

Evaluating the User Experience of a Photorealistic Social VR Movie

Jie Li¹ Shishir Subramanyam¹ Jack Jansen¹ Yanni Mei¹ Ignacio Reimat¹
Kinga Ławicka¹ Pablo Cesar^{1,2}

¹Centrum Wiskunde & Informatica, Amsterdam, The Netherlands

²Delft University of Technology, Delft, The Netherlands

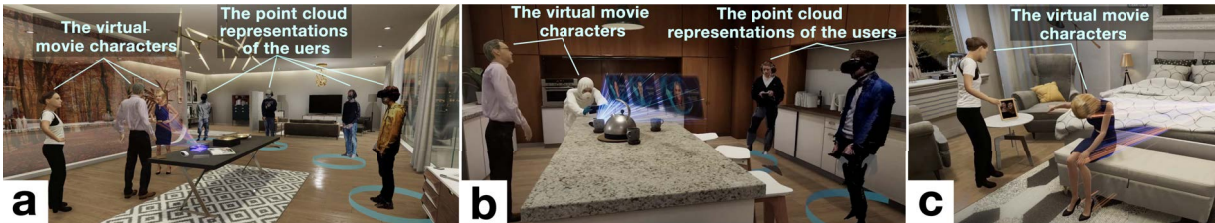


Figure 1: The luxury apartment of the victim Elena Armova: (a) the living room with four users represented as point clouds and three virtual characters: (from left to right) Rachel, Sarge and the hologram of Elena; (b) the kitchen with two users, Sarge and Evans; (c) the bedroom with Rachel and Elena.

ABSTRACT

We all enjoy watching movies together. However, this is not always possible if we live apart. While we can remotely share our screens, the experience differs from being together. We present a social Virtual Reality (VR) system that captures, reconstructs, and transmits multiple users' volumetric representations into a commercially produced 3D virtual movie, so they have the feeling of "being there" together. We conducted a 48-user experiment where we invited users to experience the virtual movie either using a Head Mounted Display (HMD) or using a 2D screen with a game controller. In addition, we invited 14 VR experts to experience both the HMD and the screen version of the movie and discussed their experiences in two focus groups. Our results showed that both end-users and VR experts found that the way they navigated and interacted inside a 3D virtual movie was novel. They also found that the photorealistic volumetric representations enhanced feelings of co-presence. Our study lays the groundwork for future interactive and immersive VR movie co-watching experiences.

Index Terms: Human-centered computing [Human computer interaction (HCI)]; HCI design and evaluation methods—User studies; Human-centered computing [Human computer interaction (HCI)]; Interaction paradigms—Virtual reality.

1 INTRODUCTION

Hunkering down on the couch and watching a movie together with friends or family is not only a nice cozy thing to do, but also enables us to share emotions, increase engagement and social bonds with people we love [12]. However, this is not always possible if we live apart. Although people at distance can text each other, video call and share their screens, or even use video stream synchronization

applications (e.g., *Teleparty*¹), it is still far from the feeling of being together. As an emerging immersive remote communication tool, social VR has the potential to afford face-to-face-like social interactions than video calls, enabling users to feel co-present and interact with virtual objects [7, 24, 31, 35]. The virtual space can be a computer-generated 3D scene or a 360° scene captured by an omnidirectional camera. Each user can be represented as a computer-generated avatar [39, 44, 56] or, in recently proposed systems, a user's virtual representation was live captured by depth cameras [14, 18].

Recently, VR films are becoming popular thanks to the market available and affordable head-mounted displays (HMDs). The *Oculus* platform offers a vast variety of immersive content, from 360° videos, immersive 3D wonderlands to interactive replicas of historical monuments. The 360° video documentary *Rebuilding Notre Dame*² immerses VR viewers with the footage from before and after the April 2019 blaze at the same locations, capturing the majestic architecture from the angles where visitors usually cannot see. *The Under Presents*³ takes users to a new virtual world, where they enjoy live immersive theatres and explore novel interface-free interactions such as the "scrunch" technique. Users move forward by reaching out their virtual arms and pulling the destination towards them. The *Anne Frank House VR*⁴ reconstructed the "Secret Annex" where Anne spent two years of her life hiding in. The experience invites users to wander through the rooms, immerse themselves in Anne's thoughts and interact with Anne's belongings.

VR has increasingly become a sophisticated tool for storytelling, which guides viewers through the narrative in a novel way and invites viewers to participate [4]. Imagine a near future scenario, where you and your friends or family who live apart can walk into the same virtual movie together and see each other as holograms. You co-present with the movie characters, interact with them, and influence

e-mail: jie.li@cwi.nl

e-mail: s.subramanyam@cwi.nl

e-mail: jack.jansen@cwi.nl

e-mail: yanni.mei@cwi.nl

e-mail: nacho.reimat@cwi.nl

e-mail: kinga.j.lawicka@gmail.com

e-mail: p.s.cesar@cwi.nl

¹Teleparty is an application that synchronizes video playback and adds group chat to multiple over-the-top movie/TV content platforms such as Netflix, Disney, Hulu: <https://www.netflixparty.com>

²Rebuild Notre Dame is available at <https://www.oculus.com/experiences/media/1353452644677196/210792686621494>

³The Uder Presents is available at https://www.oculus.com/deeplink/?action=view&path=app/1917371471713228&ref=oculus_desktop

⁴The Anne Frank House VR is available at https://www.oculus.com/deeplink/?action=view&path=app/1596151970428159&ref=oculus_desktop

the movie storylines without interrupting the watching experiences. This new type of interactive movie that supports immersive social interaction brings the co-watching experience to the next level. The conversations between us would no longer be “The detective found three finger prints”, but be “My mother and I saw the forensic report held by the detective, saying ‘three finger prints’ ”.

In this exploitative work, we propose such an immersive and interactive experience, where four users join a 10-minute virtual movie together either using an HMD or a 2D screen with a game controller. Each user is captured in real time by three *Microsoft Azure Kinect DK* depth cameras (hereinafter referred to as *Kinect* depth cameras)⁵. The volumetric videos of users were transmitted into the virtual movie scene so that they felt like walking into the movie and being together with each other. We conducted a 48-user lab experiment and semi-structured interviews to compare users’ experiences between different user groups (i.e., devices, gender and age groups). We also invited 14 VR experts from 9 companies/institutes to evaluate the social VR movie experiences. We aimed at answering the following two research questions:

- **RQ1:** How do users experience the social VR movie in terms of quality of interaction, (social) presence, workload, and visual quality of the photorealistic representations?
- **RQ2:** What are users’ and VR experts’ attitudes toward this novel immersive social VR movie co-watching experiences?

The results show that both users and VR experts found photorealistic volumetric representations enhanced the co-presence. While HMD users reported a higher sense of presence and immersion, screen users reported lower workload and could more easily explore the virtual environment. This paper made two main contributions to the ISMAR community. First, it created an innovative, immersive and interactive social VR movie experience that brings distant users to co-present and interact in the movie. Second, the user and VR expert evaluations provided valuable insights for the further development of photorealistic social VR systems.

The paper is organized as follows: Sec. 2 discusses related work about media co-watching, volumetric video conferencing, and measuring social VR experiences. Sec. 3 illustrates the virtual movie and the study protocol. Sec. 4 presents the results of evaluations. Sec. 5 discusses the limitations, opportunities for new production workflows, and commercial impact. Finally, Sec. 6 concludes with a summary of the findings and future work.

2 RELATED WORK

This section starts with an overview of the media co-watching use cases, followed by a discussion of existing volumetric video conferencing systems that are the core technology for immersive and interactive co-watching experiences and relevant studies and questionnaires for measuring social VR experiences.

2.1 Watching Media Together

Watching movies together is an important social gathering that enables people to share emotions and enhance their social bonds. However, it is not always possible for people to co-watch at the same place. Many technologies have been developed to enable people to remotely share media consumption, such as text/audio chat [11, 17], video calls [50] and media synchronization for television (TV) services [52]. In the last decade, research on social TV has received a lot of attention. Cesar and Geerts [5] analyzed the advances in social TV in terms of content selection and sharing, direct communication, community building, and status updates. Huang et al. [17] studied the preference of social communication modalities that were added

⁵Microsoft Azure Kinect DK developer kit:<https://azure.microsoft.com/en-us/services/kinect-dk/#industries>

on TV programs and found that text chat was largely preferred and could foster the sense of connectedness among users. Geerts et al. [11] studied the synchronization requirements and their influence on users’ feeling of togetherness for watching online videos together. Similarly, more studies have shown the benefits of social TV in terms of togetherness, intimacy and improved relationships (e.g., [28, 30]).

Given the benefits of social TV, the research community recently shifted the focus to explore how to support shared media consumption through social VR platforms. *Mozilla Hubs*⁶ is an open source social VR platform that enables people to be represented as customized avatars, hang out in a shared virtual space, watch videos, and play with virtual objects. *Mozilla Hubs* has been used for many academic events, including the virtual IEEE VR 2020 conference⁷, the virtual ACM IMX 2020 conference⁸, and the ACM CHI 2020 Social VR Workshop [27], where scholars can co-watch the keynote speeches, paper/poster presentations, and demos in a virtual conference hall and network with each other. There are many other commercial social VR platforms supporting shared media consumption, such as Facebook Horizon⁹ and AltSpaceVR¹⁰. In addition, McGill et al. [30] showed that the adoption of HMDs with RGB-D cameras for users’ capturing can lead to an increased engagement, feeling of immersion and enjoyable embodied telepresence compared to video conferencing tools. De Simone et al. [8] compared shared video watching experiences in one avatar-based social VR system with a photorealistic user social VR systems. The results showed that the perceived quality of interaction in the photorealistic system is better than the avatar system, and is comparable with the face-to-face experiences. Wang et al. [54] proposed *ReliveInVR*, a 3D social VR experience that allows users to replay their previous VR experience, and immersively share it with other remote VR users. They compared *ReliveInVR* experience with desktop video sharing and VR 360° video sharing, and found that *ReliveInVR* enabled users to have a higher level of social presence, immersion, and feel more fulfilling in sharing experiences.

2.2 Volumetric Video Conferencing

Video technologies are constantly evolving towards higher quality, more efficient and immersive formats. For instance, 3D videos provide a higher sense of immersion and interactivity compared to 2D videos [13, 45]. Users cohabiting in a shared virtual space can be represented using volumetric (3D) representations, which are either computer generated (CG) or real-time reconstructed. CG avatars offer low rendering and transmission overhead as only joint poses need to be transmitted to remote users. CG avatars can also be animated by tracking user movements [2]. Wei et al [55] proposed a bidirectional system that can animate photorealistic avatar heads based on both users’ full likenesses using consumer-friendly HMD mounted cameras. However, subjective assessments of user experience have shown that CG avatars offer lower presence and state recognition as compared to 3D user reconstructions [32, 33].

Reconstructing users with photorealistic representations in real time requires additional computational and bandwidth for transmission and rendering. Recently, point clouds have emerged as a popular format for real-time reconstructions owing to their simplicity and versatility [40, 41]. On-demand streaming of pre-recorded point cloud sequences has been explored in previous research [16, 34, 51]. Jansen et al. [18] built and tested a pipeline that allows for low-latency capture and reconstruction of 3D volumes as point clouds, based on a setup of several RGB-D cameras. It also supports the compression and delivery of the point clouds, and the rendering at the receiving side. This pipeline enables the 3D representations of distant users to

⁶Mozilla Hubs: <https://hubs.mozilla.com>.

⁷The IEEE VR 2020 virtual conference: <https://ieeevr.org/2020/>.

⁸The ACM IMX 2020 conference: <https://imx.acm.org/2020/>.

⁹Facebook Horizon: <https://www.oculus.com/facebookhorizon/>.

¹⁰AltSpaceVR: <https://altvr.com>.

co-present and walk around in virtual spaces. *Microsoft Research* proposed an end-to-end system “*Holoportation*” [33]. It captures and transmits in real time a 3D projection of remotely captured 3D videos of users and their surroundings to another physical or virtual space. Likewise, a volumetric display system for holographic conferencing through VR/AR spaces has been developed by Mimesys¹¹. It relies on the use of *Intel RealSense* depth cameras. Pietroszek and Eckhardt [36] proposed a volumetric video capture system for narrative films, and claimed that the technology has transformed the passive movie audience to active participants, since they can choose the viewpoints and interact with the movie scenes.

2.3 Measuring Social VR Experiences

Social VR is promising in approximating face-to-face interactions, and has the potential to better support remote communication in terms of social presence, rich non-verbal communications, and immersive realistic interactions [7, 25, 44]. Instead of watching a film together on a screen, social VR can be experienced as if viewers are co-present in the same space. Although research interests in understanding social VR experiences are growing, there is no theoretical frameworks or experimental protocols to depict what factors influence social VR experiences and how to measure them. Many studies identified the importance of user representations for providing immersive experiences. Latoschik et al. [23] found that realistic avatars were rated significantly more human-like and evoked a stronger acceptance of the virtual body. Similarly, Waltemate et al. [53] concluded that personalized avatars significantly increase the sense of body ownership, presence, and dominance. Cho et al. [6] compared the actor captured by volumetric videos with the actor captured in 2D videos and another 3D avatar obtained by pre-scanning the actor. The results show that users have the highest sense of social presence with volumetric actor when performing dynamic tasks.

Apart from user representations, there are metrics (e.g., surveys, questionnaires) and experimental protocols that can be adapted to understand social VR experiences. Metrics for evaluating presence and immersion have been developed and widely validated, such as the presence questionnaire by Witmer and Singer [58] and the Slater-Usuh-Steed questionnaire [43]. Jennett et al. [19] suggested in their immersion questionnaire to include factors like lack of the awareness of time and involvement. Li et al. [25] proposed an experimental protocol and a social VR questionnaire for measuring quality of interaction, social meaning and presence/immersion.

Some other studies have explored user experiences in VR using different devices (e.g., 2D screens, HMDs). Subramanyam et al. [48] invited users to evaluate the visual quality of point clouds digital humans in two VR viewing conditions (e.g., 3- and 6 degrees of freedom) and on a 2D screen. The results suggest that the visual quality is affected by the view conditions. Srivastava et al. [46] examined how HMD and desktop would affect spatial learning when the ambulatory locomotion in HMD was restricted. They found that users spent more time and perceived less motion sickness and task effort using desktop than HMD. In their virtual earthquake training, Shu et al. [42] found that users reported a higher sense of spatial presence and immersion while using HMD than using desktop.

3 METHOD

This section presents the virtual movie production, the questionnaires and the intended measures, and the setups, participants, procedure of the user and the expert evaluations.

3.1 User Evaluation Setup

3.1.1 Virtual Movie Production

The 10-minute virtual movie is about the investigation of the murder of Ms. Armova, which was professionally produced by The Modern

Cultural Productions (Madrid, Spain) [37]. The virtual movie invites four users to join in simultaneously, who form the Civilian Oversight Committee as the witnesses of the crime solving process, and allows interaction with the movie characters to help with the process. There are six movie characters: Sarge Hoffsteler (detective), Elena Armova (victim), Rachel Tyrell (policewoman), Evans Young (forensic technician). Christine Gerard (Elena’s assistant) and Ryan Zeller (Elena’s ex-boyfriend) are the two suspects. The movie characters were generated in three steps (Fig. 2): (1) record the full-body acting of the real actors and actresses; (2) capture movements of their faces using an iPhone and Reallusion’s Live Face application and Character Creator¹²; (3) perform a body MoCap (full-body motion capture animation) during post production.



Figure 2: The generation of the virtual movie characters: (a&b) Capture the movements of the actor’s or actress’s face using the Reallusion’s Live Face application on an iPhone that is attached to a specially designed helmet; (c) Perform a full-body motion capture animation (MoCap) of the actor or actress; (d) The generated virtual movie character based on the face and full-body capture.

3.1.2 Locomotion and Interaction in the Movie

Users can access to the virtual movie either by using an HMD or using a screen with a game controller. Both HMD and screen users can use voice to answer the questions raised by the movie characters. However, the HMD users can only teleport between blue circles inside the virtual apartment, but can interact with the environment such as switching on the light and click on buttons (Fig. 3). The screen users can use the game controller to freely navigate inside the apartment, but cannot interact with the environment. These differences were pre-defined by the movie production company, aiming at (1) reducing the motion sickness of the HMD users by limiting their movement to teleportation (e.g., joystick walking in VR may largely increase motion sickness [22]), and (2) increasing the collaboration opportunities between HMD and screen users, because they have to find out who can interact with the environment or talk to the movie characters to move on in the story.

The virtual movie takes place in the luxury apartment of the victim Elena. The first part of the movie happens in the living room (Fig.1a), where four users are observing the crime solving and interacting with the movie characters (e.g., help switch on the light or click on the phone finder). In the second part, the four users are separated in to two groups. Two users follow Sarge to the kitchen where technician Evans is checking the evidence (Fig.1b). The other two users follow Rachel to the bedroom where the hologram of Elena confesses some secrets (Fig.1c). The users are represented as hologram-like point clouds (Fig.1a).

3.1.3 Four Connected Rooms

The user evaluations were conducted in four connected rooms, Room A, B, C and D, located on the 3rd floor of the CWI building (Science Park, Amsterdam). Room A and B had an Oculus Rifts HMD. Room C and D had a desktop computer, a 50-inch monitor and a game controller (Fig. 4). Each Room had three *Kinect* depth cameras to capture users’ volumetric representations, and deliver them to the virtual movie as point clouds. For each evaluation session, we

¹¹<https://www.mimesysvr.com>

¹²Reallusion’s Character Creator: <https://www.reallusion.com/character-creator>



Figure 3: The user-movie character interaction in the movie: (a) detective Sarge instructed one user to look for the phone finder and switch it off; (b) one HMD user found the phone finder on a lower table next to the window; (c) she switched it off.

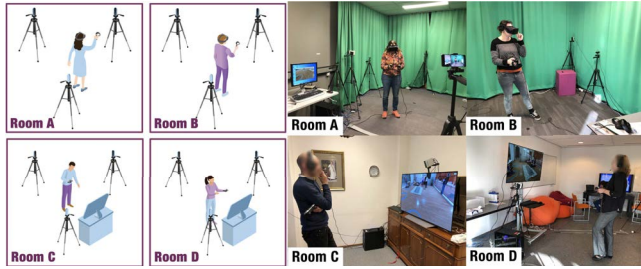


Figure 4: The (left) illustrations and (right) photos of the 4 rooms. Room A and B had an Oculus Rifts HMD. Room C and D had a screen computer, a 50-inch monitor, and a game controller. Each room had three *Kinect* depth cameras to capture users' volumetric representations.



Figure 5: The four starting positions for the users and the interactive objects (i.e., the light switch and the phone finder) in the virtual movie.

invited four users and assigned starting positions to them (Fig.5). The two HMD users (in Room A and B) were assigned to Position 1 and Position 2, respectively. The two screen users (in Room C and D) joined the virtual scene at Position 3 and 4. There are two interactive objects in the virtual movie: a light switch and a phone finder. HMD users must interact with them as instructed by the detective to move forward the story.

3.2 Participants

We recruited 48 participants (23 males, 25 females), who aged between 21-56 ($M = 34.9$, $SD = 10.3$). 12 females and 12 males were HMD users; 13 females and 11 males were screen users. All of them read, spoke, wrote fluently in English and had no visual/hearing/motor impairments. The long-/short-sighted participants were instructed to come to the experiment with their lenses or glasses. 13 of them had never used VR, 33 had used 1 to 3 times and two were experienced VR users. They came to the evaluation in groups of four persons of friends or acquaintances.

3.3 Questionnaires and Measures

We used the simulator sickness questionnaire (SSQ) [20], the social VR questionnaire [25], the presence questionnaire [58], the NASA Task Load Index (NASA TLX) [15], and the visual quality questionnaire. We intended to measure user experiences in terms of motion sickness, quality of interaction, social connectedness, presence, workload and visual quality of the volumetric representations when users were using different devices (i.e., HMD and 2D screen). In addition, inspired by previous works that have discussed the gender differences in communication [21], motion sickness [47] and presence experience in VR [9], we would like to explore the possible gender differences in the above-mentioned user experiences.

The **SSQ** was used to measure users' levels of cybersickness symptoms [20]. It was filled in by participants both before and after the virtual movie to measure their baseline states and reactions to the social VR experiences. The SSQ has three sub-scales, namely Nausea (N), Oculomotor (O), and Disorientation (D). Total Score (TS) is the score representing the overall severity of cybersickness experienced by the users of virtual reality systems [3].

The same **social VR questionnaire** was filled in by participants twice after the movie ended. The first time was to evaluate the experiences at the Part 1 of the movie and the second time was for Part 2. In this way, we explore the potential differences in the Social VR experience of Part 1 and Part 2 because Part 1 involved 4 users together in a spacious virtual living room while Part 2 had 2 users in a much smaller virtual space (kitchen or bedroom). The social VR questionnaire has 32 question items (Q1-Q32), where Q1-Q11 are intended to measure Quality of Interaction (QoI) in the social VR experience. Q12-Q22 are for Social Meaning (SM), which measures the social connectedness between users [25]. Q23-Q32 are for experiences of Presence/Immersion (PI) [25]. The questionnaire analysis is based on the sum of the scores given to the questionnaire items. The scores of items Q8, Q12-13, Q16, Q19-20, Q30 and Q32 are reversed, since they are all negative statements.

The Presence Questionnaire [58] was used to measure the participants' "sense of being in there" in the virtual movie. We removed Items 23 and 24 from the questionnaire, since the virtual movie did not include any haptic experiences. The 22 items are constructed by the following six factors [57]: (1) Realism: Items 3-7, 10 and 13; (2) Possibility to act (PtA): Items 1, 2, 8, and 9; (3) Quality of interface (QoI): Items (all reversed) 14, 17, and 18; (4) Possibility to examine (PtE): Items 11, 12, 19; (5) Self-evaluation of the performance (SEoP): Items 15 and 16; (6) Sounds: Items 20-22.

The NASA TLX assess the workload of the assigned tasks and has six items, measuring mental, physical and temporal demand, performance, effort and frustration on a 21-point Likert scale [15]. **The visual quality questionnaire** asked users to rate their own volumetric representations and those of others based on a 5-point Likert scale (1=bad, 2=poor, 3=fair, 4=good, 5=excellent).

3.4 Procedure

We have obtained the approval from the institute to run the study and strictly followed the national COVID-19 protocols. All researchers were wearing a face mask and hand gloves. All the rooms are large enough to ensure there was at least 2-meter distance between researchers and participants. All the handheld or wearable equipment were thoroughly disinfected by 70% alcohol after every usage.

There were 12 user evaluation sessions. Each session invited 4 users, took about 60 minutes, and consisted of the following steps: **Step 1 (10 minutes)**: The researcher explained the study. The participants signed an consent form and understood that they have the right to quit at any time. Then, they filled in two questionnaires: (1) a background questionnaire including demographics questions, a colour blindness screening and a vision acuity test; (2) a pre-study SSQ. **Step 2 (10 minutes)**: The researchers trained both the HMD and screen participants. **Step 3 (10 minutes)**: Each participant was taken to a separate room and the virtual movie was launched. **Step 4 (20 minutes)**: After the movie ended, the participants filled in the following questionnaires: (1) a post-study SSQ; (2) Social VR questionnaire for Part 1 (living room, 4 users); (3) Social VR questionnaire for Part 2 (kitchen/bedroom, 2 users); (4) Presence Questionnaire; (5) Visual Quality Questionnaire; (6) NASA TLX¹³. **Step 5 (10 minutes)**: The researcher conducted a semi-structured group interview with 4 users (audio recorded).

3.5 Evaluation and Focus Groups with Industry Experts

14 VR experts, from 9 companies/institutes¹⁴ were invited to evaluate the photorealistic social VR movie. We set up two rooms equipped with an HMD and one room with a screen. The 14 experts came in two groups: Group 1 had 8 experts, and Group 2 had 6 experts. After all the experts rotated and experienced both the HMD and the screen version of the virtual movie, they were gathered in a spacious meeting room (about 90 m²) for a 30-minute focus group discussion (audio recorded) about the potentials and challenges of the photorealistic social VR experiences. We followed the same COVID-19 protocol as the user evaluation.

4 ANALYSIS AND RESULTS

This section presents the statistics methods and the results of the user evaluation and the VR expert evaluation.

4.1 Simulator Sickness Questionnaire (SSQ)

For the SSQ, we checked the normality of the scores of Nausea (N), Oculomotor (O), Disorientation (D), and Total Score (TS) obtained before and after the virtual movie using Shapiro-Wilk normality test. None of them were normally distributed. We compared the scores of each category (i.e., N, O, D and TS) before and after the virtual movie experiences within each device group (i.e., HMD and screen), using Wilcoxon signed rank test with continuity correction. None of the categories show significant differences. We compared the N, O, D and TS scores given after the experiment between the HMD and the screen groups, all four categories show no significant difference.

4.2 Social VR Questionnaire

We first checked the reliability of the social VR questionnaire by calculating the Cronbach's alpha reliability coefficient. The Cronbach's alpha values for the three sets of questionnaire items (i.e., QoI, SM and PI) are 0.81, 0.74 and 0.79, respectively, indicating that the questionnaire has a good reliability [10]. A Levene's test

¹³The questionnaires used in the experiment are available at <http://bit.ly/2XBPJy1>

¹⁴The 9 companies/institutes are The Netherlands Sound and Vision Institute, MedicalVR, The Virtual Dutch Men, NEMO Science Museum, Erasmus University Medical Center, Sensiks, PostNL, BuitenboordMotor, and Interface

Part	Mean	Std.Error	95% CI
1	35.89	0.86	[34.17, 37.62]
2	37.20	0.83	[35.53, 38.88]

Table 1: PI at Part 1 and Part 2 of the movie.

Device	Mean	Std.Error	95% CI
HMD	39.50	1.15	[37.18, 41.82]
Screen	33.60	1.16	[31.27, 35.93]

Table 2: PI rated by HMD and screen users.

for equality of variances indicated that, for all the scores given on the QoI, SM and PI scales at the two parts of the experiment, the variances are equal. So, we applied the mixed-design ANOVA for the analyzing the social VR questionnaire.

4.2.1 Quality of Interaction (QoI) and Social Meaning (SM)

A mixed-design ANOVA with QoI as the dependent variable, the two movie parts as a within-subjects factor, and gender, devices as two between-subjects factors was conducted. Similarly, another mixed-design ANOVA with SM as the dependent variable was conducted as well. Both ANOVAs revealed no significant differences between HMD and screen users in terms of QoI and SM.

4.2.2 Presence/Immersion (PI)

A mixed-design ANOVA with PI as the dependent variable, the two experiment parts as a within-subjects factor, and gender, devices as two between-subjects factors was conducted. The results revealed a significant main effect of two parts of the movie on the PI ($F(1,44) = 9.32, p < .005, \eta_p^2 = .180$). Regardless which device they used, participants rated PI higher at Part 2 of the movie where two users were either in the virtual kitchen or in the virtual bedroom than Part 1 where four users were together in the virtual living room (Table 1). There is another significant main effect of the devices ($F(1,44) = 13.05, p < .001, \eta_p^2 = .229$). Participants who used an HMD rated PI higher than those who used the screen (Table 2).

There is a significant interaction effect between two movie parts and the devices (Fig. 6, $F(1,44) = 5.68, p < .05, \eta_p^2 = .114$). The HMD users increased the scores for PI when they moved from Part 1 to Part 2 (Part 1: $M = 38.33, SE = 1.21$, Part 2: $M = 40.67, SE = 1.18$). The screen users rated PI similarly at Part 1 and Part 2 (Part 1: $M = 33.45, SE = 1.21$; Part 2: $M = 33.74, SE = 1.18$).

4.3 Presence Questionnaire

A Levene's test for equality of variances indicated that, for all the scores given on the sub-scales (i.e., realism, PtA, QoI, PtE, SEoP,

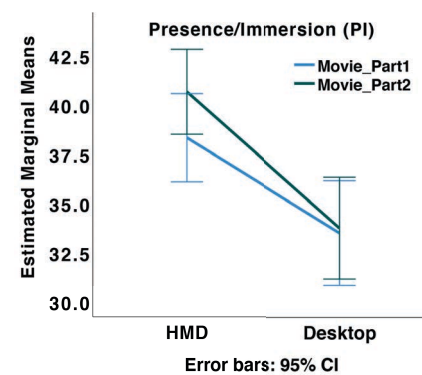


Figure 6: The significant interaction effect: two experiment parts × devices on the PI.

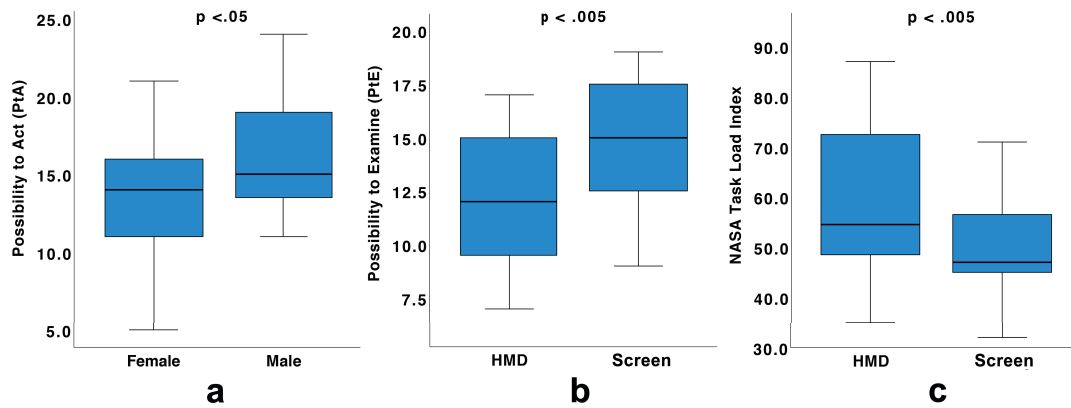


Figure 7: The two significant main effects of the Presence Questionnaire: (a) gender on PtA and (b) devices on PtE. (c) HMD users reported significantly heavier task load compared to the screen users.

and Sounds), the variances are equal. So, we conducted a MANOVA with the sum of the scores of six presence factors (i.e., realism, PtA, QoI, PtE, SEoP and Sounds) as dependent variables (presence variables), and gender (male, female), devices (HMD, screen) as two independent variables. Using Pillai’s trace, there was a significant effect of devices on the presence variables, $V = 0.47$, $F(6, 39) = 5.80$, $p < .001$, $\eta_p^2 = .472$.

Separate univariate ANOVAs on the six presence variables revealed two significant effects (Fig. 7). Male participants rated “possibility to act (PtA)” significantly higher than females (male: $M = 16.10$, $SE = .80$; female: $M = 13.55$, $SE = .76$; $F(1,44) = 5.34$, $p < .05$, $\eta_p^2 = .108$). Screen users rated “possibility to examine (PtE)” significantly higher than HMD users (screen: $M = 14.70$, $SE = .63$; HMD: $M = 11.92$, $SE = .63$; $F(1,44) = 9.76$, $p < .005$, $\eta_p^2 = .182$). No significant interaction effects are found.

Noticeably, one sub-scale of the social VR questionnaire (PI) is also meant to measure presence and immersion experiences. Therefore, we performed a Pearson correlation analysis to check the correlations between the sum of the scores of the PI sub-scale at Part 1 and Part 2 of the virtual movie, and the sum of the scores of the 22 items of the presence questionnaire. The results indicated a strong positive correlation between PI at Part 1 and PI at Part 2 of the movie, $r = .886$, $n = 48$, p (2-tailed) $< .001$; a moderate positive correlation between PI at Part 2 of the movie and the presence questionnaire, $r = .408$, $n = 48$, p (2-tailed) $< .005$; and a non-significant but positive correlation between PI at Part 1 and the presence questionnaire, $r = .284$, $n = 48$, p (2-tailed) $= .050$.

4.4 NASA Task Load Index (TLX)

We first checked the normality of the data by using the Shapiro-Wilk normality test. The results show that the overall TLX scores are not normally distributed, but the scores given by two device groups are normally distributed. Those given by males are normally distributed, but by females are not.

We applied independent-sample t-test and identified a significant difference between the HMD users and the screen users in terms of task load (HMD: $M = 60.17$, $SD = 15.30$; screen: $M = 50.00$, $SD = 9.32$, $p < .005$, Cohen’s $d = .803$). HMD users reported heavier task load compared to the screen users. We used the independent-samples Mann-Whitney test to check the differences of TLX between the gender groups. No significant effects are found.

4.5 Visual Quality of User Representations and Virtual Characters

For the visual quality questionnaire, we first examined the normality of the self-visual quality ratings and the ratings of other users’ repre-

sentations, by using the Shapiro-Wilk test. Most of the data are not normally distributed except “the ratings of the others’ representations” given by the screen users. Therefore, we used non-parametric tests for analyzing the visual quality data.

We performed an independent-samples Mann-Whitney test to check the differences of self ratings between two devices groups, and the same test to check the differences of ratings of others between two device groups. In both case, no significant differences were found. However, a related-samples Wilcoxon Signed Rank test showed that, regardless of gender or devices, the self ratings ($Mdn = 3.00$) were significantly lower than the rating of others ($Mdn = 3.33$), $Z = -3.14$, $p < .005$, $r = -.45$. However, within the HMD user group, a related-samples Wilcoxon Signed Rank test showed that, the self ratings ($Mdn = 3.00$) were also significantly lower rated than the ratings of others ($Mdn = 3.33$), $Z = -2.39$, $p < .05$, $r = -.49$. Within the screen user group, no significant difference was found between the ratings of self and others (Fig. 8).

Using the Shapiro-Wilk test, we found the scores given to visual quality of virtual movie characters by two device groups are not normally distributed. So, we performed Mann-Whitney test and found no difference between the HMD and screen user groups in terms of the quality of the virtual movie characters. Noticeably, both device groups gave very high ratings towards the quality of virtual characters (HMD: $Mdn = 5.00$, $M = 4.50$, $95\%CI = [4.20, 4.81]$; screen: $Mdn = 5.00$, $M = 4.33$, $95\%CI = [3.93, 4.74]$).

4.6 Summary of the Questionnaires Results

In summary, both HMD and screen users did not report any difference in cybersickness. We found differences in user experiences between the device and gender groups. Regardless which device the users used, they reported a higher sense of presence and immersion (PI) at Part 2 than Part 1 of the movie. We also found that while HMD users had higher sense of PI, screen users experienced lower workload and more freedom to explore the virtual movie scene (“possibility to examine (PtE)”, i.e., examining objects from multiple viewports, easy to control). In addition, we found that males rated the “possibility to act (PtA)” higher than females, indicating that they could more easily and more actively control and interact with the virtual environment. For the evaluation of the volumetric representations, we found that, within the HMD group, the ratings for self representations were worse than the ratings for others’. However, no differences were found within the screen users between the ratings for self and others’ representations. We noticed that there was a moderate positive correlation between the PI sub-scale of the social VR questionnaire and the presence questionnaire.

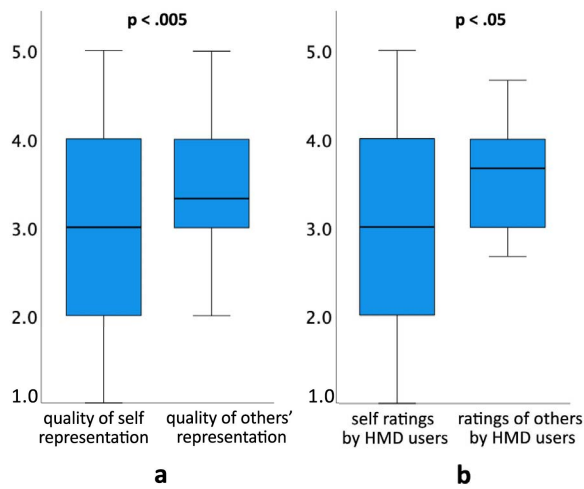


Figure 8: The significant differences of visual quality of self and others' representations: (a) Regardless of gender and devices, the self ratings were significantly lower than the rating of others; (b) within the HMD user group, the self ratings were also significantly lower rated than the ratings of others.

4.7 Interview Results

The audio recordings of the group interviews (4 users together) were transcribed. Then, the transcripts were open coded and categorized by two researchers [49]. We label the 12 groups of 48 participants as G1 to G12. A square bracket with the device name is added after the group label to indicate whether the quotes were from the HMD or screen users in that group (e.g., G1[HMD] indicating one HMD user from G1). The resulted categories are summarized as follows:

4.7.1 New Narrative Experiences

Among 12 user groups, 10 groups were very positive towards the novel social VR movie and the photorealistic representations (G4[HMD] said, “[With photorealistic representations], I felt more like I was talking with real people instead of virtual avatars.”). All the users were excited about the details of movie scenes and characters (G2[Screen] said, “The details of movie characters were pretty good especially when they were moving their fingers.”). However, all the user groups mentioned that there was not much communication and interaction between them (G7[screen] mentioned that, “I was mainly listening to the characters. There were no collaborative tasks for us to do together.”). This aspect is reflected in the questionnaire as well. The social VR questionnaire results showed that there was no difference in “Quality of Interaction (QoI)” between screen and HMD users. The mean scores that users gave to the 11 items of the QoI sub-scale were mostly below 3 (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree), indicating users’ negative feedback towards the QoI (Fig. 9). The users also commented that this new type of movies require new narrative production that supports immersive interactions, but should not interrupt the narratives (G7[HMD] said that, “It would be cool to have kind of virtual escape room experiences, but meanwhile I wanted to concentrate on the conversations of the characters.”).

In terms of locomotion, users pointed out that screen users had better mobility than HMD users, since they can move freely in the virtual scene and examine all the details in the virtual apartment (G5[screen] said, “It was nice to move everywhere and even read the book covers on the shelf.”). Although the HMD users felt immersed in the movie scene, they felt limited in movement, because they could only walk within the limited room space and teleport among pre-defined positions. (G2[HMD] said, “I was too stuck at my

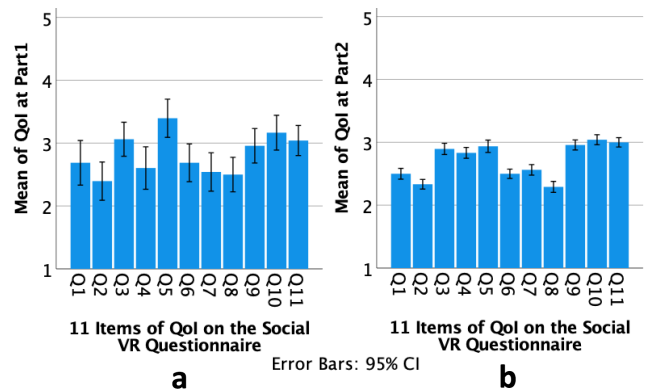


Figure 9: The mean scores given to the 11 items of the QoI sub-scale at (a) Part 1 and (b) Part 2 of the movie. Most of them are below 3, indicating users’ negative feedback towards QoI (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree).

position, could not move closer to the detective to see what was in his hands.”). However, screen users felt less immersed and were very curious to experience the HMD version of the movie (G5[screen] said, “I was jealous of the HMD users, because they were actually there [in the virtual scene], and they could switch on the light.”). On the questionnaires, we also discovered that screen users reported that they had more “possibility to examine” in the virtual movie than the HMD users, but lower sense of presence and immersion.

4.7.2 Visual Quality of the Representations

All the HMD users agreed that the photorealistic representations enhanced their co-presence feeling (G11[HMD] said: “I can see my friend’s [another HMD user’s] clothes and the way she moved. I could tell it was definitely her.”). Another interesting finding was that the HMD users were not satisfied with their self representations. They found it blurry and could barely see their own fingers. However, they thought the representations of other users were better since they were at distance, and the overall quality of their clothes and gestures was good. (G2[HMD] said, “I saw myself as a crowd of pixels.”; G8[HMD] said, “The colors of his clothes were realistic. The quality of his whole outfit was satisfactory.”). The visual quality difference was also found in the questionnaire results, where HMD users rated their own representation significantly worse than those of others.

4.7.3 Recommendation for Improvement

The participants provided insightful recommendations for improving the social VR movie experiences. First, they suggest including multi-sensory interactions (e.g., haptics). Second, they would like to have game-like experiences that enable them to collaboratively solve a problem and actively influence the movie storyline. Third, they want to have more interactions between users and the virtual characters (G12[HMD] said, “I want to have more natural conversations with the detective, like I am helping him find the clues. Now, I am only answering ‘yes’ or ‘no’ questions asked by him.”). Last but not least, they would like to explore in the virtual environment (G3[screen] mentioned, “I would like to grab the things in the living room, and participated in the crime solving like a virtual escape room.”).

4.8 Evaluation Results from VR Experts

All 14 experts (E1-E14) were impressed by the simple setup of the “hologram” capturing system and were excited to see the photorealistic representations, which enabled them to feel co-present in the same space. As we observed, they were waving at each other in the virtual space, and talking about the texture of their clothes, and the possible scenarios for applying the social VR system. The virtual

movie was short (about 10 minutes), but all the experts spent much longer time in exploring all the details. They saw the full potential of our social VR system in many market sectors, such as medical care, education, immersive meetings, family reunion, virtual dating. As E8 commented, “It doesn’t feel like an avatar, but the real person. Despite the quality, obviously, still needs a lot of improvement, but you go beyond uncanny valley, and it’s really the person there. That’s amazing. That’s more than I expected. I expected it to be a nice virtual environment with a Skype-like interface, but you go way beyond that.” The experts also pointed out challenges that we need to consider as our future work. E6 is a physician, she mentioned that, “The realism level of the virtual objects needs to be much higher in clinical context. Suppose we are going to reconstruct the breast of a cancer patient, not only the visual quality needs to be fully realistic, the haptic feelings of the 3D reconstruction also need to be realistic. In this way, patients can have a correct expectation towards the surgery.” E11 suggested based on the comparison of the two devices, “The HMD was very immersive, but on the other hand, the screen with a game controller was more practical, perhaps also more addictive. I am curious to see the effects on 3D screens.”

In summary, the experts are very positive towards the social VR platform. They see its potential not only in supporting remote experiences, but also augmenting the co-present experiences (e.g., hyper-realistic medical objects). The improvement suggestions mainly cover three aspects: (1) the visual quality of the volumetric representations (e.g., less noisy, higher resolution), (2) the possibility to interact (e.g., develop new interaction techniques dedicated to virtual worlds), and (3) the accessibility of the social VR platform (e.g., decentralized at home or centralized at public hubs).

5 DISCUSSION

We aimed to answer two research questions (RQ1 and RQ2, see Sec. 1). For RQ1, we found that HMD users reported a higher sense of presence and immersion than screen users, but screen users reported lower workload and could more easily explore the virtual environment than HMD users. We also found that gender influenced interactions with the movie. For RQ2, both end-users and VR experts found that the photorealistic volumetric representations enhanced feelings of co-presence, and they foresaw that this novel interactive and immersive VR movie is the future of movie co-watching.

5.1 Limitations

We are aware of the study limitations. The novel virtual movie was professionally produced by cinematography experts, which was not intended for laboratory experiments. Therefore, not all aspects were fully controlled. The locomotion and interaction techniques used by HMD and screen users are not all the same. However, the goal of our study is to exploratively evaluate a novel immersive movie with photorealistic user representations. We had to find the right balance between the “controlled experiment” and the “a real movie produced by professionals”. We primed the aesthetics and narrativity aspects over the perfectly controlled aspects. We find this really important, since we are able to explore a first-class production and derive important insights about user experiences.

5.2 Opportunities for Controlled Experiments

We identified many potential factors that require further investigation in a controlled manner, such as the influence of locomotion methods on users’ cybersickness, (social) presence and quality of interaction. Many users mentioned that they would have more interaction possibilities with the movie characters, the other users and the virtual environment, indicating another research opportunity to investigate the influence of interactions on users’ experiences. In addition, we also observed that there were differences in users’ movement trajectories and differences in users’ experiences using the devices. For instance, we noticed that male screen users moved more quickly

and more frequently than female screen users, and their trajectories covered the whole virtual apartment. It also seemed that experienced VR/game users finished the tasks faster than inexperienced users (e.g., find and switch on the light). We would like to further collect and analyze objective data from the users, such as movements, trajectories, gaze directions, operation errors, completion time and audio sentiments to understand users’ proxemics, social interactions and emotions in virtual spaces. To do so, we need new production workflows that can create virtual movies that allows more interactions and can be instrumented to run controlled experiments.

5.3 Production Opportunities for Immersive Movies

Recently, many novel media watching experiences have been developed, such as interactive narratives that invite viewers to choose paths for lead characters (e.g., Black Mirror: Bandersnatch [38]); social TV or multiscreen TV that enable viewers to customize their viewing content and to comment during a show [26]; and cinematic VR with 360 ° videos that allows users to choose viewpoints [29]. However, these experiences were either limited at the interaction level or lacking narrativity. We expect that immersive and interactive movies with volumetric user representations will be the next innovation of media watching, where viewers are represented realistically and have the opportunity to sing along with the artists, re-watching a movie “inside” it, or even become a character in it. Crafting such new experiences requires to incorporate interactive narratives, viewers’ co-presence, interactive virtual environments, and social communications into production workflows.

5.4 Commercial Impact

Despite of the limitations, the social VR movie experiences received very positive feedback from both users and industry experts. The users thought the walk-into-movie experience is novel, especially with the photorealistic representations. They would imagine to install such a light weight volumetric social VR system at home in the near future when it is market available, so they can immerse themselves into a movie and interact with their favourite characters. Industry experts foresee that our social VR system has the potential to open a new era of media consumption. With the arrival of 5G mobile networks [1], the experts expect to see volumetric representations with significantly improved visual quality and invisible latency, which would lead to a seamless social VR experience. They also predict that the volumetric social VR system can be adopted in many market sectors such as education, museums, medical care, remote collaboration, gaming, online dating, and virtual tourism.

6 CONCLUSION

We proposed a novel walk-into-movie experience supported by a volumetric social VR system. We evaluated the novel movie experience and its commercial impact through a 48-user evaluation study and two focus groups with VR experts. The results showed that while HMD users reported a higher sense of presence and immersion, screen users reported lower workload and could more easily explore the virtual environment. Overall, both users and VR experts found that photorealistic volumetric representations enhanced co-presence. We additionally found that gender influenced interactions with the movie. In the future, we would like to analyze objective data such as gaze directions, trajectories, audio sentiments to have an in-depth understanding of users’ behavior in virtual spaces.

ACKNOWLEDGMENTS

This research work is funded by the EU Horizon 2020 program (grant agreement 762111), as part of the project “VRtogether” (<http://vrtogether.eu/>). We would specially thank theMo and Artanim Foundation for creating the virtual movie.

REFERENCES

- [1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang. What will 5g be? *IEEE Journal on selected areas in communications*, 32(6):1065–1082, 2014.
- [2] P. Bourdin, J. M. T. Sanahuja, C. C. Moya, P. Haggard, and M. Slater. Persuading people in a remote destination to sing by beaming there. In *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology*, VRST '13, p. 123–132. Association for Computing Machinery, New York, NY, USA, 2013. doi: 10.1145/2503713.2503724
- [3] K. Brunnström, K. Wang, S. Tavakoli, and B. André. Symptoms analysis of 3d tv viewing based on simulator sickness questionnaires. *Quality and User Experience*, 2(1):1, 2017.
- [4] J. Bucher. *Storytelling for virtual reality: Methods and principles for crafting immersive narratives*. Taylor & Francis, Milton Park, Abingdon, UK, 2017.
- [5] P. Cesar and D. Geerts. Past, present, and future of social tv: A categorization. In *2011 IEEE consumer communications and networking conference (CCNC)*, pp. 347–351. IEEE, New York, NY, USA, 2011.
- [6] S. Cho, S.-w. Kim, J. Lee, J. Ahn, and J. Han. Effects of volumetric capture avatars on social presence in immersive virtual environments. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 26–34. IEEE, New York, NY, USA, 2020.
- [7] E. F. Churchill, D. N. Snowdon, and A. J. Munro. *Collaborative virtual environments: digital places and spaces for interaction*. Springer Science & Business Media, Berlin/Heidelberg, Germany, 2012.
- [8] F. De Simone, J. Li, H. G. Debarba, A. El Ali, S. N. Gunkel, and P. 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)*, pp. 890–891. IEEE, New York, NY, USA, 2019.
- [9] A. Felnhofner, O. D. Kothgassner, L. Beutl, H. Hlavacs, and I. Kryspin-Exner. Is virtual reality made for men only? exploring gender differences in the sense of presence. *Proceedings of the International Society on presence research*, pp. 103–112, 2012.
- [10] A. Field. *Discovering statistics using SPSS 3rd edition*. Sage, London, UK, 2009.
- [11] D. Geerts, I. Vaishnavi, R. Mekuria, O. Van Deventer, and P. Cesar. Are we in sync?: synchronization requirements for watching online video together. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 311–314. ACM, New York, NY, USA, 2011.
- [12] S. Gomillion, S. Gabriel, K. Kawakami, and A. F. Young. Let's stay home and watch tv: The benefits of shared media use for close relationships. *Journal of Social and Personal Relationships*, 34(6):855–874, 2017.
- [13] M. Gross, S. Würmlin, M. Naef, E. Lamboray, C. Spagno, A. Kunz, E. Koller-Meier, T. Svoboda, L. Van Gool, S. Lang, K. Strehlke, A. V. Moere, and O. Staadt. Blue-c: A spatially immersive display and 3d video portal for telepresence. *ACM Trans. Graph.*, 22(3):819–827, July 2003. doi: 10.1145/882262.882350
- [14] S. N. Gunkel, M. Prins, H. Stokking, and O. Niamut. Social vr platform: Building 360-degree shared vr spaces. In *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*, pp. 83–84. ACM, New York, NY, USA, 2017.
- [15] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, vol. 52, pp. 139–183. Elsevier, Amsterdam, NL, 1988.
- [16] M. Hosseini and C. Timmerer. Dynamic adaptive point cloud streaming. In *Proceedings of the 23rd Packet Video Workshop*, pp. 25–30. ACM, New York, NY, USA, Jun 2018. doi: 10.1145/3210424.3210429
- [17] E. M. Huang, G. Harboe, J. Tullio, A. Novak, N. Massey, C. J. Metcalf, and G. Romano. Of social television comes home: a field study of communication choices and practices in tv-based text and voice chat. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 585–594. ACM, New York, NY, USA, 2009.
- [18] J. Jansen, S. Subramanyam, R. Bouqueau, G. Cernigliaro, M. Martos Cabré, F. Pérez, and P. Cesar. A Pipeline for Multiparty Volumetric Video Conferencing: Transmission of Point Clouds over Low Latency DASH. In *Proceedings of the 11th ACM Multimedia Systems Conference, MMSys '20*, pp. 341–344. ACM, New York, NY, USA, June 2020.
- [19] C. Jennett, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, and A. Walton. Measuring and defining the experience of immersion in games. *International journal of human-computer studies*, 66(9):641–661, 2008.
- [20] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.
- [21] A. M. Kimbrough, R. E. Guadagno, N. L. Muscanell, and J. Dill. Gender differences in mediated communication: Women connect more than do men. *Computers in Human Behavior*, 29(3):896–900, 2013.
- [22] E. Langbehn, P. Lubos, and F. Steinicke. Evaluation of locomotion techniques for room-scale vr: Joystick, teleportation, and redirected walking. In *Proceedings of the Virtual Reality International Conference-Laval Virtual*, pp. 1–9, 2018.
- [23] M. E. Latoschik, D. Roth, D. Gall, J. Achenbach, T. Waltemate, and M. Botsch. The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, p. 39. ACM, New York, NY, USA, 2017.
- [24] J. Li, G. Chen, H. De Ridder, and P. Cesar. Designing a social vr clinic for medical consultations. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–9, 2020.
- [25] J. Li, Y. Kong, T. Röggl, F. De Simone, S. Ananthanarayan, H. de Ridder, A. El Ali, and P. Cesar. Measuring and understanding photo sharing experiences in social virtual reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19*, p. 1–14. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300897
- [26] J. Li, T. Röggl, M. Glancy, J. Jansen, and P. Cesar. A new production platform for authoring object-based multiscreen tv viewing experiences. In *Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video, TVX '18*, p. 115–126. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3210825.3210834
- [27] J. Li, V. Vinayamoorthy, R. Schwartz, W. IJsselsteijn, D. A. Shamma, and P. Cesar. Social vr: A new medium for remote communication and collaboration. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–8. ACM, New York, NY, USA, 2020.
- [28] D. Marfil, F. Boronat, M. Montagud, and A. Sapena. Idms solution for hybrid broadcast broadband delivery within the context of hbbtv standard. *IEEE Transactions on Broadcasting*, 65(4):645–663, 2018.
- [29] J. Mateer. Directing for cinematic virtual reality: how the traditional film director's craft applies to immersive environments and notions of presence. *Journal of Media Practice*, 18(1):14–25, 2017.
- [30] M. McGill, J. H. Williamson, and S. Brewster. Examining the role of smart tvs and vr hmds in synchronous at-a-distance media consumption. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 23(5):33, 2016.
- [31] Y. Mei, J. Li, H. de Ridder, and P. Cesar. Cakevr: A social virtual reality (vr) tool for co-designing cakes. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2021.
- [32] R. Mekuria, P. Cesar, I. Doumanis, and A. Frisiello. Objective and subjective quality assessment of geometry compression of reconstructed 3D humans in a 3D virtual room. In A. G. Tescher, ed., *Applications of Digital Image Processing XXXVIII*, vol. 9599, pp. 537–549. International Society for Optics and Photonics, SPIE, Bellingham, Washington, USA, 2015. doi: 10.1117/12.2203312
- [33] S. Orts-Escolano, C. Rhemann, S. Fanello, W. Chang, A. Kowdle, Y. Degtyarev, D. Kim, P. L. Davidson, S. Khamis, M. Dou, et al. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pp. 741–754. ACM, New York, NY, USA, 2016.
- [34] J. Park, P. Chou, and J.-N. Hwang. Rate-utility optimized streaming of volumetric media for augmented reality. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, PP:1–1, 02 2019. doi: 10.

- 1109/JETCAS.2019.2898622
- [35] C. Pidel and P. Ackermann. Collaboration in virtual and augmented reality: a systematic overview. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pp. 141–156. Springer, Cham, Switzerland, 2020.
- [36] K. Pietroszek and C. Eckhardt. Volumetric capture for narrative films. In *26th ACM Symposium on Virtual Reality Software and Technology, VRST '20*. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3385956.3422116
- [37] A. Revilla, S. Zamarvide, I. Lacosta, F. Perez, J. Lajara, B. Kevelham, V. Juillard, B. Rochat, M. Drocco, N. Devaud, et al. A collaborative vr murder mystery using photorealistic user representations. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 766–766. IEEE, 2021.
- [38] C. Roth and H. Koenitz. Bandersnatch, yea or nay? reception and user experience of an interactive digital narrative video. In *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*, pp. 247–254. ACM, New York, NY, USA, 2019.
- [39] D. Roth, J.-L. Lugrin, D. Galakhov, A. Hofmann, G. Bente, M. E. Latoschik, and A. Fuhrmann. Avatar realism and social interaction quality in virtual reality. In *2016 IEEE Virtual Reality (VR)*, pp. 277–278. IEEE, New York, NY, USA, 2016.
- [40] R. B. Rusu and S. Cousins. 3d is here: Point cloud library (pcl). In *2011 IEEE International Conference on Robotics and Automation*, pp. 1–4. IEEE, New York, NY, USA, 2011. doi: 10.1109/ICRA.2011.5980567
- [41] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokuća, S. Lasserre, Z. Li, J. Llach, K. Mamou, R. Mekuria, O. Nakagami, E. Siahaan, A. Tabatabai, A. M. Tourapis, and V. Zakharchenko. Emerging mpeg standards for point cloud compression. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 9(1):133–148, 2019. doi: 10.1109/JETCAS.2018.2885981
- [42] Y. Shu, Y.-Z. Huang, S.-H. Chang, and M.-Y. Chen. Do virtual reality head-mounted displays make a difference? a comparison of presence and self-efficacy between head-mounted displays and desktop computer-facilitated virtual environments. *Virtual Reality*, 23(4):437–446, 2019.
- [43] M. Slater, M. Usoh, and A. Steed. Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 3(2):130–144, 1994.
- [44] H. J. Smith and M. Neff. Communication behavior in embodied virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 289. ACM, New York, NY, USA, 2018.
- [45] A. Smolic, K. Mueller, P. Merkle, P. Kauff, and T. Wiegand. An overview of available and emerging 3d video formats and depth enhanced stereo as efficient generic solution. In *2009 Picture Coding Symposium*, pp. 1–4. IEEE, New York, NY, USA, 2009.
- [46] P. Srivastava, A. Rimzhim, P. Vijay, S. Singh, and S. Chandra. Desktop vr is better than non-ambulatory hmd vr for spatial learning. *Frontiers in Robotics and AI*, 6:50, 2019.
- [47] K. Stanney, C. Fidopiastis, and L. Foster. Virtual reality is sexist: but it does not have to be. *Frontiers in Robotics and AI*, 7:4, 2020.
- [48] S. Subramanyam, J. Li, I. Viola, and P. Cesar. Comparing the quality of highly realistic digital humans in 3dof and 6dof: A volumetric video case study. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 127–136. IEEE, New York, NY, USA, 2020.
- [49] D. R. Thomas. A general inductive approach for analyzing qualitative evaluation data. *American journal of evaluation*, 27(2):237–246, 2006.
- [50] M. F. Ursu, M. Groen, M. Falelakis, M. Frantzis, V. Zsombori, and R. Kaiser. Orchestration: tv-like mixing grammars applied to video-communication for social groups. In *Proceedings of the 21st ACM international conference on Multimedia*, pp. 333–342. ACM, New York, NY, USA, 2013.
- [51] J. van der Hooft, T. Wauters, F. De Turck, C. Timmerer, and H. Hellwagner. Towards 6dof http adaptive streaming through point cloud compression. In *Proceedings of the 27th ACM International Conference on Multimedia, MM '19*, p. 2405–2413. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3343031.3350917
- [52] M. O. van Deventer, M. Probst, and C. Ziegler. Media synchronisation for television services through hbbtv. In *MediaSync*, pp. 505–544. Springer, Cham, Switzerland, 2018.
- [53] T. Waltemate, D. Gall, D. Roth, M. Botsch, and M. E. Latoschik. The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE transactions on visualization and computer graphics*, 24(4):1643–1652, 2018.
- [54] C. Y. Wang, M. Sakashita, U. Ehsan, J. Li, and A. S. Won. Again, together: Socially reliving virtual reality experiences when separated. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–12. ACM, New York, NY, USA, 2020.
- [55] S.-E. Wei, J. Saragih, T. Simon, A. W. Harley, S. Lombardi, M. Perdoch, A. Hypes, D. Wang, H. Badino, and Y. Sheikh. Vr facial animation via multiview image translation. *ACM Trans. Graph.*, 38(4), July 2019. doi: 10.1145/3306346.3323030
- [56] J. Williamson, J. Li, V. Vinayagamoorthy, D. A. Shamma, and P. Cesar. Proxemics and social interactions in an instrumented virtual reality workshop. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2021.
- [57] B. G. Witmer, C. J. Jerome, and M. J. Singer. The factor structure of the presence questionnaire. *Presence: Teleoperators & Virtual Environments*, 14(3):298–312, 2005.
- [58] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3):225–240, 1998. doi: 10.1162/105474698565686