Manipulating 3D Objects in Mid-Air using Virtual Rails

Rui Rodrigo Batista da Cunha e Lorena
rui.rodrigo.lorena@ist.utl.pt

Instituto Superior Técnico, Lisboa, Portugal
May 2017

Abstract
In almost every immersive virtual environment (IVE), object manipulation is a key feature. Nonetheless, it is hard to place an object with precision using mid-air gestures, that mimic interactions in the physical world, without compromising task completion time, although being a direct and natural approach. Previous studies concluded that separation of degrees-of-freedom (DOF) and scaled transformations led to improvements on object placement precision. But this precision benefits increase the operations number, which costs additional completion time for complex tasks. Having in mind previous researches, we believe that transformations over custom axes could help decreasing task completion time. In this work, we proposed and implemented an approach, MAiOR, based on scaled movements over custom axes, allowing DOF separation and direct manipulation. In order to validate our proposal, we conducted a user evaluation, comparing it against two baseline techniques, 6DOF and Widgets. Our results suggest that MAiOR grant a higher object placement precision, but for inexperienced users it might add some difficulty when tasks are complex, i.e., require both translations and rotations.

Keywords: 3D user interface, Mid-air object manipulation, Arbitrary manipulation axis, DOF separation, Virtual reality.

1. Introduction
Virtual Reality (VR) is an Immersive Virtual Environment (IVE) and have been subject of increasing interest in the latest years. The emergence of new affordable commercial Head-Mounted Display (HMD)’s like the Oculus Rift\(^1\) and HTC Vive\(^2\) have allowed the development of techniques and applications that make use of these environments, being Three-Dimensional (3D) object interaction a key feature.

Traditional object manipulation interfaces use a Two-Dimensional (2D) screen, mouse and several keyboard shortcuts for views changing and to constraint settings. Multi-touch devices, have become a part of our daily lives and provide new ways of interacting with 3D objects. Both benefit from Degrees of Freedom (DOF) separation, but are less natural. VR can change the visualization and manipulation paradigm in a more natural way, because 3D input is used to generate 3D output. Still, it is difficult to place a object in the desired place with a high degree of accuracy, due to some technology limitations such as equipment imprecision, hand shaking and lack of haptic feedback or limited human dexterity. A way to increase precision in these environments is using DOF separation. But to execute a complex manipulation in two or more axes, we need to perform multiple operations in a row. This can take longer than manipulating multiple DOFs. Contrarily, using approaches that allows 6-DOF or just 3-DOF transformations can be very natural and be very fast, as we would do in the physical world, at the cost of manipulation precision. So, we pretend to answer the following question:

How to increase object manipulation precision without compromising task completion time?

We focused on the problems that mid-air manipulation techniques currently have. We hypothesize if transformations over custom axes with DOF separation using simple gestures in VR can provide mid-air object manipulation precision. While not increasing the time that users need to complete complex tasks, by making use of 6-DOF input. We also aim to have the option to scale down movements when performing translations and rotation, to provide better fine-grained adjustments.

The main goal of this work is to assess whether mid-air transformations over custom axes increases object manipulation precision without compromising task completion time.

\(^1\)https://www.oculus.com/en-us/rift/
\(^2\)https://www.vive.com/eu/
2. Related Work

Traditional interfaces for 2D interaction usually use mouse and keyboard as indirect manipulation input and a 2D screen for output. In most approaches, users must use keyboard shortcuts and manipulation widgets for selecting a transformation and an axis, to perform simple action over an object. Generally, DOF separation and snapping either through gravity\cite{3} or friction\cite{2} is needed to increase manipulation precision. However, to execute complex manipulations in two or more axes require multiple operations in a row, which can be harder through traditional interfaces.

Multi-touch surfaces allow a more direct approach. Users can interact directly with the object along the surface. The Rotate-Scale-Translate (RST) approach is a standard for 2D manipulation in multi-touch surfaces. It is direct, based on physical models and the way objects behave in real world, which makes it easy to learn and fast to perform. Direct manipulation is fast but may lead to unwanted transformations and offers a low precision. DS3\cite{9} focused DOF separation. Many Techniques tackled this matter through virtual widgets and hand gestures. Users are able to control translation, rotation and scale independently. GimbalBox\cite{4} and tBox\cite{7} use widgets adapted to solve the fat fingers and occlusion, problems that happens when trying to select objects in a surface. Multi Touch Gesture\cite{1} rely only on hand gestures to either apply constrains and manipulate objects. Touch Sketch\cite{17} has a constrain menu to separate DOF, which turn manipulation less natural. LTouchIt\cite{11} provides direct control of the objects position keeping DOF separation. Studies concluded that in many cases, DOF separation is better than DOF integration. Even though it is less natural. This separation benefits tasks that require some degree of precision, as there are less DOF modified simultaneously. Multi-touch interaction is also possible over stereoscopic environments. In Triangle Cursor\cite{16} is a widget-based technique and the perceived space is in 3D, which gives a better perception to users.

Multi-touch 3D interaction has the problem converting 2D input to 3D output. Mid-air manipulation techniques are usually not restricted by the screen-space limitations and grant 6-DOF or 2x3-DOF (3-DOF in each hand) input, which is the best way to simulate object manipulation in real life. Handle-Bar\cite{15} and Spindle-Wheel\cite{6} require that users hold both hands in the air and not only it cause fatigue, but also is difficult to stay still in a place or moving in a unique direction, specially after long use. Mendes et al.\cite{10} demonstrate that DOF separation is helpful again to avoid unwanted manipulations, but this time in mid-air manipulation approaches. However, in a docking task when the goal is to place an object on an exact position. DOF separation is not enough, because it requires fine-grained adjustments. User’s hand shaking can be a major problem is this situation. HOMER\cite{5} and PRISM\cite{8} scaled down dynamically the user’s virtual hand movement reducing the effect of hand instability increasing the precision rate.

Similarly to the approaches previously described in multi-touch environments, virtual widgets allow users to treat translation and rotation separately. 7-Handle\cite{13} benefit from these aspects, but the widget’s design was too complex, leading to a steeper learning curve.

Lastly, WISDOM\cite{14} improved position and rotation error compared to other techniques, but did not reduced task completion time because it had a significant learning curve due to its complexity to new users.

In this work we will have in consideration DOF separation with scaled down movements over custom axes to aim for better object manipulation precision without compromising the task completion time.

3. Proposed Approach: MAiOR

MAiOR has four different methods to interact with objects. Translating only, rotating only, scaling only or directly through a 6-DOF approach. Users can choose any of them regarding the way they interact with the object.

Translation Translating operation consists in grabbing directly the desired object. The object will be constrained to 3-DOF translation. A blue transparent axis is drawn from the object initial position passing through the actual object position (Figure 1(a)). This feature allows to draw an custom translation axis. If the axis has a 10 degrees or less deviation from any world’s axes or object’s axes, it suffers a snap to the closest one and changes its color to green or yellow, respectively. This axis can be selected or locked at any time, this means constrain the object translation to 1-DOF. When it is locked it becomes opaque (Figure 1(b)). The object center point also suffers a snap to coincide with the axis. Thereafter, object translation is done

![](image1.png)  
(a)  
(b)  
Figure 1: MAiOR’s translation axis.
only over the locked axis following the concept of an object on a rail.

We noticed from other studies[5, 8, 14] that scaled movements improve precision rate in fine-grain adjustments. Thus, we implemented a scaled translation mode with fixed scale offset. This feature can be in both 3-DOF and 1-DOF translation states by closing the non-dominant when moving an object, as if the non-dominant hand was applying a friction force. The movement speed will be scaled to a quarter from the actual user’s hand speed. Taking into account that on previous study [12] widgets (the DOF separation approach) average position error was about 4.5 millimeters and we wanted to achieve a 1 millimeter average position error, we calculated our translation scale factor to 0.25. The longer this mode is used, the bigger the offset between the grabbing hand position and the object position will be.

Direct Manipulation As mentioned before, MAiOR features direct manipulation with 6-DOF. From previously studies [12, 14] we can ensure that direct manipulation decrease completion time in tasks that require a bigger number of operations. So that we decided to implement this feature in our approach. Users can access this mode by rotating their hand over than 90 degrees (left or right), while grabbing the object with that same hand (Figure 2). We choose this gesture to unlock the 6-DOF as metaphor of unlocking a door handle, since the object is constrained to 3-DOF right after been grabbed by users (Figure 2(a)). Whenever 6-DOF is unlocked, the object suffers a rotation jump with the same degrees which the grabbing hand performed to do the unlock gesture (Figure 2(b)). This snap guarantees that the object rotation is coherent to the grabbing hand rotation.

![Figure 2: Hand 90° rotation to unlock MAiOR’s 6-DOF approach.](image)

Rotation Rotation follows the same principles from translation. Users start rotating in 3-DOF and then can select a axis to rotate in 1-DOF. Although, rotation has a two previous steps. First, users must close one hand to cast a virtual bar from it. This bar work as an handle to rotate the object. Second, they shall attach the handle to the desired object, this is selecting the object. Thereafter, the object is in 3-DOF rotation. This was based on the Handle-Bar[15], which had good results in previous researches. Contrarily to Handle-Bar, our approach substitutes one of the hands by the object’s center, to keep this point as the rotation center. A blue transparent circumference is shown around the object (Figure 3(a)), illustrating the rotation around the actual rotation axis. Similar to translation, if the rotation axis has a 10 degrees or less deviation from any world’s axes or object’s axes, it suffers a snap to the closest one and changes its color to green or yellow (Figure 3(b)), respectively. Users can seek and select the desired axis, constraining the rotation to 1-DOF along it. When it is locked it becomes opaque. Also, we implemented a wrist rotation, which uses the handle as a rotation axis and the object rotates around it.

Differently from translation, we did not implemented a discrete precision mode in rotation. However is still possible to scale rotation almost in a free way, having in mind the mathematics behind the circular motion. Moving the hand away from the object while grabbing the handle attached to the object, increases the distance between the object and the hand and the handle stretches dynamically to fit in that distance. The longest the distance is, the more the hand must move to rotate the object and slower it will rotate (Figure 4). This concept can improve precision rate for fine-grain adjustments.

![Figure 3: MAiOR’s rotation axis.](image)

![Figure 4: The physics used for MAiOR’s scaled rotation feature. Users can scale a α degrees rotation the longest the distance D is. In this example D’ is greater than D, thus (b) is more scaled than (a).](image)
**Scaling** Also based on the Handle-Bar[15]. It provides uniform and non-uniform scaling. It has three previous steps before scaling the object. Firstly, users have to gently close both hand to cast a virtual cursor (Figure 5(a)). It is always casted on the midpoint between the hands. Secondly, they shall select the desired object, colliding the cursor with it and tightly close one of the hands. The cursor turns into a blue bar that stretches between the hands (Figure 5(b)). That bar is the axis over which the object will be scaled. Similar previous transformation, if the axis is close enough from any world’s axes or object’s axes, it suffers a snap to the closest one and changes its color to green or yellow, respectively. The last step is selecting the axis tightly closing the other hand. Selecting the bar when it is blue performs uniform scaling. Otherwise, non-uniform scaling is activated along the chosen axis. Thereafter, moving both hands apart scales the object up (Figure 5(c)), while moving them together scales it down (Figure 5(d)).

4. **User Evaluation**

To validate our proposed technique, MAiOR, we implemented a prototype and conducted a user evaluation.

4.1. Baseline Mid-air Manipulation Techniques

The two next mid-air manipulation approaches were the baselines to compare our proposed solution, MAiOR. One is a direct manipulation technique, 6DOF, the other is an indirect manipulation technique based on 3D widgets.

**6DOF** In the 6DOF technique, object manipulation is done by grabbing directly the desired object, moving it to another location and/or rotating it and then releasing it. This simulates as closely as possible the interactions with physical objects. Any hand translation or wrist rotation has the same effect on the object (Figure 6). This technique offers no restriction, which means that there is no translation and rotation separation. The grabbed point in the object remains the center of all transformations during the entire manipulation, until the object is released. This approach is often used as a baseline for evaluations of other techniques [10, 12, 14].

**Widgets** A mid-air technique through widgets that provide only DOF separation. This is the same widget approach used in [14], which had good average object placement results (position error was 4.5 millimeters and rotation error was 2.3 degrees). The widget is composed by three axis. Each one is represented by a cylinder shaped object with two handles, spheres on the axis tips. It follows a RGB coding for XYZ axes. Rotations are done by grabbing an axis handle and moving the hand along that same axis. Rotation proceeds in a similar way, however the rotation gesture that allows this operation is always around the axis over which is intended to rotate.

The decision to either perform a translation or rotation (Figure 7), is taken by the algorithm based on the first 10 centimeters from the hands path. After this distance is covered it calculates the dot product between the hands path vector and the handles forward, up and right directions. Then it chooses the maximum absolute value from these calculations and it is used as an estimation to reach a decision. For example, if X axis’ handle is grabbed it will decide between translate over this axis or rotate around Y axis or Z axis. The transformation and axis resulting from that decision remain locked until a release gesture.

![Figure 6: 6DOF translation(a) and rotation(b).](image6.png)

![Figure 7: Widgets translation(a) and rotation(b).](image7.png)
4.2. Prototype

In order to compare the three manipulation techniques and achieve our goals, we developed a prototype that enables head and hand tracking alongside with hand input to interact in IVEs.

**Setup** Our prototype setup comprises non-invasive and affordable head and hands tracking with the HTC Vive setup (Figure 8(a)). The head-set is designed to work in "room scale" technology to turn a room into a 3D space via infra-red sensors. This allows users to navigate naturally, walking around the room and use motion tracked handheld controllers to manipulate objects and experience IVE’s. We placed the HTC Vive sensors about 4 meters from each other defining the room diagonal distance (Figure 8(b)). These sensors track headsets and hand controllers position and orientation, sending data to the computer which the HTC Vive is connected through the HMD. Each controller has a trigger with 10 levels of pressure and a circular track-pad. We use the trigger to detect if the hand is opened (no pressure), gently closed (any pressure level from level 1 to 9) or tightly closed (full pressure, level 10). The last pressure level is perceived by a slight click on the trigger. Also, clicking the circular track-pad is the input to perform the prototype’s undo feature, implemented in all techniques.

**Virtual Environment** The IVE scene is a wide plane area without any scenery or objects to avoid any distraction that may influence the tests. The environment has no gravity force nor object collisions. The virtual floor has a grid and receive object’s shadows to provide a better depth perception and help participants navigate along the area. When participants are immersed in VR, they visualize virtual 3D controllers with the same position and orientation as the physical ones that they are holding. The objects became transparent whenever users insert a controller inside it, and became opaque as soon as it is grabbed/selected in order to provide visual feedback for the grabbing/selection action.

4.3. Methodology

All tests were performed individually, followed the same structure. They could last a maximum of 70 minutes in the worst case. Initially, the experiment was introduced to the participants. After this briefing, they executed a set of six tasks thrice, each time with a different technique. The technique’s order were sorted in Latin square format, to avoid biased results from task repetition. For each approach, the test was split in four steps:

1. Watch a technique demo video;
2. Try the actual technique;
3. Execute six tasks;
4. Answer a questionnaire.

First, the testing approach was demonstrated to the participants through a demo video alongside a recorded explanation to grant that they receive the same exactly information. In the next step the participants had a maximum of 5 minutes to try out the technique and adapt themselves to the environment. Thus, they could freely try any operation on a dummy object and remove any doubt about something that was not well understood on the previous step. In the third step, users had a maximum duration of 15 minutes to execute six tasks. All needed data was gradually and automatically recorded into a file. In the last step, participants fulfilled a questionnaire regarding the tested technique.

After completing those steps for each of the three approaches, we did a briefly evaluation about MAiOR’s scaling feature. Participants had 5 min to try it and answer a questionnaire with their opinion about it. To end the user session, they had to fill up a final questionnaire with their profiling data and their overall opinion about the prototype.

**Tasks** Participants completed a set of six tasks for each technique. This set was based on previous studies[12, 14]. After that, they execute one more task to test the MAiOR’s scaling feature. The set of six tasks consisted in a docking task, which aimed to place a carbon component on its matching docking model in a protein compound (Figure 9(a)). When the carbon component was placed on the docking model within the error boundaries (1 millimeter for position and 1 degree for orientation), its color turns green and the task goal was achieved (Figure 9(b)). We chose this boundaries values, so that we could simulate a high precision docking task and test the precision rate offered by each technique. To avoid long user sessions, each task had a maximum time of two minutes and a half. After reaching time limit we considered the attained position and orientation as final. Also, we registered if the task goal
Figure 9: (a) Protein compound missing a carbon component. (b) Task goal achieved.

was achieved or not. Although some tasks required translation or rotation only, none of those operations were restricted on each task.

Isolated Translations The first task, the carbon needed to be translated only along X axis. In the second one its position was incorrect in X, Y, Z.

Isolated Rotations The third task pretended to rotate the carbon around the Z axis, while the fourth task requested a rotation around X, Y and Z.

Composed Transformations The fifth task required to rotate the object around the Z axis and translated along X and Y axes. On the six task, participants had to apply a full 6-DOF transformation on the component.

Scaling Transformation Users were asked to deform three distinct objects to the shape of the colored object in front of each one of those. Two objects were differently scaled in X, Y, Z axes, the other one was uniform scaled.

4.4. Apparatus and Participants

The user sessions were performed in Instituto Superior Técnico (IST) Taguspark campus, room 1.21, also known as João Lourenço Fernandes Laboratory (Figure 8(b)). This room is equipped with all the gadgets needed to execute the tests: HTC Vive and Desktop computers. This room has restricted access, thus providing a calm and controlled environment without external perturbation.

We performed the test with 24 people, 16 males and 8 female, between the ages of 17 and 33 years old. Most of them had at least a BSc degree (80%), while the remainder are finishing it. Almost everyone never had a VR experience (83%), and 50% play video games at least once a month. However, 70% never used any kind of gesture recognition systems, neither the ones like Xbox Kinect, Nintendo Wiimote or Playstation Move. Only 16% of the participants use 3D modeling systems at least once a month.

5. Results and Discussion

During the user sessions, we gathered objective data through a logging mechanism that was running in parallel with our prototype and collected subjective data from the questionnaires filled up with participant’s opinions. We used Shapiro-Wilk test to assess data normality. Then, if it is a normal distributed data ($p > 0.05$) we ran the repeated measures ANOVA test with a Greenhouse-Geisser correction to find significant differences between the approaches, checking them on the post-hoc table. Otherwise, we ran Friedman non-parametric test with Wilcoxon Signed-Ranks post-hoc test. In both cases, post-hoc tests used Bonferroni correction (corrected sig. = sig. x 3).

5.1. Objective Data

We measured the completion time for each task, both the time that participants were idle and were manipulating the object, the object placement error, position error in millimeters and orientation error in degrees. As well as the tasks success rate.

Completion Time Only for tasks completion time statistical calculations, we just considered the times obtained by the users whom achieved the task goal, i.e., placed the object within the error boundaries and did not reach the time limit (2 minutes and a half). So, we did not find a significant differences between tasks completion times (Task 1: $\chi^2(2)=3$, $p=.223$; Task 2:F(1.236,4.942)=3.993, $p=.1$; Task 3:F(1.684,5.052)=4.598, $p=.076$; Task 4:F(1.1)=2.412, $p=.364$). Nevertheless, on task 2 and 4, we verified a tendency that MAiOR (Task 2:avg=47.87s; Task 4:avg=39.23s) was faster than Widgets (Task 2:avg=92.31s; Task 4:avg=72.58s). Taking into account a pairwise comparison, 8 in 9 and 6 in 6 participants had better performance with MAiOR than Widgets on task 2 and 4, respectively. But this needed to be assured with a more significant number of participants.
Position Error (mm)

<table>
<thead>
<tr>
<th>Task</th>
<th>Widgets</th>
<th>MAiOR</th>
<th>6DOF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Orientation Error (deg)

<table>
<thead>
<tr>
<th>Task</th>
<th>Widgets</th>
<th>MAiOR</th>
<th>6DOF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 11: Position error obtained in the six tasks using the three approaches, in millimeters. The graphic presents the median, first and third inter quartile ranges (boxes) and 95% confidence interval (whiskers).

**Placement Error** Moreover, we found statistical significance in object placement error. Regarding position error (Task 1: $\chi^2(2) = 29.84$, $p = 0$; Task 3: $\chi^2(2) = 20.118$, $p = 0$; Task 4: $\chi^2(2) = 19.5$, $p = 0$; Task 5: $F(1.363, 19.085) = 10.075$, $p = .003$; Task 6: $\chi^2(2) = 9.264$, $p = .01$), post-hoc tests showed that for the first task, 6DOF (avg=4.65mm) was less precise than both Widgets (avg=6.517mm, $Z = -3.015$, $p = 0$) and MAiOR (avg=5.5mm, $Z = 3.921$, $p = 0$). In task 3, 6DOF (avg=3.443mm) still worst than Widgets (avg=4.32mm, $Z = -4.108$, $p = 0$) and MAiOR (avg=4.108, $Z = 4.108$, $p = 0$). In both fourth and fifth tasks, Widgets (Task 4: avg=0mm; Task 5: avg=6.94mm) outperformed 6DOF (Task 4: avg=3.387mm, $Z = -3.725$, $p = 0$; Task 5: avg=2.609mm, $p = .001$) and MAiOR (Task 4: avg=2.938mm, $Z = -2.803$, $p = .015$; Task 5: avg=5mm, $p = .006$). However on the sixth task, 6DOF (avg=2.609mm) proven to be better than Widgets (avg=6.517mm, $Z = -3.015$, $p = .009$) and MAiOR (avg=4.409mm, $Z = -2.521$, $p = .036$). On rotation only tasks, third and fourth, users inducted an additional position error with MAiOR, which means they performed translations. This happened mostly by mistake and oblivion of the undo feature. Yet, separated transformations was also a key feature on this tasks.

Upon orientation error (Task 1: $\chi^2(2) = 27.9$, $p = 0$; Task 3: $\chi^2(2) = 13.875$, $p = .001$; Task 4: $\chi^2(2) = 8$, $p = .018$; Task 5: $\chi^2(2) = 8.4$, $p = .015$). On task one, 6DOF (rate=17%) had a lower success than MAiOR (rate=79%, $Z = -3.638$, $p = 0$) and Widgets (rate=92%, $Z = 4.243$, $p = 0$). On task three, Widgets (rate=75%) outperformed 6DOF (rate=29%, $Z = -3.051$, $p = .006$) and MAiOR (rate=33%, $Z = -2.887$, $p = .012$). On task four and five, Widgets (Task 4: rate=67%; Task 5: rate=54%) were better than 6DOF (Task 4: rate=25%, $Z = -2.673$, $p = .024$) and better than MAiOR (Task 4: rate=42%, $Z = -2.714$, $p = .021$), respectively.

Those values demonstrated that MAiOR’s suc-

Figure 12: Orientation error obtained in the six tasks using the three approaches, in degrees. The graphic presents the median, first and third inter quartile ranges (boxes) and 95% confidence interval (whiskers).

**Task Success Rate** Finally, we analyzed tasks success rate and identified significant differences (Task 1: $\chi^2(2) = 27.9$, $p = 0$; Task 3: $\chi^2(2) = 13.875$, $p = .001$; Task 4: $\chi^2(2) = 8$, $p = .018$; Task 5: $\chi^2(2) = 8.4$, $p = .015$). On task one, 6DOF (rate=17%) had a lower success than MAiOR (rate=79%, $Z = -3.638$, $p = 0$) and Widgets (rate=92%, $Z = 4.243$, $p = 0$). On task three, Widgets (rate=75%) outperformed 6DOF (rate=29%, $Z = -3.051$, $p = .006$) and MAiOR (rate=33%, $Z = -2.887$, $p = .012$). On task four and five, Widgets (Task 4: rate=67%; Task 5: rate=54%) were better than 6DOF (Task 4: rate=25%, $Z = -2.673$, $p = .024$) and better than MAiOR (Task 4: rate=42%, $Z = -2.714$, $p = .021$), respectively.
cess rate decreased as the task complexity increased, almost in a linear proportion. Final tasks had increased complexity, since they required both translations and rotations over the object. The task success percentages showed that for 6DOF it was regardless, because more complex tasks had the same success rate as simpler tasks. We can verify that 6DOF was the best on the last task, but the results were not at the same level as the other tasks. DOF separation harmed the object manipulation with the Widgets and MAiOR techniques. Widgets prevented participants from taking a direct path to the docking location. MAiOR suffered with its own rotation feature complexity by confusing the participants. In both cases, they required to perform several transformations in separated axes, which made it slower to reach the task goal. Perhaps if the task duration was not limited to 2 minutes and a half, Widgets and MAiOR would outperform 6DOF. Widgets had a better overall success, because it had a lower learning curve than MAiOR. Although, MAiOR had achieved similar success as Widgets when the task was based on simple translations as task one.

5.2. Subjective Data
This data was gathered through questionnaires. We asked participants about their experience with each technique. This included general ease of use, fun factor, ease of change object position/orientation, the translation/rotation recall operation and the usage tiredness. In all questions it was given a Likert Scale to answer from 1 to 5, being 5 the favorable value. Answer’s average are in Table 1.

### Manipulation Techniques
We found differences in ease of use ($\chi^2(2)$=6.685, $p=.035$), fun factor ($\chi^2(2)$=7.69, $p=.021$), ease of change object orientation ($\chi^2(2)$=14.427, $p=.001$), translation recall ($\chi^2(2)$=11.541, $p=.003$) and rotation recall ($\chi^2(2)$=20.738, $p=0$). Participants agreed that MAiOR was generally harder to use and less fun than Widgets (ease of use: $Z=-2.623$, $p=.027$; Fun: $Z=-2.599$, $p=.009$). Users indicated that was harder either to perform rotations and to remember it with MAiOR than 6DOF (Ease of change orientation: $Z=-3.407$, $p=.003$; Rotation recall: $Z=-3.578$, $p=0$) or Widgets (Ease of change orientation: $Z=-2.99$, $p=.009$; Rotation recall: $Z=-3.691$, $p=0$). Participants thought that it was easier to remember how to translate objects using Widgets than MAiOR ($Z=-3.246$, $p=.003$). There was no significant difference about ease of change position and usage tiredness. Summing up, participants preferred 6DOF approach. It was fun as Widgets, but slightly easier to rotate objects. According to users opinions, they needed more time to get used to MAiOR.

### MAiOR’s Scaling Feature
As mentioned before, we did a briefly evaluation about MAiOR’s scaling feature. We requested the participants to answer a questionnaire about its usage difficulty, fun factor, ease of change object’s scale and of remember its operation and finally the usage tiredness. The answer’s average are in Table 2. Generally, they had a good impression about this feature, since all averages were above value 4. They also stated that it was a feature easy to understand because it worked like pinching, as they do in smartphones or tablets. But instead of using two fingers to scale the object they used both hands.

5.3. Observations
We observed that on second and fourth tasks, 8 in 9 and 6 in 6 participants had better performance with MAiOR than Widgets, respectively. On task 2, users had several distinct backgrounds. On task 4 they included more experienced users like designers and architects. This could lead us to conclude that MAiOR’s translation feature is widely accessible for common users. On the other side, MAiOR’s rotation feature complexity makes it only understandable for object manipulation experienced users. Which means that who handled well MAiOR approach, tended to perform better with this technique than the others. Although this must be assured with a more significant number of tests. Also, we verified that most people already have difficulties on perceiving rotation motions and whatever metaphor used for it seems to be harder to understand than a direct one (6DOF). So, it is important to explore new rotation metaphors if we want to provide a better understanding of DOF separation for those people. We also observed that participants had difficulties understanding MAiOR’s circumference widget motion. One possible explanation for this may be the fact that wrist rotation was not accounted for that motion calculations.

Besides task completion time, we logged the time when users were manipulating the object, separating either when they were translating, rotating or direct manipulating (6-DOF). People tried MAiOR’s direct manipulation feature for a while.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>MAiOR</th>
<th>SixDof</th>
<th>Widgets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ease of use*</td>
<td>2 (1)</td>
<td>2 (1)</td>
<td>3 (2)</td>
</tr>
<tr>
<td>Fun*</td>
<td>3 (1)</td>
<td>4 (1)</td>
<td>4 (2)</td>
</tr>
<tr>
<td>Ease of change Position</td>
<td>3.5 (2)</td>
<td>4 (2)</td>
<td>4 (2)</td>
</tr>
<tr>
<td>Ease of change Orientation*</td>
<td>2.5 (1)</td>
<td>3.5 (2)</td>
<td>3 (2)</td>
</tr>
<tr>
<td>Translation Recall*</td>
<td>3.5 (1)</td>
<td>5 (1)</td>
<td>5 (1)</td>
</tr>
<tr>
<td>Rotation Recall*</td>
<td>3 (1)</td>
<td>5 (1)</td>
<td>4.5 (1)</td>
</tr>
<tr>
<td>Tiredness</td>
<td>3 (2)</td>
<td>3 (2)</td>
<td>3 (1)</td>
</tr>
</tbody>
</table>

* indicates statistical significance

Table 1: Answers to questionnaires, regarding each criteria (Median, Inter-quartile Range).
Table 2: Answers to MAiOR’s Scaling feature questionnaire, regarding each criteria (Median, Interquartile Range).

<table>
<thead>
<tr>
<th></th>
<th>MAiOR’s Scaling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ease of use</td>
<td>4 (1)</td>
</tr>
<tr>
<td>Fun</td>
<td>4.5 (1)</td>
</tr>
<tr>
<td>Ease of change</td>
<td>4 (1)</td>
</tr>
<tr>
<td>Ease of remember</td>
<td>4 (1)</td>
</tr>
<tr>
<td>Tiredness</td>
<td>5 (1)</td>
</tr>
</tbody>
</table>

Figure 13: The time percentage spent in 3-DOF and in 1-DOF during MAiOR’s translation time in each task.

Figure 14: The time percentage spent in 3-DOF and in 1-DOF during MAiOR’s rotation time in each task.

6. Conclusions and Future Work

In this work, we addressed the challenges identified in mid-air object manipulation. So that we proposed and developed a mid-air manipulation technique, MAiOR, combining a direct manipulation approach with DOF separation through custom axes and scaled translations/rotations in order to reduce task completion time and provide better object placement precision. To validate our proposal we conducted user sessions and compared MAiOR against two baseline approaches, 6DOF and Widgets. MAiOR did not compromise task completion time, but improved placement precision regarding 6DOF and was slightly similar to Widgets on translation isolated tasks. MAiOR’s rotation feature was hard to understand for inexperienced users, perhaps due to its metaphor complexity affecting the learning curve. Contrary to this, MAiOR’s scaling feature was well perceived by users and naturally easy to apply on objects. Although it needed a more rigorous assessment to be validate. Overall, MAiOR has a better trade off between task completion time and placement precision rate than 6DOF and we observed that it surpasses Widgets on more complex isolated transformation tasks. Nonetheless, it must be assured with more tests for experienced users.

Summing up, custom transformation axes (1-DOF) were not so relevant when compared with transformations separation on IVEs. Besides that, they might add some confusion for inexperienced users, when manipulation task needs both translations and rotations. Still, scaled movements combined with DOF separation provide better precision rate than a direct 6-DOF approach in tasks that require only translation or only rotation.

In this work there were some aspects that can help themselves achieving the task goal. For example, they usually held tight the dominant arm using the non dominant arm as a support. They mentioned that was useful to decrease hand-shaking and dominant arm tiredness.

Our prototype’s hand input implementation was based on pressure levels in order to simulate the physical hand opening/closing pressure. Also it is adaptable to other kind of setups[12] avoiding the use of multiple buttons. However, it led participants to perform selection by mistake, because they usually full pressed the controller’s trigger.

Some participants used particular gestures to on the third and fourth tasks, but they preferred transformation separation to manipulate the object. Consequently, they had to perform translations to fix the error caused by direct manipulation. Another evidence showing that they might forgot to use the undo feature. On the more complex tasks, exists an almost perfect balance between time spent on translation and on rotation.

People tended to use transformation separation rather than a direct 6-DOF approach. Although, we verified that they preferred to perform translations and rotations using MAiOR in 3-DOF than in 1-DOF, as shown in figures 13 and 14. Participants found 3-DOF translations faster to use than 1-DOF translation. Since MAiOR provided scaled movements, the precision attained in 3-DOF translation was good enough. For rotations, we observed the same, but it denoted the opposite from previous studies. The use difficulty found in the MAiOR’s rotation metaphor may explain this results. However, is needed more researches to assure the 1-DOF relevance in rotations.

Some participants used particular gestures to...
be considered worthy of improvements or alternatively can originate future work. We describe the following:

Search for another metaphors for rotation. We believe that part of our approach success was depending from its rotation capability, which demonstrated to be too complex to be easily understood at first touch for inexperienced users. Perhaps, searching for new metaphors and testing it with people that have some object manipulation background, such as designers and architects, could bring more conclusive results.

Validate MAiOR’s scaling feature in mid-air. Our technique provides uniform and non-uniform scaling but it was only subjectively tested, since our baseline techniques did not provide this feature. It is important to validate scaling, because when combined with translation and rotation it grants 9-DOF object manipulation in mid-air.

References