

Separation, Composition, or Hybrid? – Comparing Collaborative 3D Object Manipulation Techniques for Handheld Augmented Reality

Jonathan Wieland*

Johannes Zagermann†

Jens Müller‡

Harald Reiterer§

HCI Group, University of Konstanz

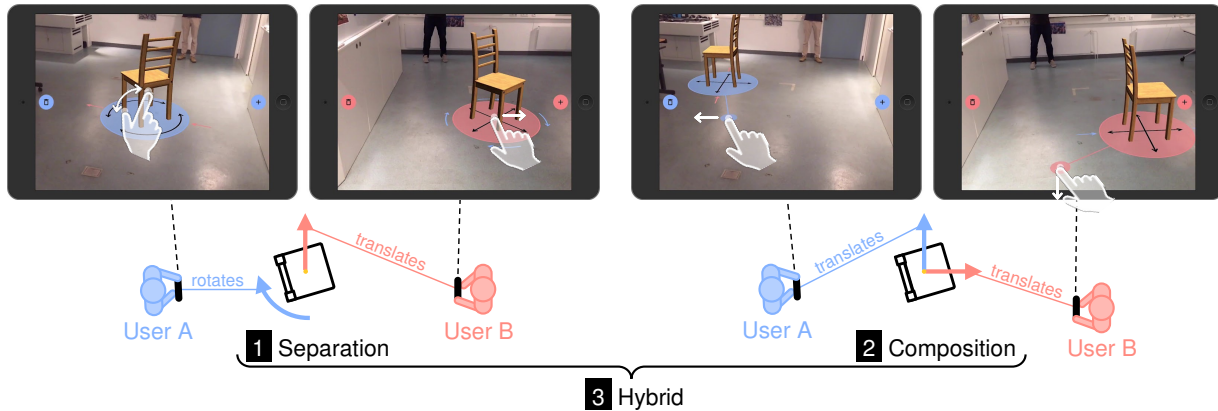


Figure 1: We compared three collaborative 3D object manipulation techniques for handheld AR displays. (1) Separation: Only different manipulation tasks can be performed simultaneously. (2) Composition: Only the same manipulation tasks can be performed simultaneously. (3) Hybrid: Any manipulation tasks can be performed simultaneously. White hands and arrows indicate user input.

ABSTRACT

Augmented Reality (AR) supported collaboration is a popular topic in HCI research. Previous work has shown the benefits of collaborative 3D object manipulation and identified two possibilities: Either separate or compose users' inputs. However, their experimental comparison using handheld AR displays is still missing. We, therefore, conducted an experiment in which we tasked 24 dyads with collaboratively positioning virtual objects in handheld AR using three manipulation techniques: 1) Separation – performing only different manipulation tasks (i. e., translation or rotation) simultaneously, 2) Composition – performing only the same manipulation tasks simultaneously and combining individual inputs using a merge policy, and 3) Hybrid – performing any manipulation tasks simultaneously, enabling dynamic transitions between Separation and Composition. While all techniques were similarly effective, Composition was least efficient, with higher subjective workload and worse user experience. Preferences were polarized between clear work division (Separation) and freedom of action (Hybrid). Based on our findings, we offer research and design implications.

Keywords: Augmented reality, mobile devices, collaborative 3D object manipulation.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / Augmented Reality; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Collaborative Interaction

*e-mail: jonathan.wieland@uni-konstanz.de

†e-mail: johannes.zagermann@uni-konstanz.de

‡e-mail: jens.mueller@uni-konstanz.de

§e-mail: harald.reiterer@uni-konstanz.de

1 INTRODUCTION

Augmented Reality (AR) applications can create interactive experiences that enable users to see virtual objects seamlessly integrated into their physical 3D environment [4]. Technological advances have made it possible to provide such AR experiences on regular smartphones and computer tablets. Due to the visualization and interaction qualities of AR, a growing number of applications has applied the technology in various areas, like education, architecture, marketing and entertainment [5]. Additionally, research has identified AR as a promising candidate for next-generation collaborative systems [6, 29]. For example, considering 3D object manipulation, every collaborator adds an additional perspective on the scene that can reduce problems regarding, e. g., the positioning of distant objects, occlusion by other virtual or physical elements, or limitations of interaction techniques [23]. Also, previous research has shown that collaborative 3D object manipulation can increase precision, performance [13], efficiency, and usability [23].

However, the collaborative manipulation of virtual objects lacks one key characteristic of physical object manipulation: When collaborators manipulate a physical object together, the object rigidly connects them and, thus, helps to coordinate the collaborators' joint actions [27] – when collaborators manipulate a virtual object, the object does not provide such mediation [12]. Consequently, manipulation techniques for virtual objects need to address the question of how to integrate collaborators' actions [26] that can have conflicting directions and forces. Previous work [23] has identified two possibilities for this action integration depending on the task related-degrees of freedom (DOFs) (e. g., for position and rotation): 1) DOF separation – collaborator's input is separated (i. e., manipulating mutually exclusive DOFs) or 2) composition of users' actions – collaborators manipulate the same DOFs and their individual inputs are combined using a merge policy. While prior work investigated how to assign and integrate collaborators' individual inputs (e. g., [11, 24, 25]), they mainly focused on immersive and projection-based virtual environments (VEs). Only Grandi et al. [11, 12] considered settings with

handheld AR displays but exclusively studied the composition of users' actions. Consequently, there is, to the best of our knowledge, no experimental comparison of the named possibilities for handheld AR displays being the most prevalent AR devices [10, 28] to date.

In a user study with 48 participants, we, therefore, investigated how different 3D object manipulation techniques influence collaborators' activities when positioning virtual 3D objects. We designed and implemented the three techniques *Separation*, *Composition*, and *Hybrid* that were inspired by previous work – however, in contrast to them, we included mobile AR displays. To provide a realistic scenario, we created a collaborative AR furnishing task (i. e., a 3D docking task) inspired by IKEA's app "Place" [15]. We contribute a deeper understanding of how the techniques influence task *Performance* (RQ1), perceived *Workload* (RQ2), *Participation* (RQ3), and *User Experience* (RQ4). Based on our quantitative and qualitative findings, we provide implications for future research and the design of applications that involve collaborative 3D object manipulation.

2 RELATED WORK

Collaborative 3D object manipulation involves two or more people simultaneously manipulating the same virtual object [7, 20]. Since collaborators are not physically connected by the shared virtual object, collaborative manipulation techniques need to define rules to integrate multiple users' actions [26]. Literature analysis reveals that there are two possibilities to perform this action integration: 1) Separate the manipulation of the object's DOFs between the users [23, 24] or 2) allow simultaneous manipulations of the same DOFs and use a merge policy to compose the users' actions to one resulting manipulation (e. g., [1, 13, 26]). In the following, we discuss related work categorized into these two strategies and their comparison. We include research on augmented and virtual environments as the techniques' underlying mechanics are the same.

2.1 DOF Separation

Pinho et al. [23, 24] explored different variants to perform DOF separation in immersive VEs. One prominent variant for dyads is to separate the DOFs of rotation and position between the users. The authors showed that DOF separation can increase performance and usability compared to single-user manipulation. Similarly, DOF separation was also investigated for single-user techniques. Here, related work [22, 32, 34] showed that a separation of translation and rotation actions can increase performance, outperforming a simultaneous manipulation of all DOFs.

2.2 Composition of Users' Actions

Research that investigates the composition of users' actions allows concurrent manipulations of the same DOFs by implementing a merge policy. These merge policies compose the collaborators' inputs to generate one transformation by typically either computing the (weighted) average [1, 2, 9, 25, 26], the common component [26], or the sum [8, 11–13] of users' inputs.

Ruddle et al. [26] compared participants' performance using the two merge policies average and common component: Using spatially tracked controllers, dyads had to maneuver virtual avatars that carried a large virtual object through VEs. Here, the authors have shown that the policy of common component should be used when users need to execute similar actions and the policy of average when they need to perform different movements. The SkeweR technique by Duval et al. [9] also relies on the concept of multiple people carrying an object. It allows two users to grab a virtual object with a virtual cursor. The users' grabbed points then act like the extremities of a skewer on which the virtual object is impaled. SkeweR then uses the translation information of these points to average the new position and rotation of the "skewed" object. Related to SkeweR, Riege et al. [25] proposed the Bent-Pick Ray as a ray-casting-based

technique for VEs that bends the users' rays during opposing manipulations to continuously preserve the visual connection between their input devices and the manipulated object. As merge policies, they investigated averaging and weighted averaging. They report that most users perceived the averaging as intuitively usable and preferred it over the weighted averaging.

Grandi et al. [13] proposed a technique for handheld devices that computes the sum of users' inputs as merge policy and does not limit the number of collaborators. During the collaboration, users can either manipulate the same DOFs simultaneously or split the manipulation of the DOFs between them. In a first publication, Grandi et al. [13] compared their technique's performance for different group sizes. Participants used the handheld device as an input device to collaboratively manipulate a virtual object in a VE presented on a shared screen. Their study results showed that larger groups tend to specialize in one transformation, which in turn increases accuracy. As future work, they suggest investigating different levels of controlled labor division and hypothesize that forcing a role division (comparable to DOF separation) could increase the accuracy for all team sizes. In a second publication, Grandi et al. [11] introduced a modified version of their technique for handheld AR displays, again using the sum of inputs as merge policy. They compared different input modalities and investigated strategies of pairs during unrestricted collaboration. The authors report that participants were fastest using a hybrid combination of touch gestures and device movements for manipulation. Further, they revealed two main strategies: More than half of the pairs manipulated the objects simultaneously, while the remaining pairs manipulated them individually. Regardless of the strategy, task completion times were similar, but groups that decided to manipulate the objects simultaneously felt more engaged in the task. In a third publication, Grandi et al. [12] adopted their technique for virtual reality (VR) head-mounted displays (HMDs). In a study, they compared an asymmetric setting with one user in immersive VR and the other one in handheld AR with symmetric AR-AR and VR-VR settings. Participants had to solve the same manipulation task with the two strategies the authors revealed in their previous study: Manipulating the objects only independently and only simultaneously. This time, pair performance was significantly lower when participants needed to manipulate the same object simultaneously. However, participants also perceived a significantly lower workload in this condition, which indicates that participants collaborated in a way that distributed the amount of work between them.

2.3 Comparison of both Approaches

Aguerreche et al. [1, 2] introduced a tangible device for a collaborative 3D object manipulation by two users in a projection-based VE. The tangible device can be reconfigured to match the shape of the manipulated virtual object. Participants grab it at different ends as if they would carry a physical object. Movements and rotations of this tangible device are directly coupled to the virtual object. They compared it with a DOF separation approach and an approach using the mean as the merge policy, each relying on spatially tracked controllers as the input device. The authors report that participants were faster with the mean approach than with the tangible device. Further, participants were more accurate with the mean approach and the tangible device than with the DOF separation approach. Interestingly, this finding contrasts with Grandi et al. [13], who concluded that accuracy increases with user specialization.

2.4 Conclusions from Related Work

Our analysis of related work revealed that there are partly contrasting results ([13] vs. [1, 2]) regarding users' performance with the two possibilities for action integration. Furthermore, prior work mainly focused on performance measurements (e. g., task completion time or accuracy) in immersive and projection-based VEs. Only Grandi et al. [11, 12] investigated settings with handheld AR displays and

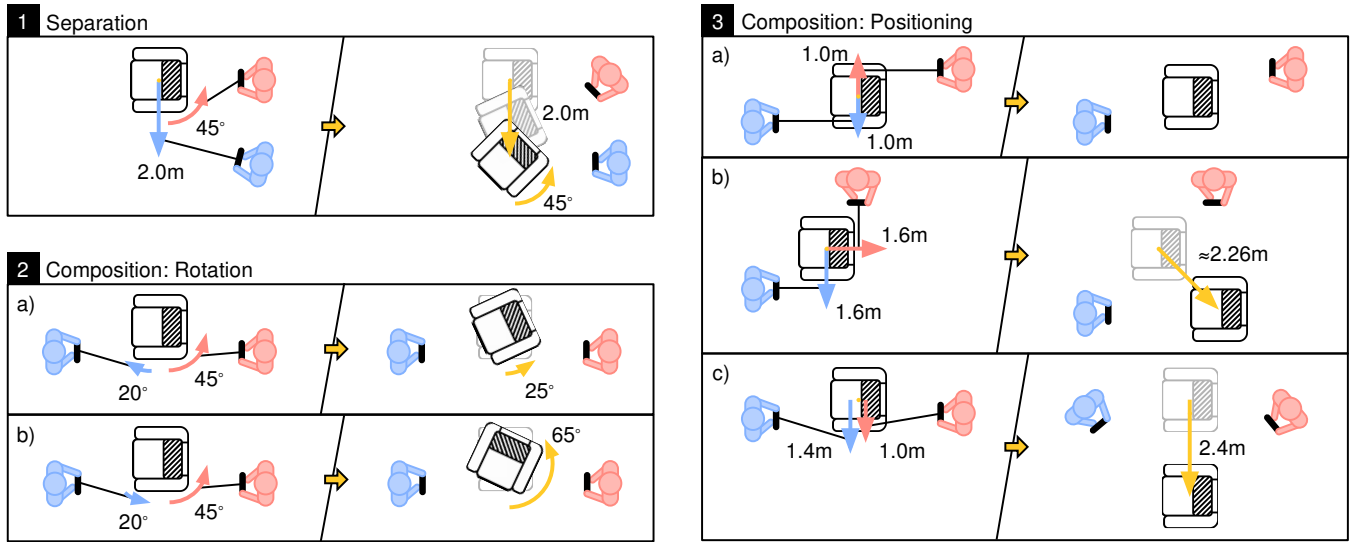


Figure 2: (1) Separation implements a DOF separation by separating the manipulation tasks rotation and positioning between users. Composition allows simultaneous (2) rotations or (3) translations and combines users' inputs with the computation of the sum as the merge policy.

included additional metrics (e. g., subjective workload or usability) but focused exclusively on the composition of actions. Consequently, there is, to the best of our knowledge, no experimental comparison of both approaches for handheld AR displays being the most prevalent AR devices [10, 28]. We aim to address this research gap by studying the manipulation techniques in handheld AR and investigating quantitative and qualitative measures (cf. Sect. 4, RQ1-4).

3 DESIGN OF OUR COLLABORATIVE 3D OBJECT MANIPULATION TECHNIQUES

We designed three collaborative 3D object manipulation techniques for dyads based on the proposed approaches by previous work: 1) *Separation* – performing only different manipulation tasks (i. e., translation or rotation) simultaneously, 2) *Composition* – performing only the same manipulation tasks simultaneously and combining their individual inputs using a merge policy, and additionally 3) *Hybrid* – performing any manipulation task simultaneously, enabling dynamic transitions between *Separation* and *Composition*. We chose these techniques as they constitute asymmetric manipulation (*Separation*, different actions simultaneously), symmetric manipulation (*Composition*, same actions simultaneously) [26], and a hybrid approach (*Hybrid*) that allows to dynamically transition between the other types of manipulation. Additionally, they reflect different symmetric collaboration styles ranging from loosely-coupled (*Separation*) to closely-coupled (*Composition*) and how users fluidly switch among them (*Hybrid*) (cf. coupling styles [16, 19, 30]).

Similar to popular apps (e. g., IKEA Place [15], AR in Google Search [17]) where AR objects are anchored to surfaces and sticking to the realistic scenario of furnishing a place, we limited the techniques to a) moving objects along the floor (2 DOF) and b) rotating them around the yaw axis (1 DOF). We relied on well-known touch gestures (instead of e. g., device movements) to reduce the perceived workload [11], which ensures a higher validity of our experiment. To start the manipulation of an object, users tap to select it. This highlights the selected object with a semi-transparent circular handle below it. Objects can be rotated with two fingers, and a single finger allows to move an object using a drag-and-drop operation. We used different gestures for the manipulation actions, as this can increase user performance [22]. For both actions, we implemented awareness cues: The circular handle shows the currently possible individual manipulation as rotation or positioning icon or remains empty if both

are possible. During manipulation it is also colored in the user's color (i. e., red or blue). Color-coded arrows around the circular handle indicate the type and direction of the team partner's activities (cf. Fig. 1). This double-coding (color and the icons' layout) helped to avoid disadvantages for color-blind people. Tapping outside of the circular handle releases an object and ends the manipulation.

In the following, we describe our implementations of the three collaborative techniques in greater detail and explain how related work specifically informed their individual design. For the implementation we used Unity 2018.2 [31] and Apple's ARKit 2.0 [3].

3.1 Technique 1: Separation

Similar to Pinho et al. [23, 24] and Aguerreche et al. [1, 2], this technique implements a DOF separation by separating the manipulation tasks positioning and rotation between the users (cf. Fig. 1.1). While one user controls the position of the object, the other one is responsible for its rotation. Consequently, users can only perform different manipulation tasks simultaneously (cf. Fig. 2.1). Thereby, the first-come, first-served (FCFS) principle takes effect: The user who initially starts manipulating the object determines what the other user is allowed to do. Users can freely switch the manipulation tasks when both stop manipulating the object. This enforced type of labor division fosters working in parallel (loosely-coupled collaboration), potentially leading to shorter task completion times and benefits regarding accuracy due to a higher specialization [13].

3.2 Technique 2: Composition

This technique implements a composition of users' actions. It allows users to perform only the same manipulation tasks simultaneously (cf. Fig. 1.2). This enforced type of labor division fosters utilizing multiple individual perspectives for each manipulation task (closely-coupled collaboration), potentially leading to higher accuracy. The user who starts manipulating the object determines what the other user is allowed to do, e. g., if the first user starts rotating the object, the second user can only rotate it. Switching is possible when both stop manipulating the object. Like Grandi et al. [11–13], we decided to use the computation of the sum as the merge policy, since they have shown that this is a suitable approach for handheld AR. Consequently, actions towards the same direction reinforce each other, while actions towards opposite directions equalize each other (cf. Fig. 2.2 and Fig. 2.3). Due to this, also the one-to-one mapping

between finger and handle can be suspended. As a solution, the concept introduces a proxy. A smaller circular handle appears that always stays underneath the user's finger (cf. Fig. 1.2). Inspired by Riege et al. [25], it is linked to the bigger circular handle by a line that dynamically changes its length. When the user lifts the finger to end moving the object, the small circular handle disappears again.

3.3 Technique 3: Hybrid

This technique is inspired by the technique of Grandi et al. [11–13]. It combines the other two techniques and allows users to perform any manipulation actions simultaneously. Consequently, users can dynamically decide whether they separate rotation and positioning between them (*Separation*) or perform these manipulation tasks together (*Composition*). In contrast to the other two techniques, *Hybrid* does not foster any type of labor division. While this allows reaping the potential benefits of *Separation* (working in parallel) and *Composition* (multiple individual perspectives for each action), it provides less structure and requires more coordination, which, in turn, could suspend the other techniques' advantages.

4 STUDY DESIGN

We conducted the study as a controlled lab experiment. Action integration represented the completely counterbalanced within-subjects factor with the three techniques *Separation*, *Composition*, and *Hybrid* being the conditions. To assess which of the three techniques should be used for a collaborative 3D object manipulation on hand-held AR displays, we designed the study to answer the following four research questions:

RQ1 Performance: How do the techniques differ in accuracy and task completion time?

RQ2 Workload: How do the techniques differ in subjectively perceived workload?

RQ3 Participation: How do the techniques relate with users' participation?

RQ4 User Experience: Which technique do users prefer?

4.1 Participants

We recruited 48 participants (28 female, 20 male) between 18 and 40 years ($M = 22.96$, $SD = 4.15$) from our local campus in pairs. We included only pairs that confirmed that they already knew each other. Forty-five participants were students and three employees, having a mixed background of different areas (e. g., biology, psychology, economics, or law). The participants formed 24 dyads, of which 9 consisted of two females, 5 of two males, and 10 of both one female and one male participant. Only 4 of them had made experiences with AR in advance. Twenty-six participants previously furnished an apartment together with a familiar person confirming the ecological validity of our task. Further, we asked participants to rate their tablet experience ($M = 3.98$, $SD = 0.79$) and gaming experience ($M = 2.40$, $SD = 1.18$) from 1 (very inexperienced) to 5 (very experienced).

4.2 Task

In line with previous research [11, 12, 33], we studied our 3D object manipulation techniques using a 3D docking task that required participants to position virtual objects at predefined target positions and orientations. The design of our docking task was inspired by IKEA's furnishing app "Place": For all conditions, dyads had to collaboratively furnish a room with seven virtual pieces of furniture anchored to the floor and having different shapes and sizes (smallest: $33 \times 33 \times 45\text{cm}$, largest: $110 \times 183 \times 102\text{cm}$). We did not restrict participants to position themselves in predefined arrangements (e. g., face-to-face or side-by-side) and allowed them to move freely through the room. We decided for this furnishing task with virtual furniture matching real-world counterparts to provide our

participants with a realistic and authentic scenario and, with this, increase the ecological validity of our user study. Target positions and orientations were the same for all conditions to reduce search activities. They were visualized with virtual masking tape on the floor (cf. video in supplemental material). Additionally, labels indicated the required target object (e. g., the table *Olmsted*). We also added virtual masking tape arrows inside the floor markings if items had a defined front side to avoid misunderstandings. Items without a defined front side (e. g., the table *Olmsted* or the stool *Odwar*) could be positioned in different orientations, i. e., turned by 90 or 180 degrees. All items were listed in a catalog within the prototype, and participants were allowed to pick their preferred items until the catalog was empty. There was no predefined order of items, and participants decided on their own who picks each individual item. Items were placed at a distance of 2 meters in front of the participant who picked them. Participants were tasked to furnish the room using one of the manipulation techniques as fast and accurately as possible. An initial training task with virtual cardboard moving boxes helped participants to understand and get used to the different manipulation techniques. The virtual cardboard moving boxes replicated the properties of the furniture (i. e., defined front side and size).

4.3 Dependent Variables and Operationalization

Accuracy and task completion time were measured to operationalize the *Performance (RQ1)* of the three approaches. As task completion, we measured the time between the first object's instantiation and the end of the last object's final manipulation based on our data logs, i. e., excluding participants' negotiation times at beginning and end of the task. For accuracy, the error between every piece of furniture and its associated floor marking was measured separately for position and rotation. For the position errors, the euclidean distances between the actual positions and the target positions of the pieces of furniture were calculated. The smallest angle difference between the rotation of floor marking and piece of furniture determined the rotation error. The raw, unweighted NASA Task Load Index (NASA TLX) [14] was used to measure subjective *Workload (RQ2)*. To estimate participants' *Participation (RQ3)*, we calculated the participation scores proposed by Grandi et al. [12], which are based on the equivalence of participants' manipulation times, i. e., the times each individual user manipulated the objects. *User Experience (RQ4)* was assessed via the User Experience Questionnaire (UEQ) [18]. Additionally, a semi-structured interview was conducted with each dyad to gain further insights. Participants were asked which condition they preferred, which they would use for furnishing their place, and to mention the advantages and disadvantages of each technique. We transcribed all interviews and clustered participants' answers thematically regarding statements describing advantages and disadvantages of each technique including general positive and negative sensations. All interviews were transcribed and sorted by the same person to ensure consistency.

4.4 Procedure

Participants were welcomed and asked to fill out a demographic questionnaire first. Then, the examiner explained the task and prototype using a slide show presentation with step-by-step instructions including images and videos showing the virtual cardboard moving boxes. After that, the first manipulation technique – according to the counterbalanced order – was presented as part of the slide show. This way, both participants were equally able to follow the instructions, which minimized problems of understanding. Then, participants started the training phase with the first manipulation technique and the virtual cardboard moving boxes until they felt comfortable using the prototype and the technique. After that, participants started the actual task with the remark to be as fast and precise as possible (cf. [11]). For each manipulation technique, participants repeated the task (i. e., the placement of all seven pieces of

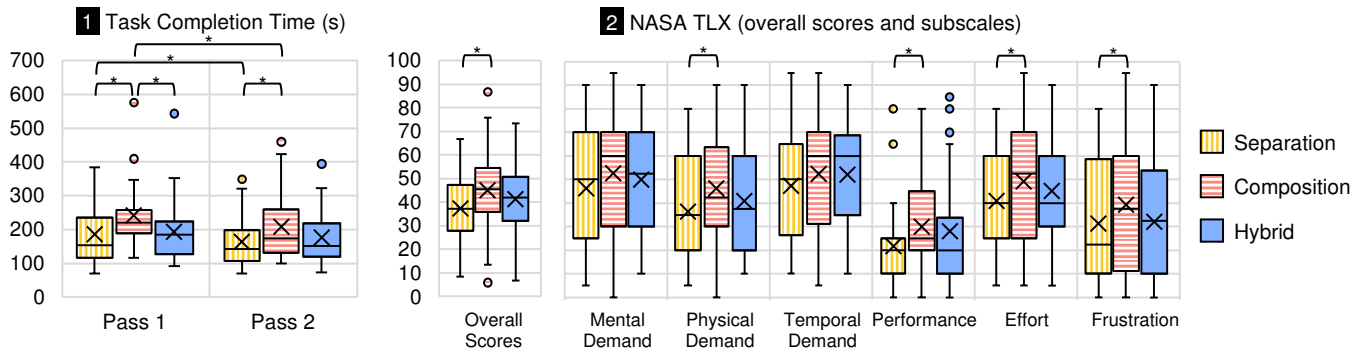


Figure 3: Boxplots for (1) task completion time in seconds (2) the overall scores of the NASA TLX, and its scores broken down into subscales.

furniture at the same predefined targets) twice. We decided for these two consecutive passes per technique to identify intended learning effects within each condition. Then, participants filled out the questionnaires (TLX + UEQ). The entire procedure was repeated for the other two conditions, starting with the slide show presentation of the specific manipulation technique. Each session was closed with the concluding semi-structured interview. Sessions took approximately one hour in total and participants were compensated for their time. We followed all ethical and sanitary guidelines provided by our institution at the time of the study.

4.5 Apparatus

Each participant was provided with an Apple iPad Pro (9.7 inches, 437 grams) with a resolution of 2048×1536 pixels (264 pixels per inch). They used their color-coded tablet to fill out the questionnaires and to work on the study task. The tablet's rear-facing camera with a resolution of 12 megapixels was used to create the AR experience. We used a Windows 10 high-end desktop computer as a server to a) collect data recorded by the tablets and b) to control the procedure and sequence of the individual study sessions (e. g., to start the training phase on each tablet). We used a dedicated wireless network with no connection to the internet to connect all devices. This way, we minimized any disturbances in both, the recording of the data and the experience of participants. We allotted a walkable area of approximately 4.0×7.0 meters in our lab for the user study. We hung up landscape prints on the walls of our lab as pretests showed that ARKit required additional visual features to ensure stable tracking.

5 FINDINGS

Following the order of the research questions, this section presents the findings of the user study. To indicate the conditions, subscript S is used for *Separation*, subscript C for *Composition*, and subscript H for *Hybrid*. The Shapiro-Wilk test was used to test if the results were normally distributed. As the data violated the assumption of normal distribution, a non-parametric approach was used to analyze it. Further, medians (Mdn) are reported. To detect differences between the three conditions for each pass, Friedman's ANOVA was used. If this test showed differences, post hoc analysis was performed with Dunn-Bonferroni-Tests. The Wilcoxon signed-rank test was used for comparing the passes within each condition (i. e., pass 1 vs. pass 2). Generally, an alpha level of 0.05 was assumed for all statistical tests. For the pairwise comparisons, the Bonferroni correction required adjusting the alpha level to $(.05/3) \approx .016$.

5.1 Performance (RQ1)

We measured task completion times and accuracy to study different aspects of performance. As participants had two passes per condition, we are able to compare their performance between conditions as well as the intended learning effects within each condition.

5.1.1 Task Completion Times

The task completion times for all three conditions and both passes are shown in Fig. 3.1. For the first pass, there was a statistically significant difference between the three conditions ($\chi^2(2) = 12.583$, $p = .002$). Task completion times for *Separation* ($Mdn_S = 153.22s$, $z = -3.320$, $p = .001$) and *Hybrid* ($Mdn_H = 184.96s$, $z = 2.742$, $p = .006$) were statistically significantly lower than for *Composition* ($Mdn_C = 220.38s$). For the second pass, task completion times were also statistically significantly different ($\chi^2(2) = 8.583$, $p = .014$). Here, participants were statistically significantly faster with *Separation* ($Mdn_S = 141.45s$, $z = -2.887$, $p = .004$) compared to *Composition* ($Mdn_C = 174.02s$). Comparing the two passes of each condition with each other, it showed that for *Separation* ($z = -2.829$, $p = .005$) and *Composition* ($z = -2.800$, $p = .005$) participants were statistically significantly faster in the second pass.

5.1.2 Accuracy

The analysis of the position errors in millimeters showed that the differences for both, the first pass ($\chi^2(2) = 3.083$, $p = .214$) and the second pass ($\chi^2(2) = 5.083$, $p = .079$) were not statistically significant. The comparison of the passes within each condition also showed no statistically significant differences.

Also for the rotation errors in degrees, no statistically significant differences could be revealed – neither for the first pass ($\chi^2(2) = 1.750$, $p = .417$) nor for the second pass ($\chi^2(2) = 0.250$, $p = .882$). Also, the comparison of the passes within each condition showed no statistically significant differences.

5.2 Workload (RQ2)

Fig. 3.2 shows the boxplots for the mean overall scores of the NASA TLX questionnaire for each condition. Analysis revealed statistically significant differences between the three conditions ($\chi^2(2) = 9.937$, $p = .007$). Overall scores for *Separation* were statistically significantly lower than those for *Composition* ($Mdn_S = 37.08$, $Mdn_C = 45.42$, $z = -3.113$, $p = .002$). Fig. 3.2 shows the breakdown of the overall scores into the different subscales of the NASA TLX questionnaire. Analysis revealed statistically significant effects for the dimensions *Physical Demand* ($\chi^2(2) = 9.337$, $p = .009$), *Performance* ($\chi^2(2) = 10.683$, $p = .005$), *Effort* ($\chi^2(2) = 6.356$, $p = .042$), and *Frustration* ($\chi^2(2) = 8.503$, $p = .014$). The dimensions *Mental Demand* ($\chi^2(2) = 2.958$, $p = .228$) and *Temporal Demand* ($\chi^2(2) = 5.035$, $p = .081$) showed no statistically significant differences. Post hoc analysis revealed, that *Separation* was statistically significantly ranked better than *Composition* for the dimensions *Physical Demand* ($Mdn_S = 35$, $Mdn_C = 42.5$, $z = -2.909$, $p = .004$), *Performance* ($Mdn_S = 20$, $Mdn_C = 25$, $z = -2.858$, $p = .004$), *Effort* ($Mdn_S = 40$, $Mdn_C = 52.5$, $z = -2.398$, $p = .016$), and *Frustration*

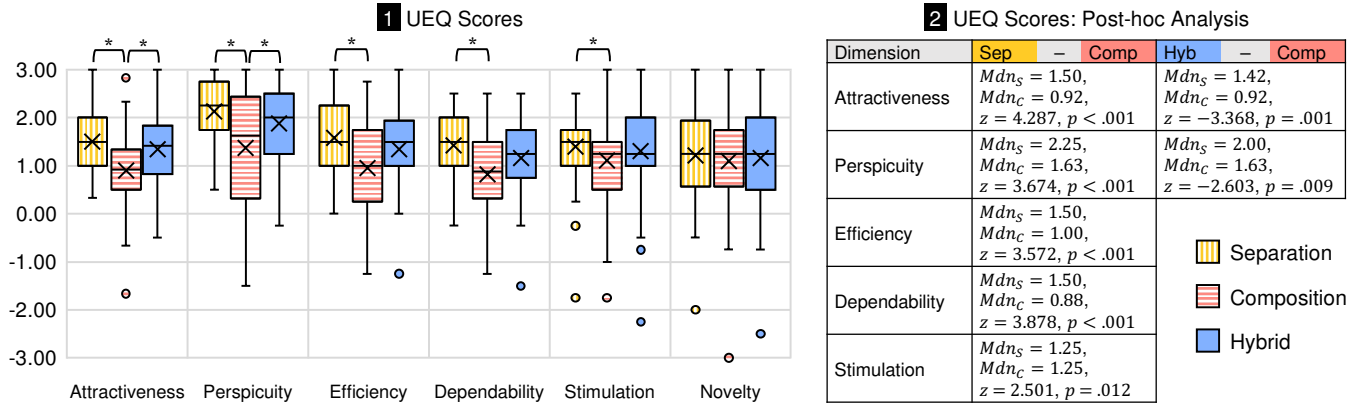


Figure 4: (1) Boxplots of the UEQ scores. (2) Detailed results of the post hoc analysis for the UEQ scores. Only significant results are listed.

($Mdn_S = 22.5$, $Mdn_C = 37.5$, $z = -2.501$, $p = .012$). The other pairwise comparisons showed no statistically significant results.

5.3 Participation (RQ3)

We calculated the participation scores [12], which are based on the equivalence of participants' manipulation times. A score of 0 means that a single participant performed the manipulations alone, while a score of 1 means that both participants contributed equally to the manipulation. Median scores were similarly high (pass 1: $Mdn_S = 0.78$, $Mdn_C = 0.83$, $Mdn_H = 0.88$; pass 2: $Mdn_S = 0.81$, $Mdn_C = 0.77$, $Mdn_H = 0.81$) and statistical analysis revealed no statistically significant differences between the scores.

5.4 User Experience (RQ4)

We assessed user experience via the UEQ [18] and a final semi-structured interview. At the beginning of the interview, participants were asked to separately rank the three techniques according to their preference and the amount of coordination they required. Afterward, they were asked to reason both of their rankings and name advantages and disadvantages of the techniques.

5.4.1 User Experience Questionnaire (UEQ)

The responses to the UEQ are visualized in Fig. 4.1. Analysis revealed statistically significant differences for the dimensions *Attractiveness* ($\chi^2(2) = 22.613$, $p < .001$), *Perspicuity* ($\chi^2(2) = 16.926$, $p < .001$), *Efficiency* ($\chi^2(2) = 14.619$, $p = .001$), *Dependability* ($\chi^2(2) = 17.190$, $p < .001$), and *Stimulation* ($\chi^2(2) = 7.560$, $p = .023$). No difference could be revealed for the dimension *Novelty* ($\chi^2(2) = 3.574$, $p = .167$). For the dimensions *Attractiveness* and *Perspicuity*, post hoc analysis showed that participants scored the conditions *Separation* and *Hybrid* significantly better than *Composition*. In terms of *Efficiency*, *Dependability*, and *Stimulation*, only *Separation* was scored significantly better than *Composition*. Fig. 4.2 lists the results of the post hoc tests that showed significant differences.

5.4.2 Subjective Rankings

Participants were asked to individually rank the conditions separately according to their general preference ("Which of the variants did you like most?") and the required amount of coordination ("Which of the variants required the least coordination?"). For analyzing the results of the participants, a fractional ranking was used. Therefore, scores were assigned to the conditions based on their ranked position (i. e., 1-3 from best to worst). If participants ranked two conditions equal, the mean of two scores was assigned. For example, if a participant ranked *Separation* and *Hybrid* equal but better than *Composition*, a score of 1.5 was assigned to *Separation* and *Hybrid*, and

Composition received a score of 3. Statistical analysis showed significant differences in the ranking of participant's general preferences ($\chi^2(2) = 57.770$, $p < .001$). Post hoc analysis revealed that participants ranked the conditions *Separation* ($Mdn_S = 1$, $z = -7.195$, $p < .001$) and *Hybrid* ($Mdn_H = 2$, $z = 5.664$, $p < .001$) statistically significantly better than *Composition* ($Mdn_C = 3$). Also, for the ranking according to the required amount of coordination, the analysis revealed statistically significant differences ($\chi^2(2) = 21.618$, $p < .001$). *Separation* ($Mdn_S = 1.25$, $z = -4.389$, $p < .001$) and *Hybrid* ($Mdn_H = 2$, $z = 2.960$, $p = .003$) were ranked to require less coordination compared to *Composition* ($Mdn_C = 3$).

5.4.3 Concluding Interview

We also asked participants to list the advantages and disadvantages of each condition and summarize their answers in the following:

Concerning positive aspects of *Separation*, the majority of participants valued the clear division of work ($n = 37$) which required only an initial communication, had the positive effect that "you don't get in each other's way" (#15b) and allowed to focus and specialize on one manipulation task ($n = 16$). Further, some participants ($n = 4$) highlighted that this division of work allowed them to perform different manipulation tasks simultaneously, which decreased the time required to complete the study task. Finally, participants ($n = 16$) expressed general positive sensations, e. g., perceiving the condition as "fast", "accurate" (#17a), "fluent" (#19b), and "structured" (#9a/b). When asked about negative aspects of *Separation*, participants remarked that the freedom of action was restricted ($n = 14$), which made it harder to correct each other ($n = 5$). Finally, some participants ($n = 4$) mentioned that the accuracy of *Separation* could suffer from the fact that they often "looked more at [their] own [task] and less at whether the other really aligned it completely correctly" (#4b). To compensate for this problem, "you had to walk more around the object" (#11b).

Asked about positive aspects of *Composition*, participants ($n = 6$) mentioned that because "both participants focused on the same task, [they] were more accurate" (#2b). Others valued increased communication ($n = 4$) and argued that the condition would be faster for large-scale actions ($n = 2$). However, there were also participants ($n = 3$) who said that *Composition* "has practically no advantages" (#24b). They found it hard to assess the addition of their actions, which lead to overshooting ($n = 21$), e. g., "you quickly turned or moved too far when both did something" (#11b). This overshooting resulted in the feeling of working against each other ($n = 16$), e. g., "like [in] a tug war" (#6b), and had the consequence that some participants ($n = 8$) decided to work alternating. Participants ($n = 8$) faulted that they could not do different tasks simultaneously, which decreased the time to complete the task. Consequently,

participants described this condition rather negatively ($n=25$), e. g., “inefficient” (#17a), “complicated” (#12a), and “confusing” (#10b).

Asked about *Hybrid*, participants ($n=25$) mainly valued the freedom they had with this technique (e. g., “you are free to do whatever you want” (#1b)) and noted ($n=13$) that they could quickly correct each other (#17b). They highlighted the higher individual share of each collaborator ($n=12$), e. g., “you could always do something” (#5a), and stated that they were faster with the approach ($n=4$) since they “could perform both different and same actions simultaneously” (#2a). Participants ($n=2$) also said that they were more accurate “because everyone felt responsible for both actions” (#4b) and expressed general positive sensations ($n=8$) describing the approach, e. g., as “fluent” (#19a), “flexible” (#13a), and “being fun” (#9b). Regarding negative aspects, participants complained about the possibility of hindering each other ($n=14$) and stated ($n=6$) that it required a high amount of coordination. As general negative sensations, they ($n=7$) mentioned that it was, e. g., “complicated”, “unclear” (#14b), and “uncoordinated” (#24a).

At the end of the concluding interview, we asked participants if the collaborative manipulation of the same object would be relevant for them if they used the app to plan the furnishing of a room together with a flatmate. For 27 of them it would be not relevant in such a scenario. They reasoned that they would have “no time pressure” (#11b), would “do it with one tablet” (#10a), or would prefer “to do it on their own” (#9a) and “compare their settings afterward” (#15b). However, for 21, it would be relevant. They reasoned their decision with the importance of agreeing on a furnishing and the different points of view they have during the collaborative manipulation. We further asked these 21 participants which of the three conditions they would like to use in the described scenario. Fifteen participants indicated *Hybrid*, only four *Separation*, and two were undecided between *Hybrid* and *Separation*. No participant mentioned *Composition* as the preferred approach.

6 DISCUSSION

The structure of the discussion follows our research questions.

6.1 Performance (RQ1)

Participants were able to achieve the same level of accuracy with all approaches – indicating a fair comparison of techniques. This means that none of the investigated types of labor division resulted in a higher accuracy – neither only allowing different tasks simultaneously (*Separation*) nor only the same tasks simultaneously (*Composition*) nor the dynamic transition between them (*Hybrid*). Interestingly, we expected either a higher accuracy for *Separation* (due to the specialization [13]) or *Composition* (due to multiple individual perspectives for each task). However, the lack of differences might be related to the differences in task completion times: Here, participants needed significantly longer to finish passes for *Composition*. Although participants were tasked to be as accurate and fast as possible, some of them ($n=14$) primarily focused on accuracy: “Accuracy does not depend on the variant. Either you are an awfully accurate person or a sloppy person” (#16a).

During the first pass, participants were significantly faster with *Separation* and *Hybrid* than with *Composition*. This is in line with our expectations: During *Composition*, participants could only perform one manipulation at a time, i. e., either rotating or positioning. During the other two conditions, however, they could perform both manipulations simultaneously – theoretically, allowing participants to perform twice as many manipulations in the same time. This is to some extent compensated by faster movements and rotations during *Composition* as a result of adding up forces. However, this advantage was less relevant for vernier adjustments: Here, participants ($n=21$) mentioned during the concluding interview that they could not estimate the addition of forces, leading to target overshoots.

Comparing the first with the second pass, participants were significantly faster with *Separation* and *Composition* during the second pass – suggesting an intended learning effect. For *Hybrid*, there was no significant difference when comparing both passes. This missing learning effect resulted in no statistically significant differences between *Composition* and *Hybrid* for the second pass. As *Hybrid* allowed participants to dynamically transition between the other techniques, they potentially were encouraged to try out different strategies to solve the task, which might have hindered the learning effect: They said that “you can work out your own strategy with this condition” (#4a) and “we wanted to try out several [strategies]” (#17b). On the contrary, participants noted that *Separation* and *Composition* already took off some coordination, e. g., “[during *Separation*] the strategy was clear, [since] we had no other option” (#14a), or “the roles are clear [during *Composition*] – if one moves the other moves too.” (#7b).

Qualitative statements from the concluding interview further imply that with dedicated training required for finding the perfect strategy, participants could even perform best with *Hybrid*. One participant said: “If you know each other well and practice often, this condition [Hybrid] has the highest potential for speed and accuracy. But this requires more training” (#17a). However, to be able to make reliable statements, future work needs to have a more in-depth look into the strategies participants developed during *Hybrid*. Notably, the results imply that the learning effect was higher for *Composition* than for *Separation*. During the first pass, medians of the task completion times of *Separation* and *Composition* differed by 67.2 seconds. For the second pass, this difference reduces to 32.6 seconds. Future work is needed to investigate if more training decreases this difference further.

6.2 Workload (RQ2)

Participants performed the same physical activities during all conditions. They needed to physically locomote in the room, hold the tablet, and perform the gestures for rotation and positioning with their fingers. However, participants rated the NASA TLX dimension *Physical Demand* higher for *Composition* than for *Separation*. One explanation could be that participants included the virtual pushing and rotating of objects into their ratings – guided by the description of the dimension *Physical Demand*, which explicitly mentions these terms: “How much physical activity was required (e. g., pushing, pulling, turning, controlling, activating, etc.). [...]” [14]. Alternatively, this might be related to participants complaining about overshooting during working with *Composition*. Here, due to overshooting, participants had to perform the gesture more often or for a longer duration, which eventually resulted in longer task completion times. Correcting this overshooting required additional mental and physical effort that was reflected in the dimension *Effort*. Overshooting also resulted in the impression of working against each other ($n=16$), the expression of negative sensations for *Composition* ($n=25$), and thus, in a higher level of subjective *Frustration*. The scores for the dimension *Performance* reflect the differences in task completion times: Participants rated their performance better for *Separation* than for *Composition*. Thus, participants interpreted their performance rather in terms of efficiency (task completion times) than in terms of effectiveness (accuracy), which reflects the results regarding the techniques’ performance.

6.3 Participation (RQ3)

For both passes, statistical analysis of the participation scores showed no significant differences between the conditions. The overall high participation scores indicate that – regardless of the condition – participants were equally involved in the task, which confirms the collaborative character of the techniques. Additionally, no significant differences between the two passes could be revealed, suggesting that participants did not change their participation in

the second pass. We initially expected to find differences in dyads' participation as *Hybrid* requires additional coordination and communication to decide for a manipulation procedure – a step that was implied in *Separation* and *Composition* by enforcing different types of labor division. Consequently, we assume that participants' participation is less related to the different techniques but influenced by the dyads' attitudes, work behavior, and how they positioned themselves to the objects.

6.4 User Experience (RQ4)

Participants preferred *Separation* and *Hybrid* over *Composition*. Between the rankings of *Separation* and *Hybrid* no statistically significant differences could be revealed. These results indicate that participants' preferences were polarized between *Separation* and *Hybrid*. The explanation for this polarization can be found in qualitative statements: On the one hand, they valued the clear division of work *Separation* promotes ($n=37$). On the other hand, they complained about restricted freedom of action during this condition ($n=14$). *Hybrid* exactly provided this freedom of action, which was highly valued by participants ($n=25$). Consequently, participants were polarized between the “organized and structured” (#9a/b) labor division of *Separation* and the “unbent and flexible” (#13a) experience of *Hybrid* not fostering any type of labor division. However, this polarization might only apply to the specific study task. When asked which approach they would like to use for furnishing their own home, 15 out of the 21 participants who found the collaborative manipulation relevant, reported that they would prefer *Hybrid*. Their qualitative statements suggest that in this scenario, the freedom of *Hybrid* outweighs the structuredness of *Separation*.

The participants' polarized preferences between *Separation* and *Hybrid* also become apparent in the results of the UEQ: Participants rated *Separation* and *Hybrid* significantly more attractive than *Composition*. We found additional statistical differences in other dimensions of the UEQ: Participants struggled with assessing the addition of their inputs in *Composition* and described it as “complicated” (#12a), a finding that is also reflected in the dimensions *Perspicuity*, *Dependability*, and *Efficiency*. The latter confirms the finding regarding the performance dimension of the subjective workload: Here, participants rated their performance mainly with respect to the efficiency. While there were no differences regarding *Novelty*, participants perceived the *Stimulation* of each technique differently: Here, *Separation* was perceived statistically different in comparison to *Composition*. We attribute this finding to the increased level of frustration of *Composition*, which in turn can be discouraging.

7 IMPLICATIONS

Regarding task completion time and subjective workload, the results of the experimental comparison suggest using *Separation* for a collaborative 3D object manipulation on handheld AR displays: Participants achieved their level of accuracy not only faster but also with less subjective workload. However, during the concluding interview, it turned out that participants' preferences were polarized between *Hybrid* and *Separation*. They highly valued the clear labor division of *Separation*; and they liked the freedom of action during *Hybrid*. These polarized preferences of the participants imply choosing the approach depending on scenario and task. If the required time is a decisive factor (as in work-related tasks like, e. g., 3D construction), interaction designers possibly should rely on the clear division of work *Separation* provides. In contrast, if the user experience is of higher priority and the time required a secondary factor (as in casual and more exploratory activities like, e. g., furnishing a place), the freedom of action offered by *Hybrid* might be more suitable. Future research should, therefore, have a more in-depth look at different tasks to investigate this trade-off between a clear division of work and freedom of action.

8 LIMITATIONS AND FUTURE WORK

We investigated specific variants of the two general approaches for action integration revealed by related work. However, also other implementations of the two general approaches are possible. Using the mean value instead of the sum of inputs seems to be a promising alternative merge policy since for projection-based VEs, Aguerreche et al. [1, 2] showed that this can lead to a higher accuracy than DOF separation. Also, one could think about a merge policy that dynamically switches between mean and sum depending on the participants' view on the object to be manipulated.

Our experimental comparison considered 3 DOFs for manipulation based on the real-world task of furnishing a place. Participants could rotate the virtual object around the yaw axis and move it along the floor. Rotations around the other two axes were not possible. Besides, participants could not lift, lower, or scale the objects. Future work could investigate different scenarios that require the manipulation of all 9 DOFs. Another limitation concerns the number of collaborators. The study reported by this work investigated dyads. But one could also consider groups with three or even more members. Especially, if more than 3 DOFs are investigated, *Separation* would allow splitting the DOFs of one canonical task between more than two users. For example, in terms of the canonical task rotation, one user could then rotate the object around the yaw axis, another around the roll axis, and a third around the pitch axis.

Future work should further investigate the differences in participants' participation between the dyads, which could be caused by differences in their work behavior, strategies, or how they positioned themselves during manipulation. For the latter, one could classify f-formations as proposed by Marquardt et al. [21] and investigate their influence on task performance. Finally, qualitative statements suggest that knowing each other well and sufficient training could further improve participants' performance using *Hybrid*. Consequently, future work could investigate the effects and influences of long-term use of the techniques (e. g., to identify learning curves) and how group constellations (e. g., (not) knowing each other) impact the techniques' performance.

9 CONCLUSION

With this work, we investigated three collaborative 3D object manipulation techniques for handheld AR displays: 1) *Separation* – separating the DOFs of positioning and rotation between the collaborators, 2) *Composition* – allowing simultaneous manipulations only for either the DOFs of positioning or rotation, and 3) *Hybrid* – combining the other techniques by enabling dynamic transitions between them. In an experimental study with 24 dyads, we compared the three techniques regarding *Performance* (RQ1), *subjective Workload* (RQ2), *Participation* (RQ3), and *User Experience* (RQ4). Based on the results, we provide first insights into their strengths and weaknesses: Equally high participation scores confirmed the collaborative character of the techniques. Regarding task completion time and subjective workload, the study results suggest using *Separation* since participants performed better while perceiving less workload with this technique. However, participants' preferences were polarized between *Separation*'s clear division of work and *Hybrid*'s freedom of action.

Our findings imply choosing the approach depending on scenario and task: If time is the decisive factor, we suggest using *Separation*. If the focus lies on user experience, we suggest using *Hybrid*.

We hope that our offered implications help to guide the design of future applications and inspire future research to further investigate collaborative 3D object manipulation using handheld devices in AR.

ACKNOWLEDGMENTS

This research was funded by the *Deutsche Forschungsgemeinschaft* (DFG, German Research Foundation) – Project-ID 251654672 – TRR 161 (Project C01).

REFERENCES

- [1] L. Aguerreche, T. Duval, and A. Lécuyer. Reconfigurable tangible devices for 3d virtual object manipulation by single or multiple users. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*, VRST '10, p. 227–230. Association for Computing Machinery, New York, NY, USA, 2010. doi: 10.1145/1889863.1889913
- [2] L. Aguerreche, T. Duval, and A. Lécuyer. Evaluation of a Reconfigurable Tangible Device for Collaborative Manipulation of Objects in Virtual Reality. In I. Grimstead and H. Carr, eds., *Theory and Practice of Computer Graphics*. The Eurographics Association, 2011. doi: 10.2312/LocalChapterEvents/TPCG/TPCG11/081-088
- [3] Apple Inc. Augmented reality – apple developer. <https://developer.apple.com/augmented-reality/>. Last accessed: September 16, 2020.
- [4] R. T. Azuma. A survey of augmented reality. *Presence: Teleoperator. Virtual Environ.*, 6(4):355–385, Aug. 1997. doi: 10.1162/pres.1997.6.4.355
- [5] M. Billinghurst, A. Clark, and G. Lee. A survey of augmented reality. *Foundations and Trends® in Human–Computer Interaction*, 8(2-3):73–272, 2015. doi: 10.1561/11000000049
- [6] M. Billinghurst and H. Kato. Collaborative augmented reality. *Commun. ACM*, 45(7):64–70, July 2002. doi: 10.1145/514236.514265
- [7] W. Broll. Interacting in distributed collaborative virtual environments. In *Proceedings Virtual Reality Annual International Symposium '95*, pp. 148–155, March 1995. doi: 10.1109/VRAIS.1995.512490
- [8] M. L. Chenechal, J. Lacoche, J. Royan, T. Duval, V. Gouranton, and B. Arnaldi. 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)*, pp. 18–22, March 2016. doi: 10.1109/3DCVE.2016.7563562
- [9] T. Duval, A. Lecuyer, and S. Thomas. Skewer: a 3d interaction technique for 2-user collaborative manipulation of objects in virtual environments. In *3D User Interfaces (3DUI'06)*, pp. 69–72, March 2006. doi: 10.1109/VR.2006.119
- [10] E. S. Goh, M. S. Sunar, and A. W. Ismail. 3d object manipulation techniques in handheld mobile augmented reality interface: A review. *IEEE Access*, 7:40581–40601, 2019. doi: 10.1109/ACCESS.2019.2906394
- [11] J. G. Grandi, H. G. Debarba, I. Bemdt, L. Nedel, and A. Maciel. 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)*, pp. 49–56, March 2018. doi: 10.1109/VR.2018.8446295
- [12] J. G. Grandi, H. G. Debarba, and A. Maciel. Characterizing asymmetric collaborative interactions in virtual and augmented realities. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 127–135, March 2019. doi: 10.1109/VR.2019.8798080
- [13] J. G. Grandi, H. G. Debarba, L. Nedel, and A. Maciel. 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, p. 5881–5891. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3025453.3025935
- [14] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. 52:139 – 183, 1988. doi: 10.1016/S0166-4115(08)62386-9
- [15] Inter Ikea Systems B.V. Ikea launches ikea place, a new app that allows people to virtually place furniture in their home. <https://newsroom.inter.ikea.com/News/ikea-launches-ikea-place-a-new-app-that-allows-people-to-virtually-place-furniture-in-their-home/s/f5f003d7-fcba-4155-ba17-5a89b4a2bd11>. Last accessed: September 8, 2020.
- [16] P. Isenberg, D. Fisher, S. A. Paul, M. R. Morris, K. Inkpen, and M. Czerwinski. Co-located collaborative visual analytics around a tabletop display. *IEEE Transactions on Visualization and Computer Graphics*, 18(5):689–702, May 2012. doi: 10.1109/TVCG.2011.287
- [17] A. Kannan. Make at-home learning more fun with 3d and ar in search. <https://blog.google/products/google-lens/make-home-learning-more-fun-3d-and-ar-search/>, 05 2020. Last accessed: September 8, 2020.
- [18] B. Laugwitz, T. Held, and M. Schrepp. Construction and evaluation of a user experience questionnaire. In A. Holzinger, ed., *HCI and Usability for Education and Work*, pp. 63–76. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. doi: 10.1007/978-3-540-89350-9_6
- [19] C. Liu, O. Chapuis, M. Beaudouin-Lafon, and E. Lecolinet. Shared interaction on a wall-sized display in a data manipulation task. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, p. 2075–2086. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2858036.2858039
- [20] D. Margery, B. Arnaldi, and N. Plouzeau. A general framework for cooperative manipulation in virtual environments. In M. Gervautz, D. Schmalstieg, and A. Hildebrand, eds., *Virtual Environments '99*, pp. 169–178. Springer Vienna, Vienna, 1999. doi: 10.1007/978-3-7091-6805-9_17
- [21] N. Marquardt, K. Hinckley, and S. Greenberg. Cross-device interaction via micro-mobility and f-formations. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, p. 13–22. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2380116.2380121
- [22] A. Martinet, G. Casiez, and L. Grisoni. The effect of dof separation in 3d manipulation tasks with multi-touch displays. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*, VRST '10, p. 111–118. Association for Computing Machinery, New York, NY, USA, 2010. doi: 10.1145/1889863.1889888
- [23] M. S. Pinho, D. A. Bowman, and C. M. 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, p. 171–178. Association for Computing Machinery, New York, NY, USA, 2002. doi: 10.1145/585740.585769
- [24] M. S. Pinho, D. A. Bowman, and C. M. D. S. Freitas. Cooperative object manipulation in collaborative virtual environments. *Journal of the Brazilian Computer Society*, 14(2):53–67, 06 2008. doi: 10.1007/BF03192559
- [25] K. Riege, T. Holtkamp, G. Wesche, and B. Frohlich. The bent pick ray: An extended pointing technique for multi-user interaction. In *3D User Interfaces (3DUI'06)*, pp. 62–65, 2006. doi: 10.1109/VR.2006.127
- [26] R. A. Ruddle, J. C. D. Savage, and D. M. Jones. Symmetric and asymmetric action integration during cooperative object manipulation in virtual environments. *ACM Trans. Comput.-Hum. Interact.*, 9(4):285–308, Dec. 2002. doi: 10.1145/586081.586084
- [27] R. A. Ruddle, J. C. D. Savage, and D. M. Jones. Levels of control during a collaborative carrying task. *Presence: Teleoperators and Virtual Environments*, 12(2):140–155, 2003. doi: 10.1162/105474603321640914
- [28] D. Schmalstieg and T. Höllerer. *Augmented Reality: Principles and Practice*. Addison-Wesley Professional, 6 2016.
- [29] M. Sereno, X. Wang, L. Besancon, M. J. McGuffin, and T. Isenberg. Collaborative work in augmented reality: A survey. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2020. doi: 10.1109/TVCG.2020.3032761
- [30] A. Tang, M. Tory, B. Po, P. Neumann, and S. Carpendale. Collaborative coupling over tabletop displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, p. 1181–1190. Association for Computing Machinery, New York, NY, USA, 2006. doi: 10.1145/1124772.1124950
- [31] Unity Technologies. Unity. <https://unity.com/>. Last accessed: September 16, 2020.
- [32] M. Veit, A. Capobianco, and D. Bechmann. Influence of degrees of freedom's manipulation on performances during orientation tasks in virtual reality environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, VRST '09, p. 51–58. Association for Computing Machinery, New York, NY, USA, 2009. doi: 10.1145/1643928.1643942
- [33] V. Vuibert, W. Stuerzlinger, and J. 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, p. 44–52. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2788940.2788950

- [34] X. Wang, L. Besançon, M. Ammi, and T. Isenberg. Augmenting tactile 3d data navigation with pressure sensing. *Computer Graphics Forum*, 38(3):635–647, 2019. doi: 10.1111/cgf.13716