DOF Separation in Mid-air 3D Object Manipulation

Filipe Relvas
filiperelvas@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

Abstract

Object manipulation is a key feature in almost every virtual environment. However, it is difficult to accurately place an object in immersive virtual environments using mid-air gestures that mimic interactions in the physical world, although being a direct and natural approach. Previous research studied mouse and touch based interfaces concluding that separation of degrees-of-freedom (DOF) led to improved results. In this work, we assess the impact of explicit 6 DOF separation in mid-air manipulation tasks. We implemented a technique based on familiar virtual widgets that allow single DOF control, and compared it against a direct approach and PRISM, which dynamically adjusts the ratio between hand and object motions. Our results suggest that full DOF separation benefits precision in spatial manipulations, at the cost of additional time for complex tasks. From these results we drew some guidelines for object manipulation in mid-air. Based on the guidelines we developed a technique, WISDOM, that combined the positive aspects of the three techniques used before. Our technique provides direct manipulation and scaled movements while maintaining DOF separation offered by our Widgets implementation. In order to validate our proposal we conducted another user evaluation in order to compare it against 6DOF and Widgets. While WISDOM didn't reduce task time completion, it improved position and rotation error in translation and rotation only tasks and more complex tasks that required both translation and rotation simultaneously compared to the other techniques.

Keywords: 3D User Interfaces, Immersive Virtual Environments, Spatial Interactions, Mid-air Object Manipulation, DOF Separation

1. Introduction

We are currently witnessing a huge interest in virtual reality (VR), mainly due to the recent technological advances that made head-mounted displays (HMD) affordable and widely available. Immersive virtual environments (IVE) that were made possible with such technologies are being used for several purposes, like engineering, architecture, game development and so forth, offering unique capabilities. To interact within these virtual environments (VE), the ability to manipulate virtual objects is a key feature. While direct approaches that mimic interactions in the physical world are the most natural, it is still difficult to accurately place a virtual object with the desired transformations. These difficulties may arise from different factors, such as limited human dexterity for mid-air gestures and lack of precision from tracking systems.

In this work we study the impact of DOF separation in mid-air object manipulation for improved object manipulation, comparing three manipulation techniques based on existing literature. We draw some guidelines from the assessment and propose a technique that combines the distinct aspects of each one of the three techniques.

2. Related Work

Being of such importance, object manipulation in VEs has been subject of research for long, covering different kinds of interaction paradigms. To overcome the difference between input and output DOF, most mouse-based techniques rely on widgets, which reduce simultaneous DOF being controlled [9, 4, 18]. These techniques attained such popularity they are still used in common commercial applications. Distinctly from mouse techniques, touch surfaces allow users to directly touch objects displayed. Aiming for more natural interactions, researchers proposed techniques for controlling several DOF at the same time [8, 17]. Nonetheless, reduction of simultaneous DOF controlled have been later suggested [11] and followed by several authors. Eden [10] and iTouchIt [13] resorted to a direct drag approach to translate objects, while separating horizontal and vertical movements, and perform rotations through virtual widgets. tBox [3], a 3D transformation widget that appears as a wireframe bounding box, favors independent 9DOF control. Similarly, Gimbal Box [2] also uses a cube shaped widget to separate DOF manipulation, and its authors concluded that adapted widgets are superior
to other approaches for multi-touch interactions. Semi-immersive stereoscopic tabletops present different challenges, since imagery appears on a volumetric space. Benko and Feiner [1] decomposed 3DOF tasks into a set of 2DOF and 1DOF tasks, using a balloon metaphor. Triangle cursor [20] follows a similar approach but uses two touches to define a triangle’s base, with a cursor in the top vertex. To manipulate virtual objects in full 9DOF, Toucheo [7] relies on widgets on the multi-touch surface.

Having an input with higher DOF, most current mid-air approaches for 3D virtual object manipulation try to mimic physical world interactions [21, 19]. However, fine-grained manipulation and precision tasks are hard to perform with these techniques. To overcome this, Frees et al. [6] proposed PRISM, that scales down hands’ movement to increase precision. Switching between precise and direct mode occurs according to the current velocity of the hand. When moving an object from one general place to another, the user is not necessarily interested in being precise and moves relatively rapidly. When users are focused on accurately moving an object to very specific locations, they normally slow their hand movements down and focus more on being precise. It increases the control/display ratio, which causes the object to move more slowly than users hand, reducing the effect of hand instability and creating an offset between the object and the hand. User evaluations results show faster performance and higher user preference for PRISM over a traditional direct approach for translation tasks. For object rotation, PRISM uses the angular speed of the hand, which authors concluded to be confusing to users. Osawa [15] proposed a position adjustment that consists in a scale factor for slowing hand movement, similar to PRISM, and viewpoint adjustment, that automatically approaches the viewpoint to the grabbed point so that the object being manipulated appears larger. Through a user evaluation, these techniques showed improvements for small targets. The 7-Handle manipulation technique [14] consists of a triangle shaped widget with seven points. User evaluation results showed that 7-Handle is only better suited than the traditional direct 6 DOF approach for manipulating large objects.

In short, research showed that for mouse and touch based interfaces, even in stereoscopic scenarios, DOF separation through widgets led to better performance. Such approaches allow to clearly and undoubtedly select transformations and axes. To improve users’ accuracy in mid-air interactions, researchers already tried to scale down hand motions or move the viewpoint closer to the object, but without regard to DOF separation. Approaches based on virtual widgets have already been proposed, but did not achieve promising results, possibly because they are very different from those used in other interfaces, being more complex and not using common reference frames, such as object or world axes. We believe that clear DOF separation in mid-air, using familiar virtual widgets, might improve users’ performance in object manipulation tasks.

3. Assessment of DOF separation in mid-air
Since DOF separation showed positive results in both mouse and touch interaction for virtual 3D object manipulation, we conducted a user evaluation to assess if it also benefits spatial interactions in IVEs. We implemented three techniques based on the literature. The first is a direct approach in which all transformations performed by the user’s hand are directly applied to the object, the second follows scaled transformations based on the user movement’s speed, and the third consists of spatial widgets for separating DOF. All implemented techniques provide 6 DOF transformations: three for translation and three for rotation.

6DOF or Direct Manipulation To mimic interactions with physical objects as closely as possible, direct manipulation uses all 6 DOF information from users’ hands [21]. It is often used as a baseline for evaluations of other techniques [6, 12, 14]. This technique consists of grabbing an object directly, moving it to a new location and/or rotating it, and then releasing. After being grabbed, the object directly follows the movement of the hand dragging changes object’s position and wrist’s rotation controls object’s rotation. All transformations are simultaneously applied to the object, as pictured in Figure 1b. The grabbed point in the object will remain the center of all transformations during the entire manipulation, until the object is released.

PRISM We implemented the PRISM technique as presented by Frees et al. [5]. This technique aims in improving accuracy of direct manipulation, switching between a precise and a direct mode according to the current speed of users’ hands. Hand’s movement in each coordinate axis is scaled down.
when users move their hands slower than a pre-defined threshold (SC) in that axis. We used the threshold value proposed by the original authors. This scaling results in an offset between the hand and the object being manipulated, that can be canceled by moving hands faster than the same threshold, translating into a 1 to 1 motion. We also included rotations later proposed by same authors [6], which follows the same premise from translations, scaling down slow wrist rotations around each coordinate axis. As suggested by the authors, resulting offsets are represented by a white line for translations, and two sets of axis for rotations. Similarly to 6DOF technique, both translations and rotations can be performed simultaneously, as exemplified in Figure 1c.

Widgets for DOF separation Widget based manipulations are widely used in mouse and keyboard 3D user interfaces. Our widget based approach strictly follows DOF separation. Not only translation and rotation operations are treated independently, users can only manipulate 1 DOF at a time. We used a representation similar to that introduced in [4], illustrated in Figure 2. Users can grab the sphere connected to the desired axis and move the hand along the axis to trigger object translation. For rotations, the approach is similar, but the hand movement is performed around the target axis. The decision to either perform a translation or rotation is made based on the hand’s path after 10 cm. Selected transformation and axis remain locked until a release gesture.

3.1. Methodology
Each user session lasted approximately 45 minutes. We started by introducing the experiment, followed by a brief technique description. Techniques were performed in alternated order, to assure every possible permutation and avoid biased results. For each technique we played a video tutorial after which, participants had a maximum training period of 3 minutes, in a dedicated environment. Following this period, participants were asked to perform a set of tasks. After completing each technique’s tasks, participants fulfilled a questionnaire regarding distinct aspects of the interaction. The session concluded with a profiling questionnaire.

3.2. Tasks
We requested participants to complete a set of six tasks for each technique. All consisted in a docking task [6, 11, 12], where participants had to put the exhaust pipes in the right place of a car engine. That component of the engine was the only object in our virtual environment that could be grabbed and transformed. Engine's model had a semi-transparent replica of the pipes showing the only possible target position and orientation. To prevent excessively long sessions, each task was limited to a maximum of three minutes. After reaching time limit we informed participants they could stop, and we considered the attained position and orientation as final.

For the first task (Figure 3a), the object to be manipulated begun with the correct position along both YY and ZZ scene axes and orientation, only with an incorrect position according to the X coordinate. Similarly to the previous task, the second task (Figure 3b) object started with the correct orientation, however its position was incorrect along all three coordinates. The third task (Figure 3c) consisted in only rotating the object around the Z axis, while the fourth task (Figure 3d) implied rotation around an arbitrary axis, requiring no translation as well. The fifth task (Figure 3e) required the object to be rotated around the Z axis and translated along both XX and YY axes. Finally, in the last task (Figure 3f), participants had to apply full

![Figure 2: Widgets technique.](image)

(a) Initial State.  (b) Translation.  (c) Rotation.

Figure 3: Tasks performed by participants.

1Original 3D model of the used engine uploaded to Sketchup’s 3D Warehouse by user M-Speed.
6 DOF transformations to the object. Although some tasks required only one kind of transformation (translation or rotation), none was restricted, as we did not intend to modify any technique in order to accommodate a specific task.

3.3. Setup and Prototype

Our setup comprises non-invasive and affordable full body user tracking with three depth cameras Microsoft Kinect v2. One was placed facing the user while the remaining lied on each side, 90 degrees from the first. Since Microsoft Kinect fails in providing reliable hand orientation data, we developed an wireless custom made device to better acquire such data. It uses an Arduino based circuit board, which incorporates an IMU for accurate 3 DOF orientation tracking, and a Bluetooth LE module. We attach the device to the user’s dominant hand using an acrylic clip, which assures it does not fall when the hand is opened. A pressure pad detects if the hand is open or closed. For viewing the VE, we used a Gear VR with a Samsung Galaxy S6, connected via Wi-Fi to our tracking server. We developed our prototype using Unity3D, with gravity and objects’ collision disabled. For improved user feedback while grasping, the object becomes transparent, revealing the penetrating portion of the hand [16]. To guide participants during the tasks, the object gradually turns green as it approaches the target position and orientation (Fig. 3d).

3.4. Results and Discussion

The experiment was performed in our laboratory with a controlled environment, using the setup previously detailed. We counted with the participation of 21 people (5 female), between the ages of 18 and 50 years old, with the great majority (62%) between 18 and 25. Most had at least a BSc degree (86%), while the remainder are finishing it. More than half (52%) had never experienced a VR setting, and 43% use some kind of gesture recognition systems more than once a month, such as XBox Kinect, Wii Remote or Playstation Move. Only 28% of participants use 3D modelling systems at least once a month.

During our experiment, we collected both objective data, through logging mechanisms, and subjective data, asking participants to fill out questionnaires. We used Shapiro-Wilk test to assess data normality. We then ran the repeated measures ANOVA test with a Greenhouse-Geisser correction to find significant differences in normal distributed data, and Friedman non-parametric test with Wilcoxon Signed-Ranks post-hoc test. In both cases, post-hoc tests used Bonferroni correction (corrected sig. = sig. × 3).

3.4.1 Objective Data

We measured time taken by participants to fulfil each task, as well as object placement error, depicted in the graphs of Figure 4. We registered both position error and rotation error.

For the translation only tasks, we found statistically significant differences in completion time (Task 1: $\chi^2(2)=25.368$, $p<.0005$; Task 2: $F(1.611,30.604)=9.025$, $p=.002$). For the first task, post-hoc test revealed Widgets approach (avg=25s) to be faster than both 6DOF (avg=59s, $Z=-3.542$, $p<.0005$) and PRISM (avg=90s, $Z=-3.823$, $p<.0005$), and 6DOF to be faster than PRISM (Z=-3.267, $p=.003$). In the second task, PRISM (avg=102s) was significantly slower than Widgets (avg=49s, $p=.008$) and 6DOF (avg=71s, $p=.028$).

For position error, differences were also found (Task 1: F(1.851,24.066)=17.474, $p<.0005$; Task 2: F(1.359,14.946)=6.653, $p=.015$), with Widgets (Task 1 avg=3.3mm; Task 2 avg=5.2mm) outperforming 6DOF (avg=15.0mm, $p<.0005$) in the first task and PRISM on both first (avg=10.7mm, $p=.002$) and second (avg=12.2mm, $p=.003$) tasks. The technique used also influenced rotation error (Task 1: $\chi^2(2)=24.500$, $p<.0005$; Task 2: $\chi^2(2)=15.000$, $p=.001$), with Widgets (Task 1 avg=0.0, Task 2 avg=0.0) achieving lower error than 6DOF (Task 1: avg=11.7, $Z=-3.724$, $p<.0005$; Task 2: avg=9.8 and PRISM (Task 1: avg=7.3, $Z=-3.408$, $p=.003$; Task 2: 7.1, $Z=-2.803$, $p=.015$).

Widgets might have outperformed both 6DOF and PRISM in the first task, due to its DOF separation. Since this task required translating the object along a single axis, the ability to manipulate with such constraint allowed users to avoid unexpected rotations and translations, thus preventing error. The same principle applies to time completion, because users did not need to correct mistakes. Similarly, the second task saw better results with Widgets in both translation and rotation error, although the time taken by users had no significant difference against 6DOF. We believe this occurred because transformation separation found in the Widgets technique made it impossible to take a direct path, requiring users to move in all three axes separately.

In the second pair of tasks we focused on rotations. Significant differences for execution time were only found for the third task ($\chi^2(2)=20.985$, $p<.0005$), in which the use of Widgets (avg=27s) reduced time needed when compared to 6DOF (avg=53s, $Z=-3.053$, $p=.006$) and PRISM (avg=58s, $Z=-3.823$, $p<.0005$). For both tasks, position error revealed significant differences according to the technique used (Task 3: $\chi^2(2)=16.545$, $p<.0005$; Task 4: F(1.619,14.575)=6.586, $p=.012$). Widgets (Task
3 avg=0.9mm, Task 4 avg=9.7mm) led to better positioning than 6DOF in both tasks (Task 3: avg=13.3mm, Z=-3.296, p=.003; Task 4: avg=15.7mm, p=.008) and than PRISM in the third task (avg=16.6mm, Z=-3.059, p=.006). Rotation error was also significantly affected by the techniques (Task 3: \(\chi^2(2)=20.118\), p<.0005, Task 4: \(\chi^2(2)=16.545\), p<.0005). Once again, Widgets (Task 3 avg=1.8, Task 4 avg=5.3) performed better than 6DOF in both tasks (Task 3: avg=8.8, Z=-3.547, p<.0005; Task 4: avg=8.7, Z=-2.868, p=.012) and than PRISM in the third task (avg=7.1, Z=-3.574, p<.0005). Alike the first pair, third and fourth tasks revealed advantageous results for Widgets in both translation and rotation error. Even though the focus of these tasks shifted from translation to rotation only, the ability to separate transformations proved to be, once again, significant. The increased completion time found in the fourth task was a consequence of rotations around all axes. Users felt confused and unable to easily figure out the necessary rotations to reach the desired orientation.

The last pair of tasks required both translations and rotations. In both cases, techniques had an effect on the time participants took to complete tasks (Task 5: F(1.422,27.021)=12.645, p<.0005; Task 6: F(1.671,23.391)=5.232, p=.017). Widgets (avg=6.6mm) reduced distance to target in the fifth task when compared to 6DOF (avg=15.1mm, Z=-2.809, p=.015) and PRISM (avg=21.4mm, Z=-3.010, p=.009). In the last task, 6DOF (avg=11.4mm) allowed users to place the object closer to its target position than PRISM (avg=21.2mm, p=.048). Analysing rotation error, we only found significant differences in the fifth task (\(\chi^2(2)=22.625\), p<.0005), in which Widgets (avg=1.1) attained better results than 6DOF (avg=9.2, Z=-3.823, p<.0005) and PRISM (avg=8.9, Z=-3.464, p=.003). Final tasks had an increase in complexity, since they required participants to apply translations and rotations to the object. Time taken to complete these tasks was negatively affected due to the necessary increased number of operations. As a consequence, translation and rotation error presented worse results when compared to previous tasks, because the time limit prevented participants to make final adjustments.

It is also worth of notice that both 6DOF and PRISM did not have major variations along all tasks, with no regard to its difficulty. For these techniques, after grabbing an object all tasks are alike, since there is no constraint in transformations being applied to the object. Taking the first and last task as an example, we used a Paired-Samples T Test and no significant differences were found in both tasks, there were differences regarding error in object positioning (Task 5: \(\chi^2(2)=8.533\), p=.014, Task 6: F(1.671,23.391)=5.232, p=.017). Widgets (avg=6.6mm) reduced distance to target in the fifth task when compared to 6DOF (avg=15.1mm, Z=-2.809, p=.015) and PRISM (avg=21.4mm, Z=-3.010, p=.009). In the last task, 6DOF (avg=11.4mm) allowed users to place the object closer to its target position than PRISM (avg=21.2mm, p=.048). Analysing rotation error, we only found significant differences in the fifth task (\(\chi^2(2)=22.625\), p<.0005), in which Widgets (avg=1.1) attained better results than 6DOF (avg=9.2, Z=-3.823, p<.0005) and PRISM (avg=8.9, Z=-3.464, p=.003). Final tasks had an increase in complexity, since they required participants to apply translations and rotations to the object. Time taken to complete these tasks was negatively affected due to the necessary increased number of operations. As a consequence, translation and rotation error presented worse results when compared to previous tasks, because the time limit prevented participants to make final adjustments.

It is also worth of notice that both 6DOF and PRISM did not have major variations along all tasks, with no regard to its difficulty. For these techniques, after grabbing an object all tasks are alike, since there is no constraint in transformations being applied to the object. Taking the first and last task as an example, we used a Paired-Samples T Test and no significant differences were found in both tasks, there were differences regarding error in object positioning (Task 5: \(\chi^2(2)=8.533\), p=.014, Task 6: F(1.671,23.391)=5.232, p=.017). Widgets (avg=6.6mm) reduced distance to target in the fifth task when compared to 6DOF (avg=15.1mm, Z=-2.809, p=.015) and PRISM (avg=21.4mm, Z=-3.010, p=.009). In the last task, 6DOF (avg=11.4mm) allowed users to place the object closer to its target position than PRISM (avg=21.2mm, p=.048). Analysing rotation error, we only found significant differences in the fifth task (\(\chi^2(2)=22.625\), p<.0005), in which Widgets (avg=1.1) attained better results than 6DOF (avg=9.2, Z=-3.823, p<.0005) and PRISM (avg=8.9, Z=-3.464, p=.003). Final tasks had an increase in complexity, since they required participants to apply translations and rotations to the object. Time taken to complete these tasks was negatively affected due to the necessary increased number of operations. As a consequence, translation and rotation error presented worse results when compared to previous tasks, because the time limit prevented participants to make final adjustments.

3.4.2 Subjective Data

We asked the participants how they felt about each technique. This included general easiness of use, translation and rotation difficulty and fun factor. A Likert Scale from 1 to 5 was used to answer our questions, being 5 the favorable value. Answers are depicted in Table 1.

Analysing attained results, we identified significant differences in ease of use (\(\chi^2(2)=19.547\), p<.0005), rotation difficulty (\(\chi^2(2)=25.352\), p<.0005) and fun factor (\(\chi^2(2)=13.216\), p=.001). Participants strongly agreed that PRISM was generally harder (Widgets: Z=-3.716, p<.0005, 6DOF: Z=-3.157, p=.003; Task 6: F(1.671,23.391)=5.232, p=.017). Widgets (avg=13.3mm, Z=-3.920, p<.0005) and PRISM (avg=112s, Z=-2.520, p=.014) achieved faster results than 6DOF (avg=9.2, Z=-3.823, p<.0005) and PRISM (avg=8.9, Z=-3.464, p=.003). Widgets and PRISM consistently shared similar results. As the authors pointed out, PRISM rotations were confusing for some users, which might have had a negative impact in tasks overall performance.
to perform object rotation than 6DOF (Z=-2.863, p=.012) and PRISM (Z=-3.874, p<.0005). Also, participants agreed that it is easier to rotate objects using 6DOF than PRISM (Z=-2.708, p=.021). There was no difference in translation difficulty, even though PRISM sacrifices directness and time over enhanced precision. The Widgets approach, although requiring more effort for complex movements, was as appealing as other techniques, being as much as direct manipulation, but with increased final placement.

3.5. Guidelines for Mid-air Object Manipulation

As a result of our evaluation, we were able to draw some guidelines for object manipulation in IVE. These should aid researchers and developers in creating better techniques that can combine the better aspects of each evaluated approach:

- Direct manipulation (6DOF) is well suited for coarse transformations. It allows fast and natural interactions, although not offering accurate placement;

- It should be possible to perform translation and rotation operations independently. We found that, in both 6DOF and PRISM, unwanted transformations happen when a simple translation or rotation is in order, which negatively impacts performance;

- Single DOF separation is very desirable for precise transformations, typically for fine-grain adjustments. This separation, more than separating translation and rotation, constrains transformations to a single dimension, preventing additional unwanted actions;

- Scaled transformations, as proposed in PRISM, are appealing only for translation. Separated scaled rotation in each coordinate axis confused participants, but they found scaled translations to be helpful in improving accuracy. Combining scaled translations with other approaches might improve their overall performance.

4. Proposed approach - WISDOM

Based on the previous results and the guidelines we drew, our proposal improves upon the three techniques described before by combining the strongest aspects of each one. WISDOM provides the direct manipulation found in 6DOF and the scaled movements we observed in PRISM, while maintaining DOF separation offered by our widget implementation.

4.1. Technique overview

Users can switch between different manipulation modes by enabling or disabling widgets at any moment. This behaviour can be achieved thanks to the pressure pad on our custom device, that we use to identify two different pressure levels. Direct object manipulation is set off by grabbing the desired object when widgets are disabled, while a similar behaviour with widgets enabled, triggers isolated scaled translation similar to PRISM [5]. Scaled movements are also applied during widget translations, allowing for fine-grained adjustments after users lock the desired axis.

An additional step to the transformation decision process was added as well. Previously users needed to cover a distance of ten centimeters after which a decision would be made based on the hand’s path. In our proposal the decision distance to either perform a translation or rotation was increased to fifteen centimeters and decision regions were added to every available handle. These regions are three-dimensional cone-shaped objects not visible to users, that allow decisions to be made without the necessity to describe a certain hand path. In addition we implemented uniform and non-uniform object scaling, a feature not present in most mid-air techniques, and 2D TRS (Fig. 5). Users can trigger scaling and TRS by selecting two handles from the same axis or from different axes respectively.

5. WISDOM Validation

In order to validate our proposal we conducted another user evaluation, where we compared it against 6DOF and Widgets. We followed the same method-

![Image](image.png)

(a) Grab two handles from different axes.
(b) Translate along the plane.
(c) Rotate around the (d) Scale along both plane normal. axes.

Figure 5: 2D TRS.
ology described in Subsection 3.1 with a minor difference in the session duration, which lasted approximately 40 minutes. The setup and prototype can also be found in Subsection 3.3.

5.1. Tasks
Participants were asked to complete a set of four tasks for each technique. All consisted in a docking task, where participants had to place a Tetris L-shaped piece in the same position and orientation of a transparent copy. Only the original piece could be grabbed and transformed in our virtual environment. Each task was limited to a maximum of three minutes in order to prevent excessively long sessions. After reaching time limit we informed participants they could stop, and we considered the attained position and orientation as final.

For the first task (Figure 6a), the object to be manipulated begun with the correct orientation, but with a wrong position along all three coordinates. The second task (Figure 6b) implied only rotation around an arbitrary axis, as the object was already in the right position. In the third task (Figure 6c) users were required to rotate the object around the Y axis and translated along both XX and ZZ axes. Finally, in the last task (Figure 6d), participants had to apply full 6 DOF transformations to the object. Although some tasks required only one kind of transformation, none was restricted.

5.2. Results and Discussion
The experiment was performed in our laboratory with a controlled environment, using the setup detailed in Subsection 3.3. We counted with the participation of 20 people (3 female), between the ages of 18 and 40 years old, with the great majority (65%) between 18 and 25. Most had at least a BSc degree (80%), while the remainder are finishing it. More than half (60%) rarely or have never experienced a VR setting, and 45% never used any kind of gesture recognition systems, such as XBox Kinect, Wii Remote or Playstation Move. Only 25% of participants use 3D modelling systems at least once a month.

We collected both objective data, through logging mechanisms, and subjective data, asking participants to fill out questionnaires, during our experiment. We used Shapiro-Wilk test to assess data normality. We then ran the repeated measures ANOVA test with a Greenhouse-Geisser correction to find significant differences in normal distributed data, and Friedman non-parametric test with Wilcoxon Signed-Ranks post-hoc test. In both cases, post-hoc tests used Bonferroni correction (corrected sig. = sig. × 3).

5.2.1 Objective Data
We measured time taken by participants to fulfil each task, as well as object placement error, depicted in the graphs of Figure 7. We registered both position error and rotation error.

For the translation only and rotation only tasks, we found statistically significant differences in completion time (Task 1: $\chi^2(2)=14.282$, $p=.001$; Task 2: $\chi^2(2)=14.282$, $p=.001$). For the first task, post-hoc test revealed 6DOF approach (avg=53.7s) to be faster than both Widgets (avg=108.4s, Z=-3.310, $p=.003$) and WISDOM (avg=117.8s, Z=-3.354, $p=.003$). In the second task, 6DOF (avg=49.1s) was also faster than Widgets (avg=107s, Z=-3.332, $p=.003$) and WISDOM (avg=113.5s, Z=-3.332, $p=.003$). The technique used also influenced rotation error in the first task ($\chi^2(2)=11.804$, $p=.003$), with 6DOF (avg=5.8) performing worse than Widgets (avg=3, Z=-2.534, $p=.033$) and WISDOM (avg=3.1, Z=-2.900, $p=.012$), and position error in the second task ($\chi^2(2)=7.964$, $p=.019$) with WISDOM (avg=9.6mm) outperforming 6DOF (avg=16.3mm, Z=-2.605, $p=.027$).

Even though 6DOF was the faster approach in both tasks, it didn’t achieve the same level of precision regarding translation and rotation error. The first task, which could be completed without applying any rotation to the object, showed that separating transformations benefited both Widgets and WISDOM. The same principle applies to the second task, where users were required to rotate the object around an arbitrary axis.Translations were inevitable due to the nature of 6DOF as opposed to the possibility of separating translation from rotation in WISDOM.

The last pair of tasks required both translations and rotations. In both cases, techniques had an effect on the time participants took to complete tasks (Task 3: $\chi^2(2)=18$, $p<.0005$; Task 4: $\chi^2(2)=24.108$, $p<.0005$). 6DOF (avg=56.6s) outperformed Widgets (Task 3: avg=103.4s, Z=-3.623, $p<.0005$; Task 4: avg=146.2s, Z=-3.824, $p<.0005$) and WISDOM (Task 3: avg=106.8s,
analysing attained results, we identified no significant differences in any of the preferences.

Table 2: Participants preference for each technique, regarding different criteria (Median, Inter-quartile Range).

<table>
<thead>
<tr>
<th></th>
<th>6DOF</th>
<th>Widgets</th>
<th>WISDOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easiness</td>
<td>4 (1)</td>
<td>4 (2)</td>
<td>4 (1)</td>
</tr>
<tr>
<td>Translation</td>
<td>4 (1)</td>
<td>3 (2)</td>
<td>4 (1)</td>
</tr>
<tr>
<td>Rotation</td>
<td>4 (2)</td>
<td>4 (2)</td>
<td>4 (1)</td>
</tr>
<tr>
<td>Fun</td>
<td>4 (2)</td>
<td>4 (1)</td>
<td>4 (1)</td>
</tr>
</tbody>
</table>

5.3. Observations
At first we believed WISDOM would present lower completion times from providing both 6DOF and Widgets to users without compromising placement error. Our initial thought was to use 6DOF to place the object closely to the target and then switch to Widgets to increase object placement accuracy. In some cases we observed this didn’t happen, as users took the time initially gained from a more direct manipulation to make adjustments in the end. As we did not impose any specific transformation order, users were able to make a decision based on their needs resulting in different approaches and distinct levels of performance. In more extreme situations, some users even took an exclusive approach to one of the two methods of manipulation, which might have caused some discrepancies in regards to time completion and object placement error.

As we mentioned earlier our technique also provided 2D TRS, which could be used at any time during tasks. However users forgot or didn’t see the benefits of making use of it during object manipulation. It is important to notice that we included this feature in the explanation of our technique, but the large number of options and ways to interact with the object might have dissuaded users from making use of 2D TRS.

6. Conclusions
As we are witnessing a huge interest in virtual reality (VR), mainly due to the recent technological advances that made head-mounted displays (HMD) affordable and widely available, mid-air interactions became a subject of interest in various research topics. Immersive virtual environments (IVE) that
were made possible with such technologies are being used for several purposes, like engineering, architecture, game development and so forth, offering unique capabilities. To interact within those virtual environments (VE), the ability to manipulate virtual objects is a key feature.

Given its importance, object manipulation in virtual environments has been subject of research for long, covering different kinds of interaction paradigms. While direct approaches that mimic interactions in the physical world are the most natural, as the dimensional difference between input and output does not exist, it is still difficult to place a virtual object in the desired place with a high degree of accuracy. These difficulties may arise from different factors, such as limited human dexterity for mid-air gestures and lack of precision from tracking systems. For both mouse and touch interfaces, separation of degrees-of-freedom (DOF) led to better users’ performance when compared to direct approaches, mainly due to the required mapping between the 2D input and 3D output.

In this work, we conducted an evaluation to assess the benefits of DOF separation in mid-air, after it has been proved useful in other interaction paradigms by previous research. We implemented three mid-air object manipulation techniques, one follows a direct approach (6DOF), the second scales users’ movement (PRISM) and the third is our implementation of mid-air virtual handles for DOF separation. We concluded that indeed DOF separation through virtual widgets led to error reduction, at the cost of increased time for more complex tasks. Drawn from our results we drew some guidelines for future work on object manipulation in mid-air.

Following the guidelines we drew from the DOF separation assessment, we proposed a technique called WISDOM that combines different aspects of the techniques we used to assess the importance of DOF separation in mid-air. It provides the direct manipulation found in 6DOF and the scaled movements we observed in PRISM, while maintaining DOF separation offered by our Widgets implementation. In order to validate our proposal we also conducted a user testing in this work. WISDOM improved position and rotation error in some instances compared to the other techniques, but didn’t reduce time completion as opposed to our initial thoughts. It is also the first mid-air technique that provides 1DOF, 2DOF and 3DOF object scaling.

6.1. Future Work
As a result of the work presented in this dissertation, we believe some aspects should be further studied as they might benefit future work.

Our technique originated some improvements namely position and rotation error in certain tasks, but failed at reducing time completion. A distinct application from our guidelines might be an approach based on mid-air hand gestures to specify the desired transformations.

We believe it might be interesting to allow users to define custom arbitrary transformation as an addition to the Widgets approach. By doing so, object manipulation keeps the DOF separation benefits, but reduces substantially the number of operations needed. This will possibly allow for quick transformations in complex tasks.

Our technique provides 1DOF, 2DOF and 3DOF scaling but it wasn’t enabled during user testing, since the techniques we compared it with didn’t provide this capability. It is important to validate scale as it might be a compelling addition for object manipulation in mid-air.

References


