Evaluating Shared Surfaces for Co-Located Mixed-Presence Collaboration

Christian Mai
LMU Munich, Germany
Christian.Mai@ifi.lmu.de

Sarah Aragon Bartsch
LMU Munich, Germany
Sarah.Aragon.Bartsch@ifi.lmu.de

Lea Rieger
LMU Munich, Germany
Lea.Rieger@campus.lmu.de

Figure 1. Collaborative task used in the study: (a) *Real_World*: shared surface without HMD (baseline), (b) *No_Avatar*: mixed-presence collaboration with a shared surface only and (c) *Avatar*: mixed-presence scenario with a rendered point-cloud avatar of the real world instructor

ABSTRACT
When wearing a head-mounted display (HMD) in everyday environments, interactions with real world bystanders often fail due to the visual barrier. As a result, the HMD user takes off the headset, which ends the virtual reality (VR) experience. We address this problem by providing a shared surface with the same content for both users, which is located at the same physical position in the real and the virtual world. In a between-subject user study ($N = 40$), we investigate the effects of a shared surface for short-term collaboration in co-located mixed-presence scenarios. We compare (a) real-world collaboration, (b) having a shared surface only and (c) combining the shared surface with an avatar representation of the real world user in VR. We could show that shared surfaces are helpful for mixed-presence collaboration. Adding an avatar in VR improves performance measures such as task-completion time, error rate and number of clarifying questions. To support future work in this field, we finally propose design implications and research directions.

CCS Concepts
•**Human-centered computing** → **Empirical studies in collaborative and social computing;**

Author Keywords
Head-Mounted Displays; Collaboration; Mixed-Presence

INTRODUCTION
The growing distribution of affordable head-mounted displays (HMDs) introduces the technology of virtual reality (VR) to everyday environments, e.g. living rooms or shops[1]. In contrast to lab environments, in which the rooms are specially designed for VR experiences, these everyday environments are prone to interference or interruptions. These interruptions, to name but a few, can be caused by physical objects in the walking area of the VR user, noise originating from the real world or other persons being around. In this work, we focus on interruptions caused by real world bystanders, trying to collaborate with the HMD user.

We call the situation of people being physically in the same place but visually in different worlds **co-located mixed presence**. The HMD user becomes part of the social community when wearing the HMD in an everyday environment. Therefore, many situations occur in which mixed-presence collaboration is needed, as for example in the following scenario:

Scenario – A mother takes part in a business VR tele-presence conference from home. Her five year old son wants to show her a picture of his favorite dog. The kid will not accept a denial of attention, but the mother has professional obligations towards her colleagues to attend the meeting.

Similar to our scenario, McGill et al. [7] could show that there is a need for HMD users to communicate with the world surrounding them. In most cases, this communication fails due to the visual barrier. As a result, the user has to take off the HMD, which is not only perceived as very annoying but also hinders her from carrying out the VR task [7]. Methods to overcome this problem are augmenting the real world into the virtual [7], sharing the physical space for playful asymmetric collaboration [3] or providing insight into the HMD user's environment on a head mounted screen that simultaneously serves as an input and output device for the real world user [4]. In contrast to previous work, we suggest to avoid mixing the two presence states and rather keep them separated. From research on presence in VR, we derive that VR users will

MUM '18, November 25–28, 2018, Cairo, Egypt

ISBN 978-1-4503-6594-9/18/11...$15.00

DOI: https://doi.org/10.1145/3282894.3282910

[1] https://www.audi-mediacenter.com/en/audi-at-the-ces-2016-5294/the-audi-vr-experience-5304, accessed 30/08/18

benefit from this, as they are not reminded about the real world, keeping the focus on the virtual stimuli.

We propose to use a shared surface as a mediation between the virtual and the real world (see figure 1). For the real world user, the shared surface can be any digital device providing a screen and being able to transmit or receive pictures, e.g. a tablet, a micro projector[2] or a tabletop display. The shared surface is rendered in VR as a “digital twin”, located at the same physical position as in the real world. The concept of a shared surface is similar to the idea of using tabletop displays for tele-presence tasks [11]. However, in our co-located scenario, the real world user can see the HMD user and they are able to touch each other and shared objects. This might affect the collaboration leading to design opportunities different from tabletops.

In this work, we address the following **research questions**, focusing on **short-term collaboration in co-located mixed-presence scenarios**:

1. What is the effect of having a shared surface on user behavior, task performance and user experience?
2. Collaboration in this scenario is asymmetric, as the HMD user cannot see the real world user. Is an additional augmentation of the bystander needed, as it is known from the work on tabletop displays for tele-presence?

To answer our research questions, we conducted a between-subject user study ($N = 40$). We compared the following conditions: (a) *Real_World*: both collaborators in the real world (baseline), (b) *No_Avatar*: shared surface without rendering the real world user in VR and (c) *Avatar*: shared surface with a point cloud representation of the real world bystander in the VR (see figure 1).

Altogether our **contributions** are:

- Insights on the usage of a shared surface in co-located mixed-presence scenarios
- Design implications for co-located mixed-presence collaboration using a shared surface

BACKGROUND AND RELATED WORK

Shared surfaces are well known to support collaborative tasks in real world [2, 13] as well as in tele-presence scenarios [8]. Both of these situations are called “symmetric”, as the visual information and the possibility to interact with the virtual content are the same for both collaborators. In our work, we deal with an asymmetric situation, in which the real world collaborator can only see the HMD user’s body, but not their eyes. The HMD user, on the other hand, is mentally in a remote situation and cannot see the real world user. To maintain awareness for each other, Tang et al. introduced shadow techniques to present the abstract arm position of the collaborators on the remote screen [11]. In a follow-up study, they used a video overlay of a real arm [12]. They found that the more realistic the virtual representation, the better the possibility to perform and understand directed gestures.

[2] https://www.sonymobile.com/global-en/products/smart-products/xperia-touch, accessed 30/08/18

The work on co-located mixed-presence collaboration is highly diverse. Gugenheimer et al. [3] present a system that augments the collaboration in both directions in a gaming context. The real world user gets an indication about the HMD user’s VR experience by a projection on the floor. In contrast, we propose to create a shared frame of reference for both users at the same physical position in the real and the virtual world. Furthermore, we want to analyze short-term collaboration in a different context and therefore use a standardized task and quantitative as well as qualitative measures.

Visual integration of the real world user in VR is often achieved by using green screen technology [7, 15]. The mixture of both worlds generates a more symmetric interaction, in which the real world and the VR user have the same information about each other. In general, this additional information proofed beneficial for the collaboration. However, green screen technology is not part of everyday environments and probably not wanted in living rooms or shops. More sophisticated tracking technologies, e.g., Kinect, are neither well-suited for ad hoc collaboration in these contexts. There are not only technical challenges when rendering real world objects in VR, but also some drawbacks for the HMD user. Users might feel frightened when an avatar suddenly appears, it can lead to illogical situations and have a negative impact on the feeling of being present in VR [7]. In this work, we want to find out whether a shared surface, that can more easily be rendered in VR, can bridge the gap between both worlds by itself or if a representation of the real world user in VR is still needed.

USER STUDY

The main goal of our study was to analyze the applicability of a shared surface for short-term co-located mixed-presence collaboration. We therefore collected data on user behaviour, performance and user experience in three different conditions. The structure of the user study will be explained in detail in the following.

Design

We used a paper prototype on a table as a shared surface, which was located at the same physical position in both the virtual and the real world. To achieve generalizability to a variety of tasks, we chose a generic sandwich-assembly-task as proposed by Andrist et al. [1]. Pairs of two participants collaboratively had to build a total amount of eight sandwiches made of different ingredients, presented on the shared surface. The real world user was the instructor and the HMD user was the worker. The task fosters gestural interaction as some ingredients have high similarity in shape and/or color.

We conducted the user study with three different conditions: (a) *Real_World*: no visual restrictions, (b) *No_Avatar*: shared-surface only and (c) *Avatar*: shared surface and VR representation of the real world user (see figure 1).

A between-subject design was chosen to avoid carry-over effects from condition to condition. To monitor user behaviour, the participants were filmed from two different angles throughout the study. In addition, we measured the amount of time it took the participants to assemble each sandwich, the number of errors and the number of clarifying questions the worker asked. An error occurred when the worker selected a wrong ingredient even if the choice was corrected afterwards. We defined clarifying questions as every question the worker asked to ensure the ingredient they were about to select was the right one, e.g., "This one?". For a further analysis, we used the igroup presence questionnaire (IPQ) [10] and the user experience questionnaire (UEQ) [6].

Participants

40 participants were recruited (21 male and 19 female) aged from 18 to 53. Most of them were students with a technical background. The participants had to perform the study in pairs of two. Nine pairs did not know each other beforehand. Six pairs took part in the *Real_World* condition, seven pairs respectively in condition *No_Avatar* and *Avatar*. The roles of instructor and worker were assigned randomly. Subjects with known simulator sickness were assigned to be the instructor. There was no training for the participants with the system before the study. Participants were compensated with five Euros cash.

Apparatus

The paper prototype was presented on a table with a Microsoft Kinect v2 mounted above. A HTC Vive was used as an HMD. To stream the point cloud data to Unity 5.5, the point cloud library (PCL) was used [9]. All points above and behind the table were rendered in VR (see figure 1). The visualization was detailed enough to distinguish single fingers and subtle movements. The set and arrangement of ingredients changed with every sandwich to build. When the worker used the HMD, they saw the tabletop display and its ingredients in original size at the same position as the real world instructor.

Procedure

We welcomed the participants, explained the study procedure and the task and let them sign a consent form. The participants did not know about the other conditions in advance. They had to stand on each side of the table facing each other. An instructional sheet with ten different ingredients was handed to the instructor. In each round, 18 out of 28 ingredients were presented. The instructor gave directives to the worker on how to assemble the sandwich. In all three conditions, the participants were told that any form of communication or action was allowed. The worker's job was to build the sandwich by tapping on the correct ingredients in the right order. After performing the task, both participants were interviewed and the questionnaires were filled out.

RESULTS

All 20 pairs were able to build all eight sandwiches. We conducted a one-way between-subjects ANOVA to compare the effect of the collaboration condition on the respective measured variable for the conditions *Real_World*, *No_Avatar* and *Avatar* on a 5% alpha level. All were tested for homogeneity in variance with a Levene-Test, without any salience. For significant outcomes of the ANOVA, we used the Tukey post hoc test to compare the conditions for significant differences.

	Condition	Mean	Std. Deviation
Time	Real_World	15s	1s
	No_Avatar	35s	5s
	Avatar	26s	6s
Errors	Real_World	2.7	1.5
	No_Avatar	3.3	1.4
	Avatar	2.1	2
Clarifying Questions	Real_World	0.2	0.5
	No_Avatar	1.7	1.4
	Avatar	1	1.2

Table 1. Means and Standard Deviations for the time to complete a task, the error rate and the number of questions to clarify the situation

Outliers were identified by a BoxPlot, which lead to the removal of one subject pair in the *Real_World* baseline and one in the *No_Avatar* condition.

Task Performance

We measured task performance with the variables "task completion time", "error rate" and "number of clarifying questions". The results are shown in table 1. There was a significant effect on the completion time ($F(2,17) = 13.643; p = .001$) and error rate ($F(2,17) = 7,983; p = .001$). The post-hoc comparison showed a highly significant difference between *Real_World* and *No_Avatar* ($p = .001$), *Real_World* and *Avatar* ($p = .01$) and *No_Avatar* and *Avatar* ($p = .01$). For the error-rate, there was a significant difference between *No_Avatar* and *Avatar* ($p = .01$). As we expected the users to adapt to the given system, we analyzed the data once for all eight built sandwiches and once for the last six sandwiches. In general, the users got faster, but the relative comparison between the conditions stayed the same. The mean completion-time decreases to $M = 14.9s$ ($SD = 0.1s$) for the *Real_World* condition, to $M = 33.6s$ ($SD = 5.2s$) for *No_Avatar* and $M = 23.9s$ ($SD = 4.6s$) for *Avatar* when looking at the last 6 sandwiches only.

Presence

The IPQ presence questionnaire [10] was analyzed with a t-test between the conditions *No_Avatar* and *Avatar*. It is not applicable to the *Real_World* condition [14]. One rating was excluded as an outlier for the *Avatar* condition. The IPQ consists of the three independent subscales "spatial presence" (*SP*), "involvement" (*INV*) and "experienced realism" (*REAL*) and a general presence item (*G*). On a significance level of 5%, no difference could be found between the *No_Avatar* and *Avatar* condition. The means of *No_Avatar* are $G = 3.7$ ($SD = .9$), $SP = 4.3$ ($SD = .5$), $INV = 2.7$ ($SD = .9$), $REAL = 2.5$ ($SD = .4$). The *Avatar* condition shows the means $G = 4(SD = 1.4)$, $SP = 4$ ($SD = .2$), $INV = 2.7$ ($SD = 1.4$), $REAL = 2.7$ ($SD = 1.5$).

User Experience

The user experience based on the UEQ rating scale was at least good in all hedonic and pragmatic categories for all conditions, based on Laugwitz's suggestion to interpret the results [5]. No significant difference in the hedonic quality between the conditions could be shown. Significant difference could be found in the dependability (D) ($F(2,17) = 4.97, p = .02$) (see figure 2).

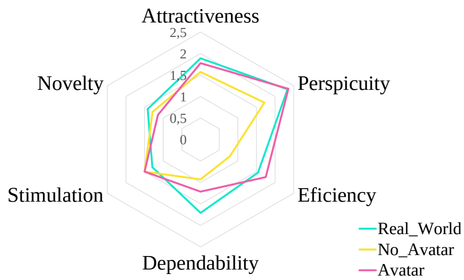

Figure 2. The results from the UEQ.

The post hoc comparison showed a significant difference between the *Real_World* ($M = 1.5, SD = .8$) and the *No_Avatar* ($M = .8, SD = .6$) condition.

However, for the pragmatic quality, in particular the means of perspicuity (P), efficiency (E) and dependability (D), the condition *No_Avatar* ($P = 1.7; E = .8; D = .9$) is worse than the conditions *Real_World* ($P = 2.3; E = 1.5; D = 1.7$) and *Avatar* ($P = 2.4; E = 1.8; D = 1.2$).

Interviews

The interviews gave some interesting insights into the users' subjective experience. All workers reported for the *Real_World* condition that they relied on the gestures and did not look into the instructors face. Speech was used for clarification in the *Real_World* condition.

For the *No_Avatar* condition, users completely relied on oral instructions. The collaborators created different strategies to get faster. In particular, they used oral codes to describe the position of an ingredient. Examples are a chain of parameters (*the red - round thing - near you*), creating codes for the objects (*dark or pale mushroom*) or coding the lines and rows with numbers. Two subjects reported experiencing the instructor as speaking from "the off" or from "another world". This was perceived as distracting from the virtual world. One other reported to imagine the relative location of the instructor by the direction of the voice.

Statements on the *Avatar* condition showed many similarities to the *Real_World* condition. Two participants reported that they felt a connection to the point cloud avatar like to a human being. All subjects reported a strong focus on the hands of the instructor's avatar. Three pairs used oral instructions to overcome the delay in task completion due to latency. They used oral commands to search for the possible target at which the hand of the instructor would appear one second later. Latency in the point cloud rendering was reported by all HMD users and half of the real world users. Two of the HMD users were not able to cope with the latency until the end.

Limitations

Although our study was carefully designed, we want to point out some limitations in our results that should be addressed in future research. Our system had a latency between $0.7s$ and $1s$ in the *Avatar* condition, underlining the fact that real-time object augmentation is still complex. Furthermore, our results are limited to a small sample size and a simple paper prototype, as we only wanted to gain first insights on the applicability of a shared surface for co-located mixed-presence collaboration. Results could also be different for other task-types.

DISCUSSION AND IMPLICATIONS

A Shared Surface is Beneficial for the Users – With the help of a shared surface, we could enable mixed-presence collaboration and prevent users from taking off the VR headset. The concept provided good user experience and efficiency in all conditions. In the sandwich-assembly task, adding an avatar improved the completion time about 25% and decreased the error rate by 30%. We propose to limit the concept of shared surfaces to comparably simple and short tasks, so that the users can easily adapt. The collaborators in the *No_Avatar* condition overcame the restriction that the HMD cannot see the real world user by giving oral descriptions. The voice coming from the real world was reported as distracting the attention from the VR. To avoid this, we propose considering an integration of the instructor's voice into the virtual world by using a microphone.

Give Visual Feedback to the HMD User – The visibility of the real world user for the HMD user almost halves the need for clarifying questions. In addition, providing an avatar leads to a user experience similar to the *Real_World* and significantly better than the *No_Avatar* condition. The live-tracking of a person to present an avatar in VR is technically challenging in everyday environments. Yet users reported in the *Real_World* and the *Avatar* condition to primarily focus on the hands of the instructor. Therefore, a visualization of the touch-points on the shared surface might be enough to improve collaboration.

Keep in Mind Possible Behavior Changes – We found that the system influences the users' behavior during the collaboration. Latency in the *Avatar* condition lead to increased usage of oral instructions. We could observe the biggest behaviour changes in the *No_Avatar* condition, as people needed to overcome the visual restrictions. They started to use codes over time to reduce verbal commands. This means, that the users' performance is not only system-dependent but also influenced by different types of communication strategies. We assume that people need to adapt their strategy for every new task, which leads to an increase of effort and mental demand. This should be kept in mind when designing for mixed-presence scenarios.

CONCLUSION & FUTURE WORK

We could show that a shared surface enables collaboration between the real and the virtual world. The use of an additional avatar of the real world user improved performance measures. However, real-time augmentation is still complex to be deployed in everyday environments. As we found promising results in both conditions *No_Avatar* and *Avatar*, future work should have a closer look on the design of these conditions. For example, the tracking of the real world user could be simplified to the touches on the shared surface. As our study was limited to a paper prototype and a generic task, future work should focus on a wider range of devices and use-cases to explore the possibilities of a shared surface in-the-wild. In doing so, the effort to use such a system should be kept at a minimum to enable ad hoc collaboration.

REFERENCES

1. Sean Andrist, Michael Gleicher, and Bilge Mutlu. 2017. Looking Coordinated: Bidirectional Gaze Mechanisms for Collaborative Interaction with Virtual Characters. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 2571–2582. DOI: http://dx.doi.org/10.1145/3025453.3026033

2. Jacqueline Brodie and Mark Perry. 2002. Mobile Collaboration at the Tabletop in Public Spaces. In *ACM CSCW 2002 Workshop on Co-Located Tabletop Collaboration: Technologies and Directions (CSCW 2002)*.

3. Jan Gugenheimer, Evgeny Stemasov, Julian Frommel, and Enrico Rukzio. 2017a. ShareVR: Enabling Co-Located Experiences for Virtual Reality Between HMD and Non-HMD Users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 4021–4033. DOI: http://dx.doi.org/10.1145/3025453.3025683

4. Jan Gugenheimer, Evgeny Stemasov, Harpreet Sareen, and Enrico Rukzio. 2017b. FaceDisplay: Enabling Multi-User Interaction for Mobile Virtual Reality. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, New York, NY, USA, 369–372. DOI: http://dx.doi.org/10.1145/3027063.3052962

5. Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and Evaluation of a User Experience Questionnaire. In *HCI and Usability for Education and Work*, Andreas Holzinger (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 63–76.

6. Bettina Laugwitz, Martin Schrepp, and Theo Held. 2006. Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten. In *Mensch und Computer 2006: Mensch und Computer im Strukturwandel*, Andreas M. Heinecke and Hansjürgen Paul (Eds.). Oldenbourg Verlag, München, 125–134. DOI:http://dx.doi.org/10.1524/9783486841749.125

7. Mark McGill, Daniel Boland, Roderick Murray-Smith, and Stephen Brewster. 2015. A Dose of Reality: Overcoming Usability Challenges in VR Head-Mounted Displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2143–2152. DOI: http://dx.doi.org/10.1145/2702123.2702382

8. Alexandre Pauchet, Francois Coldefy, L. Lefebvre, Stphane Louis-Dit-Picard, Louis Perron, A. Bouguet, Michel Collobert, Joris Guerin, and D. Corvaisier. 2007. TableTops: Worthwhile Experiences of Collocated and Remote Collaboration. In *Horizontal Interactive Human-Computer Systems, 2007. TABLETOP '07. Second Annual IEEE International Workshop on*. 27–34. DOI:http://dx.doi.org/10.1109/TABLETOP.2007.13

9. Radu Rusu and Steve Cousins. 2011. 3D is here: Point Cloud Library (PCL). In *2011 IEEE International Conference on Robotics and Automation (ICRA)*. 1–4. DOI:http://dx.doi.org/10.1109/ICRA.2011.5980567

10. Thomas Schubert. 2003. The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Zeitschrift für Medienpsychologie* 15, 2 (2003), 69–71. DOI: http://dx.doi.org/10.1026//1617-6383.15.2.69

11. Anthony Tang, Michael Boyle, and Saul Greenberg. 2004. Display and Presence Disparity in Mixed Presence Groupware. In *Proceedings of the Fifth Conference on Australasian User Interface - Volume 28 (AUIC '04)*. Australian Computer Society, Inc, Darlinghurst, Australia, Australia, 73–82. http://dl.acm.org/citation.cfm?id=976310.976320

12. Anthony Tang, Carman Neustaedter, and Saul Greenberg. 2007. VideoArms: Embodiments for Mixed Presence Groupware. In *People and Computers XX — Engage: Proceedings of HCI 2006*, Nick Bryan-Kinns, Ann Blanford, Paul Curzon, and Laurence Nigay (Eds.). Springer London, London, 85–102. DOI: http://dx.doi.org/10.1007/978-1-84628-664-3{_}8

13. John Tang. 1991. Findings from Observational Studies of Collaborative Work. *Int. J. Man-Mach. Stud.* 34, 2 (1991), 143–160. DOI: http://dx.doi.org/10.1016/0020-7373(91)90039-A

14. Martin Usoh, Ernest Catena, Sima Arman, and Mel Slater. 2000. Using Presence Questionnaires in Reality. *Presence: Teleoper. Virtual Environ.* 9, 5 (Oct. 2000), 497–503. DOI: http://dx.doi.org/10.1162/105474600566989

15. Valve. 2016. Virtual Reality - SteamVR featuring the HTC Vive. (2016). https://www.youtube.com/watch?v=qYfNzhLXYGc