ICS-MR: Interactive Conversation Scenarios for Assessment of Mixed Reality Communication

Felix Immohr[*,†]
Chair of Signal Processing in Communication Engineering
RWTH Aachen University
Aachen, Germany
immohr@ient.rwth-aachen.de

Gareth Rendle[*]
Virtual Reality and Visualization Research Group
Bauhaus-Universität Weimar
Weimar, Germany
gareth.rendle@uni-weimar.de

Annika Neidhardt
Institute of Sound Recording
University of Surrey
Guildford, United Kingdom
a.neidhardt@surrey.ac.uk

Anton Benjamin Lammert
Virtual Reality and Visualization Research Group
Bauhaus-Universität Weimar
Weimar, Germany
anton.benjamin.lammert@uni-weimar.de

Bernd Froehlich
Virtual Reality and Visualization Research Group
Bauhaus-Universität Weimar
Weimar, Germany
bernd.froehlich@uni-weimar.de

Alexander Raake
Chair of Signal Processing in Communication Engineering
RWTH Aachen University
Aachen, Germany
raake@ient.rwth-aachen.de

(a) Scenario 1: Floor Plan Negotiation. **(b) Scenario 2: Spot the Difference.** **(c) Scenario 3: Survival Game.**

Figure 1: The ICS-MR dataset contains three conversation scenarios for assessing communication in Mixed Reality contexts. Full descriptions, flexible and extensible study materials, and multi-user Unity implementations (pictured) are provided, allowing straightforward application of scenarios in user studies conducted in immersive communication systems.

Abstract

We present ICS-MR, a dataset containing three conversational scenarios designed for the evaluation of communication quality in Mixed Reality (MR) systems. Along with detailed descriptions of the conversation tasks, we provide all the materials required to incorporate the tasks into MR user studies. The materials also support application of the scenarios in real-world and video-conferencing contexts for studies that, for example, call for comparison of immersive systems against reference communication media. Open-source Unity implementations of the scenarios are also made available, supporting direct usage of the scenarios in distributed, multi-user experiments. The conversation tasks have all been administered in recent scientific works that address the evaluation of user experiences in immersive communication systems, allowing analysis and comparison of each scenario's evoked behavioral properties. The ICS-MR dataset therefore contributes valuable resources for further research on communication in immersive systems.

CCS Concepts

• **Human-centered computing** → **Virtual reality**; **Collaborative interaction**.

Keywords

Conversation, communication, task, scenario, mixed reality, video conferencing, collaboration, telemeeting, quality of experience

ACM Reference Format:

Felix Immohr, Gareth Rendle, Annika Neidhardt, Anton Benjamin Lammert, Bernd Froehlich, and Alexander Raake. 2025. ICS-MR: Interactive Conversation Scenarios for Assessment of Mixed Reality Communication.

[*]Both authors contributed equally to this research.
[†]Also with Audiovisual Technology Group, Technische Universität Ilmenau.

MM '25, Dublin, Ireland

ACM ISBN 979-8-4007-2035-2/2025/10
https://doi.org/10.1145/3746027.3758247

In *Proceedings of the 33rd ACM International Conference on Multimedia (MM '25), October 27–31, 2025, Dublin, Ireland*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3746027.3758247

1 Introduction

Social Mixed Reality (MR) systems allow users to enter immersive virtual environments where they can interact with remote conversation partners. The benefits of immersive systems, which include the creation of a shared spatial context for collaborative work and support for rich communication by conveying non-verbal cues like gestures and facial expressions, mean that they have the potential to complement (or even succeed) traditional (video-)conferencing as widely adopted remote communication media. This apparent promise has led to research evaluating the quality and effectiveness of social MR systems in different contexts, with work examining how mediated and unmediated communication differ [31, 35, 6], as well as investigating how social interaction is affected by different system or service properties (e.g. avatar realism [10, 31, 35, 24], audio spatialization [13, 8], and network parameters [26, 5]).

A key consideration when designing an experiment to assess communication quality in a given medium is the choice of the conversation scenario that is prescribed to participants. Scenarios can be designed to assess suitability of the medium for conveying certain cues (e.g. pointing gestures or gaze movements) or for supporting different conversation goals (sharing information, discussing emotions, facilitating negotiation or collaboration). While research into communication quality evaluation in audio- and video-conferencing systems has led to the recommendation of a set of standardized conversation scenarios (c.f. ITU-T Rec. P.8XX Series [18], P.9XX Series [20] and P.13XX Series [15, 16]), methodological guidelines for communication assessment in immersive systems are still under development [17, 29]. Conversation scenarios that were established for telephony or video-conferencing contexts may not transfer well to MR applications, where certain actions (e.g. writing with a pen) may not be well supported, or where the shared interaction space inhabited by all participants opens possibilities for tasks involving shared visual references. Many realizations of standardized and non-standardized scenarios are developed for and used in numerous studies, but their materials and implementations are rarely provided for further study. Accordingly, researchers have little information on how potential conversation scenarios and specific implementations affect the resulting participant interaction.

In an effort to aid researchers in choice and implementation of experimental scenarios that support controlled, systematic and comparable evaluation, we present ICS-MR, a dataset containing three conversational scenarios designed for the evaluation of communication quality in MR systems. The scenarios represent a range of situations that encourage different types of interaction (negotiation, collaboration, discussion), emphasize different non-verbal cues and gestures, and encourage varying levels of participant movement. Along with detailed descriptions of the conversation scenarios, we provide all materials required to incorporate the scenarios into conversational MR user studies, including open-source Unity implementations which enable direct application of the scenarios in distributed, multi-user experiments. In addition, all necessary materials are provided for employing the scenarios in real-world and video-conferencing contexts, for studies that call for a comparison of MR systems against reference media. The materials have been developed to be flexible and extensible, allowing the scenarios to be tailored to a wide range of experimental scenarios and media.

The presented scenarios have all been administered in recent scientific works, published by the authors, that address the evaluation of user experiences in immersive communication systems [14, 13, 12, 31]. We analyze data generated during those studies to inform a comparison of the behavior evoked by each scenario in terms of participant locomotion and movement. This work therefore provides researchers with a valuable tool when selecting a conversation scenario for communication assessment studies, and also informs the development of recommendations for standardized scenarios for assessment of immersive communication systems more generally.

In summary, our work includes the following contributions:

- The ICS-MR dataset, comprising three conversation scenarios for communication quality assessment in MR;
- Flexible, extensible, open-source materials[1] for administering each of the described scenarios in virtual environments and real-world reference contexts;
- Unity implementations[2] of the scenarios, supporting straightforward application in distributed, immersive communication studies with minimal development overhead;
- A comparison of the behavioral properties evoked by each scenario, informing scenario selection for future research evaluating communication in MR systems.

We review related work in Section 2 before introducing the scenarios in the ICS-MR dataset in Section 3. The analysis of the behavioral properties evoked by the scenarios and a discussion of their relative characteristics can be found in Sections 4 and 5.

2 Related Work

In the context of communication systems, many conditions influence user experience, including system, human and context-influencing factors [34]. To assess such systems or services in interactive and conversational multi-user contexts, users are typically placed in conversation test scenarios that evoke engagement with the system or service depending on the subject of investigation. Such scenarios can target spontaneous free-form conversations or administer goal-oriented experimental tasks to users.

Existing conversation test scenarios and methodology have been proposed or standardized for application in different contexts, and can exhibit varying degrees of structure in the conversation. Structured [27] and interactive [30] short conversation tests were proposed for audio contexts and standardized in ITU-T Rec. P.805 [19]. As an extension of the P.8XX Series for audiovisual conversation tests, P.920 [20] recommends the name-guessing task (structured question and answer), unstructured story or picture comparison tasks and a building block task with focus on the visual terminal. With a focus on multi-party telemeeting assessment, P.1301 [15] recommends the unstructured survival task (see Section 3.3), the Leavitt task (see Section 3.2 and brainstorming tasks. With the increase of considered modalities introduced by system affordances,

[1] https://github.com/Telecommunication-Telemedia-Assessment/ics-mr-communication-scenario-materials
[2] https://github.com/vrsys/ics-mr-unity-communication-scenarios

attentional allocation on the modality of interest is methodologically recommended (e.g. keeping users' focus on the screen during video conferencing). If transmission delay is to be investigated, ITU-T Rec. P.1305 [16] recommends tasks exhibiting high delay sensitivity, including role playing game, navigation and random number verification tasks, as well as modifications of name-guessing, block building and survival tasks. In the context of immersive communication systems, for example those based on MR technologies, Pérez et al. [29] review related studies and propose a taxonomy for evaluation. Based on that, tasks of related user studies are classified into deliberation, exploration and manipulation tasks. While many study realizations of such tasks exist (new and based on the tasks proposed for traditional media), validation and standardization is still in development for immersive systems.

In addition to holistic evaluation of experience, research studies often target modality-specific investigations that pose additional requirements for test methodology and scenario. Modalities could include auditory, haptic, and visual aspects on the basis of which, for example, aspects of user representation, the virtual environment, virtual objects and different interaction methods are investigated. Desirable properties that scenarios should evoke could include manipulation, exploration, attentional allocation, speaker activity/distribution, movement (e.g. head, hands, gesturing, facial expressions, pose), technical setup integrability, scalability towards multiple users [34, 9]. While not restricted to the use in social MR, immersive scene datasets are available that implement varying degrees of complexity, for example in the auditory scene [32].

The use case, conversation situation and task are context factors that should be controlled as they influence evoked behavior and (quality) perception [34, 36, 17]. In an effort to support in-depth systematic and comparable task characterization, possible only with shared reference implementations and material, we make our resources openly available in this work.

3 The ICS-MR Dataset

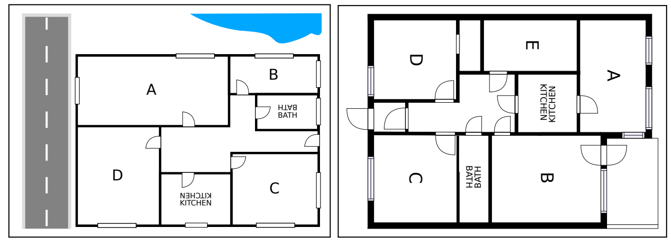
The dataset consists of three conversational scenarios, described in detail below. Each scenario is associated with the following:

- Exemplary task instructions to participants during the study;
- Task materials required to present the scenario in virtual contexts, which can also be printed for use in real-world contexts (or for use as real stimuli in MR contexts);
- Tools for flexible extension of the provided materials;
- A Unity implementation of the scenario that can be directly employed in distributed user studies.

The Unity implementations of each scenario can be found in the ICS-MR Unity repository, while the scenario materials can be found in the ICS-MR scenario material repository. Note that in both repositories, the materials and scenes are organized by scenario. Please see the repositories' README files for more information.

3.1 Scenario 1: Floor Plan Negotiation

This negotiation task is based on the version introduced by Smith and Neff [35] and further developed by Abdullah et al. [1], and was employed in a study conducted by Immohr et al. [14].



(a) With environment (2P) **(b) No environment (3P)**

Figure 2: Example floor plans for Scenario 1.

3.1.1 Description. Participants play the role of roommates looking for a new apartment. They are shown a floor plan of a potential apartment and are instructed that their goal is to agree on the assignment of the rooms. The discussion relies on a shared visual reference in the form of a fictional floor plan of the apartment, which elicits referential talking and gesturing. The plans include some pre-assigned rooms, including kitchens and bathrooms, and several unassigned rooms. Participants are required to assign shared living and dining rooms, and a bedroom for each participant.

By creating a situation where limited resources must be shared between the participants, the role-play scenario contains elements of conflict, negotiation, and reaching a consensus. Optionally, participants can be prescribed conflicting room preferences, emphasizing the need for discussion and compromise from at least one party.

3.1.2 Included Materials and Variations. We provide floor plans for variants of the task suitable for two- and three-party scenarios (with four or five rooms to be assigned, labeled with letters A-E for referencing). In total, four plans with four assignable rooms and six plans with five rooms are available. In a provided variation, features of the surrounding environment, such as nearby roads and lakes, are shown on the floor plan, influencing the desirability of the rooms. All included floor plans have variants with and without surrounding environment features. Example floor plans are shown in Figure 2. In addition to exemplary participant instructions, participant preference sheets are available for each floor plan. The scenario can be easily administered in a real-world communication context by printing the floor plans on large sheets of paper.

3.1.3 Extensions. The task can be scaled to multi-party experiments by increasing the number of unassigned rooms. The provided material can be extended by adapting the floor plan images (included in the dataset with a vector graphics format). Further floor plan designs can be sourced from open databases [37, 7].

3.2 Scenario 2: Spot the Differences

This task is inspired by the *Leavitt task*, originally introduced by Leavitt [25] for the assessment of text-based group communication and subsequently recommended for the assessment of traditional multi-party telemeetings in ITU-T Rec. P.1301 Appx. V.2 [15]. In the recommended task, sets of shape symbols are individually shown to each participant on paper cards. Participants are then instructed to identify the item that is common to all participants' sets of shapes through conversation. While information is inherently separated between users in traditional audiovisual and text-based communication systems, users of MR systems typically inhabit a common interaction space and share visual references, diminishing the need to exchange information verbally. To transfer this paradigm to a social MR context, a modified version was realized as a *spot the difference* task [13], which is described below.

3.2.1 Description. In the MR version of the task, the participants' objective is to identify the *differences* between their respective sets of colored shape symbols. Each participant's set of symbols is represented on a set of boxes that are visible in the virtual environment (see 3a and 3b). Each box shows exactly one shape, on the side of the box that faces the participant. Both sets of boxes are laid out in the same position relative to the participant, resulting in a clear correspondence between pairs of boxes from each set. The symbols on corresponding boxes may differ in form, orientation, or color.

Participants are instructed to discuss with their conversation partner to identify corresponding boxes that display different symbols, and to mark the relevant cubes by intersecting them with their virtual hand (Figure 3b). The task is completed once all differing shapes were found and marked. To prevent participants from solving the task without communicating, the boxes should be positioned such that no position in the virtual environment allows the user to see both sets of shapes simultaneously. To avoid the possibility of solving the task by moving to view the other participant's shapes, participant movement should be restricted (for example, by instructing participants not to cross a line that divides the room).

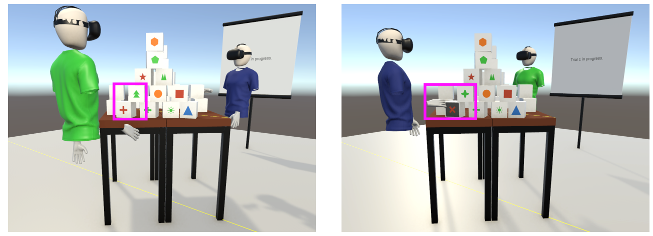
3.2.2 Included Materials and Variations. The implementation of the *spot the difference* task provided in the ICS-MR dataset for use in social MR systems is designed for dyadic communication scenarios. The materials specify four sets of 14 symbol pairs, each with three non-matching pairs. The shapes are attached to a set of boxes that are distributed on and around a table in the virtual environment. Two different variations of how the boxes are distributed are included in the provided implementation (c.f. Figures 1b and 3).

The materials include the symbol sets as SVG image files for representation in arbitrary formats. This also allows application of the task in paper form, similar to ITU-T Rec. P.1301 Appx. V.2 [15], as well as the implementation of a real-world version of the task that displays shapes on paper attached to cardboard boxes.

3.2.3 Extensions. The provided shape sets can be trivially edited and extended to create new combinations of symbols with different numbers of non-matching symbol pairs. We include a script to create shape images for edited symbol combinations. Further extensions could include increasing the visual complexity of shapes to increase task difficulty, and scaling towards higher numbers of interlocutors. We note that the boxes must be carefully arranged when scaling the number of participants, such that participants are only able to see their own shape sets.

3.3 Scenario 3: Survival Game

The survival game is a group discussion task that was originally proposed to assess collective decision-making performance [11], before being adapted as a tool for evoking conversation and collaboration in interactive social systems. In the game, the topic of conversation is a *survival scenario*, where participants are presented



(a) Participant 1's view, with differing shapes highlighted. **(b) Participant 2 marks one of the boxes with a differing shape.**

Figure 3: Spot the Difference (Scenario 2). Participants identify and mark corresponding boxes with different shapes.

with a fictional, life-threatening situation (e.g. being stranded in a desert after a plane crash). The participants are also informed that they have access to a number of items which could help them to survive, and are required to discuss the relative utility of the items in the context of the given survival scenario.

Realizations of the survival game exist in a wide range of formats, depending on research goals, system affordances, and the communication modalities under consideration. In the context of multi-party telemeeting assessment for telephony and video conference systems, the task has been recommended in ITU-T Rec. P.1301 Appx. V.4-VI [15] and provided therein in a paper-based format (c.f. [33] for original version in german language). The recommendation adopts four existing survival tasks in *desert* [23], *sea* [28], *winter* [21] and *moon* [11] environments, which are complemented with three further scenarios, namely the *mountains*, *swamp* and *cave labyrinth* scenarios. An adaptation of the task, building on the same scenarios, is also recommended for investigation of the effect of transmission delay in ITU-T Rec. P.1305 Annex A.7 [16]. Furthermore, implementations have been developed for use in immersive environments, with some versions placing participants in an (audio-) visual recreation of a survival scenario [2], while survival items are represented as abstract (manipulable) virtual objects in others [4, 3, 22]. The task materials presented in this dataset represent an update to ITU-T Rec. P.1301 Appx. VI [15] and were used in two recent studies in different MR systems [12, 31].

3.3.1 Description. In the survival game realization included in this dataset, participants are presented with the survival scenario on a text display in the virtual environment. In addition, a number of survival items are distributed throughout the environment, represented as boxes with labels and images depicting the items (see Figure 4). Participants are instructed to select a subset of the items that they believe will help them survive in the given scenario, by moving boxes to a specified target area (the blue grid shown in Figure 5c). The distribution of boxes in the virtual environment aims to encourage verbal information exchange between participants.

3.3.2 Materials and Variations. The ICS-MR dataset includes all seven survival scenarios from ITU-T Rec. P.1301 [15]. In addition to the scenario descriptions, labels and high resolution images are provided for 12 survival items for each scenario. These can be printed and attached to cardboard boxes to create a real-world replication of the scenario. In an alternative paper-based variation, used in MR by

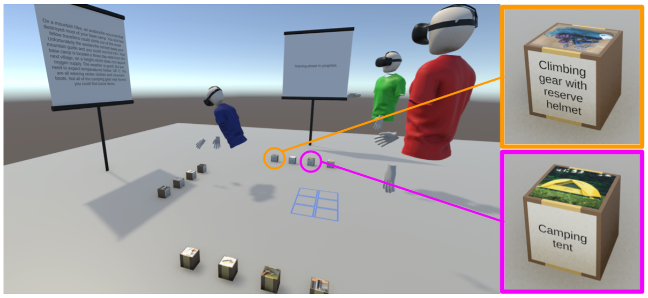


Figure 4: Survival Game (Scenario 3). Participants discuss and select a subset of items, represented as manipulable boxes, that will help them to survive in the given scenario.

Rendle et al. [31], each participant is provided with a list containing a subset of the items. This approach encourages all participants to contribute to the conversation, since the group can only find out which objects are available by exchanging information.

3.3.3 Extensions. Additional survival scenarios (with item lists and images) can be added to the implementation as required. To assess the performance of participants in the task, the selection can be evaluated against expert opinion [11, 21, 23]. In some realizations of the task, participants must rank the items by importance, which introduces more discussion and increases task completion time.

3.4 Implementation Notes

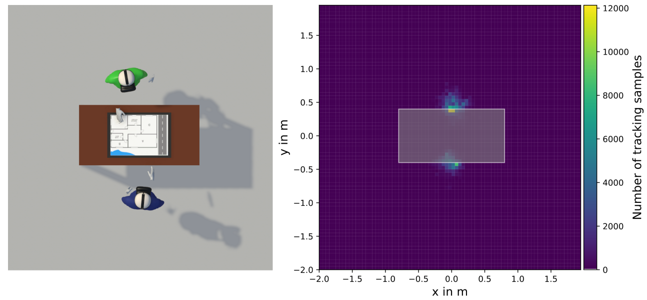
The scenario implementations included in the ICS-MR dataset are realized with the Unity game engine (version 6000.0.35f1). The VRSYS-Core framework[3] was used to create distributed, multi-user scenes, meaning that participants can join the study from remote locations. VRSYS-Core distributes scene state with Unity Netcode and supports voice communication through the *4Players ODIN* SDK. Unity scenes corresponding to each scenario are included in the repository. The scenes allow configuration of experiment properties, including duration and number of trials, as well as enabling selection of scenario variations where appropriate.

4 Behavioral Analysis of Scenarios

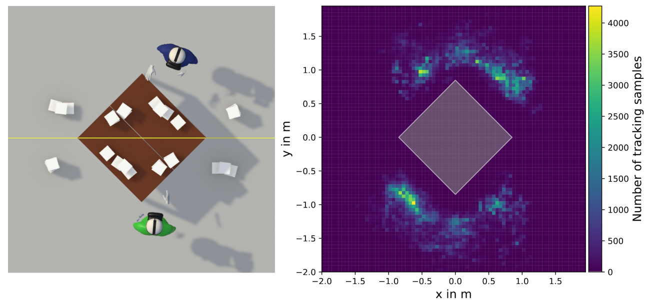
In this section, we analyze data captured during user studies that have applied the three presented scenarios (hereafter S1, S2 and S3), enabling the scenarios to be characterized in terms of their evoked behavior. The data was acquired during three studies with varying sample sizes: n=16 in dyads for S1 [14], n=32 in dyads for S2 [13], and n=66 in triads for S3 [12]. Please see the publications for details of the implementation, technical setup and research aims.

During the studies the position and orientation of the employed Head-Mounted Displays (HMDs) were recorded. The recordings were resampled to 25Hz as required. To quantify the amount of participant movement evoked by each of the scenario implementations, we analyze the distribution of user positions captured during all mediated conditions in each study, visualized in Figure 5 as 2D histograms. In addition, we calculate the mean translation and rotation speed, which indicates the degree to which scenarios affect

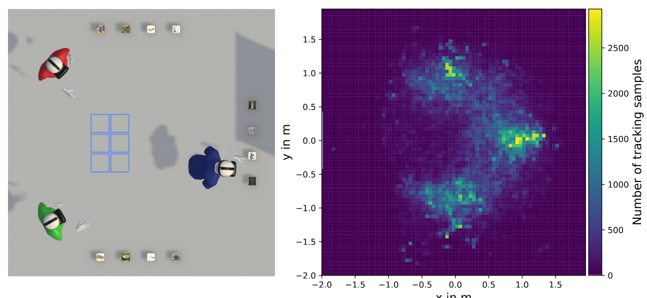
[3]https://github.com/vrsys/vrsys-core



(a) Scenario 1: Floor Plan Negotiation.



(b) Scenario 2: Spot the Difference.



(c) Scenario 3: Survival Game.

Figure 5: Top-down view of scenarios (left) and 2D histograms showing the participants' position distribution (right).

interaction behaviors (e.g. nodding, visual search). Distributions of mean head translation and rotation speed statistics (calculated per participant for each trial) are shown in Figure 6.

S1 evoked the least translational movement and exhibits static positioning near the floor plan. While this scenario shows little head rotation around the vertical axis (yaw), rotation speed around the lateral axis (pitch) is the highest across the scenarios, potentially indicating shifts of attentional allocation between the plan on the table and the standing participants. S2 evoked more movement in the virtual environment, as participants adjusted their position to investigate and compare the shapes. Overall head translation and rotation speeds are comparable to S1, although rotation speeds were higher for yaw rotation and lower for pitch rotation, reflecting the lateral layout of the boxes with respect to the participants. S3 also shows wide distribution of positions, as well as the highest mean translation and rotation speeds, due to the movement that was required to inspect survival items in the analyzed implementation.

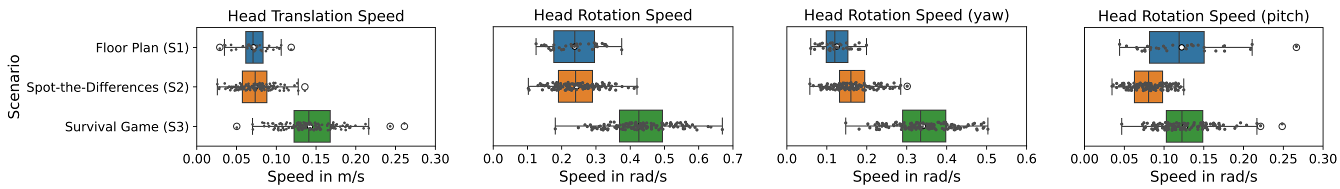


Figure 6: Movement data evoked by the three presented scenarios, through analysis of mean hand and head translation and rotation speed. Rotations are represented for all axes combined and separately for vertical (yaw) and lateral axis (pitch).

We note that the evoked behavior is not only subject to the administered scenario, but also depends on the chosen variation, the characteristics of the mediating system, and the number of users, among other context factors. The application context should therefore be considered when generalizing the presented evaluation.

5 Discussion of Scenario Properties

The scenarios comprising the ICS-MR dataset represent a range of different conversation and interaction situations. In this section, we discuss and compare the characteristics of each scenario, and address their suitability for use in various experiment contexts.

The *conversational structure* that results from a given scenario can vary between interactions with regular, predictable turn-taking on one end, and more natural, less regular conversations on the other. S2 is the most structured of the tasks, with participants likely to take a systematic approach to exchanging information when identifying differences. S3 may start with a semi-structured phase during which participants share nearby items, but becomes a more natural discussion, similar to S1, thereafter. Therefore, scenarios S1 and S3 are more appropriate for studies where the natural flow of conversational states is an important property.

S1 requires *conflict resolution* through negotiation. S2 does not introduce conflict between the participants, and is therefore purely collaborative. S3 represents a middle ground where conflict may occur, and if so must be resolved collaboratively, but is not prescribed by the scenario. S1 and S3 are therefore more fitting for evaluating the ability of a system to transmit cues that contribute to building trust and conveying emotions. While S2 allows *task performance* to be quantified by measuring task completion time, it is a poor measure of system effectiveness for S1 and S3, as conflict resolution strategies and the resulting duration may differ between groups.

As shown in Section 4, S2 and S3 require *locomotion* (i.e. moving through the scene) to solve the task. S3 also evoked a high degree of head *movement*, quantified as translation and rotation speed. Scenarios that evoke more motion are suitable for evaluating system support for navigation (e.g. with redirected walking) or spatial awareness (e.g. through directional audio cues).

In terms of *scalability*, S2 is the least suitable for larger groups (e.g. 4+ participants), since participants are more likely to be able to see others shapes as group size increases. S1 can be scaled by increasing the number of unassigned rooms, but conflict resolution with additional participants becomes more time-consuming. Scaling S3 is straightforward for up to 12 participants, at the cost of an extended duration of the discussion phase.

6 Summary and Outlook

This work presented the ICS-MR dataset, which is comprised of three conversation scenarios suitable for assessing communication in MR systems. The dataset contains all the materials required for incorporation into user studies, including participant instructions, Unity implementations, and resources for realization in real-world and video-conferencing reference contexts. To characterize the behavior evoked by each of the described tasks, we analyze movement recorded from recent applications. The results allow a detailed comparison of the scenarios that considers their scalability, applicability to research aims and resulting participant behaviors.

While the ICS-MR dataset contains a range of conversation scenarios, many other experimental scenarios have been employed in the context of assessing communication quality. The availability and analysis of additional scenarios not covered by this dataset, particularly those that are commonly used for assessing communication in non-immersive contexts, would provide the community with the tools and knowledge to assess communication in immersive systems in a reliable, reproducible manner. In addition to experimental scenarios, evaluation methods, including behavioral analysis approaches and standardized questionnaires, form a vital part of the assessment of quality of experience, and require careful development and application in parallel to conversation scenarios.

The scenarios, materials, and implementations provided by the ICS-MR dataset will serve to inform researchers when designing experiments for future communication evaluation studies, as well as representing a step towards standardization of conversational scenarios for novel immersive communication media.

Acknowledgments

This work was funded by Deutsche Forschungsgemeinschaft (DFG) under the project APlausE-MR (444831328) as part of DFG SPP2236-AUDICTIVE. Further, this work was supported by the project "Co-Presence of Humans and Interactive Companions for Seniors" funded by Carl-Zeiss-Stiftung within the "Durchbrüche 2020" program (https://www.carl-zeiss-stiftung.de/themen-projekte/uebersichtprojekte/detail/co-humanics) and the UKRI Engineering and Physical Sciences Research Council (EPSRC) under the 'Challenges in Immersive Audio Technology (CIAT)' grant EP/X032914/1. The contributions of Felix Immohr and Alexander Raake were mainly carried out at Technische Universität Ilmenau. The authors would also like to thank Christian Kehling, Steve Göring, Rakesh Rao Ramachandra Rao, Robert Gräfe, Victoria Meyer zur Heyde and the colleagues from AUDICTIVE for fruitful related discussions.

References

[1] Ahsan Abdullah, Jan Kolkmeier, Vivian Lo, and Michael Neff. 2021. Videoconference and embodied vr: communication patterns across task and medium. *Proceedings of the ACM on Human-Computer Interaction*, 5, CSCW2, (Oct. 2021), 1–29. doi:10.1145/3479597.

[2] Alex Adkins, Ryan Canales, and Sophie Jörg. 2024. Hands or controllers? how input devices and audio impact collaborative virtual reality. In *30th ACM Symposium on Virtual Reality Software and Technology* (VRST '24). ACM, (Oct. 2024), 1–12. doi:10.1145/3641825.3687718.

[3] Sahar Aseeri and Victoria Interrante. 2021. The influence of avatar representation on interpersonal communication in virtual social environments. *IEEE Transactions on Visualization and Computer Graphics*, 27, 5, (May 2021), 2608–2617. doi:10.1109/tvcg.2021.3067783.

[4] Zhuang Chang, Huidong Bai, Li Zhang, Kunal Gupta, Weiping He, and Mark Billinghurst. 2022. The impact of virtual agents' multimodal communication on brain activity and cognitive load in virtual reality. *Frontiers in Virtual Reality*, 3, (Nov. 2022). doi:10.3389/frvir.2022.995090.

[5] Carlos Cortés, Irene Viola, Jesús Gutiérrez, Jack Jansen, Shishir Subramanyam, Evangelos Alexiou, Pablo Pérez, Narciso García, and Pablo César. 2024. Delay threshold for social interaction in volumetric extended reality communication. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 20, 7, (Apr. 2024), 1–22. doi:10.1145/3651164.

[6] Elena Dzardanova, Vlasios Kasapakis, Damianos Gavalas, and Stella Sylaiou. 2022. Virtual reality as a communication medium: a comparative study of forced compliance in virtual reality versus physical world. *Virtual Reality*, 26, 2, 737–757. doi:10.1007/s10055-021-00564-9.

[7] Zhiwen Fan, Lingjie Zhu, Honghua Li, Xiaohao Chen, Siyu Zhu, and Ping Tan. 2021. Floorplancad: a large-scale cad drawing dataset for panoptic symbol spotting, (May 2021). arXiv: 2105.07147 [cs.CV]. doi:10.48550/ARXIV.2105.07147.

[8] Daniel Immanuel Fink, Moritz Skowronski, Johannes Zagermann, Anke Verena Reinschluessel, Harald Reiterer, and Tiare Feuchtner. 2024. There is more to avatars than visuals: investigating combinations of visual and auditory user representations for remote collaboration in augmented reality. *Proceedings of the ACM on Human-Computer Interaction*, 8, ISS, 540–568. doi:10.1145/3698148.

[9] Rachel L. Franz, Sasa Junuzovic, and Martez Mott. 2024. A virtual reality scene taxonomy: identifying and designing accessible scene-viewing techniques. *ACM Transactions on Computer-Human Interaction*, 31, 2, (Feb. 2024), 1–44. doi:10.1145/3635142.

[10] Maia Garau, Mel Slater, Vinoba Vinayagamoorthy, Andrea Brogni, Anthony Steed, and M. Angela Sasse. 2003. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '03). Association for Computing Machinery, New York, NY, USA, 529–536. ISBN: 1581136307. doi:10.1145/642611.642703.

[11] Jay Hall and W. H. Watson. 1970. The effects of a normative intervention on group decision-making performance. *Human Relations*, 23, 4, (Aug. 1970), 299–317. doi:10.1177/001872677002300404.

[12] Felix Immohr, Gareth Rendle, Christian Kehling, Anton Lammert, Steve Göring, Bernd Froehlich, and Alexander Raake. 2024. Subjective evaluation of the impact of spatial audio on triadic communication in virtual reality. In *2024 16th International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, (June 2024), 262–265. doi:10.1109/qomex61742.2024.10598292.

[13] Felix Immohr, Gareth Rendle, Anton Lammert, Annika Neidhardt, Victoria Meyer Zur Heyde, Bernd Froehlich, and Alexander Raake. 2024. Evaluating the effect of binaural auralization on audiovisual plausibility and communication behavior in virtual reality. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, (Mar. 2024), 849–858. doi:10.1109/vr58804.2024.00104.

[14] Felix Immohr, Gareth Rendle, Annika Neidhardt, Steve Göring, Rakesh Rao Ramachandra Rao, Stephanie Arevalo Arboleda, Bernd Froehlich, and Alexander Raake. 2023. Proof-of-concept study to evaluate the impact of spatial audio on social presence and user behavior in multi-modal vr communication. In *Proceedings of the 2023 ACM International Conference on Interactive Media Experiences* (IMX '23). ACM, (June 2023), 209–215. doi:10.1145/3573381.3596458.

[15] ITU-T Rec. P.1301. 2017. Subjective quality evaluation of audio and audiovisual multiparty telemeetings. (2017).

[16] ITU-T Rec. P.1305. 2016. Effect of delays on telemeeting quality. (2016).

[17] ITU-T Rec. P.1320. 2022. Quality of experience assessment of extended reality meetings. (2022).

[18] ITU-T Rec. P.800. 1996. Methods for subjective determination of transmission quality. (1996).

[19] ITU-T Rec. P.805. 2007. Subjective evaluation of conversational quality. (2007).

[20] ITU-T Rec. P.920. 2000. Interactive test methods for audiovisual communications. (2000).

[21] David W Johnson and Roger T Johnson. 1987. *Learning together and alone: Cooperative, competitive, and individualistic learning*. Prentice-Hall, Inc.

[22] Kangsoo Kim, Celso M. de Melo, Nahal Norouzi, Gerd Bruder, and Gregory F. Welch. 2020. Reducing task load with an embodied intelligent virtual assistant for improved performance in collaborative decision making. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, (Mar. 2020). doi:10.1109/vr46266.2020.00074.

[23] J Clayton Lafferty, Patrick M Eady, and J Elmers. 1974. The desert survival problem. *Experimental learning methods*.

[24] Marc Erich Latoschik, Daniel Roth, Dominik Gall, Jascha Achenbach, Thomas Waltemate, and Mario Botsch. 2017. The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology* (VRST '17). Association for Computing Machinery, New York, NY, USA. ISBN: 9781450355483. doi:10.1145/3139131.3139156.

[25] Harold J. Leavitt. 1951. Some effects of certain communication patterns on group performance. *The Journal of Abnormal and Social Psychology*, 46, 1, (Jan. 1951), 38–50. doi:10.1037/h0057189.

[26] Karan Mitra, Henrique Souza Rossi, Justin Gavrell, and Christer Åhlund. 2025. Qoe assessment of cloud-based social extended reality applications over heterogeneous access networks. In *2025 IEEE 22nd Consumer Communications & Networking Conference (CCNC)*. IEEE, (Jan. 2025), 1–2. doi:10.1109/ccnc54725.2025.10975911.

[27] Sebastian Möller. 2000. *Assessment and Prediction of Speech Quality in Telecommunications*. Springer US. ISBN: 9781475731170. doi:10.1007/978-1-4757-3117-0.

[28] Paul M Nemiroff and William A Pasmore. 1975. Lost at sea: a consensus-seeking task. *The annual handbook for group facilitators*, 28–34.

[29] Pablo Pérez, Ester Gonzalez-Sosa, Jesús Gutiérrez, and Narciso García. 2022. Emerging immersive communication systems: overview, taxonomy, and good practices for qoe assessment. *Frontiers in Signal Processing*, 2, (July 2022). doi:10.3389/frsip.2022.917684.

[30] Alexander Raake. 2006. *Speech Quality of VoIP: Assessment and Prediction*. Wiley, (Sept. 2006). ISBN: 9780470033005. doi:10.1002/9780470033005.

[31] Gareth Rendle, Felix Immohr, Christian Kehling, Anton Benjamin Lammert, Adrian Kreskowski, Karlheinz Brandenburg, Alexander Raake, and Bernd Froehlich. 2025. Influence of Audiovisual Realism on Communication Behaviour in Group-to-Group Telepresence. In *2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*.

[32] Thomas Robotham, Daniela Rebmann, Dominik O Fintineanu-Anghelescu, Alexander Raake, and Emanuël AP Habets. 2024. Cgi scenes for interactive audio research and development: cave, cinema, and mansion. In *Audio Engineering Society Conference: AES 2024 International Audio for Games Conference*. Audio Engineering Society.

[33] Janto Skowronek. 2017. Quality of experience of multiparty conferencing and telemeeting systems. en. (2017). doi:10.14279/DEPOSITONCE-5811.

[34] Janto Skowronek et al. 2022. Quality of experience in telemeetings and videoconferencing: a comprehensive survey. *IEEE Access*, 10, 63885–63931. doi:10.1109/access.2022.3176369.

[35] Harrison Jesse Smith and Michael Neff. 2018. Communication behavior in embodied virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18). ACM, (Apr. 2018), 1–12. doi:10.1145/3173574.3173863.

[36] Sam Watson, A Josefine Munch Sørensen, and Ewen MacDonald. 2020. The effect of conversational task on turn taking in dialogue. In *International Symposium on Auditory and Audiological Research: Auditory Learning in Biological and Artificial Systems*. The Danavox Jubilee Foundation, 61–68.

[37] Wenming Wu, Xiao-Ming Fu, Rui Tang, Yuhan Wang, Yu-Hao Qi, and Ligang Liu. 2019. Data-driven interior plan generation for residential buildings. *ACM Transactions on Graphics*, 38, 6, (Nov. 2019), 1–12. doi:10.1145/3355089.3356556.