QoE Evaluation of Remote Physiotherapy in Volumetric Video and Video-Based Real-Time Communication

Ashutosh Singla*, Irene Viola*, Jack Jansen*, Pablo Cesar*†
*Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
†TU Delft, Delft, The Netherlands
Email: irene.viola@cwi.nl, pablo.cesar@cwi.nl

***Abstract*—In recent years, video conferencing platforms have become powerful tools for remote communication. There has also been an increase in the use of VR systems for communication. However, very few of these systems utilize photorealistic human representation. This paper investigates the strengths, challenges, and limitations of a novel 3D communication prototype (VR2Gather) and a well-established video conferencing system (Zoom). Specifically, we explore whether the 3D communication prototype can achieve comparable performance levels in a remote physiotherapy use case. By assessing various aspects, such as audio-visual quality, presence, and interaction, we aim to determine if the current prototype is comparable with commercial systems in some dimensions while exceeding expectations in others. Our results indicated that VR2Gather has the potential for a better sense of connection and higher concentration. However, challenges like improving 3D rendering quality and communication ease still need to be overcome to make it suitable for physiotherapy.**

***Index Terms*—Social XR, Point Cloud, Zoom, Physiotherapy, Communication, eXtended Reality**

I. Introduction

Over the years, video communication has evolved into modern platforms such as Zoom, WhatsApp, WebEx, and Microsoft Teams. These platforms enable us to communicate in real-time with multiple persons in Full High Definition (FHD) quality [1]. Since the COVID-19 outbreak, online video communication and meetings have grown in importance, finding applications in various domains [2], [3]. However, traditional 2D video communication platforms are limited in immersiveness, spatial presence, and the ability for users to feel like they are sharing a space together and rely on 2D visual representations [4]. Recent advancements in eXtended Reality (XR) address these limitations by offering 3D representations of individuals and natural interaction, enhancing the sense of presence and creating the impression of sharing a space together [4], [5].

XR collaboration systems are still in the early stages of adoption and development [6]. Exploring the impact of XR on use cases that require good spatial understanding and visual representation besides audio quality, such as physiotherapy, is important to understand where XR technology stands today compared to classical video communication. To ensure a good physiotherapy session, it is important to have real-time communication with minimal latency, natural interactions, and the ability of patients to observe precise movements from different angles and distances to fully understand them. While 2D video communication platforms fulfill the requirement of real-time communication, they lack immersive interactivity and spatial presence. XR platforms [7], [8] address the limitations by providing all of the above requirements, offering enhanced spatial presence and interactivity [4], [5].

Some studies in the literature evaluate physiotherapy [9], TaiChi [10] training and remote healthcare [11] in social VR systems. However, none of these works explore the user experience with different user representations. Moreover, no studies have investigated how XR communication and collaboration tools perform in comparison to their 2D counterparts in physiotherapy. Hence, this paper explores whether a 3D communication prototype (VR2Gather) can deliver a user experience comparable to that of well-established video conferencing tools like Zoom. We selected Zoom as the 2D video conferencing tool due to its widespread use and popularity[1]. In this paper, we make two primary contributions: (1) We design an experiment to thoroughly investigate the use of two systems—VR2Gather and Zoom—for a physiotherapy use case. We provide a detailed protocol tailored to this scenario, which will serve as input to the International Telecommunication Union (ITU) standard P.IXC[2]. (2) We conduct the experiment and present a comprehensive analysis of key metrics, including quality of interaction, presence, cybersickness, workload, and visual quality for both systems.

Our results indicated that VR2Gather has the potential for a better sense of connection and higher concentration. However, challenges like improving 3D rendering quality and communication ease still need to be overcome to make it suitable for physiotherapy.

II. Related Work

A. Social XR Systems

Some social VR systems represent users with simple avatars that consist of the head, hands, and upper body [12], [13]. In contrast, systems like embodVR [14] track full body using

This work was supported through the European Commission Horizon Europe program, under the grant agreement 101070109, *TRANSMIXR* https://transmixr.eu/. Funded by the European Union.

[1]https://www.emailtooltester.com/en/blog/video-conferencing-market-share/
[2]https://www.itu.int/ITU-T/workprog/wp_item.aspx?isn=20848

Motion-capture suits and the OptiTrack system. These systems enable real-time communication among participants, who are represented as avatars. These systems lack photorealistic representation and tracking of facial features. Social VR system like ImmerTai [10] focuses on Chinese Taichi training. Their system is not live, as students learn Taichi by observing the pre-recorded video of the Taichi expert's motion. The system lacks real-time feedback and interaction capabilities.

Some XR communication systems also incorporate photorealistic representations [7], [8], [15]–[18]. The XR systems developed in [15] and [17] use mesh representations of users. Beck et al. [16] telepresence system uses projection-based 3D screens and represents users as photorealistic 3D video avatars. These conferencing systems allow to have real-time remote participant communication. In [7], an asymmetric collaboration system is presented. The teacher is presented in photorealistic representation, where the student is represented as a viewpoint avatar within the shared virtual environment. VR2Gather [8], [18] is a symmetric communication system where both users are presented in photorealistic representations (point cloud). They can interact with each other in a virtual environment in real-time. Additionally, it is open source and easy to modify for use. Hence, we selected VR2Gather as an XR communication platform for our experiment.

B. Evaluation Methods

Several studies measure user experience in social VR systems using task-specific methodologies [9], [10], [13], [19]–[22]. These studies use different test protocols tailored to their specific research questions or task requirements. Common evaluation constructs include presence [10], [13], [19]–[22] and Quality of Interaction (QoI) [9], [13], [20]–[22] to understand participants' sense of being there and the effectiveness of user communication. Visual quality and overall QoE are also evaluated to understand users' experience of visual representations [9], [19], [21]. Additionally, some researchers have also measured users' Task-Related Experience while completing tasks in social VR environments [9], [21].

For our task, we adapted our test protocols based on these works [19], [21] and selected questionnaires to measure presence, QoI, audio and video quality, cybersickness and to assess workload.

C. Different Representations

In [20], Li et al. compare photo-sharing experiences on three platforms: face-to-face (F2F), Skype, and Facebook Spaces (FS). They found that Social VR (FS) can closely approximate the F2F photosharing experience. In [21], two platforms, HMD and 2D screens, were evaluated for watching a virtual movie together. All four users were represented with photorealistic representations. Their results indicated that HMD users experienced greater presence and immersion, while screen users reported lower workload and more ease in exploring the environment. De Simone et al. [22] compared video-watching experiences on F2F, FS, and in a video-based Social VR system (Photorealistic). Their results indicated that the video-based social VR system provides a similar experience to F2F. In [23], three different modes of interaction were studied F2F, 2D video conferencing, and Horizon Workrooms (avatars) in the idea generation and decision-making task. They observed that F2F outperforms the other two modes of interaction. However, communication in VR has some advantages over 2D, especially in terms of collaboration and engagement.

No studies in the literature compare different representations of participants in physiotherapy [9]–[11] in social VR systems. In our work, we explore two platforms: a 3D communication prototype with photorealistic user representations and the Zoom system, which uses 2D video representations to learn the exercises.

III. Experimental Setup

The experiment consisted of two sessions, each using a different test system. The sessions differed only in terms of which system was used first. The participants were equally divided into two groups: the first group performed the exercises using the HMD first and then Zoom. In contrast, the second group completed the exercises using Zoom and then the HMD. This design was used to control for session order effects.

A. Selection of Exercises

A real physiotherapist initially selected the exercises, excluding any balancing exercises. This is because participants would wear HMDs, making it hard to balance without seeing the outside world. We conducted pre-tests and found that these initial exercises were too simple for participants to perform, requiring no physical or mental effort. To address this, we consulted a Kung-Fu expert to introduce more complicated exercises. This complexity was designed to encourage participants to interrupt the physiotherapist (referred to as the confederate user), ask questions, or seek clarification to perform the exercises correctly.

For this experiment, the final selection included six different exercises. Two of these were normal exercises, which some participants might already be familiar with. The remaining exercises were selected from Kung-Fu to introduce intentionally complex movements, ensuring participants would experience some difficulty. On average, the time taken to complete one exercise was 2–3 minutes. These exercises were evenly distributed between the two sessions: Zoom and VR2Gather. Participants randomly performed one regular exercise and two Kung-Fu exercises in each session. A separate exercise was used for training.

More information about the exercises can be found here: https://github.com/cwi-dis/vr2gather-zoom-physio-qoe

B. Technical Setup

To perform the exercises in the Social XR environment, we selected VR2Gather [8], [18]. This software allows multiple people to be present in the same virtual space. VR2Gather uses CWIPC[3] system to generate the point cloud using Azure

[3]https://github.com/cwi-dis/cwipc

Fig. 1: Participant's view on the screen and a real-life view of participant replicating confederate user's movement.

Kinect or Intel Realsense cameras. Our experiment used four Azure Kinect cameras to create point clouds for the physiotherapist and participants. Fig. 1 shows the setup used to capture participants' movements and a screen displaying the participant's view.

On top of VR2Gather, a virtual room with grey walls and a door was built, including furniture such as tables and flower pots. This virtual room resembled the participant's physical room in terms of dimensions. The tables and flower pots were strategically placed to align with the Kinect cameras' positions so that participants would not move closer to the cameras. This space served as the meeting point where the physiotherapist and participants interacted and where participants learned the exercises. The distance between participants and the physiotherapist in the virtual space was 2 meters. This distance was chosen to ensure that both parties were not too close to each other and could move within the physical space to observe each other's movements. Once the exercise was completed, the physiotherapist pressed a button, and both were transported to another room where the participant filled in questions related to QoI, presence, sickness, and quality online. After completing the questionnaires, they were automatically transported back to the exercise room.

During the pre-tests, we noticed that the back-and-forth transportation and filling out questionnaires in the virtual environment were too much for the participants. Therefore, we decided to simplify the process by using only one exercise room. After the exercise, participants removed the HMD to complete the questionnaires on paper. The final exercise virtual room had grey walls and a door but no furniture. This simple setup ensured minimal distractions and allowed participants to focus entirely on the exercise session. The VR2Gather application ran on two identically configured Windows 10 machines, each equipped with an i9 CPU @ 3.6GHz, 64GB RAM, and a Nvidia GeForce RTX 2080 Ti GPU.

Participants and physiotherapist wore Quest Pro HMD, which were wirelessly connected to the machine. However, due to the use of HMDs, some facial information, particularly from the face region, was lost. During the pre-tests, participants complained that the lack of visibility of the physiotherapist's face made it difficult to understand the exercises. To address this, VR2Gather was used in an asymmetric setup, where only the participant was immersed in a 6 DoF environment using an HMD, while the physiotherapist only wore headphones, ensuring the participants could see the physiotherapist's face and facial expressions.

To perform exercises in the 2D video communication platform, we selected Zoom[4]. In our experiment, we used Full HD webcams to capture the participants and the physiotherapist, ensuring their movements were clearly visible. The physiotherapist and participants were instructed to maintain an approximate distance of 2 meters from the screen to give the other person an almost complete view of their body. Headphones equipped with microphones were used for audio transmission and reception. The headphones were connected to the machine with wires, so there was no additional delay in audio transmission. Participants viewed the physiotherapist's video feed on a 165 cm monitor to effectively observe and replicate the exercises.

C. Test Method

Before starting the experiment, each participant was screened for correct visual acuity using Snellen charts (20/25) and for color vision using Ishihara charts. After passing the pre-screening, participants were given written instructions to understand the experiment's aim. They filled in the Simulator Sickness Questionnaire (SSQ) [24] before starting the experiment, at the end of the first session, and at the end of the experiment. Furthermore, the NASA Task Load Index (NASA-TLX)[5] was administered at the end of the first session and again at the end of the experiment. To evaluate the presence, Quality of Interaction, and perception of interruptions, we included questions from [19], which were also asked at the end of the first session and the experiment.

When performing the exercises using an HMD, participants removed the device and filled out a questionnaire on paper regarding their experience, audio, and video quality, using a 5-point Absolute Category Rating scale [25], [26]. They also rated their ease of communication using the system [25], [27]. Once they completed the questionnaire, they could take a short one-minute break or continue directly with the following exercise if they felt ready. Before each exercise, it was ensured that their point cloud representation was perfectly aligned with their body. Once the first session was complete, they had a 10-minute break before moving on to the next system.

When learning the exercises in Zoom, participants only wore wired headphones and a large screen was placed in front of them so they could see the physiotherapist from a distance. After the second session, the experimenter conducted a brief interview with them to ask specific questions about their experience during the entire experiment. The entire data collection is available here: https://github.com/cwi-dis/vr2gather-zoom-physio-qoe

[4]https://zoom.us/

[5]https://humansystems.arc.nasa.gov/groups/tlx/

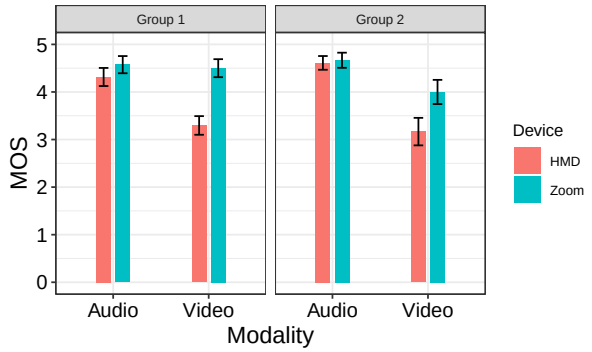


Fig. 2: MOS for audio and video quality.

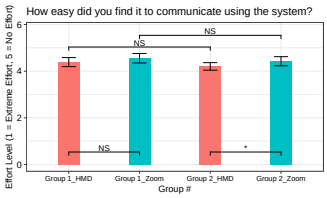


Fig. 3: MOS for ease of communication.

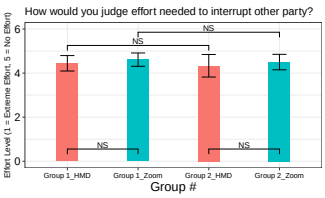


Fig. 4: MOS for an effort to interrupt others.

The total duration of the experiment was $\approx$ 45 minutes. When performing exercises with an HMD, participants entered a virtual room where they and the physiotherapist were represented as point clouds. Participants were instructed to move closer to the physiotherapist if they needed a better view of the movements in any system. They were encouraged to ask questions or interrupt the physiotherapist if instructions were unclear. In total, 36 participants (25 males and 11 females) participated in the experiment. These participants were equally divided into two groups. Their average age was 34 years, with a median age of 27.5 years.

IV. Experimental Results

We performed the Shapiro-Wilk normality tests on the ratings for each question and each group. The results suggested that the ratings are not normally distributed (p-values < 0.05).

A. Audio and Video Quality

Figure 2 shows the mean opinion scores (MOS) with their 95% confidence intervals (CI) for audio and video quality averaged over all exercises and participants for each group. It can be observed that Zoom and VR2Gather have a comparable level of audio quality, irrespective of the group. For video quality, Zoom is always rated higher than VR2Gather, which is expected because the 3D quality of the confederate user (Physiotherapist) has some visual artifacts, and some fine details were missing. There are visible distortions on the face and body, and the technology is still developing. VR2Gather's point cloud format limits visual quality due to low resolution, lighting sensitivity, and unstable depth data. Rendering issues lead to flickering and loss of detail, undermining the visual clarity essential for observing physiotherapy movements. We conducted a Mann-Whitney U test (Wilcoxon rank-sum test) to compare Zoom and HMD (VR2Gather) within and between groups and for audio and video quality. We observed both between-group and within-group differences. For video quality, Zoom provides statistically significantly better quality than HMD for both groups, with p-value < 0.01. A significant difference was observed between both groups for Zoom ($p < 0.01$). This highlights that session order may have influenced the evaluation of Zoom, as participants in Group #2 rated video quality significantly lower. For audio quality, a significant difference was observed for Group #1 ($p < 0.05$). A significant difference was also observed between Group #1 and Group #2 for HMD ($p < 0.05$).

Figures 3 and 4 show the MOS for ease of communication and for the effort required to interrupt others for each group. Both groups show a similar trend for the effort levels, which are lower when using Zoom than HMD. This could be attributed to the fact that most people nowadays are more familiar with video conferencing on Zoom than with HMDs. Additionally, it could be possible that wearing an HMD for a long time may lead to discomfort, thereby increasing the perceived effort. Using the Mann-Whitney U test, we observed no significant between-group and within-group differences, except for a significant effect of the device on Group #2 ($p < 0.05$) for ease of communication.

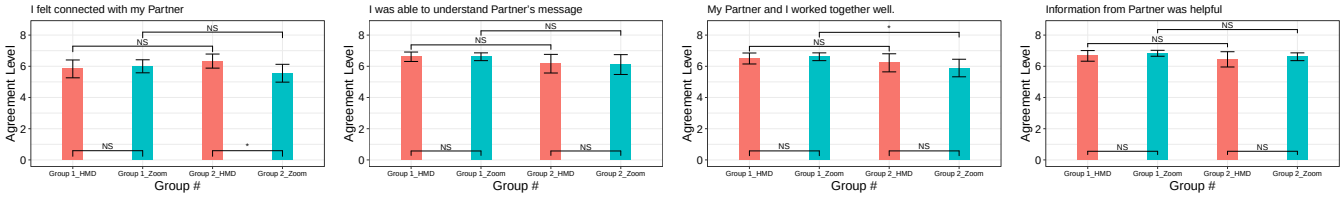
B. Quality of Interaction

Figures 5 a, b, c, and d show the MOS for different questions related to the Quality of Interaction, averaged over all participants for each group. Figure 5a indicates that, for Group #1, participants felt slightly more connected using Zoom; however, this difference is nonsignificant. In contrast, HMD provides a stronger and statistically significant sense of connection in Group #2, with $p < 0.05$. The possible reason could be that they felt they were sharing the same space and could see the 3D representation of the Confederate user, which made them feel more connected to the Confederate user.

From Fig. 5b, it can be observed that HMD and Zoom provide similar levels of communication clarity, irrespective of the group. Users could understand or follow the confederate user's instructions with both devices. We observed no significant differences between the group and the within-group.

From Fig. 5c, HMD promotes better collaboration for Group #2. This could be attributed to the fact that participants and the Confederate user can observe each other from different angles or distances, similar to real life. However, this is not possible with Zoom. The same effect was not observed for Group #1. A significant difference was observed between Group #1 and Group #2 for Zoom ($p < 0.05$), with Group #2 participants providing significantly lower ratings than those in Group #1.

Figure 5d indicates that Zoom is more effective in providing information for both Group #1 and Group #2. This could be attributed to the visual quality of the Confederate user (see Fig. 2). While doing exercises, participants need to observe the confederate user very closely to replicate the movements. In VR2Gather, the 3D representation of the confederate user contains some visual artifacts, and possibly some important details are also missing. This is not the case with Zoom, where participants can see the physiotherapist's movements.



(a) (b) (c) (d)

Fig. 5: MOS for Quality of Interaction questions for different groups for Zoom and HMD.

We observed no significant differences between the group and the within-group.

C. Presence

Figures 6 a, b, and c show the MOS with their 95% confidence intervals for different questions related to the Presence, averaged over all participants for each group. Figure 6a indicates that participants found their experience in Zoom to be more aligned with real-world experiences, irrespective of the group. This could be attributed to participants being more familiar with Zoom than HMD or the Confederate user's visual representation being clearer than in HMD. For Group #2, a significant difference was observed ($p < 0.05$).

It can be interpreted from Fig. 6b that, for Group #1, participants have a slightly higher concentration level while performing exercises using HMD. However, this difference is only marginally significant (p-value $= 0.078$). Participants in this group may have been able to focus more on the tasks because, after wearing HMD, they became unaware of their surroundings. In contrast, for Group #2, Zoom provides higher concentration levels. A possible reason could be the familiarity and simplicity of Zoom or the additional mental load introduced by wearing HMD, which may have made it more challenging for them to concentrate. This difference is significant, with a p-value < 0.05. It is interesting to note that concentration level significantly impacts HMD between groups.

From Fig. 6c, it can be observed that HMD and Zoom provide similar levels of confidence in completing the task correctly for Group #1. In contrast, for Group #2, participants feel more confident completing the task correctly in HMD. The possible reason could be that participants could observe the Confederate user from different angles and distances, which may have enhanced their sense of presence and confidence in task performance. No significant differences between the group and the within-group were observed.

D. Summary

The results show that Zoom provides better visual representation, lower communication effort, and participants found their experience with Zoom to be more aligned with real-world interactions. A possible reason for this is that, after COVID, most people are familiar with Zoom. The limitation is that participants cannot view the other person from multiple angles and are always aware of their surroundings.

In contrast, VR2Gather provides a better sense of connection with the other person. Furthermore, some participants showed a tendency to feel a higher concentration level and greater confidence in completing tasks when using HMD. However, VR2Gather is limited by the quality of 3D rendering, the discomfort caused by the HMD, and the fact that most people have not used an HMD before.

These findings emphasize the strengths, limitations, and challenges of the novel 3D communication prototype (VR2Gather) in comparison to the well-established video conferencing system (Zoom).

E. Cybersickness and Task Load

We evaluated participants' experience using two measures: Simulator Sickness and NASA Task Load Index (NASA TLX), for both Group #1 and #2 across devices (HMD and Zoom).

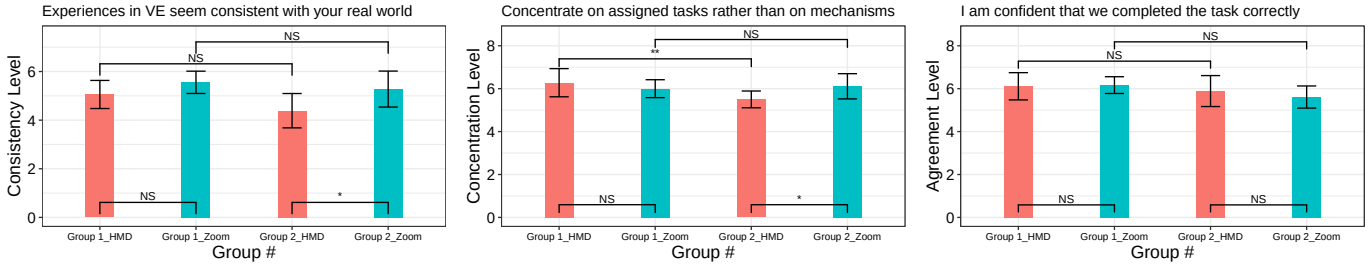
Simulator sickness scores for all factors, Disorientation, Nausea, Oculomotor, and Total Score, were consistently higher for the HMD compared to Zoom, irrespective of the group. A Mann-Whitney U test on Participants' Total Score for both devices found the difference insignificant for both groups.

Similarly, NASA TLX scores across all factors, Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration, were comparable between HMD and Zoom, irrespective of the group. A Mann-Whitney U test on each factor for both devices found the difference insignificant for both groups.

V. Discussion and Conclusions

This paper explores the limitations, challenges, and strengths of the XR communication prototype (VR2Gather) and a video conferencing system (Zoom) in a physiotherapy use case. An experiment was designed with a tailored protocol to evaluate key metrics including QoI, presence, audio and visual quality, cybersickness, and workload. The results showed that audio quality is comparable in both systems. However, the video quality is significantly better in Zoom as participants can see the confederate user more clearly. QoI-related analysis showed similar levels of communication clarity, collaboration, and the ability to deliver information effectively, with no significant differences observed between Zoom and VR2Gather. In some cases, VR2Gather demonstrated a statistically significant sense of connection compared to Zoom. Presence-related results were mixed; both groups preferred different devices for higher concentration levels and confidence in completing the task correctly. The experience in Zoom was found to be more aligned with real-world experiences.

If the XR communication prototype is expected to surpass the effectiveness of video conferencing systems, future work should concentrate on improving the 3D rendering quality of



(a) (b) (c)

Fig. 6: MOS for Presence questions for different groups for Zoom and HMD.

users and addressing the discomfort caused by the HMD to improve the user experience.

REFERENCES

[1] Hyunseok Chang, Matteo Varvello, and et al., "A tale of three videoconferencing applications: Zoom, webex, and meet," *IEEE/ACM Transactions on Networking*, vol. 30, no. 5, pp. 2343–2358, 2022.

[2] Hiroki Kashiwazaki, Takuro Ozaki, Hajime Shimada, Yusuke Komiya, Eisaku Sakane, and Kazuhiro Mishima et al., "Japanese activities to bring online academic meetings against covid-19: How we learned to stop worrying and love the online meetings," in *Proceedings of the 2021 ACM SIGUCCS Annual Conference*, New York, NY, USA, 2021, SIGUCCS '21, p. 54–59, Association for Computing Machinery.

[3] Janto Skowronek, Alexander Raake, Gunilla H. Berndtsson, and et al., "Quality of experience in telemeetings and videoconferencing: A comprehensive survey," *IEEE Access*, vol. 10, pp. 63885–63931, 2022.

[4] Caroline Kuhne, Eda D. Kecelioglu, Steven Maltby, Rebecca J. Hood, Brendon Knott, Elizabeth Ditton, Frederick Rohan Walker, and Murielle G. Kluge, "Direct comparison of virtual reality and 2d delivery on sense of presence, emotional and physiological outcome measures," *Frontiers in Virtual Reality*, vol. 4, pp. 1211001, 2023.

[5] Simon NB Gunkel, Sylvie Dijkstra-Soudarissanane, Hans M Stokking, and Omar A Niamut, "From 2d to 3d video conferencing: Modular rgb-d capture and reconstruction for interactive natural user representations in immersive extended reality (xr) communication," *Frontiers in Signal Processing*, vol. 3, pp. 1139897, 2023.

[6] Alexander Schäfer, Gerd Reis, and Didier Stricker, "A survey on synchronous augmented, virtual, andmixed reality remote collaboration systems," *ACM Computing Surveys*, vol. 55, no. 6, pp. 1–27, 2022.

[7] Jason W. Woodworth, Sam Ekong, and Christoph W Borst, "Virtual field trips with networked depth-camera-based teacher, heterogeneous displays, and example energy center application," in *IEEE Virtual Reality (VR)*, 2017, pp. 471–472.

[8] Irene Viola, Jack Jansen, Shishir Subramanyam, Ignacio Reimat, and Pablo Cesar, "Vr2gather: A collaborative, social virtual reality system for adaptive, multiparty real-time communication," *IEEE MultiMedia*, vol. 30, no. 2, pp. 48–59, 2023.

[9] Shishir Subramanyam, Irene Viola, Jack Jansen, Evangelos Alexiou, Alan Hanjalic, and Pablo Cesar, "Evaluating the impact of tiled user-adaptive real-time point cloud streaming on vr remote communication," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, pp. 3094–3103.

[10] Xiaoming Chen, Zhibo Chen, Ye Li, Tianyu He, Junhui Hou, Sen Liu, and Ying He, "Immertai: Immersive motion learning in vr environments," *Journal of Visual Communication and Image Representation*, vol. 58, pp. 416–427, 2019.

[11] Klara Nahrstedt, "3d teleimmersion for remote injury assessment," in *Proceedings of the 2012 International Workshop on Socially-Aware Multimedia*, New York, NY, USA, 2012, SAM '12, p. 21–24, Association for Computing Machinery.

[12] Sebastian J Friston, Ben J Congdon, David Swapp, Lisa Izzouzi, Klara Brandstätter, and Daniel Archer et al., "Ubiq: A system to build flexible social virtual reality experiences," in *Proceedings of the 27th ACM symposium on virtual reality software and technology*, 2021, pp. 1–11.

[13] Felix Immohr, Gareth Rendle, Annika Neidhardt, Steve Göring, Rakesh Rao Ramachandra Rao, and Stephanie Arevalo Arboleda et al., "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*, New York, NY, USA, 2023, IMX '23, p. 209–215, Association for Computing Machinery.

[14] Harrison Jesse Smith and Michael Neff, "Communication behavior in embodied virtual reality," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2018, CHI '18, p. 1–12, Association for Computing Machinery.

[15] Rufael Mekuria, Michele Sanna, Stefano Asioli, Ebroul Izquierdo, Dick C. A. Bulterman, and Pablo Cesar, "A 3d tele-immersion system based on live captured mesh geometry," in *Proceedings of the 4th ACM Multimedia Systems Conference*, 2013, pp. 24–35.

[16] Stephan Beck, André Kunert, Alexander Kulik, and Bernd Fröhlich, "Immersive group-to-group telepresence," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, pp. 616–625, 2013.

[17] Nikolaos Zioulis, Dimitrios Alexiadis, Alexandros Doumanoglou, Georgios Louizis, Konstantinos Apostolakis, Dimitrios Zarpalas, and Petros Daras, "3d tele-immersion platform for interactive immersive experiences between remote users," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 365–369.

[18] Jack Jansen, Thomas Röggla, Silvia Rossi, Irene Viola, and Pablo Cesar, "Open-sourcing vr2gather: A collaborative social vr system for adaptive multi-party real time communication," in *Proceedings of the 32nd ACM International Conference on Multimedia*, New York, NY, USA, 2024, MM '24, p. 11210–11213, Association for Computing Machinery.

[19] Carlos Cortés, Irene Viola, Jesús Gutiérrez, Jack Jansen, Shishir Subramanyam, Evangelos Alexiou, Pablo Pérez, Narciso García, and Pablo César, "Delay threshold for social interaction in volumetric extended reality communication," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 20, no. 7, pp. 1–22, 2024.

[20] Jie Li, Yiping Kong, Thomas Röggla, Francesca De Simone, Swamy Ananthanarayan, Huib de Ridder, Abdallah El Ali, and Pablo Cesar, "Measuring and understanding photo sharing experiences in social virtual reality," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–14.

[21] Jie Li, Shishir Subramanyam, Jack Jansen, Yanni Mei, Ignacio Reimat, Kinga Ławicka, and Pablo Cesar, "Evaluating the user experience of a photorealistic social vr movie," in *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2021, pp. 284–293.

[22] Francesca De Simone, Jie Li, Henrique Galvan Debarba, Abdallah El Ali, Simon N.B Gunkel, and Pablo Cesar, "Watching videos together in social virtual reality: An experimental study on user's qoe," in *2019 IEEE Conference on virtual reality and 3d user interfaces (VR)*. IEEE, 2019, pp. 890–891.

[23] Gregorio Macchi and Nicola De Pisapia, "Virtual reality, face-to-face, and 2d video conferencing differently impact fatigue, creativity, flow, and decision-making in workplace dynamics," *Scientific Reports*, vol. 14, no. 1, pp. 10260, 2024.

[24] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," *The International Journal of Aviation Psychology*, vol. 3, no. 3, pp. 203–220, 1993.

[25] ITU-R BS.1284-2, *Recommendation ITU-R BS.1284-2: General methods for the subjective assessment of sound quality*, 2019.

[26] ITU-T P.910, *Recommendation ITU-T P.910: Subjective video quality assessment methods for multimedia applications*, 2008.

[27] ITU-T P.1305, *Recommendation ITU-T P.1305: Effect of delays on telemeeting quality*, 2016.