

Merge Policy Comparison for AR HMD Collaborative Manipulation

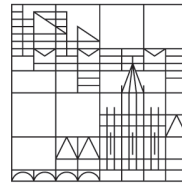
Master Thesis

submitted by

Pablo Martinez Blasco (01/1052824)

at the

Universität
Konstanz



Human-Computer Interaction Group

Department of Computer and Information Science

1st Supervisor: Prof. Dr. Harald Reiterer

2nd Supervisor: TT.-Prof. Tiare Feuchtner

Konstanz, 1. December 2022

Abstract

Merge Policy Comparison for AR HMD Collaborative Manipulation

Augmented Reality incorporates both virtual and virtual worlds. Head mounted displays provide a hands-free interaction with the virtual environment, without forcing the user's attention to a screen like hand-held devices. For true immersion to occur, virtual objects need to possess the affordances of real objects, one of them being the concurrent manipulation from multiple users. Literature reveals two approaches for this collaboration: 1. *Separation of DOFs* splits the object's degrees of freedom among the users, and 2. *Combination of user actions* aggregates the user manipulations into a final transform applied to the object. Since any mathematical function can be calculated for *Combination*, related work has implemented a variety of them: 1. *Sum* adds all the user inputs together, 2. *Mean* averages user actions, and 3. *Weighted mean* assigns weights to the averaging process. Nonetheless, little comparison among these merge policies is provided in literature and none is available on HMD devices.

Therefore, this thesis provides a baseline comparison for the three approaches in a AR HMD docking task. To that end, a study prototype which implements the aforementioned merge policies is developed. A controlled lab experiment with 36 participant was conducted, where the performance, subjective workload and user experience for each of the conditions were measured using the prototype. While there were no significant differences among conditions in regards to performance, user preferences were polarized between the *Sum* and *Weighted* conditions. The *Average* was preferred by users to make any accurate placement, and the *Sum* needed the most coordination due to overshooting issues. Based on the obtained results, merge policy selection for future tasks in addition to a comparison frame for future collaborative AR HMD research are presented.

Contents

- Abstract i

- Contents iii

- 1 Introduction 1

- 2 Foundations 3
 - 2.1 Four Canonical Tasks 3
 - 2.2 Collaboration in AR 3

- 3 Related Work 6
 - 3.1 Action Integration 6
 - 3.2 Merge Policy Conditions 12
 - 3.3 Conclusions from Related Work 14

- 4 Implementation 15
 - 4.1 Requirements 15
 - 4.2 Merge Policy Conditions 16
 - 4.3 Networking 20
 - 4.4 Coordinate System Synchronization 21
 - 4.5 Cues 21

- 5 Experimental Comparison 23
 - 5.1 Study Design 23
 - 5.2 Apparatus 23
 - 5.3 Task 23
 - 5.4 Pre-study and Tolerance level 24
 - 5.5 Main Study 26
 - 5.6 Results 27

- 6 Discussion 34
 - 6.1 Pre-study 34
 - 6.2 Main study 35
 - 6.3 Limitations 38
 - 6.4 Future Work 39

- 7 Conclusions 40

Contents

References

iv

1 Introduction

Augmented Reality (AR) offers a seamless immersion of the real and virtual world. With content being added to our own viewport (our eyes), any 3D virtual object coexists in the same space as real objects. Single-user applications show great promise, but for a complete integration of the virtual world collaboration among users needs to be supported [1]. The combination of both virtual and real spaces allow for the affordances of the real world to be applied to virtual objects (i.e use as reference frames [2]). Consequently, computer supported collaborative work (CSCW) can be enhanced considering how users perform the task in the real world, while improving the sense of presence and performance [3].

Hand-held devices (HDD) have been often used in recent literature, mainly due to their market availability, “with 31% of papers relying partially or fully on them” [3]. While their uses vary from complementing a larger system, or as input devices for Head Mounted Displays (HMDs), HDDs require the users to change focus between the device and their collaborators. Furthermore, the use of HDDs constrains the hands of the users as they hold the device, depicted in Figure 1.1. HMDs are preferred to HDDs, as they allow users to interact with other people and the physical workspace [4].

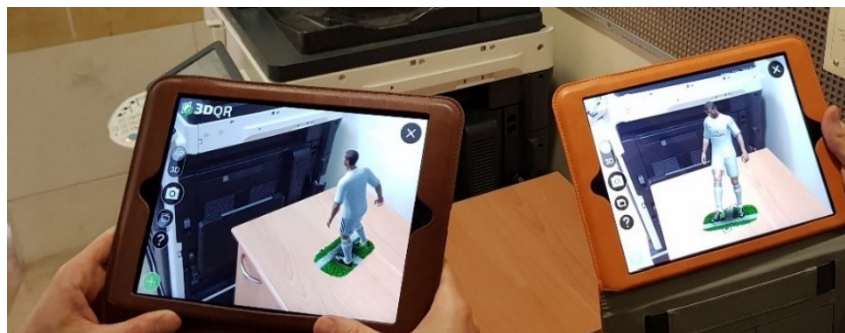


Figure 1.1: Hand-held devices hand restrictions and focus requirement.
Taken from [5].

Due to hardware restrictions, namely the smaller field of view (FOV) and gesture recognition difficulties, there is not much research that focuses on collaborative manipulation with HMD devices. With the release of the HoloLens2 and its increased FOV, the development of baselines is required in order to compare different implementations of collaborative manipulation systems, as the findings obtained from Virtual Environments [6, 7, 8] and HDDs [9] may not apply.

Nonetheless, literature defines the requirements for any collaborative manipulation task, independent of the device employed. First of all, users must be able to simultaneously manipulate the objects. While action on independent degrees of freedom (DOFs) can be readily integrated as there is no collision on the manipulations, what happens if users act on the same degrees of freedom? In this regard, related work makes a clear distinction on whether users are allowed to access the same DOFs, or they are subdivided among the participants. The latter is defined as *Separation of DOFs* [10], while the former is called *Combination of user actions* [11]. *Combination* incorporates the transformation requests of the users and an aggregated final transform is then applied to the object. There is available literature comparing both of the action integration approaches, on the other hand, little has been conducted on comparing the different mathematical functions (merge policies) to combine the user inputs. While different merge policies like the *Sum*, *Mean* or *Weighted Mean* have appeared in related work, there is only one direct comparison of the *Mean* and *Common Component* [8].

In order to provide a baseline performance comparison for different merge policies, three approaches were implemented into a collaborative docking task [12], with extended use in literature, developed into a study prototype. The prototype was used to evaluate an experimental comparison among the conditions. Results obtained are presented, analyzed and discussed in this thesis.

The structure of the thesis is comprised of 5 chapters: the next chapter lays out the foundations of collaborative work and object manipulation required to understand the literature introduced in Chapter 3. Chapter 4 extends from the previous chapters and describes the system requirements for the comparison of the techniques, in addition to describing the implementation of the related work into the study prototype. Results obtained and their analysis are presented in Chapter 5, and further discussed and analyzed in Chapter 6, together with limitations and future work. The last chapter offers a summary of the most meaningful information from the thesis.

2 Foundations

2.1 Four Canonical Tasks

Bowman et al. [13] decompose 3D manipulations into basic tasks, defined as four canonical tasks: *selection/deselection*, *positioning*, *rotation* and *scaling*. All manipulations must start with a selection, allowing the user to acquire or identify the object to manipulate. Any manipulation task is ended via release or deselection, which represents the opposite to selection, and is thus considered the same canonical task. Positioning refers to any alteration of the 3D position of the object, taking the object from its position to a target position which is user defined. Rotation is the task of “changing the orientation of an object” [13], where the rotation magnitude is again user defined. Scaling encompasses the alteration of the object’s size, and is the only canonical task without a real world counterpart. A typical sequence of an object manipulation is shown in Figure 2.1.

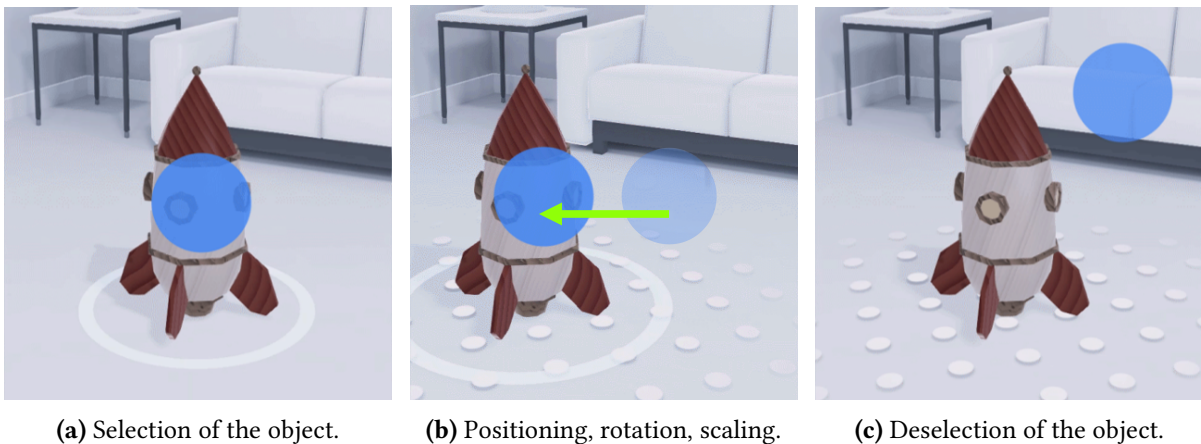


Figure 2.1: Manipulation task subdivided in the canonical tasks that comprise it. The object is first selected. With the object acquisition, any desired transformation change can be performed (i.e positioning, rotation, scaling). The manipulation ends with the deselection of the object. Images from [14].

2.2 Collaboration in AR

Collaborative object manipulation is a subsection of collaboration. While collaboration also entails the task subdivision among users, their communication, joint visual attention, etc. this thesis will

only touch on the knowledge required in order to have a correct understanding of collaborative object manipulation and its requirements.

2.2.1 Cooperation Levels

An interesting cooperation taxonomy was presented by Margery and Arnaldi [15]. They split cooperation in 3 different levels, each building on top of the previous one by adding more limitations, which are pictured in Figure 2.2. *Level 1* is defined as the basic cooperation level, which only requires the ability for “users to perceive each other in the virtual world ... and providing ways of communicating between these users”. *Level 2* requires the possibility for the users to perform changes on the scene e.g. move an object. *Level 3* is the top level of the taxonomy. While users can perform changes on the objects in Level 2, they cannot collaboratively realize them. Thus, this level is divided in two further sublevels. *Level 3.1* is achieved when users can “act on the object in independent ways”, therefore never altering the same properties of the object. *Level 3.2* “enables two users to act in a codependent way”, with the allowance to act on the same properties collaboratively and simultaneously, thus being the highest attainable degree of collaboration.

This thesis will focus on the third level and more specifically on the 3.2 sublevel, due to the action integration policy selected which will be explained in the Related Work chapter.

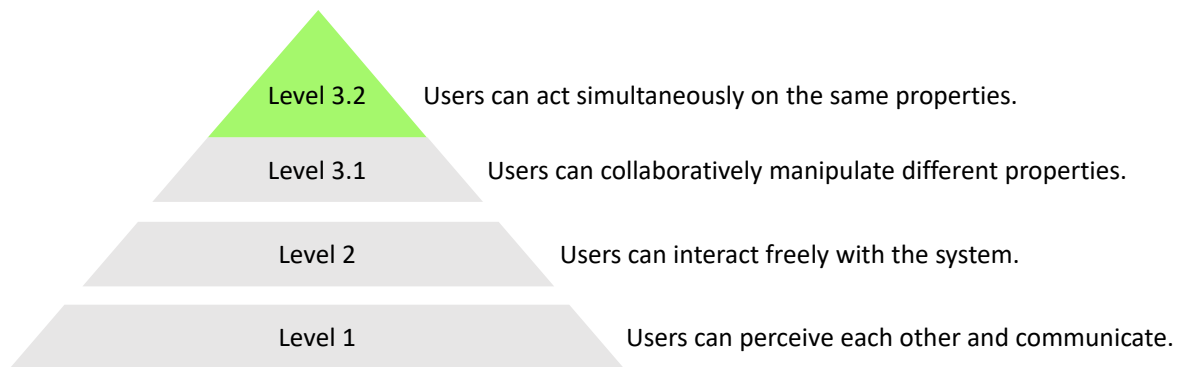


Figure 2.2: Different cooperation levels presented by Margery and Arnaldi [15]. Diagram extended from [16]

2.2.2 Concurrent inputs

Broll’s [11] *Detecting and Synchronizing Concurrent Interactions* section will be summarized as it provides a really good definition of “concurrent requests”.

Concurrency in a distributed system needs to be mathematically defined. This is achieved by defining as concurrent any two request that come within a period of $\Delta\epsilon$. Therefore, any collaborative manipulation has an innate time window that needs to be set. Broll provide a list of options to process the interaction requests that are received during this time window:

2 Foundations

- ignore further requests
- use new requests instead of preceding ones
- store and process new requests in a later cycle
- combine all of the requests

As soon as two or more users perform concurrent interaction requests, these are no longer processed instantly, and a timer has to be started and the requests stored. When the timer runs out, or a request from each client has been received, the timer is reset and the requests processed by the selected method.

3 Related Work

This chapter will outline the literature meaningful to the topic. First, the available different approaches for action integration will be looked at, in addition to comparisons between them. Common merge policies used in literature will then be discussed. At the end, relevant information from this chapter will be summarized, together with its impact on this work.

3.1 Action Integration

Based on Margery and Arnaldi's level taxonomy [15], and more specifically the level 3 sublevels, two branches on how to combine the user manipulations become apparent. One in which users can only interact with different properties of the object, and another one where modification of the same properties is allowed. Broll identify the same two possibilities: one that they name *constraint based*, which "also includes the separation of independent interaction requests for the same object.", and the second one named *combining interaction requests*, and it "involves calculating a new total request from a series of single requests." [11].

Pinho, Bowman, and Freitas name the presented approaches as *Separation of DOFs* and *Composition of user actions* [10], nomenclature that will be borrowed for this thesis, along with the shortened versions in *Separation* and *Combination*. The term *Combination* will be used over *Composition* as the former has a more spread out use [15, 11, 17, 18].

3.1.1 Separation of Degrees of Freedom

Motivated by the possibility to have users employ different interaction techniques (Simple Virtual Hand and Ray-cast), Pinho, Bowman, and Freitas presented a framework that allowed cooperative manipulation via the use of Separation, allowing users to control exclusive degrees of freedom at a time [10]. They identified different DOF splits between the users that may prove helpful depending on the task, available in Table 3.1.

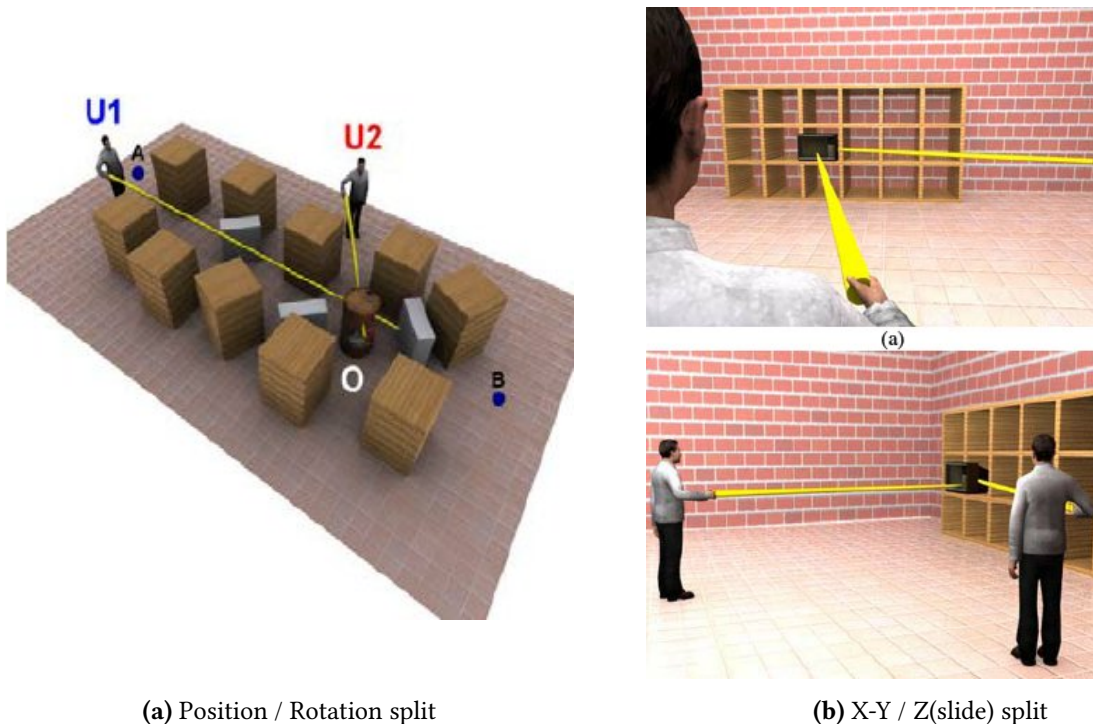
The first subdivision performed (Figure 3.1a) splits the control over different canonical tasks, with one user controlling the translation while the other has access to the rotation. Pinho, Bowman, and Freitas state that this separation "has proven very interesting when small adjustments are necessary" and also very useful when the user rotating can see parts that the one translating the object cannot. The second

3 Related Work

approach provided (Figure 3.1b) splits the translation DOFs into X-Y and Z axis, thus having one user in charge of vertical and horizontal alignment, with the second taking care of the depth positioning. In cases where the target position is far from the user that controls the 2D placement, or when there are objects blocking the view of the final position for them, this approach was “quite helpful” [10].

IT A	DOF A	IT B	DOF B	Comments
SVH	Position	SVH	Rotation	Useful for docking tasks and small adjustments. Good when one user cannot see parts of the object.
SVH	X, Y	SVH	Z	Facilitates precise positioning
RC	Position	RC	Rotation	Useful for rotations that are difficult with RC
RC	Position	RC	Rotation, Slide	Useful for distant placement and rotations
SVH	Rotation	RC	Position	Useful for rotations that are difficult with RC
SVH	Rotation, Slide	RC	Position	Useful for distant placement and rotations

Table 3.1: Pinho, Bowman, and Freitas present different combinations of interaction techniques (IT) by the users, as well as performing a separation of DOF that each controls. From [10].



(a) Position / Rotation split

(b) X-Y / Z(slide) split

Figure 3.1: DOF split approaches provided by Pinho, Bowman, and Freitas

3.1.2 Combination of User Actions

As hinted by Margery and Arnaldi [15], *Combination* is the more general approach, as applying constraints to it would lead to the *Separation* method. Since multiple users have simultaneous access to the same DOFs of an object, a decision has to be made on how the different transformation requests are merged together. The mathematical function that combines the requests is called merge policy (term coined by Chenechal et al. [19]), with the most common choices being the *Sum* and *Mean*. Different merge policies will be described in depth in Section 3.2.

Ruddle, Savage, and Jones [8] implement a piano mover's problem, where users have to maneuver a large object through a narrow space (Figure 3.2). The study compared two merge policies, the *mean* and the *common component*. With the *common component*, transformations were only applied if both users performed actions that were within a small tolerance of each other, as only the shared component from the users' input is used. On the other hand, the *mean* allowed for any transformation to be made, in turn supporting different types of maneuver among the users.

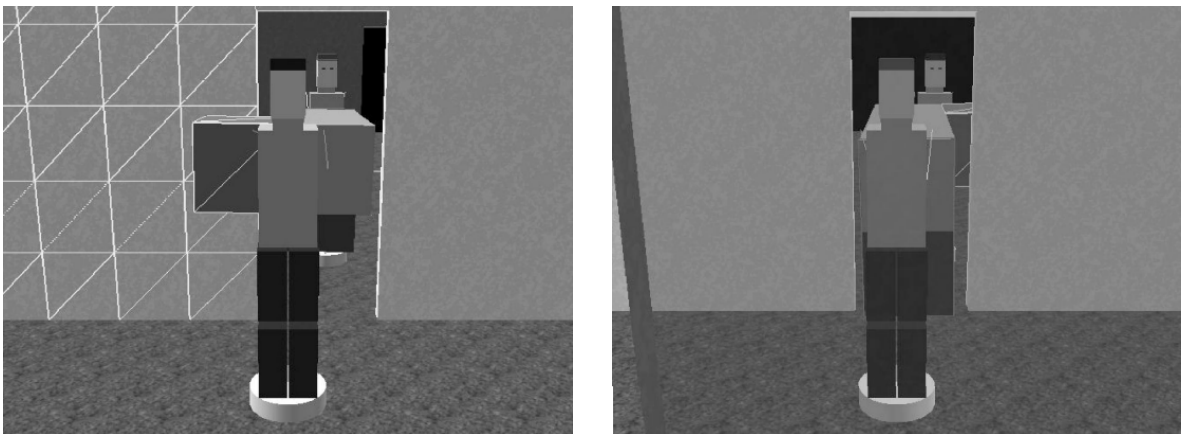


Figure 3.2: Piano mover's problem implementation by Ruddle, Savage, and Jones [8].

No statistical significance was shown on the task completion time between both of the merge policies, depicted in Figure 3.3. However, when the task was subdivided in different sub-tasks, significant results were obtained. When the phase required the users to perform similar actions (e.g. rotate the object), the *common component* was quicker likely due to the constraints applied to the inputs, as with the *mean* policy "slight differences in their manipulations caused the object to collide" [8]. On the other hand, when the task required different types of movement (e.g. one had to maneuver through an opening), the *mean* policy obtained a significantly better performance.

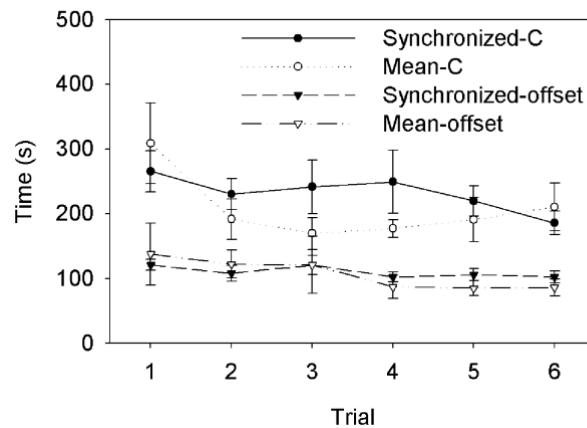


Figure 3.3: Task completion times for both merge policies and environment setups. Taken from [8].

Grandi et al. [6] had users perform an obstacle crossing task, depicted in Figure 3.4, where they had to guide an object inside a winding tunnel where adjustments in 7DOF need to be performed. In the study, how the number of users that conformed a group affected the performance was analyzed, ranging from 2 users to 4. The *Sum* was used as a merge policy, obtaining a transform matrix from each user and applying all of them to the task object. Group size was found to affect the accuracy but not the completion time of the task. They also report that, as the number of group members increases, so does their task subdivision (lower allotted DOFs per user), indicated by the number of role swaps. It is also interesting to note that there is a significance in completion time between the first task and second performed by the group, but accuracy presents no statistically improvements.



Figure 3.4: Sum as merge policy and obstacle crossing task. Taken from [6].

Salzmann, Jacobs, and Froehlich [20] evaluate two collaborative tasks, an assembly planning scenario and a windshield assembly task. The former will be ignored as the roles for each user were asymmetric and thus no collaborative object manipulation was performed. On the other hand, the windshield

assembly task consisted on the users picking up a windshield and transporting it to the corresponding position on the car, both virtually and with a real prop. Movement of the windshield in the virtual reality environment was performed via the *mean* merge policy, “averaging of translations and rotations of the two hands” [20]. Results will not be discussed as they focus on the use of HMDs or prop based interactions.



Figure 3.5: Virtual and real views of the windshield assembly task, where users have to carry a windshield from a rack to the corresponding position on the car. Taken from [20].

3.1.3 Comparison of Approaches

Aguerreche, Duval, and Lécuyer [18] compare a Reconfigurable Tangle Device (RTD-3) with both the *Separation* and *Combination*. The task to be completed was “pick-and-place” task similar to the one used by Salzmann, Jacobs, and Froehlich [20], with users having to maneuver a car hood out of a Z-shape and place it on a visual support as shown in Figure 3.6. The combination merge policy implemented in this study was the *Mean*.

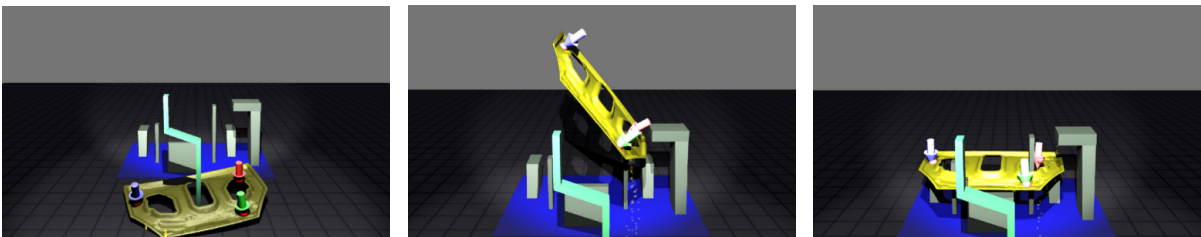


Figure 3.6: Starting, intermediate and end positions for the pick and place task implemented by Aguerreche, Duval, and Lécuyer. Taken from [18].

Results obtained by Aguerreche, Duval, and Lécuyer are depicted in Figure 3.7. In regards to completion time, only differences between the RTD-3 and the *Mean* were of statistical significance. Results for the number of collisions (accuracy) showed significance between the RTD-3 and *Mean* in addition to in

between the *Separation* and *Mean*. The subjective ratings revealed no statistical significance between the *Separation* and *Mean*, or if they were, only the RTD-3 post-hoc test were reported.

	Time (in seconds)		Number of collisions		RTD-3		Mean		Separation	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
RTD	26.22	$\sigma = 9.7$	151.88	$\sigma = 38.59$	5.83	$\sigma = 1.37$	4.54	$\sigma = 1.32$	4.00	$\sigma = 1.72$
Mean	18.34	$\sigma = 7.38$	166.38	$\sigma = 51.63$	5.88	$\sigma = 0.99$	4.42	$\sigma = 0.88$	3.63	$\sigma = 1.64$
Separation	22.44	$\sigma = 8.39$	227.54	$\sigma = 59.03$	5.71	$\sigma = 1.12$	4.88	$\sigma = 0.85$	4.54	$\sigma = 0.83$
					4.79	$\sigma = 1.64$	4.88	$\sigma = 1.45$	5.13	$\sigma = 1.68$
					4.96	$\sigma = 1.49$	5.04	$\sigma = 1.57$	5.08	$\sigma = 1.47$

Figure 3.7: Results from Aguerreche, Duval, and Lécuyer [18]. On the left, completion time and number of collisions for the three conditions. On the right the qualitative evaluation results from a 7-point Likert scale. Taken from [18].

Wieland et al. [9] use hand-held displays in a collaborative furnishing task. The compared approaches include the *Separation* and the *Combination* (defined as Hybrid), in addition to a constrained version of the *Combination* (defined as Composition) in which users may only perform the same canonical task simultaneously. These approaches are depicted in Figure 3.8. The merge policy selected in this case was the *Sum*. Wieland et al. implement a 3D docking task where participants needed to place virtual objects in predefined positions and rotations. Seven pieces of furniture needed to be placed correctly for the task to be completed.

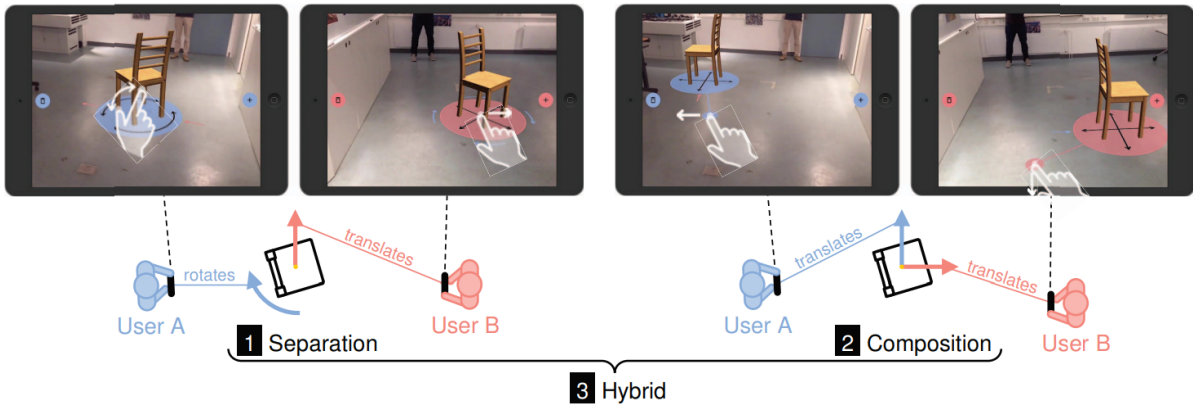


Figure 3.8: Action integration methods implemented by Wieland et al. [9]. The presented *Composition* technique is a constrained version of the *Combination* (Hybrid in this work) where users can only simultaneously manipulate the same canonical task. Taken from [9]

Wieland et al. measure performance via task completion time and accuracy. Completion time analysis (pictured in Figure 3.9) showed significance between the *Separation* and *Composition*, and likewise between *Hybrid* and *Composition*. Due to *Composition* constraining the users to perform the same canonical task, this behaviour was expected. Interestingly, this increase in completion time is also due to participants overshooting the intended target transform, having to later perform corrections. The NASA_TLX results, also illustrated in Figure 3.9, support the aforementioned issues with the *Composition*, reporting significance in the *Physical Demand*, *Effort* and *Frustration* dimensions between the

Separation and the *Composition*. No significance was found between the *Separation* and *Hybrid* conditions. In regards to accuracy, there was no significance between any of the conditions.

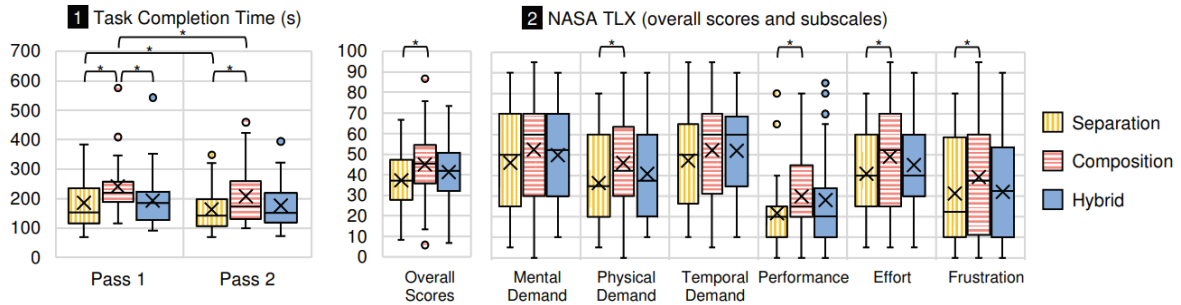


Figure 3.9: Task completion times and NASA_TLX results. Participants completed the task faster with the *Separation* and *Hybrid* techniques than with the *Composition*. Taken from [9]

3.2 Merge Policy Conditions

Since composition allows for the manipulation of multiple (or all) DOFs at the same time for a user, when multiple users act on the same one, the system needs to decide how to combine their respective inputs. While any mathematical function can be used and argued for or against [17], related work has focused on a few options, which will be described in this section.

3.2.1 Sum

The sum is the simplest of the merge policies, simply due to all user inputs getting added together sequentially. This behaviour allows for this merge policy to work server-less, as network authority over objects is not needed and clients just sum all of the manipulations from other users to their local copy of the object. This merge policy can be seen in use by Margery and Arnaldi [15], Grandi et al. [6, 21, 22] and Wieland et al. [9]. Assuming u_i is the transformation for user i , and n is the total number of users, the sum function is defined by Equation 3.1.

$$T_r = \sum_{i=1}^n U_i \quad (3.1)$$

3.2.2 Mean / Average

The mean is another very common merge policy [8, 18, 23]. The mean, intuitively enough, combines the inputs from the users by averaging them. As mentioned in the previous section 2.2.2, there are multiple considerations that need to be made to implement it properly. The literature approach averages

participants' inputs one to one, with little information on the decisions taken whenever input from a user is not received for the current calculation. While there are multiple methods on how to deal with this (assume input is 0, have a bigger window to accept an input, etc.), they will not be covered here. The average of user inputs is calculated using Equation 3.2, in which n again represents the number of users and u_i the transformation performed by a user.

$$T_r = \frac{u_1}{n} + \frac{u_2}{n} + \dots + \frac{u_n}{n} \rightarrow \frac{1}{n} \sum_{i=1}^n u_i \quad (3.2)$$

3.2.3 Weighted Average

The only use of the *Weighted Average* as a merge policy is by Riege et al. [23], although no evaluation of the technique was performed. Their implementation only uses the weighted average for the translations and handled rotations via the standard average. The weighted average functions like the ordinary average, except that each input does not contribute equally to the final transformation, and instead does so via an assigned weight (which sum equals 1). The *Weighted Average* as a merge policy provides an in-between of the *Sum* and *Average*, as it behaves like the latter when coordinated and similar manipulations are performed, while performing like the former as the input magnitudes among the users differ more. Given n representing the total users, with u_i being the transformation from a user and w_i their assigned weight, Equation 3.3 depicts the weighted average translation calculation. Here P_r is used since the result is not applied to the transformation but only to the position of it (as rotation is performed with the *Mean*).

$$P_r = w_1 \cdot u_1 + w_2 \cdot u_2 + \dots + w_n \cdot u_n \rightarrow \sum_{i=1}^n w_i \cdot u_i \quad (3.3)$$

For the calculation of the weights, Riege et al. use the formula shown in Figure 3.10. This approach does not provide the results described, as they state that “A larger amount of hand movement, corresponding to a larger object drag, results in a bigger weight”, demonstrated in the table in Figure 3.10.

$$\omega_1 = \begin{cases} L_1 > L_2 : & \frac{1}{4} \cdot \left(\frac{L_1 - L_2}{L_2} + \frac{1}{2} \right) \\ L_1 < L_2 : & \frac{1}{4} \cdot \left(\frac{L_2 - L_1}{L_1} + \frac{1}{2} \right) \end{cases}$$

$$\omega_2 = 1 - \omega_1$$

L_1	L_2	ω_1	ω_2
3	2	1/4	3/4
4	2	3/8	5/8
6	2	5/8	3/8
8	2	7/8	1/8

Figure 3.10: Riege et al.'s weight calculation formula on the left, the table shows weights obtained for different user inputs.

3.3 Conclusions from Related Work

Whenever two or more users manipulate the same object, unlike in the real world, there is no physical restrictions applied to the interaction, and it needs to be decided how the manipulations will be combined (action integration). To that end, related work proposes two options, *Separation* and *Combination* [11], the latter being the more general approach as constraints can be applied to transform it into the former [11]. *Separation* performs a subdivision of the available degrees of freedom among the users, be it using canonical tasks (i.e. position and rotation) or axis (i.e. X-Y and Z) [10]. On the other hand, *Combination* allows simultaneous manipulations of any degree of freedom, therefore requiring a mathematical function to process user manipulations, define as merge policy. Although multiple merge policies have been implemented in related work (i.e. *Sum*, *Mean*, *Weighted average*, *Common component*), only one comparison between the *Mean* and *Common component* was performed [8] which showed no statistical significance. Comparisons between the action integration approaches are not exceedingly more common. Wieland et al. [9] and Aguerreche, Duval, and Lécuyer [18] both contrast the *Separation* and *Combination*, although they implement different merge policies in the *Sum* and *Average* respectively. This thesis aims to bring a baseline comparison among literature's most commonly used merge policies, analyzing them quantitatively in addition to qualitatively (cf. Section 5.1).

4 Implementation

In this chapter, the system's implementation will be described, starting with the requirements and handling topics like networking, coordinate system synchronization and concurrent interactions. Lastly, the different conditions, as well as their mathematical formulas, will be defined.

4.1 Requirements

Any collaborative work, and more specifically collaborative object manipulation, is heavily dependent on a low latency network environment, as delays in the messaging lead to the concurrency problems described previously in 2.2.2.

- R1 Networking:** The HoloLens from the users need a low-latency connection between one another and with the server in order to have an accurate and real-time state of the system.

Additionally, users must be able to interact with the objects present in the system. The implementation of the canonical tasks is thus necessary. In order to simplify the task, scaling will be the one canonical task left out.

- R2 Selection/Release:** Users should be able to select and release the objects comfortably and at any wanted point.
- R3 Translation:** Translation of the objects by each user should be supported by the study prototype.
- R4 Rotation:** Users need to be able to apply any rotation to the objects.

Furthermore, the study prototype needs to implement *Combination* as an action integration and *Sum*, *Mean* and *Weighted* as merge policies in order to compare them.

- R5 Composition:** The study prototype needs to support simultaneous access to the same DOFs by different users.
- R6 Sum:** The study prototype needs to implement the merge policy *Sum* to combine users' inputs.
- R7 Mean:** Transformations based on the *Mean* merge policy from users' inputs needs to be implemented by the study prototype.
- R8 Weighted:** The combination of user actions using the *Weighted* merge policy needs to be implemented by the study prototype.

Lastly, due to HMDs, a fewer amount of visual feedback is needed for coordination. Nonetheless, feedback of the partner's actions within the system as well as feedback from the system are needed for the correct performance of the task.

- R9 Visual Cues:** The study prototype needs to implement visual cues to give users feedback of the actions of their partner as well as the state of the system.

4.2 Merge Policy Conditions

The most commonly used merge policies from literature have been discussed in the previous section 3.2. The goal of the study is to provide a baseline performance comparison of them in a HMD setting, in order to narrow down performance discrepancies as well as identify where each policy may excel.

4.2.1 Sum

As described in section 3.2.1, this merge policy directly applies all transformations made by the users to the object. Antagonistic transforms, seen on the left side in Figure 4.1, take on the direction of the higher magnitude input, reducing it by the smaller's magnitude. On the other hand, transformations in the same degree of freedom with the same direction quickly amplify the input made by any user as their magnitudes combine together, visible on the right side graphs in Figure 4.1. A behaviour similar to what is pointed by Ruddle, Savage, and Jones [8] in their comparison of the *mean* and the *common component* is expected, where users try to perform the same action that would be correct individually but due to both users performing that same action the object overshoots.

4 Implementation

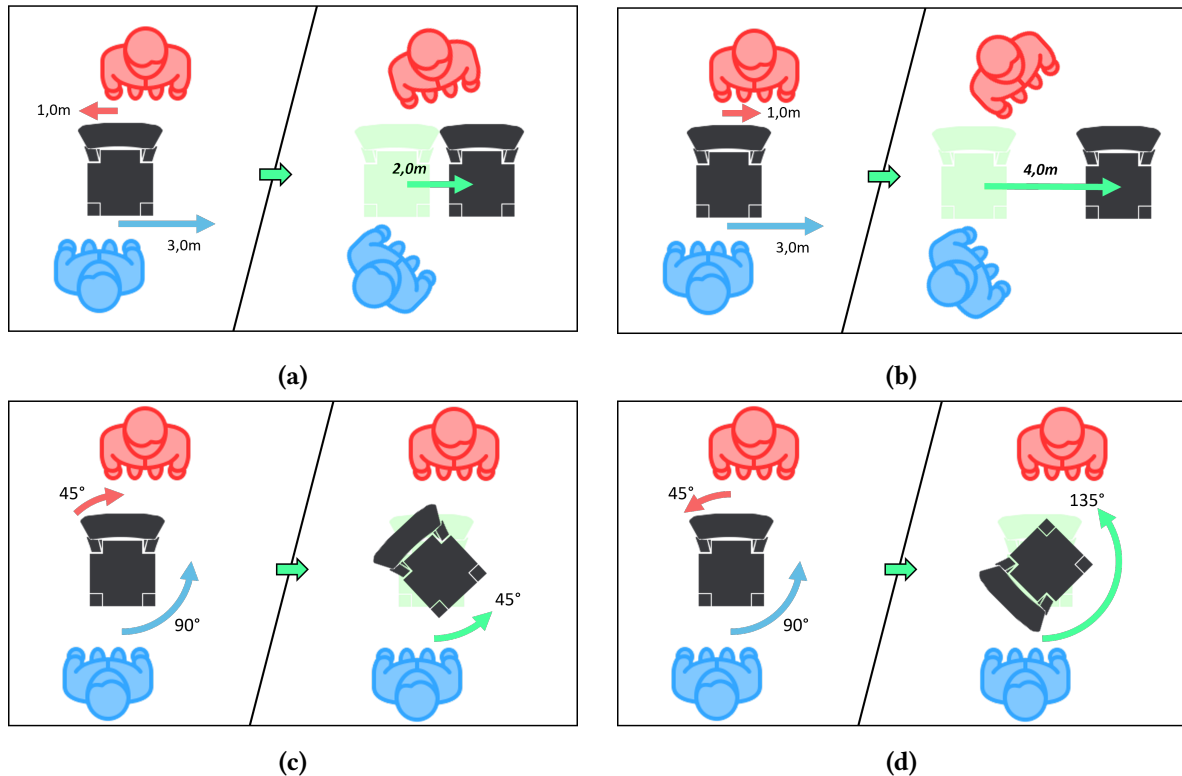


Figure 4.1: *Sum* merge policy examples. In (a), users perform antagonistic translations which are subtracted and the bigger magnitude one dictates the main direction of movement. (b) represents the users performing more synchronized movements, which enhance one another resulting in a bigger magnitude transformation in the direction both were translating it. Same is visible with rotations, antagonistic inputs (c) result in a smaller rotation in the direction of the higher magnitude input. Synchronized rotations (d) add to each other and quickly accelerate the object in one direction.

4.2.2 Mean / Average

In regards to the 3.2.2, instead of users enhancing each other's transformations, they limit them as long as the transformation requests differ. If users were to perform the same action at the same time, that individual transformation would be applied, instead of duplicating the magnitude like in the sum. Manipulations on different DOFs get halved as the input from the other user in that specific DOF for the calculation is 0. Figure 4.2 depicts the same cases that were shown with the *Sum*, where the mentioned reduction in magnitude is apparent. Translations in opposite directions quickly hinder the partner's manipulation, while synchronizing on the direction smooths the magnitudes between the users. The same behaviour occurs with rotations, motivating group members to synchronize their actions. Since users reduce each other, the overshooting issues mentioned by Ruddle, Savage, and Jones [8] should be less noticeable for this merge policy.

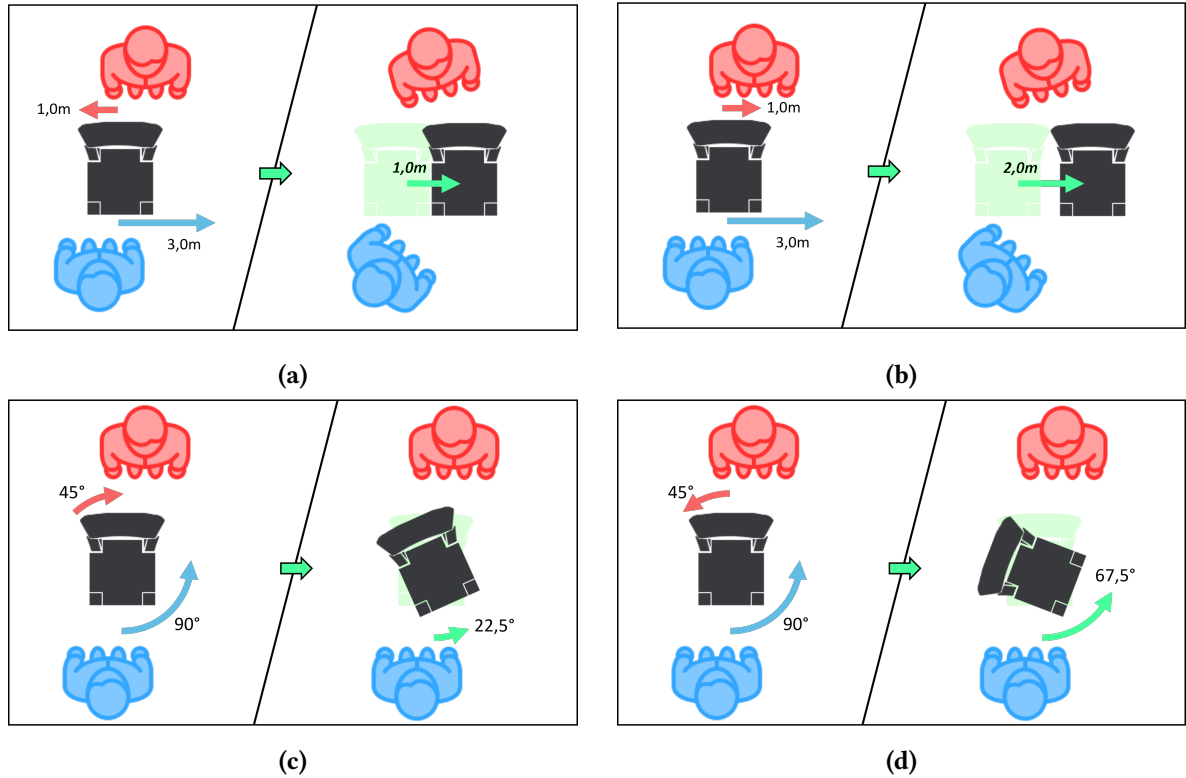


Figure 4.2: Average policy transformation examples. Antagonistic transformations between the users (a,c) quickly hinder the partner’s transform. Synchronized actions (b,d) are more gradual than the *Sum* and similar to the individually performed action as they get closer in magnitude.

4.2.3 Weighted Average

In the 3.2.3 section, issues with the formula used by Riege et al. [23] were brought up as it did not behave as intended, which would be to increment a user’s weight relative to the magnitude of their performed transformation. Consequently, this thesis implements a linear weight assignment formula, described in Equation 4.1. Given u_i is the input for a specific user, $|u_i|$ the magnitude of the corresponding transform, and n the total number of users, the weight w_i of each individual user is given by the percentage of their magnitude over the sum of magnitudes from all users. The weight for a user is maximized when the other user is not performing any manipulation, behaving exactly like the *Sum* for translations in this given case (rotations will still be calculated by the average).

$$w_i = \frac{|u_i|}{\sum_{j=1}^n |u_j|} \quad (4.1)$$

In figure 4.3, the translational behaviour of the *Weighted average* policy is represented, as rotations are calculated using the *Average*, thus using the same implementation used in Riege et al. [23]. Antagonistic manipulations are not as detrimental as they would be with the *Average*, since if magnitudes differ enough the larger one will be predominant over the other. On the other hand, synchronized manipulations do not propel the object as much in comparison to the *Sum*, but if magnitudes differ it will not reduce the result as much as the normal *Average*.

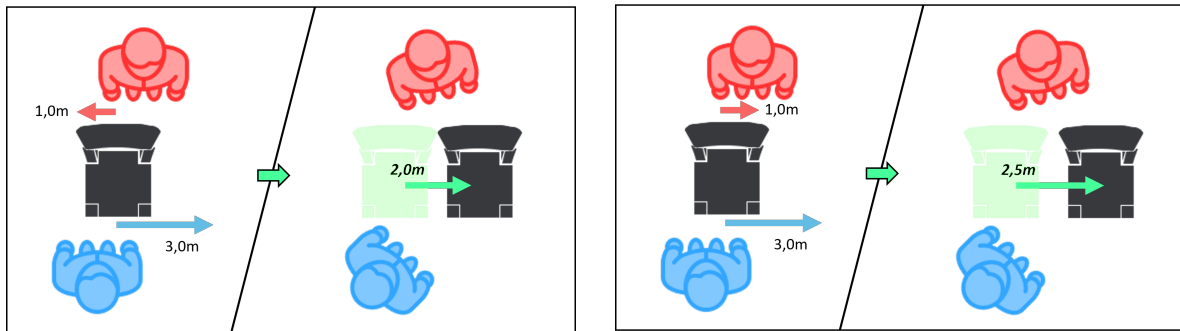


Figure 4.3: Weighted average translation examples. Antagonistic translations give priority to the user with a bigger magnitude, thus reducing their impact. Coordinated translations in which magnitudes are different reduce the resulting magnitude less than the *Average*, while working in the exact way were the actions of the users equal.

4.2.4 Comparison of merge policies

Figure 4.4 provides a comparison between all of the implemented merge policies. When user's inputs work on different degrees of freedom or together on the same one (Figures 4.1, 4.2 and 4.3), the conditions provide different results. On the other hand, if the inputs from the users are antagonistic, all merge policies provide the same result. The *Sum* policy adds the user's input together so its resulting magnitude is the biggest among the implemented policies. When working on different degrees of freedom, the *Average* performs exactly like half of the *Sum*, as users do not share any DOF and thus their input is 0 when averaging the transforms. The *Weighted Average* policy stands in between of the other two, behaving closer to the *Sum* as input magnitudes differ more, and akin to the *Average* as the magnitudes get similar in value.

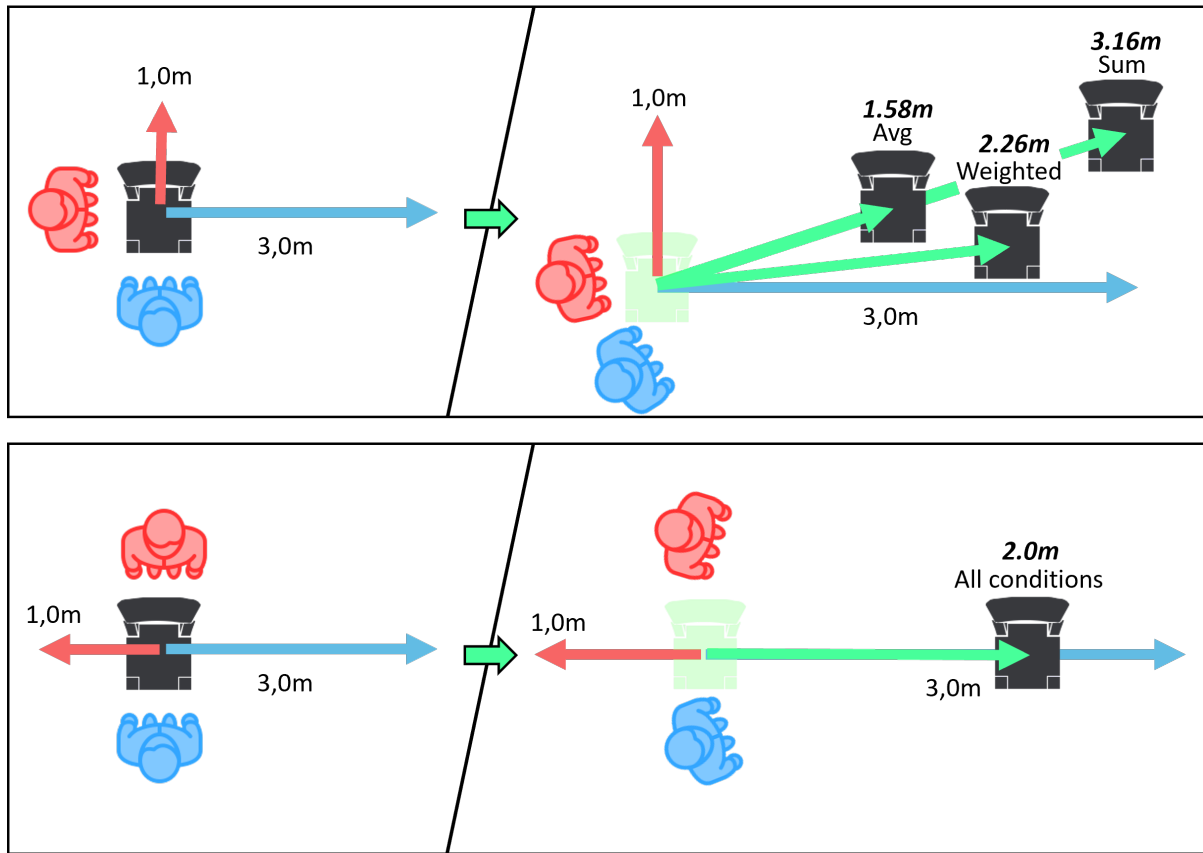


Figure 4.4: Comparison of all implemented merge policies. Antagonistic manipulations will result in the same end transformation for all the conditions (bottom), while actions on different degrees of freedom have diverse outcome.

4.3 Networking

The system is setup following a server/client structure, while using MQTT and a pub/sub architecture to communicate with one another. All the topics are described in Table 4.1. Clients (Hololenses) send the input performed in each frame to the server, which performs the corresponding calculations based on the current condition and broadcasts the task object transformation back to the Hololenses. This creates an *active replication* [11] structure, in which the object with network authority (master copy [15] or entity [11]) is on the server and is replicated in all of the clients, thus any alterations made always have the same result in all of the clients.

Client	Server	Topic	Description
publisher	subscriber	deltaTransform	Framewise input from client for merge calculations
subscriber	publisher	transform	Task object current position broadcast.
subscriber	publisher	instantiate	Task object and target object instantiation position/rotation (trial start).
subscriber	publisher	delete	Task object and target object deletion (trial end).
publisher	subscriber	camera	Camera position for logging purposes.
publisher	subscriber	grab	Grab distance information.
both		cues	Local task object information for visualization on the other client.
both		interaction	Grab status information to other client to control object lock.

Table 4.1: MQTT topic list

4.4 Coordinate System Synchronization

A Vuforia image target was used to synchronize the coordinate systems of both HoloLens. Users had to look at a printed QR code which creates a synchronized region where any child object will appear in the same place in real space for any of the users.

4.5 Cues

Different visual cues are provided to the users. Microsoft's Mixed Reality Toolkit (MRTK) already implements a raycast pointer as well as object selection feedback which indicates whether the user has the object selected. Due to the inability for users to move the object unless both have it selected, a colored outline is shown depicting which of the users has the object currently selected (Figure 4.5). Outline colors are local and will be opposite for the users at the same time: if a user has the object selected, the outline will be blue for that user, and orange for their partner, who does not have it selected. As both users select the object the outline disappears in order to not interfere with the docking task. Furthermore, when both users have the task object selected, colored lines connecting the task object with each of their local copies are shown, which provide continuous feedback on the other user's actions, depicted in Figure 4.6.

4 Implementation



Figure 4.5: Colored outline when only one user has the object selected. An orange outline signifies the other user has it selected and that user is missing, and a blue outline the opposite case.

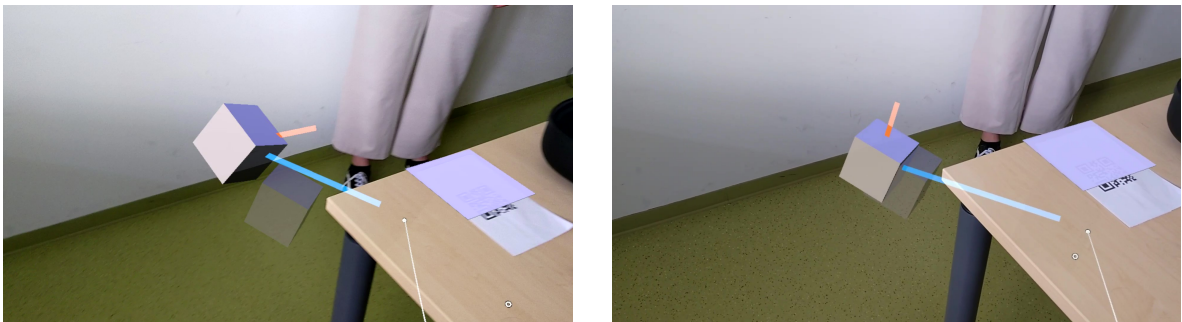


Figure 4.6: Colored lines connecting the task object with each of the user's local copies. Blue signifies the same user's local copy and orange points to the other user's.

5 Experimental Comparison

This chapter describes the task used, then specifies how the pre-study and the main study were conducted. Lastly, it reports the results obtained and their corresponding statistical analysis.

5.1 Study Design

The study was conducted as a controlled lab experiment. Merge policy was the evaluated within-target factor, with the commonly used approaches *Sum*, *Mean* and *Weighted Average* as the conditions. In order to provide a sturdy baseline for comparison in Augmented Reality, the study was designed to answer the following research questions:

- RQ1 Performance:** How do the different merge policies differ in terms of completion time?
- RQ2 Subjective Workload:** Does the perceived workload vary among the conditions?
- RQ3 User Experience:** Which is the preferred policy?

5.2 Apparatus

Each participant used a Hololens2. The Unity server and the docker MQTT server were run in a high end pc with Windows 10 installed. The QR code that participants needed to scan was placed on top of a table at a height of 110cm. When participants scanned the code, the table was moved away in order to prevent any movement obstruction it could cause during the task. Both the Hololenses and the server were connected to a private 5GHz WiFi to ensure bandwidth availability.

5.3 Task

As both the pre-study and the main study tasks are akin to each other, the common components will be described here and the task completion criteria will be further discussed on each of the corresponding sections.

For all the conditions, participants had to perform a docking task . They had to collaboratively move a chair from a fixed starting position to a randomly chosen target. The chair had a bounding box of

5 Experimental Comparison

$5cm \times 10cm \times 5cm$. The task object always appeared in the same fixed position, with no rotation, on top of the previously scanned QR code, marked with a fixed white rectangle to the users. A chair was selected as the object to be manipulated as chairs have been found to have similar accuracy and better time performance to previously used colored edge tetrahedra in docking tasks, probably due to a reduction of perceptual complexity due to being more familiar and less symmetrical [7]. On the same note, Schneider et al. note that "filled shapes contain perceptual information that can be used to facilitate visual recognition" [24].

Target object positions were calculated individually per group and their order randomized among conditions. The eight target positions were selected from a $1m \times 1m \times 1m$ cube with center on the fixed spawn position. Each position was randomized inside each one of the cube's eight quadrants, obtained by bisecting each of its faces. There was a minimum distance of $50cm$ from the fixed spawn position and rotations were randomized for each target. Figure 5.1 shows an example of the all the target positions for a group, with one being selected per trial.



Figure 5.1: An example of the task target positions. Target positions were situated in the quadrants of a cube with edges of size $1m$. The target positions cannot be closer than $50cm$ from the start position of the task object. Only one of them was selected for each trial.

As previous research has pointed out, "pairs that adopted the shared interaction felt more involved in the task than pairs that adopted the independent strategy" [21]. Participants that made use of shared interaction also had a significantly higher number of simultaneous manipulations, as seen in Figure 5.2. Therefore, as we are interested in combination as action integration, task collaboration between participants is forced by only allowing object movement upon its selection by both users. Color coded visual feedback was provided to indicate when only one user was grabbing the object at a time [25], on top of the feedback locally provided by MRTK upon object selection.

5.4 Pre-study and Tolerance level

Previous work has shown that "some participants were more accurate than others, although at the cost of longer trial completion times" [7]. For the sake of a better comparison, a pre-study was conducted in order to determine the appropriate tolerance level for both position and rotation errors, making the main study focus on completion times with a comparable accuracy. The participants of the pre-study

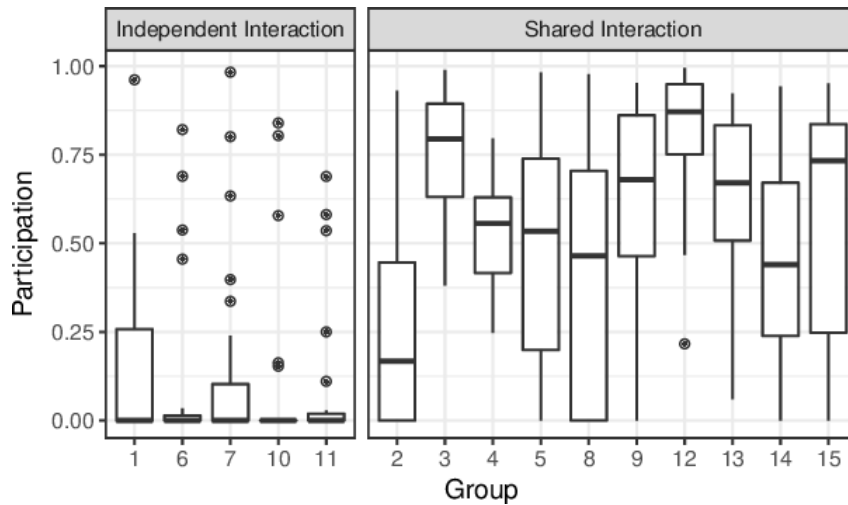


Figure 5.2: Participation score (higher means more simultaneous manipulations). Taken from Grandi et al. [21]

were three dyads formed of unpaid university members. Two were female-male and one female-female. Five of the participants had sizeable previous HoloLens experience and only one did not have any.

The pre-study consisted of eight trials for each of the conditions, for a total of 24 trials. Condition order was defined by the use of a Latin square to counterbalance learning curve. In order to minimize its effects even more, a tutorial/test environment for each condition was first presented to the users, with a simplified docking task using two cubes. Users were able to repeat the tutorial task until comfortable with the merge policy (dyads repeated the tutorial up to a maximum of two times, depending on their previous HoloLens experience). Participants were asked to perform the docking task as accurately as possible, with no regards to completion time. Trial completion was indicated by the participants via vocal feedback when they felt happy with the positioning.

The high accuracy tolerance threshold selected for the main study was based on the worst performing condition defined by its average position and rotation errors. The largest errors in both position and orientation were found in the Sum condition. From the overall combined errors for the trials in the condition, the third quartile (Q3) was selected as the threshold ($8mm$ and 4°).

We will also take into account a second threshold, defined as *low accuracy threshold*, focused on finding when the participants have moved the object "close enough" to the target object, as some related work has pointed to the docking task being divisible in two sub-tasks [20]: **1.** Moving the object close to the target and **2.** accurately docking it from this nearby position. Ruddle, Savage, and Jones [8] also sub-divide their task into smaller phases in order to compare performances in different sub-tasks. The threshold is once again calculated by the worst performing condition, and it is derived from the accuracy obtained at the knee point of the graph described by the position error. Once again we select the third quartile of the data as the threshold for the study ($16cm$). This threshold does not include a rotation check as it will be highly influenced by the target rotation and alter the objective of using this additional threshold.

5.5 Main Study

5.5.1 Participants

Thirty-six participants (15 female, 21 male) between 19 and 33 years ($M = 24.33$, $SD = 4.00$) were recruited from our local university. Thirty two identified themselves as students, two as research assistants and two as PhD from a wide range of fields including politics, economics, biology, computer science, law, psychology, etc. Participants formed eighteen dyads in which all of them knew their partner beforehand. The gender distribution of the dyads was 7 male-male, 6 female-male and 5 female-female. Participants rated their AR experience ($M = 1.86$, $SD = 1.12$) and AR HMD experience ($M = 1.41$, $SD = 0.90$) on a Likert scale from 1 ("No experience") to 5 ("Very experienced"). Seventeen participants normally wear glasses or contacts but only ten of them used them with the Hololens.

5.5.2 Task

Analog to the pre-study, each condition consists of eight trials, for a total of 24. Condition order is again defined by the use of a balanced Latin square to counterbalance learning curve. Preceding the trials for each condition, a training task was presented to the users in order for them to get used to the merge policy. Therefore, cubes were presented as the object to dock and the target position alternated between two set positions. The task has the same completion conditions (thresholds) as the actual trials. Participants were able to perform this training task until they felt comfortable to move onto the task.

In the main study, trial completion is achieved when participants dock the task object within the high accuracy threshold from the target object. When this is achieved, both the target object and task object disappear and the next trial for the current condition is presented after a countdown.

5.5.3 Dependent Variables

As accuracy was analyzed in the pilot study and a threshold is set based on previous work for the study, performance will only be evaluated based on completion time. Completion times to reach each of the thresholds as well as the difference between them are analyzed in order to determine if a merge policy performs better for a specific sub-task of the docking task. Subjective workload for each of the conditions are measured via the NASA Task Load Index (NASA TLX). On the same page, the User Experience Questionnaire (UEQ) is used to evaluate user experience. On top of this, a short structured interview is conducted for each dyad after all the trials are performed to gain insights on more subjective matters (i.e. task preference, overall opinion of collaboration needed, strategies used).

5.5.4 Procedure

The procedure took place according to Figure 5.3. Participants were welcomed and presented with the purpose and procedure of the study as well with the consent form and demographic questionnaire to fill. After that, they were given a short presentation on the basics of the Hololens 2, as well as the task to be performed. A second presentation was used to explain the first condition to the participants. After adjusting the Hololens for their eyes, they started the training task for the corresponding condition, during which they could get comfortable with the selected merge policy. Subsequently, eight trials of the actual task were performed, measuring completion time for each. After the task was complete, the participants filled a NASA TLX as well as an UEQ in order to rate the corresponding condition. This procedure was then repeated for each of the remaining conditions, starting with their respective presentation. After all conditions were completed, a short structured interview was performed.

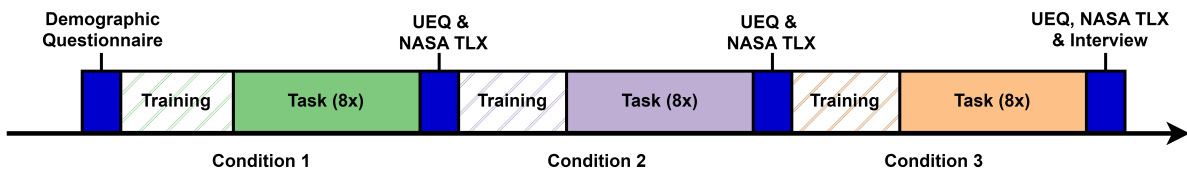


Figure 5.3: Study participants first filled a demographic questionnaire. Hololens adjustment was performed and for each condition they completed a introductory training task, after which 8 docking trials were completed. When the task was concluded, participants evaluated each condition via the NASA TLX and UEQ questionnaires. After all conditions were performed, a structured interview was conducted. Diagram adapted from [16]

5.6 Results

5.6.1 Performance

Performance in the task is measure via completion time. Both thresholds (low accuracy and high accuracy) were calculated separately, furthermore, the time between the low accuracy threshold and the high one that dictates task completion was also accounted for. Interesting metrics such as the number of re-grabs and the learning rate will also be analyzed.

Total completion time

Total task completion times for the accurate threshold can be visualized in Figure 5.4. As the completion times are not normally distributed (Shapiro-Wilk test), Friedman’s test was used for statistical analysis. The test showed no significant differences among the three conditions ($\chi^2(2) = 1.333, p = 0.513$).

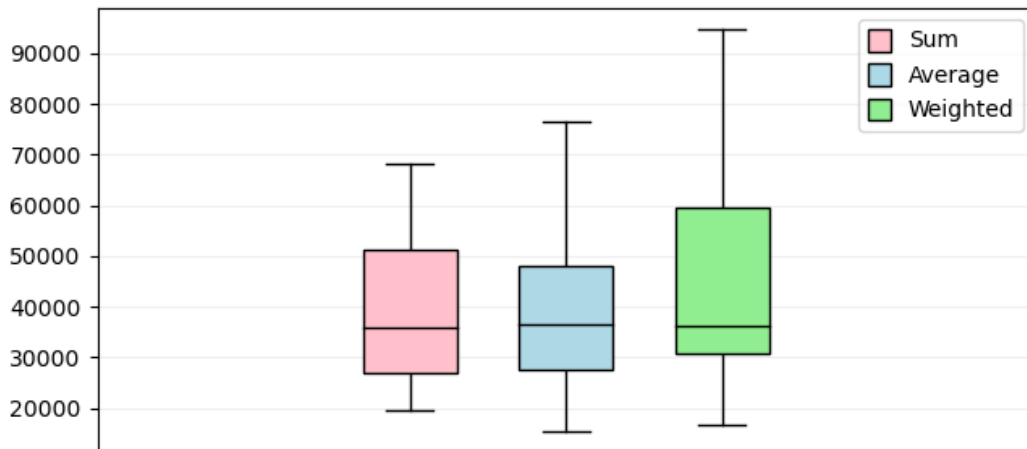


Figure 5.4: Total task completion time (high accuracy threshold).

Big threshold completion time

Figure 5.5 depicts the completion times for the low accuracy threshold. Since the normal distribution of the data was rejected (Shapiro-Wilk test), statistical significance was analyzed via Friedman’s test. Results for the test showed no significant differences among the three conditions ($\chi^2(2) = 0.777, p = 0.678$).

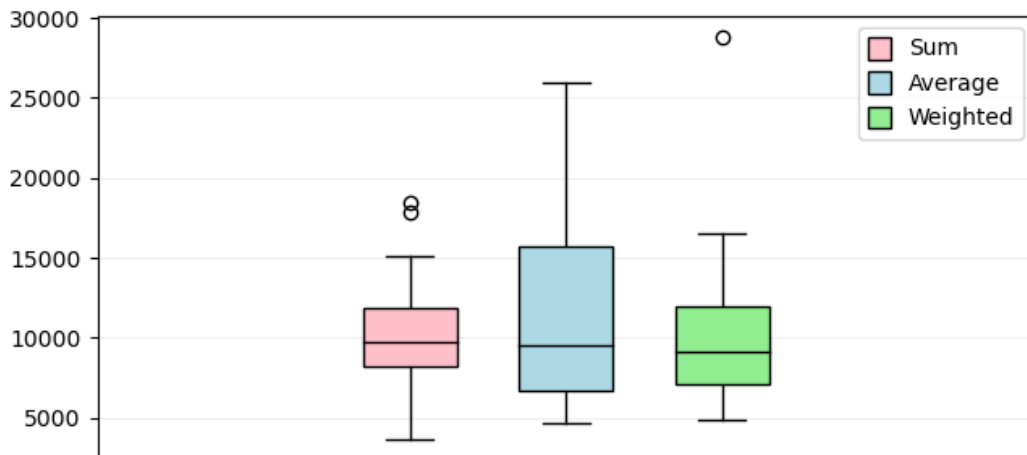


Figure 5.5: Low accuracy threshold completion time.

Accurate completion time

Completion time between both thresholds can be visualized in Figure 5.6. As the completion times are not normally distributed (Shapiro-Wilk test), Friedman's ANOVA was used for statistical analysis. The test showed no significant differences among the three conditions ($\chi^2(2) = 1.444, p = 0.485$).

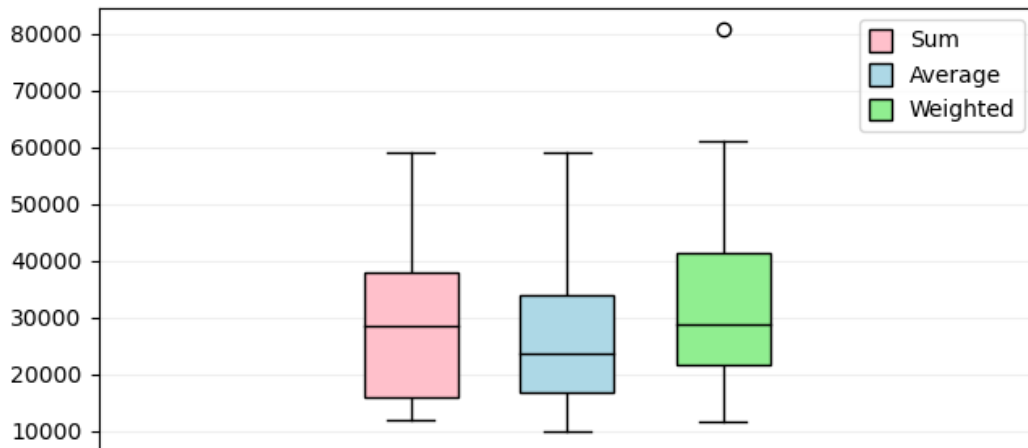


Figure 5.6: Completion time between low and high accuracy threshold.

5.6.2 Number of re-grabs

The number of re-grabs performed by the users is represented in Figure 5.7. A re-grab is counted when any of the users releases the task object to later select it again. Since normality for the data was rejected, a Friedman's test was conducted. No statistical significance was found among the conditions ($\chi^2(2) = 1.444, p = 0.485$).

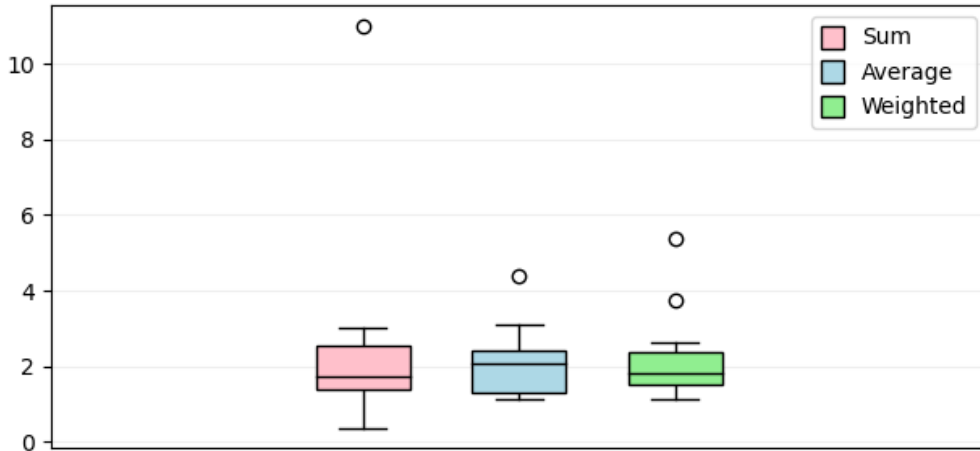


Figure 5.7: Average number of re-grabs per trial for each condition.

5.6.3 Learning rate

Learning rate is measured in two ways, first comparing the task completion times based on the order and secondly by comparing the trials for each of the tasks once again based on the order they were presented. Since the normality distribution of the data is rejected, Friedman tests will be used for statistical analysis. Significance was found for the completion times based on condition order ($\chi^2(2) = 14.778, p = 0.000618$), as well as the in-between trial completion time ($\chi^2(7) = 24.1111, p = 0.00109$). Post-hoc analysis revealed significance between the 1st and 3rd presented conditions ($p = 0.0028$), as well as the 1st and 8th trials of the first presented condition ($p = 0.0115$). No statistical significance among the trials in the 2nd or 3rd presented conditions were found, with the same being true between the 1st and 2nd presented conditions, or the 2nd and 3rd.

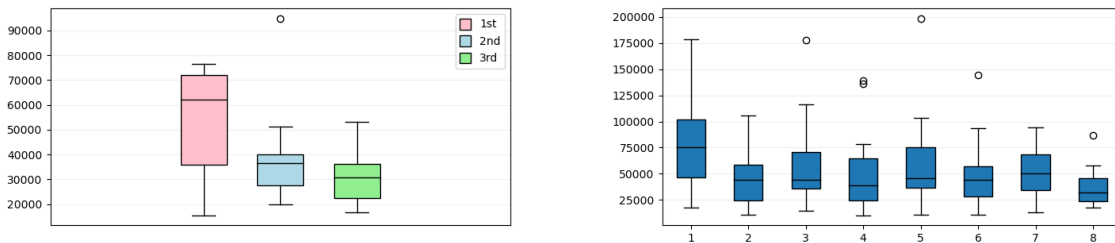


Figure 5.8: Learning rate results. The first plot (left) shows the completion times based on condition order (i.e. 1st, 2nd, 3rd). Second graph (right) plots the task completion time for each of the trials of the first condition presented to each group.

5.6.4 Subjective Workload

Friedman’s test shows no statistical significance among conditions in terms of the global subjective workload ($\chi^2(2) = 0.394, p = 0.821$), as represented in Figure 5.9. Further inspection of the individual dimensions shows again no statistical significance among any of the conditions in any of them, shown in Figure 5.10.

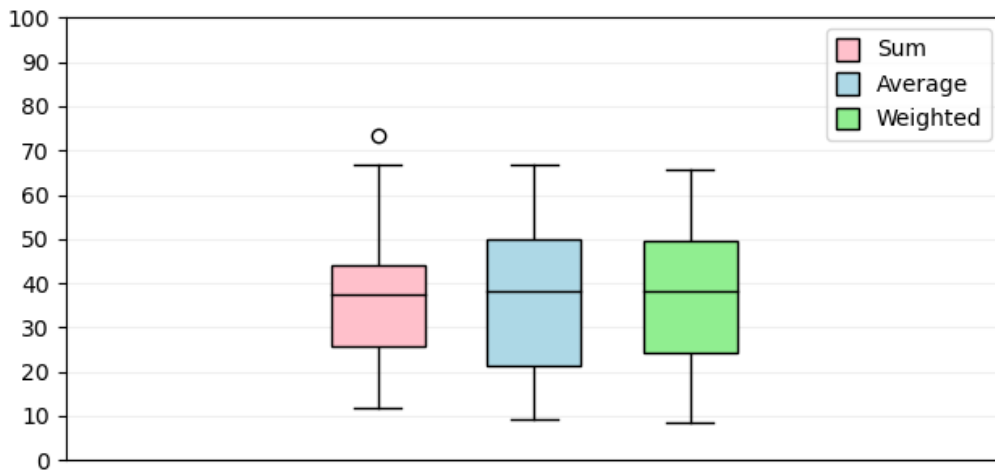


Figure 5.9: Global results for NASA TLX.

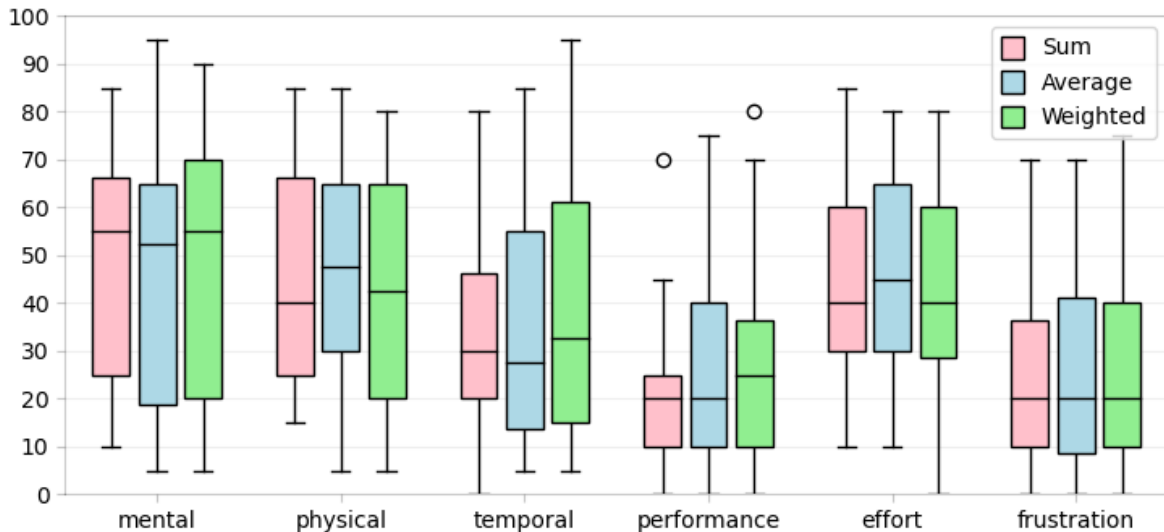


Figure 5.10: Dimension results for NASA TLX.

5.6.5 User Experience

Similar to the Subjective Workload results, Friedman's test showed no statistical significance in any of the dimensions of the UEQ. The overall results are illustrated in Figure 5.11.

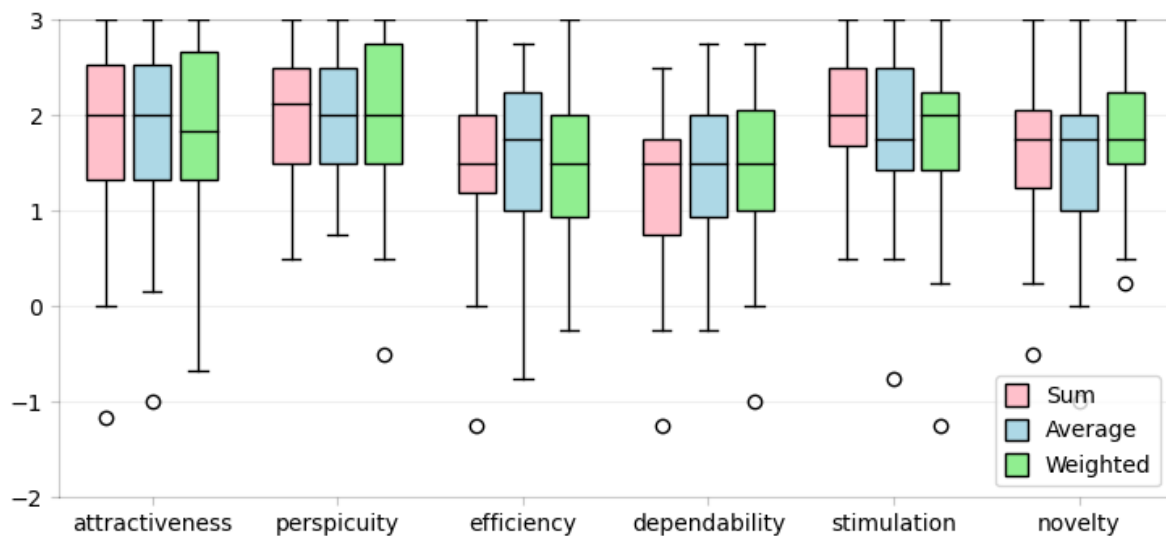


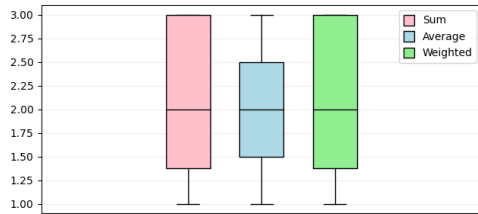
Figure 5.11: Dimension results for UEQ.

5.6.6 Closing Interview

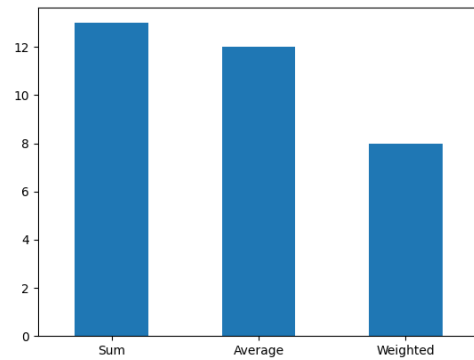
In the closing interview, participants were asked to rank the three merge policies based on their general preference of use. Scores were assigned to the conditions based on the ranking received, with the first one receiving 3, the second one 2 and the worst one 1. Any tie in scores assigns half the points of the sum of the tied places (i.e 2.5 for a tie for best, 1.5 for a tie for second). The results for this evaluation can be observed in Figure 5.12. Friedman's test shows no statistical significance for the condition rating ($\chi^2(2) = 0.0624, p = 0.9692$).

Further questions had the participants evaluate the conditions in regards to different criteria: completion time, accuracy performance and coordination. Given that most participants had a hard time getting to a conclusion or only mentioned the best, scores will not be assigned to the corresponding criteria, but the sum of all preferred conditions will be counted instead. the bar charts for each of the criteria can be found in Figure 5.12.

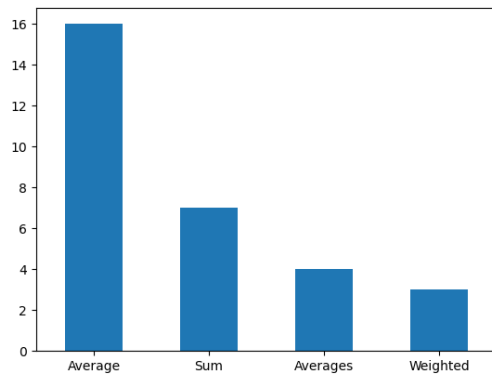
5 Experimental Comparison



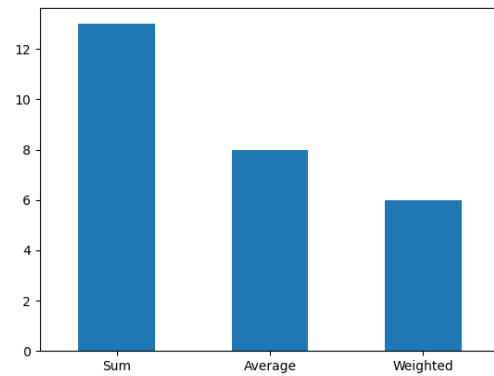
(a) Merge policy preference by users.



(b) User evaluation in regards to completion time.



(c) User evaluation in regards to accuracy.



(d) User evaluation in regards to coordination needed.

Figure 5.12: Users evaluated each merge policy in the closing interview. Different metrics were evaluated such as completion time (b), accuracy (c) and coordination (d), as well as their overall preferred technique (a). The overall preferred condition is scored on a scale from 1 to 3 with 3 being the best value, while the rest are a count of the mentioned preference.

6 Discussion

6.1 Pre-study

The pre-test only looked at the quantitative data as to find the correct threshold for the main study. Therefore, only performance, and more specifically accuracy, will be discussed. In this case, completion time is non-comparable since pairs had different perceptions of what a "good enough" placement is. While the number of participants does not allow for any meaningful statistical comparison, as accuracy will be a threshold and thus non-measurable in the main study, some degree of comparison of the results may prove helpful.

As the participants were overall very experienced with the Hololens, there were no issues with the gestures that needed to be performed in order to grab the task object and the training task seemed sufficient for them.

Table 6.1 shows the overall quantitative results that will be referenced. There are three main results that will be looked at: position and rotation accuracy, knee timestamp and position, and number of re-grabs. Position and rotation accuracy refer to the error of the task object respective to the target object when participants deemed the trial done. Knee position is calculated as the point of maximum curvature of the (usually) downwards trending position accuracy, examples of this position can be observed in Figure 6.1. Number of re-grabs indicates the amount of times the object was released and grabbed again by one or both of the participants, indicating a period of non-interaction.

For the accuracy error, conditions performed in accordance to the hypotheses. The Average condition had the best accuracy, while the Sum had the worst, with the Weighted average being in between and pretty comparable to the Average. A similar conclusion can be drawn from the knee results, where the Sum condition was the fastest to achieve it, the Average the slowest, and the Weighted once again in between but closer to the better performing condition. On the other hand, the knee position indicates that the Average condition had a smoother movement path from starting position to target, while the Sum and Weighted conditions trace back their movements in some way. This behaviour can be observed on the position distance being bigger or smaller, since a sudden movement that makes the task object move further away from the target object will be interpreted as the knee position.

The number of re-grabs is quite similar in all conditions, with Weighted having the least. There seems to be a minimum number of re-grabs for the task as it is shared by all conditions, as positioning the task object precisely requires a few perspective changes. The weighted condition's difference could be due to the inherent transition from a faster method when making bigger moves to a more precise one when performing smaller moves, as can be seen in Figure 6.1 on the right side plot.

	Sum	Average	Weighted
Position accuracy	6.103 <i>mm</i>	4.548 <i>mm</i>	4.685 <i>mm</i>
Rotation accuracy	3.119°	2.350°	2.544°
Q3 position	8.361 <i>mm</i>	5.621 <i>mm</i>	6.599 <i>mm</i>
Q3 rotation	4.166°	2.994°	3.182°
Knee timestamp	44.47 <i>s</i>	64.57 <i>s</i>	49.55 <i>s</i>
Knee Q3 position	123.023 <i>mm</i>	83.188 <i>mm</i>	158.857 <i>mm</i>
# of re-grabs	3.148	2.740	2.296

Table 6.1: Pre-test quantitative results

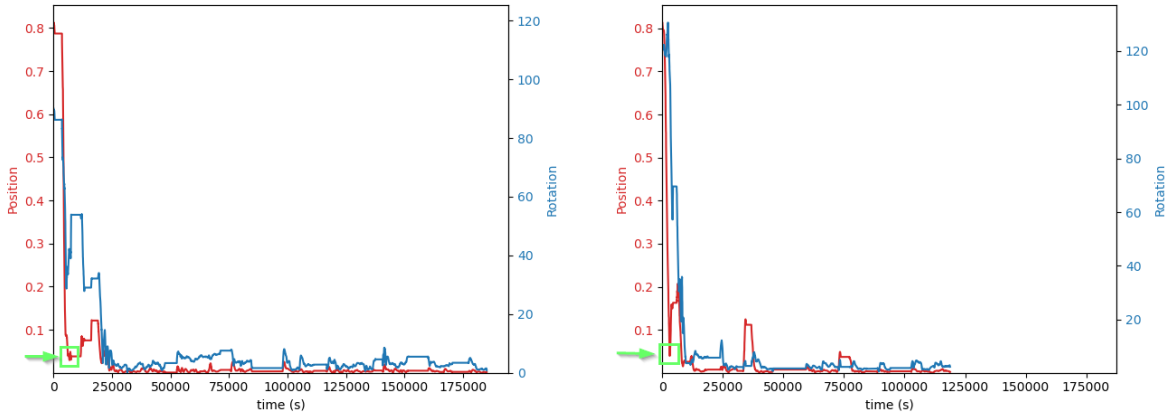


Figure 6.1: Knee position examples of trials.

6.2 Main study

6.2.1 Performance

Surprisingly, no statistical significance was found among the conditions for the task completion time. The same behaviour was observed by Ruddle, Savage, and Jones [8], the only other study that compares different merge policies. Following their approach, the task was sub-divided in two by the *low accuracy threshold*, as described in section 5.4.

For the low accuracy completion time, no significant results are shown. Nonetheless, the interquartile range (IQR) of the *Average* condition is bigger than for the other conditions. This variability is likely due to group coordination, as the *Average* policy was the most hindering when not performing similar actions, therefore being highly dependent on how coordinated participants were. For the accurate completion time, while no significance was again found. With no visible differences in IRQ or range for the conditions. The overall percentage of time employed to achieve each of the thresholds is depicted in Figure 6.2.

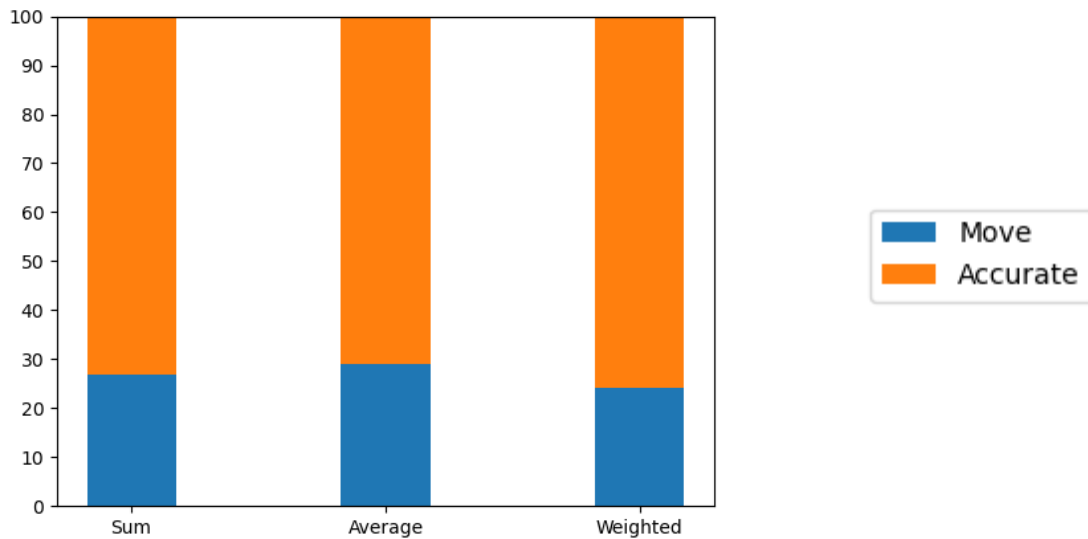


Figure 6.2: Percentage of total completion time taken for each of the docking sub-tasks.

In regards to learning rate, improvement over the trials of the first condition is apparent, likewise over the conditions which was to be expected [12]. Improvement over the trials has also been reported by Grandi et al. [6], Wieland et al. [9] and Ruddle, Savage, and Jones [8]. While analysis of the later trials was useful in the latter, in this case there are no significant differences among trials outside the first and last of the first presented condition. Since positions and rotations are randomized among the conditions, removing this trial would mean taking the object from all conditions, and thus opens the possibility of introducing a bias towards the other quadrants.

6.2.2 Subjective Workload

No statistical differences were found among the conditions in regards of subjective workload. Very much like in the quantitative analysis, users perceive the workload among conditions to be about the same. While non-significant, in the dimensions of *Temporal Demand* as well as *Performance*, the *Sum* has a smaller interquartile range (IQR), most likely due to transformations performed with this conditions being faster than for the averages.

Results of the NASA TLX can still be analyzed in regards to the implemented task. Users evaluate the *Frustration* dimension as low, likely due to the fast nature of the trials, both the *Temporal Demand* and the *Performance* dimensions further support that possibility. There are no significant differences between *Mental Demand* and *Physical Demand*, which indicates similar requirements for the mental processing of the target position together with the collaborative system and the physical performance of the transformation.

6.2.3 User Experience

Users rated all conditions similarly for all of the dimensions of the questionnaire. Once again there seems that the overall perception of the techniques is the same in terms of the user experience. These results could be explained by the simplicity of the docking task that the users needed to perform, which did not provide users with the time needed to form a more distinct opinion.

6.2.4 Qualitative results

User ratings of their favourite condition follow the same non-statistical significance trend of the questionnaires. However, the reason behind the non-significance for the condition preference is due to the polarization among the users, which have very strong feedback on both the *Sum* or the *Weighted* policies, with the *Average* being their overall second choice. Both the *Sum* and *Weighted* conditions have 12 first places, compared to the 8 obtained by the *Average*.

The closing questionnaire also reveals differences on user perception versus the quantitative results obtained. For completion time, users perceived the *Sum* and *Average* merge policies as overall faster than the *Weighted*. When asked about their condition of choice for accurate placement, the *Average* comes at an overwhelming advantage compared to the other conditions, with some users interestingly grouping both average policies together. The *Average* as a better condition for precise positioning does not come as a surprise, since this idea is prompted in related work by Aguerreche, Duval, and Lécuyer where “some users described the Mean technique as a kind of lowpass filter” [18]. Regarding the amount of coordination needed for each condition, users reported that the *Sum* policy needed a higher degree of collaboration as user’s manipulation escalate one another. This is in line with the results reported by Aguerreche, Duval, and Lécuyer [18], where more difficult conditions (as perceived by the users) motivate a higher degree of communication between them.

Thematic analysis

Users responses and feedback in the closing interview were clustered thematically in the most recurring topics. Since the ratings in the interview were closed-ended questions, analysis on those will be mainly deductive with the conditions as clusters. On the other hand, other open-ended questions were clustered inductively.

Most recurring themes for the *Sum* merge policy were that it was “faster”, “more efficient”, “fun”, “overshooting” and “harder”. *Faster* was used both as an advantage and a disadvantage by users, in a similar manner *harder* made reference mainly to the accurate placement, but some users indicated that the challenge provided entertainment. In regards to the *Average*, common themes are that it was “slow”, “predictable”, “precise” and “intuitive”. Once again, “slow” was used in a positive and negative manner. For the *Weighted* condition, “overshooting”, “no mental image”, “easier” were the most mentioned themes. This analysis supports the polarization present in the preferred merge policy, with the *Sum*

and *Weighted* obtaining a lot of positive and negative feedback, while the *Average* received a moderated one, with the salient feature that it was slower than the others.

“*Learning curve*” was another highly mentioned topic, with users referring to it multiple times when asked about any ranking of the conditions represented previously in Figure 5.12. Regarding this theme, users also indicated that they developed a strategy during a certain period of time from the start of the study, improving in performance and collaboration when this method was integrated.

Multiple different “*strategies*” were developed. Users mainly subdivided the accurate placement of the object to one person holding as steady as possible while the other one manipulates it. Most also reported that it was better to start the task at opposite sides (180°) to have a faster target acquisition, and then move to either 90° or 120° to perform the accurate placement, with the angles being dependent on the geometric shape of the object’s bounding box. On the other hand, a handful of users indicated that their strategy was to just follow intuition or do so randomly.

Regarding the presented “*visual cues*”, a majority of the participants reported that the outline feedback when individually grabbing the object was most helpful. Since a lot of users had issues with having the object forcefully deselected when moving their hand out of the tracking area, or were not performing the selection gesture correctly, the colored feedback allowed them to communicate on who had the issue and promptly solve it. The lines from the task object to the local user copies were mostly not used or it was done subjectively to check the DOFs available to their partner or know what they were doing. A lot of users thought the lines had some significance with the forces being applied to the object.

Another common theme was to perform the task in “*two subtasks*”, the first which consisted in the users moving the task object as fast as possible to a nearby location to the target, to later re-grab/reposition and perform the accurate adjustments. This feedback further signifies the subdivision of the docking task into a fast manipulation to reach a position close to the target and a precise one, to complete the docking, as indicated by Salzmann, Jacobs, and Froehlich [20].

6.3 Limitations

In order to improve the learning rate and ideally eliminate it, a longer introductory task would be needed. This would be best performed by having the users interact with the system individually first before of any of the conditions, in order to get used to the Air Tap gesture as well as the field of view of the HoloLens. There were a lot of user errors due to the novelty of the HoloLens as they tried to perform the Air Tap gesture, mainly done with a hand position that felt comfortable for them but unrecognized by the HoloLens as the selection gesture. The most prominent examples observed are depicted in Figure 6.3. A longer training task was not possible here as participants would have to perform the study over multiple days due to its length.

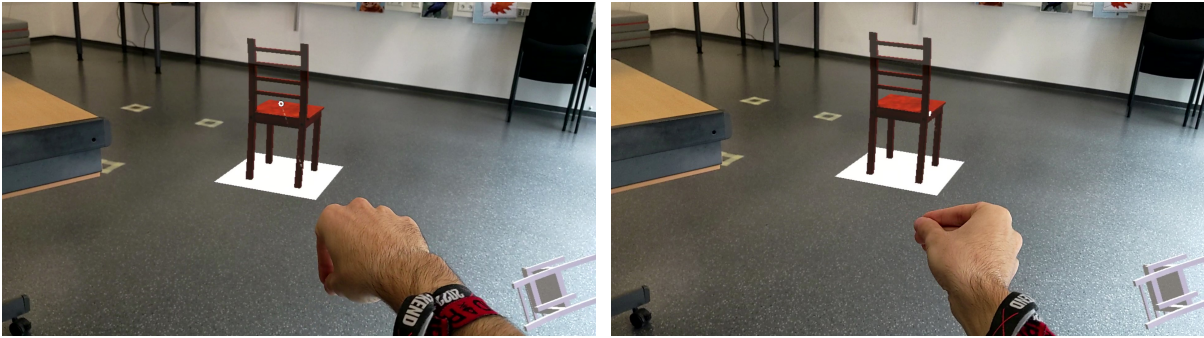


Figure 6.3: Incorrect selection gestures performed by the users.

6.4 Future Work

Due to time constraints, the Bent Pick Ray [23] for visual feedback was unable to be implemented. Multiple participants reported that they thought the feedback lines had any relation to the forces applied to the object, while they were merely position and distance indicators. The Bent Pick Ray would most likely present the users with better feedback, as shown by Riege et al.

Aguerreche, Duval, and Lécuyer [18] indicate that the use of bimanual interactions techniques “is also responsible for the feeling of a better preparation for the real task” [18]. Since other literature points towards bimanual techniques performing worse in tasks where no scaling is needed, and the RTD-3 from Aguerreche, Duval, and Lécuyer is fixed on scaling, it would be interesting to compare multiple interactions techniques to identify which one feels more natural for manipulation.

As users made a clear distinction between the subtasks of the docking task, merge policies that exploit the subtasks may have increased performance. Hybrid merge policies that change depending on context, in addition to ones that combine them depending on the manipulated DOFs of each user (e.g. *Sum* non-colliding DOFs and *Mean* the ones with conflicts) will be interesting to analyze. The implementation of more complex merge policies is also motivated, which would allow to understand what the tradeoffs of intuitiveness and task compliance are.

7 Conclusions

Computer supported collaborative work (CSCW) has recently been drawing more attention to AR setups. With the HoloLens2 release, head mounted displays are now in a functional place in terms of AR development, as they provide a hands-free interaction with a now reasonable field of view. While literature presents a lot of comparisons between the action integration methods, there is barely any research on which merge policy works better for the *Combination* approach. Nonetheless, multiple merge policies have been used in literature, even if no direct comparison has been analyzed. The three most common ones are *Sum*, *Mean* and *Weighted Average*. The *Sum* merge policy combines all user input together, the *Mean* averages the manipulations performed among the number of users, and *Weighted Average* first assigns weights to the users to later average their actions based on them. Related work which compares action integration approaches uses different merge policies for the comparison, and while obtaining non-significant results in their comparisons, conclusions cannot be extended to the merge policies themselves.

In order to provide a valid comparison, thus setting a baseline for future and more complex mathematical functions, this thesis implements the three aforementioned merge policies into a study prototype in order to evaluate them. A docking task was performed by the participants in order to measure the performance, subjective workload and user experience for each condition.

In terms of performance, all merge policies performed similarly, demonstrating no statistical significance. While task subdivision into phases has shown significance in literature, due to the simplicity of the docking task implemented in the study prototype, completion times of all the conditions remained non-significant. A minimum number of re-grabs was observed for all condition, indicating that multiple perspective and gesture changes are required for the the docking task in HMDs. Both subjective workload and user experience were rated similarly among conditions, showing no user preference by those metrics. On the other hand, analysis of the closing interview provided interesting insights that need to be considered in future work. Merge policy preference was polarized between the *Sum* and the *Weighted* conditions, with the *Average* being constantly placed as second. The *Average* condition was preferred by users when making more precise adjustments, and the *Sum* required the most coordination due to constant overshoots. Multiple users reported that they were incapable of mentally portraying the *Weighted* policy. More complex, less intuitive merge policies should be studied carefully as their perceived performance can highly depend on the ability to understand them.

This work provides the first head mounted display comparison among merge policies, in addition to interesting subjective observations that contrast the quantitative results. A general comparison and baseline is thus provided for the future implementation of more complex or hybrid techniques, on top of research questions and paths for future work.

References

- [1] Mark Billinghurst and Hirokazu Kato. “Collaborative Augmented Reality”. In: *Commun. ACM* 45.7 (2002), 64–70. ISSN: 0001-0782. DOI: 10.1145/514236.514265. URL: <https://doi.org/10.1145/514236.514265>.
- [2] Jens Müller, Roman Rädle, and Harald Reiterer. “Virtual Objects as Spatial Cues in Collaborative Mixed Reality Environments: How They Shape Communication Behavior and User Task Load”. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI ’16. New York, NY, USA: Association for Computing Machinery, May 7, 2016, pp. 1245–1249. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858043. URL: <https://doi.org/10.1145/2858036.2858043> (visited on 10/22/2022).
- [3] Mickael Sereno et al. “Collaborative Work in Augmented Reality: A Survey”. In: *IEEE Transactions on Visualization and Computer Graphics* (2020), pp. 1–1. ISSN: 1077-2626, 1941-0506, 2160-9306. DOI: 10.1109/TVCG.2020.3032761. URL: <https://ieeexplore.ieee.org/document/9234650/> (visited on 10/14/2022).
- [4] Susanna Nilsson, Bjorn Johansson, and Arne Jonsson. “Using AR to support cross-organisational collaboration in dynamic tasks”. In: *2009 8th IEEE International Symposium on Mixed and Augmented Reality*. 2009, pp. 3–12. DOI: 10.1109/ISMAR.2009.5336522.
- [5] Lev Poretzki, Joel Lanir, and Ofer Arazy. “Normative Tensions in Shared Augmented Reality”. In: *Proc. ACM Hum.-Comput. Interact.* 2.CSCW (2018). DOI: 10.1145/3274411. URL: <https://doi.org/10.1145/3274411>.
- [6] Jerônimo Gustavo Grandi et al. “Design and Evaluation of a Handheld-based 3D User Interface for Collaborative Object Manipulation”. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI ’17. New York, NY, USA: Association for Computing Machinery, May 2, 2017, pp. 5881–5891. ISBN: 978-1-4503-4655-9. DOI: 10.1145/3025453.3025935. URL: <https://doi.org/10.1145/3025453.3025935> (visited on 10/15/2022).
- [7] Vanessa Vuibert, Wolfgang Stuerzlinger, and Jeremy R. Cooperstock. “Evaluation of Docking Task Performance Using Mid-air Interaction Techniques”. In: *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*. SUI ’15. New York, NY, USA: Association for Computing Machinery, Aug. 8, 2015, pp. 44–52. ISBN: 978-1-4503-3703-8. DOI: 10.1145/2788940.2788950. URL: <https://doi.org/10.1145/2788940.2788950> (visited on 10/15/2022).
- [8] Roy A. Ruddle, Justin C. D. Savage, and Dylan M. Jones. “Symmetric and asymmetric action integration during cooperative object manipulation in virtual environments”. In: *ACM Transactions on Computer-Human Interaction* 9.4 (Dec. 1, 2002), pp. 285–308. ISSN: 1073-0516. DOI: 10.1145/586081.586084. URL: <https://doi.org/10.1145/586081.586084> (visited on 10/15/2022).

- [9] Jonathan Wieland et al. “Separation, Composition, or Hybrid? – Comparing Collaborative 3D Object Manipulation Techniques for Handheld Augmented Reality”. In: *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 2021, pp. 403–412. DOI: 10.1109/ISMAR52148.2021.00057.
- [10] Márcio S. Pinho, Doug A. Bowman, and Carla M.D.S. Freitas. “Cooperative object manipulation in immersive virtual environments: framework and techniques”. In: *Proceedings of the ACM symposium on Virtual reality software and technology*. VRST ’02. New York, NY, USA: Association for Computing Machinery, Nov. 11, 2002, pp. 171–178. ISBN: 978-1-58113-530-5. DOI: 10.1145/585740.585769. URL: <https://doi.org/10.1145/585740.585769> (visited on 10/15/2022).
- [11] W. Broll. “Interacting in distributed collaborative virtual environments”. In: *Proceedings Virtual Reality Annual International Symposium ’95*. Virtual Reality Annual International Symposium ’95. Research Triangle Park, NC, USA: IEEE Comput. Soc. Press, 1995, pp. 148–155. ISBN: 978-0-8186-7084-8. DOI: 10.1109/VRAIS.1995.512490. URL: <http://ieeexplore.ieee.org/document/512490/> (visited on 11/27/2022).
- [12] Shumin Zhai and Paul Milgram. “Quantifying coordination in multiple DOF movement and its application to evaluating 6 DOF input devices”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’98. USA: ACM Press/Addison-Wesley Publishing Co., Jan. 1, 1998, pp. 320–327. ISBN: 978-0-201-30987-4. DOI: 10.1145/274644.274689. URL: <https://doi.org/10.1145/274644.274689> (visited on 10/15/2022).
- [13] Doug A. Bowman et al. *3D User Interfaces: Theory and Practice*. USA: Addison Wesley Longman Publishing Co., Inc., 2004. ISBN: 0201758679.
- [14] ARCore. *ARCore Content Manipulation*. URL: <https://developers.google.com/ar/design/content/content-manipulation>.
- [15] David Margery and B. Arnaldi. “A General Framework for Cooperative Manipulation in Virtual Environments”. In: (Jan. 1, 1999). ISSN: 978-3-211-83347-6. DOI: 10.1007/978-3-7091-6805-9_17.
- [16] Jonathan Wieland. “Comparing Different Approaches for the Collaborative Manipulation of Virtual 3D Objects in Augmented Reality”. MA thesis. 2019. URL: <https://hci.uni-konstanz.de/teaching/theses-and-current-topics-for-projects/theses/master-theses/overview/#c444691>.
- [17] Thierry Duval, Anatole Lecuyer, and Sebastien Thomas. “SkeweR: a 3D Interaction Technique for 2-User Collaborative Manipulation of Objects in Virtual Environments”. In: *3D User Interfaces (3DUI’06)*. IEEE Computer Society, Mar. 1, 2006, pp. 69–72. ISBN: 978-1-4244-0225-0. DOI: 10.1109/VR.2006.119. URL: <https://www.computer.org/csdl/proceedings-article/3dui/2006/02250069/120mNBV9Ibb> (visited on 10/15/2022).
- [18] Laurent Aguerreche, Thierry Duval, and Anatole Lécuyer. *Evaluation of a Reconfigurable Tangible Device for Collaborative Manipulation of Objects in Virtual Reality*. Accepted: 2013-10-31T10:30:56Z. The Eurographics Association, 2011. ISBN: 978-3-905673-83-8. DOI: 10.2312/LocalChapterEvents/TPCG/TPCG11/081-088. URL: <https://diglib.eg.org/443/xmlui/handle/10.2312/LocalChapterEvents.TPCG.TPCG11.081-088> (visited on 10/18/2022).

- [19] Morgan Le Chenechal et al. “When the giant meets the ant an asymmetric approach for collaborative and concurrent object manipulation in a multi-scale environment”. In: *2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE)*. 2016, pp. 18–22. DOI: 10.1109/3DCVE.2016.7563562.
- [20] Holger Salzmann, Jan Jacobs, and Bernd Froehlich. “Collaborative Interaction in Co-Located Two-User Scenarios”. In: *Joint Virtual Reality Conference of EGVE - ICAT - EuroVR (2009)*. Artwork Size: 8 pages ISBN: 9783905674200 Publisher: The Eurographics Association, 8 pages. ISSN: 1727-530X. DOI: 10.2312/EGVE/JVRC09/085-092. URL: <http://diglib.eg.org/handle/10.2312/EGVE.JVRC09.085-092> (visited on 10/20/2022).
- [21] Jeronimo G Grandi et al. “Design and Assessment of a Collaborative 3D Interaction Technique for Handheld Augmented Reality”. In: *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). Reutlingen: IEEE, Mar. 2018, pp. 49–56. ISBN: 978-1-5386-3365-6. DOI: 10.1109/VR.2018.8446295. URL: <https://ieeexplore.ieee.org/document/8446295/> (visited on 10/21/2022).
- [22] Jerônimo Gustavo Grandi, Henrique Galvan Debarba, and Anderson Maciel. “Characterizing Asymmetric Collaborative Interactions in Virtual and Augmented Realities”. In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). ISSN: 2642-5254. Mar. 2019, pp. 127–135. DOI: 10.1109/VR.2019.8798080.
- [23] Kai Riege et al. “The Bent Pick Ray: An Extended Pointing Technique for Multi-User Interaction”. In: *3D User Interfaces (3DUI’06)*. 2006, pp. 62–65. DOI: 10.1109/VR.2006.127.
- [24] Bertrand Schneider et al. “3D Tangibles Facilitate Joint Visual Attention in Dyads”. In: (), p. 8.
- [25] Daniel Mendes et al. “Mid-air interactions above stereoscopic interactive tables”. In: *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. 2014 IEEE Symposium on 3D User Interfaces (3DUI). MN, USA: IEEE, Mar. 2014, pp. 3–10. ISBN: 978-1-4799-3624-3. DOI: 10.1109/3DUI.2014.6798833. URL: <http://ieeexplore.ieee.org/document/6798833/> (visited on 10/15/2022).

List of Figures

1.1	Hand-held devices hand restrictions and focus requirement. <i>Taken from [5].</i>	1
2.1	Canonical tasks.	3
2.2	Cooperation levels	4
3.1	DOF separation by Pinho, Bowman, and Freitas	7
3.2	Piano mover’s problem [8]	8
3.3	Completion times [8]	9
3.4	Obstacle crossing task [6]	9
3.5	Windshield assembly task [20]	10
3.6	Pick and place task [18]	10
3.7	Results of Aguerreche, Duval, and Lécuyer [18]	11
3.8	Furnishing task [9]	11
3.9	Results of Wieland et al. [9]	12
3.10	Riege et al. weight formula	13
4.1	Sum merge policy	17
4.2	Average merge policy	18
4.3	Weighted average merge policy	19
4.4	Comparison of all implemented merge policies.	20
4.5	Selection outline	22
4.6	Visual feedback	22
5.1	Example of task target positions	24
5.2	Participation score (higher means more simultaneous manipulations). Taken from Grandi et al. [21]	25
5.3	Study procedure	27
5.4	Total task completion time (high accuracy threshold).	28
5.5	Low accuracy threshold completion time.	28
5.6	Completion time between low and high accuracy threshold.	29
5.7	Average number of re-grabs per trial for each condition.	30
5.8	Learning rate results	30
5.9	Global results for NASA TLX.	31
5.10	Dimension results for NASA TLX.	31
5.11	Dimension results for UEQ.	32
5.12	Qualitative results	33
6.1	Knee position examples of trials.	35

List of Figures

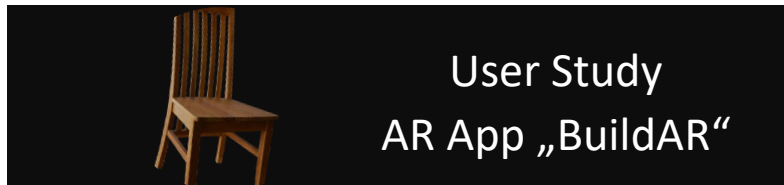
6.2 Time used in each sub-task. 36
6.3 Selection issues 39

List of Tables

- 3.1 DOF separation Pinho, Bowman, and Freitas 7
- 4.1 MQTT topic list 21
- 6.1 Pre-test quantitative results 35

Appendix

A. Welcome Letter



Welcome!

Thank you for participating in our study. In doing so, you are supporting our research significantly. Before we get started, we would like to briefly explain what the study is about and what role you play in it.

Study Aims and Procedure

In this study, we investigate to what extent different mathematical functions have an impact in how user input is combined. For this purpose, you will solve several tasks. After these tasks, we will ask you to tell us about your experiences and impressions. Before the tasks, there will be an introduction to the HoloLens and the specific task to be solved. You can ask questions about the general procedure or the system during this briefing. However, please understand that we cannot answer any questions during the actual task to prevent data bias.

Your interaction data will be saved but no video or audio recordings will be done. In this context, we have prepared a consent form, which is enclosed with this letter. At this point we would like to point out that we are not evaluating you or your performance but are only interested in the suitability of the application.

While not common, it is possible for you to feel “motion sickness”, which starts as a mild headache that keeps growing as you continue interacting with the AR glasses.

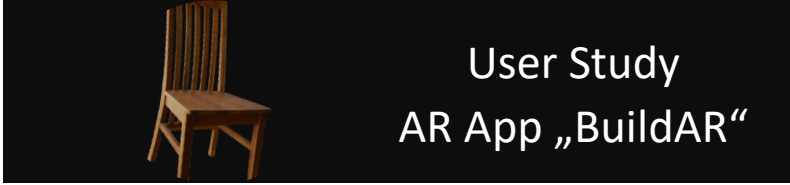
Duration and Compensation

Participation in the study takes approximately one hour and a half. If at any time you feel unwell and wish to end your participation, this is of course possible without giving any reason. In that case, please contact the investigator.

After the study is completed, you will be rewarded with 18€ for your help. We you once again want to thank you for your support!

Human-Computer Interaction Group,
University of Konstanz

B. Consent Form



**User Study
AR App „BuildAR“**

Informed Consent ID: _____

Information about the Investigators

Investigator: _____
Institution: Human-Computer Interaction Group, University of Konstanz

Declaration

I was informed about the aim, content, and duration of the study. Within the scope of this study, personal data will be collected in questionnaires, as well as movement data.

I am hereby informed that the personal data will be treated confidentially and will not be passed on to third parties. After recording, the data will be evaluated by our research team. The publication of the research results in publications or at conferences takes place exclusively in pseudo-anonymized form and at no time allows conclusions to be drawn about you as a person.

According to the GDPR, you can ask for your data to be deleted at any point before it has been processed and used in the report, since we will be unable to do so after this point.

Optional Points *(Please mark with a cross if you agree)*

I agree that recorded data may additionally be used for internal presentation purposes.

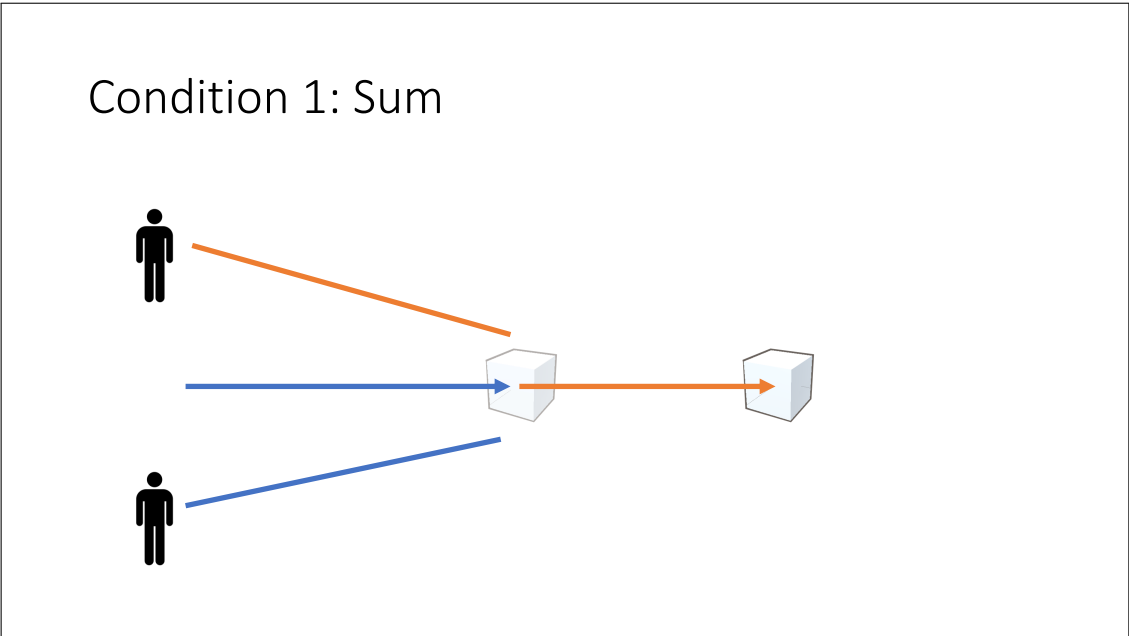
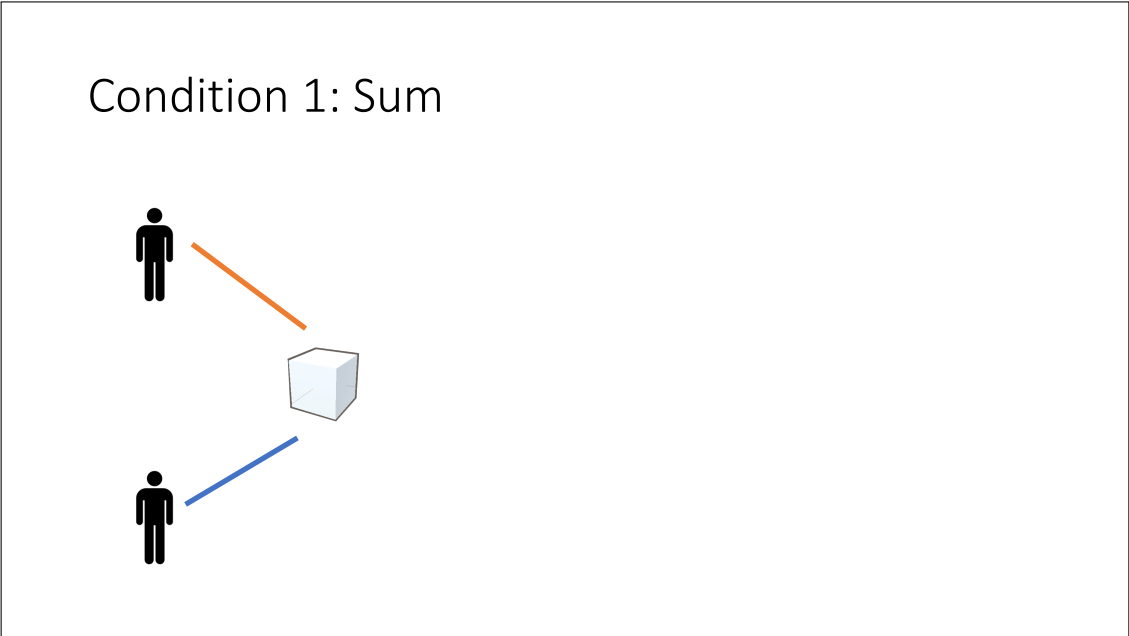
I hereby agree to the items listed under "Declaration" and the optional items I marked with a cross:

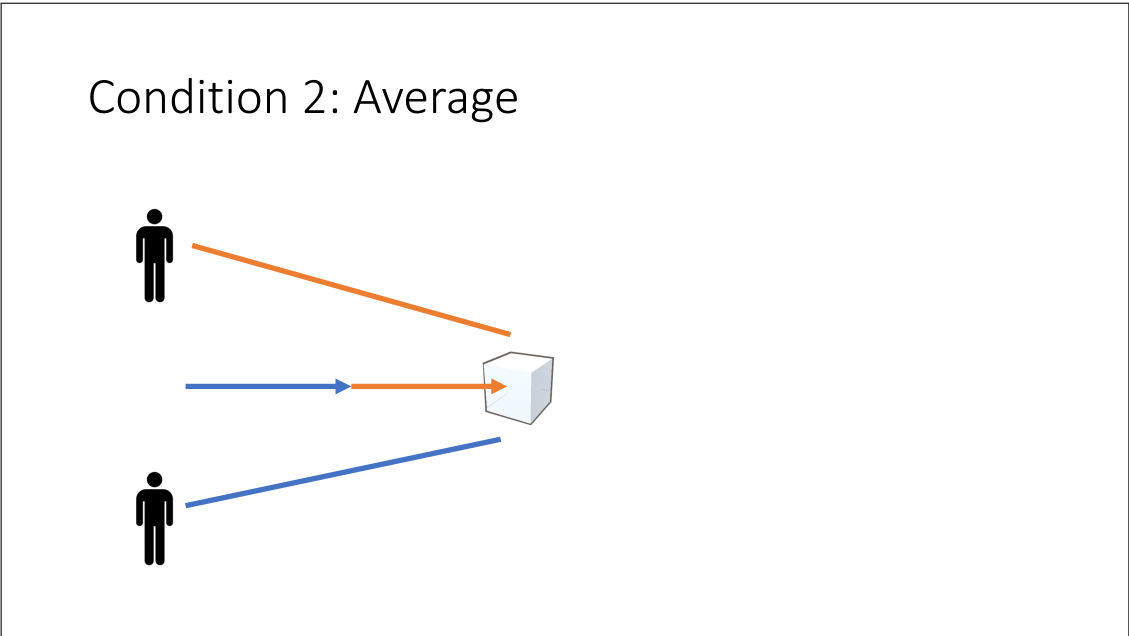
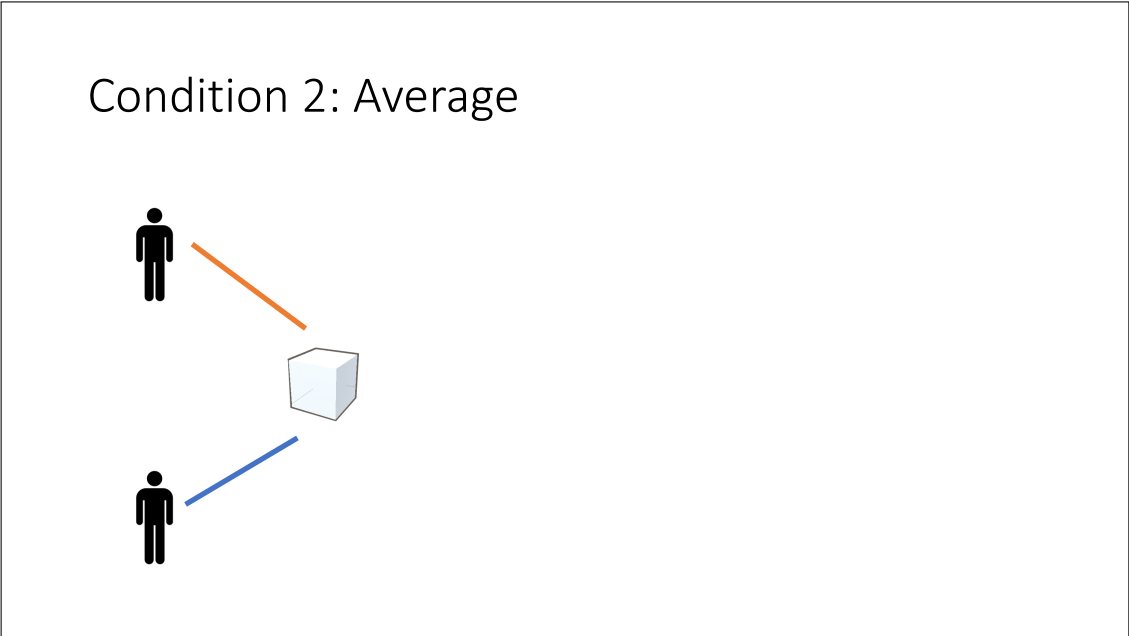
_____ (Name) _____ (Place, Date) _____ (Signature)

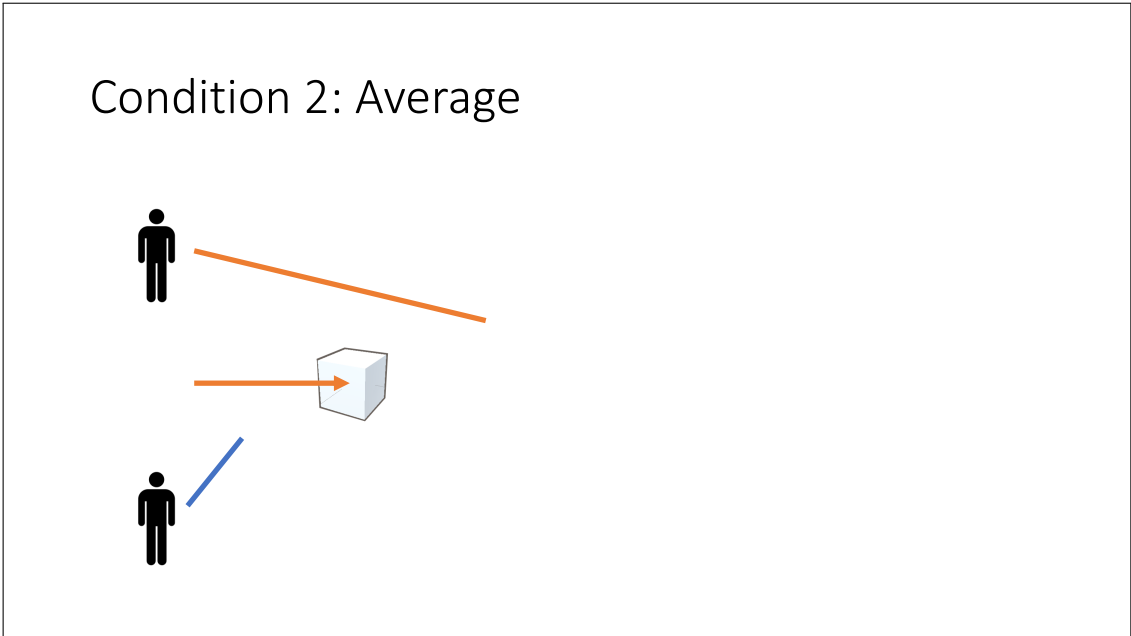
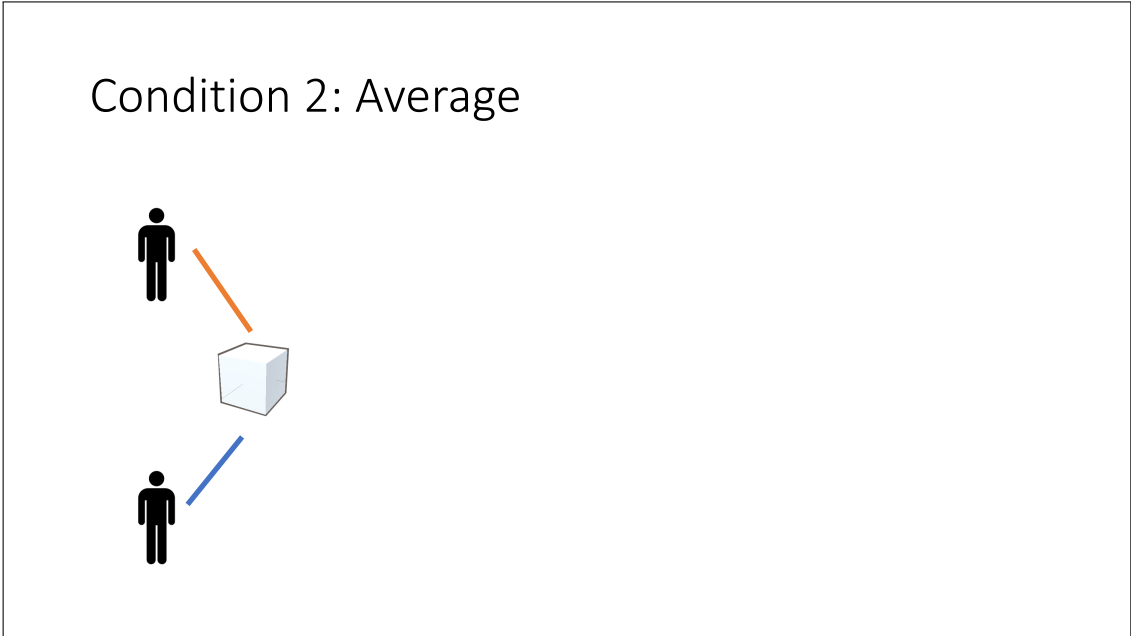
The study investigator hereby agrees to use any data obtained solely for evaluation purposes in the context of this study:

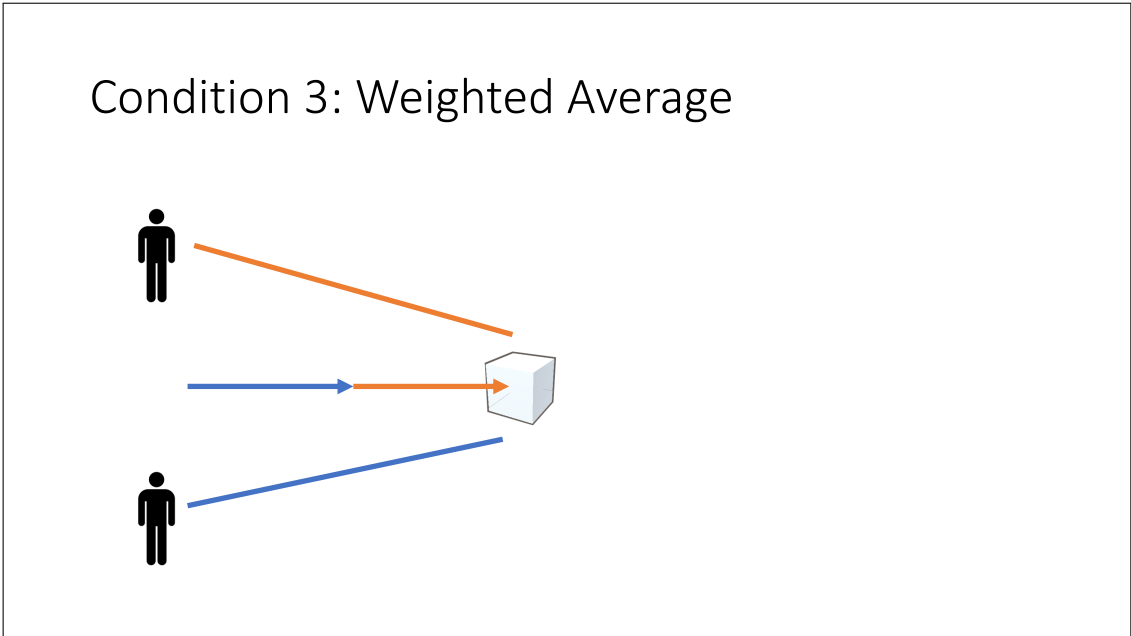
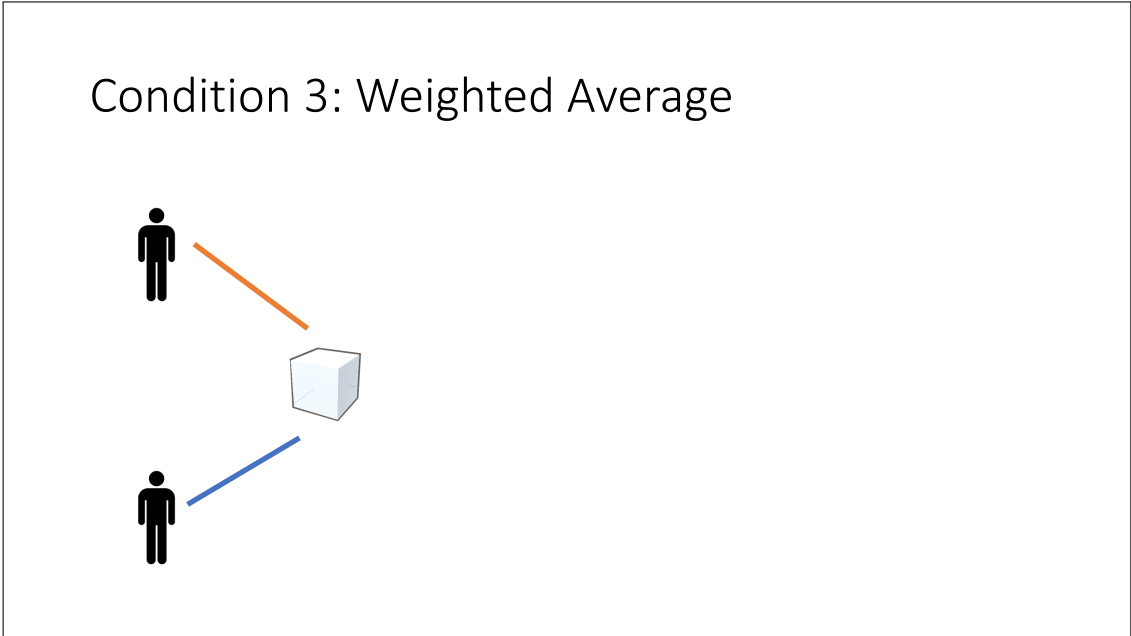
_____ (Name) _____ (Place, Date) _____ (Signature)

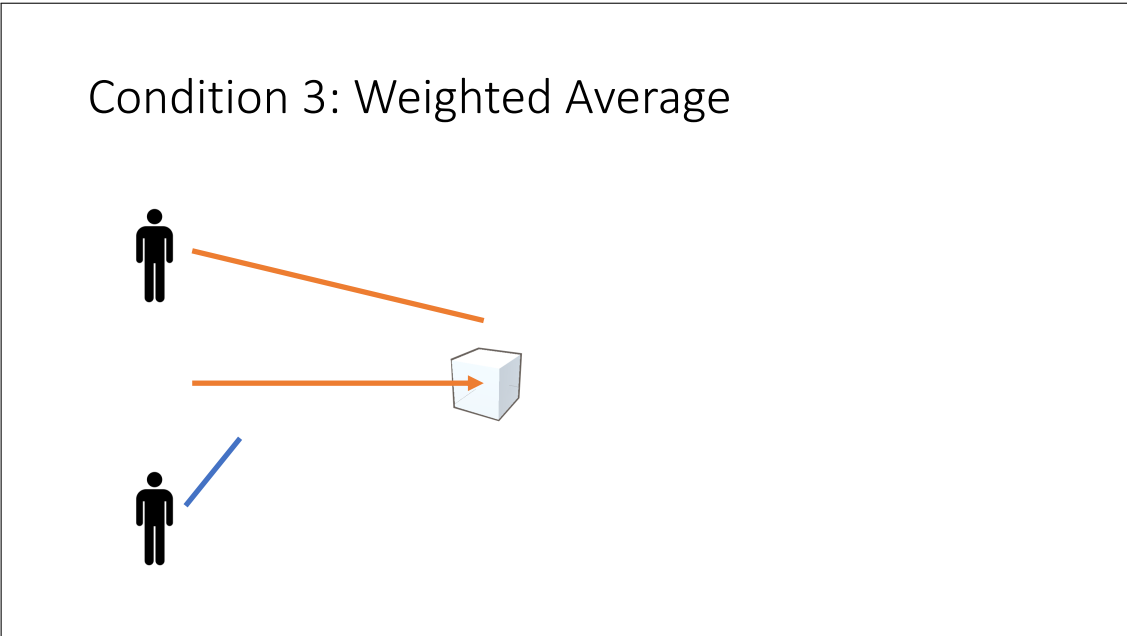
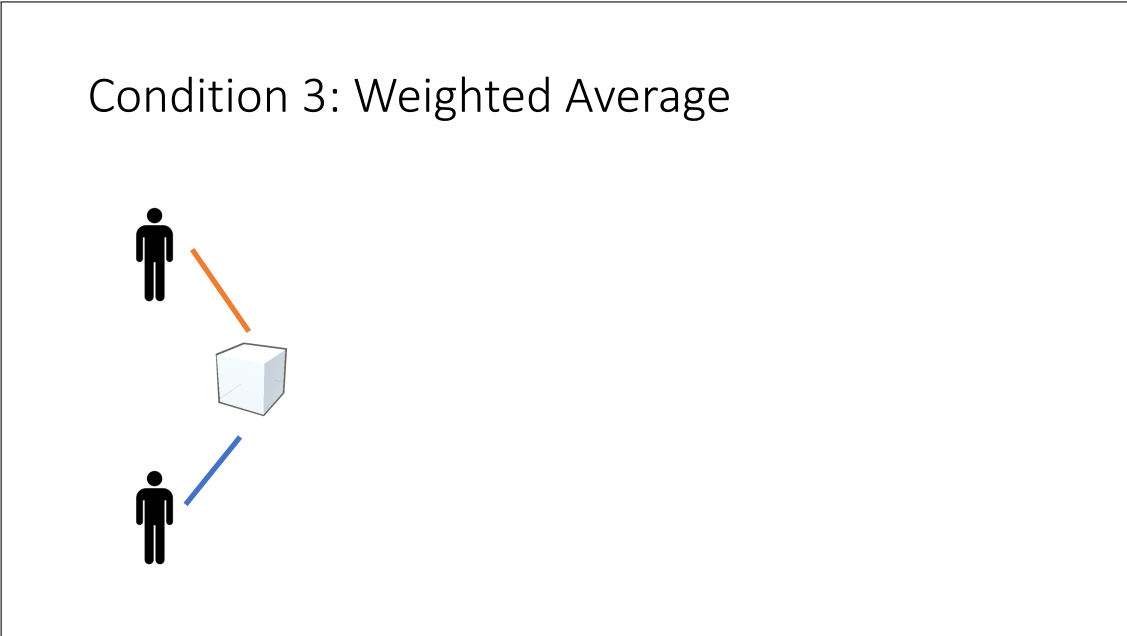
C. Presentation of the Approaches











D. Structured Interview Questions

Observations Study “Collaborative Manipulation”

ID:	ID:
Sum	Sum

ID:	ID:
Avg	Avg

List of Tables

ID:	ID:
Weighted	Weighted

Remind they can answer individually

Favourite Condition. Why?:

Coordination. Which condition did you have to coordinate the least?. Why?:

Speed. Which condition were you fastest with?. Why?:

Accuracy. Which condition were you the most accurate with?. Why?:

Advantages and disadvantages.

Sum	Sum
Avg	Avg
Weighted	Weighted

Strategies. What strategy did you use for each condition. Was it the same for them? E.g. "you rotate I place it"

Sum	Sum
Avg	Avg
Weighted	Weighted

Visual feedback. Was it useful? Did you use it at all? Any ideas on how to improve it?

--	--

Relative positioning. How did you position relative the other person? Did you do so intentionally? Why?

Sum	Sum
Avg	Avg
Weighted	Weighted

Ideas. In the real world you move stuff together and have physical restrictions. This does not happen in the virtual world, what do you think this could be useful for?

--	--