Understanding Users' Capability to Transfer Information between Mixed and Virtual Reality: Position Estimation across Modalities and Perspectives

Joan Sol Roo*
Inria, France

Jean Basset*
ENSC, France

Pierre-Antoine Cinquin*
Bordeaux University, France

Martin Hachet*
Inria, France

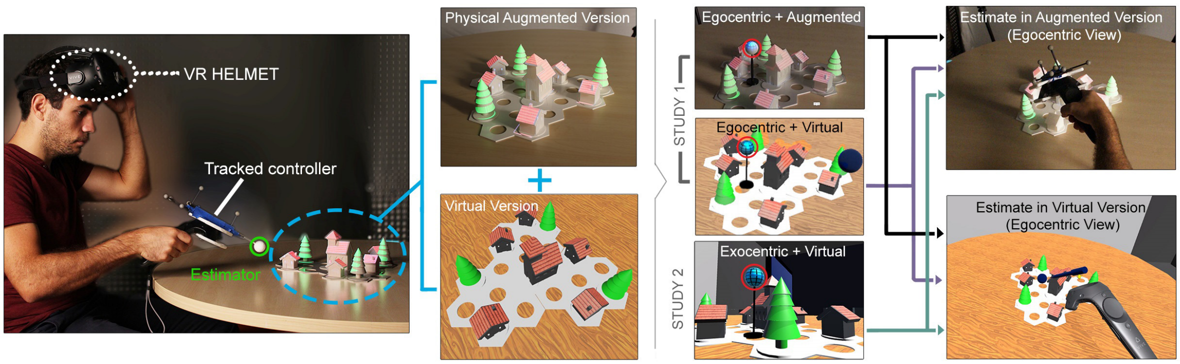

Figure 1. Setup and Tasks involved in the studies presented in this paper. Participants interacted with a spatially augmented mock-up and its virtual counterpart (left). They were iteratively presented with a spherical target from either egocentric (Study 1) and exocentric perspectives (Study 2), and then asked to estimate the position of the target (right) using an estimator attached to a controller. The objective of this protocol is to quantitatively measure the participants' capability to transfer information between physical and virtual spaces, and between egocentric and exocentric perspectives.

ABSTRACT

Mixed Reality systems combine physical and digital worlds, with great potential for the future of HCI. It is possible to design systems that support flexible degrees of virtuality by combining complementary technologies. In order for such systems to succeed, users must be able to create unified mental models out of heterogeneous representations. In this paper, we present two studies focusing on the users' accuracy on heterogeneous systems using Spatial Augmented Reality (SAR) and immersive Virtual Reality (VR) displays, and combining viewpoints (egocentric and exocentric). The results show robust estimation capabilities across conditions and viewpoints.

Author Keywords

Mixed Reality; Spatial Augmented Reality; Virtual Reality; Evaluation; Quantitative Methods

*email: FirstName[-MiddleName].LastName@inria.fr

CHI'18, April 21–26, 2018, Montreal, QC, Canada

ACM ISBN 978-1-4503-5620-6/18/04…$15.00

DOI: 10.1145/3173574.3173937

ACM Classification Keywords

H.5.1 Multimedia Information Systems: Artificial, augmented, and virtual realities; H.5.2 User Interfaces: Evaluation/methodology

INTRODUCTION

The constant quest for improving the human interaction with physical and digital worlds has led to the emerging of numerous display technologies, input devices, and interaction techniques. One of the earlier methods is to immerse users in a virtual space by overriding the perception of physical space, called Virtual Reality (VR). As Milgram and Kishino [31] stated back in 1994, physical and digital are not opposites, but instead sit at a continuum of Mixed Realities (MR). A clear example of this is Augmented Reality (AR), in which the physical world is complemented with digital information. Being a continuum, it is possible to consider the navigation between degrees of virtuality, or to use complementary mixed reality modalities. This provides not only the flexibility to choose the best modality for a given task, but also allows to overcome technical trade-offs.

Concretely, this work focuses on the complementarity between spatial augmentation and immersive displays: Spatial Augmented Reality (SAR) [34] places the augmentation directly onto the environment (in contrast with see-through based AR). SAR is of great interest since it does not require to instrument the users, while also providing a unified experience for multiple users close to the physical scene. However, SAR is not without limitations, since it is constrained by the available physical geometry, and neither SAR nor optical See-Through AR (ST-AR) can completely override what the user sees. On the other hand, immersive Head Mounted Displays (HMD-VR) provide complete freedom from physical constraints. The main problem with HMDs is that they isolate the users from the physical world around them. It has been recently shown that these technologies complement well with each other from both conceptual and technical points of view [38]; however, they provide very heterogeneous representations. As technology becomes mature enough to allow these hybrid MR systems to become more common, it is necessary to better understand the ability of humans to interact with them. Indeed, once the technical issues are addressed, the success of a hybrid MR system will rely on users being able to create unified mental models based on heterogeneous representations.

The main question addressed by this work is: *are users able to correctly complement digital and physical information in hybrid mixed reality systems?* To answer this question, we designed an experimental protocol that requires participants to transfer information between spatially augmented and virtual views of a single scene. This allows us to build knowledge to help the development of future hybrid systems that jointly stand on both the forces of VR and AR technologies.

The contributions of the presented work are: 1) design of the experimental protocol, used on 2) a first user study focusing on egocentric (i.e., the natural point of view) position estimation task in SAR, VR and their combination, and 3) a second user study focusing on mixed egocentric and exocentric (i.e., from outside of the user's body) position estimation in MR and VR.

RELATED WORK

This work focuses on the empirical evaluation of participants' performance when using MR systems. As such, it is based on the 1) literature on MR systems, and differs from 2) previous evaluations of MR systems by 3) deriving inspiration from experimental approaches in Cognitive Science and VR.

Interacting with Mixed Reality Systems

The best way to understand the potential of hybrid MR systems is under Benford's Artificiality and Transportation [4]. Artificiality is closely related with the degree of virtuality [31], while transportation refers to the user(s) degree of presence in either the local or a remote location (by "leaving your body behind"). Some MR systems cover not a single point of the artificiality-transportation space, but areas instead. MagicBook [8] is a physical book that supports different degrees of artificiality (from purely physical to purely virtual), allowing the transition from an egocentric viewpoint (no transportation) to a location inside the book (partial transportation, since the body is still visible). Similarly, Kiyokawa extensively studied collaborative MR systems [22], where the focus could alternate between the local scene (no transportation) to a remote location (full transportation), and this remote scene could be obtained through scanning (no artificiality), virtually created (full artificiality), or a combination of both. Rekimoto studied dynamically augmenting the local space [36, 35]. More recently, Rekimoto explored the potential of taking different perspectives of the physical world [21, 20, 25] and asymmetric perspectives of a virtual scene [17]. Moving towards the physical world, Ullmer and Ishii combined complementary displays and tangible interfaces [42]. In this line of work, we previously explored complementary display modalities in combination with tangible interfaces, to support increasing artificiality and transportation when needed [38, 37].

Previous Evaluations of Mixed Reality Systems

When looking at the MR literature, the users' performance is rarely a studied factor [12], even if it is critical for the system's success. Most evaluations focus on either i) the quality of the subjective experiences generated [7, 6], ii) their impact on learning [13], or iii) the communication between users [32, 23, 24]. To our knowledge, only a handful of evaluations explore the performance on perception tasks, perhaps given the difficulties of building MR experimental protocols [3]. The existing evaluations focus on depth estimation, asking the participants: i) if a virtual object is in front or behind a wall [14], ii) asking to estimate the distance to a target either verbally or with an object (error in meters) [19, 30, 10], or iii) the depth of a virtual floating object (error in cm) [5]. In all cases where accuracy was measured, the targets were visible when the estimations were made, thus evaluating perception and not explicitly evaluating the mental representation (i.e., *could the users operate once the information is not available?*).

Understanding Spaces

When looking at the existing MR systems it is possible to observe that some of them involve different views of a single space. This has been previously explored in VR: Stoakley's World-in-Miniature [41] allows an immersed to observe an additional external view of their surroundings, while CALVIN [28] allows the asymmetric collaboration between users inside and outside a single scene. It would be of great interest to know more about the users' performance when combining the information of these heterogeneous views, and perhaps know more of the underlying mental processes. For this, we can look at the approaches used in Cognitive Science and VR.

Wang and Simons [44] studied the capability of participants to perceive changes on a physical scene, between changes in object orientation and viewpoint (the latter being more accurate). M.A. Amorim [1] evaluated the capability of participants to orient themselves in relation with an object, and vice versa, while in VR. Steinicke et al. [40] showed that using a virtual replica of the physical space as transitional environment can reduce depth compression in VR, evaluated through blind walking. In order to evaluate the construction of a correct spatial model, these experiments share the a three step structure: the participants are provided with information (1), and shortly after removing the information (2), participants are asked to operate based on what they were shown (3). This same structure is the one selected for our studies, yet the task is completed by operating in the peripersonal space and based on both the physical and digital worlds.

STUDY DESIGN

We designed a Mixed Reality experimental protocol that requires participants to estimate the location of a previously presented target. This experiment was used on two different studies: the first study involves an egocentric task (i.e., from the participant's perspective) in order to characterize users' accuracy, while the second study combines both egocentric and exocentric views. This section describes the details of the environment where the task takes place, and the details regarding implementation and calibration.

Scene and Task

We tested the interaction with an augmented physical mock-up and its virtual counterpart (Figure 1). For this, we used the 3D printed mock-up of a small town with 3 types of landmarks (5 houses, 3 trees and a church) over lasercutted hexagonal bases (18cm diameter, 6 cm diameter per cell); the landmarks were distributed on a non-symmetrical layout, and the mock-up was placed at the center of a circular table (135cm diameter). The SAR version provided basic texture mapping. The virtual version reproduced not only the augmented mock-up and table, but also the room where the experience took place. To increase reproducibility, the used 3D models are available for 3D printing, and as assets for Unity3D [1].

Both studies involved a position estimation task. Participants were sitting facing the mock-up, and were iteratively shown a spherical target (3cm in diameter) in a location either inside or around the mock-up at one of three possible heights (3.0, 6.0 or 9.0cm). After the target was hidden, the participants were asked to place an estimation using a sphere (also 3cm of diameter) attached to a wand controller (Figure 3-left), and to confirm the estimation by pressing controller's trigger (using the soft "hair-trigger" trigger mode to mitigate unintended movement when clicking).

Homogenization of Conditions

The researcher conducting the study was sitting next to the participant (90 degrees to the participant's right, also facing the mock-up, Figure 2). The conditions involving physical targets required the researcher to manually place and remove the target, the position indicated using projection. During this time (when they had to interact with the mock-up), the participants were asked to close their eyes, while a board was placed in front of their eyes to prevent peeking. To keep the conditions as similar as possible, the VR counterpart steps displayed a black screen, using a fade to black to make it less abrupt.

To standardize the time measurements and workload, each of the steps that required attention were preceded by a 3 second countdown, and this was indicated visually on the helmet, but also verbally for all conditions. The time per trial was recorded to evaluate its impact on the performance.

Ethical Authorization

The study design as well as data managing was approved beforehand by the ethics committee of Inria France (COERLE).

[1] http://team.inria.fr/potioc/accuracy-sar-vr/

Apparatus

The software was implemented using the approach described in [38]. The hardware layout was comprised of off-the-shelf components: 4 Optitrack flex 3 cameras, an LG projector PF80G and a HTC Vive set (helmet, controllers and lighthouses). The tracking of the controllers was performed using Optitrack (figure 2) instead of the HTC tracking. This decision was based on the fact that HTC Vive uses sensor fusion for tracking, and as a result the world position of the components is not known accurately enough for our case.

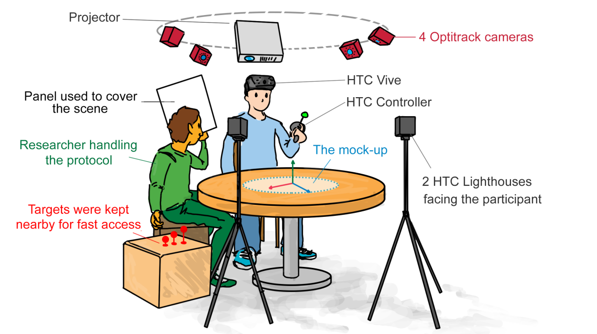

Figure 2. Experiment setup: HTC Vive lighthouses were placed in front of the participant, while the Optitrack cameras and the LG projector where placed around and over the participant.

Calibration

To guarantee the quality of the results, the system was calibrated at least once per day. First, the Optitrack volume was calibrated over and around the table, and the origin was set to the center of the table. Then, the projector was calibrated using OpenCV camera calibration functionality[2], by matching 3D points in the Optitrack frame of reference with 2D points in the projector's image plane. The OpenCV perspective matrix was transformed into a projection matrix [26] by taking into account resolution and near and far planes. Finally, the alignment between Optitrack coordinate system and HTC Vive was performed computing the 3D to 3D transform using a controller with infrared markers as reference (Figure 3-right).

The Optitrack calibration reported an error at sub-millimeter scale (under 0.2mm), while the projection calibration showed in average a reprojection error of 3.8px (reflected on 1.8mm of error at the center of the mock-up in average). The HTC

[2] CalibrateCamera http://docs.opencv.org/2.4/doc/tutorials/calib3d/camera_calibration/camera_calibration.html

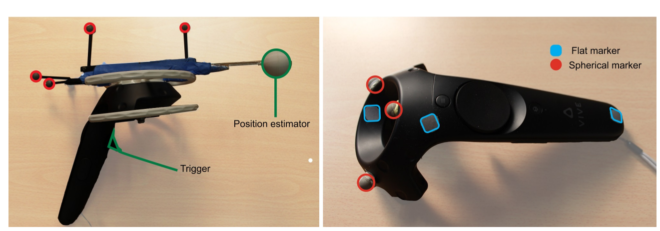

Figure 3. Controller with estimation attached to it (left). Controller used to align Optitrack and HTC Vive spaces, markers placed to easily identify origin and orientation of the controller (right).

Vive to Optitrack registration error was in the order of 1cm, which only affected the head position, since the controller was tracked using Optitrack to guaranty higher precision.

STUDY 1: EGOCENTRIC ESTIMATION

The first study focused on comparing performances and subjective similarities between the physically augmented and virtual scenes, in order to provide a baseline for more complex tasks (Study 2). The objective of Study 1 was to test if:

H1: People using a mixed reality environment are able to perceive information in one space, virtual or physical, and then use this information to accurately operate in the other space.

Operational hypothesis: The accuracy on a position estimation task (i.e., placing an element at a previously indicated location) in hybrid conditions (combining SAR and VR) should be *not bigger than* in pure conditions (either SAR or VR).

Participants

In order to recruit participants for the study we made a public announcement and posted it on the mailing lists of both the institute and the university. A total of 18 participants (11 male, 7 female) volunteered to take part in the study. Their ages ranged from 21 to 51 (Mean = 26±9), all of them except one were right handed, and they all had normal or corrected-to-normal vision. Most of the participants were students or members of the research institute. Most of the participants play video games: 6 of them play frequently, 9 of them play occasionally, while 3 of them never play. Most of them (11) played sports frequently while growing up, 6 occasionally, and one never. 10 still do sports frequently now, 5 occasionally, and 3 never. 13 of them already had experience with HMD based virtual reality, and 10 of them with AR applications. This demographic information was obtained in order to detect influencing factors.

Procedure

Participants were welcomed and asked to sign a consent form that explained the objectives of the study along with clarifications regarding the anonymity of the data, known risks of VR and that they were volunteers (i.e., free to stop at any time). Once the consent form was signed, participants filled a demographic questionnaire : age, gender, sports and video game habits, and experience with AR and VR. Then, they filled two mental task tests to know more about their profile: 3D mental rotation test [43], and 2D spatial orientation [16]. These questionnaires took around 20 minutes to complete.

The evaluation involves a position estimation task, comprised of 4 runs with 12 trials each, using a within-participant counterbalanced conditions. Four conditions were considered (Table 4.2): seeing the target either in SAR or VR, and pointing either in SAR or VR; the hybrid conditions (SAR_VR and VR_SAR) required the participants to either place or remove the helmet, and for this reason they were alternated to reduce discomfort (i.e., one trial of SAR_VR, then one trial of VR_SAR, and so on). To keep the trial length homogeneous, then the combined SAR_VR and VR_SAR was divided in 2 runs, for a total of four runs, in 3 groups of conditions to counterbalance (Table 4.2). Of the 12 trials, 2 were discarded leaving a total of 10 usable trials per run: the first trial was explicitly mentioned to the participant as a practice trial, while the seventh trial had to be discarded to compensate for the fact that SAR_VR and VR_SAR where divided on two runs. Each run was followed by the NASA TLX [15] (standard questionnaire used to estimate the effort required to complete a given task), and a custom questionnaire for the conditions involving the HMD (Figure 7).

CONDITION		See target	Make estimation
Type	Code	(Egocentric view)	(Egocentric view)
SAR	SAR_SAR	SAR	SAR
MR	SAR_VR	SAR	VR
	VR_SAR	VR	SAR
VR	VR_VR	VR	VR

Table 1. Conditions for Study 1. Four conditions were grouped in three runs, based on their type. Only egocentric perspective was involved.

At each trial, the participant was presented with a target for 5 seconds, and then had to estimate where they considered the target was previously located, by placing an estimation artifact of the same size (Figure 3). Once the estimation was confirmed, the target was displayed again, along with an indicator of the estimation position at the table level. The reason why the estimation feedback was placed at the table level was to keep both SAR and VR conditions equivalent (since it is not possible to make 2 physical objects intersect to provide feedback for the SAR condition).

Finally, a short unstructured interview was conducted, to know more about how the participants felt during the experience. The evaluation took around 45 minutes, given a total time of around 65 minutes for the whole experience.

Measurements

For each trial we registered in a log file the positions (target, estimation, head position) and times (total, estimation_start and estimation_end). From this, all the metrics were obtained.

Accuracy - Absolute error: The absolute error is the distance between the target location and the estimation location in world coordinates (expressed in centimetres (*cm*)).

Accuracy - Failure count: we consider failures those trials with estimation error over 6*cm*. This decision was based on the properties of scene (6*cm* is the size of one cell of the mock-up, and twice the diameter of the target), over that limit we consider the participants forgot or confused landmarks. Failures were counted, but not considered when computing mean error distance.

Accuracy - Depth error: The depth error is computed as the signed distance between the target location and the estimation, taken from the participant's point of view.

Cognitive Load: After each run, the participants were asked to complete the NASA TLX.

Subjective Experience: evaluated with a custom 7-point Likert scale questionnaire, and completed after the conditions involving the HMD (Question listed in Figure 7).

Results - Study 1

In this section, we present the analysis of the data collected, including accuracy, workload and subjective experience.

Data Analysis

To ensure independence, trials were reduced to one sample per participant per factor combination (using mean). In the cases the data presented a non-normal distribution, we used the Aligned Rank Transform [45] to correct our data (indicated with a ART subscript when presenting the results). Then, we used ANOVA on the data (corrected or otherwise), and used Bonferroni as post-hoc analysis. Bivariate correlations were computed using Pearson when non-categorical variables where involved. All the data analysis was performed using SPSS 23. The obtained results are displayed at two levels: i) p-values for statistically significant differences, paired with mean values and confidence intervals (grouping the trials per participant), and ii) distribution box-plots (not grouped). This was done as an effort to complement the p-values [11], and allow the reader to have their own interpretation of the data.

Accuracy Results

Failure count: The number of estimations with an error over 6cm (i.e., over twice the diameter of the target) was overall low (5.8%): 3.9% in SAR_SAR, 5.6% in VR_SAR, 6.7% in SAR_VR, and 7.2% in VR_VR. As mentioned in measurements, these estimations were not considered when computing the mean error.

Accuracy The accuracy of the subjects did not seem to be affected by the condition, according to a one-way ANOVA ($F(3,68)_{ART} = 0.491, p = 0.690$). To better understand the results, the targets and their estimations were divided in clusters based on their distance from the mock-up (inside or outside) and angle from the center (left, right, top), giving a total of 6 clusters (Figure 4). Results are presented in Figure 5. We conducted a two-way ANOVA explaining the absolute error by the condition and the region. The ANOVA did not show a significant effect of condition on the absolute error ($F(3,380)_{ART} = 1.724, p = 0.162$). It showed a significant difference in absolute error between the targets' regions ($F(5,380)_{ART} = 9.793, p < 0.001$). Finally, the ANOVA rejected the interaction effect between condition and region for the absolute error ($F(15,380)_{ART} = 1.129, p = 0.328$). Figure 5 presents the error distribution per region and condition. For the region, the absolute error on targets outside the mock-up is significantly higher than for the targets inside, while

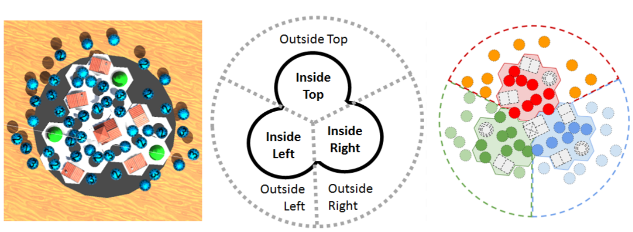

Figure 4. 48 target locations (left), 6 regions (center), and the 48 targets clustered by region (right). The participant (not represented) would be placed at the bottom of the picture.

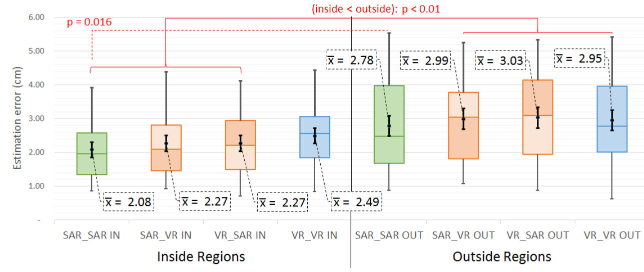

Figure 5. Absolute error: The effect of region by condition. The charts present both mean with confidence 95% intervals at one sample per-participant (black), and distribution at one sample per-trial (colored).

estimations for the same region show similar results for different conditions (except for VR_VR). The failures match this result: out of the estimations with an error over 6*cm*, most are positioned in the outside clusters (37/42), half of which are in the outer top cluster (18/37).

Depth error: There is a strong effect of depth in error on the condition, particularly for VR_VR. A one-way ANOVA showed a significant difference in depth error between conditions ($F(3,60)_{ART} = 5.285, p = 0.003$). It is possible to look at the depth error as a signed variable (Figure 6). In our case a positive depth error means that the estimation is between the real target position and the user. Participants tended to estimate the target closer to them than it really is, even more in the VR_VR condition.

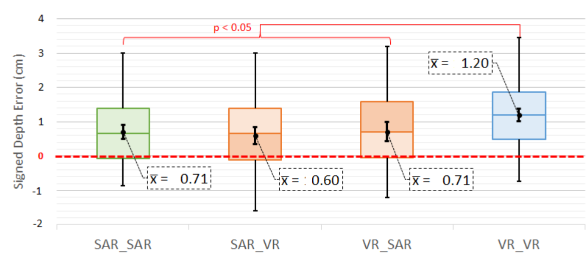

Figure 6. Depth error: A positive error in depth implies that the participants estimated closer to themselves (presented this way for clarity).

Subjective Experience Results

The results obtained by the subjective experience questionnaire (Figure 7, 7-Likert scale) and the comments obtained during the interview were similar.

First and foremost, no statistical differences were found between mean scores for MR and VR conditions ($F(1,34) = 0.002, p = 0.969$, inverting the values of negative questions Q6 and Q7). A descriptive analysis shows that both modalities received rather positive scores (mean scores, MR: 1.65±1.50; VR: 1.64±1.55, values ranging between -3 and 3), that the participants perceived both spaces as the same once they understood the mapping (Q4), and felt the mock-up was in front of them even while immersed (Q3). Note that even for the VR_VR condition, participants were sitting in front of the mock-up, thus seeing it before and after the run. All but one of them felt overall precise, and accredited the estimation error to

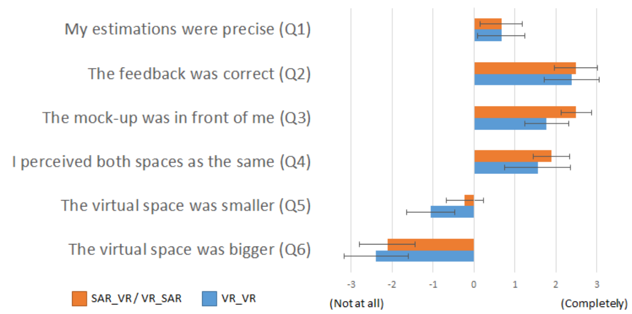

Figure 7. Subjective experience based on a 7-Likert scale questionnaire, values between -3 and 3. Error bars indicate confidence intervals.

themselves (Q1), rather than the system (Q2); this difference was confirmed verbally during the interview (the participant which questioned the system accuracy tended to face directly down during the experience, thus occluding the HMD tracking). When considering the fidelity of the registration between SAR and VR modalities, the answers present a high variance (Q5), and participants mentioned that noticed this effect only after reading the question.

Both during the protocol and the following interview, participants mentioned difficulties with the helmet or the VR rendering. Several participants were initially disoriented by the lack of feedback on where their hands were. Most participants mentioned that the illumination of the virtual scene was different than the physical one, and in particular the shadows were too strong; participants reported that the height of the estimator was harder to see while in VR. Some participants (in particularly, females) mentioned that the helmet felt heavy, and we had difficulties with some of participant's haircuts.

Workload Results

Regarding the NASA TLX workload, only 17/18 subjects were evaluated, since one of them provided an incomplete questionnaire and had to be discarded. The tasks measured by the NASA TLX are the three groups of the study (SAR, VR, and MR). We conducted a two-way ANOVA on the workload estimated by the subjects explained by the task and the order in which the tasks were passed. The ANOVA showed no effect of the task on the workload ($F(3,54) = 0.453, p = 0.716$), no effect of the order ($F(3,54) = 0.253, p = 0.859$), and no interaction between task and order ($F(7,54) = 0.400, p = 0.898$). The obtained results are displayed later on when discussing the workload for both studies (Figure 11).

Influencing Factors

We found some influence of the demographic data collected and the mental rotation tasks on the accuracy of the subjects. The questionnaires were used as a way to know more about the population rather than to correct the results and further experimentation should be conducted to obtain reliable conclusions; still, tendencies can be observed.

Mental tests: The mean error shows an inverse correlation with *spatial orientation* ($r = -0.387, p = 0.01$), and no significant correlation with *mental rotation* capabilities ($r = -0.045, p = 0.721$); as reported [16], both tests show correlation with each other ($r = 0.601, p = 0.01$). Even when both mental rotation tests show correlation with playing sports (particularly *playing while growing up*, Mental Rotation: $r = 0.425, p = 0.01$, Spatial Orientation: $r = 0.530, p = 0.01$), no correlation was found between *sports* and *estimation error*.

Gender: Females obtained higher values for absolute error (females: 2.76±0.43 vs males: 2.38±0.50), to a significant extent (2-way ANOVA gender and condition: $F(1,64) = 10.586, p = 0.002$), while there was no interaction between gender and condition. The same effect appears for the signed depth error (females: 0.98±0.57 vs males: 0.64±0.53), also significant ($F(1,56) = 7.482, p = 0.008$). This is consistent with other studies involving depth estimation in VR [2]. Additionally, differences between genders were found in relationship with the mental tasks: the correlation with mental rotation between spatial orientation and estimation error is significant only for males (males: $r = -0.705, p = 0.015$; females: $r = -0.087, p = 0.852$); this could indicate a sample effect explaining the difference in performance between genders. We consider the differences in accuracy between genders not large enough in practice, even when significant.

No trial duration influence, no order effect: Even when the time taken to complete one trial was different for each condition ($F(3,68) = 21.692, p < 0.001$), no correlation was found between this time and the estimation error ($r = -0.022, p = 0.852$), perhaps because the differences are small in practice (time between hiding the target and confirming estimation: SAR_SAR 11.8±1.8s; SAR_VR 15.4±5.7s; VR_SAR 12.7±2.6s; VR_VR 9.6±1.4s). Skill transfer between VR and physical environments has been found in the past for other tasks (e.g., [39, 27]), yet such an effect was not detected ($F(2,60) = 0.798, p = 0.455$), nor an interaction between order and condition ($F(6,60) = 0.689, p = 0.659$).

Study 1 - Conclusion

The results of the first study indicate that both spaces are perceived and interacted-with in a complementary manner. The error is mostly influenced by the region where the participant is targeting, more so than the condition involved. Regarding the hybrid conditions (SAR_VR and VR_SAR), the participants' accuracy was not significantly different with the SAR condition (no HMD), nor between each other.

It is also worth noting that even when in average participants estimated closer to themselves, the VR condition is the only showing a significant depth compression [18] when compared to the control condition (SAR); this is the case even when both mixed conditions required to perceive the space using VR (either to memorize the target location, or to estimate the location).

This first study supports the veracity of **H1** (*participants can transfer information from the physical space into the virtual space, and vice versa*), both objectively and subjectively. This result allow us to evaluate more complex tasks, such as the combination of egocentric and exocentric viewpoints.

STUDY 2: COMPLEMENTARY VIEWS

In the second study we focused on the change of scale and point of view, to test if:

H2: Users are able to complement their perception of a space when being provided information from different viewpoints.

H3: Complementary views can reduce the estimation error, particularly when dealing with far objects / lack of landmarks from the user's POV.

Once again, the operational hypothesis is that similar estimation error between conditions would imply similar perception, using the mixed condition as a referent to study the complementarity of modalities. To this end, accuracy should be: 1) not worse for MR than for VR, 2) not worse than for the first study (H2), and potentially 3) better than in study one (H3).

Participants

For Study 2, we wanted to ensure that it was possible to compare the obtained results with Study 1, so we performed recruitment from 2 sources: **new participants** obtained through a public announcement published on institute and university mailing lists, and **repeating participants** obtained by contacting the participants that performed the first study.

A total of 20 participants (15 male, 5 female) volunteered to take part in the study, 9 of which were part of Study 1. Their ages ranged from 20 to 58 (26.8±8.4), two of them were left handed, and they all had normal or corrected-to-normal vision. Most of the participants were students or members of the research institute. Around half of participants play video games: 6 of them play frequently, 5 of them play occasionally, while 9 of them almost never play. Part of them (7) played sports frequently while growing up, 11 occasionally, and two never. Four of the participants still do sports frequently, 12 occasionally, and 4 never. 12 of them already had experience with HMD-VR, and 12 of them with mobile AR.

Procedure

The second study followed the procedure used in Study 1, with some minor corrections. The protocol involved showing from an egocentric point of view a location on the table where the user will be "teleported", as an arrow oriented towards the center of the mock-up (Figure 8-left). The number of locations were 6, located every 60 degrees around the mock-up at a distance of 25 cm from its centre. Once the location was presented, the participant was then teleported to that location thanks to the VR helmet, where he or she could see the target location for 7 seconds. After the 7 seconds, the participant was presented with the scene once again from an egocentric point of view, and then asked to estimate the target's location. As with Study 1, feedback was displayed regarding the target location and the participant's estimation.

Since changing the point of view can only be done while wearing the helmet, we considered as independent variable the display modality used outside (i.e., to see the target location, and to perform the estimation). As a result, 2 conditions were considered (Table 5.2), each of them consisting of two series of 12 trials each. The extension to 2 series was in order to observe if there is a learning effect. Given that the task was considered harder than in the first study, the first 2 trials were explicitly discarded as rehearsal trials, given a total of 10 usable trials per run. These conditions were counterbalanced within participant, and within group (i.e., the conditions were alternated for *NEW* and *REPEAT* participants independently).

Most of the questionnaires and forms were shared with the first study. The half of the participants that took part of the first study did not fill the entry questionnaires, only the consent form. The only different questionnaire was the subjective experience questionnaire, which was extended to include questions regarding the change in point of view (Figure 12).

CONDITION		See location	See target	Make estimation
Type	Code	(Egocentric)	(Exocentric)	(Egocentric)
MR	SAR_VR_SAR	SAR	VR	SAR
VR	VR_VR_VR	VR	VR	VR

Table 2. Conditions for the second study.

Scene Corrections

Based on the comments from participants of the first study, the virtual scene was improved by decreasing the intensity of shadows. An iconic avatar was added at the location of the participant, to give a frame of reference when immersed.

In order to keep a constant distribution of landmarks, the mock-up was rotated 180 degrees, while the target locations was re-randomized (Figure 8-left). Each target was associated with one of the nearest POV locations, taking special care on preventing total occlusions. Since the change of perspective added a digital arrow to provide orientation, this could modify the available landmarks around the mock-up (Figure 8). This was taken into account when considering the accuracy similitude between studies, as explained in the *Results* section.

Measurements

As with Study 1, for each trial we registered on a log file the positions (target, estimation, head position, *and teleport position*) and times (total, estimation_start and estimation_end). All the metrics were computed as in Study 1 (namely, absolute and depth error, error count, time, and cognitive load). The *subjective experience* was evaluated with an extended version of the questionnaire used for Study 1 (Figure 12).

Results

This section presents the obtained results for the second study, and when relevant, it compares these results with the counterparts from Study 1.

Accuracy Results

Failures: Regarding the failures (estimations with an error above 6cm), the total count adds up to 46 of the estimations. Participant 19 accounted a total of 11 of these errors (27.5% of their own estimations), presenting an outlier behaviour, and for this reason was excluded from the evaluation. The remaining 35 failures (4.6% of the total measurements) were similarly distributed among SAR_VR_SAR (18/35) and VR_VR_VR conditions (17/35).

Condition, viewpoint and target location: The accuracy of the subjects was not significantly different between conditions ($F(1,36) = 0.397, p = 0.533$, Figure 9-top). The impact of which viewpoint was used to observe the target showed no statistical significance ($F(5,210) = 0.996, p = 0.421$). When looking at the spatial distribution of the error, results show a more uniform distribution than in the first study. No significant statistical differences were found for the accuracy between targets inside and outside the mock-up ($F(1,72) = 0.005, p = 0.943$), nor detectable interactions between condition and location ($F(1,72) = 0.088, p = 0.768$). Regarding the failures, they happened similarly often outside (17/35) and inside (18/35).

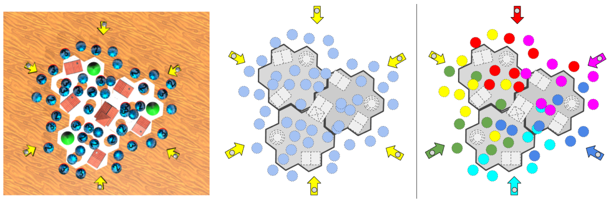

Figure 8. target locations (48) and 6 POVs, regions (6).

Participant group and accuracy between studies: There were no statistical differences between *NEW* and *REPEAT* participants ($F(1,34)_{ART} = 0.007, p = 0.933$), nor for between REPEAT group and the other participants of Study 1 ($F(1,16) = 0.007, p = 0.933$). When comparing accuracy, a significant difference in accuracy was found between studies ($F(1,35)_{ART} = 4.569, p = 0.04$): overall, the first study showed a higher estimation error than the second study. We remind the reader that for Study 1 a strong region effect was found; when comparing both studies, no statistical differences can be found for the internal regions ($F(5,104)_{ART} = 0.898, p = 0.485$), while the external regions present significant differences ($F(5,104)_{ART} = 4.296, p = 0.001$), up to various extents. The accuracy in Study 2 (for both internal and external targets) is similar to the accuracy for internal targets in Study 1.

Depth error: The error from the participant's perspective did not show significant differences between conditions

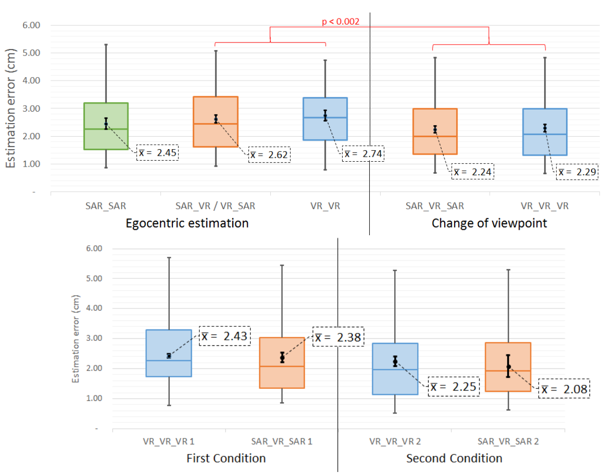

Figure 9. Estimation error per condition in comparison with Study 1 (top), and the tendency towards order effect (bottom).

($F(1,36) = 2.048, p = 0.161$), nor from the exocentric perspective (Figure 10). When comparing the results with Study 1, a significant difference can be found ($F(5,104) = 16.114, p < 0.005$), given that Study 2 presents less depth compression on all conditions. This can be observed when comparing Figure 10 and Figure 6: both studies present similar distributions from the egocentric perspective, yet there is a shift towards zero for the Study 2.

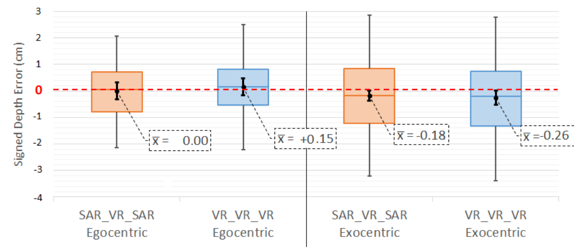

Figure 10. Depth error: The error per condition for Study 2, computed from both the egocentric (left) and exocentric (right) viewpoints. Note the distribution around zero, in contrast with Study 1 (see Figure 6).

Order effect: The order of the conditions has a significant impact on the accuracy ($F(1,34) = 7.901, p = 0.008$, Figure 9-bottom). The post-hoc analysis shows tendencies towards improvement for the second condition, both for SAR_VR_SAR ($p = 0.054$) and VR_VR_VR ($p = 0.074$).

Workload Results

When studying the impact of condition on the workload, no significant differences were found by *condition* ($F(2,72) = 1.221, p = 0.273$), nor by *order* ($F(3,72) = 0.078, p = 0.972$), nor an interaction *between order and condition* ($F(3,72) = 0.753, p = 0.524$). When comparing against the workload reported in the first study, a strong tendency was found ($F(12,127) = 2.380, p = 0.055$); the post-hoc analysis reflected this tendency only between SAR_SAR and VR_VR_VR ($p = 0.099$), as seen in Figure 11.

Subjective experience

As with the first study, no statistical differences were found on mean scores between conditions ($F(1,38) = 1.099, p = 0.301$). A descriptive analysis of the results (Figure 12) shows that both conditions present positive mean scores (MR: 1.04±0.53; VR:0.9±0.50, on a scale from -3 to 3, inverting scores for negative questions).

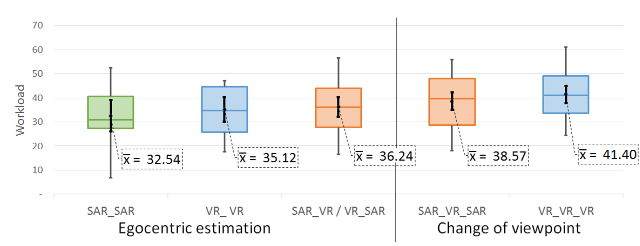

Figure 11. Workload results from the NASA-TLX questionnaire. Both the results for the first and second study are presented.

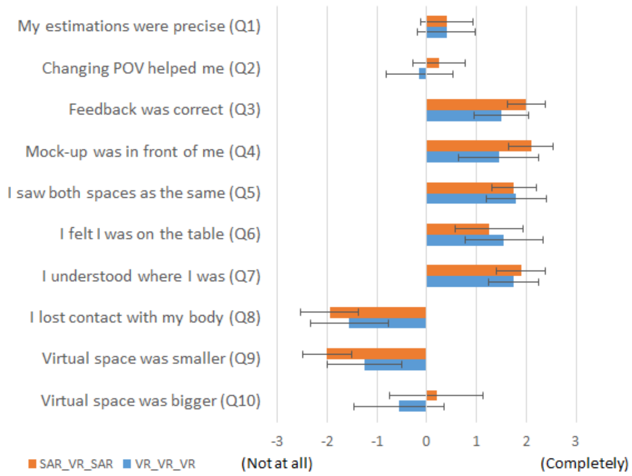

Figure 12. Results obtained for the subjective experience questionnaire for the second study, using a 7-Likert scale. The results are presented on a scale between -3 to 3 for clarity.

Questions regarding the feeling of presence (Q6) and complementarity between scenes (Q5) scored consistently high. Participants did not feel they lost contact with their body (Q8) or location (Q4), albeit the virtual condition scored slightly worse on these questions. The feedback was considered precise (Q3), yet slightly less precise in the virtual condition. Participants did not feel particularly precise (Q1), nor considered the change in POV was particularly useful (Q2). When studying the answers for Q2 (*changing scale is useful*) by participant, *NEW* participants tended to give a slightly positive answer (MR:0.7±1.3, VR:0.6±1.1), while *REPEAT* gave a slightly negative answer (MR:−1.2±1.1, VR:0.2±1.2).

Influencing Factors

Mental tasks, video games: Opposite to the findings from Study 1, an inverse correlation was found between *estimation error* and *mental rotation* only for the virtual condition (VR: $r = -0.564, p = 0.012$), and no significant correlation was found between *estimation error* and *spatial orientation* capabilities. In this study, no significant correlation was found between the tests ($p = 0.452$). Additionally, the estimation error presented an inverse correlation with playing video-games ($r = -0.539, p = 0.017$), which extended mostly to VR (VR: $r = -0.475, p = 0.04$), as only a tendency towards significance was found for MR (MR: $r = -0.394, p = 0.095$).

Order effect and trial duration: The tendency towards an order effect was previously discussed. Time varied between conditions ($F(1,36) = 4.5, p = 0.041$; time between hiding the target and confirming the estimation: MR 14.3±2.2s; VR 12.9±1.7s), and an inverse correlation between time and error was found ($r = -0.4, p = 0.013$). It seems that the longer the participants take to estimate, the more precise they are.

Study 2 - Conclusion

Participants seem to be able to correctly complement the viewpoints. Accuracy was not worse than for Study 1 (considering the SAR_SAR condition as a control task), even when the task involved can be considered harder. The similitude in accuracy between internal and external regions can be caused by the digital arrows acting as effective landmarks. This is supported by comparing the results with the ones obtained for Study 1: accuracy was found overall higher, and a tendency towards accuracy improvement for external targets (i.e., those placed outside the mock-up) was found. Even when participants were overall at least as precise, they did not report a subjective improvement regarding Study 1 (not only in average, but particularly the participants that were part of both studies). This seems to indicate that they expected higher accuracy, perhaps caused by seeing the scene from closer.

Overall, the obtained results support **H2** (*Users are able to complement their perception of a space when being provided information from different viewpoints*). It is not possible to make strong statements regarding **H3** (*Complementary views can reduce the estimation error, particularly when dealing with far objects / lack of landmarks from the user's POV*); instead, it can be said that the change in perspective *in combination with the addition of digital landmarks* seem to reduce the estimation error for far away objects / lack of *available* landmarks from the users POV.

DISCUSSION

Understanding the accuracy results

In order to put the obtained results in perspective, Figure 13 show three cases of distances between target and estimation. For both studies, when presented with enough landmarks, participants were able to obtain some degree of intersection between target and estimation (between case A and B of Figure 13). This was consistently achieved in most cases for most participants, disregarding modality (SAR or VR) or the viewpoint used to obtain the information (egocentric, exocentric). For both studies, 25% of the estimations were under 1.5cm (case A), around 75% of the estimations were under 3cm of error (case B), and the remaining estimations were in most cases under 6cm (between 93% to 97% estimations are better than case C, depending of the condition).

In addition to the rather robust estimation capabilities for all conditions, there could be in place a skill transfer, as shown by Study 2. This might imply that operating in different modalities could be equivalent in practice. It is worth noticing that

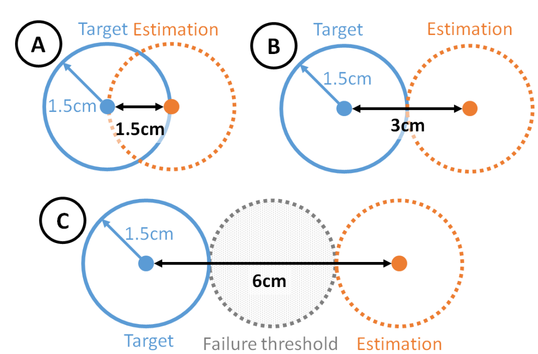

Figure 13. Three instances of estimation: touching the center of the target (A), touching the target (B), and the failure threshold (C).

participants considered their performance not good enough, in contrast with the perceived high precision of the system. This indicates that there is room for improvement at the users' end before the precision of the system becomes an issue.

Finally, pure VR conditions tended to have slightly higher error, and a significantly higher depth compression. It seems that alternating physical and digital can reduce this effect. This aligns with the findings of [40], and it is a strong indication towards the construction of a unified mental model.

Study Design - Considerations

This section briefly describes the rationale behind the most important decisions taken during the study design.

Homogeneity: At the protocol design stage, when facing a trade-off caused by differences between SAR and HMD-VR, we opted for the feature that could be implemented in both, prioritizing homogeneity between conditions over usability. For instance, when showing the estimation feedback only the base was indicated (since it is not possible to intersect two solid physical spheres), or the lack of explicit height indication when making the estimation while in VR (since it would provide an advantage over the SAR condition). Such self-imposed limitations do not need to be preserved when designing hybrid interfaces, allowing designers to improve each modality by addressing their limitations independently.

Non-ecological, yet prioritizing comfort: We are aware that putting and removing the helmet is uncomfortable and rather non-ecological, and we are not proposing this as the correct usage of this kind of systems, but instead as a way of evaluating the capabilities of the participants of combining both spaces (it would be better to use ST helmets [8], yet the render quality is not yet comparable to HMD-VR). For this reason, during the first study we alternated SAR_VR and VR_SAR conditions to minimize discomfort. As a consequence of this decision, we lack individual answers for the post-run questionnaires (subjective questions, and workload); no significant differences were found for the other available measurements. Even when our objective is to move towards a unified space (and bidirectional interaction would be ideal), these conditions could be studied independently in the future.

Risk of survival bias: For the second study we decided to also contact the participants from the first study, which had the risk of presenting a survival bias. We consider that the impact of this decision is negligible, since we are not asking questions regarding the enjoyment of the experience (e.g., we did not use the SUS), and we found no statistical differences for accuracy between *NEW* and *REPEAT* participants, nor between *REPEAT* and other participants of Study 1.

Study Design - Limitations

Some limitations should be explicitly mentioned, to place the obtained results in context and to improve future evaluations.

The depth estimation error is a known effect in VR called distance compression [18]; participants reported virtual elements as smaller while in VR. The rendering matched the field of view of the HMD, yet we used a default InterPupillary distance (IPD); in the future, taking into account per-participant IPD could reduce this effect. As mentioned on the *subjective experience results* of Study 1, ergonomics are also a factor. Even when the HMD used is an improvement over previous generations, it seems to still present difficulties for the general population, and females in particular.

There were several effects detected (gamers, sports, gender), but the protocol was not prepared to take them into account, and the population was not balanced in order to reach conclusions. Still, these effects resonate with the literature [2, 29]. A particular effect that could have been explicitly taken into account is the variability in working memory of the participants. The protocol was designed to keep trials short, with less than 15 seconds between the moment the target was hidden and the estimation was made. Additional questionnaires could be used to measure per-participant working memory (e.g., the complex figure test [9]), and longer memorization times per run could be also used, at the cost of longer sessions.

CONCLUSION

In this work we presented the results of two user studies that focus on the user's accuracy in mixed reality systems contrasted with pure VR versions. The first study considered variations – and similarities – between conditions from an egocentric viewpoint, while the second study evaluated the complementarity of egocentric and non-egocentric viewpoints for target estimation.

The obtained results indicate that, as with other spatial tasks, the accurate perception of the space is supported by the presence of landmarks. Participants showed a remarkable capability to transfer information between SAR and VR modalities, even between ego/exocentric POVs. Additionally, perceiving the scene from closer seems to increase the participants' expectations on their accuracy. It is worth mentioning that depth compression was significantly higher than in purely physical scenarios only for egocentric tasks that happened solely in VR. Hybrid MR conditions do not seem to suffer from this any more than in purely physical tasks in the case of egocentric estimation. Additionally, the lack of depth compression when changing scale (Study 2) challenges previous work [33], perhaps caused by the range (peripersonal space) and the availability of landmarks (both physical and virtual). These results indicate that the participants were able to construct a unified mental model from heterogeneous representations and views.

The presented research follows a rich history of perception and cognition studies, proposing the evaluation of mixed and hybrid systems from both a perceptual and cognitive standpoints. In the future, it would be of interest to study up to which extent the complementarity of modalities extends to real world scenarios, by exploring more complex and ecological tasks. Finally, it would be important to test the limits of systems with heterogeneous representations, looking for cases where users cannot transfer knowledge between display modalities.

ACKNOWLEDGMENTS

This work was supported by the ISAR project ANR-14-CE24-0013. We want to thank the study participants, and Thibault Laine for his help with fabrication.

REFERENCES

1. Michel-Ange Amorim. 2003. "What is my avatar seeing?": The coordination of" out-of-body" and" embodied" perspectives for scene recognition across views. *Visual Cognition* 10, 2 (2003), 157–199.
2. Robert S Astur, Maria L Ortiz, and Robert J Sutherland. 1998. A characterization of performance by men and women in a virtual Morris water task:: A large and reliable sex difference. *Behavioural brain research* 93, 1 (1998), 185–190.
3. Cédric Bach and Dominique L Scapin. 2004. Obstacles and perspectives for evaluating mixed reality systems usability. In *Acte du Workshop MIXER, IUI-CADUI*, Vol. 4.
4. Steve Benford, Chris Greenhalgh, Gail Reynard, Chris Brown, and Boriana Koleva. 1998. Understanding and constructing shared spaces with mixed-reality boundaries. *ACM Transactions on computer-human interaction (TOCHI)* 5, 3 (1998), 185–223.
5. Hrvoje Benko, Ricardo Jota, and Andrew Wilson. 2012. MirageTable: freehand interaction on a projected augmented reality tabletop. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 199–208.
6. Mark Billinghurst, Hirokazu Kato, Kiyoshi Kiyokawa, Daniel Belcher, and Ivan Poupyrev. 2002. Experiments with face-to-face collaborative AR interfaces. *Virtual Reality* 6, 3 (2002), 107–121.
7. Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. 2001a. The MagicBook: a transitional AR interface. *Computers & Graphics* 25, 5 (2001), 745–753.
8. Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. 2001b. The magicbook-moving seamlessly between reality and virtuality. *IEEE Computer Graphics and applications* 21, 3 (2001), 6–8.
9. P Caffarra, G Vezzadini, F Dieci, F Zonato, and A Venneri. 2002. Rey-Osterrieth complex figure: normative values in an Italian population sample. *Neurological Sciences* 22, 6 (2002), 443–447.
10. Arindam Dey, Andrew Cunningham, and Christian Sandor. 2010. Evaluating depth perception of photorealistic mixed reality visualizations for occluded objects in outdoor environments. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*. ACM, 211–218.
11. Pierre Dragicevic. 2016. Fair statistical communication in HCI. In *Modern Statistical Methods for HCI*. Springer, 291–330.
12. Andreas Dünser, Raphaël Grasset, and Mark Billinghurst. 2008. *A survey of evaluation techniques used in augmented reality studies*. Human Interface Technology Laboratory New Zealand.
13. Andreas Dünser, Karin Steinbügl, Hannes Kaufmann, and Judith Glück. 2006. Virtual and Augmented Reality As Spatial Ability Training Tools. In *Proceedings of the 7th ACM SIGCHI New Zealand Chapter's International Conference on Computer-human Interaction: Design Centered HCI (CHINZ '06)*. ACM, New York, NY, USA, 125–132. DOI: http://dx.doi.org/10.1145/1152760.1152776
14. Chris Furmanski, Ronald Azuma, and Mike Daily. 2002. Augmented-reality visualizations guided by cognition: Perceptual heuristics for combining visible and obscured information. In *Mixed and Augmented Reality, 2002. ISMAR 2002. Proceedings. International Symposium on*. IEEE, 215–320.
15. Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology* 52 (1988), 139–183.
16. Mary Hegarty and David Waller. 2004. A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence* 32, 2 (2004), 175–191.
17. Hikaru Ibayashi, Yuta Sugiura, Daisuke Sakamoto, Natsuki Miyata, Mitsunori Tada, Takashi Okuma, Takeshi Kurata, Masaaki Mochimaru, and Takeo Igarashi. 2015. Dollhouse vr: a multi-view, multi-user collaborative design workspace with vr technology. In *SIGGRAPH Asia 2015 Emerging Technologies*. ACM, 8.
18. Victoria Interrante, Brian Ries, and Lee Anderson. 2006. Distance perception in immersive virtual environments, revisited. In *Virtual Reality Conference, 2006*. IEEE, 3–10.
19. J Adam Jones, J Edward Swan II, Gurjot Singh, Eric Kolstad, and Stephen R Ellis. 2008. The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization*. ACM, 9–14.
20. Shunichi Kasahara, Shohei Nagai, and Jun Rekimoto. 2017. JackIn Head: Immersive Visual Telepresence System with Omnidirectional Wearable Camera. *IEEE transactions on visualization and computer graphics* 23, 3 (2017), 1222–1234.
21. Shunichi Kasahara and Jun Rekimoto. 2014. JackIn: integrating first-person view with out-of-body vision generation for human-human augmentation. In *Proceedings of the 5th Augmented Human International Conference*. ACM, 46.
22. Kiyoshi Kiyokawa. 2007. 3d collaboration using mixed reality technology. In *Proc. of the First International Symposium on Universal Communication*. 100–109.
23. Kiyoshi Kiyokawa, Mark Billinghurst, Sohan E Hayes, Anoop Gupta, Yuki Sannohe, and Hirokazu Kato. 2002. Communication behaviors of co-located users in collaborative AR interfaces. In *Mixed and Augmented Reality, 2002. ISMAR 2002. Proceedings. International Symposium on*. IEEE, 139–148.

24. Kiyoshi Kiyokawa, Hidehiko Iwasa, Haruo Takemura, and Naokazu Yokoya. Collaborative immersive workspace through a shared augmented environment.

25. Ryohei Komiyama, Takashi Miyaki, and Jun Rekimoto. 2017. JackIn space: designing a seamless transition between first and third person view for effective telepresence collaborations. In *Proceedings of the 8th Augmented Human International Conference*. ACM, 14.

26. Robert Kooima. 2008. Generalized perspective projection. *School of Elect. Eng. and Computer Science* (2008), 1–7.

27. Kai S Lehmann, Joerg P Ritz, Heiko Maass, Hueseyin K Çakmak, Uwe G Kuehnapfel, Christoph T Germer, Georg Bretthauer, and Heinz J Buhr. 2005. A prospective randomized study to test the transfer of basic psychomotor skills from virtual reality to physical reality in a comparable training setting. *Annals of surgery* 241, 3 (2005), 442.

28. Jason Leigh, Andrew E Johnson, Christina A Vasilakis, and Thomas A DeFanti. 1996. Multi-perspective collaborative design in persistent networked virtual environments. In *Virtual Reality Annual International Symposium, 1996., Proceedings of the IEEE 1996*. IEEE, 253–260.

29. Marcia C Linn and Anne C Petersen. 1985. Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child development* (1985), 1479–1498.

30. Mark A Livingston, Catherine Zanbaka, J Edward Swan, and Harvey S Smallman. 2005. Objective measures for the effectiveness of augmented reality. In *Virtual Reality, 2005. Proceedings. VR 2005. IEEE*. IEEE, 287–288.

31. Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems* 77, 12 (1994), 1321–1329.

32. Jens Mueller, Roman Rädle, and Harald Reiterer. 2017. Remote Collaboration With Mixed Reality Displays: How Shared Virtual Landmarks Facilitate Spatial Referencing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 6481–6486.

33. Tien Dat Nguyen, Christine J Ziemer, Timofey Grechkin, Benjamin Chihak, Jodie M Plumert, James F Cremer, and Joseph K Kearney. 2011. Effects of scale change on distance perception in virtual environments. *ACM Transactions on Applied Perception (TAP)* 8, 4 (2011), 26.

34. Ramesh Raskar, Greg Welch, Kok-Lim Low, and Deepak Bandyopadhyay. 2001. Shader lamps: Animating real objects with image-based illumination. In *Rendering Techniques 2001*. Springer, 89–102.

35. Jun Rekimoto, Yuji Ayatsuka, and Kazuteru Hayashi. 1998. Augment-able reality: Situated communication through physical and digital spaces. In *Wearable Computers, 1998. Digest of Papers. Second International Symposium on*. IEEE, 68–75.

36. Jun Rekimoto and Katashi Nagao. 1995. The world through the computer: Computer augmented interaction with real world environments. In *Proceedings of the 8th annual ACM symposium on User interface and software technology*. ACM, 29–36.

37. Joan Sol Roo, Renaud Gervais, Jeremy Frey, and Martin Hachet. 2017. Inner Garden: Connecting Inner States to a Mixed Reality Sandbox for Mindfulness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1459–1470. DOI: http://dx.doi.org/10.1145/3025453.3025743

38. Joan Sol Roo and Martin Hachet. 2017. One Reality: Augmenting How the Physical World is Experienced by Combining Multiple Mixed Reality Modalities. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*. ACM, New York, NY, USA, 787–795. DOI: http://dx.doi.org/10.1145/3126594.3126638

39. Neal E Seymour, Anthony G Gallagher, Sanziana A Roman, Michael K OâĂŹbrien, Vipin K Bansal, Dana K Andersen, and Richard M Satava. 2002. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Annals of surgery* 236, 4 (2002), 458.

40. Frank Steinicke, Gerd Bruder, Klaus Hinrichs, Markus Lappe, Brian Ries, and Victoria Interrante. 2009. Transitional environments enhance distance perception in immersive virtual reality systems. In *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization*. ACM, 19–26.

41. Richard Stoakley, Matthew J Conway, and Randy Pausch. 1995. Virtual reality on a WIM: interactive worlds in miniature. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press/Addison-Wesley Publishing Co., 265–272.

42. Brygg Ullmer and Hiroshi Ishii. 1997. The metaDESK: models and prototypes for tangible user interfaces. In *Proceedings of the 10th annual ACM symposium on User interface software and technology*. ACM, 223–232.

43. SG Vandenberg. 1971. Mental rotation test. *Boulder: University of Colorado* (1971).

44. Ranxiao Frances Wang and Daniel J Simons. 1999. Active and passive scene recognition across views. *Cognition* 70, 2 (1999), 191–210.

45. Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 143–146.