Science Arts & Métiers (SAM)

This is an author-deposited version published in: https://sam.ensam.eu
Handle ID: .http://hdl.handle.net/10985/24564

To cite this version :

Guillaume BATAILLE, Abdelhadi LAMMINI, Jean-Rémy CHARDONNET - ARPuzzle: Evaluating the Effectiveness of Collaborative Augmented Reality - In: 22nd IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Australie, 2023-10-16 - IEEE International Symposium on Mixed and Augmented Reality (ISMAR) - 2023

2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)

ARPuzzle: Evaluating the Effectiveness of Collaborative Augmented Reality

Guillaume Bataille*
Orange XDLab

Abdelhadi Lammini†
Orange XDLab
Arts et Metiers Institute of Technology, LISPEN,
HESAM Université

Jean-Rémy Chardonnet‡
Arts et Metiers Institute of Technology, LISPEN,
HESAM Université

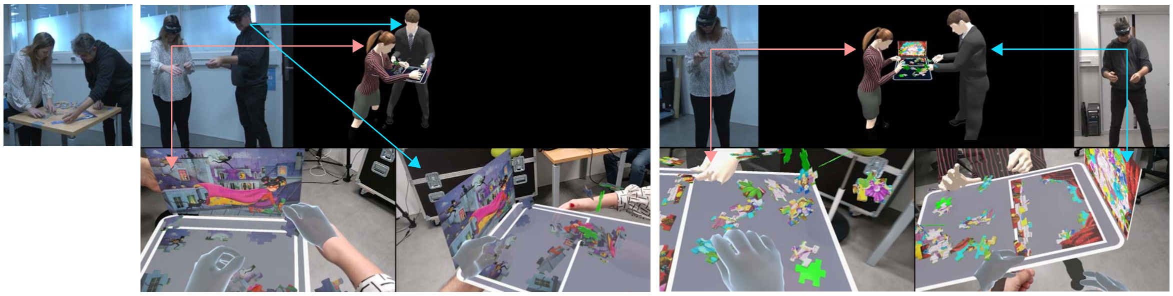


Figure 1: Our ARPuzzle experiment. On the left, two collaborators solve a real puzzle. In the middle, mixed collaborators colocally solve a virtual replica of a real puzzle. On the right, mixed collaborators remotely solve a virtual replica of a real puzzle.

ABSTRACT

Collaborative Augmented Reality (CAR) offers disruptive ways for people to collaborate. However, this emerging technology must improve its acceptance, efficiency, and usability to scale up and, for example, support augmented operations executed by technicians. This paper presents our CAR system and its experimentation during a cooperative puzzle-solving task. Our system provides collaborators with a shared virtual space allowing verbal and non-verbal interpersonal communications, and intuitive interactions with shared virtual replicas of real objects. Our system also integrates avatars embodied by remote users. We conducted a dual-user study comparing colocated and remote solving of a puzzle virtual replica with its real solving. We evaluated task performance, collaboration, mutual awareness, spatial presence, and copresence, usability, and preference. We found that, if real is preferred and more efficient than our CAR system, CAR is reaching favorable usability levels. We also found that remote augmented reality including full-body avatars offers similar results to colocated augmented reality. This preliminary work paves the way for future research aiming to support and enhance the design and making of Collaborative Augmented Reality systems dedicated to augmented operations.

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

1 INTRODUCTION

Collaborative Augmented Reality (CAR) allows multiple users to collaborate on tasks involving virtual objects. CAR users perceive a mixed environment, composed of their real environment and a Shared Virtual Space (SVS) [8]. Colocated CAR users collaborate in the same mixed environment. Remote CAR users, incarnated by avatars mimicking their movements, ubiquitously communicate verbally and non-verbally thanks to their avatar providing copresence. Task objects can be real, mixed, or virtual. Virtual objects are virtual replicas [30] or virtual twins [25] of real objects. For example, a virtual mockup of an industrial machine synchronized with its real twin allows its ubiquitous control by multiple collaborators.

Sanitary and environmental crises highly raised the need for collaborative tools like emails, instant messages, phones, and videoconferences. But these tools separate task and communication spaces and restrict communication cues exchanged between users [16]. A SVS merges task and communication spaces in a common frame of reference enhancing awareness and understanding between collaborators [10]. For Kim et al. [22], CAR has the potential to improve or even replace typical collaborative technologies. Tech Giants are concurrently working on this potential growth driver. Microsoft presented the Mesh platform[1], enabling ubiquitous copresence and shared experiences through Mixed Reality (MR). Meta launched the Horizon Worlds platform[2], dedicated to Virtual Reality-based social experiences. The Metaverse hype shows the potential of SVS-based technologies to replace real, non-intermediated human activities with immersive technologies [46].

Few studies have investigated CAR, as shown in Table 1. Researchers typically associate an AR local technician with a Virtual Reality (VR) remote expert. But compared to VR, AR keeps users aware of their real surroundings, is safer to use in public and industrial environments, and reduces the risk of cybersickness [42]. For these reasons, CAR systems offer a wider range of uses for better scalability. For example, augmented technicians may synchronously intervene on real objects or machines at different locations. For these reasons, we used Optical See-Through Head Mounted Displays (OST-HMD). We also enabled collaborators to embody full-body avatars interacting intuitively with SVS. Compared to prior work, the primary novel contributions of this paper include:

- A novel CAR system enabling mixed users to intuitively collaborate on tasks remotely or colocally through symmetric OST-HMDs in an SVS. Remote participants embody avatars mimicking their head, eyes, hands and fingers movements, and voice. Bare hands users intuitively interact with virtual objects

---

*e-mail: guillaume2.bataille@orange.com
†e-mail: abdelhadilammini@gmail.com
‡e-mail: jean-remy.chardonnet@ensam.eu

[1] https://www.microsoft.com/en-us/mesh
[2] https://www.meta.com/horizon-worlds/

2473-0726/23/$31.00 ©2023 IEEE
DOI 10.1109/ISMAR59233.2023.00079

by touching, pushing, or grabbing them. Our CAR system provides a high degree of combined user tracking, affordance, and intuitiveness,

- A formal user study ($n = 36$) compares the solving of real puzzles with the solving of their virtual replicas in the cases of two colocated and remote participants. We compared task performance, collaboration, mutual awareness, spatial presence, copresence, usability, and preferences. We also evaluated the usability of our system. To our knowledge, only one user study compared real collaboration with colocated CAR, and none compared real collaboration with colocated and remote CAR.

2 RELATED WORK

Intermediated collaboration, i.e., collaboration based on the use of interpersonal communication systems, is widely explored for decades. Recent **surveys** show a growing interest in CAR. While CAR is sparsely studied [13, 37], this technology has the potential to overcome current collaborative tools for Ladwig et al. [26]. Ens et al. [15] pointed out that MR has recently reached the maturity required to enhance environmental awareness and communication between collaborators.

Projected **video** and voice were initially employed for remote collaboration [16, 19, 21, 24]. These works proved the benefits of video sharing as an effective remote communication technology in collaborative tasks, such as memory stimulation, communication grounding, workspace perception, and task completion swiftness, compared to an audio-only experience. However, video separates task and communications spaces. Billinghurst et al. [7] compared real, non-intermediated solving of a real puzzle as a baseline with colocated AR and video technologies. AR increased communication compared to videos. AR performance was lower than non-intermediated collaboration. They also compared remote video avatars with desktop video conferencing [6] and noticed that spatialized avatars enhance co-presence and understanding between collaborators.

Remote CAR users need **communication cues** to collaborate. Piumsomboon et al. [32] studied collaboration between augmented and virtual users during a search task. Eye and head tracking provided remote collaborators with a higher co-presence than a baseline providing only verbal cues. For Masai et al. [28], communication and empathy between remote collaborators are improved by facial expressions, eye gaze, and internal parameters such as heart rate. Gupta et al. [18] showed that gaze combined with pointing increased communication quality and task performance. These studies reveal the importance of providing remote collaborators with non-verbal communication cues, notably head, and eye gaze, to increase co-presence and task performance.

A commonly studied Extended Reality (XR) setup consists of two users with asymmetric **roles and technologies** collaborating synchronously. A technician wearing an AR-HMD is assisted by a remote VR expert. In that typical case, the expert provides specific knowledge to the technician to accomplish a task in the real world. The expert perceives the technician's environment, talks with him, and may augment the technician's environment with annotations or virtual objects. Piumsomboon et al. [31, 33] stated that awareness cues are important for enhancing usability, user experience, and efficiency in such setups. Wang and Dunston [45] compared a pen-and-paper collaborative task with colocated and remote mixed reality setups. However, their experimental design provided limited user tracking and interpersonal communication. Users benefited in similar proportions with both colocated and remote mixed reality compared to the pen and paper condition. Smith et al. [39] compared colocated asymmetric tasks in real and virtual conditions, with and without avatars in the VR condition. Both embodied VR and real conditions provided a high level of social presence, unlike the no avatar VR condition. Waldow et al. [44] compared remote and colocated asymmetric tasks involving partial outline avatars. They stated that seeing such avatars had similar effectiveness to real users. Chen et al. [11] studied visual cues with two gaze-pointing techniques in colocated CAR. However, **symmetric technologies and roles** are rarely studied simultaneously in CAR systems, as shown in Table 1, while they are technically simpler to create, provide equal interpersonal communication cues, and facilitate scalability thanks to their homogeneity. For these reasons, we focus on a symmetric CAR system and its experimentation in this paper.

We listed the main **data types and variables** of previous work studying collaborative extended reality in Table 1. This table lists experimental technologies and task symmetries, studied conditions and criteria of previous studies, and emphasizes our experimental design specificities described in section 4. The most redundant objective data is task completion time. Non-verbal communication is also analyzed, mainly the number and type of gestures, walking distance, and mutual gaze, as shown in Table 1. Verbal communication is also studied in the literature, considering the number of words per sentence, speaker turns, and deictic sentences. The most commonly studied subjective metric is usability. Mainly used questionnaires are the System Usability Scale (SUS) questionnaire [9], the Single Ease Question (SEQ) questionnaire [34], and the Subjective Mental Effort Question (SMEQ) questionnaire [34, 47]. User preference appears to be the second most used metric (Table 1). User preference questionnaires are useful to compare the acceptance of experimental conditions [40]. Copresence and social presence are major variables regarding CAR aims [3, 20, 28]. The spatial presence questionnaire is also helpful to evaluate the efficiency of an SVS [27, 41, 43]. The acceptance of collaborative XR systems is another important aspect to expand and scale up the use of CAR systems [12, 14, 36].

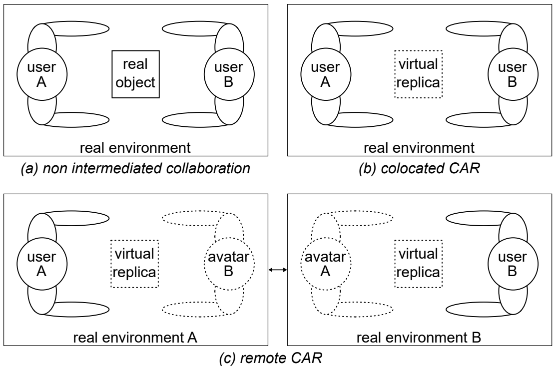


Figure 2: Non-intermediated, colocated, and remote augmented collaboration. (a) two users collaborate face-to-face on a task involving a real object with no intermediation. (b) two augmented users collaborate face-to-face on a synchronous task involving a virtual replica of the real object. (c) two augmented users remotely collaborate on a synchronous task involving a virtual replica of the real object. Each user embodies his humanoid avatar in the other user environment.

As emphasized in Table 1, no previous research implemented and compared **real, colocated, and remote** CAR with **symmetric** task and extended tracking technologies. Figure 2 presents the spatial configurations we address in this paper. Using head, eyes, hands, and fingers tracking allows users to embody **realistic full-body avatars** able to **intuitively** interact with an SVS. A CAR system enabling these features would allow us to evaluate their maturity, usability, and efficiency. This is important to estimate the capability of CAR systems to replace non-intermediated face-to-face collaboration. Comparing non-intermediated with colocated and remote CAR would help us concentrate on the most problematic aspects of this technology.

Table 1: Data type and variables used in collaborative extended reality studies. Bold aspects are similar to our CAR system. Non-bold aspects emphasize the differences between our experimental setup and related work. Device acronyms are Monoscopic (Mo), Video Pass-Through Head-Mounted Display (VPT-HMD), Virtual Reality Head-Mounted Display (VR-HMD), and Optical See-Through Head-Mouted Display (OST-HMD). Compared conditions acronyms are Egocentric (Ego), Exocentric (Exo), Colocated (Co), Remote (Re), Local (Lo), and Facial Expression (Fe). Objective acronyms are Task Completion Time (TcT), Accuracy (Ac), Number and type of Gestures (Ng), Number of Words per Phrase (NwP), Number/type of Speaker Turns (NsT), Number of Deictic Phrases (NdP), Object Positioning Time (OpT), Walking Distance (Wd), Number of Phrases (Np), Head Movements (Hm), Mutual gaze (Mg), and Inter-Personal Distance (IpD). Subjective measured data: Observations (O), Usability (U), Interview (I), Collaboration and communication feeling (C), Mutual Understanding (Mu), User Preference (Up), Phrase Awareness (Pa), Perceived Effort (Pe), Mutual Awareness (Ma), Communication (Cm), Empathy (Ep), Copresence (Co), Social Presence (Sp), Level of Focus (Lf), Feedback (F) User Experience (Ux), Mental Effort (Me), Motion Sickness (Ms), Task Load (Tl), Presence (P), Avatar Perception (Ap), Spatial Presence (SaP), Task Difficulty (Td).

Year	**Reference**	**Device**	**Task symmetry**	**Tracking**	**Compared conditions**	**Measured data**	
						objective	**subjective**
2000	Billinghurst et al. [8]	VPT-HMD (2)	**sym**	both (QRCode)	**Co AR w.o. avatar**		**O**
2002	Billinghurst et al. [7] [5]	VPT-HMD (2) 2D projector	**sym**	QRCode	**Co AR w.o. avatar** Co projection **Co Real**	**TcT** Ng NwP NsT NdP	**U** I **Mu Up** **C** Pa
2004	Schafer et al. [35]	desktop	asym		Ego Exo	**TcT** OpT	C Pe **Ma**
2004	Benko et al. [4]	VPT-HMD tactile table screen - tablet	asym	head+hand gestures			**O**
2005	Grasset et al. [17]	VR-HMD VPT-HMD	asym	both (head) +QRCode	single Ego VR Ego VR+Exo VR Ego VR+Exo AR	**TcT** Hm Wd	**Ma** Cm
2009	Wang et al. [45]	VPT-HMD (2)	**sym**	both (tracker ball+ QRCode)	benchmark Co AR - Re AR	**TcT**	
2016	Masai et al. [28]	OST-HMD laptop	**sym**	eye+face	all (Lo video) +Lo gaze +Lo Fe +Lo (gaze+Fe)		Ep **C Ma**
2016	Gupta et al. [18]	Mo OST-HMD screen	asym	eye	all (Lo video) +Re pointer +Lo eye +Lo eye+Re pointer	**TcT** Np	**Co Mu Up**
2017	Piumsomboon et al. [31] [32]	VR-HMD OST-HMD	**sym** asym	**both (head+eye+ hands+fingers)**	both (Re (head+hands)) +Re FoV	Mg Wd	**U**
2018	Piumsomboon et al. [33]	VR-HMD OST-HMD	**sym**	both (head+ hand gestures)	all (Re full-body avatar) + Re mini-avatar	**TcT**	**U** F **Up** Lf Sp
2018	Smith et al. [39]	VR-HMD (2)	asym	both (head+ hand gestures)	Co VR full-body avatar Co VR w.o. avatar **Co Real**	Ng NdP NsT	**Cp** Sp **Ma** Cm **Up**
2019	Kim et al. [23]	OST-HMD VR-HMD	asym	both (head) hands+fingers	all (Lo video+Re hands) +Re pointer +Re sketch +Re (pointer+sketch)	**TcT**	**U Up** Me **Co**
2019	Teo et al. [40]	VR-HMD OST-HMD	asym	hands+fingers+eye 360° camera	Lo 360° Lo 3D Lo (3D+360°)	**TcT**	**U Up** Sp Ux Ms
2019	Waldow et al. [44]	**OST-HMD (2)**	asym	both (head+ hand gestures)	Re AR w.o. avatar Re AR partial avatar **Co AR w.o. avatar**	**TcT**	Tl P Ap
2020	Bai et al. [1]	OST-HMD VR-HMD	asym	eye+hand gesture	all (voice) +Re eye +Re hand gesture +Re (eye+hand gesture)	**TcT**	Sp **SaP** Ti
2021	Chen et al. [11]	**OST-HMD (2)**	**sym**	head	pointing techniques object density+movement	**TcT** Ac	Sp **U Up**
	Our approach	**OST-HMD (2)**	**sym**	**both (head+eye+ hands+fingers)**	**Re AR full-body avatar** **Co AR w.o. avatar** **Co Real**	**TcT**	**U O C Mu Up Td Co Ma SaP**

3 SYSTEM OVERVIEW

In this section, we present our novel CAR system enabling intuitively interactive SVS with symmetric tasks and technologies. Our system provides the following features. Collaborators wear an OST-HMD ubiquitously immersing them in a mixed environment composed of an SVS and their real environment. They see the SVS scaled at real-world dimensions over the real world from a first-person Point of View (PoV). OST-HMDs provide intuitive interactions between collaborators and the SVS, and both verbal and non-verbal communication between remote collaborators. OST-HMDs enable users to touch, push or grab with one or two bare hands the virtual objects that are part of the SVS. OST-HMDs synchronize in real-time the pose of the virtual entities composing the SVS through the Internet. We use a client/server service named Photon PUN 2[3] to synchronize in real-time the pose of virtual objects and avatars composing the SVS. OST-HMDs send the coordinates of their user's avatar and of the virtual objects their user interacts with to the server. The server broadcasts this data to other clients to update the SVS in real-time. Fig. 3 describes this architecture for a pair of collaborators and a single mixed object, but more users can collaborate on tasks involving multiple mixed objects synchronously. Colocated collaborators naturally communicate without any intermediation. In that case, the SVS is initially anchored to the real environment by each OST-HMD client, providing spatial consistency between colocated collaborators. OST-HMD clients transform their local coordinate system into the global SVS coordinate system. In remote cases, the OST-HMD additionally provides intermediated verbal and non-verbal communication between collaborators, incarnated by realistic full-body avatars. We use a voice chat service named Dissonance[4] for verbal communication. We implemented our OST-HMD client with Unity[5] 2020.3.36.

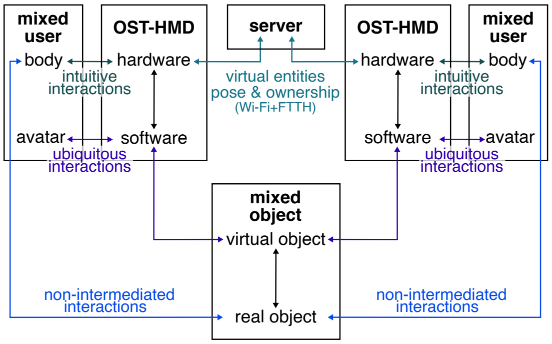


Figure 3: The architecture of our CAR system is composed of two mixed users and one mixed object. Mixed users intuitively interact with the SVS thanks to their OST-HMD. The pose of each dynamic virtual entity composing the SVS is transmitted between a server centralizing and distributing this data and OST-HMD clients.

Mixed collaborators wear an OST-HMD, a Microsoft Hololens 2. OST-HMDs track their own user's head, eyes, hands and fingers, and instantiate and animate the incarnated full-body avatar of their user in the SVS. In remote cases, OST-HMDs also record their own user voice, play the voice of remote collaborators, and display embodied avatars mimicking the real movements of remote users in the SVS. Embodied avatars are initially generated by the Make Human Project[6], a user-friendly avatar editor. This open-source editor allows its users to easily personalize the parameters of an avatar, like its height, width, hand length, or eye color, and to dress it in virtual clothes. Avatars are animated by inverse kinematics based on the head, eye, hand, and finger tracking of the real user. The VRIK plugin from RootMotion[7] provides inverse kinematics. Fig. 4 shows the mixed collaborator lifecycle in ARCollab.

Mixed collaborators **directly interact** with the virtual world thanks to their displayed virtual hands. Users see their own virtual hands in any AR case, awaring them of tracking failures and inaccuracy due to real occlusions. Virtual hands also provide occlusion management between real hands and virtual objects. An occluding shader may be employed to render virtual hands at the cost of hand-tracking awareness and accuracy. Intuitive interactions allow collaborators to touch virtual objects based on simulated physics with the Unity physics engine, and to grasp and manipulate them with one or two hands by using the MRTK grasping technique[8]. We cut off gravity in the SVS to save users from failing to grasp virtual objects felt on the ground since Hololens 2 does not detect hands near surfaces. Virtual objects are slowed down by parameterized friction to avoid infinite moves and also collide together.

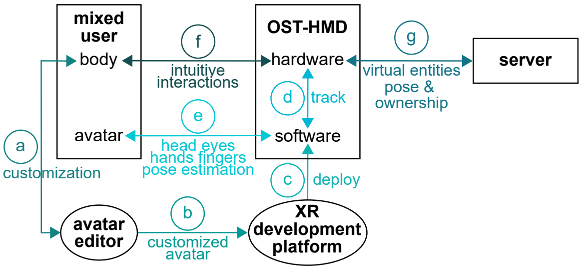


Figure 4: The mixed user lifecycle. (a) A user creates his/her personalized avatar. (b) The avatar is integrated into the OST-HMD application, a client of the synchronization server. (c) The client is deployed to the OST-HMD. (d) The OST-HMD tracks the user's external activity. (e) The tracked data is used to support the virtual twin of the collaborator and animate his/her avatar. (f) The OST-HMD provides intuitive interactions between the mixed collaborator and the SVS. (g) The OST-HMD exchanges with the server the position and state of virtual entities constituting the SVS in order to update and synchronize in real-time this data between clients.

Our novel CAR system enables colocated and remote CAR users to intuitively interact with SVS through their embodied full-body avatar. Table 1 shows that our system differs from previous work, with none to less than average similarities.

4 USER STUDY

We describe in this section our user study called ARPuzzle. This experimental design was initially inspired by a study from Billinghurst at al. [5] comparing a collaborative real task with its simulation in tangible colocated augmented reality and video projection. We chose to compare a non-intermediated collaboration involving a real object with colocated and remote AR-based collaboration involving the virtual replica of the real object. This experimental design allowed us to evaluate separately and compare the impacts of simulating a real object with its virtual replica and intermediating a natural face-to-face interpersonal communication with its simulation based on embodied full-body human avatars.

4.1 Apparatus

Each participant wore a Hololens 2 during AR-based conditions. We employed real puzzles and their virtual replicas as non-synchronized mixed objects. The experimenter used a laptop to observe and record the SVS during AR-based conditions. The OST-HMDs and the laptop communicated with the server through a Wi-Fi router connected to the Internet by an optical fiber connection (FTTH).

4.2 Task

The dual-user task consisted of solving a puzzle. We used one 6-piece virtual puzzle for training, and three 36-piece real puzzles and

---

[3] https://www.photonengine.com/en/PUN

[4] https://placeholder-software.co.uk/dissonance/

[5] https://unity.com/

[6] http://www.makehumancommunity.org/

[7] http://root-motion.com/

[8] https://github.com/Microsoft/MixedRealityToolkit-Unity

their virtual replicas for collecting data. All 36-piece puzzles were designed for at least four years old players by the same manufacturer. The puzzles were all 16.5” x 11.8”. Each 36-piece puzzle was 2.75” by 1.96” average. For this reason, we considered all 36-piece puzzles to be equally difficult to solve. Fig. 5 presents the puzzles used.

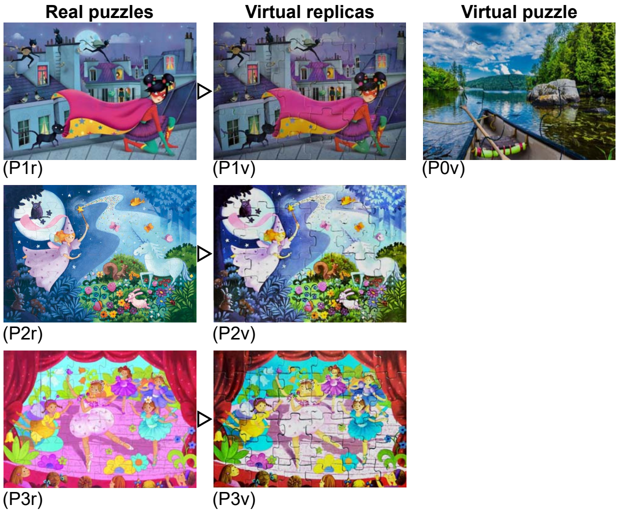


Figure 5: Puzzles used during the experiment. On the left, is a picture of the real 36-piece puzzles P1r, P2r, and P3r. In the middle, the 36-piece virtual replicas P1v, P2v, and P3v of the real puzzles. On the right, is the 6-piece virtual puzzle P0v used during training.

Solving a real puzzle is a non-intermediated task providing a ground truth as a baseline since we want to compare CAR with real collaboration. Solving a puzzle virtual replica occurs in an SVS containing a virtual table upon which its virtual pieces stand. The final puzzle to obtain is rendered in front of the virtual table. Collaborators must place pieces in a solving frame in order to validate the task achievement. Fig. 6 shows the SVS we designed for this experiment. Users simultaneously collaborate on solving these puzzles and their virtual replicas with their bare hands. They move the virtual puzzle pieces by touching, pushing, or grabbing them with one or two hands. When a user positions a puzzle piece close to its final position a magnetic effect freezes the puzzle piece at its final position. Both position and rotation are finely thresholded in order to avoid hacks consisting in moving pieces over the solving frame without seeking their position as detected in preliminary tests. Puzzle piece ownership is managed by allowing ownership to the first participant grabbing the piece, and freeing it when the participant releases it.

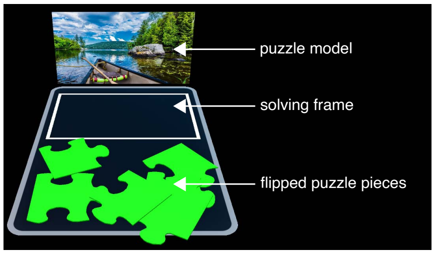


Figure 6: The ARPuzzle SVS with the training puzzle P0v.

4.3 Conditions

We considered three conditions applied to two collaborators in pairs. The first one is the Real (R) condition, consisting of a real puzzle-solving with no intermediated interaction. The colocated AR (ARc) and the remote AR (ARr) conditions both required the use of an OST-HMD by each participant to interact with the virtual replica of the real puzzle and solve it. In the ARc condition, both participants were in the same environment and directly interacted together with no intermediation. A QR code placed on the ground provided a spatial anchor ensuring that both participants viewed the SVS at the same location. In the ARr condition, remote participants were in adjoining rooms and could see their collaborator’s avatar and talk together. Participants were free to move and walk. In each room and the SVS, a camera recorded the experiments. Fig. 1 shows these conditions in action.

4.4 Hypotheses

We expected the real condition (R) to be the most efficient one. The colocated AR (ARc) condition provides real interpersonal communication cues, while the remote AR condition (ARr) provides partial intermediated communication cues. Therefore we expected ARc to perform better than ARr. From these assumptions, we made the following hypotheses:

H1 the performance of the R condition is the best, followed by the ARc condition, and the ARr performance is the worst,
H2 the R condition is the easiest and the most usable, the ARc condition is easier and more usable than the ARr condition,
H3 spatial presence, copresence, awareness and collaboration, and usability in the remote AR condition are lower than in the colocated AR condition, both AR conditions having lower results than the R condition,
H4 R is the most preferred condition.

4.5 Metrics

We selected from our literature review a subset of the most commonly used metrics in collaborative XR studies. We also evaluated our interactive system acceptance. Our objective metric was the task completion time. Our subjective metrics were task difficulty, collaboration and mutual awareness, spatial presence, copresence, usability, and preference. Task difficulty was evaluated with the Single Ease Question questionnaire (SEQ) by Sauro et al. [34]. It consisted of the phrase ”Overall, this task was:”, asking to rate the task difficulty on a 7-point Likert scale. We used the collaboration and mutual awareness questionnaire created by Masai et al. [28]. It consisted of seven questions rated on a 7-point Likert scale. Regarding spatial presence, Self Location (SPSL) and Possible Actions (SPPA) were measured with the MEC spatial presence questionnaire (MEC-SPQ) by Vorderer et al. [41]. Each questionnaire consisted of eight items, each of them rated on a 7-point Likert scale. We evaluated copresence with the questionnaire used by Basdogan et al. [3] and based on the Slater et al. presence and co-presence questionnaire [38]. It consisted of eight questions rated on a 7-point Likert scale. We quantified usability with the System Usability Scale (SUS) questionnaire by Brooke et al. [9]. It consisted of ten questions rated on a 7-point Likert scale. The Preference questionnaire was adapted from Teo et al. [40]. We asked participants to state their preferred condition between R, ARc, and ARr against six criteria.

4.6 Subjects

42 subjects participated in this experiment. Because of partially collected data, we removed 3 pairs from the collected data, and 3 other pairs of participants ran the same conditions in the same order. 36 participants provided the final dataset, 21 males (58.3%) and 15 females (41.7%). Participants were aged between 13 and 61 ($M = 48 \pm 11.88$). Student’s t-tests did not reveal any significant

difference between participants in AR ($t(34) = 1.0, p = .32$) and VR familiarity ($t(34) = 1.32, p = .20$).

4.7 Procedure

Participants experimented with all conditions and puzzles balanced with Latin Square. Participants were split into six groups composed of three pairs of participants each. Each pair of a group experienced the same puzzles in the same condition order. We assumed that all pairs of participants had the same level of expertise in solving puzzles and that all 36-piece puzzles were of equal difficulty. All participants first filled out information and consent forms. Then they learned how to use the OST-HMD to launch, configure and close the application. After launch, participants chose between a male and a female avatar, and selected the puzzle to solve and the condition to consider under the experimenter's control. Next, participants trained to solve the simple virtual 6-piece puzzle P0v during the training phase. They trained with the ARc condition before the ARr one and learned to register their real environment with a QR code pose estimation. For the remote condition, each participant was located in an adjacent room. Table 2 summarizes the order of conditions and puzzles during the experiment. Finally, participants solved a 36-piece puzzle under each condition. After each solving, they completed the subjective questionnaires presented above, except the preference one. When participants experimented with all conditions, they completed the preference questionnaire. Cameras recorded all conditions. During both AR conditions, the SVS was also recorded. Task completion time measurement started when both users were immersed in the shared virtual space and observed the virtual puzzle table. Positioning the last puzzle piece ended measurement.

4.8 Results

We measured our system's latency around 33ms by measuring the latency between the movement of one user grabbing a virtual puzzle piece and its movement in the SVS observed by another user facing him. We observed that no participant walked around the real puzzle. The ARc condition generated walks because collaborators avoided colliding together. Few participants expressed fatigue or a temporarily affected perception after HMD use. Participants regularly expressed that they refined their understanding of the interaction techniques after training. This may be due to 36 puzzle pieces being smaller than 6 puzzle pieces, requiring more accuracy. For all data analyses, normality checks and homogeneity of variances tests were first performed. The significance level was set to .05 for all analyses. Results are depicted in Fig. 7.

Completion time

First, we checked that all three puzzles were equally difficult. A Friedman's test did not reveal any significant difference in completion time between each puzzle ($M_1 = 3min11s \pm 30s, M_2 = 2min58s \pm 34s, M_3 = 4min9s \pm 1min52s$), $\chi^2(2) = 0.33, p = .85$. In the following analyses, we will thus consider the puzzles as not being influential on any result. A Friedman's test found significant differences between each condition, $\chi^2(2) = 27.11, p < .001$ (Fig. 7a). Post-hoc Wilcoxon signed-rank tests using a Bonferroni correction ($\alpha = .05/3 = .017$) revealed that the R condition ($M = 3min26s \pm 1min12s$) was significantly faster to complete than the ARc condition ($M = 9min57s \pm 3min34s$), $T = 0, z = 3.70, p < .001$, and the ARr condition ($M = 8min51s \pm 2min21s$), $T = 0, z = 3.70, p < .001$. However, no significant difference was found between the ARc and the ARr conditions, $T = 49, z = 1.57, p = .12$.

SEQ

A Friedman's test revealed significant differences between the conditions, $\chi^2(2) = 49.78, p < .001$ (Fig. 7b). Post-hoc Wilcoxon signed-rank tests using a Bonferroni correction revealed that the R condition ($M = 1.5 \pm 0.91$) was significantly easier than the ARc condition ($M = 3.92 \pm 1.52$), $T = 0, z = 5.04, p < .001$, and the ARr condition ($M = 3.72 \pm 1.39$), $T = 14.5, z = 4.87, p < .001$. However, we found no significant difference between both AR conditions, $T = 140.5, z = 0.90, p = .37$. As we asked participants to rate their familiarity with AR and VR technologies, we analyzed whether familiarity could influence the SEQ scores. Since there was no significant difference in the SEQ scores between both AR conditions, we grouped them into a single AR condition taking the average of the scores in both AR conditions. Then Pearson correlation tests did not reveal any significant correlation between familiarity and SEQ scores ($r_{famAR} = .22, p = .20$ and $r_{famVR} = -.05, p = .76$). We performed the same with gender. However, no correlation was found with SEQ scores ($r = .24, p = .16$). We then computed an average SEQ score for each pair of participants and checked whether the level of perceived difficulty in both AR conditions could be correlated with the time to complete the puzzles in these conditions. Pearson correlation tests revealed no significant correlations ($r_{ARc} = -.13, p = .62$ and $r_{ARr} = -.23, p = .35$).

SUS

We recalculated the scores to get total SUS scores out of 100. A Friedman's test revealed significant differences between the conditions, $\chi^2(2) = 31.82, p < .001$ (Fig. 7c). Post-hoc Wilcoxon signed-rank tests using a Bonferroni correction showed that the R condition ($M = 90.71 \pm 12.65$) was significantly more usable than the ARc condition ($M = 66.94 \pm 14.92$), $T = 21.5, z = 4.89, p < .001$, and the ARr condition ($M = 68.17 \pm 14.67$), $T = 14, z = 4.92, p < .001$. However, we found no significant difference between AR conditions, $T = 274, z = 0.92, p = .36$. We analyzed further whether familiarity with immersive technologies could influence the SUS scores. Since there was no significant difference in the SUS scores between both AR conditions, we grouped them into a single AR condition taking the average of the scores in both AR conditions. Pearson correlation tests did not reveal any significant correlation between familiarity and SUS scores ($r_{famAR} = -.11, p = .54$ and $r_{famVR} = -.05, p = .78$).

Copresence

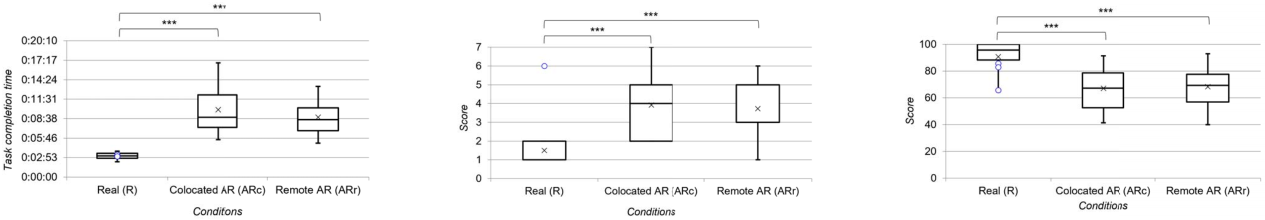
Following Basdogan et al.'s procedure [3], we counted the number of questions rated more than 6. A Friedman's test revealed significant differences between the conditions, $\chi^2(2) = 23.92, p < .001$ (Fig. 7d left). Post-hoc Wilcoxon signed-rank tests using a Bonferroni correction revealed that the R condition ($M = 5.47 \pm 1.49$) generated significantly higher copresence than the ARc condition ($M = 3.19 \pm 2.07$), $T = 60, z = 4.19, p < .001$, and the ARr condition ($M = 3.19 \pm 2.08$), $T = 41.5, z = 4.17, p < .001$. However, we found no significant difference between AR conditions, $T = 146, z = 0.10, p = .92$. We then checked for differences in the number of ratings over 6 between participants of each pair in each condition, to study whether participants within a pair could have a different perception of copresence. A Friedman's test found significant differences among the conditions, $\chi^2(2) = 10.87, p = .004$ (Fig. 7d right). From post hoc Wilcoxon signed-rank tests using a Bonferroni correction, the number of ratings over 6 revealed not to be significantly different between participants within a pair in the R condition ($M = 1.17 \pm 1.15$) and in the ARc condition ($M = 2.39 \pm 1.54$), $T = 34.5, z = 2.00, p = .045$, compared to the ARr condition ($M = 2.39 \pm 1.54$), $T = 6, z = 2.68, p = .007$. Furthermore, we found no significant difference in the number of ratings over 6 within pairs between both AR conditions, $T = 57.5, z = 0.11, p = .91$. We further checked whether copresence scores could be correlated to completion times. Pearson correlation tests did not reveal any significant correlation in all conditions ($r_R = -.26, p = .30$, $r_{ARc} = .05, p = .85$ and $r_{ARr} = -.22, p = .38$).

Collaboration and mutual awareness

Table 3 summarizes the results. For all questions except Q4, we found no significant differences between conditions. Pairwise comparisons with Wilcoxon signed-rank tests revealed that overall, the real condition provided higher collaboration and mutual awareness

Table 2: Order of phases per group from left to right during experimentation. Each phase is composed of a condition and a puzzle. ARc is the colocated AR condition. ARr is the remote AR condition. R is the real condition. P0v is the virtual 6-piece puzzle used for training. P1r, P2r, and P3r are real 36-piece puzzles. P1v, P2v, and P3v are their virtual replica.

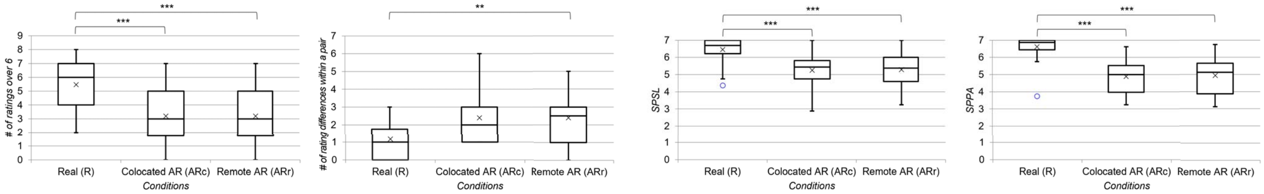
Group	**Phase (condition+puzzle)**				
	Training		**Measures**		
A	ARc+P0v	ARr+P0v	R+P1r	ARc+P2v	ARr+P3v
B	ARc+P0v	ARr+P0v	R+P1r	ARr+P2v	ARc+P3v
C	ARc+P0v	ARr+P0v	ARc+P1v	ARr+P2v	R+P3r
D	ARc+P0v	ARr+P0v	ARc+P1v	R+P2r	ARr+P3v
E	ARc+P0v	ARr+P0v	ARr+P1v	ARc+P2v	R+P3r
F	ARc+P0v	ARr+P0v	ARr+P1v	R+P2r	ARc+P3v



(a) Task completion time

(b) SEQ scores

(c) SUS scores



(d) Copresence scores (left: number of ratings over 6 per participant; right: differences between the number of ratings over 6 within pairs)

(e) Spatial presence scores (left: SPSL; right: SPPA)

Figure 7: Task completion, SEQ, SUS, Copresence and Spatial Presence scores in all conditions.

than both AR conditions, except for Q1 (“My partner and I worked well together on the task”) for which pairwise comparisons did not find any difference, Q5 (“I understood how my partner was feeling”) with no significant difference between the R and the ARc conditions, and Q7 (“I was satisfied with the output of the task”) with no significant difference between the R and the ARr conditions.

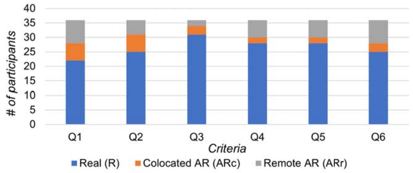


Figure 8: Participants’ preferred condition against six criteria.

Spatial presence

For SPSL, a Friedman’s test found significant differences between the conditions, $\chi^2(2) = 35.75, p < .001$ (Fig. 7e left). Post-hoc Wilcoxon signed-rank tests using a Bonferroni correction revealed that the R condition ($M = 6.46 \pm 0.67$) enabled significantly higher self-location than the ARc condition ($M = 5.25 \pm 0.95$), $T = 45.5, z = 4.41, p < .001$, and the ARr condition ($M = 5.28 \pm 0.91$), $T = 56.5, z = 4.34, p < .001$. However, we found no significant difference between AR conditions, $T = 216, z = 0.33, p = .74$. Similarly, for SPPA, a Friedman’s test showed significant differences between the conditions, $\chi^2(2) = 47.03, p < .001$ (Fig. 7e right). Post-hoc Wilcoxon signed-rank tests using a Bonferroni correction revealed that the R condition ($M = 6.63 \pm 0.62$) enabled significantly more possible actions than the ARc condition ($M = 4.89 \pm 0.97$), $T = 3, z = 5.18, p < .001$, and the ARr condition ($M = 4.94 \pm = 1.07$), $T = 7, z = 5.12, p < .001$. However, we found no significant difference between AR conditions, $T = 227, z = 0.68, p = .49$.

Preference

A Kruskal-Wallis test found significantly different numbers of participants between each preferred condition, $H(2) = 12.32, p = .002$ (Fig. 8). Overall, from pairwise Mann-Whitney tests using a Bonferroni correction, a significantly higher number of participants preferred the real condition ($M = 26.5 \pm 3.15$) over both AR conditions ($M_{ARc} = 3.67 \pm 1.86$, $M_{ARr} = 5.83 \pm 2.23$), $U_{R-ARc} = 0, p = .005, U_{R-ARr} = 0, p = .005$, while no significantly different number of participants was found between the AR conditions, $U = 9, p = .16$.

5 DISCUSSION

We did find significant differences between real and AR conditions but no significant difference between colocated and remote AR. These results indicate that participants were mainly impacted by the gap between the real task and its simulation in an SVS. This statement is aligned with the results obtained by Billinghurst et al. [7].

Table 3: Results of the collaboration and mutual awareness questionnaire (the significance threshold for pairwise comparisons integrates a Bonferroni correction $\alpha = .05/3 = .017$).

Question	**p-values**			
	Overall	**R-ARc**	**R-ARr**	**ARc-ARr**
Q1-My partner and I worked well together on the task	.007**	.070	.096	.74
Q2-It was easy to be aware of what my partner was doing	$<$ .001***	$<$ .001***	$<$ .001***	.81
Q3-I felt connected with my partner	.001***	.008**	$<$ .001***	.92
Q4-My partner and I communicated together well	.27	.82	.10	.20
Q5-I understood how my partner was feeling	.004**	.026	$<$ .001***	.28
Q6-My partner understood how I was feeling	.005**	.014*	.003**	.21
Q7-I was satisfied with the output of the task	$<$ .001***	$<$ .001***	.032	.12

Our results partially validate our hypotheses H1, H2, and H3. As expected, the real condition is the preferred one, which validates H4. We expected colocated AR to benefit from better interpersonal communication than remote AR. However, experimental results did not validate this hypothesis. Both AR-based conditions obtained average SUS scores around 68 with non-significant differences, which is a marginally high acceptance score slightly under good acceptance [2]. The ARr condition even obtained slightly better results than the ARc condition. This appears to be due to ARc drawbacks, like a perfectible accuracy of the SVS, disturbed hand tracking due to confusion between users' hands, or users physically interfering while moving in the SVS. We also analyzed the evolution of the results according to the order of phases performed (Fig. 9). Overall, we observe that, regardless of the order of the AR condition, swapping AR conditions provided the same evolution of the results and confirms analyses made above. Most of the time, participants who finished the experiment with the real condition did not achieve better scores than those who started with this condition, except for spatial presence. We suppose that they might have experienced fatigue due to HMD use. The second AR condition performed by participants generally led to better results, suggesting a learning effect during the whole experience. Training puzzle pieces were larger than 36-piece puzzles, making training puzzle pieces easier to manipulate, probably taking part in this learning effect during the experiment. We did not find any significant difference considering gender. Surprisingly, XR familiarity neither impacted results, indicating that our system provides enough intuitiveness and affordance to get equally used by users of any XR familiarity.

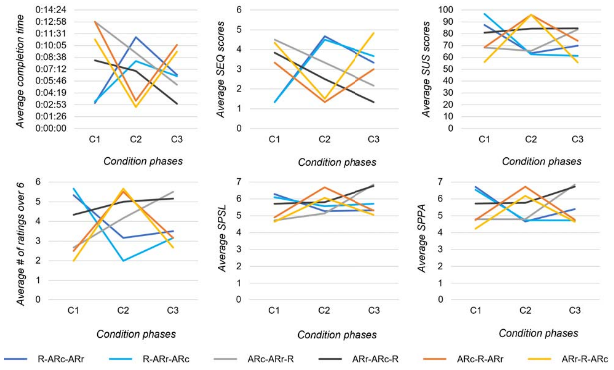


Figure 9: Evolution of the average results according to the order of phases (from left to right: completion time, SEQ scores, SUS scores, copresence scores, SPSL, and SPPA scores).

Limitations

The ARc condition caused additional drawbacks. The hands and fingers of colocated participants regularly occluded, overloading hand detection and tracking while perceiving three or four hands. OST-HMDs attribute the same perceived hands to their holder, provoking failures in hand tracking and ownership management of manipulated puzzle pieces. For example, when one hand was perceived by both OST-HMDs, participants concurrently owned the same virtual puzzle piece. In any AR-based condition, we observed that a couple of participants failed in grabbing virtual puzzle pieces by trying to respect very precisely their geometry and thickness. In such cases, they circumvented grabbing failures by pushing the puzzle pieces. These participants often expressed their frustration and tended to contribute less to the task, preferring to not interfere with their collaborator. From time to time, participants naturally aligned their eyes, their wrist, and their fingers to grab objects. Their wrist occluded their fingers from the frontal depth camera, blocking finger tracking. About half of the participants initially failed to stop grabbing virtual puzzle pieces when forgetting to separate their fingers due to the lack of tangible feedback. Puzzle pieces then kept "glued" to their fingers. Comparing the Uncanny Valley effect [29] with the struggle to interact with AR objects and with tracking failures would be useful to better identify the issues causing frustration and submissive behavior in participants. About half of the participants initially complained about the limited field of view, similar to observations made by Billinghurst et al. [7].

Lessons learnt

This experiment showed us that avatar embodiment is a secondary aspect of CAR systems. Colocated CAR does not perform better than remote CAR, despite providing optimal interpersonal communication. Bare-hand interaction techniques take precedence over avatar embodiment. We also notice that our CAR system reaches an acceptable usability, and provides enough intuitiveness and affordance to be equally used by users of any level of XR familiarity. Colocated AR requires non-ambiguous and accurate hand and finger tracking robust to multiple user presence, and a precise SVS colocation.

6 CONCLUSION

In this paper, we presented our ARPuzzle experiment, designed for collaborative virtual puzzle solving in colocated and remote AR. Our CAR system provides shared virtual spaces, verbal and nonverbal interpersonal communication between remote collaborators incarnated as full-body avatars, and intuitive interactions between collaborators and virtual objects. We experimented ARPuzzle with 36 participants by comparing real, colocated AR, and remote AR conditions. We studied performance, collaboration and awareness, spatial presence and copresence, difficulty, usability, and preference. We found the real condition to outperform both AR-based conditions. Remote and colocated AR conditions performed similarly despite different interpersonal communication efficiency.

Future work will extend task objects to synchronous mixed objects and their use during augmented operations use cases. Intuitive interactions and spatial anchoring need improvements in efficiency, accuracy, and tracking. Finally, we aim at studying multiple remote groups of colocated mixed collaborators.

REFERENCES

[1] H. Bai, P. Sasikumar, J. Yang, and M. Billinghurst. A User Study on Mixed Reality Remote Collaboration with Eye Gaze and Hand Gesture Sharing. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13. ACM, Honolulu, HI, USA, Apr. 2020. doi: 10.1145/3313831.3376550

[2] A. Bangor, P. Kortum, and J. Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *J. Usability Studies*, 4(3):114–123, 2009.

[3] C. Basdogan, C.-H. Ho, M. A. Srinivasan, and M. Slater. An experimental study on the role of touch in shared virtual environments. *ACM Transactions on Computer-Human Interaction*, 7(4):443–460, Dec. 2000. doi: 10.1145/365058.365082

[4] H. Benko, E. Ishak, and S. Feiner. Collaborative Mixed Reality Visualization of an Archaeological Excavation. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 132–140. IEEE, Arlington, VA, USA, 2004. doi: 10.1109/ISMAR.2004.23

[5] M. Billinghurst, D. Belcher, A. Gupta, and K. Kiyokawa. Communication Behaviors in Colocated Collaborative AR Interfaces. *International Journal of Human-Computer Interaction*, 16(3):395–423, Dec. 2003. doi: 10.1207/S15327590IJHC1603_2

[6] M. Billinghurst and H. Kato. Collaborative augmented reality. *Communications of the ACM*, 45(7):64–70, July 2002. doi: 10.1145/514236.514265

[7] M. Billinghurst, H. Kato, K. Kiyokawa, D. Belcher, and I. Poupyrev. Experiments with Face-To-Face Collaborative AR Interfaces. *Virtual Reality*, 6(3):107–121, Oct. 2002. doi: 10.1007/s100550200012

[8] M. Billinghurst, I. Poupyrev, H. Kato, and R. May. Mixing Realities in Shared Space: An Augmented Reality Interface for Collaborative Computing. In *2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings. Latest Advances in the Fast Changing World of Multimedia (Cat. No.00TH8532)*, vol. 3, pp. 1641–1644, 2000. doi: 10.1109/ICME.2000.871085

[9] J. Brooke. SUS: A quick and dirty usability scale. *Usability Eval. Ind.*, 189, Nov. 1995.

[10] A. Chellali, I. Milleville-Pennel, and C. Dumas. Elaboration of a common frame of reference in collaborative virtual environments. In *Proceedings of the 15th European conference on Cognitive ergonomics the ergonomics of cool interaction - ECCE '08*, p. 1. ACM Press, Funchal, Portugal, 2008. doi: 10.1145/1473018.1473045

[11] L. Chen, Y. Liu, Y. Li, L. Yu, B. Gao, M. Caon, Y. Yue, and H.-N. Liang. Effect of Visual Cues on Pointing Tasks in Co-located Augmented Reality Collaboration. In *Symposium on Spatial User Interaction*, pp. 1–12. ACM, Virtual Event USA, Nov. 2021. doi: 10.1145/3485279.3485297

[12] F. D. Davis. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3):319, Sept. 1989. doi: 10.2307/249008

[13] A. Dey, H. Chen, C. Zhuang, M. Billinghurst, and R. W. Lindeman. Effects of Sharing Real-Time Multi-Sensory Heart Rate Feedback in Different Immersive Collaborative Virtual Environments. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 165–173, 2018.

[14] P. Disztinger, S. Schlögl, and A. Groth. Technology Acceptance of Virtual Reality for Travel Planning. In R. Schegg and B. Stangl, eds., *Information and Communication Technologies in Tourism 2017*, pp. 255–268. Springer International Publishing, Cham, 2017. doi: 10.1007/978-3-319-51168-9_19

[15] B. Ens, J. Lanir, A. Tang, S. Bateman, G. Lee, T. Piumsomboon, and M. Billinghurst. Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies*, 131:81–98, Nov. 2019. doi: 10.1016/j.ijhcs.2019.05.011

[16] S. R. Fussell, R. E. Kraut, and J. Siegel. Coordination of communication: effects of shared visual context on collaborative work. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work - CSCW '00*, pp. 21–30. ACM Press, Philadelphia, PA, USA, 2000. doi: 10.1145/358916.358947

[17] R. Grasset, P. Lamb, and M. Billinghurst. Evaluation of mixed-space collaboration. In *Proceedings - Fourth IEEE and ACM International Symposium on Symposium on Mixed and Augmented Reality, ISMAR 2005*, vol. 2005, pp. 90– 99, 2005. doi: 10.1109/ISMAR.2005.30

[18] K. Gupta, G. A. Lee, and M. Billinghurst. Do You See What I See? The Effect of Gaze Tracking on Task Space Remote Collaboration. *IEEE Transactions on Visualization and Computer Graphics*, 22(11):2413–2422, Nov. 2016. doi: 10.1109/TVCG.2016.2593778

[19] P. Gurevich, J. Lanir, and B. Cohen. Design and Implementation of TeleAdvisor: a Projection-Based Augmented Reality System for Remote Collaboration. *Computer Supported Cooperative Work (CSCW)*, 24(6):527–562, Dec. 2015. doi: 10.1007/s10606-015-9232-7

[20] C. Harms and F. Biocca. Internal Consistency and Reliability of the Networked Minds Measure of Social Presence. In *Seventh Annual International Workshop: Presence*, p. 7, 2004.

[21] H. Ishii, M. Kobayashi, and J. Grudin. Integration of interpersonal space and shared workspace: ClearBoard design and experiments. *ACM Transactions on Information Systems (TOIS)*, 11(4):349–375, 1993. Publisher: ACM New York, NY, USA.

[22] K. Kim, L. Bölling, S. Haesler, J. Bailenson, G. Bruder, and G. F. Welch. Does a Digital Assistant Need a Body? The Influence of Visual Embodiment and Social Behavior on the Perception of Intelligent Virtual Agents in AR. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 105–114. Munich, 2018. doi: 10.1109/ISMAR.2018.00039

[23] S. Kim, G. Lee, W. Huang, H. Kim, W. Woo, and M. Billinghurst. Evaluating the Combination of Visual Communication Cues for HMD-based Mixed Reality Remote Collaboration. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–13. ACM, Glasgow, Scotland, UK, May 2019. doi: 10.1145/3290605.3300403

[24] D. Kirk and D. Stanton Fraser. Comparing remote gesture technologies for supporting collaborative physical tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1191–1200. ACM, Montréal, Québec, Canada, Apr. 2006. doi: 10.1145/1124772.1124951

[25] M. Kritzler, M. Funk, F. Michahelles, and W. Rohde. The Virtual Twin: Controlling Smart Factories Using a Spatially-correct Augmented Reality Representation. In *Proceedings of the Seventh International Conference on the Internet of Things*, IoT '17, pp. 38:1–38:2. ACM, New York, NY, USA, 2017. doi: 10.1145/3131542.3140274

[26] P. Ladwig and C. Geiger. A Literature Review on Collaboration in Mixed Reality. In M. E. Auer and R. Langmann, eds., *Smart Industry & Smart Education*, vol. 47, pp. 591–600. Springer International Publishing, Cham, 2019. Series Title: Lecture Notes in Networks and Systems. doi: 10.1007/978-3-319-95678-7_65

[27] M. Lombard, T. Bolmarcich, and L. Weinstein. Measuring Presence: The Temple Presence Inventory. In *Proceeding of Presence 2009 : The 12th International Workshop on Presence*, Jan. 2009.

[28] K. Masai, K. Kunze, M. Sugimoto, and M. Billinghurst. Empathy Glasses. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 1257–1263. ACM, San Jose, CA, USA, May 2016. doi: 10.1145/2851581.2892370

[29] M. Mori, K. MacDorman, and N. Kageki. The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine*, 19(2):98–100, June 2012. doi: 10.1109/MRA.2012.2192811

[30] O. Oda, C. Elvezio, M. Sukan, S. Feiner, and B. Tversky. Virtual Replicas for Remote Assistance in Virtual and Augmented Reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pp. 405–415. ACM, Charlotte, NC, USA, Nov. 2015. doi: 10.1145/2807442.2807497

[31] T. Piumsomboon, A. Day, B. Ens, Y. Lee, G. Lee, and M. Billinghurst. Exploring Enhancements for Remote Mixed Reality Collaboration. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*, SA '17. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3132787.3139200

[32] T. Piumsomboon, A. Dey, B. Ens, G. Lee, and M. Billinghurst. The Effects of Sharing Awareness Cues in Collaborative Mixed Reality. *Frontiers in Robotics and AI*, 6:5, Feb. 2019. doi: 10.3389/frobt.2019.00005

[33] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman,

B. H. Thomas, and M. Billinghurst. Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–13. ACM, Montreal, Québec, Canada, Apr. 2018. doi: 10.1145/3173574.3173620

[34] J. Sauro and J. S. Dumas. Comparison of three one-question, post-task usability questionnaires. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1599–1608. ACM, Boston MA USA, Apr. 2009. doi: 10.1145/1518701.1518946

[35] W. Schafer and D. Bowman. Evaluating the effects of frame of reference on spatial collaboration using desktop collaborative virtual environments. *Virtual Reality*, 7(3-4), June 2004. doi: 10.1007/s10055-004-0123-3

[36] R. Schoenfelder and D. Schmalstieg. Augmented Reality for Industrial Building Acceptance. In *2008 IEEE Virtual Reality Conference*, pp. 83–90. IEEE, Reno, NV, USA, 2008. doi: 10.1109/VR.2008.4480755

[37] 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

[38] M. Slater, A. Sadagic, M. Usoh, and R. Schroeder. Small-Group Behavior in a Virtual and Real Environment: A Comparative Study. *Presence: Teleoperators and Virtual Environments*, 9(1):37–51, Feb. 2000. doi: 10.1162/105474600566600

[39] H. J. Smith and M. Neff. Communication Behavior in Embodied Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12. ACM, Montreal QC Canada, Apr. 2018. doi: 10.1145/3173574.3173863

[40] T. Teo, L. Lawrence, G. A. Lee, M. Billinghurst, and M. Adcock. Mixed Reality Remote Collaboration Combining 360 Video and 3D Reconstruction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–14. ACM, Glasgow, Scotland, UK, May 2019. doi: 10.1145/3290605.3300431

[41] P. Vorderer, W. Wirth, F. Gouveia, F. Biocca, T. Saari, L. Jäncke, S. Böcking, H. Schramm, A. Gysbers, T. Hartmann, C. Klimmt, J. Laarni, N. Ravaja, A. Sacau, T. Baumgartner, and P. Jäncke. MEC spatial presence questionnaire (MEC-SPQ): Short documentation and instructions for application. *Report to the European Community, Project Presence: MEC (IST-2001-37661)*, June 2004.

[42] A. Vovk, F. Wild, W. Guest, and T. Kuula. Simulator Sickness in Augmented Reality Training Using the Microsoft HoloLens. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–9. ACM, Montreal, Québec, Canada, Apr. 2018. doi: 10.1145/3173574.3173783

[43] I. Vrellis, M. Delimitros, P. Chalki, P. Gaintatzis, I. Bellou, and T. A. Mikropoulos. Seeing the unseen: user experience and technology acceptance in Augmented Reality science literacy. In *2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT)*, pp. 333–337. IEEE, Tartu, Estonia, July 2020. doi: 10.1109/ICALT49669.2020.00107

[44] K. Waldow, A. Fuhrmann, and S. M. Grünvogel. Investigating the Effect of Embodied Visualization in Remote Collaborative Augmented Reality. In *Virtual Reality and Augmented Reality: 16th EuroVR International Conference, EuroVR 2019, Tallinn, Estonia, October 23–25, 2019, Proceedings*, pp. 246–262. Springer-Verlag, Berlin, Heidelberg, 2019. doi: 10.1007/978-3-030-31908-3_15

[45] X. Wang and P. S. Dunston. Comparative effectiveness of Mixed Reality based virtual environments in collaborative design. *2009 IEEE International Conference on Systems, Man and Cybernetics*, pp. 3569–3574, 2009.

[46] Y. Wang, Z. Su, N. Zhang, R. Xing, D. Liu, T. H. Luan, and X. Shen. A Survey on Metaverse: Fundamentals, Security, and Privacy. *IEEE Communications Surveys & Tutorials*, p. 1, 2022. doi: 10.1109/COMST.2022.3202047

[47] F. Zijlstra and L. Doorn. The Construction of a Scale to Measure Perceived Effort. *Department of Philosophy and Social Sciences*, Jan. 1985.