Understanding Online Education in Metaverse: Systems and User Experience Perspectives

Ruizhi Cheng* Erdem Murat† Lap-Fai Yu‡ Songqing Chen§ Bo Han¶

George Mason University

Abstract

Thanks to recent advances in immersive technologies, virtual reality (VR) is becoming increasingly popular in online education, particularly in light of the rise of the Metaverse. However, there is currently no in-depth investigation of the user experience of VR-based online education and the comparison of it with video-conferencing-based counterparts. To fill these critical gaps, we conduct multiple sessions of two courses in a university with 10 and 37 participants on Mozilla Hubs (Hubs for short), a social VR platform that is deemed as one of the early prototypes of the Metaverse, and let them compare the classroom experience on Hubs with Zoom, a popular video-conferencing application. In addition to employing traditional analytical methods to understand user experience, we benefit from an end-to-end measurement study of Hubs to corroborate our findings and systematically detect its performance bottlenecks. Our study leads to the following key observations. First, the scalability issue of Hubs makes it inadequate for accommodating large courses. Second, compared to Zoom, Hubs can offer a better sense of place presence and social presence to students, thanks to its avatar-based interactions and the hand and head tracking enabled by headsets. Third, even though VR headsets help students concentrate in class, effectively utilizing learning tools through them remains a challenge.

1 Introduction

In recent years, with the continuing development of virtual reality (VR) technologies and the emergence of the Metaverse [12], utilizing VR for online education has gained increasing popularity [72]. VR aims to provide users with a fully immersive experience that can boost their sense of presence and immersion and create an authentic learning environment, which in turn may potentially enhance their learning experience [36,48]. Although VR has been demonstrated to be a potentially viable option for online education in a few pilot studies [14, 20, 23, 47, 64], the majority of online courses remain to rely on video-conferencing applications, such as Zoom [71]. However, the prevalence of "Zoom fatigue" [81] and the potential benefits of VR in providing better immersion and learning experiences for students demand an in-depth understanding of VR in real-world online education settings in order for us to recognize its advantages and limitations, particularly when compared with existing video-conferencing-based platforms.

Current research on VR-based online education, however, is limited in the following aspects. First, existing studies rely mainly on qualitative comparison and/or quantitative analysis without providing any insights into the complex interplay between the performance of the underlying systems, the number of students, and their experience in VR classrooms. Due to the complexity of 6DoF (six degrees of freedom) motion enabled by immersive content and the proximity of VR display to the user's eyes, the frame rate, end-to-end throughput/latency, and on-device resource utilization of VR systems can have a profound impact on user experience. For example, rendering VR content on headsets demands at least 60 FPS (frames per second) [7]. A low frame rate will negatively affect user experience, even leading to motion sickness [85]. On the other hand, modern video-conferencing applications, such as Zoom and Google Meet, which display 2D content, set their target FPS to 30 [43]. Second, comparisons of the learning experience with VR and video-conferencing applications are still rudimentary, with only qualitative analysis [19,29] and small-scale tests (*e.g.,* 17 participants in Ripka *et al.* [69]). A comprehensive comparison in larger-scale courses would provide a deeper understanding of the strengths and limitations of these two types of online education platforms.

To fill these critical gaps, we conducted IRB-approved user studies on multiple sessions of two courses in a university using an open-source social VR platform, Mozilla Hubs (Hubs for short) [39]. One course had 37 participants, and the other had 10. We chose Hubs over other social VR platforms such as Rec Room [68] and Horizon Worlds [51] because: 1) Hubs has been demonstrated to be a suitable choice for hosting online courses [9]; 2) Given that students may use PCs with different operating systems to attend online courses, Hubs, as a Web-based application that supports access from any operating system, allows us to compare user experience on VR-based online education for both PCs and headsets; and 3) Hubs is open-source, which greatly facilitates our measurement studies via source code instrumentation (§3.3).

We provided students with the Oculus Quest 2 VR headset [52], which is a popular choice for conducting out-of-lab VR experiments [55] and arguably the most popular VR headset [78]. We did not choose tethered headsets such as HTC VIVE because they require high-end PCs with powerful CPUs and GPUs, which may not be available to most students. Moreover, due to its immersive nature, the VR-based classroom can be interactive. For instance, students may move around to observe 3D models from different angles. Cables attached to tethered headsets can be dangerous for them to trip over [42]. For Hubs-based classes, 17 students in the large course and 7 in the small one used Oculus Quest 2 while others used PCs. Moreover, we asked all students to compare their classroom experience on Hubs with Zoom, which is widely used in online education [71]. Through a questionnaire after the lectures, we analyzed and compared their ratings of the overall experience, visual and audio quality, and the sense of place presence, social presence, and co-presence (defined in §3.4) for both platforms. In addition, we interviewed participants to gain a deep understanding of their ratings and learning experience, and whether they have suffered motion sickness on Hubs. We summarize our key findings as follows.

- In the classroom with 37 participants, headset users on Hubs encounter serious performance issues with only ∼20 FPS, which significantly raises the risk of experiencing motion sickness (§4.1) and substantially impacts the sense of presence on Hubs (§4.2).

*e-mail: rcheng4@gmu.edu
†e-mail: emurat@gmu.edu
‡e-mail: craigyu@gmu.edu
§e-mail: sqchen@gmu.edu
¶e-mail: bohan@gmu.edu

2642-5254/24/$31.00 ©2024 IEEE
DOI 10.1109/VR58804.2024.00080

• Our measurement study and source code analysis of Hubs indicates that its low frame rate is attributed to one of its key design decisions: on-device local rendering[1]. Thus, the inferior rendering capability of untethered VR headsets becomes the performance bottleneck when the number of students is large (§4.3).
• In the small course with 10 students, the VR headset's frame rate on Hubs is satisfactory, offering a better overall experience and sense of place presence and social presence than the large course (§4.4).
• Compared to Zoom, avatar-based interactions on Hubs can enhance students' sense of place presence and social presence, and the head and hand tracking of VR headsets can further boost their sense of co-presence (§5).
• Immersive interactions on Hubs may not always be conducive to learning, as they could divert students' attention. Moreover, although wearing a headset makes students concentrate in class, utilizing tools such as search engines and note-taking on headsets is inconvenient and may reduce learning efficiency (§5).

We make the following contributions in this paper. First, we conduct an end-to-end measurement study of VR classrooms on Hubs to better understand user experience from a systems perspective. Second, we identify the performance bottleneck of Hubs, which substantially impacts large-course students' sense of presence and learning experience. Third, we demonstrate the effects of various features, such as 6DoF motion tracking and spatial sound on the headset, as well as avatar-based embodiment, on students' sense of presence and learning experience in a VR classroom on Hubs and compare them with those of Zoom-based online education. Finally, we point out opportunities to further improve the user experience of VR classrooms on Hubs, such as enhancing the design of avatars and learning tools.

Our findings have broad implications for the future development of online VR classrooms. To effectively accommodate a large number of students in regular teaching activities, the system architecture of VR classrooms must be designed with scalability as a primary consideration. Moreover, the content of VR-based online courses must be meticulously devised to leverage the immersive interaction nature of VR technology to enhance educational outcomes. Additionally, the incorporation of effective learning tools into VR classrooms and emotional reflection (*e.g.,* facial expressions during conversation) in the avatar design is essential to enhance the sense of presence and learning experience for students. This improvement, however, may exacerbate the challenges associated with computational resource utilization on VR headsets, making the scalability issues more pronounced. Careful consideration of these trade-offs must be taken in the development of VR classrooms for online education. We have released the source code and questionnaire used in this paper at `https://github.com/felixshing/Hubs_VR2024`. This work does not raise any ethical issues.

2 Background

2.1 Social VR

By integrating online social media and VR technologies, social VR becomes a major component of the Metaverse [11]. It facilitates social interaction and communication among users in a shared virtual space through their avatar embodiments, which are controlled, for example, by headset controllers or keyboards. This interaction differs greatly from video-conferencing applications, in which users interact with each other via microphones and cameras. Thus, these two types of platforms have different user embodiments: avatar-based on social VR *vs.* video-based for video-conferencing applications. Moreover, social VR platforms typically employ spatial audio [33], which creates a realistic auditory experience by dynamically adjusting the volume level users can hear based on the relative positions of their avatars. VR headsets are a common choice for accessing social VR platforms, which provide users with an immersive experience by tracking their hand and head motion. Besides headsets, PCs are another option to access social VR platforms [11]. However, PC users typically consume only traditional 2D content and their bodies are not tracked to drive the experience.

2.2 Mozilla Hubs

Mozilla Hubs is a browser-based and open-source social VR platform, which is built with A-Frame, a WebVR framework [73]. It adopts the local rendering technique, and thus the rendering of avatars and scenes is done on the client side. The servers of Hubs are responsible for exchanging data among users [58]. The Mediasoup server [57] utilizes Web Real-Time Communication (WebRTC) [25] to deliver video (*e.g.,* shared PC screen) and audio data. WebRTC integrates several network protocols. Among them, the Real-time Transport Protocol (RTP) [70], which runs over User Datagram Protocol (UDP), is used to carry audio and video content. The Phoenix server [59] uses the Hypertext Transfer Protocol Secure (HTTPS) protocol, which runs over the Transmission Control Protocol (TCP), for updating the position and orientation of avatars among users [60]. User devices render the updated scene, including other avatars, based on the updated coordinates received from the Phoenix server.

3 Experiment Setup

3.1 Participants and Course Design

We conducted our IRB-approved user studies in two graduate-level classes offered by a university. We started the study in a course with 36 students (13 females and 23 males) and one instructor (male). Among them, 29 students are 18–24 years old, and the others are 25–34 years old. 17 students (7 females and 10 males) used Oculus Quest 2 headsets. The rest of them and the instructor used their own PCs to access Hubs. During the above user study, we found that the frame rate of headsets on Hubs was significantly low, negatively affecting user experience. This was due to the large number of participants in the course, which significantly increases the rendering overhead of headsets (§4.3). Thus, to examine the performance of headsets and experience of headset users in a smaller setting, we conducted another user study with 9 students (5 females and 4 males) and the same instructor in a small-scale course. Among them, 2 students are 18–24 years old, and the rest are 25–34 years old. We provided 7 students with Oculus Quest 2. We let the remaining 2 students use their PCs to access Hubs to examine how the frame rate of PCs changes in the small course. Note that two students in the small course also participated in the large course. Thus, our user studies involved different 44 participants in total.

To ensure a fair comparison between online education on Hubs and on Zoom, the course content on Hubs was designed to be similar to that of traditional courses on Zoom. The structure of the courses on Hubs and Zoom consisted of a lecture stage (45 minutes), during which the instructor introduced a research topic on VR by sharing his PC screen, followed by breakout sessions (15 minutes) in which students discussed the material presented. We allowed students the flexibility to ask questions at any time and the instructor to mute students who make noise to ensure a fair comparison between Hubs and Zoom. We use the following notations to denote participants: H_i – the i^{th} headset user and P_i – the i^{th} PC user; L – large course user and S – small course user. For example, $H_4 - S$ represents the fourth headset user in the small course, and $P_1 - L$ represents the first PC user in the large course. We omit L and S when discussing the user experience on a specific course.

3.2 VR Classroom Design

Figure 1(a) shows the overall structure of the VR classroom we designed on Hubs. It consisted of two parts: a lecture room where the instructor and students can conduct classes and three breakout

[1]We have verified that local rendering is widely used in popular social VR platforms [13].

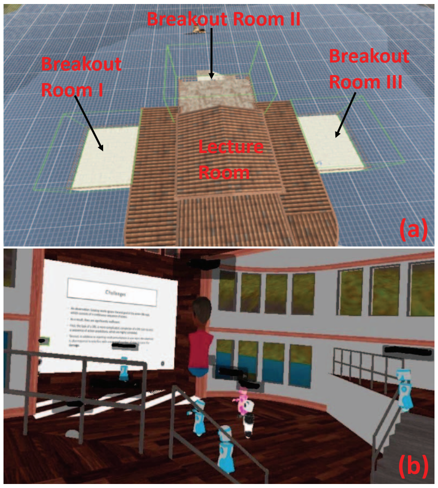


Figure 1: The virtual classroom we designed on Hubs. (a) Aerial view of the entire classroom (divided into two components, one lecture room and three breakout rooms). (b) The internal structure of the lecture room.

rooms where students can engage in discussions. In the lecture stage, the instructor shared the slides from his PC on a board mounted on the wall of the room, as shown in Figure 1(b). Students can "stand"[2] anywhere in the classroom to listen to the lecture. Similar to other social VR platforms, Hubs employs spatial sound that adjusts the volume of users' speech heard by others based on their distance [56], which can enhance their immersive experience and communication efficiency. To allow students to hear the instructor clearly during the breakout session, who may not be in the same room, we set up an audio enhancement zone on Hubs for the instructor. In this zone, the instructor's voice can be heard equally loud from all locations of the classroom. In the breakout session, a student can enter any breakout room to engage in discussions with others. We set up three rooms to prevent students from interfering with communication due to overcrowding in a single room.

3.3 Customized Setup of Hubs

Server Setup and Data Collection. Given that Hubs is an open-source platform, we deployed a private Hubs server on an Amazon AWS [3] EC2 instance (t3.medium), which was the recommended AWS server configuration when we conducted our user study. Note that our experiments showed the AWS instance with t3.medium configuration is not the performance bottleneck (§4.3). On the AWS server, we used Glances [22] to monitor its resource utilization. Moreover, we used tcpdump [77] to capture and analyze the network traffic on the server side. We also collected participants' public IP addresses. Thus, we can identify them in the network traces collected on the server side.

Client-side Data Collection. We instrumented the client-side source code of Hubs for data collection. We implemented a script to collect the following information from participants: 1) ID generated by Hubs; 2) what devices they use, 3) whether they have entered the classroom, and 4) whether they are muted, all of which are provided by A-Frame [73]. In addition, we collected the frame rate of Hubs on each participant's device as it is an important indicator of the system performance of Hubs. Since Hubs renders the virtual scene on the client side (§2.2), the frame rate reflects the local rendering quality. In order to measure and compare the performance of headsets, we let two users (H_1 and H_2) join Hubs in a controlled lab environment when participating in both the large-group and the small-group courses. Their headsets were connected to a WiFi access point (AP) attached to a campus network. The throughput of this AP was ∼350 Mbps for uplink and ∼700 Mbps for downlink. We utilized tcpdump [77] on the AP to capture and analyze network traffic. Meanwhile, we ran the OVR Metrics Tool [62], an official performance monitoring tool from Oculus, to measure the CPU, GPU, and memory utilization of these two headsets. We fully charged both headsets and terminated all background processes except the OVR Metrics Tool before data collection. The remaining participants joined the Hubs sessions from different locations of their choice, congruent with the typical setting of online education.

We conducted multiple sessions and measured the server-side and client-side performance for each course (small or large) on Hubs. Since the measurement results are similar, we present only those from the most representative session of each course in this paper.

[2]Due to the lack of full-body tracking, users' avatars always "stand" on Hubs, as shown in Figure 1(b), regardless of their pose in the real world.

3.4 Survey and User Interview Design

We designed a survey to understand participants' experience of taking classes on Hubs and Zoom, respectively. After demographic questions, participants filled out several 5-point Likert scale questions. We first let them rate their overall experience on Hubs and Zoom. We then asked them to rate the visual and audio quality of the two platforms, indicating the platforms' system performance. In the questionnaire, we defined visual quality as the system's frame rate and audio quality as the clarity of other users' voices. Next, we asked participants to rate their sense of place presence, social presence, and co-presence on Hubs and Zoom, which are distinct dimensions of presence in the virtual world. Bulu [8] defined the three dimensions of presence as: 1) Place presence: "*the sense of being there*"; 2) Social presence: "*the degree of salience of the other person in the interaction and the consequent salience of the interpersonal relationships*"; and 3) Co-presence: "*the degree of psychological connection of minds with others*". We asked participants to rate the three dimensions of presence since: 1) avatar-based embodiment in VR can enhance place presence, which is the foundation of social presence and co-presence [30]; and 2) social presence and co-presence are crucial aspects that influence the learning experience of VR-based online education [36, 44], resulting in deeper cognitive processing and enhancing learning outcomes [48]. After participants filled out the survey, we interviewed them to understand the reasons behind their answers. Moreover, we asked them two additional questions: 1) *When you attend the class on Mozilla Hubs, have you experienced motion sickness? If so, please describe the scene that caused motion sickness.* 2) *Comparing Mozilla Hubs and Zoom, which platform gives you a better learning experience, and why?*

When measuring users' sense of place presence, social presence, and co-presence, we chose not to use standardized surveys because completing all of them for each session on Hubs and Zoom would take excessive time. In our pilot study, we recruited five users, separate from the 44 participants in the main study. They were tasked to attend the same number of sessions on Hubs and Zoom, the same as the main study. After each session, we let them finish the *presence questionnaire* [83], the *social presence module of the game experience questionnaire* [28], and the *Co-presence of Other People* part of the survey designed by Poeschi *et al.* [66] to measure their place presence, social presence, and co-presence of each session on Zoom and Hubs, respectively. The average completion time to finish all surveys was 31.2 minutes ($SD: 6.8$). Given that such a long completion time may negatively impact response quality [16], instead of relying on questions in standardized surveys, we adopted the definition of Bulu [8] to create more concise questionnaires for measuring users' sense of place presence, social presence, and co-presence. By doing this, we optimized the average survey completion time to 13.7 minutes ($SD: 2.8$).

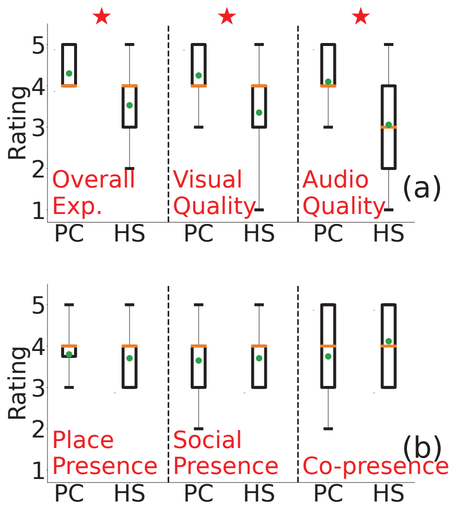


Figure 2: Comparison of the ratings of (a) the overall experience, visual and audio quality, and (b) the sense of presence, social presence, and co-presence on Hubs from PC and headset (HS) users in the large course. $\star: p \leq 0.05$. The box plot shows 95, 75, 25, and 5 percentiles, medium, and mean (green dots).

4 User Experience On Hubs

4.1 Overall Experience, Visual Quality, and Audio Quality

Figure 2(a) shows the ratings from PC and headset users regarding the overall experience, visual quality, and audio quality of Hubs for the large course. The Shapiro-Wilk test shows that the ratings of overall experience and audio quality from headset users are normally distributed, while the other four are not. Thus, we apply the Mann-Whitney U test to show the statistical significance [84]. We have the following observations.

• First, the ratings from the 17 headset users for the overall experience, visual quality, and audio quality are significantly lower than the 19 PC users (Overall experience: headset: $M = 3.53$, $SD = 1.09$; PC: $M = 4.30$, $SD = 0.55$; $p < 0.05$[3]; Rank-biserial correlation $r = 0.40$[4]; Visual quality: headset: $M = 3.35$, $SD = 1.18$; PC: $M = 4.25$, $SD = 0.43$; $p < 0.05$; $r = 0.40$; Audio quality: headset: $M = 3.05$, $SD = 1.30$; PC: $M = 4.10$, $SD = 0.73$; $p < 0.05$; $r = 0.45$).

• Second, the Spearman correlation test shows that the rating of the overall experience of headset users has a high correlation with those of visual quality ($k = 0.85$)[5] and audio quality ($k = 0.78$). Such correlations are not present in the ratings of PC users (both k values are less than 0.3).

• Third, the ratings of headset users have a larger variance than those of PC users. The standard deviation of ratings is 1.09 (0.55), 1.18 (0.43), and 1.30 (0.73) for headset (PC) users, regarding the overall experience, visual quality, and audio quality, respectively.

The above observations lead to our first hypothesis: *The visual and audio quality of headset VR on Hubs affects participants' overall experience*. Our interviews with headset users who rated 5 (four participants) or <3 (six participants) of their overall experience on Hubs confirm this hypothesis. All six headset users who gave a low rating attributed to the poor visual quality of Hubs. Moreover, four participants mentioned that the poor audio quality of Hubs made it difficult for them to understand the class content and communicate with others, negatively affecting their overall experience. For example, H_4, who rated 1 for overall experience, visual quality, and audio quality on Hubs, described that *"The experience on Mozilla Hubs was terrible. The screen lag was very severe, and I felt severe vertigo. Also, the others' voices were broken most of the time."* In contrast, all four headset users who rated 5 responded that the visual and audio quality of Hubs did not negatively impact their experience.

For the six participants who rated their overall experience on Hubs less than 3, while describing situations in which they experienced motion sickness, three of them directly attributed to the "low frame rate" or "low FPS" of Hubs. The other three described the low frame rate of Hubs as "picture lag" (H_8), "screen lag" (H_4), and "image is frozen" (H_9). Previous work has demonstrated that a low frame rate will significantly increase the probability of users experiencing motion sickness [85]. Thus, we further analyze the ratings of eight headset users who experienced motion sickness on Hubs. They rated their overall experience of Hubs ($M = 2.22, SD = 0.62$) below the average ($M = 3.53, SD = 1.09$). These findings confirm that motion sickness, which is induced by low frame rates (*i.e.,* poor visual quality), is an important factor that contributes to the inferior overall experience of headset users on Hubs, compared to PC users.

4.2 Place Presence, Social Presence, and Co-presence

Figure 2(b) shows the ratings of place presence, social presence, and co-presence on Hubs from PC and headset users. All ratings are not normally distributed based on the Shapiro-Wilk test. Moreover, the Mann-Whitney U test shows that there is no significant difference between ratings from PC and headset users of the three dimensions of presence on Hubs. This contradicts previous studies that headset can greatly enhance users' sense of presence in the VR environment [49, 61]. These results, combined with our previous finding that headsets' poor visual and audio quality negatively affects participants' overall experience, lead to our second hypothesis: *Poor visual and audio quality of Hubs deteriorates participants' sense of presence*.

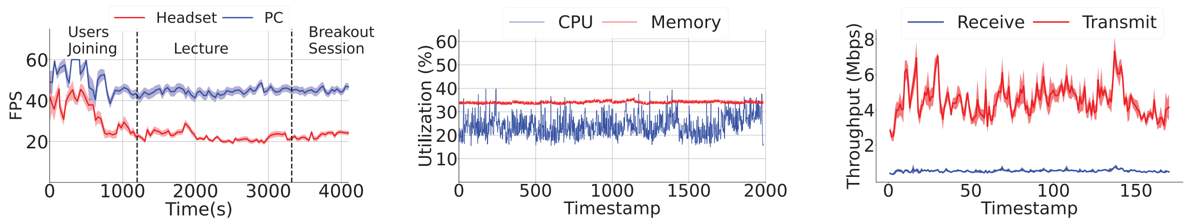
We first apply the Spearman correlation test to analyze the correlation between headset users' ratings of visual quality and audio quality and those of place presence, social presence, and co-presence. There exist correlations ($k > 0.5$) for all six combinations except for the combination of visual quality and place presence. User interviews explain the reason for these correlations. Poor audio and video quality caused participants to communicate with others less frequently, weakening their sense of social presence and co-presence. For example, H_8, who rated 2 for both social presence and co-presence, mentioned that *"When I am around a lot of people, the picture lags. So, I have to stay in a corner away from everyone with very little social interaction."* H_9, who rated 3 to social presence and 2 to co-presence, commented that *"The sound and vision were so bad that I could barely concentrate in class or interact with others."* H_4, who rated 2 for both social presence and co-presence, further explained that *"I experienced graphics and audio issues on the headset. So, I had to quit and rejoin the session and did not stay with others for very long."*

In addition, poor audio quality limited the efficiency of participants' conversations and impacted their sense of place presence. For example, H_5, who rated 1 to place presence, stated that *"It is difficult for me to converse with others because their voice is broken. And since I was not communicating well with others, they were unwilling to talk to me. It made me feel like I was not in the classroom."*. However, poor visual quality does not necessarily affect participants' sense of place presence, thanks to the immersive experience provided by headset VR. For example, H_{13}, who rated 2 for visual quality while 5 for place presence, explained that *"Since*

[3] p is the result of the statistically significant test. $p < 0.05$ is considered to be statistically significant [84].

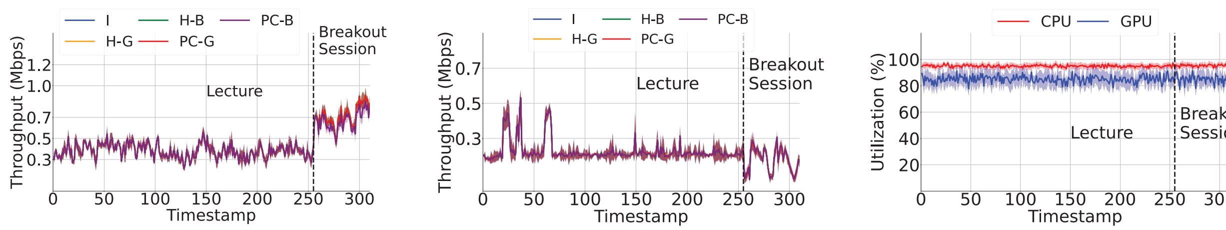
[4] Rank-biserial correlation r is the effect size of the Mann-Whitney U test, evaluating the strength of the claim of the significant test. $r > 0.1$ indicates the claim is valid [46].

[5] k is the correlation coefficient of the Spearman correlation test. $k \geq 0.3$ indicates two variables can be considered correlated and $0.7 \leq k \leq 1$ indicates that variables are highly correlated [34].



(a) Average FPS (b) Resource Utilization (c) Throughput

Figure 3: The average FPS of headset and PC users, the resource utilization on the customized Hubs server, and the throughput of the server in the large course. The bands in (a) represent 95% confidence intervals. The data in (b) and (c) was captured from the beginning of the lecture (each timestamp containing data collected by Glance [22] in the last 1–3 seconds).



(a) Downlink TCP traffic. (b) UDP traffic. (c) CPU and GPU utilization.

Figure 4: (a) The throughput of downlink TCP traffic for five selected participants. (b) The throughput of uplink UDP traffic from the instructor and downlink UDP traffic to the other four students. (c) The average CPU and GPU utilization on the headset of H_1. All data is captured in the large course. Each timestamp contains the records of the last 10 seconds. The bands represent 95% confidence intervals.

I was wearing a VR headset, I could always feel like I was in the classroom, even though sometimes I felt dizzy."

Summary. Based on the above qualitative and quantitative analysis, we have the following three findings: 1) headset users rated the visual and audio quality on Hubs significantly lower than PC users; 2) headset users had varying perceptions of the visual and audio quality on Hubs; and 3) headset users had a poor overall experience and sense of presence on Hubs due to its poor visual and audio quality. Since visual quality and audio quality are tied to the system performance of Hubs, we next explain these findings by analyzing the results of our measurement study.

4.3 Understanding User Experience on Mozilla Hubs: A Systems Perspective

We systematically measure the performance of Hubs to understand the bottlenecks that impact user experience. We begin by measuring the FPS of user devices on Hubs, which is the visual-quality indicator of system performance. We have the following key findings. The FPS of headset users is much lower than that of PC users, as shown in Figure 3(a). For example, during the lecture phase, the average FPS of headsets ($M = 22.58$, $SD = 3.91$) is 49.5% lower than PCs ($M = 44.68$, $SD = 3.45$). Moreover, the average FPS of headsets during the whole class is only $\sim$20, which is far below the minimum requirement of 60 FPS [7]. The average FPS of PCs, on the other hand, is larger than 30, satisfying the requirement for displaying 2D content on PCs [43]. We then explore the following possible reasons for this finding: 1) the computation and network utilization on the server is too high to sustain the required FPS of headset VR; 2) the server has different data forwarding policies for headset and PC users, resulting in varying amounts of data received by these two groups; and 3) different client devices have different capabilities when rendering content locally.

Resource Utilization on Server. Figures 3(b) and (c) show the CPU and memory utilization and the throughput of our customized Hubs server during the large course. We have the following observations. First, the CPU and memory utilization did not exceed 40% during the whole class, demonstrating that the computing capability of our server is adequate for tasks such as data forwarding. Second, the maximum throughput was less than 10 Mbps during the entire course, which is much lower than the maximum throughput allowed by the AWS EC2 `t3.medium` instance (5 Gbps) [2]. Third, the downlink (from the server to users) throughput was higher than the uplink (from users to the server) throughput. This is due to the fact that, for each piece of data uploaded by a user, the server must distribute it to all others. To summarize, the server's resource utilization and throughput are within the acceptable ranges, indicating that the server was not the performance bottleneck.

Data Forwarding Policy on Server. Next, we investigate whether the server employs a policy for transmitting different amounts of data to headset and PC users. This technique is widely used for adaptive content delivery [1], which aims to modify digital content, for example, by reducing its size, to suit the limited computational resources of mobile devices. To verify if Hubs utilizes this technique, we select five representative users: I (instructor), H-G/H-B (headset user whose device had good/bad system performance), and PC-G/PC-B (PC user whose device had good/bad system performance). We separate the network traffic of these five users collected on the server side to explore how the server exchanged data with them. Figure 4(a) shows the throughput of downlink TCP traffic, which is used to transmit avatar motion data (§2.2). We note that the server transmitted almost the same amount of data for each user. Moreover, the server transmitted more data in the breakout session stage than in the lecture stage. This is because, when listening to the lecture, users spent the majority of their time remaining stationary. During the breakout session, however, they freely moved around, which led to more data exchanges with the Phoenix server of Hubs (§2.2). Figure 4(b) shows the throughput of uplink UDP traffic received by the server from the instructor and the downlink UDP

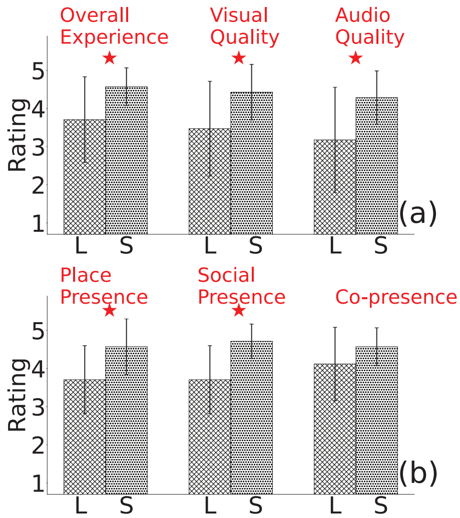


Figure 5: Comparison of the ratings of (a) the overall experience, visual quality, and audio quality, as well as (b) the presence, social presence, and co-presence on Hubs from large-course and small-course headset users. $\star : p \leq 0.05$.

traffic transmitted to the four selected students. Similar to the TCP traffic, the server transmitted almost the same amount of data to those participants, indicating that the server had the same data forwarding strategy for each participant. We identify that most of the UDP traffic was for the PC screen shared by the instructor based on the "Payload Type (PT)" field in the RTP packets [70]. Thus, the spikes in Figure 4(b) were likely introduced by the content of the shared screen (*e.g.,* interactive animations in the slides). In summary, the data forwarding policy of the server is not the cause of the headset's FPS being significantly lower than that of the PC.

Local Rendering. We next examine whether the quality of local rendering for headset VR is the root cause of its low frame rate. During our interviews with all headset users who rated the visual quality of Hubs less than 3 (six users), all of them mentioned that they could perceive a drop in frame rate with the joining of additional users, which is consistent with our measurement result, as shown in Figure 3(a). The reason is that with an increase in the number of users, there is a concomitant escalation in the volume of content that necessitates updates [60], leading to a heightened rendering workload. Considering their weak rendering capability [42], we hypothesize that untethered VR headsets such as Oculus Quest 2 could not handle such high rendering demand in the large course. To validate this hypothesis, we analyze the resource utilization of H_1's headset since it had a low frame rate (*e.g.,* during the lecture stage, $M = 18.24, SD = 6.21$). This is feasible because this participant took the course in our controlled lab environment (§3.3). Figure 4(c) shows the CPU and GPU utilization of H_1's headset, which, during the entire class, remained consistently high, especially for the CPU utilization (nearly 100%). Such a high CPU and GPU utilization indicates that the hardware capability of untethered headsets was unable to efficiently handle the rendering task in the large course, resulting in a drop in FPS. This explains why the headsets' FPS was significantly lower than that of PCs.

Summary. The above measurement study demonstrates that the low frame rate of untethered headsets was caused by the high rendering pressure imposed by a large number of participants (37) in the large course. We will discuss potential solutions to this issue in §6.1.

4.4 User Experience of Hubs for a Small Course.

To figure out whether the frame rate of headsets will improve with fewer participants, we conducted another user study on Hubs in a small-group course, which had seven headset users and three PC users. The frame rate of headset users in this small course (M: 61.2; SD: 16.4) was 164.9% higher than the large course (M: 23.1; SD: 7.8), resulting in significantly higher ratings on the overall experience, visual quality, audio quality, place presence, and social presence, as shown in Figure 5 (Overall experience: large course: $M = 3.53$, $SD = 1.09$; small course: $M = 4.57$, $SD = 0.49$; $p < 0.05$; $r = 0.46$; Visual quality: large course: $M = 3.35$, $SD = 1.18$; small course: $M = 4.42$, $SD = 0.72$; $p < 0.05$; $r = 0.46$; Audio quality: large course: $M = 3.05$, $SD = 1.30$; small course: $M = 4.28$, $SD = 0.69$; $p < 0.05$; $r = 0.53$; Place presence: large course: $M = 3.70$, $SD = 0.89$; small course: $M = 4.63$, $SD = 0.89$; $p < 0.05$; $r = 0.38$; Social presence: large course: $M = 3.70$, $SD = 0.96$; small course: $M = 4.71$, $SD = 0.45$; $p < 0.05$; $r = 0.47$). We use the bar plot instead of the box plot due to the limited number of headset users in the small course (*i.e.,* only seven). Although the low frame rate deteriorated headset users' sense of co-presence to some extent in the large course (§4.2), their ratings of co-presence were still satisfactory ($M = 4.11; SD = 0.96$), thanks to the hand and head motion tracking offered by VR headsets, which can increase the users' sense of co-presence (§5). Thus, there is no significant difference in the ratings of co-presence between the participants of the small and large courses. Moreover, the average FPS of two PCs ($M = 51.74$, $SD = 3.73$) is larger than 30, indicating that PCs can satisfy rendering requirements for both large and small courses.

Takeaways. We summarize our key findings by comparing the experience of PC and headset users in a VR classroom on Hubs.

- In the large class with 37 participants, the frame rate of headsets was much lower than that of PCs, which increased the probability that users felt motion sickness and negatively affected their overall experience and sense of presence.
- The low frame rate of untethered headsets was caused by the fact that the hardware was incapable of rendering high-quality VR content under the stressful workload caused by a large number of participants.
- In the small class with 10 participants, the frame rate of Hubs was satisfactory, enhancing participants' overall experience and their sense of place presence and social presence.

5 Online Education: Hubs *vs.* Zoom

5.1 Comparing User Experience on Hubs and Zoom

After investigating the user experience of online education on Hubs, we compare it with that on Zoom. Since the low frame rate of headsets affects the user experience on Hubs, we compare the ratings of PC and headset users separately.

Figures 6 and 7 show the rating of Hubs *vs.* Zoom on the overall experience, visual quality, audio quality, and the sense of place presence, social presence, and co-presence for the large course, given by the headset and PC users, respectively. The Shapiro-Wilk test shows that only the ratings of overall experience and audio quality from headset users are normally distributed, while others are not. From these figures, we have the following observations.

- First, the Wilcoxon signed-rank test indicates that headset users rated the overall experience, visual quality, and audio quality of Hubs significantly lower than Zoom, as shown in Figure 6(a) (Overall experience: Hubs: $M = 3.53$, $SD = 1.09$; Zoom: $M = 4.35$, $SD = 0.68$; $p < 0.05$; Rank-biserial correlation $r = 0.38$; Visual quality: Hubs: $M = 3.35$, $SD = 1.18$; Zoom: $M = 4.35$, $SD = 0.58$; $p < 0.05$; $r = 0.47$; Audio quality: Hubs: $M = 3.06$, $SD = 1.30$; Zoom: $M = 4.52$, $SD = 0.60$; $p < 0.01$; $r = 0.36$). This is because the visual and audio quality of the headsets on Hubs is poor (§4.1).

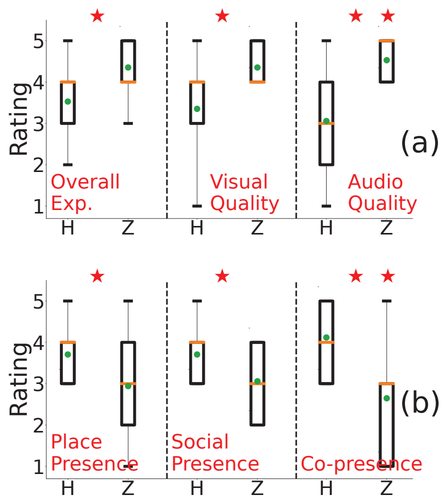


Figure 6: Comparison of the ratings of overall experience, visual and audio quality (a), place presence, social presence, and co-presence (b) on Hubs (H) and Zoom (Z) from headset users in the large course. $\star : p \leq 0.05$. $\star\star : p \leq 0.01$.

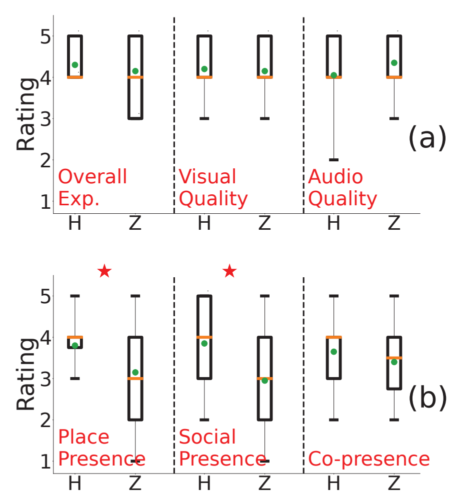


Figure 7: Comparison of the ratings of overall experience, visual and audio quality (a), place presence, social presence, and co-presence (b) on Hubs (H) and Zoom (Z) from PC users in the large course. $\star : p \leq 0.05$.

Conversely, there is no significant difference in the ratings from PC users, as shown in Figure 7(a).

• Second, Figure 6(b) and Figure 7(b) show that headset and PC users perceive a stronger place presence and social presence on Hubs than Zoom, thanks to avatar embodiment and spatial sound on Hubs (§5.2) (Headset users: place presence: Hubs: $M = 3.70$, $SD = 0.89$; Zoom: $M = 2.94$, $SD = 1.30$; $p < 0.05$; $r = 0.58$; social presence: Hubs: $M = 3.70$, $SD = 0.89$; Zoom: $M = 3.05$, $SD = 0.93$; $p < 0.05$; $r = 0.48$. PC users: place presence: Hubs: $M = 3.80$, $SD = 0.87$; Zoom: $M = 3.15$, $SD = 1.38$; $p < 0.05$; $r = 0.48$; social presence: Hubs: $M = 3.85$, $SD = 1.10$; Zoom: $M = 2.95$, $SD = 1.35$; $p < 0.05$; $r = 0.32$).

• Third, as shown in Figure 6(b), headset users also have a stronger sense of co-presence on Hubs than Zoom (Hubs: $M = 4.11$, $SD = 0.96$; Zoom: $M = 2.64$, $SD = 1.32$; $p < 0.01$; $r = 0.48$). Through user interviews, we find this is because they can benefit from the 6DoF hand and head motion enabled by headsets. For example, $H_{11} - L$, who rated 5 to co-presence on Hubs and 2 on Zoom, shared that *"I think the co-presence in VR is better because it is more realistic. I can use my body language to interact with others, like waving my hands to say Hi."* However, since PC users cannot enjoy such motion tracking on Hubs (§2.1), it may explain why they could not feel a stronger sense of co-presence on Hubs than Zoom, as shown in Figure 7(b).

We then analyze the ratings from headset users in the small-group course, as shown in Figure 8. The Shapiro-Wilk test reveals that the ratings of place presence and social presence on Zoom are normally distributed, while others are not. We apply the Wilcoxon signed-rank test and observe that there is no significant difference in the ratings of overall experience and visual and audio quality between the two platforms. In terms of the ratings of place presence, social presence, and co-presence, headset users in the small course, consistent with those in the large course, considered that Hubs is significantly better than Zoom (Place presence: Hubs: $M = 4.00$, $SD = 0.90$; Zoom: $M = 2.94$, $SD = 1.30$; $p < 0.05$; $r = 0.42$; Social presence: Hubs: $M = 4.11$, $SD = 0.85$; Zoom: $M = 3.05$, $SD = 0.93$; $p < 0.01$; $r = 0.32$; Co-presence: Hubs: $M = 4.23$, $SD = 0.80$; Zoom: $M = 2.64$, $SD = 1.32$; $p < 0.01$; $r = 0.31$). These findings indicate that headset users can perceive a stronger sense of all dimensions of presence on Hubs compared to Zoom.

5.2 Avatar Embodiment, Spatial Sound, and Presence

We next explore why users on Hubs could have a better sense of place presence and social presence than Zoom. User interviews reflect this benefits from avatar embodiment and spatial sound offered by Hubs. Avatar embodiment could give participants the impression that they are in a real classroom, boosting their sense of place presence on Hubs. For example, $PC_8 - L$, who rated 5 to place presence on Hubs and 3 on Zoom, stated that *"Having an avatar makes the sense of place presence pretty strong. I feel like I am really in a classroom and not in a web app."* $H_{17} - L$, who rated 4 to place presence on Hubs and 1 on Zoom, told us *"The place presence on Hubs is good because I can control my avatar to attend the lecture from many angles. It is more authentic than Zoom."* Furthermore, $H_4 - S$, who rated 5 to place presence on Hubs and 2 on Zoom, mentioned that *"Place presence of Mozilla Hubs is better than Zoom since we can explore the classroom as an avatar like in an in-person classroom."*

In addition, both spatial sound and avatar embodiment could boost participants' sense of social presence. For instance, $H_9 - S$, who rated 5 to the social presence on Hubs and 2 on Zoom, said that *"What I like the most about Hubs is that we can cluster together to talk and discuss. And I do not feel crowded or disturbed since others' voice is changed based on how close they are to me. This experience is very authentic."* $H_4 - S$, who rated 4 to the social presence on Hubs and 2 on Zoom, stated that *"I rated the social presence based on my group discussion experience on these two platforms. I think the spatial sound of Hubs is especially useful during group discussions because I can know who is talking. While on Zoom, if there are a lot of people talking at the same time, it will be very noisy."* $PC_{16} - L$, who rated 5 to the social presence on Hubs and 3 on Zoom, further commented that *"Hubs provides a much stronger social presence since I can stay in a hall with co-workers. This gives me a more comfortable environment to talk with others than Zoom."*

However, the avatar embodiment of Hubs could not enhance the sense of co-presence for participants mainly due to the following two reasons. First, the avatars of Hubs are cartoon-shaped instead of human-like. Some participants believe that only face-to-face

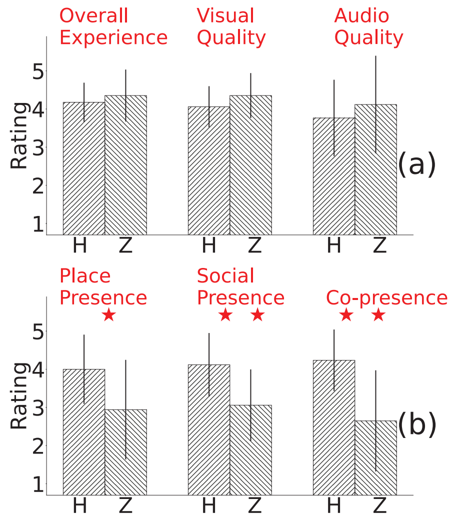


Figure 8: Comparison of the ratings of overall experience, visual and audio quality (a), place presence, social presence, and co-presence (b) on Hubs (H) and Zoom (Z) from headset users in the small course. $\star : p \leq 0.05$. $\star\star : p \leq 0.01$.

interactions can provide them with a sense of co-presence. For example, $H_9 - L$, who rated 2 to co-presence on Hubs and 4 to Zoom, stated that *"Mozilla Hubs had avatars to represent the users, while Zoom allows us to share the video of the actual user when attending a session. I prefer looking at the actual person when they are speaking rather than a user-generated avatar, which gives me a closer co-presence feel."* $H_5 - S$, who rated 1 to co-presence on Hubs and 4 to Zoom, further commented that *"I think the avatar of Hubs is pixelated and has a retro feel to it. I have a stronger co-presence on Zoom over Hubs since I can see people's faces on Zoom rather than a cartoon figure on Hubs."* Second, compared to video-based embodiments on Zoom, avatar-based embodiments on Hubs make it difficult for participants to observe the facial expressions and emotions of others. For instance, $H_{13} - L$, who rated 2 to co-presence on Hubs and 5 to Zoom, told us *"On Zoom, you can see others' facial expressions. This helps a lot to get a feeling of the tone of a situation and what another person's intentions are, which gives me a better co-presence."* $PC_{11} - L$, who rated 2 to co-presence on Hubs and 4 to Zoom, shared that *"I have a better co-presence on Zoom since I can get others' expressions and emotions on it."*

5.3 Comparing Learning Experience on Hubs and Zoom

Next, we analyze the learning experience of students on Hubs and Zoom. Prior work has demonstrated that a strong sense of social presence and co-presence can lead to a better learning experience in the VR environment [44, 48]. However, our study discovers that while the avatar-based interaction on Hubs boosts social presence for students, it could also distract them and deteriorate their learning experience. This observation was made by PC and headset users in both large and small courses on Hubs. For instance, $H_5 - S$ said that *"The movements of other students will distract me."* $PC_{20} - L$ stated that *"... the ability to move around and explore the VR world puts me off and doesn't feel like being in a classroom."* H_2 told us *"When I was in the large course with 30+ students, I felt very crowded, distracted, and annoyed by people walking around all the time."* Moreover, in the large course, the poor video and audio quality of headsets on Hubs makes it hard for students to focus. H_1 commented that *"I had to stay in the corner because I would get dizzy in a crowded room, which caused me to have trouble seeing the slides."* $H_8 - L$ shared that *"Motion sickness prevented me from thinking about the class content."* $H_4 - L$ stated that *"Low frame rates made me severely distracted."* $H_{12} - L$ told us *"I could not hear clearly what the lecturer was saying, so I did not understand what was going on in the class."*

Our interviews further reveal that compared to PCs, using learning tools on headsets is inconvenient, which lowers the learning effectiveness of students. For example, $H_{13} - L$ said that *"Taking Zoom classes allows me to google for information, take notes, and check the resources shared by the teacher. While using Hubs for them is not convenient."* $H_{11} - L$ shared that *"I like typing to ask questions or discuss with others since I think word is more understandable than voice. However, typing on VR headsets is not as convenient as on a PC."* $H_6 - L$ mentioned that *"I cannot take notes and feel hard to take a screenshot for slides when wearing the headset."* Nevertheless, wearing headsets can force students to concentrate. For example, $H_{10} - L$ explained that *"With the Oculus device, it was almost impossible to do anything else other than pay attention unless one took the device off."* $H_{14} - L$ commented that *"I believe the VR headset is a great teaching tool because you cannot do anything but listen to the lesson after wearing it. This is important in the online classroom because most students using PCs in class get distracted by other content."* In addition, several PC and headset users from both large and small courses reported that classes on Hubs were more engaging than those on Zoom, resulting in increased motivation for learning. For example, $PC_{17} - L$ told us *"I am tired of taking classes on Zoom. Attending classes on the social VR platform is a fun experience to indulge in the classroom."* $H_{15} - L$ shared that *"Attending classes on Hubs is definitely a new experience and makes me feel excited in taking classes"*. $H_6 - S$ stated that *"I felt tired and often distracted on Zoom. But on Hubs I felt interested and happy to listen to the lectures."*

Takeaways. Through comparing the user experience of attending courses on Hubs and Zoom, we have the following key findings.

- Compared to Zoom, avatar-based embodiment and spacial sound on Hubs make students feel a stronger sense of place presence and social presence. Moreover, the hand and head movement tracking of the headset further enhances their sense of co-presence.
- Although avatar-based interaction on Hubs improves students' sense of social presence, it may distract their attention, negatively impacting their learning experience. In addition, the poor video and audio quality in the large course on Hubs makes it difficult for students to focus.
- Wearing headsets can force students to pay attention to the class. However, using study tools such as search engines and note-taking on headsets is inconvenient and decreases their learning efficiency.

6 Discussion

In this section, we present the implications of our findings on the development of online education in the Metaverse and discuss our future work.

6.1 Online Education in Metaverse

Scalability. Our measurement results demonstrate that the frame rate of headsets on Hubs declines as the number of users increases, indicating that Hubs faces the scalability issue for large-group classes. We have verified that such scalability issues also exist on other popular commercial social VR platforms [13], such as Horizon Worlds [51], VRChat [79], and Rec Room [68]. The root cause is that they all employ the local rendering technique, in which the platform server is solely responsible for forwarding data without further processing. As the rendering overhead increases with the number of users, the weak computation capability of mobile VR headsets is unable to maintain high-quality rendering. In this case, networking has a limited impact on the scalability issue. As explained in §3.3, $H1$ accessed Hubs in a lab environment with a high-bandwidth network.

Remote rendering [42] that offloads the rendering burden to the server is a promising solution to address the scalability issue. For remote rendering, the server renders the whole scene in the user's viewport as a 2D video frame. Consequently, the amount of transmitted data is independent of the number of users. The key challenge of remote rendering is to accurately predict the future viewport, as the rendering and delivery of video frames take time.

Avatar Embodiment. Our work reveals that avatars on Hubs cannot enhance users' sense of co-presence compared to Zoom since they are cartoon-shaped and cannot reflect users' emotions. This finding promotes the creation of more realistic avatars for VR classrooms. However, there is a trade-off between avatar embodiment, user experience, and system performance. To improve user experience, detailed avatar embodiment requires more sensors to capture users' motion through kinematics [15, 80], which leads to higher network bandwidth for transmitting data and higher rendering overhead [40]. For instance, Holoportation [63], a system that reconstructs a photorealistic 3D model of the human body in real-time, requires a 10 Gbps network connection for data transmission and four PCs with high-end GPUs for rendering. Thus, when designing high-fidelity avatars, the balance between user experience and system overhead should be carefully considered.

Teaching Material Creation. Our findings indicate that students can perceive a stronger sense of place presence and social presence for both headset and PC users and co-presence for headset users on Hubs than on Zoom, mainly due to the immersive VR environment. Thus, we should build highly realistic 3D models for teaching in VR and simulate the learning environment with objects in the real world, for example, in surgical rooms, automotive workshops, and museums [26]. Displaying such models, however, remains a challenge due to the heavy rendering load they impose [35]. In addition, immersive technologies such as haptic feedback sensors may also be able to enhance teaching effectiveness [26]. Therefore, how to design practical and effective teaching materials deserves an in-depth study.

6.2 Future Work

Survey Validity. As delineated in §3.4, we did not use existing surveys to measure users' sense of presence because our pilot study revealed that it is overly time-consuming, diminishing the response quality. Instead, we designed a more concise survey grounded in well-established definitions of presence [8], thereby potentially preserving its validity. In our future work, we will demonstrate the validity of our designed survey by applying the exploratory factor analysis [18]. Additionally, we intend to incorporate existing surveys of measuring presence [6, 83] into ours, for example, by selecting and summarizing the relevant items from them, to enrich our survey without unduly extending the completion time for participants.

Deeper Understanding of Learning Experience. In this work, we investigated the learning experience of students by letting them describe their feedback in interviews. We did not specifically examine the effects of VR classrooms on various aspects of the learning experience, such as interest and motivation [27], as well as outcomes [41]. Thus, we plan to leverage widely used online education surveys such as the online learning enrollment intentions (OLEI) scale [37] to conduct a comprehensive user study and gain a deeper understanding of the impact of VR-based online education on the learning experience of students.

Other Social XR Platforms. We conducted our user studies on Mozilla Hubs, an open-source platform. There are other social VR platforms, such as Horizon Workrooms [50] and Rec Room [68], as well as multi-user collaborative AR/MR platforms, such as Microsoft Mesh [53], which can be utilized for educational purposes. We will investigate the user experience on these platforms for online education and compare the differences in supporting teaching and learning activities across them.

7 Related Work

7.1 Social VR

Social VR [74] has gained great interest from HCI, VR, and network communities. HCI researchers have investigated a vast array of themes in social VR, such as interpersonal relationships [75, 76], event organization [65, 82], non-verbal communication [45, 76], and design principles [32, 17]. In the VR community, Friston *et al.* [21] built an open-source social VR system. Sykownik *et al.* [74] conducted a user survey to understand the motivation for using social VR platforms. In the network community, Cheng *et al.* [11, 13] performed network measurement studies on several popular commercial social VR platforms. In this paper, we utilize Mozilla Hubs, a social VR platform, to study online education in the Metaverse.

7.2 VR-based Learning

VR technology is becoming increasingly popular in online education because of its potential to boost students' learning motivation and performance [5, 14, 23, 47, 64]. It has also been utilized in several areas for training purposes, such as the automotive industry and medicine [10, 67]. Previous work has demonstrated the benefits of VR in learning activities by understanding the perspectives of various stakeholders, including students [47], instructors [14, 23], and VR developers [31]. The VR community mainly focuses on using qualitative and/or quantitative analysis to understand user experience in VR classrooms. Parmar *et al.* [64] examined how avatars in VR affect the learning experience of middle school education. Belani *et al.* [5] investigated the impact of different spatial representations of learning content on user experience. In contrast to the above work, we investigate the user experience in VR classrooms from a systematic measurement perspective.

7.3 Online Learning and Education

Online learning [4] removes geographic barriers and allows students to enjoy a desirable education remotely. Kizilcec *et al.* [37] adopted a quantitative analysis to understand students' motivations for choosing online education. Kwon *et al.* [38] compared the appeal of multiple online teaching methods. Martinez *et al.* [54] examined how anonymity and identifiability affect the interaction of students in large multicultural classrooms. Hamilton *et al.* [24] proposed an online learning platform for context sharing and participation. In this work, we study user experience in the emerging online educational platform – social VR.

8 Conclusion

In this paper, we presented a novel analytic method that combines qualitative comparison, quantitative analysis, and systematic measurements to understand user experience in VR classrooms, by using Mozilla Hubs as a case study. Our work began with qualitative comparison and quantitative analysis, indicating that headset users had a poorer experience on Hubs than PC users in a large course. We then leveraged end-to-end measurement studies to figure out the root cause, which is the high rendering demand with 30+ participants on Hubs that cannot be handled by untethered headsets. Moreover, we found that avatar-based interaction can improve students' sense of place presence and social presence, while the hand and head movement tracking of headsets can future enhance their sense of co-presence. We hope our work can shed light on the system design of social VR platforms to further boost the user experience of online education in the Metaverse.

Acknowledgments

We extend our thanks to the anonymous reviewers for their insightful comments, and to the user study participants for their valuable contributions. This work was partially supported by the U.S. National Science Foundation, under Grants CNS-2007153 and CNS-2235049.

REFERENCES

[1] V. Adzic, H. Kalva, and B. Furht. A Survey of Multimedia Content Adaptation for Mobile Devices. *Multimedia Tools and Applications*, 51:379–396, 2011. 5

[2] Amazon. Amazon EC2 T3 Instances. `https://aws.amazon.com/ec2/instance-types/t3/`, 2022. [accessed on 31-January-2024]. 5

[3] Amazon. Amazon Web Services (AWS). `https://aws.amazon.com/`, 2022. [accessed on 31-January-2024]. 3

[4] T. Anderson. *Theory and Practice of Online Learning*. Athabasca University Press, 2008. 9

[5] M. Belani, H. V. Singh, A. Parnami, and P. Singh. Investigating Spatial Representation of Learning Content in Virtual Reality Learning Environments. In *Proceedings of IEEE VR*, 2023. 9

[6] F. Biocca, C. Harms, and J. Gregg. The Networked Minds Measure of Social Presence: Pilot Test of the Factor Structure and Concurrent Validity. In *Proceedings of Annual International Workshop on Presence*, 2001. 9

[7] K. Boos, D. Chu, and E. Cuervo. FlashBack: Immersive Virtual Reality on Mobile Devices via Rendering Memoization. In *Proceedings of ACM International Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2016. 1, 5

[8] S. T. Bulu. Place Presence, Social Presence, Co-presence, and Satisfaction in Virtual Worlds. *Computers & Education*, 58(1):154–161, 2012. 3, 9

[9] G. Burnett, R. P. Kay, and C. Harvey. Future Visions for Higher Education: An Investigation of the Benefits of Virtual Reality for Teaching University Students. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 2021. 1

[10] F. Caputo, A. Greco, E. D'Amato, I. Notaro, and S. Spada. On the Use of Virtual Reality for a Human-centered Workplace Design. *Procedia Structural Integrity*, 8:297–308, 2018. 9

[11] R. Cheng, N. Wu, S. Chen, and B. Han. Reality Check of Metaverse: A First Look at Commercial Social Virtual Reality Platforms. In *Proceedings of IEEE Workshop for Building the Foundations of the Metaverse (Metabuild), co-located with IEEE VR*, 2022. 2, 9

[12] R. Cheng, N. Wu, S. Chen, and B. Han. Will Metaverse be NextG Internet? Vision, Hype, and Reality. *IEEE Network*, 36(5):197 – 204, 2022. 1

[13] R. Cheng, N. Wu, M. Varvello, S. Chen, and B. Han. Are We Ready for Metaverse? A Measurement Study of Social Virtual Reality Platforms. In *Proceedings of ACM SIGCOMM Conference on Internet Measurement (IMC)*, 2022. 2, 8, 9

[14] G. Cooper, H. Park, Z. Nasr, L. P. Thong, and R. Johnson. Using Virtual Reality in the Classroom: Preservice Teachers' Perceptions of Its Use as a Teaching and Learning Tool. *Educational Media International*, 56(1):1–13, 2019. 1, 9

[15] H. Dai, A. Valenzuela, and R. Tedrake. Whole-body Motion Planning with Centroidal Dynamics and Full Kinematics. In *Proceedings of IEEE-RAS International Conference on Humanoid Robots*, 2014. 9

[16] E. Deutskens, K. De Ruyter, M. Wetzels, and P. Oosterveld. Response Rate and Response Quality of Internet-based Surveys: an Experimental Study. *Marketing letters*, 15:21–36, 2004. 3

[17] R. Du, D. Li, and A. Varshney. Geollery: A Mixed Reality Social Media Platform. In *Proceedings of ACM CHI*, 2019. 9

[18] L. R. Fabrigar and D. T. Wegener. *Exploratory Factor Analysis*. 2011. 9

[19] M. M. Fisher and D. E. Baird. Humanizing User Experience Design Strategies with NEW Technologies: AR, VR, MR, ZOOM, ALLY and AI to Support Student Engagement and Retention in Higher Education. In *International perspectives on the role of technology in humanizing higher education*, volume 33, pages 105–129. 2020. 1

[20] G. Fransson, J. Holmberg, and C. Westelius. The Challenges of Using Head Mounted Virtual Reality in K-12 Schools From a Teacher Perspective. *Education and Information Technologies*, 25(4):3383–3404, 2020. 1

[21] S. J. Friston, B. J. Congdon, D. Swapp, L. Izzouzi, K. Brandstätter, D. Archer, O. Olkkonen, F. J. Thiel, and A. Steed. Ubiq: A System to Build Flexible Social Virtual Reality Experiences. In *Proceedings of ACM VRST*, 2021. 9

[22] Glances. `https://nicolargo.github.io/glances/`, 2015. [accessed on 31-January-2024]. 3, 5

[23] D. Hamilton, J. McKechnie, E. Edgerton, and C. Wilson. Implementing Immersive Virtual Reality in Higher Education: A Qualitative Study of Instructor Attitudes and Perspectives. *Innovative Practice in Higher Education*, 4(2), 2021. 1, 9

[24] W. A. Hamilton, N. Lupfer, N. Botello, T. Tesch, A. Stacy, J. Merrill, B. Williford, F. R. Bentley, and A. Kerne. Collaborative Live Media Curation: Shared Context for Participation in Online Learning. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*, 2018. 9

[25] C. Holmberg, S. Hakansson, and G. Eriksson. Web Real-Time Communication Use Cases and Requirements. `https://rfc-editor.org/rfc/rfc7478.txt`, 2015. [accessed on 31-January-2024]. 2

[26] H. Huang and C.-F. Lee. Factors Affecting Usability of 3D Model Learning in a Virtual Reality Environment. *Interactive Learning Environments*, 30(5):848–861, 2022. 9

[27] H.-M. Huang, U. Rauch, and S.-S. Liaw. Investigating Learners' Attitudes toward Virtual Reality Learning Environments: Based on a Constructivist Approach. *Computers & Education*, 55(3):1171–1182, 2010. 9

[28] W. A. IJsselsteijn, Y. A. De Kort, and K. Poels. The Game Experience Questionnaire. 2013. 3

[29] J. Iwanaga, Y. Kamura, Y. Nishimura, S. Terada, N. Kishimoto, T. Tanaka, and R. S. Tubbs. A New Option for Education During Surgical Procedures and Related Clinical Anatomy in a Virtual Reality Workspace. *Clinical Anatomy*, 34(3):496–503, 2021. 1

[30] L. Jensen and F. Konradsen. A Review of the Use of Virtual Reality Head-mounted Displays in Education and Training. *Education and Information Technologies*, 23(4):1515–1529, 2018. 3

[31] Q. Jin, Y. Liu, S. Yarosh, B. Han, and F. Qian. How Will VR Enter University Classrooms? Multi-stakeholders Investigation of VR in Higher Education. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*, 2022. 9

[32] M. Jonas, S. Said, D. Yu, C. Aiello, N. Furlo, and D. Zytko. Towards a Taxonomy of Social VR Application Design. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play Companion (CHI PLAY), Extended Abstracts*, 2019. 9

[33] J.-M. Jot, R. Audfray, M. Hertensteiner, and B. Schmidt. Rendering Spatial Sound for Interoperable Experiences in the Audio Metaverse. In *Proceedings of International Conference on Immersive and 3D Audio (I3DA)*, 2021. 2

[34] Keith G. Calkins. Correlation Coefficients. `https://www.andrews.edu/~calkins/math/edrm611/edrm05.htm`, 2022. [accessed on 31-January-2024]. 4

[35] B. Kim. Virtual Reality for 3D Modeling. *Beyond Reality: Augmented, Virtual, and Mixed Reality in the Library*, pages 31–46, 2019. 9

[36] D. Kim and T. Im. A Systematic Review of Virtual Reality-Based Education Research Using Latent Dirichlet Allocation: Focus on Topic Modeling Technique. *Mobile Information Systems*, 2022, 2022. 1, 3

[37] R. F. Kizilcec and E. Schneider. Motivation as a Lens to Understand Online Learners: Toward Data-Driven Design with the OLEI Scale. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 22(2):1–24, 2015. 9

[38] B. C. Kwon and B. Lee. A Comparative Evaluation on Online Learning Approaches Using Parallel Coordinate Visualization. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*, 2016. 9

[39] A. L. Lamont Gardenhire. Which devices can i use to join a hubs room? `https://support.mozilla.org/en-US/kb/which-devices-can-i-use-join-hubs-room/`, 2022. [accessed on 31-January-2024]. 1

[40] M. E. Latoschik, F. Kern, J.-P. Stauffert, A. Bartl, M. Botsch, and J.-L. Lugrin. Not Alone Here?! Scalability and User Experience of Embodied Ambient Crowds in Distributed Social Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):2134–2144, 2019. 9

[41] E. A.-L. Lee, K. W. Wong, and C. C. Fung. How Does Desktop Virtual Reality enhance Learning Outcomes? A Structural Equation Modeling Approach. *Computers & Education*, 55(4):1424–1442, 2010. 9

[42] L. Liu, R. Zhong, W. Zhang, Y. Liu, J. Zhang, L. Zhang, and M. Gruteser. Cutting the Cord: Designing a High-quality Untethered VR System with Low Latency Remote Rendering. In *Proceedings of ACM International Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2018. 1, 6, 9

[43] K. MacMillan, T. Mangla, J. Saxon, and N. Feamster. Measuring the Performance and Network Utilization of Popular Video Conferencing Applications. In *Proceedings of ACM Conference on Internet Measurement (IMC)*, 2021. 1, 5

[44] G. Makransky, P. Wismer, and R. E. Mayer. A Gender Matching Effect in Learning with Pedagogical Agents in an Immersive Virtual Reality Science Simulation. *Journal of Computer Assisted Learning*, 35(3):349–358, 2019. 3, 8

[45] D. Maloney, G. Freeman, and D. Y. Wohn. "Talking without a Voice" Understanding Non-verbal Communication in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–25, 2020. 9

[46] E. L. Martin and D. A. Martinez. The Effect Size in Scientific Publication. *Educacion XX1*, 26(1):09–17, 2023. 4

[47] T. J. Matome and M. Jantjies. Student Perceptions of Virtual Reality in Higher Education. In *Balancing the Tension between Digital Technologies and Learning Sciences*, pages 167–181. Springer, 2021. 1, 9

[48] R. E. Mayer. Principles Based on Social Cues in Multimedia Learning: Personalization, Voice, Image, and Embodiment Principles. *The Cambridge Handbook of Multimedia Learning*, 16:345–370, 2014. 1, 3, 8

[49] R. P. McMahan, D. A. Bowman, D. J. Zielinski, and R. B. Brady. Evaluating Display Fidelity and Interaction Fidelity in a Virtual Reality Game. *IEEE Transactions on Visualization and Computer Graphics*, 18(4):626–633, 2012. 4

[50] Meta. Horizon Workrooms. `https://www.oculus.com/workrooms/`, 2022. [accessed on 31-January-2024]. 9

[51] Meta. Horizon Worlds. `https://www.oculus.com/horizon-worlds/`, 2022. [accessed on 31-January-2024]. 1, 8

[52] Meta. Oculus Quest 2. `https://www.meta.com/quest/products/quest-2/`, 2022. [accessed on 31-January-2024]. 1

[53] Microsoft. Microsoft Mesh. `https://www.microsoft.com/en-us/mesh`, 2022. [accessed on 31-January-2024]. 9

[54] G. Morales-Martinez, P. Latreille, and P. Denny. Nationality and Gender Biases in Multicultural Online Learning Environments: The Effects of Anonymity . In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*, 2020. 9

[55] A. Mottelson and K. Hornbæk. Virtual Reality Studies Outside the Laboratory. In *Proceedings of ACM Symposium on Virtual Reality Software and Technology (VRST)*, 2017. 1

[56] Mozilla. Avatar Audio Source of Mozilla Hubs. `https://bit.ly/47UJHup`, 2022. [accessed on 31-January-2024]. 3

[57] Mozilla. Mediasoup based WebRTC SFU for Mozilla Hubs. `https://github.com/mozilla/dialog`, 2022. [accessed on 31-January-2024]. 2

[58] Mozilla. Mozilla Hubs System Overview. `https://hubs.mozilla.com/docs/system-overview.html`, 2022. [accessed on 31-January-2024]. 2

[59] Mozilla. Phoenix web server for Mozilla Hubs. `https://github.com/mozilla/reticulum`, 2022. [accessed on 31-January-2024]. 2

[60] Mozilla. Source Code of Updating Avatars' Coordinates on Mozilla Hubs. `https://github.com/mozilla/hubs/blob/master/src/network-schemas.js`, 2022. [accessed on 31-January-2024]. 2, 6

[61] M. Murcia-López and A. Steed. The Effect of Environmental Features, Self-Avatar, and Immersion on Object Location Memory in Virtual Environments. *Frontiers in ICT*, 3:24, 2016. 4

[62] Oculus. OVR Metrics Tool. `https://developer.oculus.com/documentation/unity/ts-ovrmetricstool/`, 2022. [accessed on 31-January-2024]. 3

[63] 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 ACM Symposium on User Interface Software and Technology (UIST)*, 2016. 9

[64] D. Parmar, L. Lin, N. Dsouza, S. Joerg, A. E. Leonard, S. B. Daily, and S. Babu. How Immersion and Self-Avatars in VR Affect Learning Programming and Computational Thinking in Middle School Education. *IEEE Transactions on Visualization and Computer Graphics*, 29(8):3698–3713, 2022. 1, 9

[65] R. Piitulainen, P. Hämäläinen, and E. D. Mekler. Vibing Together: Dance Experiences in Social Virtual Reality. In *Proceedings of ACM CHI*, 2022. 9

[66] S. Poeschl and N. Doering. Measuring Co-presence and Social Presence in Virtual Environments–Psychometric Construction of a German Scale for a Fear of Public Speaking Scenario. *Annual Review of Cybertherapy and Telemedicine*, pages 58–63, 2015. 3

[67] Y. Pulijala, M. Ma, M. Pears, D. Peebles, and A. Ayoub. An Innovative Virtual Reality Training Tool for Orthognathic Surgery. *International journal of oral and maxillofacial surgery*, 47(9):1199–1205, 2018. 9

[68] Rec Room. `https://recroom.com/`, 2022. [accessed on 31-January-2024]. 1, 8, 9

[69] G. Ripka, S. Grafe, and M. E. Latoschik. Peer Group Supervision in Zoom and Social VR-Preparing Preservice Teachers for Planning and Designing Digital Media Integrated Classes. In *EdMedia+ Innovate Learning*. Association for the Advancement of Computing in Education (AACE), " O'Reilly Media, Inc.", 2021. 1

[70] H. Schulzrinne, S. L. Casner, R. Frederick, and V. Jacobson. RTP: A Transport Protocol for Real-Time Applications. RFC 3550, 2003. [accessed on 31-January-2024]. 2, 6

[71] R. M. Simamora. The Challenges of Online Learning during the COVID-19 Pandemic: An Essay Analysis of Performing Arts Education Students. *Studies in Learning and Teaching*, 1(2):86–103, 2020. 1

[72] R. P. Singh, M. Javaid, R. Kataria, M. Tyagi, A. Haleem, and R. Suman. Significant Applications of Virtual Reality for COVID-19 Pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(4):661–664, 2020. 1

[73] Supermedium. A-frame. `https://aframe.io/`, 2022. [accessed on 31-January-2024]. 2, 3

[74] P. Sykownik, L. Graf, C. Zils, and M. Masuch. The Most Social Platform Ever? A Survey about Activities & Motives of Social VR Users. In *Proceedings of IEEE VR*, 2021. 9

[75] P. Sykownik and M. Masuch. The Experience of Social Touch in Multi-User Virtual Reality. In *Proceedings of ACM VRST*, 2021. 9

[76] T. J. Tanenbaum, N. Hartoonian, and J. Bryan. "How do I make this thing smile?": An Inventory of Expressive Nonverbal Communication in Commercial Social Virtual Reality Platforms. In *Proceedings of ACM CHI*, 2020. 9

[77] tcpdump. `https://www.tcpdump.org/`, 1999. [accessed on 31-January-2024]. 3

[78] Tech Times. Meta's Oculus Quest 2 is the Top Selling VR Headset of 2021. `https://bit.ly/3Ltruuq`, 2022. [accessed on 31-January-2024]. 1

[79] VRChat. VRChat. `https://hello.vrchat.com/`, 2022. [accessed on 31-January-2024]. 8

[80] K. J. Waldron and J. Schmiedeler. Kinematics. In *Springer Handbook of Robotics*, pages 11–36. Springer, 2016. 9

[81] Wikipedia. Zoom fatigue. `https://en.wikipedia.org/wiki/Zoom_fatigue`, 2022. [accessed on 31-January-2024]. 1

[82] 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 ACM Conference on Human Factors in Computing Systems (CHI)*, 2021. 9

[83] B. G. Witmer and M. J. Singer. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence*, 7(3):225–240, 1998. 3, 9

[84] K. Yatani. Statistical Methods for HCI Research. `https://yatani.jp/teaching/doku.php?id=hcistats:start`, 2023. 4

[85] D. J. Zielinski, H. M. Rao, M. A. Sommer, and R. Kopper. Exploring the Effects of Image Persistence in Low Frame Rate Virtual Environments. In *Proceedings of IEEE VR*, 2015. 1, 4