Remote Cross-platform Instructions between MR and VR

Mathieu Lutfallah
Innovation Center Virtual Reality
ETH Zurich
Zurich, Switzerland
mlutfallah@ethz.ch

Tamara Gini
Innovation Center Virtual Reality
ETH Zurich
Zurich, Switzerland
tgini@ethz.ch

Christian Hirt
Innovation Center Virtual Reality
ETH Zurich
Zurich, Switzerland
hirtc@ethz.ch

Kordian Caplazi
Rimon Technologies
Zurich, Switzerland
kordian@rimon-ar.ch

Andreas Kunz
Innovation Center Virtual Reality
ETH Zurich
Zurich, Switzerland
kunz@ethz.ch

***Abstract*—This paper introduces a collaboration setting where an expert in VR gives instructions to a technician using MR being in front of the real object. For this study, we develop a system in which users communicate while building a specific structure using Lego Duplo blocks. Here, only the expert knows the final configuration and instructs the technician equipped with an MR headset about the assembly. Four communication methods are evaluated: voice chat, avatars, the possibility of moving virtual blocks similar to the real ones, and an arrow functionality. This work proves the usefulness of such a VR/MR instruction system and provides insights into the advantages of each proposed communication method.**

***Index Terms*—Human Computer Interaction, Virtual Reality, Mixed Reality**

I. Introduction

In the last decades, the frequency at which new machine tools are released has increased significantly. Moreover, machines are tailored to the customers' individual needs and each individual machine embraces broader functionalities. At this pace, it becomes impossible for technicians to build and retain expertise on all individual machines on the market. For supporting the technicians, experts cannot travel the world to provide that tailored knowledge for the specific product to the technicians. A more feasible approach to solving the issue at hand is to use remote instruction sharing. In the scenario described above, an expert can instruct technicians on-site while being at a different location. They share and exchange information through a common virtual space, which is entered individually, with both partners using different technologies. Whereas the remote expert uses Virtual Reality (VR), the technicians on-site use Mixed Reality (MR), in which they see virtual instructions generated by the remote expert. Relying on MR technology, these instructions can be conveniently projected to reality. Such a VR/MR system allows for straightforward three-dimensional instructions and easier access to spatial information compared to a 2D layout plan, written instructions, or simple verbal guidance.

A. Related Work

Applications allowing multiple users to access a shared virtual space over the internet from identical or varying immersive platforms, such as VR and PC, have received substantial attention in recent years. This interest, primarily driven by the pandemic, pre-dates the pandemic itself [1]. It was demonstrated that there is growing economic interest in integrating remote expert support for task resolution. Early work mainly involved the transfer of camera images alongside users' hands, as exemplified by ReMoTe [2]. While these systems required intricate technical setups, the increasing availability of mobile devices gave rise to collaboration systems like Semarbeta [3]. However, these remote collaboration systems relied on transferring video content together with pointing gestures.

In recent years, use cases for MR applications in industry and manufacturing sectors have garnered more attention [4], and the discourse surrounding the interoperability and immersion levels of a "Metaverse" has intensified [5]. A comprehensive review of various remote systems is presented in [6], [7]. Additionally, a taxonomy of collaboration system components is presented, classifying systems into "Environment", "Avatars", and "Interaction" categories. This work will focus primarily on VR-MR collaboration, a topic already explored by many researchers. Several studies have found that MR and VR collaboration create a better sense of co-presence than traditional video conferencing [8], [9].

Jones et al. [10] investigated two collaborative systems: the first one allows a user to interact with a remote individual through a 2D video call and control a robot equipped with a screen displaying the other person, while the second system provides the expert in VR with a first-person view from the robot, and the other participant uses MR to visualize a hologram of the expert using the Hololens. Gasques [11] researched a collaboration between medical experts and trainees utilizing a complex system allowing a reconstruction of the human body in VR. Other researchers focused on investigating

2642-3596/23/$31.00 ©2023 IEEE
DOI 10.1109/CW58918.2023.00045

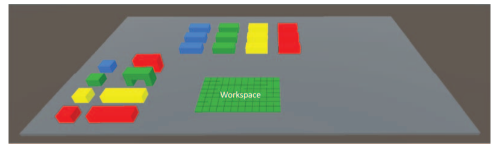
the impact of gaze and avatars on VR-MR collaboration for improved natural visual communication cues [12], [13]. However, an investigation into the efficiency of the basic communication tools in such systems is currently lacking.

II. System Description

Our research aims to present communication tools that can help enable cross-platform MR applications. We investigate how instructions can be effectively transferred between an expert and a trainee, what tools can be made available and how well they work in a practical scenario. A VR/MR communication framework is implemented, where the information how to solve a given task is unevenly distributed. The VR user is considered the expert and is provided with the solution to a Lego assembly task, while the MR user, acting as the technician, builds the Lego configuration in reality. Since the VR user has the complete information about how to solve the task, the MR user receives instructions from the VR user. This setup was chosen because it can be widely applied since the technician needs to see and manipulate the real object. For passing the instructions of the given task, the users can use four different communication tools:

- A bidirectional voice chat
- Possibility of moving the virtual objects by both users
- An arrow functionality
- Non-verbal communication and pointing through the avatars

The main goal is to determine which communication tools work best. The task to solve should require no prior knowledge, and the individual process steps should be clearly describable by the introduced communication tools. On the other hand, the task should also not be logically self-explanatory since otherwise, the remote technician would not need help from the expert. Therefore, an assembly task with Lego bricks was chosen, where the remote expert sees the complete assembly and can instruct the MR user which brick of which color they should place next. The MR user is supposed to build this in real life. In a shared workplace, the users stand on opposite sides of a virtual table and can see each other's avatars consisting of head and hand positions, speak to each other, and share virtual objects. The real setup was duplicated for the MR and VR users, as seen in Figure 1. The MR user sees the virtual scene and the real table setup, while the VR user has the solution and an extra button allowing him to spawn arrows.



(a) MR perspective, real table underneath is not shown

(b) VR perspective, Lego configuration shown

Fig. 1. Virtual table shown from different perspectives

III. Implementation

Our pilot application ran on the Microsoft Hololens 2 (HL2) and the HTC Vive Pro and was developed in Unity. We use Microsoft's Mixed Reality Toolkit (MRTK), which allows for simple implementation for spatial interaction, and Photon Unity Networking 2 (PUN2), providing lower-level network primitives effectively connecting the two headsets to the shared virtual space.

The application follows a peer-to-peer network topology: a single code base is used, and every device participating in the network runs the same program. A single user owns each element, meaning that a single user can update the shared state of this element. The ownership can be transferred from one user to another, thus making it possible for the second user to modify the shared state of the element. In order to make interactions feel natural, the transfer of ownership was automated: any user can interact with any object in the scene and modify its state automatically, taking ownership in the process. This automation was made possible by coordinating MRTK concepts with PUN2 concepts. This allowed both users to move any of the virtual objects quickly. Enabling both users to manipulate the blocks is crucial, as it facilitates the expert in constructing the assembly, while the trainee can rotate the structure and dismantle it to gain a better understanding of its internal components.

A. Communication Tools

Four different tools were chosen to facilitate communication as much as possible.

1) Avatar and Hands: To create a sense of co-presence, avatars representing the users are a key requirement for collaborative VR/MR experiences. The avatars consist of an abstract, monochrome head shape and articulated hands, mimicking the head and hand movements of the users. The avatar allows for adding limited non-verbal communication by allowing the users to see where they are looking and giving them the possibility of nodding.

The hands are fully articulated in MR, as the HL2 natively tracks them in the field of view. Every finger that is moved can also be seen in the virtual space. Therefore, it is possible to point at certain objects. In VR, the user is a bit more restricted when it comes to different gestures. While the hands still look realistic, only a handful of gestures are possible. In fact, the virtual hands can only change their gesture if the respective button on the HTC Vive controller is pressed.

2) Voice Chat: Another indispensable feature of natural communication is a low-latency voice chat. It can be implemented with an additional Unity asset by PUN2, which presents a free asset and functions across multiple VR devices.

3) Workspace Visualization: The virtual table is a shared space where the users can collaborate using the same set of objects, in this case, Lego blocks. On the table, we have all the virtual counterparts of the real-world Lego blocks. The remote expert may assemble the blocks on the workspace to show how the blocks have to be put together in the real world. In contrast, manipulating the virtual blocks allows the MR user to rotate and dismantle the structure for better visualization.

4) Arrow: The "Arrow"button is only available to the VR user to provide specific instructions (e.g., highlighting or pinpointing a precise location). Despite the possibility of using the hands to point, the arrow button could still show some advantages since many of those can be placed and left in the scene highlighting some specific areas of the VE.

IV. User Study

We designed and conducted a user study to investigate the system, the different communication tools, and how well the information exchange worked between the two platforms. Specifically, we are interested in how the different communication tools work together and which ones are the most promising ones for future work.

A. Study Design

The study procedure is summarized in Figure 2. The light blue steps are done by the experimenter, while the dark blue ones are tasks for the participants. The participants are welcomed and informed about the study, and sign a consent form confirming that they know what the study is about. After being informed about the general process of the study, the participants fill out a pre-questionnaire with general information about the participant's demographics. Another part of the pre-questionnaire is about personal innovativeness and how quick a person is to try out new information technologies and how much they enjoy experimenting with them.

After completing the task, or after a maximum of ten minutes, the study is stopped, and both participants had to fill out another questionnaire. It consisted of the system usability score (SUS) questionnaire [14], the simulation task load index by Harris et al. [15], and a part with specific questions about the usefulness of the implemented communication tools. General information is gathered about the system and the communication tools by specifying a value between 0 and 10 for each tool, where 0 signals finding the tool unhelpful, and 10 means the tool was very beneficial. Next, the participants switched roles and solved the task again with a different LEGO constellation. After completing the second trial, they fill out the main questionnaire again.

V. Results and Discussion

The user study was carried out with eleven groups of two participants each. Every participant solved the task once in VR and once in MR. Each group managed to solve the task before reaching the time limit of ten minutes. Since we did not give an incentive to solve the task quickly, the groups took their time to explore the different tools.

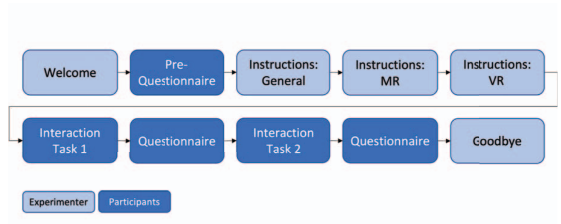


Fig. 2. User study procedure

A. Demographics

Of the 22 participants, six were female. Their ages ranged from 21 to 32 years, and the participants were recruited from the student staff. Twelve participants spent in total five hours or less using VR equipment. Most users had to adapt to the system's controls and learn how to handle grasping gestures.

B. Standardized Questionnaires

The results of the task load index show that the task did not overstrain the participants. It was neither mentally nor physically fatiguing, the participants did not feel stressed, and felt immersed in the virtual world.

The SUS resulted in a score of 79 out of 100. This is a reliable hint that the tools were well integrated and the system itself is not too complex, even for novice users.

C. Evaluation of the Communication Tools

The scores given by the participants for the different communication tools are presented in Figure 3.

1) Moving Lego Blocks: Being able to move virtual Lego blocks is a valuable feature, as this again makes it simple to give instructions on the exact location a block needs to be placed ($score = 7.23 \pm 2.68$). One way how certain groups solved the task was that the VR user built the solution virtually and let the MR user build the same thing in real life. This way, the MR user did not need to move any virtual blocks except for some cases where he used this feature to rotate or dismantle some parts of the structure. Then, this tool becomes useless to the MR user, which could explain those outliers.

2) Pointing Arrow: The feedback for the arrow was mixed ($score = 3.78 \pm 2.89$). Most people did not see much use in the arrow, as they could use their hands to point somewhere in the scene. However, there was still a number of users who did find it helpful, since there is a potentially large number of arrows available. Another advantage of the arrow is that it stays where it was put so the user can talk about several locations simultaneously.

3) Voice Chat: The voice chat was evaluated to be the most helpful communication tool ($score = 9.23 \pm 1.04$). This result was expected, as verbal communication is most natural, and it is simple to give specific instructions. This tool got the highest scores from all participants, as seen in Figure 3.

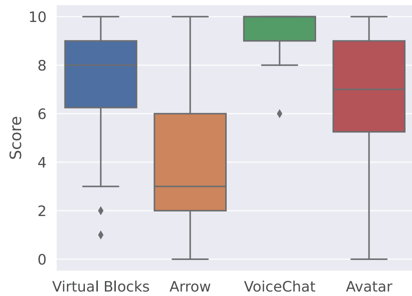


Fig. 3. Evaluation of the different communication tools

4) Avatar Representation: Most users found the head and hand movements helpful ($score = 6.77 \pm 2.76$). However, the results in this section are mixed because the hands were much more helpful to see than the head, as it cannot express much information. More particulary, using hands to gesture or point to specific locations was seen to be very useful. In the informal feedback, some users reported that it was good to see the head specifically, even if it is to get a more realistic feeling in the virtual world and to give a sense of the other user being present in the scene.

VI. Conclusion

In this work, we presented a remote cross-platform collaboration. We studied the case that one user is in VR while the other one joins the shared virtual space through MR. For the collaboration, we provided tools such as voice connection, avatar representation, pointing arrow, and interaction capabilities with virtual objects. The goal was to evaluate which of these tools is suited most to support such a VR/MR collaboration. For the evaluation, we chose an unevenly distributed information scenario, so that an instructor using VR had to guide a technician in MR how to assemble a physical Duplo brick construction. As a result of the user study, we found that the arrow is the least useful tool, while the other three communication tools were useful and should be kept. The system has been implemented in a way that enables first-time users to get accustomed to it quickly. While not all communication tools were equally helpful, they were enough to let all groups in the user study successfully solve the task.

A. Future Work

One remaining issue is that the VR user cannot check whether the MR user is doing the correct thing in the real world. This could result in the fact that the two users may misunderstand each other and end up with different results, even if both are sure they were talking about the same thing. To alleviate this risk, an automatic feedback from the MR user is required so that the VR user can be sure they understand the instructions correctly. One way to achieve this is to show the VR user a video stream of the real world that the MR user is seeing. With this stream available, the VR user can watch what the MR user is doing and immediately intervene if they see something going wrong. Another more sophisticated solution would be to perform an automatic detection of the bricks and of the way how they are assembled. With this information in place, a digital representation on the VR side could be generated that would help the remote expert to compare "as planned" vs. "as built".

References

[1] C. Papadopoulos, "Improving awareness in mobile cscw," *IEEE Transactions on Mobile Computing*, vol. 5, no. 10, pp. 1331–1346, 2006.

[2] L. Alem, F. Tecchia, and W. Huang, "Remote tele-assistance system for maintenance operators in mines," in *2011 Underground Coal Operators' Conferenbce*, 2011, pp. 171–177.

[3] S. Bottecchia, J.-M. Cieutat, and J.-P. Jessel, "Tac: augmented reality system for collaborative tele-assistance in the field of maintenance through internet," in *Proceedings of the 1st augmented human international conference*, 2010, pp. 1–7.

[4] W. Vorraber, J. Gasser, H. Webb, D. Neubacher, and P. Url, "Assessing augmented reality in production: Remote-assisted maintenance with hololens," *Procedia CIRP*, vol. 88, pp. 139–144, 2020.

[5] J. D. N. Dionisio, W. G. B. III, and R. Gilbert, "3d virtual worlds and the metaverse: Current status and future possibilities," *ACM Computing Surveys (CSUR)*, vol. 45, no. 3, pp. 1–38, 2013.

[6] A. Schäfer, G. Reis, and D. Stricker, "A survey on synchronous augmented, virtual, andmixed reality remote collaboration systems," *ACM Comput. Surv.*, vol. 55, no. 6, dec 2022. [Online]. Available: https://doi.org/10.1145/3533376

[7] C. G. Fidalgo, Y. Yan, H. Cho, M. Sousa, D. Lindlbauer, and J. Jorge, "A survey on remote assistance and training in mixed reality environments," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 5, pp. 2291–2303, 2023.

[8] S. Kim, G. Lee, N. Sakata, and M. Billinghurst, "Improving co-presence with augmented visual communication cues for sharing experience through video conference," in *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2014, pp. 83–92.

[9] D. Jo, K.-H. Kim, and G. J. Kim, "Effects of avatar and background representation forms to co-presence in mixed reality (mr) tele-conference systems," in *SIGGRAPH ASIA 2016 Virtual Reality Meets Physical Reality: Modelling and Simulating Virtual Humans and Environments*, ser. SA '16. New York, NY, USA: Association for Computing Machinery, 2016. [Online]. Available: https://doi.org/10.1145/2992138.2992146

[10] B. Jones, Y. Zhang, P. N. Y. Wong, and S. Rintel, "Belonging there: Vroom-ing into the uncanny valley of xr telepresence," *Proc. ACM Hum.-Comput. Interact.*, vol. 5, no. CSCW1, apr 2021. [Online]. Available: https://doi.org/10.1145/3449133

[11] D. Gasques, J. G. Johnson, T. Sharkey, Y. Feng, R. Wang, Z. R. Xu, E. Zavala, Y. Zhang, W. Xie, X. Zhang, K. Davis, M. Yip, and N. Weibel, "Artemis: A collaborative mixed-reality system for immersive surgical telementoring," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: https://doi.org/10.1145/3411764.3445576

[12] 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*, ser. CHI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–13. [Online]. Available: https://doi.org/10.1145/3313831.3376550

[13] 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*, ser. CHI '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–13. [Online]. Available: https://doi.org/10.1145/3173574.3173620

[14] J. Brooke *et al.*, "Sus-a quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.

[15] D. Harris, M. Wilson, and S. Vine, "Development and validation of a simulation workload measure: the simulation task load index (sim-tlx)," *Virtual Reality*, vol. 24, no. 4, pp. 557–566, 2020.