A Research Platform for Studying Mixed-Presence Collaboration

Wolfgang Büschel*
Interactive Media Lab Dresden
TUD Dresden University of Technology

Katja Krug*‡
Interactive Media Lab Dresden
TUD Dresden University of Technology

Marc Satkowski*
Interactive Media Lab Dresden
TUD Dresden University of Technology

Stefan Gumhold†‡
Chair of Computer Graphics and Visualization
TUD Dresden University of Technology

Raimund Dachselt*‡
Interactive Media Lab Dresden
TUD Dresden University of Technology

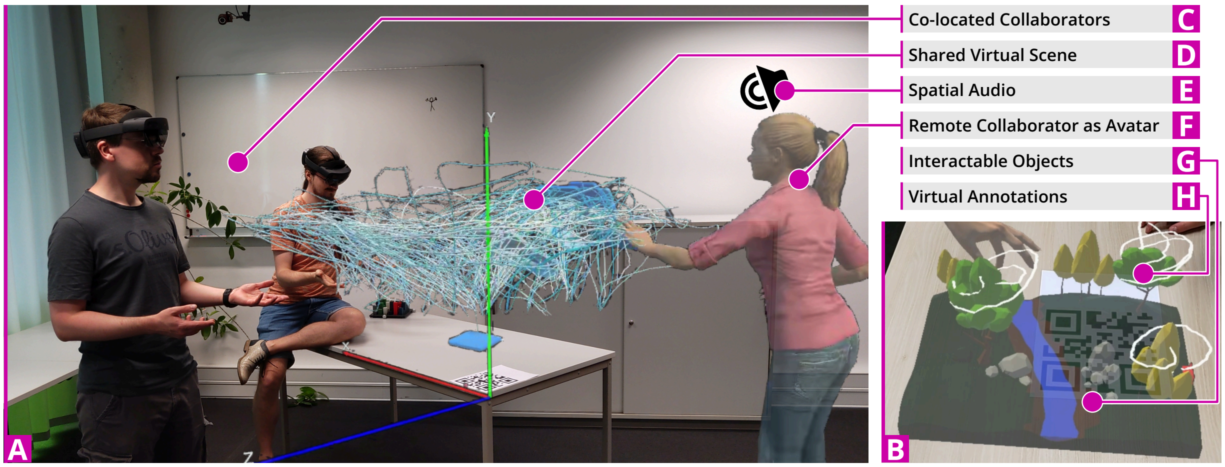

Figure 1: Our mixed-presence research platform used in different use cases: **(A)** data analysis and **(B)** public park planning. The system allows **(C)** co-located and **(F)** remote users to communicate through **(E)** spatial audio in one shared environment. Collaborators are able to **(H)** annotate or **(G)** interact with objects in the **(D)** shared scene.

Abstract

In this paper, we present a research platform to support studying collaboration in hybrid and co-located scenarios. Mixed-presence collaboration includes various novel and exciting use cases, such as situated and immersive data analysis by multiple users. However, research in this emerging field is hindered by the technical complexity of the setups and often requires re-implementation of common features. We address this issue by contributing a toolkit and research platform for mixed-presence collaboration that serves as an extensible baseline implementation and enables fast prototyping for user studies in collaborative mixed reality. Furthermore, our platform provides adjustable parameters, such as types of avatars, audio source placement, or the amount of simulated network latency. This way, developers are supported in making design choices regarding typical, re-occurring technical challenges.

Index Terms: Mixed Presence, Mixed Reality, Collaboration, Situated Analytics, Immersive Analytics

1 Introduction & Background

Immersive Analytics (IA) systems can be facilitated for *"collaborative analytical reasoning and decision-making"* [2] like collaborative sensemaking [3, 11]. At the same time, issues around collaboration have been identified as one major category of grand challenges in IA [5], including overcoming constraints of reality or supporting cross-platform collaboration. Specifically for situated analytics in Mixed Reality (MR), another essential aspect is to include remote participants. However, current research in this area is regularly stifled by many conceptual and technical challenges and requirements [14, 2] that must be addressed before research questions can be investigated.

Looking into the existing literature, several frameworks and systems exist that either inherently enable MR collaboration or were explicitly designed around this concept. For example, *Colibri* [9] is a framework for the rapid development of networked MR applications. On the other hand, *Ubiq* [7] is a framework for multiuser VR applications, such as social VR and remote collaboration. Other frameworks, like *RagRug* [6] or *MIRIA* [3] are targeting co-located MR collaboration, here specifically for situated analytics. Lastly, *CADET* [12] already presents a mixed-presence collaborative system for data exploration.

Investigating challenges within the research area of mixed-presence collaboration can be difficult due to the increasing complexity of setups. Furthermore, a lack of standardized components also means that researchers often need to re-implement common features. The aforementioned systems only target a subset of features that we believe to be essential for future mixed-presence systems. We address this by providing a research platform that serves as an open-source baseline implementation to build upon, uniting multiple core features of previous frameworks. Our research platform allows for fast prototyping and feature extension to explore mixed-presence systems through adjustable parameters, such as types of avatars, audio source placement, or adjustable simulated network latency. In the following, we describe the platform in detail. To further highlight its applicability, we will then describe a set of use cases illustrating its different features.

*e-mail: [bueschel, katjakrug, msatkowski, dachselt]@acm.org
†e-mail: stefan.gumhold@tu-dresden.de
‡also with Centre for Tactile Internet with Human-in-the-Loop (CeTI)

2 Research Platform

We designed our platform to support studies on several interesting aspects of collaboration by providing user-configurable parameters. These include, among others, simulated network latency, different avatar representations, and configurable spatial audio for real-time voice communication. Our software is open source, available on GitHub[1]. By making it available to other interested researchers, we offer our research platform as a foundation for future studies.

In the following, we describe the general architecture of our platform and highlight its main features.

2.1 Platform Architecture

We chose the Unity game engine as the base for our platform due to its widespread use in HCI research and the excellent support for different MR headsets. We currently only target Microsoft's HoloLens 2 AR glasses using MRTK v2, but our platform can easily be adapted to support other OpenXR-compatible devices fully.

Given our goal to support multiple users and, in the future, heterogeneous device ensembles, we built our platform around the concept of services, which provide an easy and scalable architecture: On the application level, components provide services that other components may use if necessary. These services are made accessible using a service locator pattern. Between devices, network services provide thematic data channels that clients can connect to, e.g., data sources like an audio stream or body tracking data. Network services announce themselves via broadcasts. Once a connection has been established, they allow bidirectional communication between all connected participants.

With this architecture, we also ensure our platform's extensibility through the simple addition of new services. Furthermore, all existing services are based on interfaces, allowing developers to add support for, e.g., new sensor hardware or a different network layer.

We encapsulated most of our platform's features in a set of prefabs that can easily be integrated into existing scenes in Unity. For easier deployment, an install dialog guides users through selecting optional features.

2.2 Key Features

Our platform supports mixed-presence sessions of multiple users in several rooms. Each room typically corresponds to a physical location while also grouping users logically. For example, by default, only audio streams of remote users (i.e., users in different rooms) are streamed.

Configuration & Logging We provide services for the persistent storage of configuration parameters and text-based logging of, e.g., study data. Furthermore, we made key parameters configurable through a body-attached AR palm interface. Each user can use this lightweight menu to control their own settings or set options globally. This currently includes choosing the user representation, spatial audio configuration, and target latency adjustments, but it can easily be extended with additional features. By allowing users to make runtime adjustments to these parameters, we support testing different options quickly.

User Representation We use the built-in tracking of the HMDs, capturing the head-pose in 3D space and enabling hand- and finger-tracking, optionally in combination with full-body tracking by Kinect Azure depth sensors. Based on this, we support per-user choices of different avatar representations. Currently, we include avatars from the *Rocketbox* [8] (see Fig. 1F), *VALID* [4] (see Fig. 2B), and *ReadyPlayerMe*[2]. Furthermore, we offer a simple head and hands avatar if no body tracking is available. Finally, our platform also streams RGB-D data of tracked users for a point-cloud-based representation (see Fig. 2A) as used in, e.g., [13].

Spatial Audio Spatial audio is an essential factor in how convincing a mixed-presence collaboration system is. Accordingly, we support recording and transmitting the users' audio input and use Microsoft's spatializer plugin for Unity to provide real-time 3D voice chat capabilities in our platform (see Fig. 1E). However, there are systems (e.g., [10]) in which researchers opted to place a stationary audio source in the room for a more straightforward setup. Thus, we allow switching between full 3D spatialization, stationary audio source placement, and classic, non-spatial sound during runtime. With this, we aim to make the impact of audio quality on presence and collaboration quality easier to study.

Interactive Objects In many use cases, the collaborative arrangement of virtual objects (including visualizations) plays a central role. In our research platform, we provide a simple network-synchronized object manipulation feature. Developers can make scene objects interactable by adding a single script to them or by adding them as children of a pre-defined parent object in the scene graph. During runtime, they can then be moved around freely (see Fig. 1G and Fig. 2E) and also snap to each other or to one or more snapping guides (e.g., a ground plane).

Shared Annotations In order to support collaborative sensemaking tasks, we also provide shared annotations. As pictured in Fig. 1H, users can draw free annotations in 3D space. These annotations are positioned in the joint virtual space and shared among all local and remote users. In our current implementation, the left hand is used to activate and deactivate the annotation mode, and the right hand controls the actual drawing action.

Coordinate Mapping We use QR codes to map coordinates between users in a room and the rooms to each other. One code per room defines a common, shared virtual space, and one code per Kinect defines the respective sensor's position in that space. Still, inaccuracies of the devices' sensors and QR-code tracking, drift, and the number of necessary coordinate transformations between multiple users can lead to reduced mapping quality. To counteract this, we added a manual calibration tool that allows users to calibrate their local Kinect to the HoloLens frame of reference at any time. To this end, the local Kinect's point cloud is visualized, effectively showing the mapping accuracy between the real and virtual worlds. A 3D widget (see Fig. 2D) can then be used to precisely translate and rotate the position of the sensor in the virtual space to optimize the mapping.

Latency Adjustment Both the literature and our own experiences underline the importance of system latency for the feeling of presence and the efficacy of communication with remote participants. The latency of a system is dependent on many factors, e.g., the hardware, the distance between remote locations, the network link quality, and the output latency of displays. To make it easier to study these effects, we included a (simulated) network latency parameter in our platform (see Fig. 2C). After setting a target latency, any network message with an estimated latency lower than the target threshold is delayed accordingly. With this, even large delays of several hundred milliseconds can be tested without additional equipment. Incidentally, this feature can also be used to synchronize between different network streams, as it allows users to set a common, lower-bound latency target.

3 Illustrating Study Scenarios

To illustrate the capabilities of our platform, we implemented three simple use-case applications based on it, incorporating the aforementioned key features. In the following sections, we will describe scenarios of potential mixed-presence user studies based on these applications using the fictitious personas, Matt, Lucy, and Taylor, who are researchers in the field of human-computer interaction.

[1] https://github.com/imldresden/mp-collab
[2] https://readyplayer.me

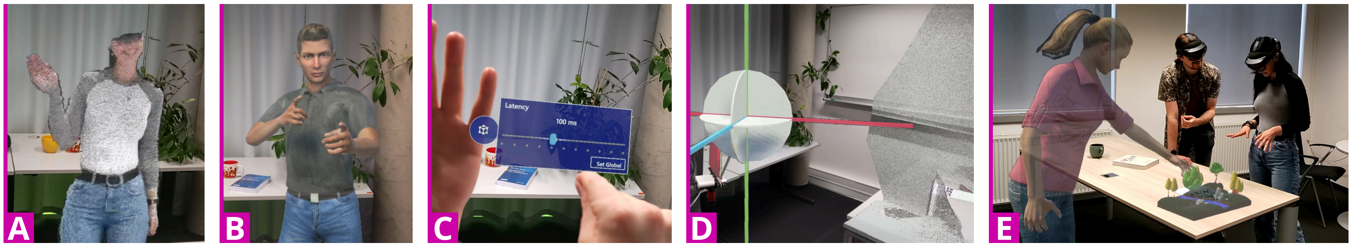

Figure 2: Different aspects of our mixed-presence research platform. **(A+B)** Two other types of avatars supported by our system: point cloud representation and a model from [4]. **(C)** Altering the target network latency live within the system. **(D)** A calibration widget allows adjusting the automatic coordinate mapping. A poorly aligned point cloud can be seen on the right. **(E)** The park planning scenario.

Communication Behavior in Hybrid Meetings Lucy, Matt, and Taylor want to study the communication behavior of remote and co-located users in hybrid meetings. Specifically, they are interested in the effect of different avatar representations. They hypothesize that more realistic avatars help to include remote participants in the discussion.

Using our research platform, they implement a small research prototype in which two participants will be in one of their meeting rooms, and a third one will join the session from a second room. For their study, they focus on providing multiple *virtual representations*. As a baseline, they choose a simple *head-and-hands avatar* that doesn't need body tracking and would work in non-instrumented settings, relying only on the tracking of the HMDs. Lucy also suggests including *point cloud avatars*, as they are a more faithful representation of a person and might circumvent the uncanny valley effect. Finally, Taylor argues for *3D mesh avatar* using a widely used library of avatars spanning diverse genders and ethnic groups so participants may choose an avatar that represents them better. For a more realistic auditory experience, they decide to use 3D spatial audio for the remote participant.

During their within-subject study, they counterbalance the order of the three avatar types. They run short discussion groups on different social topics, and having extended the *logging features* of the platform, they log where users are looking and who is talking. They also record mixed-reality captures of the study for later video coding and transcription of the discussion content.

Roles in Collaborative Data Analysis In their next study, Matt, Lucy, and Taylor are interested in role distributions between local and remote participants in a collaborative visual analytics task (see Fig. 1A). They study this using a dataset of interaction patterns above a tabletop. By integrating a framework for the situated analysis of user interaction data [3], they extend the research platform for their specific use case. A *3D visualization* of movement trajectories is placed above a table (see Fig. 1D). It can be explored from all sides, supports dynamic playback of the data over time, and supports different filters.

The three researchers again decide to utilize *spatial audio* and agree on using *mesh-based avatars*, as they are not interested in the facial features of the participants but rather the general patterns and roles shown in the study sessions. During the study, Taylor manually aligns each Kinect for each session. The local point clouds are visualized in their respective rooms, showing whether the alignment is incorrect. Taylor grabs the *calibration widget* (see Fig. 2D) and aligns the point cloud to the environment. This way, they ensure that the avatars' interaction with the virtual content is as precise as possible. Immersed in their analytics task, the participants can *walk around* the table to gain new perspectives into the data. Observing the participants carefully, Matt and Lucy sometimes watch someone step back to gain an overview and give advice to the other participants, indicating that they assumed a form of coordinator role.

Effect of Latency on Remote Guidance & Instruction In the literature, Lucy found that the effect of network problems on mixed-presence tasks is an exciting research topic, having been examined, e.g., by Ahsen et al. [1]. Therefore, the three researchers decide to examine the *influence of network latency* as one specific aspect of this larger topic.

Using our research platform and building on their prior study prototypes, they quickly develop an application for a small pre-study. With it, users can co-design gardens and parks by manipulating 3D models of plants, park furniture, etc. (see Fig. 1B and Fig. 2E). They plan to have a remote expert guide a local team of students in designing the layout of a small park under changing latency conditions. They choose *3D mesh avatars* over other avatar types, such as *point clouds*, because they require precise visualizations of the participants' hands for pointing.

During their test, they observe how the invited landscaping expert inspects the 3D scene and comments on the placement of the plants: "I would put this tree over there," she says, pointing to the object and its target location. She then encircles the location with *3D mid-air annotations* (see Fig. 1H) that are synced and displayed in situ for the other participants. The participants can *grab objects* and place them onto the park's model. Matt is excited to turn up the *simulated network latency* in the *palm menu* (see Fig. 2C). He wonders how communication with the expert will change with just a few hundred milliseconds delay...

4 Conclusion

We have presented our toolkit and modular research platform for mixed-presence collaboration. Our platform unites various key features for studying multi-user mixed-reality use cases like collaborative Immersive Analytics, such as coordinate frame mapping, spatial audio transmission, and logging functionality. It further provides adjustable parameters to preview and study commonly faced implementation considerations, such as network latency, audio placement, and diverse user representations with varying complexity and realism. With common interaction possibilities already enabled, e.g., interactive and synchronized objects and shared 3D annotations, our platform serves as a baseline implementation for future use case applications within the realm of distributed multi-user Mixed Reality. In three example use cases, we illustrate the applicability of our platform to enable fast prototyping and support decision-making when designing user studies in this context. With this platform, we support researchers in their future endeavors by decreasing the technological barriers to entry that precede the realization of mixed-presence user studies.

Acknowledgments

This work was supported by the German Federal Ministry of Education and Research (BMBF) in the program "Souverän. Digital. Vernetzt.". Joint project 6G-life, project identification number: 16KISK001K. Furthermore, the authors acknowledge the support of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy: EXC-2068, 390729961 – Cluster of Excellence "Physics of Life" and EXC 2050/1, 390696704 – Cluster of Excellence "Centre for Tactile Internet" (CeTI) of TU Dresden, by DFG grant 389792660 as part of TRR 248 – CPEC (see `https://cpec.science`) and by the German Federal Ministry of Education and Research (BMBF, SCADS22B) and the Saxon State Ministry for Science, Culture and Tourism (SMWK) by funding the competence center for Big Data and AI "ScaDS.AI Dresden/Leipzig".

References

[1] T. Ahsen, Z. Y. Lim, A. L. Gardony, H. A. Taylor, J. P. de Ruiter, and F. Dogar. The Effects of Network Outages on User Experience in Augmented Reality Based Remote Collaboration - An Empirical Study. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2), Oct. 2021. doi: 10.1145/3476054 3

[2] M. Billinghurst, Maxime Cordeil, Anastasia Bezerianos, and Todd Margolis. Collaborative Immersive Analytics. In *Immersive Analytics*, vol. 11190. Springer International Publishing, Cham, 2018. doi: 10.1007/978-3-030-01388-2 1

[3] W. Büschel, A. Lehmann, and R. Dachselt. MIRIA: A Mixed Reality Toolkit for the In-Situ Visualization and Analysis of Spatio-Temporal Interaction Data. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, pp. 1–15. Association for Computing Machinery, New York, NY, USA, May 2021. doi: 10.1145/3411764.3445651 1, 3

[4] T. D. Do, S. Zelenty, M. Gonzalez-Franco, and R. P. McMahan. VALID: A perceptually validated Virtual Avatar Library for Inclusion and Diversity. *Frontiers in Virtual Reality*, 4, Nov. 2023. doi: 10.3389/frvir.2023.1248915 2, 3

[5] B. Ens, B. Bach, M. Cordeil, U. Engelke, M. Serrano, W. Willett, A. Prouzeau, C. Anthes, W. Büschel, C. Dunne, T. Dwyer, J. Grubert, J. H. Haga, N. Kirshenbaum, D. Kobayashi, T. Lin, M. Olaosebikan, F. Pointecker, D. Saffo, N. Saquib, D. Schmalstieg, D. A. Szafir, M. Whitlock, and Y. Yang. Grand Challenges in Immersive Analytics. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 2021. doi: 10.1145/3411764.3446866 1

[6] P. Fleck, A. S. Calepso, S. Hubenschmid, M. Sedlmair, and D. Schmalstieg. RagRug: A Toolkit for Situated Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 29(7):3281–3297, July 2023. doi: 10.1109/TVCG.2022.3157058 1

[7] S. J. Friston, B. J. Congdon, D. Swapp, L. Izzouzi, K. Brandstätter, D. Archer, O. Olkkonen, F. J. Thiel, and A. Steed. Ubiq: A System to Build Flexible Social Virtual Reality Experiences. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*, VRST '21, pp. 1–11. Association for Computing Machinery, New York, NY, USA, Dec. 2021. doi: 10.1145/3489849.3489871 1

[8] M. Gonzalez-Franco, E. Ofek, Y. Pan, A. Antley, A. Steed, B. Spanlang, A. Maselli, D. Banakou, N. Pelechano, S. Orts-Escolano, V. Orvalho, L. Trutoiu, M. Wojcik, M. V. Sanchez-Vives, J. Bailenson, M. Slater, and J. Lanier. The Rocketbox Library and the Utility of Freely Available Rigged Avatars. *Frontiers in Virtual Reality*, 1, Nov. 2020. doi: 10.3389/frvir.2020.561558 2

[9] S. Hubenschmid, D. I. Fink, J. Zagermann, J. Wieland, H. Reiterer, and T. Feuchtner. Colibri: A Toolkit for Rapid Prototyping of Networking Across Realities. In *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 9–13, Oct. 2023. doi: 10.1109/ISMAR-Adjunct60411.2023.00010 1

[10] A. Irlitti, M. Latifoglu, Q. Zhou, M. N. Reinoso, T. Hoang, E. Velloso, and F. Vetere. Volumetric Mixed Reality Telepresence for Real-time Cross Modality Collaboration. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, pp. 1–14. Association for Computing Machinery, New York, NY, USA, Apr. 2023. doi: 10.1145/3544548.3581277 2

[11] W. Luo, A. Lehmann, H. Widengren, and R. Dachselt. Where should we put it? Layout and placement strategies of documents in augmented reality for collaborative sensemaking. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3491102.3501946 1

[12] J. McDade, A. Drogemuller, A. Jing, N. Ireland, J. Walsh, B. Thomas, W. Mayer, and A. Cunningham. CADET: A Collaborative Agile Data Exploration Tool for Mixed Reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 899–900, Oct. 2022. doi: 10.1109/ISMAR-Adjunct57072.2022.00195 1

[13] R. Schlagowski, D. Nazarenko, Y. Can, K. Gupta, S. Mertes, M. Billinghurst, and E. André. Wish You Were Here: Mental and Physiological Effects of Remote Music Collaboration in Mixed Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, pp. 1–16. Association for Computing Machinery, New York, NY, USA, Apr. 2023. doi: 10.1145/3544548.3581162 2

[14] M. Sereno, X. Wang, L. Besançon, M. J. McGuffin, and T. Isenberg. Collaborative Work in Augmented Reality: A Survey. *IEEE Transactions on Visualization and Computer Graphics*, 28(6):2530–2549, June 2022. doi: 10.1109/TVCG.2020.3032761 1