one hand while it was closed. For rotations, they employed the metaphor of rotating a sheet of paper. That is, the gesture of pinching with both hands allowed the user to rotate around the three primary axes.

Iacolina, Soro, and Scateni (2011) designed a mid-air manipulation analogous to an existing multi-touch technique. Contrary to the previous bimanual techniques, one-hand gestures were used for rotating around the X and Y axes. Two-hand gestures were used for translations and Z-axis rotation.

Researchers such as Bettio, Giachetti, Gobbetti, Marton, and Pintore (2007) and Hackenberg, McCall, and Broll (2011) developed basic two-hand manipulation techniques in order to validate their hand-tracking system. These techniques were improved by Song, Goh, Hutama, Fu, and Liu (2012) producing a two-hand 7-DOF manipulation technique (3 DOF for translations, 3 DOF for rotations, and 1 DOF for scale). This technique allows manipulation of multiple objects at the same time and modification of translation and rotation simultaneously. They used the metaphor of manipulating a handle bar with two hands that pierces the objects.

3 Design Choices

During the design of a manipulation technique, several key decisions have to be taken. These design choices will have a major effect on the usability of the resulting technique. Namely, usability encompasses speed, accuracy, user's error rate, ease of use, and user's level of satisfaction (Bowman, Gabbard, & Hix, 2002). In this section, we have gathered from the existing literature four design choices and embodied them in the form of questions.

Despite being fundamental questions, conclusions differ depending on the technology or from one study to another. There are studies with even opposite conclusions. The most recent studies tend to be less assertive and to conclude that there is not a definitive answer. In any case, the presented design choices can be used to classify manipulation techniques, as they have a major impact on the techniques.

3.1 Should the Technique Use One or Two Hands?

In 1986, Buxton and Myers (1986) showed that using both hands in sequential tasks can reduce the task completion time, since it avoids task switching. Later, Guiard (1987) proposed a theoretical model for human asymmetric bimanual interactions in which the non-dominant hand can cooperate with the dominant hand even when their roles were different. Nevertheless, opinions about this model are divergent.

Some researchers agree with the model and posit that using two hands is more efficient when the task associated to each hand has the same conceptual objective (Leganchuk, Zhai, & Buxton, 1998; Owen, Kurtenbach, Fitzmaurice, Baudel, & Buxton, 2005). In this case, bimanual interactions have physical and cognitive advantages.

In contrast, some researchers suggest that the non-dominant hand can complicate the interaction as the user has to synchronize both hands (Kabbash, Buxton, & Sellen, 1994; Seay, Krum, Hodges, & Ribarsky, 2000). Nonetheless, the second hand can be used in parallel for simple actions.

Nancel, Wagner, Pietriga, Chapuis, and Mackay (2011) demonstrated that for Pan and Zoom actions, bimanual interaction was faster than uni-manual interaction. Nonetheless, 3D translations and rotations require further analysis as they are more complex tasks.

3.2 Should the Technique Integrate Translation and Rotation?

From the point of view of Jacob, Sibert, McFarlane, and Mullen (1994) and Wang, Mackenzie, Summers, and Booth (1998), translation and rotation are not separable. They concluded that tasks should not be separated when they belong to the same perceptual structure. Additionally, they showed that translation and rotation have a parallel and interdependent structure, although translation is a more dominant process.

Nevertheless, subsequent studies have shown that even when the input device allows the simultaneous manipulation of translation and rotation, users frequently manipulate them separately (Masliah & Milgram, 2000). Additionally, Froehlich, Hochstrate, Skuk, and Huckauf (2006) concluded that separated manipulation is more suitable for a docking task because it has better usability and produces less manual motor fatigue.

On the other hand, an evaluation reported that the best option is to design interaction techniques that allow the user to perform both separated and simultaneous manipulations (Hancock, Carpendale, & Cockburn, 2007).

3.3 Should the Technique Decompose Rotation by Axis?

Chen, Joy Mountford, and Sellen (1988) and Jacob et al. (1994) have shown that the fastest and more intuitive way to complete a rotation task is through free rotation; for example, by not decomposing rotation.

Nonetheless, Veit, Capobianco, and Bechann (2009) showed that decomposing rotation is as precise as not decomposing it, but faster. They also detected that in composed rotation, users generally used only up to 2 DOF at the same time, even when the technique supported 3 DOF.

For both types of rotation, Parsons (1995) concluded that users encounter significant difficulties for mentally rotating objects, particularly when the rotation axis did not coincide with one of the viewer primary axes.

3.4 Could the Technique Be Described with a Metaphor?

It seems fundamental to create interaction techniques that bear some resemblance to actions already known by the users (Shank & Gebler, 2002); that is, a metaphor that naturally explains an unfamiliar domain (Bowman, McMahan, & Ragan, 2012). Fishkin (2004) highlights the importance of a metaphor as an enormously powerful component in thought and design that also plays a vital role in interaction techniques.

4 Selection of Techniques

The purpose of this section is to select and describe techniques that represent the most relevant combinations of the presented design choices. Therefore, by evaluating the selected techniques it will be possible to determine the consequences of the design choices.

For increasing the significance of the evaluation, we intend to use a within-subjects design. That is, the subjects must try all the techniques in different orders. Therefore, the number of techniques to test should be kept low in order to guarantee the evaluation feasibility. Consequently, we carried out a pilot study to narrow the selection of techniques to the three most representative ones. The pilot study took into account the techniques covered in the literature and four main conclusions were extracted from it. First, metaphors are always helpful and thus the evaluated techniques must employ a suitable metaphor. Second, if a technique integrates translation and rotation, it also should not decompose rotation. On the other hand, when translation and rotation are performed as separated actions, it is better to decompose rotation on primary axes. Finally, for two-hand techniques it is more reasonable to integrate translation and rotation.

To synthesize, we need three techniques that employ metaphors. Specifically, one technique must use one hand, separate translation from rotation, and decompose rotation. Another one also has to use one hand but integrate translation and rotation, and compose rotation. Finally, the last technique should use two hands, integrate translation and rotation, and compose rotation.

Table 1. *Classification of Mid-Air Manipulation Techniques According to the Design Choices*

	Number of hands		Translation - Rotation		Rotation axis		
	One	Two	Separated	Integrated	Decomposed	Mixed	Metaphor
O'Hagan et al. (2002)	X		X			X	-
Lu et al. (2009)	X		X			X	-
Raj et al. (2012)	X		X			X	-
Schlattmann and Klein (2009)		X		X		X	Grasp
Wang et al. (2011)		X	X			X	Sheet of paper
Iacolina et al. (2011)		X		X		X	-
Song et al. (2012)		**X**		**X**		**X**	Handle Bar
Crank Handle (created)	**X**		**X**		**X**		Rotate Cranks
Grasping Object (adapted)	**X**			**X**		**X**	Friction

A classification of the current techniques attending to the design choices (see Table 1) revealed insufficiencies in one-hand techniques. Specifically, they did not employ a metaphor and none of them separated rotation by axis. Therefore, we had to design the two one-hand interaction techniques required for the evaluation. The last three techniques are presented in this study.

In the following subsections, we describe the three manipulation techniques. For the first technique, we created the Crank Handle, a one-hand technique that separates translation from rotation and decomposes rotation into the primary axes employing the metaphor of rotating three different cranks. For the second one, we adapted the RNT algorithm (Kruger, Carpendale, Scott, & Tang, 2005) to 3-DOF inputs. It resulted in a one-hand technique that integrates translation and rotation without decomposing rotation. It is called the Grasping Object technique. Finally, for the third technique we reproduced the Handle Bar (Song et al., 2012), an existing two-hand technique that integrates translation and rotation.

4.1 Crank Handle Technique (CH)

Our main objective for this technique was to design a one-hand technique that separated translation from rotation and decomposed rotations in primary axes. Additionally, we employed the metaphor of rotating three crank handles to rotate across each of the primary axes.

4.1.1 Description. This technique has three modes: *idle mode*, *translation mode*, and *rotation mode*. A bar is rendered at one side of the virtual object to provide visual feedback about the current mode. The bar appears at the right or left side of the object depending on the dominant hand of the user.

The system starts at the *idle mode* and returns to it whenever the user opens his hand. In this mode, the user can rest or change the hand placement without changing the object transformation. Throughout this mode the bar is translucent and grey.

From the idle mode the user can close the hand to enter *translation mode*. While the hand remains closed, its movement is transferred to the virtual object. The bar turns opaque and a black handle appears during this mode. To pass to the *rotation mode* the user has to open and close his or her hand in less than 0.6 seconds. This value was chosen after various tests and seemed to be the most usable.

During the rotation mode, three translucent crank handles appear at the end of the bar, one for each primary axis (see Figure 1). Afterwards, the user has to describe a circle with the hand around the primary axis in which he or she wants to rotate the object, exactly as he or she would interact with a real crank handle. When the

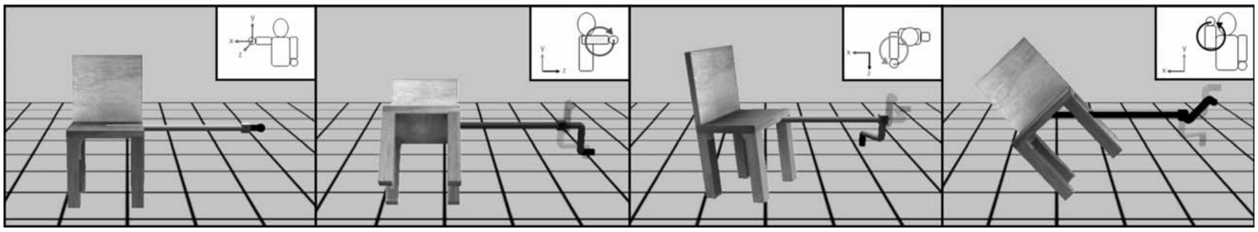

Figure 1. *Crank Handle. From left to right,* translation mode, rotation mode X-axis, rotation mode Y-axis, *and* rotation mode Z-axis.

user rotates one of the crank handles, it becomes opaque and its orientation is updated following the user's gesture. The gesture can be performed continuously and the gain factor varies depending on the gesture linear speed.

4.1.2 Crank Handle Algorithm.

The Crank Handle algorithm has three steps: the detection of the rotation axis, the detection of the rotation angle, and the choice of the gain factor to be applied to this rotation.

To detect the primary axis in which the user wants to rotate the object, the algorithm stores a trail of the last hand positions. From this trail, the average of the normals formed from subsequent triplets of points is calculated using the cross product. The algorithm considers that the user is rotating around a determinate axis when the angle between this axis and the previously calculated average vector is less than 30 degrees. This threshold was chosen after previous tests to avoid confusion between primary axes but, at the same time, to afford a certain degree of imprecision in the gesture.

Once the primary axis is detected, the curvature of the trajectory is analyzed to detect direction changes or undesired movements such as lines. Then, the algorithm computes the center and radius of the circle using the method proposed by Bourke (2014). When the center of the circle is known, the angle between subsequent points can be determined.

Finally, a gain factor is applied to the angle according to the gesture linear speed. When the hand speed is less than 10 cm/s no rotation will be applied to the virtual object, as we consider the movement indecisive. Between 10 cm/s and 65 cm/s a linear function is used; it transforms 50 loops at lower speeds or 30 loops at higher speeds to 360 degrees. Above 65 cm/s, an exponential function that transforms 10 loops at lower speed to 360 degrees is used; it was capped at 360 degrees per 2.5 loops.

4.2 Grasping Object Technique (GO)

We propose a second one-hand manipulation technique. Opposite to the Crank Handle, it combines translation and rotation and does not decompose rotation in primary axes. Its metaphor comes from the physics of moving an object against friction or through a stream.

4.2.1 Description.

The Grasping Object is based on the RNT algorithm (Kruger et al., 2005) and on its 3D extension for 2-DOF inputs (Hancock et al., 2007). In this paper, we extended the algorithm to support 3-DOF inputs.

The technique starts in the *idle mode* and returns to it whenever the user opens his or her hand. In the idle mode the user can aim a virtual ray with his or her hand. When the ray intersects a point of the object, a blue sphere is drawn in this point. Subsequently, the user can grab the object by closing the hand. The virtual ray will disappear and the blue sphere will turn green to indicate the change to the *transformation mode*. Then, the object will be modified in translation and rotation according to the hand trajectory (see Figure 2). The grabbed point will follow the trajectory of the hand as described by Hancock et al. (2007). In order to facilitate the pointing of the virtual ray, its range is limited to the bounding box of the object.

This technique manipulates translation and rotation simultaneously; however, it was not feasible for the users

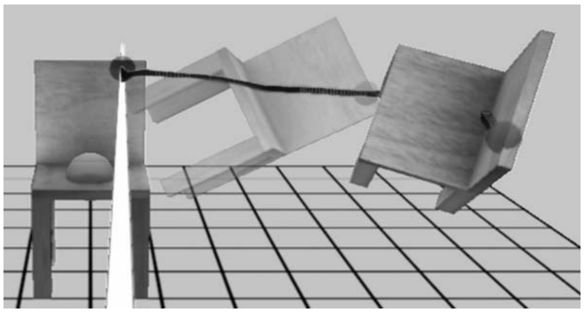

Figure 2. *Grasping Object. Transformation made by grasping the object on the left and following the black path with the hand. The object is in* idle mode *on the left, and in* transformation mode *on the center and right.*

to precisely manipulate the object in this way. A previous study (Hancock et al., 2007) reported the same issue with the technique and added a translation-only mode; we proceeded likewise. Namely, if the selection ray intersects a translucent sphere centered in the object, the system changes to *translation mode* and only translation is modified. The diameter of the sphere is 30% of the object size and was chosen to be large enough to enable a good selection and sufficiently small to allow the user to grasp any corner of the object.

4.3 Handle Bar Technique (HB)

For the third technique we decided to replicate the Handle Bar Metaphor (Song et al., 2012). It is a recent technique that can be implemented using the Kinect sensor. The employed metaphor of this technique consists in manipulating a virtual handle bar that pierces the object (see Figure 3).

4.3.1 Description. In this technique both ends of a handle bar are grabbed and controlled with each hand. The user can control the translation and rotation of the object since it behaves as if it was pierced by the bar.

With both hands open the user remains in the *idle mode*. In this mode, the user can orientate the handle bar without transforming the virtual object. The bar is displayed translucently and in grey. Once a virtual object is selected, the virtual handle bar is drawn through the virtual object with the representation of the two hands in each side.

If the user closes both hands, he or she enters the *transformation mode*, in which he or she can manipulate the 6 DOF of the object. The handle bar is displayed in blue during this mode. With a parallel movement of the hands, the user can move the virtual object without rotating it. With a horizontal or vertical asymmetric movement, the user will be able to rotate the object around the y axis or z axis, respectively. Finally, to rotate around the x axis, the user can perform a pedaling gesture with both hands.

Additionally, when the user opens only one hand, the handle bar turns violet and the system enters the *constrained rotation mode* (see Figure 3, right). In this mode the user can describe a circle with the open hand to rotate the object around the bar axis.

5 User Studies

We conducted two user studies that compared and evaluated the three techniques in order to extract insights about the effects of the design choices. The techniques were tried by 18 people (5 female and 13 male), aged between 16 and 49 ($M = 27.3$, $SD = 8.9$). Two users were left-handed but both one-hand techniques were designed to support this.

The experiment was performed using a 47″ 3D TV with a Microsoft Kinect placed over it. The software ran under Windows 7 and was developed using C#, the Microsoft Kinect SDK, and GoblinXNA. Participants were located in front of the Microsoft Kinect at a distance of 2 m. As suggested by Bowman, Coquillart, Froehlich, Hirose, Kitamura, et al. (2008), in order to enhance the spatial perception of the user we employed stereoscopy, perspective, occlusion, and shadows. One meter in the real world was equivalent to 5 m in the virtual world.

5.1 Study 1: 3D Docking Task

In this study, we used a 3D docking task (Zhai & Milgram, 1998). Participants were asked to overlap a moveable dark tetrahedron over a static pale tetrahedron

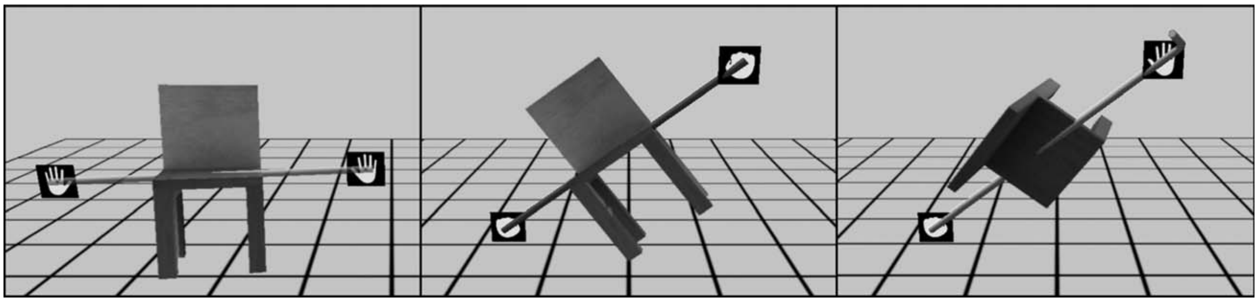

Figure 3. *Handle Bar. From left to right:* idle mode, transformation mode, *and* constrained rotation mode.

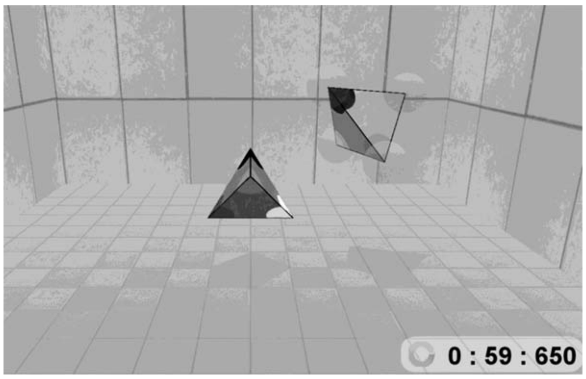

Figure 4. *3D Docking task: Participants had to move the dark tetrahedron inside the pale one in 1 min. Spheres represent the error tolerance.*

(see Figure 4). The dark tetrahedron was initially located at the center of the scene. The edges of the tetrahedrons were 1.63 m. The position of the target tetrahedron was randomly generated at a distance of 3 m from the moveable tetrahedron. Each corner of the tetrahedrons had a different color so there was only one correct orientation. The error tolerance was represented by a sphere on each corner of the target tetrahedron. The size of these spheres was 45% of the tetrahedron size. Our pilot study had shown that a lower threshold reduced considerably the number of successful dockings. The spheres were red but turned green when the correct corner was inside. Similar to Froehlich et al. (2006), the docking trial ended successfully when the four corners remained inside their corresponding spheres for 0.8 s. If a participant could not complete the trial within one minute, the docking ended in failure. To continue to the next docking trial, participants had to activate a button. Then, they had 3 seconds before the trial started to analyze the situation or to place their hands as desired. Object selection was disabled for the study and the manipulation was directly applied to the moveable object.

We used a within-subjects design. The independent measured variables were *Technique*, *Rotation*, and *Angle*. The values of the Rotation variable were *Simple* (around one primary axis) and *Complex* (around a random axis). The values of the Angle variable were *Acute* (between 30 and 90 degrees) and *Obtuse* (between 91 and 150).

Participants performed 6 blocks of 5 trials. The order of the trials was: *No Rotation*, *Simple Rotation* with *Acute Angle*, *Simple Rotation* with *Obtuse Angle*, *Complex Rotation* with *Acute Angle*, and *Complex Rotation* with *Obtuse Angle*. With this study we wanted to measure the performance of the techniques in terms of speed, accuracy (percentage of successful dockings), inefficiency (amount of misused work in rotation and in translation) (Zhai & Milgram, 1998), and reaction time. Additionally, we aimed at analyzing each technique by assessing the time passed in each mode and by gathering two specific metrics. These metrics are the coordination translation/rotation, based on the ratio between actual trajectory and the optimal one (Zhai & Milgram, 1998); and the *m*-metric, that analyzes the use and efficiency of the different combinations of DOF (Masliah & Milgram, 2000).

5.2 Study 2: Precise 3D Docking Task

In the second study, participants were asked to place the tetrahedron as precisely as possible within 2 minutes. Contrary to the first study, the target was

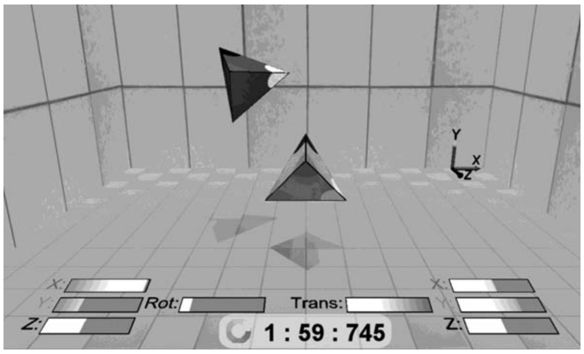

Figure 5. *Precise 3D docking task: Participants had to move the dark tetrahedron over the pale one in 2 min. Gauge bars represent the error in translation and rotation and its decomposition into primary axes.*

always placed at the center of the scene. Two main error gauges were displayed, one for the global rotation error and another for the global translation error. Additionally, each of the two gauges was decomposed into the three primary axes. The gauges had two scales going from red to yellow and yellow to green (see Figure 5). The gauge entered in the second scale when the error was less than 25 cm for translation or 30 degrees for rotation.

Participants performed 3 blocks of 2 trials. The first trial was a *simple rotation* and the second a *complex rotation*. The aim of this study was to determine the minimum error achievable in translation and in rotation.

5.3 Procedure

Participants performed the evaluation across 3 days, one day per technique. Each session lasted approximately 1 hour and consisted of two studies. First, we explained and showed to them how the technique works. Then, as training they tried a complete block without any time limit. Afterwards, they performed the first task and then the second one. They could rest between each trial. Finally, they were asked to fill out the NASA-TLX questionnaire (Hart & Staveland, 1998) and a custom questionnaire that evaluates the usability of the technique (7-point Likert scale). Moreover, at the end, the participants had to rank the three techniques according to their preferences. Participants interacted with the techniques in a different order following a Latin Square. The transformations of the trials were randomly generated and were different for each technique but the same for each participant.

To summarize, the experiment consisted of: 18 participants $\times$ 3 techniques $\times$ ([6 blocks x 5 trials] + [3 blocks $\times$ 2 trials]) = 1944 docking tasks.

6 Results

Data were analyzed using repeated measures ANOVA. For data that violated the Mauchly's test of sphericity, we reported results using the Greenhouse-Geisser correction. When a significant effect appeared, we performed a t-pair pairwise comparison with Bonferroni correction to detect significant differences. Only completed trials were included in this analysis. CH stands for Crank Handle, GO for Grasping Object, and HB for Handle Bar.

6.1 Techniques Comparison

6.1.1 Task Completion Time (TCT). For docking trials with only translation, the analysis revealed a significant effect of the technique, $F(2;34) = 5{:}780$; $p < 0{:}01$. The pairwise comparison showed that CH (7.6 s, $SD = 0.8$) was significantly faster ($p < 0{:}05$) than HB (11.8 s, $SD = 1.1$). GO average TCT was 10.4 s ($SD = 1.2$).

For *Simple Rotation* docking trials, the analysis revealed a significant effect of the technique, $F(2;34) = 50{:}506$, $p \approx 0$. The pairwise comparison showed that CH (23.2 s, $SD = 1.2$) was faster ($p < 0{:}01$) than HB (29.5 s, $SD = 2$) and faster ($p \approx 0$) than GO (36.2 s, $SD = 1.6$). HB was also faster ($p \approx 0$) than GO.

For *Complex Rotation* docking trials, the analysis revealed a significant effect of the technique, $F(2;34) = 10{:}391$, $p \approx 0$. The pairwise comparison showed that HB (31.7 s, $SD = 1.6$) was faster ($p < 0{:}05$) than CH (36.6 s, $SD = 1.7$) and faster ($p < 0{:}01$) than GO (38.8 s, $SD = 1.9$). The TCT of each technique split by trial type is shown in Figure 6.

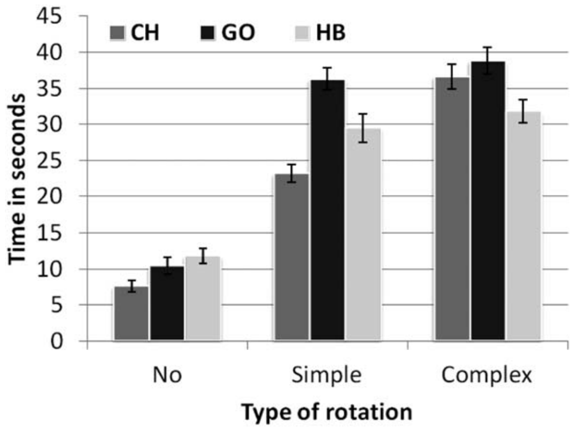

Figure 6. *Task completion time per technique and per type of trial.*

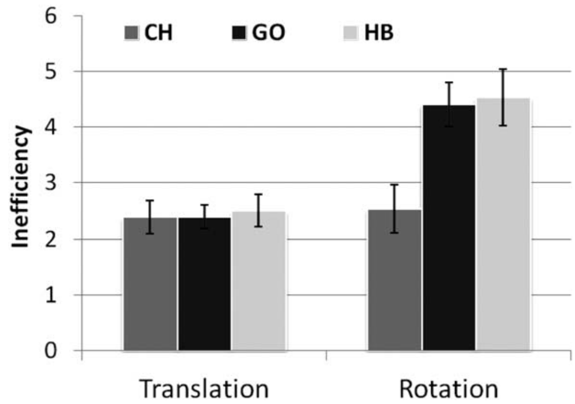

Figure 7. *Inefficiency in translation and rotation.*

6.1.2 Accuracy. For all docking trials (successful or not), the analysis showed a significant effect, $F(2;34) = 9:308$, $p < 0:01$. The pairwise comparison showed that HB (86.1%, $SD = 3.4$) and CH (83.1%, $SD = 2.6$) were significantly more accurate ($p < 0:05$) than GO (75.9%, $SD = 3.1$).

6.1.3 Inefficiency in translation. The translation inefficiency value represents the ratio between the length of the followed path and the length of the optimal one. For only translation docking trials, the analysis showed a significant effect, $F(2;34) = 4:097$, $p < 0:05$ although the pairwise comparisons did not show any significant difference. The inefficiency value was 1.43 ($SD = 0.2$) for CH, 1.83 ($SD = 0.2$) for GO and 2.12 ($SD = 0.3$) for HB. For all the other sets of docking trials, no significant difference was revealed. Inefficiency in translation is shown in Figure 7.

6.1.4 Inefficiency in rotation. The rotation inefficiency value represents the ratio between the sum of all the performed rotations and the optimal rotation. The analysis showed a significant effect, $F(2;34) = 9:731$, $p \approx 0$. The pairwise comparison showed that CH (2.54, $SD = 0.4$) was less inefficient ($p < 0:01$) than GO (4.4, $SD = 0.4$) and HB (4.5, $SD = 0.5$). Inefficiency in rotation is shown in Figure 7.

6.1.5 Reaction time. We report the data of all the docking trials, as the results of the set of *Simple* and *Complex Rotation* trials were similar. The analysis showed a significant effect, $F(1:436;24:416) = 16:336$, $p \approx 0$. The pairwise comparison showed that CH (0.7 s, $SD = 0:06$) provoked a smaller reaction time ($p < 0:01$) than HB (1.7 s, $SD = 0.3$) and GO (2.2 s, $SD = 0.1$).

6.1.6 Precision in translation. These results are derived from the second study. We report the data of all the docking trials, as the results of the set of *Simple* and *Complex Rotation* trials were similar. The analysis did not show any significant effect. The maximum precision reached in translation was 3.3 cm with HB ($SD = 1.5$), 7.2 cm ($SD = 11.4$) with CH, and 8.7 cm ($SD = 5.4$) with GO.

6.1.7 Precision in rotation. These results are also derived from the second study. For the set of all the trials that required rotation, the analysis showed a significant effect, $F(2;34) = 29:448$, $p \approx 0$. The pairwise comparison showed that CH (1.4°, $SD = 0.9$) and HB (1.47°, $SD = 0.6$) were more precise ($p \approx 0$) than GO (3.1°, $SD = 1.1$).

For all the *Simple Rotation* docking trials, the analysis showed a significant effect, $F(2;34) = 38:134$, $p \approx 0$). The pairwise comparison showed that CH (0.79°, $SD = 1.1$) was more precise ($p < 0:05$) than HB (1.4°, $SD = 0.6$) and more precise ($p \approx 0$) than GO (3.3°, $SD = 1.3$). HB was also more precise ($p \approx 0$) than GO.

For all the *Complex Rotation* docking trials, the analysis showed a significant effect, $F(2;34) = 8:43$,

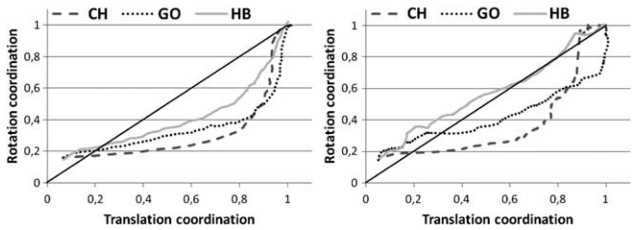

Figure 8. *Translation–rotation coordination: average of all the participants on the left, and a specific participant on the right. For both charts, the curve begins at (1, 1) and ends at (0, 0) plus the error tolerance.*

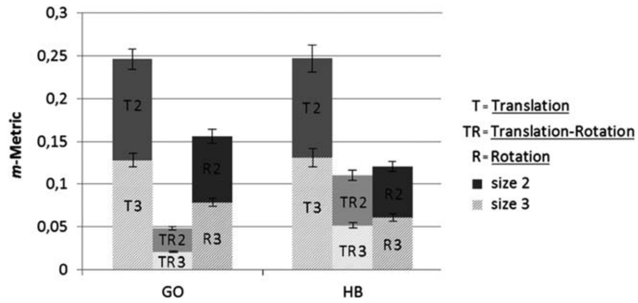

Figure 9. *m-metric values for groups of size 2 and 3 for GO and HB techniques.*

$p = 0{:}01$). The pairwise comparison showed that HB (1.5°, $SD = 0.7$) was more precise ($p < 0{:}01$) than GO (2.9°, $SD = 1.2$). CH had a maximum precision of 2 degrees ($SD = 1.3$).

6.2 Analysis of the Techniques

6.2.1 Transformation coordination. The coordination between rotation and translation (Zhai & Milgram, 1998) is plotted in Figure 8 split by technique.

6.2.2 *m*-metric. *m*-metric (Masliah & Milgram, 2000) defines the allocation of control as the product between usage and efficacy of a certain set of DOF. For our study, the 6 DOF are X, Y, Z, RX, RY, and RZ. Therefore, there are 15 metrics for groups that combine 2 DOF (size 2), 20 for groups of size 3, 15 for groups of size 4, 6 for groups of size 5, and 1 that combines the 6 DOF. As the complete report of *m*-metric is large, we report the most relevant statistical tests. The analysis was applied for each technique separately.

Attending to the groups that contain only translation (XY, XZ, and YZ), the analysis showed a significant difference between pairs of planes for every technique.

For CH, $F(2;34) = 67{:}901$, $p \approx 0$, the *m*-metric value of translations on the plane XY was 0.35 ($SD = 0.09$), 0.29 ($SD = 0.07$) on plane XZ, and 0.22 ($SD = 0.06$) on plane YZ.

For GO, $F(1{:}484;25{:}229) = 128{:}219$, $p \approx 0$, XY was 0.31 ($SD = 0.07$), XZ was 0.22 ($SD = 0.04$), and YZ was 0.19 ($SD = 0.04$).

For HB, $F(2;34) = 89{:}068$, $p \approx 0$, XY was 0.3 ($SD = 0.07$), XZ was 0.23 ($SD = 0.07$), and YZ was 0.2 ($SD = 0.06$).

More specifically, the pairwise comparisons revealed significant differences at XY > XZ ($p \approx 0$) and XY > YZ ($p \approx 0$) for the three techniques.

Comparing groups of only translation (T), coupled translation-rotation (TR), and only rotation (R), statistical tests showed a significant difference for groups of size 2 and 3 on both GO and HB techniques. In more detail:

For GO and groups of size 2, $F(2;34) = 245{:}347$, $p \approx 0$, the *m*-metric value was 0.24 ($SD = 0.05$) for T groups, 0.04 ($SD = 0.009$) for TR groups, and 0.15 ($SD = 0.03$) for R groups. For groups of size 3, $F(1{:}467;24{:}947) = 164{:}838$, $p \approx 0$), 0.12 ($SD = 0.03$) for T groups, 0.02 ($SD = 0.004$) for TR groups, and 0.08 ($SD = 0.02$) for R groups.

For HB and groups of size 2, $F(1{:}1397;19{:}356) = 113{:}658$, $p \approx 0$, the *m*-metric value was 0.24 ($SD = 0.06$) for T groups, 0.11 ($SD = 0.02$) for TR groups, and 0.12 ($SD = 0.02$) for R groups. For groups of size 3, $F(1{:}107;18{:}818) = 72{:}179$, $p \approx 0$, 0.13 ($SD = 0.04$) for T groups, 0.05 ($SD = 0.01$) for TR groups, and 0.06 ($SD = 0.01$) for R groups.

For both techniques and group sizes, the pairwise comparisons revealed that TR groups had significantly lower values than the two other groups (see Figure 9).

For GO, a *t*-paired test revealed that groups which contain RZ had significantly lower values than groups without RZ for groups of size 2 ($p \approx 0$,

Table 2. *Averaged Scores of the Usability Questionnaire*

Questions (7-point scale)	CH	GO	HB
I found the technique easy to understand	6.3	6.5	6.3
I found the technique easy to use	5.8	5.1	5.5
I would need practice to use the technique	3.6	4	3.3
The object reacted as I expected	5.6	4.9	5.8
I found rotation easy to do	5.6	4.8	5.5
I found translation easy to do	6.3	6	6
I felt precise	5.2	4.1	4.8

$\Delta M = 0.009$, $\Delta SD = 0.007$) and groups of size 3 ($p < 0{:}001$, $\Delta M = -0.003$, $\Delta SD = 0.003$).

For HB, a t-paired test showed that groups which contain RX had a significantly lower result than the groups without RX for groups of size 2 ($p \approx 0$, $\Delta M = -0.03$, $\Delta SD = 0.03$) and groups of size 3 ($p \approx 0$, $\Delta M = -0.02$, $\Delta SD = 0.012$).

6.2.3 Time spent in modes.

For CH, participants spent 10.4% of the time in *idle mode*, 34.1% in *translation mode*, and 55.5% in *rotation mode* (7.2% for the x-axis, 7.5% for the y-axis, 7.7% for the z-axis, and 33.1% without performing rotations).

For HB, participants spent 14.1% of the time in *idle mode*, 59.1% in *transformation mode*, and 26.8% in *constrained rotation mode.*

For GO, participants spent 60.4% of the time in *idle mode*, 23.3% in *transformation mode*, and 16.3% in *translation mode.*

6.3 Subjective Ratings

The scores for NASA-TLX were 35/100 ($SD = 32.7$) for HB, 38/100 ($SD = 32.8$) for CH, and 39.9/100 ($SD = 30$) for GO. A low score means that the task did not offer a meaningful mental and physical effort to the user.

The usability questionnaire is shown in Table 2. Statistical tests did not reveal any significant effects.

The ranking of techniques preference was analyzed using a Friedman test ($\chi 2 = 5{:}44$, $df = 2$, $p > 0{:}05$) and it revealed no significant effects. The ranking was 1.61 ($SD = 0.18$) for HB, 2 ($SD = 0.16$) for CH, and 2.39 ($SD = 0.2$) for GO.

7 Discussion

This section examines the results of the evaluations and it is divided into two parts. The former analyzes the effects of the different design choices. The latter focuses on the evaluated techniques.

7.1 Implications of the Design Choices

7.1.1 Should the technique use one or two hands? Regarding physical fatigue, one might think that two-hands techniques are the most tiresome, as both hands must be coordinated and kept up while interacting. However, participants expressed that they felt slightly more fatigued using one-hand techniques. This may be caused by the fact that when users interacted with only one hand, the other hand that was resting served as a point of comparison. As a result, although two-hand techniques may be more physically demanding, the subjective perception of the users is the opposite. In general, another cause of physical fatigue could be the lack of haptic feedback, as users could not rest their hands against the objects. Possibly, other types of feedback such as audio could reduce the perception of fatigue.

The two-hand technique was slower than the one-hand techniques in translation but twice as precise. Possibly, the average of the two hands softened the input sensor error. Consequently, with the current hardware, using two hands implies a slower but more precise translation.

We suggest that applications that need to manipulate a single object at time can use a one-hand technique without decreasing their performance. This recommendation arises from the fact that the Crank Handle performed in general terms as well as the Handle Bar.

7.1.2 Should the technique integrate translation and rotation? Concerning the integration of translation and rotation, it is not possible to completely generalize, as the two techniques that integrated them obtained opposite results: The Grasping Object showed the worst results but the Handle Bar was satisfactory. The poor results of the Grasping Object could be caused by the metaphor. Although it is intuitive, it is also hard to perform precise rotations with it. Nevertheless, the necessity of adding an only-translation mode, the inefficiency (see Figure 7), and the *m*-metrics (see Figure 9) suggest that the combination of rotation and translation had a negative effect in the Grasping Object.

Even though people could use up to 6 DOF in the Grasping Object technique, *m*-metric revealed that they tended to separate rotation and translation. These results coincide with Masliah et al.'s (2000) and Veit et al.'s (2009) conclusions. We suggest that as one hand represents a 3-DOF input, it should be assigned to manipulate no more than 3 DOF at the same time. Therefore, if we focus on one-hand techniques, we may argue that separation between translation and rotation is beneficial. For two-hand techniques, it is possible to satisfactorily integrate translation and rotation as the two hands represent a 6-DOF input.

7.1.3 Should the technique decompose rotation by axis? Attending to the decomposition of the rotation in primary axes, the Crank Handle (decomposes) and the Handle Bar (not decompose) had similar results. However, results imply that the Crank Handle is better for rotation around one primary axis, whereas the Handle Bar is better for rotations around a random axis (see Figure 6). Additionally, although both techniques had similar completion times and precision at rotations, the Crank Handle was more efficient (see Figure 7). This result suggests that decomposing rotation in primary axes could be more efficient. Veit et al. (2009) obtained a similar result. Nonetheless, not decomposing could be more useful for free exploration of objects or other tasks that require less efficiency (Kruger et al., 2005).

7.1.4 Could the technique be described with a metaphor? We argue that using a metaphor is always beneficial for a technique. This assertion is supported by the literature review, the pilot study, and the feedback of the users. Nonetheless, although metaphors always facilitate the interaction with a technique by making it more intuitive, they do not guarantee high performance. That was the case with the Grasping Object metaphor.

Quantifications of the metaphors' adequacy could not be made, as the subjective ratings did not reveal significant differences. Nevertheless, participants' comments coincided in indicating the Handle Bar as the most intuitive metaphor. This preference is reflected on the ranking but not in the questionnaire.

7.2 Other Implications

The *m*-metric analysis revealed that participants had a better control of translation on the plane parallel to the TV (XY). Consequently, if the manipulation technique requires manipulating only one or two DOF, it could be better to employ the X and Y position of the hand as input instead of the Z position.

The transformation coordination (see Figure 8) shows that participants started adjusting the rotation of the object and then the translation. Those results are opposite to Martinet, Casiez, and Grisoni (2012), who compared three manipulation techniques on multi-touch surfaces. We suggest that the interaction patterns of users in docking tasks are different in mid-air from those of multi-touch interactions. This fact obstructs to some point the generalization of our implications to other interaction modalities different from mid-air.

7.3 Analysis of the Techniques

In general, the Crank Handle performed as well as the Handle Bar in terms of accuracy, task completion time, and maximum precision. The Grasping Object was entirely outperformed. The results showed that the Crank Handle is operative and even within the same range of performance as the two-hand technique. The following subsections analyze individually each technique.

7.3.1 Crank Handle technique.

The time passed in each rotation mode was similar and rotations were equally difficult around all the axes. We detected that up to 33% of the time, participants unsuccessfully tried to perform a rotation. We plotted the 3D hand positions and detected that in those cases, the traced circle was irregular. The algorithm should be improved to be more adaptable at detecting circles.

7.3.2 Grasping Object technique.

Participants passed 60% of the time in the *idle mode*. We observed that during a significant amount of this time the users were targeting the ray to the desired point or thinking about from where to grasp the object. *m*-metrics revealed that rotations around the *z*-axis had a lower performance than the rest. In fact, *z*-axis manipulation in free rotations like Arcball (Shoemake, 1992) is a known issue and techniques are usually extended with specific gestures to mitigate this problem (Iacolina et al., 2011).

7.3.3 Handle Bar technique.

In the *m*-metrics charts, we observed that the Handle Bar is the only technique that uniformly uses all DOF combinations. Similarly, the Handle Bar is the closest technique to the optimal coordinated manipulation of translation and rotation (see Figure 8). *m*-metrics reported that rotations around the *x*-axis were slightly worse than in the rest of the axes. This kind of rotation is difficult to perform without entering the *Constrained Rotation mode*.

8 Conclusions

In this paper we collected four design choices for manipulation techniques. We classified the existing mid-air manipulation techniques according to these design choices. The classification revealed deficiencies in the one-hand techniques, such as the lack of interaction metaphors and the absence of techniques that decompose rotation. Consequently, we developed two one-hand manipulation techniques to address these absences.

We conducted an evaluation that compared the two developed one-hand techniques and an existing two-hand technique. These techniques can be classified attending to the design choices of using one or two hands, separating or integrating translation and rotation, and decomposing or not rotation per axis. Additionally, all the techniques had a metaphor associated with the interaction.

The selection and evaluation of the techniques were designed to analyze the implications of the design choices. Consequently, the results gave insights on how the design choices influence the performance of a technique. For instance, separating translation from rotation and decomposing rotation in primary axis improves efficiency. The priority of the design choices will depend on the final use of the technique. For example, modeling software may take advantage of an accurate manipulation technique on rotations, whereas video games may require only ease of use.

Additionally, results revealed that the Crank Handle, a novel one-hand technique, performs similarly to a two-hand technique such as the Handle Bar. Therefore, the Crank Handle technique can be used when it is required to manipulate only one object, leaving the other hand available for different tasks or resting. Finally, the results can also be used to select among the three presented techniques and as a point of comparison for future techniques.

The analysis of the techniques also showed points to improve. The Crank Handle should improve the algorithm of circle detection. Additionally, the Grasping Object would take advantage from a specific gesture for *z* rotation. Similarly, the Handle Bar could be improved with a more rapid access to *x* rotation. In the future, the new version of the Kinect will be able to detect wrist rotation and tilt. This would provide more input DOF that could be used to improve the interaction techniques.

Acknowledgments

We are grateful to all the participants who spent their time on our experiments. This work has been partially supported by the EU Alfa project Gaviota (DCI-ALA/19.09.01/10/ 21526/ 245-654/ALFA 111(2010)149). We also thank Amaia Zulet for her great labor at proofreading. Bossavit B. and Marzo A. have grants from the Universidad Pública de Navarra and Gobierno de Navarra, respectively.

References

Bettio, F., Giachetti, A., Gobbetti, E., Marton, F., & Pintore, G. (2007). A practical vision based approach to unencumbered direct spatial manipulation in virtual worlds. *Eurographics Italian Chapter Conference,* pp. 145–150.

Bourke, P. (September 2014). Circles and spheres. Retrieved from http://paulbourke.net/geometry /circlesphere/

Bowman, D. A., Kruijff, E., LaViola, J., & Poupyrev, I. (2001). An introduction to 3-D user interface design. *Presence: Teleoperators and Virtual Environments, 10*(1), 96–108.

Bowman, D. A., Gabbard, J. L., & Hix, D. (2002). A survey of usability evaluation in virtual environments: Classification and comparison of methods. *Presence: Teleoperators and Virtual Environments, 11*(4), 404–424.

Bowman, D. A., Coquillart, S., Froehlich, B., Hirose, M., Kitamura, Y., Kiyokawa, K., & Stuerzlinger, W. (2008). 3d user interfaces: New directions and perspectives. *IEEE Computer Graphics and Applications, 28*(6), 20–36.

Bowman, D. A., McMahan, R. P., & Ragan, E. D. (2012). Questioning naturalism in 3D user interfaces. *Communications of the ACM, 55*(9), 78–88.

Buxton, W., & Myers, B. (1986). A study in two-handed input. *Proceedings of CHI '86,* pp. 321–326.

Chen, M., Joy Mountford, S., & Sellen, A. (1988). A study in interactive 3-d rotation using 2-d control devices. *Proceedings of SIGGRAPH '88,* pp. 121–129.

Fishkin, K. P. (2004). A taxonomy for and analysis of tangible interfaces. *Journal of Personal & Ubiquitous Computing, 8*(5), 347–358.

Froehlich, B., Hochstrate, J., Skuk, V., & Huckauf, A. (2006). The GlobeFish and the GlobeMouse: Two new six degree of freedom input devices for graphics applications. *Proceedings of CHI '06,* pp. 191–199.

Guiard, Y. (1987). Asymmetric division of labor in human skilled bi-manual action: The kinematic chain as a model. *Journal of Motor Behavior, 19*(4), 486–517.

Hackenberg, G., McCall, R., & Broll, W. (2011). Lightweight palm and finger tracking for real-time 3D gesture control. *Proceedings of IEEE VR,* pp. 19–26.

Hancock, M., Carpendale, S., & Cockburn, A. (2007). Shallow-depth 3d interaction: Design and evaluation of one-, two- and three-touch techniques. *Proceedings of CHI '07,* pp. 1147–1156.

Hart, S., & Staveland, L. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Human Mental Workload,* 139–183.

Iacolina, S. A., Soro, A., & Scateni, R. (2011). Natural exploration of 3D models. *Proceedings of CHI '11,* pp. 118–121.

Jacob, R. J. K., Sibert, L., McFarlane, D., & Mullen, P. (1994). Integrality and separability of input devices. *ACM Transactions on Computer–Human Interaction, 1*(1), 3–26.

Kabbash, P., Buxton, W., & Sellen, A. (1994) Two-handed input in a compound task. *Proceedings of CHI '94,* pp. 417–423.

Kruger, R., Carpendale, S., Scott, S. D., & Tang, A. (2005) Fluid integration of rotation and translation. *Proceedings of CHI '05,* pp. 601–610.

Leganchuk, A., Zhai, S., & Buxton, W. (1998). Manual and cognitive benefits of two-handed input: An experimental study. *ACM Transactions on Computer–Human Interaction, 5*(4), 326–359.

Lu, G., Shark, L.-K., Hall, G., & Zeshan, U. (2009). Dynamic hand gesture tracking and recognition for real-time immersive virtual object manipulation. *Proceedings of International Conference of Cyberworlds,* pp. 29–35. Bradford, West Yorkshire.

Martinet, A., Casiez, G., & Grisoni, L. (2012). Integrality and separability of multi-touch interaction techniques in 3D manipulation tasks. *IEEE Transactions on Visualization and Computer Graphics, 18*(3), 369–380.

Masliah, M., & Milgram, P. (2000). Measuring the allocation of control in a 6 degree-of-freedom docking experiment. *Proceedings of CHI '00,* pp. 25–32.

Mine, M. R., Brooks, Jr., F. P., & Sequin, C. H. (1997). Moving objects in space: Exploiting proprioception in virtual-environment interaction. *Proceedings of SIGGRAPH '97,* pp. 19–26.

Nancel, M., Wagner, J., Pietriga, E., Chapuis, O., & Mackay, W. (2011). Mid-air pan-and-zoom on wall-sized displays. *Proceedings of CHI '11,* pp. 177–186.

O'Hagan, R. G., Zelinsky, A., & Rougeaux, S. (2002). Visual gesture interfaces for visual environments. *Journal of Interacting with Computers, 14*(3), 231–250.

Owen, R., Kurtenbach, G., Fitzmaurice, G., Baudel, T., & Buxton, B. (2005). When it gets more difficult, use both hands: Exploring bimanual curve manipulation. *Proceedings of GI '05,* pp. 17–24.

Parsons, L. (1995). Inability to reason about an object's orientation using an axis and angle of rotation. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 1259–1277.

Poupyrev, I., & Ichikawa, T. (1999). Manipulating objects in virtual worlds: Categorization and empirical evaluation of interaction techniques. *Journal of Visual Languages and Computing, 10*(1), 19–35.

Raj, M., Creem-Regehr, S. H., Rand, K. M., Stefanucci, J. K., & Thompson, W. B. (2012). Kinect based 3D object manipulation on a desktop display. *Proceedings of SAP '12,* pp. 99–102.

Schlattmann, M., & Klein, R. (2009). Efficient bimanual symmetric 3D manipulation for markerless hand-tracking. *Proceedings of VR '09.*

Seay, A. F., Krum, D. M., Hodges, L. F., & Ribarsky, W. (2000). *Direct manipulation on the virtual workbench: Two hands aren't always better than one.* (Technical Report GIT-GVU-00-07). Georgia Institute of Technology.

Segen, J., & Kumar, S. (1998). GestureVR: Vision-based 3D hand interface for spatial interaction. *Proceedings of ACM Multimedia '98,* pp. 455–464.

Shank, G., & Gebler, C. (2002). Six metaphors in search of the internet. *Journal of Teaching and Learning, 17*(1), 43–52.

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

Song, P., Goh, W. B., Hutama, W., Fu, C.-W., & Liu, X. (2012). A handle bar metaphor for virtual object manipulation with mid-air interaction. *Proceedings of CHI '12,* pp. 1297–1306.

Veit, M., Capobianco, A., & Bechann, D. (2009). Influence of degrees of freedom's manipulation on performances during orientation tasks in virtual reality environments. *Proceedings of VRST '09,* pp. 51–58.

Wang, R., Paris, S., & Popovic, J. (2011). 6D hands: Markerless hand tracking for computer aided design. *Proceedings of UIST '11,* pp. 549–558.

Wang, Y., Mackenzie, C. L., Summers, V. A., & Booth, K. S. (1998). The structure of object transportation and orientation in human-computer interaction. *Proceedings of CHI '98,* pp. 312–319.

Zhai, S., & Milgram, P. (1998). Quantifying coordination in multiple DOF movement and its application to evaluating 6 DOF input devices. *Proceedings of CHI '98,* pp. 320–327.