STAG: Smart Tools and Applications in Graphics (2023)
K. Lupinetti, F. Banterle, N. Capece, and U. Erra (Editors)

Remote and deviceless manipulation of virtual objects in mixed reality

A. Caputo, R. Bartolomioli and A. Giachetti

University of Verona

Abstract

Deviceless manipulation of virtual objects in mixed reality (MR) environments is technically achievable with the current generation of Head-Mounted Displays (HMDs), as they track finger movements and allow you to use gestures to control the transformation. However, when the object manipulation is performed at some distance, and when the transform includes scaling, it is not obvious how to remap the hand motions over the degrees of freedom of the object. Different solutions have been implemented in software toolkits, but there are still usability issues and a lack of clear guidelines for the interaction design. We present a user study evaluating three solutions for the remote translation, rotation, and scaling of virtual objects in the real environment without using handheld devices. We analyze their usability on the practical task of docking virtual cubes on a tangible shelf from varying distances.

The outcomes of our study show that the usability of the methods is strongly affected by the use of separate or integrated control of the degrees of freedom, by the use of the hands in a symmetric or specialized way, by the visual feedback, and by the previous experience of the users.

CCS Concepts

• ***Human-centered computing*** → ***Gestural input; Interaction devices;***

1. Introduction

An effective and easy-to-use virtual object manipulation is fundamental for the development of Mixed Reality interfaces. Using recent Head-Mounted Displays for Virtual and Augmented Reality (Microsoft Hololens 2, Varjo XR-2, Magic Leap One, Oculus Quest, etc.), object manipulation can be performed in a "natural" way by using hand gestures, as they feature finger-tracking capabilities.

The design and the implementation of the manipulation control, however, are not trivial, as the MR applications often require manipulating virtual things remotely, deviceless interaction cannot provide haptic feedback, hand tracking and gesture recognition are not always reliable and mid-air gesticulation is fatiguing.

Many solutions for the manipulation of virtual objects have been presented in the literature [MCG*19], with quite different characteristics (direct vs indirect, single-handed vs two-handed, integrated vs separated DOFs control), but the outcomes of the studies are often conflicting and may be biased by the use of outdated VR technology. As discussed by Bergström et al. [BDAH21], the evaluation of selection and manipulation methods in VR (and even more in AR) lacks specific guidelines, and no clear design guidelines are available. Bergström et al. [BDAH21] note that most of the literature is focused on the selection issues rather than on the design of novel manipulation methods, and consider the development of manipulation techniques an "important research direction" for the future. In the same paper, they also point out that few studies evaluate manipulation considering depth control and addressing the issues related to depth perception. Considering remote manipulation in mixed reality, no studies are available evaluating manipulation considering the effects of distance and mixed virtual/real word visualization. Yet, this task is common in a variety of mixed reality applications like immersive interior design [Jan19], fabrication [WLK*14], virtual shops [FIO19] and it is certainly useful to evaluate the possible solutions and try to define design guidelines.

Remote manipulation methods based on hand gestures are now available in many mixed-reality applications. The most popular are those created for Hololens 2 and developed with the MRTK [Mic22] toolkit. In these applications, the user can remotely select objects with hand-controlled ray casting and pinch, and then manipulate them with 7DOF transforms using a single-handed direct manipulation metaphor for translation and rotation, coupled with a separated control of rotation and uniform scaling performed by a second hand.

In our work, we considered this default method as a baseline and compared it with a novel remote manipulation solution aimed at replicating the classical 3D manipulation mechanisms performed on 2D interfaces as well as a custom handlebar [SGH*12] implementation. Our implementation of the latter method enforces sym-

DOI: 10.2312/stag.20231290

metric control and affordance with a visualization helping to focus on both hands, as suggested in [BH00]. All the methods tested are based on finger tracking, but present different characteristics (single-handed/two-handed, integrated/separate DOF control) and control mappings, allowing an interesting analysis of the roles of the different options.

The comparison is made with a specifically designed user study where subjects must complete a 7-DOF mixed reality docking task by putting virtual cubic elements in specified niches of a tangible shelf using the different remote manipulation methods. The three manipulation methods are used after a common object selection procedure (the standard remote pinch offered by MRTK 2).

The study follows most of the guidelines suggested in [BDAH21], evaluating the manipulation control only, using cubic elements, and asking to complete transformations involving depth variations. The task is low-level, and the physical setting is fixed and controlled. The task mimics a real activity on an augmented scene originally designed for an end-user augmented fair application. The user-object distances and the size of the manipulated target are both set to be consistent with the room-scale user scenario.

In this experimental context, we aim at answering a few specific research questions:

- **Q1.** Do the selected manipulation methods allow a sufficiently effective remote manipulation in mixed reality?
- **Q2.** Which of these methods is the most effective according to different criteria? (i.e. performance, fatigue, ease of use). Given the methods tested, differences in the measured values may reveal the roles of DOFs separation, symmetry, and visual feedback on the interaction.
- **Q3.** Is the distance from the manipulated object a relevant factor when choosing the manipulation technique?
- **Q4.** Does the previous experience in using 3D manipulation interfaces affect user performances?

The outcomes of the experiment can give really useful insights to improve the usability of mixed reality applications requiring an effective interaction with virtual objects.

2. Related Work and motivations

The manipulation of virtual objects in immersive environments is a widely covered topic in the scientific literature. Mendes et al. [MCG*19] recently published a survey also including manipulation control for non-immersive visualization systems. Some of the methods cited in the survey and a few ones proposed more recently can be used in mixed reality and for distant objects as well, for example, "direct" ones like HOMER [BH97] or handlebar[SGH*12], widget-based methods [BMA*14; CEG18], proxy-based techniques like Vodoo Doll[PSP99], Poros [PLMH21] or the method proposed in [KRSH22]). The use of these methods in mixed reality presents, however, non-negligible problems. Some of them provide only 6DOF control, not handling scaling. The use of hands-free setups requires accurate tracking of the hand pose to enable 6DOF direct manipulation, which is not always possible with all systems or devices. Proxy-based methods are not optimal for remotely controlling the pose of an object in a mixed environment, as the attention would need to be split between the proxy and the target location to perform the task.

In our work, we compare different approaches for 7DOF remote manipulation of virtual objects in a real environment on a docking task. Both the scenario and the task are rather common in practical applications, based on existing mixed reality tools. In particular, we work with the most popular setup, based on the Hololens 2 HMD, using the MRTK 2 toolkit [Mic22] for app development.

The default deviceless interaction method in the MRTK 2 supports a 7-DOF manipulation mixing single-handed "direct" control of translation and rotation (6DOF) and bimanual solutions for the control of rotation and uniform scaling. While it provides redundant control options and flexibility, and while potentially supporting both integration and separation of DOFs [JSMM94], it also suggests a well-defined separation of the roles of the two hands, with a paradigm related to the Kinematic Chain proposed in [Gui87]. However, according to previous studies on Virtual Reality, DOF separation may improve accuracy, but at a cost of an increased execution time, [MRFJ16]. A symmetric bimanual manipulation, as proposed in the handlebar metaphor demonstrated good usability in previous studies [SGH*12; BGG*07], and this is consistent with the findings in [JSMM94] on the fact that integral tasks are performed better with integrated controls. But, as pointed out in [BH00], to enforce the symmetry in the interaction, and thus the integration of the DOFs of translation, rotation, and scaling, visual integration of the controls in the field of view is mandatory.

For this reason, as a first alternative to the default MRTK, we implemented a handlebar interface with integrated symmetric 7-DOFs control and a clear visual integration and affordance (see Figure 4).

The second alternative we designed is based on widgets and tries to mimic the remote manipulation used in 3D editors for desktops, using an arcball-like [Sho92; CG15] control for rotation, a direct mapping for the translation and a slider control for scaling.

While the docking task and comparisons of single-handed, bimanual, and widget-based 7DOF manipulation solutions have been presented in the literature, these experiments did not compare these particular solutions and were performed with completely different setups. For example, in [MFA*14] different manipulation methods are tested on a tabletop visualization with a custom hand-tracking system. In [CW15], precision-grasp 6DOF isotonic input devices are exploited as input devices. In [BMA*14] widget-based solutions like the Crank Handle (CH) and the Grasping Object (GO) are compared with a Handlebar [SGH*12] implementation. This paper also features an interesting discussion about the inconsistent findings of previous works investigating human preferences for single-handed/two-handed methods and integrated/separated DOFs. Our work is strictly related to these contributions, but presents a relevant amount of novelty:

- We use off-the-shelf technology used in many end-user applications, comparing alternative solutions to the default manipulation method employed in most interfaces. This results in useful guidelines for practical interaction design with the current MR technology, and that is not always the case for studies based on old tracking techniques and visualization tools.

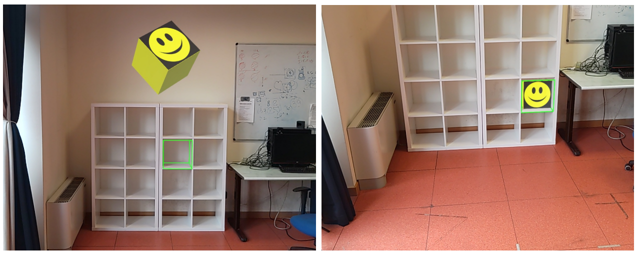


Figure 1: *The mixed environment of our task as seen from the Hololens 2 glasses: the virtual cube appears on top of the shelf (left) and should be rotated resized and inserted in the highlighted shelf compartment (right)*

- We consider remote manipulation only, evaluating the effect of the distance.
- We test the manipulation in a mixed reality environment, with a task involving both the virtual and the real objects. Few studies in the literature tested manipulation methods in this context. An attempt to find guidelines for manipulation in mixed reality has been presented in [KSP20], where the authors compared three approaches for manipulating virtual objects in a real environment. The study, however, was focused on discoverability, including selection, and the task did not consider docking objects in the real scene. The method that was found most usable (World in Miniature) is not suitable for our task, as the real part of the scene would need to be scanned and accurately duplicated in miniature in real time. Another method tested in this work was direct manipulation, which is not usable remotely.
- We focus on the evaluation of symmetric bimanual manipulation with visual feedback enforcing integration against asymmetric bimanual/single-handed "direct" control with DOFs separation. The outcomes of previous studies on these aspects, made on old setups and testing different solutions, resulted in contrasting outcomes [BMA*14; MCG*19]. It is, therefore, important to find novel insights into this.
- We propose a novel indirect manipulation method explicitly exploiting a desktop manipulation metaphor, which can benefit from the re-use of existing skills acquired on different interfaces.

3. Study design

We built our study upon a mixed-reality application designed for marketing purposes in an augmented shop scenario. The app was created with Unity [Uni] and the MRTK2 toolkit [Mic22] and runs on a Hololens 2 headset. It uses, as a real-world reference, a simple shelf with cubic cells of fixed sizes where products are displayed, and the users can interact with them.

For the study we designed a docking task where the subjects have to slot objects inside the cells, to evaluate the usability of different deviceless 7-DOF manipulation methods regardless of the selection action. After the launch, a textured cube appears on the top of the shelf featuring a preset rotation and scale (Figure 1, left), and the subjects had to select it and place it in a specific and visually highlighted cell of the shelf (Figure 1, right). After a docking completion, a new cube appears again on the top with a different pre-defined orientation, and the process is iterated. The docking is considered completed by the application when the cube is correctly slotted in the target cell with the smiley face pointing outwards, facing the user with position and orientation errors lower than fixed thresholds. The placement accuracy is considered sufficient if the distance between the centers of the cube and the cell is lower than 5 cm, the angle between the cube and the target orientation lower than 15° and the difference between the volumes less than 10 % of the cell volume.

The subjects had to repeat a task consisting of three cube docking actions starting from the same initial positions and orientations with the same target cell in all the experimental conditions tested, defined by two independent variables.

The first is the manipulation method, as the goal of the work is to derive guidelines for interaction design. We tested the three methods detailed in Section 4. The second is the distance from the shelf, to assess its effects on the usability of the methods. We considered two reasonable distance values for room-scale remote manipulation (near=1.6m, far=3.2m) as similarly done in previous work [WHB*18].

Each task has been repeated twice in each condition so that each subject completed 36 docking actions. The order of the executions with the different conditions was programmed with a Latin square scheme to avoid biases. Before the experiment, we clearly explained the manipulation methods to the subject, who had two minutes to practice with each of them without testing the docking itself.

For each task execution we measured several dependent variables. A first set is obtained directly from the application log, e.g.,

- the time required to complete the task, estimated with a stopwatch started with the first manipulation of the cube and stopped when the cube is positioned correctly;
- the number of basic movements, that is incremented each time the user starts a new manipulation action on the cube;
- the accumulated translation of the cube, calculated by adding, at each frame, the distance between the position of the center of the cube in the previous frame and its position in the current one;
- the accumulated translation of each hand, calculated in the same way considering the coordinates of the palms;
- the accumulated rotation of the cube, estimated by adding the angle differences between the cube's orientations (expressed in quaternions).

Other dependent variables were collected with questionnaires administered to the subjects after the completion of the tasks.

Each questionnaire included 12 5-point Likert scale questions. Two of them asked to assign a score from 1 (minimum) to 5 (maximum) to the difficulty in completing the task and the fatigue felt with each method. The other ten were taken from the SUS questionnaire [Bro*96], and asked to give a rate on a Likert scale from 1(totally disagree) to 5(totally agree) the agreement with the following sentences:

- I think that I would like to use this system frequently.
- I found the system unnecessarily complex.
- I thought the system was easy to use.

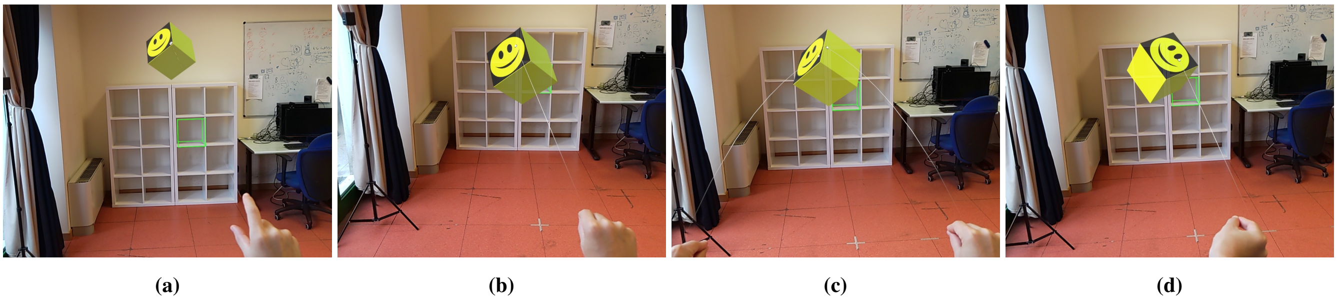


Figure 2: *Task performed with the default MRTK 2 implementation. (a) the object is selected. (b) the object is translated by mapping the hand displacement on the object's one. (c) Scaling is performed with the second hand by pinching and mapping the ratio between the hands' distance and the initial one over the scale factor. (d) The rotation is directly mapped from the dominant hand pose, or, in case of activation of the second hand, can be controlled by changing the orientation of the line joining the two hands.*

- I think that I would need the support of a technical person to be able to use this system.
- I found the various functions in this system were well integrated.
- I thought there was too much inconsistency in this system.
- I would imagine that most people would learn to use this system very quickly.
- I found the system very cumbersome to use.
- I felt very confident using the system.
- I needed to learn a lot of things before I could get going with this system.

4. Manipulation methods' details

We decided to compare three different remote manipulation methods. The first is the default solution featured in MRTK 2 (Mixed-Reality Toolkit 2), which can be considered a variation of the HOMER technique for translation/rotation with the addition of bimanual rotation/scaling. The second is a novel technique that tries to resemble a desktop-like manipulation approach, adapted to the immersive environment. The third is an implementation of the handlebar metaphor [SGH*12].

4.1. Default MRTK2 manipulation (MRTK)

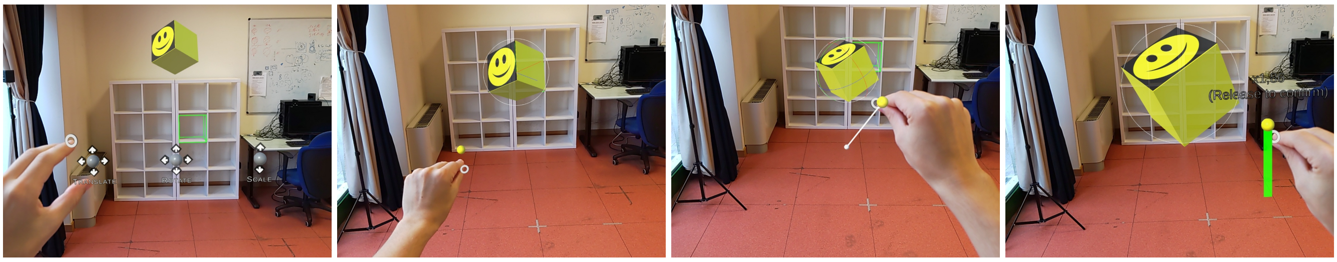
This method [Mic22] features a single-handed control of translation and rotation following the HOMER paradigm, e.g. making the object move coherently with the hand used for the selection.

The translation control features a scaled mapping where the hand motion is multiplied by a factor equal to the ratio between the object-head distance and the distance between the original hand position and the current one (Figure 2 (b)). The library provides two options for the remote rotation: around the center or around the hit point. We chose the second one as it makes the object move as if it is held by the hand, and resulted more intuitive in the preliminary tests. When the pinch with the second hand is executed, the user can also rotate the object according to the orientation change of the vector joining the hands' centers (Figure 2 (c)). This means that he can actually choose between the single-handed and the two-handed rotation control. With this bimanual solution, the user could, in principle, use an integrated (and symmetric) DOF control, even if the interface seems to suggest a DOF separation. The uniform scaling (Figure 2 (d)) is enabled by executing a pinch gesture with the second hand and then moving it. We apply a scaling factor equal to the ratio between the current distance between the hands and the initial one.

4.2. Desktop metaphor (DTP)

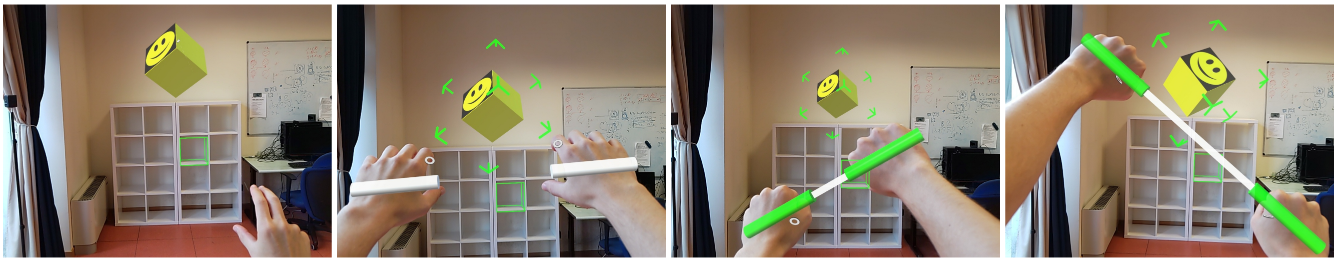
As the manipulation is performed remotely, we consider the use of a desktop-like metaphor exploiting a clear DOF separation. The idea is to replicate the controls used in 3D editors for desktop computers, with separate handling of translation and scaling. A single hand is used like the cursor in those applications, with the only difference being that its position is not constrained to the view plane, and we can use this fact to improve the rotation control. The user, after the selection action, is prompted with three buttons (Figure 3 (a)), and by pinching one of them, he can start the specific manipulation type. The translation is then directly remapped from the hand motion (Figure 3 (b)); the scaling is proportional to the displacement from the pinched point. (Figure 3 (d)).

For the rotation, the idea is to employ an Arcball-style control, projecting the motion of the hand coordinates over a virtual sphere to determine the axis and the angle of rotation of the object (Figure 3 (c)). However, instead of mapping the 2D position projected from the viewpoint onto the sphere as done in the several variations of Arcball [Sho92], the mapping here is done in 3D on a sphere with a center vertically aligned to the cursor's starting position. The solution is similar to the one proposed in [CEG18] for an indirect, widget-based rotation method for VR. Using the starting position and the current position it is possible to calculate the axis and the angle of rotation to be applied to the object. Like in the standard Arcball, the user is not able to manipulate all the degrees of freedom, but thanks to the 3D movement of the cursor he is able to obtain better control with respect to a simple porting of the 2D mapping as we assessed in preliminary testing. The lack of the third rotational DOF can cause the necessity of splitting the rotation into more steps, but this is done efficiently in desktop manipulation.



(a) (b) (c) (d)

Figure 3: *Task performed with the desktop-inspired option: after the selection, mid-air buttons appear (a) allowing the separate control of translation (b), rotation (c) and scaling (d) with remapped single-hand translation.*



(a) (b) (c) (d)

Figure 4: *Task performed with the handlebar (HB) implementation: (a) the object is selected with the pinch. (b) the handlebar appears in front of the user, that can grab them (c). Scale, translation, and rotation are then mapped from the hands' distance, average position, and relative orientation (d).*

4.3. Handlebar (HB)

By grabbing a virtual "handlebar" at its extrema, it is possible to intuitively ans simultaneously apply translation, rotation, and scaling with a total of 7DOF just controlling the hands' position. This mechanism is similar to the two-handed option of the MRTK library. The only difference in the mapping is that in the complete handlebar metaphor implementation it is possible also to rotate around the bar axis. However, the MRTK option doesn't fully exploit the affordance provided by the metaphor, not visualizing the "handles" to be grabbed and not enforcing the integral control of the DOFs.

In our implementation, when the user selects an object with the standard ray cast and pinch method, the object is highlighted, and two virtual handles appear near the user (Figure 4). Performing a grab on them, the user can apply different transformations: the translation is estimated from the translation of the hands' midpoint, with a scaled mapping similar to the one used in the MRTK solution. The orientation change is mapped from the rotation of the bar. To obtain a 3-DOF orientation control, the rotation around the bar axis is also estimated from the hands' poses. The uniform scaling factor is proportional to the hands' distance.

5. Results

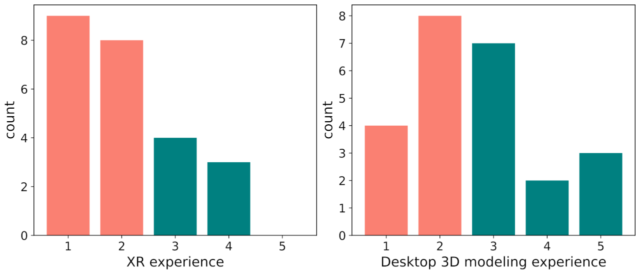


Figure 5: *Distribution of self-reported previous experience of the subjects of our user study with MR and 3D editing applications on desktop platforms. The colors represent the groups of "experts" (green) and "non-experts" (orange) considered in Section 5.3.*

Our test was completed by 24 subjects aged 22-36. The subjects featured different levels of previous experience (from 1 - no experience to 5 - extensive) in the use of MR systems and desktop software for 3D object manipulation, as shown in Figure 5.

5.1. Task completion statistics

Figure 6 shows the distribution of the completion times obtained with the three methods at both distances. We use a combination of violin plot and swarm plot to fully represent the distribution of the values, being data non-normal, as clearly shown by a Shapiro-Wilk test (p=8.16×10^{-38}). The horizontal lines in the plots represent the median values. The plots show that MRTK and DTP resulted in similar performances on average, even if the distributions are dif-

	median values (standard dev.)					
	time (s)	nmov	acc transl (m)	acc rot (deg)	LHM (m)	RHM (m)
MRTK far	26.7(17.9)	7(3.6)	7.8(4.1)	427(307)	2.9(2.4)	4.1(2.7)
MRTK near	25.0(19.2)	7(3.8)	5.6(3.6)	489(350)	3.4(3.2)	4.2(3.2)
MRTK all	25.6(18.6)	7(3.7)	6.7(4.0)	467(331)	3.1(2.9)	4.1(3.0)
DTP far	25.1 (18.5)	8(5.0)	3.3(1.7)	274(545)	0(1.3)	3.5(3.4)
DTP near	26.5(35.4)	8(7.0)	2.4(2.0)	314(620)	0(1.5)	4.1(6.1)
DTP all	26.2 (28.3)	8(6.0)	2.8(1.9)	284(894)	0(1.4)	3.7(5.0)
HB far	12.5 (10.3)	3(3.1)	3.3(2.1)	296(368)	1.9(1.8)	1.8(1.8)
HB near	12.9 (9.4)	3(3.8)	2.7(1.5)	344(527)	2.1(1.8)	2.1(1.8)
HB all	12.8 (9.9)	3(2.4)	3.1(1.8)	318(455)	2.0(1.8)	1.9(1.8)

Table 1: *Median values collected for the repeated docking tasks performed by all the subjects with the different methods at the different distances. The values of the hand motions have been estimated only on the right-handed subjects.*

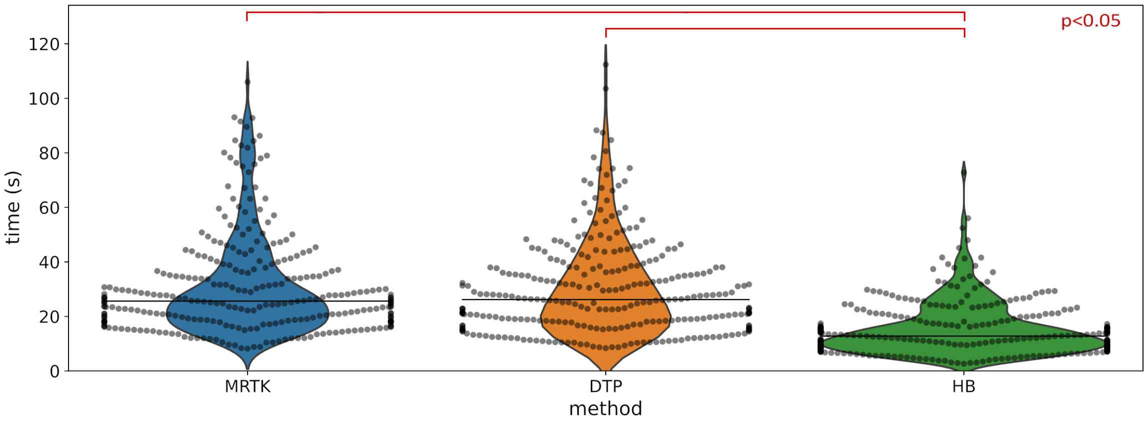


Figure 6: *Violin plots and swarm plots representing the distribution of the completion times for the three methods at the two different distances. Horizontal lines represent the median values (red lines indicate statistically significant differences).*

ferent, while the HB technique allowed most users to obtain faster docking by a large margin. Table 1 shows the median values (and the standard deviations) of the different measurements performed in the experiments with the different control methods and at the two distances. We report the values of the accumulated left/right-hand motions estimated on right-handed subjects.

The advantages of the HB method are evident, and also confirmed by the statistics: a Friedman test revealed that the methods cannot be considered equivalent (eff.size $= 0.33$, p$= 4.96 \times 10^{-42}$). Post-hoc pairwise comparisons performed with Wilcoxon rank-sum test with Bonferroni-Holm corrections show that that the null hypothesis (same median) is clearly rejected for all the comparisons between HB and other methods (eff.size> 0.81, $p < 10^{-32}$ for all pairs), while the same-median hypothesis between MRTK and DTP cannot be rejected even at the 5% confidence level.

The violin plots also show that the distributions of the times are highly overlapped and not regular, indicating that different subjects may be faster with different methods.

A clear rejection of the same median hypothesis is found comparing the number of moves performed to reach the docking (Figure 7). As shown by a Friedman test (eff.size$= 0.61$, p$= 1.41 \times 10^{-76}$), there are significant differences. The pairwise posthoc analysis shows that DTP requires more moves than HB (eff.size $= 0.93$, $p < 10^{-40}$) and MRTK (eff.size $= 0.27$, $p < 10^{-4}$)

The fact that the number of moves with HB is significantly lower than the one recorded with other techniques ($p < 10^{-40}$) may be surprising as MRTK can in principle exploit DOF integration and bimanual manipulation as HB and the latter technique requires an additional action for the handles grabbing.

The relevant difference measured indicates that the MRTK interface suggests a DOF separation with the split of the manipulation action in several steps and does not suggest an integrated, bimanual control.

Table 1 shows numerical values of the medians of all the collected data, including completion time, the number of moves, the accumulated object translation and rotation during the task, and the accumulated translation of left and right hands. It is possible to see several interesting things: the displacement of the objects is significantly higher from the higher distance (3.2m) with all the methods, and the accumulated displacement recorded with MRTK is significantly higher than those measured with the other methods ($p < 10^{-46}$ for all pairs with MRTK). This could result in higher times or perceived fatigue, but the task completion times do not confirm this hypothesis. The total displacement obtained with DTP

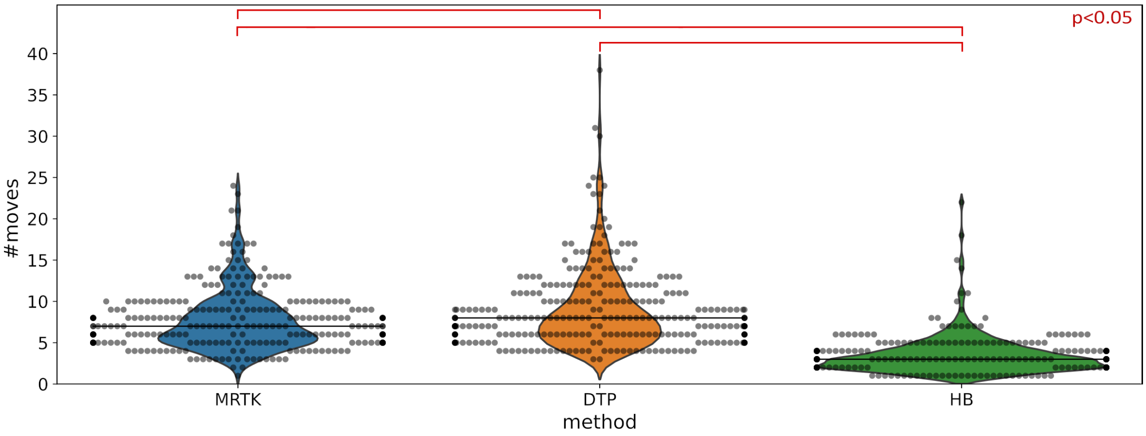


Figure 7: *Violin plots and swarm plots representing the distribution of the number of basic movement steps recorded for the three methods at both distances. Horizontal lines represent the median values (red lines indicate statistically significant differences).*

is significantly lower than that obtained with HB (p= 2.2×10^{-4}). With DTP, the accumulated rotation is significantly smaller compared with the other methods ($p < 10^{-7}$ for all pairs with DTP), while we cannot reject the equal median hypothesis in the comparison between MRTK and HB.

Another interesting fact is that there are significantly different accumulated movements of the users' hands. In particular, with the HB method, the right hand is moved far less than with the other controls, with $p < 10^{-27}$. This outcome is likely related to the reduced times obtained with this method and should also, in principle, determine reduced fatigue. With HB, the movement of the left and right hands are similar, as expected by the symmetry of the method. The fact that the movements of the secondary hand recorded with MRTK are significantly higher than those recorded with HB shows that the second hand is likely used also for rotation control in a non-symmetric and non-integral way.

Another goal of our experiments was to analyze the effect of the distance on the manipulation performances. Figure 8 shows the comparisons of the distributions of the docking times obtained with the three methods at the two different distances tested: near (1.6m) and far (3.2m). We cannot reject the hypothesis of equal medians at a confidence level of 0.001 as shown by a Wilcoxon rank sum test, even if some differences in the distribution are clearly visible.

As shown in Table 1, the accumulated translation of the object is higher when the manipulation is from the highest distance with all the methods. This effect is confirmed by a Wilcoxon test ($p < 10^{-5}$ for all the techniques) but does not influence the task completion times. This fact is likely a consequence of the translation scaling. The accumulated rotation is, instead consistently higher when the manipulation is done from the shortest distance, even if for the statistical test the same median hypothesis can be rejected with a smaller confidence only for MRTK (p= 3×10^{-3}) and Handlebar (p= 9×10^{-4}).

Also the accumulated hands movements are consistently larger for the shortest distance with all the methods, but the statistical test does not show significant effects.

5.2. Questionnaire

For the two general questions on perceived difficulty and fatigue, we did not find any statistically significant difference across the methods. The values had a huge variance for all the methods.

If we consider the SUS scores resulting from the specific questions 9 using the standard formula reported in [Bro*96], it is possible to see that the scores of HB are in general higher, but also that there is a visible outlier that had relevant problems with the DTP method and gave bad scores to all the questions. The average(median) SUS scores are 67.5(70) for MRTK, 56.3(58.8) for DTP, 72.8(73.8) for HB.

We can reject the hypothesis of different medians of SUS by performing a Friedman test ($p = 0.47$). However, it is evident that, while the system is usable for many subjects, some of them experienced relevant issues with all the methods. The SUS score was less than 70 (the average value of evaluation surveys [BKM09]) for 16 subjects with DTP, 12 subjects with MRTK, and 10 subjects with HB. It was less than 50 (acceptability threshold [BKM09]) for 10 subjects with DTP, 4 subjects with MRTK, and 3 subjects with HB. The evidency that selected users have relevant interaction issues is confirmed by the fact that 6 subjects rated the difficulty 4 or 5 using MRTK of DTP.

Looking for statistically significant differences in the single SUS questions, we found, in Friedman tests, only one p-value smaller than 0.05 in the question related to the learnability by the majority of users. The HB method was rated the most learnable, and in the posthoc comparisons with a Wilcoxon rank-sum test and Bonferroni-Holm correction, the null hypothesis of equal medians between HB and DTP is rejected with $p = 0.01$.

The fact that the average SUS score of DTP is lower than that of MRTK despite the similar efficiency is related to the fact that, on average, DTP is perceived as more cumbersome, less integrated, and requiring more learning. The judgment on being cumbersome may be related to the higher number of gestures required, even if this does not affect performances and does not require larger move-

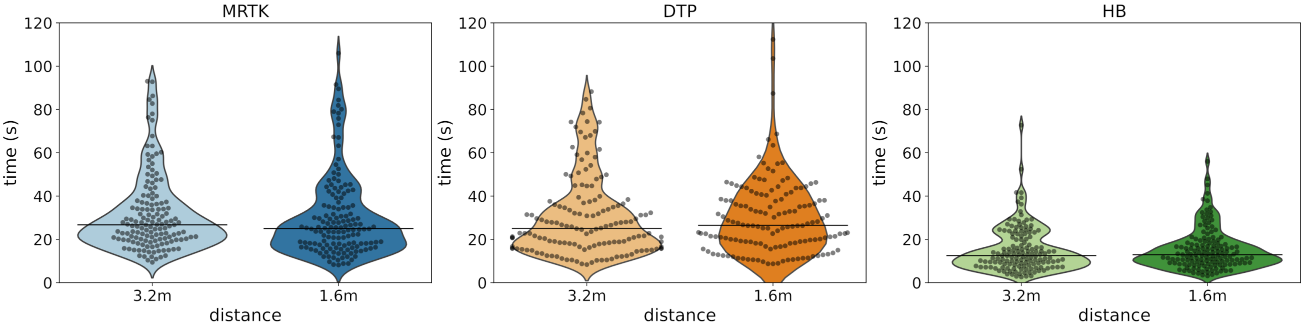


Figure 8: *Violin plots and swarm plots comparing the distribution of the completion times recorded for the three methods at varying distances. Horizontal lines represent the median values*

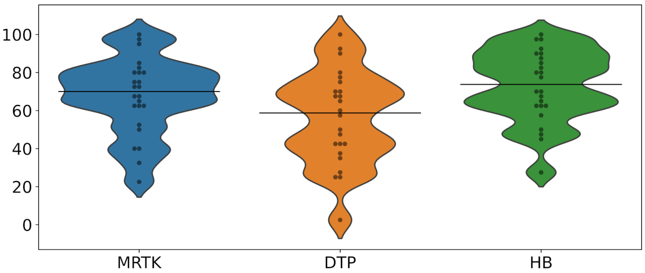


Figure 9: *Distribution of the SUS scores for the different methods. Horizontal lines represent the median values.*

ments. The lower integration is a specific design choice and is not necessarily a negative aspect.

5.3. Experienced users

To understand the effects of previous experience on the manipulation with the different methods, we analyzed the performances of specific groups of subjects. In particular, we compared the performances of the subjects declaring experience with MR systems higher than 2 (MR exp., 7 subjects) with the group of MR-non-experts (MR non-exp., 17 subjects) and the performance of the subjects declaring experience with 3D desktop editors higher than 2 (3D exp., 12 subjects) with the remaining ones (3D non-exp., 12 subjects).

Figures 10 and 11 show the distributions of the task completion times for the subgroups with the different methods. Statistical testing on these groups should be considered with care, as we have small groups with unbalanced execution orders. However, by looking at the plots and the outcome of the tests it is possible to obtain some interesting insights. The difference in the medians between the MR experts and the non-expert groups seems relevant only to the MRTK method. In fact, for this method, there is a non-negligible difference in the medians (21.1s vs. 27.9s), and an unpaired Mann-Withney test results in a low p-value ($p = 3.8 \times 10^{-4}$). For the other methods, p-values in similar comparisons are higher than 0.05. However, the experience seems to result in faster task execution: 25.4s vs. 26.5s for DTP, 11.1s vs. 13.3s for HB. This bias might depend on specific experience with Hololens 2 applications or other VR tools sharing similar interaction modes.

If we look at the subgroups of experts/non-experts of desktop 3D editing, we find that, with the MRTK method, the median time of experts is even higher than that of the non-expert (25.8s vs. 24.5s) and the p-value is high. With the HB techniques, desktop 3D experts seem to have an advantage (11.3s vs. 13.3s), but the p-value is higher than 0.05. With the DTP method there is, instead, a large improvement for the experts (23.1s vs. 28.2s) with a p-value in the unpaired Mann-Withney test equal to 0.0016. This demonstrates that the DTP method effectively re-uses the mental models employed in the desktop interaction to improve the manipulation efficiency in a mid-air interface.

6. Discussion

A large number of mixed reality applications require the remote manipulation of virtual objects, and their success depends heavily on the usability of the transform objects' control. This means that the control should be efficient, effortless, easy to learn, and not fatiguing.

We compared three solutions differing for the following key aspects: integration vs. separation of DOF, use of one/two hands, and direct manipulation vs. use of 3D widgets.

The handlebar control was the most efficient method allowing the subjects to perform the docking tasks of our study in a significantly lower average completion time. While this result was somehow expected, we did not expect such a large difference in the completion times between the handlebar and the default MRTK technique, as the latter handles the scaling in the same way and allows a similar two-handed rotation as well.

This difference suggests not only that the integral control of an integral task [JSMM94] provides increased efficiency but also that the visual enforcement of a symmetric and integrated interaction [BH00] is fundamental to obtain this result.

The number of actions recorded with the MRTK method demonstrates that while the subjects could complete the task in this condition with integrated symmetric control, they preferred to subdivide it into different steps with DOF separation.

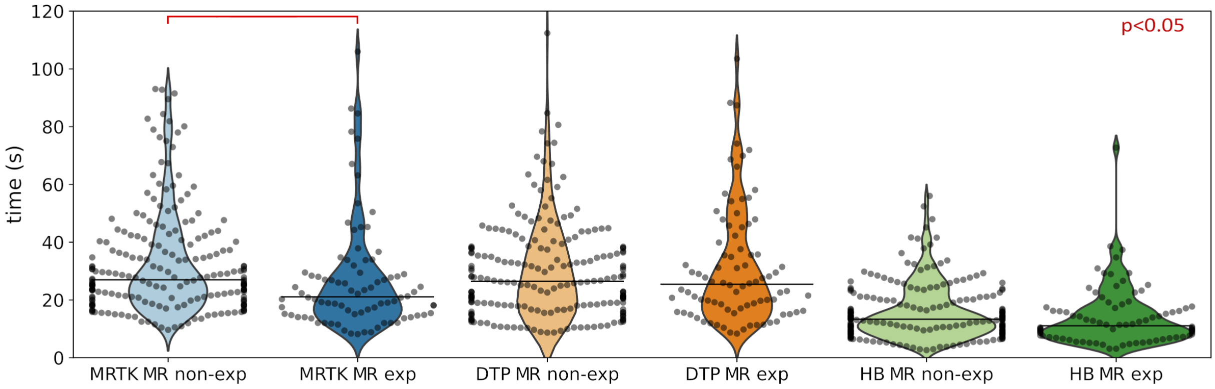


Figure 10: *Distributions of the task completion times separating the groups of MR experts from the groups of non-experts (red line indicates statistically significant difference).*

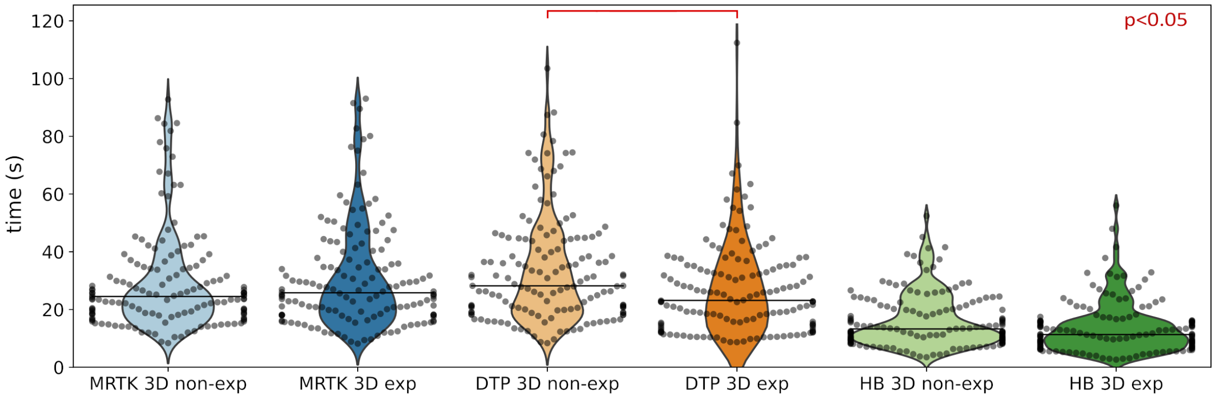


Figure 11: *Distributions of the task completion times separating the groups of experts of desktop software for 3D editing from the groups of non-experts (red line indicates statistically significant difference).*

The method based on the Desktop metaphor resulted in an efficiency comparable to the standard MRTK solution, even if it is perceived as more complex by some users and requires more actions for task completion. We expected this outcome as DTP provides a complete separation of the DOF, and requires a pinch on a mid-air button to change the manipulation mode. The same effect on the task completion time with DOF separation is also visible in [MRFJ16].

We expect, however, that by changing the method to switch between transformations with a simpler one, we could reduce the time required and the complexity perceived in the same way we can simplify the manipulation in desktop editing with shortcut keys to change the transformation mode. We plan to test a similar option employing static hand gesture recognition for mode switching.

Concerning our research questions, we found that, for Q1, all the techniques are viable for practical use. However, while the usability of all the methods is sufficient for most subjects, and the task is usually completed in a reasonably short time, a non-negligible percentage of the subjects rated all the methods difficult/fatiguing, and low SUS scores (<50) were found with all the techniques, including MRTK, even though it is the default method for Hololens 2 applications. Looking at the swarm plots in Figure 6, it is possible to see that a non-negligible amount of dockings are performed in more than one minute with DTP and MRTK. These results demonstrate the necessity to investigate further and improve the design of the techniques.

Considering Q2, we found a clear advantage in the completion time performances for HB, with the other two methods providing similar, worse results. However, when we look at the questionnaires, the preferences are not evenly in favor of a single method. While 23 subjects (96%) obtained the shortest average docking time with HB (only one performed better with DTP), the SUS scores were highest for MRTK for 9 users (37.5%), for DTP for 6 users (25%) and with HB for 7 users (29.1%). This fact strongly hint s at possible psychological factors influencing the perceived usability independently of the effectiveness.

A possible reason is related to a preference of some subjects for a more understandable manipulation provided by the DOF separation. The conscious feeling of DOF controls may have a positive bias on the SUS question for certain subjects even if they can effectively but unconsciously use existing motor programs to complete the task faster using integral, symmetric actions.

It is similarly surprising to observe that there are no statistically significant differences in the perceived difficulty in completing the task although there is a large difference in the efficiency of the methods. It is also worth noting that the perception of fatigue is not matching the differences in hand motions recorded with the different techniques. This fact might depend on the short duration of the task and will also require further investigation.

Fatigue can be a relevant issue for all methods, especially in the case of long-lasting interactions. One possible solution to reduce it could be scaling indirect motion control properly, as the control-response ratio can strongly influence the perceived fatigue [CCCS17].

Concerning the performances of the manipulation methods at different distances (Q3), we couldn't find evidence of that being a relevant factor, and we recorded only a small effect for MRTK. There are, however, sizeable differences in the amount of accumulated translation and rotation at different distances. For the translation, they can depend on the scaled mapping implemented similarly in MRTK and HB, but the difference is much more relevant for MRTK, and we think this aspect definitely needs further investigation. The object rotation is consistently higher at the shortest distance, independent of the method and possibly related to the changed visual feedback.

Finally, considering Q4, it emerged that previous experience does influence the subjects' performance. This fact suggests that the amount of acquaintance with different types of 3D editors, games, and VR/MR toolkits can affect the results of usability studies, and related information should always be collected and analyzed. The different performances and preferences of the users suggest that a good design guideline is to enable the choice of variable manipulation modes within the same MR application to better support the diversity of the subjects.

The experience-related differences suggest that specific training could make all the methods easier to use and more accepted. To verify the effects of training, however, it would be necessary to study learning curves, and this is out of the scope of our work. It is worth noting that we did not find significant improvements in the repeated trials performed in our tests. The study of the learning curves would require the design of a proper longitudinal study.

For generic users without specific experience, a simple design choice derived from our results is to choose the method according to the application constraints. MRTK may be the optimal choice for a more "natural" manipulation if scaling is not involved. DTP can be the optimal choice if a single-handed method is required and scaling is necessary. A handlebar implementation with proper visual feedback is always the best choice if fast manipulation is needed.

7. Limitations and future works

Our study provides many insights into the interaction design of MR applications but presents clear limitations. We tested only three different implementations of the transform control. The rotation mapping can be particularly critical, and past works show that subtle variations in this choice can affect the performances and user preferences of the whole method [CEG18; CG15]. In [KKF20], the authors discuss how the inability of the users to identify the target rotation they need to perform results in their action being a searching process rather than a planned movement.

Our analysis of the effects of the distance is also limited, as we compared only two offsets from the virtual object. These distances are typical of room-scale applications (e.g., interior design) but a larger variability could have determined increased effects.

The docking task depends on specific parameters (the thresholds of the position, orientation, and scale used to determine if the docking is successful), and a variation of the required accuracy could impact the final results.

We tested the method on a single task. This task is one of the most common in practical applications, but other manipulation tasks are intrinsically different. Examples can be to find an accurate alignment to a given template or to explore the object's surface. It would be interesting to compare the methods tested here on these different tasks with other evaluation metrics. We could expect, for example, that DOF separation may provide better accuracy in tasks requiring a fine-grained control of the manipulated object pose as suggested in [MRFJ16].

Finally, while our experiments are based on a "mixed" interaction task, docking virtual objects on real ones, and, as far as we know, it is the first time that this is proposed, we did not test the effects of the mixed reality implementation choices on the manipulation (e.g., HMD type, rendering, etc.) or the differences in the effectiveness of the methods between pure VR and MR as proposed in prior works [KYT*17; CJK*20]. We plan to extend our research in future works trying to address some of these issues.

8. Conclusion

The success of mixed reality technologies is not only linked to the development of sophisticated visualization and tracking tools but also to the implementation of efficient user interfaces. In this work, we have shown that an efficient remote manipulation of virtual objects in augmented environments without using handheld devices but relying instead on Hololens 2 finger tracking can be achieved with different interaction metaphors but with non-negligible usability issues. All the methods tested in our study demonstrated specific advantages and drawbacks and may be particularly suited for selected categories of users. Symmetric and integrated control of 7DOF transform is, however, faster for almost all subjects, even if it is not necessarily the preferred choice if different options involving asymmetric and separated controls are available. A proper visual affordance is important to suggest an integrated control. The efficiency of the methods is not significantly affected by the target distance, while there is a non-negligible influence of the previous experience with different interface types on the users' performance with specific techniques.

Acknowledgements

This study was partially carried out within the PNRR research activities of the consortium iNEST (Interconnected North-Est Innovation Ecosystem) funded by the European Union NextGenerationEU (Piano Nazionale di Ripresa e Resilienza (PNRR) – Missione 4 Componente 2, Investimento 1.5 – D.D. 1058 23/06/2022, ECS_00000043).

References

[BDAH21] Bergström, Joanna, Dalsgaard, Tor-Salve, Alexander, Jason, and Hornbæk, Kasper. "How to evaluate object selection and manipulation in VR? Guidelines from 20 years of studies". *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021, 1–20 1, 2.

[BGG*07] Bettio, Fabio, Giachetti, Andrea, Gobbetti, Enrico, et al. "A Practical Vision Based Approach to Unencumbered Direct Spatial Manipulation in Virtual Worlds." *Eurographics Italian Chapter Conference*. 2007, 145–150 2.

[BH00] Balakrishnan, Ravin and Hinckley, Ken. "Symmetric bimanual interaction". *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2000, 33–40 2, 8.

[BH97] Bowman, Doug A and Hodges, Larry F. "An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments". *Proceedings of the 1997 symposium on Interactive 3D graphics*. 1997, 35–ff 2.

[BKM09] Bangor, Aaron, Kortum, Philip, and Miller, James. "Determining what individual SUS scores mean: Adding an adjective rating scale". *Journal of usability studies* 4.3 (2009), 114–123 7.

[BMA*14] Bossavit, Benoit, Marzo, Asier, Ardaiz, Oscar, et al. "Design choices and their implications for 3d mid-air manipulation techniques". *Presence: Teleoperators and Virtual Environments* 23.4 (2014), 377–392 2, 3.

[Bro*96] Brooke, John et al. "SUS-A quick and dirty usability scale". *Usability evaluation in industry* 189.194 (1996), 4–7 3, 7.

[CCCS17] Chen, Li-Chieh, Cheng, Yun-Maw, Chu, Po-Ying, and Sandnes, Frode Eika. "Identifying the usability factors of mid-air hand gestures for 3D virtual model manipulation". *International Conference on Universal Access in Human-Computer Interaction*. Springer. 2017, 393–402 10.

[CEG18] Caputo, Fabio M, Emporio, Marco, and Giachetti, Andrea. "The Smart Pin: An effective tool for object manipulation in immersive virtual reality environments". *Computers & Graphics* 74 (2018), 225–233 2, 4, 10.

[CG15] Caputo, Fabio Marco and Giachetti, Andrea. "Evaluation of basic object manipulation modes for low-cost immersive Virtual Reality". *Proceedings of the 11th Biannual Conference on Italian SIGCHI Chapter*. 2015, 74–77 2, 10.

[CJK*20] Caputo, Ariel, Jacota, Sergiu, Krayevskyy, Serhiy, et al. "XR-Cockpit: a comparison of VR and AR solutions on an interactive training station". *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. Vol. 1. IEEE. 2020, 603–610 10.

[CW15] Cho, Isaac and Wartell, Zachary. "Evaluation of a bimanual simultaneous 7dof interaction technique in virtual environments". *2015 IEEE symposium on 3D User Interfaces (3DUI)*. IEEE. 2015, 133–136 2.

[FIO19] Flavián, Carlos, Ibáñez-Sánchez, Sergio, and Orús, Carlos. "The impact of virtual, augmented and mixed reality technologies on the customer experience". *Journal of business research* 100 (2019), 547–560 1.

[Gui87] Guiard, Yves. "Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model". *Journal of motor behavior* 19.4 (1987), 486–517 2.

[Jan19] Janusz, Jan. "Toward the new mixed reality environment for interior design". *IOP Conference Series: Materials Science and Engineering*. Vol. 471. 10. IOP Publishing. 2019, 102065 1.

[JSMM94] Jacob, Robert JK, Sibert, Linda E, McFarlane, Daniel C, and Mullen Jr, M Preston. "Integrality and separability of input devices". *ACM Transactions on Computer-Human Interaction (TOCHI)* 1.1 (1994), 3–26 2, 8.

[KKF20] Kulik, Alexander, Kunert, André, and Froehlich, Bernd. "On Motor Performance in Virtual 3D Object Manipulation". *IEEE Transactions on Visualization and Computer Graphics* 26.5 (2020), 2041–2050. doi: 10.1109/TVCG.2020.2973034 10.

[KRSH22] Kim, Myoung Gon, Ryu, Ji Seok, Son, Jaemin, and Han, Jung Hyun. "Virtual object sizes for efficient and convenient mid-air manipulation". *Visual Computer* 38.9-10 (2022), 3463–3474 2.

[KSP20] Kang, Hyo Jeong, Shin, Jung-hye, and Ponto, Kevin. "A comparative analysis of 3d user interaction: How to move virtual objects in mixed reality". *2020 IEEE conference on virtual reality and 3D user interfaces (VR)*. IEEE. 2020, 275–284 3.

[KYT*17] Krichenbauer, Max, Yamamoto, Goshiro, Taketom, Takafumi, et al. "Augmented reality versus virtual reality for 3d object manipulation". *IEEE transactions on visualization and computer graphics* 24.2 (2017), 1038–1048 10.

[MCG*19] Mendes, Daniel, Caputo, Fabio Marco, Giachetti, Andrea, et al. "A survey on 3d virtual object manipulation: From the desktop to immersive virtual environments". *Computer graphics forum*. Vol. 38. 1. Wiley Online Library. 2019, 21–45 1–3.

[MFA*14] Mendes, Daniel, Fonseca, Fernando, Araujo, Bruno, et al. "Mid-air interactions above stereoscopic interactive tables". *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE. 2014, 3–10 2.

[Mic22] Microsoft. *MRTK 2 - Unity documentation*. 2022. url: https://docs.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk2 (visited on 07/30/2022) 1–4.

[MRFJ16] Mendes, Daniel, Relvas, Filipe, Ferreira, Alfredo, and Jorge, Joaquim. "The benefits of dof separation in mid-air 3d object manipulation". *Proceedings of the 22nd ACM conference on virtual reality software and technology*. 2016, 261–268 2, 9, 10.

[PLMH21] Pohl, Henning, Lilija, Klemen, McIntosh, Jess, and Hornbæk, Kasper. "Poros: configurable proxies for distant interactions in VR". *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021, 1–12 2.

[PSP99] Pierce, Jeffrey S, Stearns, Brian C, and Pausch, Randy. "Voodoo dolls: seamless interaction at multiple scales in virtual environments". *Proceedings of the 1999 symposium on Interactive 3D graphics*. 1999, 141–145 2.

[SGH*12] Song, Peng, Goh, Wooi Boon, Hutama, William, et al. "A handle bar metaphor for virtual object manipulation with mid-air interaction". *Proceedings of the SIGCHI conference on human factors in computing systems*. 2012, 1297–1306 1, 2, 4.

[Sho92] Shoemake, Ken. "ARCBALL: A user interface for specifying three-dimensional orientation using a mouse". *Graphics interface*. Vol. 92. 1992, 151–156 2, 4.

[Uni] Unity Technologies. *Unity website*. https://unity.com/. Accessed: 2022-12-11 3.

[WHB*18] Whitlock, Matt, Harnner, Ethan, Brubaker, Jed R, et al. "Interacting with distant objects in augmented reality". *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2018, 41–48 3.

[WLK*14] Weichel, Christian, Lau, Manfred, Kim, David, et al. "MixFab: a mixed-reality environment for personal fabrication". *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2014, 3855–3864 1.