Understanding Social VR Streamers' Unique Challenges in Managing Cross-Reality Social Interactions Through Multi-dimensional VR Interfaces

Yang Hu
Clemson University
Clemson, South Carolina, USA
yhu3@clemson.edu

Guo Freeman
Clemson University
Clemson, South Carolina, USA
guof@clemson.edu

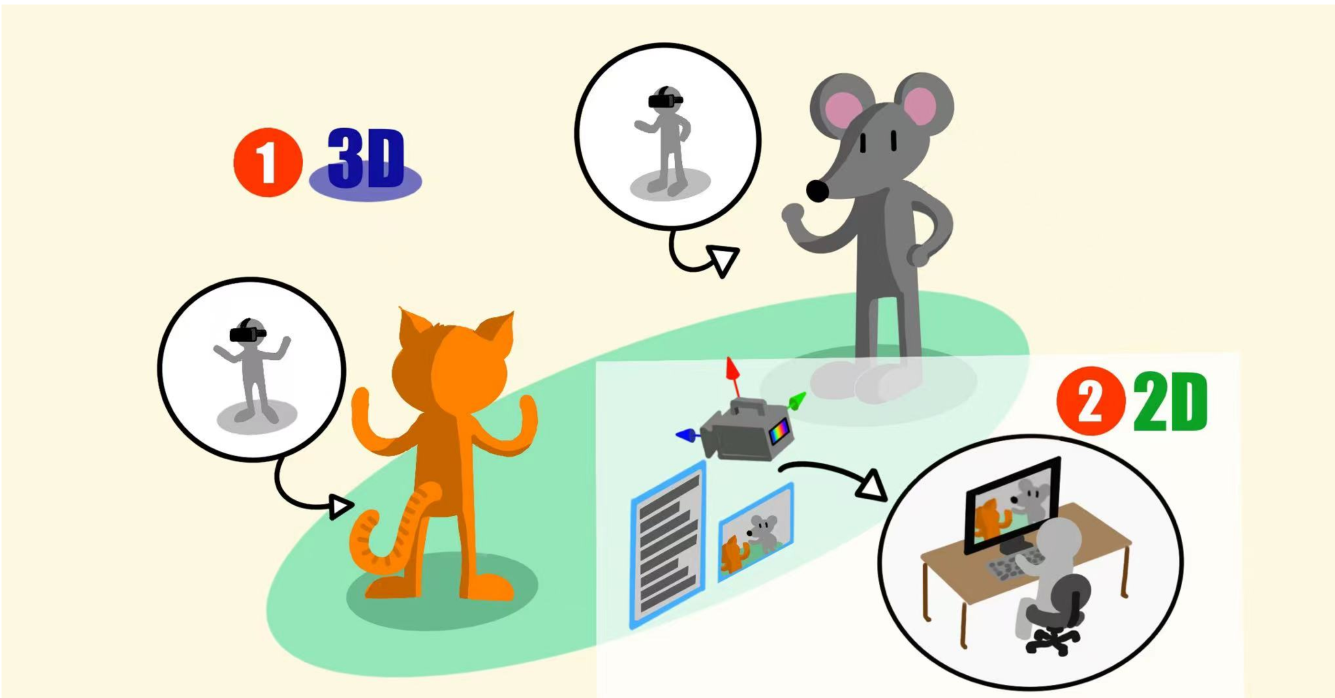

Figure 1: An example of cross-reality social interactions and multi-dimensional interfaces involved in social VR streaming. Left: A social VR streamer interacts with another social VR user using embodied avatars in social VR. Right: A social VR streamer monitors and controls 2D windows overlaid into VR spaces to manage live streaming software and read messages from their audiences outside VR; and their audiences can watch the streamer's activities in social VR through a 2D streaming video on a screen outside VR.

Abstract

Cross-reality interaction is a novel paradigm where users traverse and collaborate across virtual and physical realities. While prior research has investigated cross-reality systems through controlled experimental settings, how people navigate cross-reality interfaces in real-world contexts remains understudied. In this work, we focus on social VR streaming, an emerging practice where streamers engage in immersive VR activities while audiences view through traditional 2D interfaces. Through 17 interviews with experienced social VR streamers, we uncover their unique challenges in managing interfaces that span different realities to facilitate cross-reality interactions. We also highlight nuanced design spaces to facilitate cross-reality social engagement by addressing the difficulty of using and arranging existing 2D professional software within immersive 3D environments and the difficulty to maintain more natural and intuitive cross-reality interactions through everyday technologies rather than specialized hardware solutions. These insights advance our understanding of cross-reality interactions beyond experimental settings and can inform future interface design to better support everyday users' experiences of such interactions.

DIS '25, Funchal, Portugal

ACM ISBN 979-8-4007-1485-6/2025/07
https://doi.org/10.1145/3715336.3735825

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**.

Keywords

Cross-Reality Interactions, Social VR, Live Streaming, Interface Design

ACM Reference Format:

Yang Hu and Guo Freeman. 2025. Understanding Social VR Streamers' Unique Challenges in Managing Cross-Reality Social Interactions Through Multi-dimensional VR Interfaces. In *Designing Interactive Systems Conference (DIS '25), July 5–9, 2025, Funchal, Portugal*. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3715336.3735825

1 Introduction

Immersive technologies, such as Virtual Reality (VR) and Augmented Reality (AR), have become increasingly accessible in the consumer market and widely adopted in everyday contexts for work, learning, socializing, and entertainment. While earlier hardware implementations were typically restricted to fixed points along the Reality-Virtuality Continuum [44], more recently, there is a growing trend toward transcending these single-reality experiences into **cross-reality experiences.** Examples include applications that seamlessly transition between VR and AR modes [63] and the efforts to extend conventional mobile applications into immersive environments [7]. This technological convergence addresses the *isolation* issue often associated with immersive experiences (i.e., users in one reality are typically closed off from others, such as: people in VR cannot effectively share their experiences to people outside VR [18, 22, 27, 69]), both technically and socially. For example, on the technical side, the more open software ecosystem now enables people to access traditional 2D-based productivity tools they are familiar with in 3D virtual environments, such as using Microsoft Office applications in VR [3]. Likewise, on the social side, new technical features help people see and interact with others across different realities, including systems that render nearby physical bystanders within virtual environments [41] and "pass-through" capabilities in contemporary VR/AR devices like Meta Quest 3 [2] and Apple Vision Pro [1].

This growing convergence of different reality technologies has established **cross-reality interactions**, a novel interaction paradigm that allows individuals to navigate across different reality manifestations while facilitating gathering, communication, and collaboration with other people in different reality contexts [58]. In this sense, this emerging paradigm has potential to significantly reshape how people may connect and collaborate in future online spaces beyond single-reality environments. However, while prior research on cross-reality interactions has demonstrated how people in different realities can collaborate to combine their respective advantages (e.g., collaborators using AR to more precisely place objects to VR environments [56]), most studies tend to focus on controlled lab settings with customized software and hardware solutions that were built for the specific experimental tasks. In contrast, how to understand and address challenges in complex, real-world cross-reality scenarios remains largely unexplored.

Therefore, in this paper, we examine *social VR streaming* as a compelling research context to understand real-world users' unique challenges in cross-reality systems. Social VR streaming is an emerging phenomenon where a social VR user conducts immersive social activities through their embodied, virtual avatar **inside VR** while simultaneously broadcasting their experiences as 2D video content to viewers on live streaming platforms **outside VR** [16, 30]. We chose to focus on social VR streaming due to two key reasons. First, social VR streaming inherently demands cross-reality social interactions, as streamers must equally engage with both other users in 3D VR space and non-VR viewers through traditional text chat embedded in 2D live streaming platforms. Second, social VR streaming highlights the navigation of **multi-dimensional user interfaces**, meaning that social VR streamers often have to manage both 3D controls (e.g., using body-tracking controls for 3D social interactions and virtual camera manipulation for video streaming [67]) and 2D interfaces **in real time** (e.g., using ray-casting controls for overlaid 2D windows in VR showing audience messages and streaming controls [16, 26, 76]). Figure 1 illustrates these nuanced cross-reality social interactions and interface management. As technical or social missteps can be unpredictable and immediately visible to viewers, social VR streaming thus reveals more nuanced and natural challenges in managing cross-reality interactions compared to the pre-defined interactions in experimental cross-reality prototypes. Therefore, this complexity is especially valuable for understanding how to better support everyday users' experiences of cross-reality interactions through innovative interface design.

Using 17 in-depth interviews with experienced social VR streamers, we explore the following research question:

RQ: What are social VR streamers' main challenges in managing various interfaces in their VR streams to facilitate cross-reality social interactions and maintain stream qualities?

We make two primary contributions to the understanding and design of future cross-reality interactive systems. **First**, we expand the growing body of literature on cross-reality social interactions that mainly focus on controlled experimental settings with pre-defined interactions and purpose-built software. In doing so, we offer new insights into how real-world cross-reality practices, such as social VR streaming, present more nuanced challenges where complex workflows, professional software usage, and spontaneous social dynamics coexist across both VR and non-VR spaces. This further expands our understanding of cross-reality social interactions beyond idealized lab scenarios to practical everyday contexts. **Second**, we extend existing design principles for cross-reality systems (e.g., transitional interfaces, object substitution, and multi-user interfaces [6]) by revealing new design spaces for rethinking both *interfaces* and *interactions* to facilitate cross-reality social engagement. In particular, we highlight the critical needs to design future cross-reality systems that support complex 2D software in immersive environments and the coexistence of multidimensional interfaces. We also point out the importance of designing such systems to support more natural and intuitive cross-reality social interactions through everyday technologies (e.g., 2D text messages and videos on mobile phones) rather than specialized hardware (e.g., external physical screens for showing panoramic VR view [22, 27]). Taken together, these design suggestions can support more streamlined interface transitions and intuitive social interactions across realities in the future while mitigating identified technical and social challenges, which contributes to a broader research context around more effective, enjoyable, and accessible cross-reality environments for everyday users.

2 Related Works

2.1 Existing Practices and Challenges in Cross-Reality Interface Design

Cross-reality is an emerging practice that involves users and interaction paradigms across different realities (e.g., virtual, augmented, and physical). Its applications span both single-user (e.g., transitional interface that one user transits between VR, AR and desktop PC [20, 52]) and multi-user scenarios (e.g., hybrid collaborative workspaces where people in VR, AR and the physical world work together for shared tasks [45, 56, 59]). The recent research in cross-reality has began to contribute to various fields, including *education* (e.g., mirroring VR user's anatomy demonstration on a conventional projector [38] or 3D light field display [78]), *training* (e.g., remote experts using VR to guide a local AR user to perform accurate physical tasks on mechanical parts [47]), *filming* (e.g., an artist outside VR creates VR scenes for a director inside VR to test frames for pre-visualization [25]), and *gaming* (e.g., co-located non-VR players engage in VR players' games through external displays such as screens attached to the VR player's head [22] or projected images on the ground [21]).

Recognizing the potential of future cross-reality social interactions, previous research has established important design principles for cross-reality systems based on comprehensive experiments and lab studies, including principles for transitional interfaces (e.g., allowing smooth transitions between realities, using suitable metaphors, giving users control), substitutional interfaces (e.g., considering surrounding physical objects, integrating relevant physical objects in immersive spaces), and multi-user interfaces (e.g., including bystanders, enabling context understanding of different actualities) [6]. Collectively, these prior works have provided valuable frameworks for navigating between different reality states and highlighted two main advantages of cross-reality systems for social interactions and collaborative works.

First, cross-reality systems allow users in different realities to share visual perspectives and interact with each other, which facilitates communication and collaboration across reality boundaries. For example, they would allow non-VR users to see what VR users experience through external screens and projectors [27, 37, 71], help storyboard artists outside VR to view and control the movie director's environment in VR using tablets [25], and enable architectural space designers to design occupants' surroundings in VR using giant touch screens outside VR [59]. These cross-reality context sharing and collaboration can be used to address the isolation issue where users in one reality cannot effectively share their experience with people in other realities, which allows more natural communication and collaborative engagement beyond just a single reality. Second, cross-reality systems can combine the strengths of different interaction modalities in their respective realities. For example, traditional 2D interfaces (e.g., desktop PC) can foster precision [13, 55, 79], familiarity [7, 60, 64], and efficiency [19, 24, 32, 55], while novel VR/AR interfaces (e.g., VR headset with hand-tracking control) can better support spatial awareness [33, 48, 51, 62], immersion[15, 17, 28, 31, 65] and direct manipulation [5, 11, 29, 72]. These complementary benefits may lead to more novel interaction possibilities, such as using 2D software to precisely annotate in remote VR space [46], using AR aerial map view to arrange VR items [56], and mapping physical 2D tablet in VR for precise sketching [13] and modeling [60] in virtual environments.

However, while cross-reality systems have demonstrated promising capabilities for further innovating how we can interact and work together in future social spaces, they also introduce significant challenges in bridging different realities. Such challenges may include: unclear context awareness (e.g., understanding VR context while not in VR), difficult cross-reality communications (e.g., conveying non-verbal social cues to VR users), and incompatible interface modalities (e.g., annotating depth-sensitive information in VR while using a 2D interface) [6]. Yet, our current understanding of these challenges primarily stems from controlled experimental environments with carefully scoped tasks and customized tools [21, 22, 45, 49, 56, 59]. While these experimental studies provide valuable initial insights, they may not capture the full complexity of cross-reality interactions in real-world settings, particularly the nuanced social dynamics and emergent user behaviors [6]. Therefore, research on cross-reality interactions would benefit from empirical studies and insights beyond controlled environments on how users naturally perceive, adapt to, and overcome cross-reality challenges in authentic task scenarios. This motivates us to focus on social VR streaming as a valuable research context, which we detail in the next section.

2.2 Social VR Streaming as a Unique Context for Understanding Cross-Reality Social Interactions

Situating at the intersection of 3D Social VR and traditional 2D live streaming, social VR streaming shares significant similarities with existing cross-reality experiments in both social and technical aspects. First, similar to experiments on collaborative cross-reality workplaces where users need to maintain verbal communication with people in another reality [21, 22, 45, 56, 59], social VR streamers also need to closely communicate with people in different realities, such as bystanders within VR (e.g., other social VR users) and people outside VR (e.g., viewers and moderators) [16, 26, 76]. Second, similar to experiments on transitional interface settings where people need to shift between different realities and respective interaction modalities [20, 52], social VR streamers also need to manifest controls over both 3D space (e.g., embodied VR interactions with physical objects, immersive environments, and other people's avatars in social VR) and 2D interfaces (e.g., precise 2D interactions on overlaid windows for reading audience messages, monitoring body-tracking systems, and overlaying 2D visuals on the stream image) [16, 76]. These similarities thus make social VR streaming a valuable context for understanding how people experience cross-reality complexities in real world scenarios beyond the controlled lab settings.

In particular, while social VR streaming presents compelling parallels with existing cross-reality research, it further amplifies these cross-reality challenges both technically and socially. **On the one**

hand, social VR streamers encounter unique technical challenges in creating visually appealing content for non-VR viewers. This significantly differs from both experimental cross-reality systems and other avatar-based streaming such as virtual YouTubing [30, 35, 68]. First, while visibility across realities is a fundamental feature for experimental cross-reality systems, such visibility requirement can be limited to specific functionality needs and assisted by automated solutions (e.g., desktop users can view VR users' locations through real-time 2D map systems [56]). In contrast, social VR streamers must create engaging, creative, and situational content for their audience by themselves, which requires streamers to constantly coordinate across the virtual camera, the stream management software, and their full-body tracked virtual avatar [26, 30, 76]. Second, compared to virtual YouTubing where streamers establish their workflows in traditional 2D space (e.g., on-screen video gaming [30]), social VR streamers need to explore the 2D representations (e.g., as shown in on-screen live streaming) of their 3D avatars and immersive activities while in VR, which requires unique skills and strategies to create visually comfortable and appealing content for viewers on 2D platforms [26]. For example, to ensure a natural and consistent appearance of their virtual avatars across various VR environments and lighting conditions, social VR streamers usually need to quickly change or adjust parts of their virtual avatars in real time through in-VR menus [30].

On the other hand, social VR streamers face substantially more intensive social interactions compared to experimental cross-reality systems where interactions are typically task-centric and participants are pre-trained [6]. In contrast, for Social VR streamers, the social cues they need to process to facilitate cross-reality interactions can be more spontaneous and multifaceted. First, they need to manage voice communication, express body language through fully- or partially-body tracked avatars, and adapt to various social settings in virtual environments [17, 42, 61]. Recent studies highlight how streamers with virtual avatars must more actively maintain identity presentations (e.g., Virtual YouTube personas [35, 68]) by intentionally aligning their facial, bodily, and vocal performances with their avatar appearance and the virtual environment, adding both technical and social complexities [4, 12, 30]. Second, as live streamers, they also need to constantly attend to their audiences outside VR to maintain conversations and establish connections [23], primarily by dividing their attention to the overlaid chat window and strategically processing audience messages in VR space [16, 76]. Such audience activities include prioritizing viewers sending digital gifts or donations, balancing between general audience and a specific viewer, and following specific topics [36, 57, 74, 75, 80]. As such, social VR streamers' dual identity at the intersection of two spontaneous spaces with real time social interactions creates complexities far exceeding those in conventional cross-reality scenarios.

Therefore, given how social VR streaming presents these intensified cross-reality challenges in both social and technical dimensions, we believe that it offers a valuable and nuanced context for studying cross-reality interactions in authentic, real world settings, which may help the research gap identified in existing controlled experimental studies.

3 Methods

Recruitment and Participants. The university's Institutional Review Board (IRB) approved this study for research ethics. We recruited participants who have live-streamed themselves engaging in social VR spaces (e.g., VRChat) through three approaches. First, we directly contacted social VR streamers from our prior networks to gauge their interest in participation. Second, we identified and reached out to popular social VR streamers on major live streaming platforms (e.g., Twitch), deliberately selecting streamers with varying follower counts (from less than 1,000 to over 100,000) to ensure diverse experience levels and popularity (see Table 1). Third, we employed snowball sampling to recruit additional participants.

From October 2023 to July 2024, we conducted semi-structured interviews with 17 social VR streamers. While this sample size may appear modest due to the specialized nature of social VR To protect streamers' privacy, we did not collect personally identifiable information and used follower count ranges (<1,000, 1,000-5,000, 5,000-10,000, 10,000-50,000, 50,000-100,000, and >100,000) rather than exact numbers. These ranges reflect our participants' diverse experience levels and popularity while maintaining their anonymity. Notably, while some streamers in general may have remote moderators on live streaming platforms, all our participants operated their live streams independently, highlighting their personal challenges of managing both 3D social VR and 2D streaming simultaneously.

Interviews. Our semi-structured interviews specifically explored streamers' various challenges and struggles when conducting live streaming in social VR and their own efforts to address such challenges. Data saturation was achieved as no new major themes emerged in our final interviews. Before the interviews, we provided all participants with a consent document through their preferred communication channel (e.g., email or Discord). After obtaining consent, we scheduled and conducted interviews via either text or voice chat according to participant preference (e.g., Discord or Zoom).

We designed our interview questions using dialogic techniques to encourage deep, reflective responses [73]. In particular, we designed these interview questions mainly based on both existing literature on social VR, live streaming, and cross-reality interactions as well as our own practical experiences in this space. For example, we reviewed relevant research on live streaming (e.g., audience interaction [74] and gifting system [80] in traditional live streaming), social VR (e.g., collaboration [15] and streaming [26, 76] in social VR), and cross-reality interaction challenges (e.g., collaborative tasks across VR, AR and desktop PC [56]) to craft the initial interview guide. The first author, who has extensive experiences as an active social VR user and a live streamer, then iteratively refined the interview guide through pilot interviews and critical reflections to ensure that the interview questions captured streamers' nuanced challenges in cross-reality live streaming.

Each interview began with introductions, demographic questions, and queries about participants' experience levels with social VR platforms (e.g., VRChat, RecRoom, and Horizon Worlds) and live streaming in general. We then explored participants' technical practices and challenges specific to streaming in social VR spaces, particularly in comparison to traditional, non cross-reality streaming. Example questions included: *"How do you make your streaming*

P#	Gender	Age	Sexuality	Ethnicity	Social VR Platform	Streaming Platform	Number of Followers	Other Non-VR Content Streamed	Experience of Social VR	Experience of Social VR Streaming
P1	Man	25	Straight	Hispanic	VR Chat	Twitch	<1,000	N/A	5 years	4 years
P2	Trans Woman	26	N/A	White	VR Chat	Twitch	<1,000	Video Games	3 years	2 years
P3	Woman	26	Bisexual	N/A	VR Chat	Twitch	1,000-5,000	Video Games & Online Commentary	6 years	2 years
P4	Man	18	Straight	Asian	VR Chat	Twitch	<1,000	Video Games	3 years	2 years
P5	Woman	22	Bisexual	White	VR Chat	Twitch	1,000-5,000	Video Games	1.5 years	10 months
P6	Man	33	Bisexual	Black	VR Chat	Twitch & YouTube	1,000-5,000	3D Modeling & Video Games	6 years	3 years
P7	Non-Binary	36	Bisexual	White	VR Chat	Twitch & YouTube	1,000-5,000	Art & Video Games	1.5 years	3 years
P8	Gender Fluid	26	Demisexual	N/A	VR Chat	Twitch & YouTube	1,000-5,000	Video Games	2 years	5 months
P9	Woman	28	Lesbian	White	VR Chat	Twitch	10,000-50,000	Video Games	5 years	5 years
P10	Non-Binary	31	Biromantic Asexual	White	VR Chat	Twitch	1,000-5,000	Video Games	2 years	2 years
P11	Man	29	Pansexual	White	VR Chat	Twitch & YouTube & Kick	10,000-50,000	Video Games	5 years	5 years
P12	Non-binary	22	Bisexual	N/A	VRChat	Twitch & YouTube	1,000-5,000	Video Games	5 years	3 months
P13	Man	30	Straight	Asian	VR Chat	Bilibili	1,000-5,000	Video Games	2 years 8 months	9 months
P14	Woman	24	Straight	Asian	VR Chat	Bilibili	10,000-50,000	Video Games	2 years 4 months	7 months
P15	Gender Fluid	31	N/A	Asian	VR Chat	Bilibili	10,000-50,000	Video Games	5 years	3 months
P16	Man	20	Straight	Asian	VR Chat	Bilibili	1,000-5,000	Video Games	1 Year 2 months	3 months
P17	Man	23	Straight	Asian	VR Chat & Horizon Worlds	Bilibili & Tiktok	>100,000	N/A	2 Year 6 months	3 months

Table 1: Participants' offline demographics & online social VR streaming experiences. *N/A means participant information not provided.*

in social VR appealing to your viewers? What are the VR and non-VR tools you usually use to support your stream? What would you consider to be the main challenges in streaming in social VR compared to streaming other types of non-VR content?". Next, they were asked how they would socially interact with their audiences, moderators, and other people in social VR while streaming. Key questions relevant to this study included: *"How do you interact with your audiences and other social VR users? What are the differences between these two interactions? How do you balance between different social interactions?"* We then specifically asked about how streamers use VR and non-VR user interface for these cross-reality social interactions. Questions included: *What VR and non-VR user interfaces do you use during your stream? What are the purposes of these interfaces? What are the challenges you experienced when using these interfaces?* Finally, participants were encouraged to suggest improvements for future VR and streaming interfaces to address identified challenges and enhance their cross-reality social interactions during social VR live streaming. Interviews averaged 64 minutes in length, and participants received a$20 gift card after they completed the interview.

Data Analysis. After the interviews were complete, recordings were transcribed for data analysis. We then conducted an in-depth inductive qualitative analysis using thematic analysis [9, 10], as illustrated in Fig. 2. We chose a qualitative approach as it is particularly well-suited for investigating "how people interpret their experiences, how they construct their worlds, and what meaning they attribute to their experiences" [43]. Following McDonald et al.'s guidelines for defining reliability in qualitative analysis in CSCW and HCI practice, our analytical procedures focused on identifying concepts and themes (recurrent topics or meanings that represent

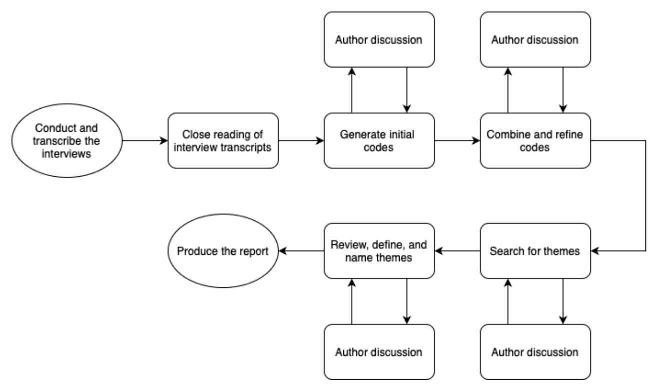

Figure 2: Overview of the data collection and analysis process in this study

phenomena) rather than seeking agreement [40]. As McDonald et al. note, coders may interpret the underlying meaning of codes differently even when they agree on the codes themselves [40]. Therefore, instead of pursuing inter-rater reliability, we focused on identifying recurring themes, detecting relationships among them, and formulating these into broader thematic clusters [40].

We followed Braun and Clarke's [10] six-step framework for thematic analysis:

1. Data Familiarization: The first author conducted a detailed line-by-line review of interview transcripts, taking notes and identifying information relevant to our research question about social VR streamers' challenges in managing interfaces for cross-reality interactions and stream quality.

2. Initial Code Generation: The first author developed preliminary codes through an iterative process. Two authors then collaborated to consolidate codes through an iterative process of open coding, discussing ambiguities and resolving disagreements at regular intervals. For example, the quote *"I have to constantly be aware of where my camera is in VR and where it's pointing at."* was initially coded as "difficulty in using virtual camera" and "not supportive camera system" which were then discussed and consolidated into "lack of assistance in camera position awareness."

3. Theme Development: These two authors categorized codes into thematic topics related to our research question and developed sub-themes from participants' descriptions. For instance, codes about managing audience chat interactions and technical streaming management in social VR were discussed and categorized under the theme *challenges in controlling external 2D software in a 3D VR environment*.

4. Theme Review: The two authors continued to review themes against the original data set to ensure they accurately represented participants' experiences, with particular attention to how these themes related to our research question.

5. Theme Definition and Naming: All authors collaborated to further refine and name the final set of themes.

6. Report Production: All authors selected compelling representative quotes and developed a logical structure for presenting the findings [10].

4 Findings

Drawing upon our interview data, we identify four critical challenges that social VR streamers face when managing cross-reality interfaces to simultaneously handle 3D social VR and 2D live streaming activities: (1) challenges in *controlling external 2D software in a 3D VR environment*, (2) challenges in *arranging 2D interfaces in a 3D VR environment*, (3) challenges in *maintaining 3D social awareness of 2D audience messages*, and (4) challenges in *maintaining 3D spatial awareness about the 2D camera system*. Table 2 summarizes our key findings.

Following established qualitative research practices (e.g., [39]), we focus on identifying and analyzing emergent themes across our dataset rather than quantifying the frequency of specific observations. As Maxwell [39] argues, numerical reporting of qualitative data can potentially mislead by suggesting unwarranted generalizability and overlooking crucial contextual factors.

4.1 Challenges in Controlling External 2D Software in a 3D VR Environment

Our participants unanimously noted the special need to overlay 2D PC interfaces within their VR environments for various streaming tasks. According to our participants, these interfaces typically include audience chat windows, streaming system managers (e.g., OBS), music players, and social VR community platforms (e.g., Discord).

To social VR streamers, these external 2D interfaces are indispensable for polished stream content. For example, P5 emphasized how music selection helps craft specific moods signified in her stream, *"I play music during my stream to get everyone into a mood, especially to match with an ongoing topic in my stream or a specific theme in a map."* According to P5, music switching occurs frequently as she navigates different game worlds to craft diverse streaming content and encounter various social VR users. This highlights how social VR streaming requires significant efforts to work not only in VR but also in 2D interfaces to create an overall engaging experience for viewers.

Figure 3: Examples of 2D window overlays social VR streamers need to manage for their streaming: (1) streaming manager (OBS) (2) audience chat, and (3) music and community software. (4) highlights the VR ray-casting control to navigate the menu items on 2D interfaces in VR.

In addition to music control, this careful curation also extends to visual elements controlled by different 2D interfaces in VR space:

"I have a cartoon television overlay in my stream. [...] So whenever I pretend to 'interview' someone in VR, I will make that visible (in OBS)." (P6)

"Before I dance, I usually pull my OBS up, turn up the BGM volume, and activate a filter image. It is like a semi-transparent image that overlays a color to my whole stream. [...] I want everything matches with my next dance." (P14)

For participants like P6 and P14, such coordination between different software applications is crucial to create *"cohesive viewing experiences"* (P14). As a result, such overlaid 2D interfaces can be intensively used by social VR streamers. However, due to **the fundamental mismatch between VR controls and 2D interfaces designed for traditional PC**, stream management in social VR presents significant challenges. P11 highlighted difficulties even in seemingly simple operations, *"It is much harder to be more impromptu with showing things on the (audience's) screen, even for simple things like playing a video."* He then elaborated on the contrast with traditional streaming,

"To show a video (in social VR), I need to reorder the visual overlay on OBS and click on the play button on my PC browser. [...] However, I cannot use the 'Alt Tab' (keyboard shortcut) to switch windows, so I have to use the VR pointer (ray-casting via VR controller) as a mouse cursor to manually find and switch windows." (P11)

In this case, because social VR streamers typically need to stand and move around during their streams, they do not have quick access to suitable PC controls (e.g., keyboard) for 2D windows in VR. As a result, compared to traditional desktop interfaces where input methods and visual feedback exist in the same dimensional space,

RQ: What unique challenges do users face when managing cross-reality interfaces for social interactions in VR environments?	**Key Findings**
Challenge 1: Controlling external 2D software in a 3D VR environment	- Social VR streaming requires managing existing **complex 2D interfaces** within VR for polished streaming image. - Streamers experience the **mismatch** between VR controls and 2D interfaces originally designed for traditional PC interaction. - Visual complexity and limited error recovery add considerable **psychological pressure** during streams. - **Identified gap:** current systems lack integration support between immersive 3D interfaces and widely adopted 2D software.
Challenge 2: Arranging 2D interfaces in a 3D VR environment	- Multiple essential 2D windows for stream management and audience engagement can cause **visual obstruction**. - **Large windows** needed for readable text and precise control can make the spatial conflict even worse. - **Identified gap:** current window overlay solution does not support flexible 2D and 3D interfaces co-display mechanisms to address the overlapping issue.
Challenge 3: Maintaining 3D social awareness of 2D audience messages	- 2D chat windows with specialized display for enhanced visibility can still be **overlooked** in VR. - Text-based communication **requires more active attention** compared to natural 3D social cues. - Current tangible notification systems **cannot support meaningful communication between streamers and viewers**. - **Identified gap:** existing systems lack translation mechanisms to display rich 2D text messages with intuitive 3D social cues.
Challenge 4: Maintaining 3D spatial awareness of 2D camera system	- Streamers need to **constantly monitor** 2D stream preview to confirm whether the viewers' viewpoint indeed matches the streamer's intention. - **Identified gap:** current virtual camera systems lack VR-native visual signifiers for easily maintaining awareness of cross-reality perspectives.

Table 2: Summary of key findings

social VR streamers must constantly switch between incongruent interaction paradigms (e.g., using controller ray-casting to click 2D user interface elements), which may cause frequent mistakes and significantly increase their cognitive loads. This challenge becomes particularly acute when streamers need to perform in an embodied approach, such as hand-tracking, during which even their VR controllers may be inaccessible. P9 explained,

"I don't use my controllers during dances, because I need to be expressive with my hand and finger movements. But, because I also need to use my controller before and after, to access menu and other stuff, there would be a big bump, as I need to firstly find my controller in my physical room."

As shown in this quote, the need to frequently switch between different interaction settings (e.g., controller-free 3D performance and controller-based 2D stream management) can add extra complexity to their streaming workflow and potentially disrupts the natural process of their content delivery (e.g., they have to stop and find controllers).

Further, while conducting simple tasks is already challenging, social VR streamers' cross-reality interactions are even more compounded by the visual complexities of some feature-rich PC interfaces they have to use in VR. Applications like OBS feature numerous small buttons and text elements designed for precise mouse interaction, such as toggle music mixers and adjustable visual element layers (See windows in Figure 3). This thus leads to more frequent errors when streamers have to manipulate this software through VR controllers. P11 shared,

"I can make several mistakes trying to click the button in OBS. [...] Also, as I cannot type, I usually keep lots of browser tabs for different videos and websites. It gets frustrating when I cannot find the right tab or even close the one I need by accident. [...] I must spend much longer time to fix it (compared to non-VR streaming)."

As social VR streamers' ability to make adjustments on 2D PC software is severely limited after entering the VR space, streamers usually prepare extensively before their stream to ensure smooth content delivery while in VR. For participants like P11, this inevitably results in a visually cluttered interface with numerous windows and tabs. Furthermore, as P11 noted, even minor mistakes (e.g., accidentally closing a pre-prepared video tab) can disrupt their carefully prepared content flow and force them to improvise solutions while still maintaining audience engagement. This combination of visual complexity and limited error recovery thus not only complicates streamers' workflow but also adds considerable psychological pressure during their streams.

4.2 Challenges in Arranging 2D Interfaces in a 3D VR Environment

Besides the difficulty of manipulating 2D controls in VR, our participants also highlighted the significant **spatial conflict** when arranging 2D interfaces in 3D social VR environments. As mentioned in the previous section, streamers need to frequently attend to several 2D interfaces in VR. While these 2D windows can be turned off and pulled up again, for easy and quick access, streamers usually keep several 2D windows around them. For example, P17 emphasized: *"I keep my OBS windows available all the time, so I can monitor my stream image. [...] I also don't want to miss anything from my viewers, so that one (audience chat window) is also up all the time."* For streamers like P17, to maintain both stream quality (e.g., through observing the real-time stream image mirrored to OBS) and audience engagement (e.g., through reading and replying to audience messages), constant presence of these 2D windows becomes essential.

However, from the streamer's view, these windows take up large amounts of visual space around them, which unfortunately makes their **surrounding work environment physically complex**. Indeed, while most window overlay solution supports transparency setting, the dense content they display (e.g., mirrored stream image and audience text) inevitably blocks streamers' vision in VR. Multiple participants described this issue:

"When I pull up my windows, I would be totally surrounded by them." (P13)

"If you were to be my POV of streaming, I'm basically looking at a wall of screens." (P11)

For participants like P11 and P13, maintaining multiple 2D windows in VR can dominate their surroundings and directly hinder their ability to view and interact with others in social VR. Furthermore, this visual challenge can be compounded by the use of large windows in VR. As mentioned in the previous section, feature-rich streaming software typically features dense controls (e.g., OBS) and text-heavy interfaces (e.g., audience chat). For these tasks, smaller windows can become impractical for social VR streamers. As P7 explained, *"when I need to read or control my PC windows in VR, I can't use small windows, or I would make mistakes. [...] This often blocks the (VR) world from my view. [...] I think I missed lots of things because of this."* As a result, this requirement for larger, legible interfaces with dense controls creates a fundamental tension between managing streaming tools and engaging with the actual virtual world, which ultimately constrains streamers' ability to create dynamic, immersive content for their viewers.

Additionally, this may extend beyond blocking static surroundings (e.g., VR environments and objects) to **affect dynamic gameplay experiences in social VR**. For instance, P11 highlighted how these windows can limit their ability to play games in social VR: *"It can sometimes be a little detrimental because I can't see a monster coming for me in a game."* For P11, these limitations not only lead to potential gameplay failures but also create unwanted interruptions for both streamers and their audience, compromising the overall streaming experience and content quality.

P3 further added that these visual obstructions would even significantly **impact streamers' social interactions with other social VR users**. As she elaborated, *"I usually have an overlay over my face, and I could only see around the edges to see other players come up to me and try to get my attention, but I would be busy watching these windows."* Therefore, when streamers are physically surrounded by multiple 2D interfaces, they are also socially isolated from surrounding social VR users. As social VR streamers are social VR users *per se*, their social interactions with other users are essential both for their own social VR experience and for content creation in their streams (e.g., interviewing other social VR users). Therefore, while 2D windows support essential stream management in VR, their current simple implementations lead to a dilemma where streamers must constantly shift between 2D and 3D interaction spaces rather than seamlessly integrating both, ultimately constraining their ability to make the best use of such cross-reality systems.

4.3 Challenges in Maintaining 3D Social Awareness of 2D Audience Messages

Among all the 2D windows social VR streamers need to manage, our participants unanimously emphasized that the audience chat window is the most crucial 2D interface, as it serves a pivotal role in building audience connections. As such, to maintain effective awareness of audience messages, streamers have developed various technological solutions for enhanced chat window visibility while navigating 3D space. For example, P8 described his approach to fix chat window in his view, *"The way I make my chat visible is to make it like a HUD (Heads-Up Display), so wherever I am looking at, it is always at the corner of my view."* Similarly, P1 explained their solution to attach the chat window on their body to maintain easy accessibility, *"All windows are in the air, but I put my chat on my wrist, just like a watch, so I can easily know what my viewers are talking about."* For both participants, they adopted specialized audience chat display different from other windows, which highlights the priority of audience interaction in their live streaming.

However, even with optimized chat window displays, streamers can still miss messages from their audiences and make delayed responses:

"My viewers usually comment more when I have something to do, like when we are having parties (in social VR). [...] But in that case, I might also miss their messages. So sometimes I have to handle several messages that stacked up during the past minute." (P7)

"When I am dancing or talking, I always put the chat window in the center (of my view). [...] But sometimes, like when I am playing a game, I cannot do that, or my view would be blocked. So I have to put it on the side and regularly check it." (P14)

Here, even with optimized display methods (e.g., using small body-fixed display), **managing audience chat and engaging in immersive VR activities can still become competing tasks**. Participants such as P4 explained why they still experience this dilemma, *"talking to people in VR is much easier, because you can see them walking to you, looking at you, and you can hear their sound. But, chat is different. You have to open it and read it more actively. [...] To me, it (chat reading) needs more energy."* In this case, compared to

3D social VR interactions where natural, multi-modal social cues can be more passively conveyed and sensed (e.g., through body languages, eye contacts, and sound), 2D chat requires more active, deliberate attention to process text messages, making it more likely for streamers to unconsciously overlook audience interactions even when they are easily visible.

It is worth noting that, to address this challenge, some streamers have begun to experiment with more tangible methods to make audience messages naturally visible in a 3D virtual space. These solutions typically involve audience *"redeems"*, a special action that viewers can trigger on 2D live streaming platforms to generate and project audiovisual effects in the streamer's 3D VR environment. These effects range from simple audio cues like *"applause sound"* (P8) and *"scream sound"* (P10) to more elaborate visual elements such as *"3D emoticon"* (P14) around streamer's avatar for showing audience support and *"steel tube"* (P16) that creates physical interactions by dropping onto the streamer's virtual head.

However, while these innovative solutions can foster streamers' awareness of audience presence, they often remain in simple forms and cannot support more meaningful and in-depth communications. As P16 reflected, *"Initially it was very popular. But now I feel it is more like a 'I'm here' message. [...] Sometimes it is just a reminder for me to read their chat."* For streamers like P16, although streamers can passively see and hear from audiences, these cues ultimately serve as notifications of the mere presence of their viewers. If streamers endeavor to engage in actual, meaningful communication with their audiences, they still need to shift their attention to carefully read the chat. In this sense, simple tangible messages still cannot support spontaneous and natural social interactions between streamers and viewers across different realities. This thus highlights a persistent gap in effectively conveying 2D social messages into the 3D immersive space.

4.4 Challenges in Maintaining 3D Spatial Awareness of the 2D Camera System

Besides the chat window, our participants emphasized another crucial interface for audience experience: the stream monitor window that displays camera-captured images (e.g., OBS). According to our participants, social VR streaming requires constant attention to this window to ensure proper framing and content delivery for their viewers:

"I have to constantly be aware of where my camera is in VR and where it's pointing at." (P10)

"To me, having a nice looking environment and having a nice looking frame is important. So, to some extent, I stream as a cinematographer [...] I spent lots of time testing those frames and compositions for my stream." (P2)

For these social VR streamers, their audiences' perspective (e.g., rectangular 2D videos) is fundamentally different from their own views in 3D immersive environments. Therefore, though streamers can directly manipulate the virtual camera as a 3D object (e.g., grabbing and moving) in VR to adjust the camera image (e.g., framing and composition), they still need to frequently inspect the 2D camera output through 2D interfaces (e.g., OBS) to ensure that their audiences' viewpoint is appropriate and aligns with their intention.

In this sense, similar to challenges in managing the chat window, streamers can also **fail to maintain attention to the camera preview during immersive activities.** This can cause frustrating moments when the camera fails to capture the correct content that streamers intend to show their viewers. For example, P12 shared their failed camera control when immersed in 3D activities, *"When I am talking to someone in VR, I sometimes refer to something or someone else around us (in VR), but my audiences actually couldn't see."* Here, P12 highlights that streamers' struggles with immersion in 3D activities and missed control through 2D camera monitoring may cause frequent uncertainties about the camera coverage while attending to immersive activities. This thus creates another persistent tension between streamers' immersive participation and technical monitoring for 2D audience experience.

Notably, virtual camera systems in social VR usually have a "following" mode that allows social VR users to keep the image focusing on themselves, which can be a viable solution for this issue. However, P12 revealed that even in the "following" mode, streamers may encounter similar challenges, *"by default, the camera points at me, but I cannot exactly know how much around me is captured, until I check my camera image."* Here, even with the feature to constantly track the virtual camera, our participants are still uncertain about the exact camera coverage for their viewers. Moreover, such a "following" mode does not support automated functions besides following the streamer's own face, which prevents streamers from crafting more creative camera usage. P15 explained,

"The camera has a follow mode, but it only track my face. [...] For example, it can't follow my hand. Or, I want it select something in VR, like tracking a car, or any other users. [...] The way to do it right now is still holding the camera and looking at the window (stream image monitor)."

P15 then referenced other VR games to illustrate how 3D camera systems should use 3D cues to support more natural usage,

"In some shooting games, you have those flash lights, and you can clearly see the light from your teammates, so you know where they are looking at."

Here, P15's example of using visible 3D frustum highlights how the visibility of VR camera viewpoint can be enhanced through immersive 3D visual cues instead of 2D windows. This thus reveals a significant design gap in generalizing intuitive VR designs for 2D content creation in VR.

5 Discussion

As mentioned in the Related Works section, existing works have established valuable design principles for cross-reality systems through controlled experiments (e.g., reality transition for single users, cross-reality objects substitutions for surrounding awareness, and multi-user context display across different realities [6]). While these works provide important foundational understandings, our findings offer new insights into how real-world cross-reality practices, such as social VR streaming, present more nuanced challenges that transcend these existing design guidelines, particularly when complex technical setups and spontaneous social interactions are involved (see Table 2).

One highlight from our findings is how streamers' diverse technical settings and audience scales might compound their challenges in managing social VR streaming identified in this paper. For example, while our participants tend to use modern commercial VR hardware (primarily Meta Quest 2/3 and Valve Index) and industry-standard streaming software (e.g., OBS), they usually customize these technologies (e.g., adding face-tracking hardware, using wrist display for the chat window) to adapt to their cross-reality environment. This suggests that the lack of universal cross-reality support pushes streamers to come up with their own workarounds, resulting in manual solutions that further expose the the fundamental mismatch between 2D and 3D interaction paradigms demonstrated in social VR streaming. Regarding audience sizes, our participants with larger followers (>10,000) seem to experience both heightened advantages and challenges. While these streamers typically have access to more sophisticated technical setups (e.g., dedicated streaming PCs and advanced body-tracking systems), they are also expected to produce higher quality content and constantly interact with a larger population of viewers, which further amplifies the difficulties for them to manage cross-reality social interactions. For instance, P9 and P14 (both have 10,000-50,000 followers) described that they had to craft more complex interface arrangements to accommodate additional audience engagement features, which also led to greater visual obstruction and attention management difficulties.

Building upon these findings, in this section, we highlight critical design spaces for addressing unique technical and social challenges in future cross-reality interactions in real world contexts. In particular, we structure our discussion around two complementary dimensions: **(1) rethinking interfaces**: cross-reality systems should require **granular interface transitions** rather than directly transitioning users between realities. This points to the urgent need to rethink and redesign established 2D tools, interfaces, and workflows to maintain compatibility with new 3D virtual spaces while avoiding potential spatial conflicts between interfaces; and **(2) rethinking interactions**: cross-reality systems should be designed to facilitate more **natural social interactions** beyond lab settings. This points to the urgent need to transform non-VR native inputs (e.g., text) into immersive 3D representations while simultaneously supporting multi-user context sharing by making 3D experiences accessible through familiar 2D formats for non-VR users (e.g., on-screen video).

5.1 Rethinking Interfaces: From 2D and 3D to Multidimensional

Our findings highlight how current cross-reality interfaces in social VR streaming are insufficient to support the complex workflows and spatial arrangements across various realities. As our finding have shown, streamers complain that many 2D desktop-based interfaces required for their streaming are not compatible in a 3D virtual space. In this section, we focus on two potential directions to rethink and redesign cross-reality interfaces. First, future cross-reality systems should support existing complex 2D software interfaces in immersive environments rather than frequently transiting users between reality spaces and developing specialized but limited cross-reality tools in experimental settings. Second, future cross-reality systems should address the fundamental tension between overlapping 2D and 3D interfaces to support the co-existence of multidimensional interfaces in cross-reality contexts.

5.1.1 Designing for Supporting Existing 2D Software Interfaces in Cross-Reality Systems. Current transitional interface principles in cross-reality systems primarily focus on shifting users smoothly between realities [20, 50, 54, 56, 58]. This means that different interaction spaces and interfaces are separate domains that require explicit transitions (e.g., using physical 2D interface would require people to exit VR first [6]). Yet, our findings highlight a more fundamental need to translate established workflows and software ecosystems into immersive environments for faster access and smoother task transitions. As cross-reality systems increasingly require coordination with traditional working devices like desktop computers [46, 52, 56], the disconnect between immersive 3D interfaces and widely adopted 2D software can be particularly pronounced.

Therefore, we suggest that future cross-reality interface design should prioritize compatibility with existing 2D-based software and workflows. This means focusing on adapting existing sophisticated 2D control systems for 3D manipulation without undermining functionality, which may help users in immersive 3D space to have easier access to established controls and complex operations that they are already familiar with. For example, future cross-reality systems could implement contextual input adaptation that dynamically transforms 2D interface elements based on VR interaction context. Instead of enlarging the whole window to ensure precision (P7), future designs can enable magnifier mechanics that automatically expand small buttons on any 2D interfaces when approached by VR controllers and provide haptic guidance for precise control. Additionally, similar to how our participants adopt various HUD and embodied displays for specific functionalities, future cross-reality systems could support more flexible interface decomposition, which allows users to extract only the most essential controls into the 3D space (e.g., camera flipping and chat transparency) while relegating other secondary functions (e.g., playing music) to auxiliary spaces. This approach would maintain the workflow continuity while mitigating the precision challenges that our participants have experienced when attempting to use complex 2D software in VR.

5.1.2 Designing for Addressing Multi-Dimensional Interface Interference in Cross-Reality Systems. Moreover, our findings reveal unique spatial challenges in social VR streaming that differ from current research approaches that mainly focus on spatial conflicts in the physical world (e.g., between VR users' body and the physical objects around them [6]). In our study, social VR streamers experience similar visual blockage even in the virtual space, as too many interfaces imported from another reality (e.g., 2D chat windows) can cause both physical and mental burdens during their cross-reality practices.

Such visual interference between 2D and 3D interfaces stems from current cross-reality systems' limited approach to window display options. For example, existing virtual window control tends to focus on individual window positioning methods (e.g., world-locked, view-locked, and body-locked [14]) instead of addressing the visual and physical interference between multiple interfaces in dynamic virtual environments. Furthermore, as users need to maintain continuous awareness of both 2D information and 3D

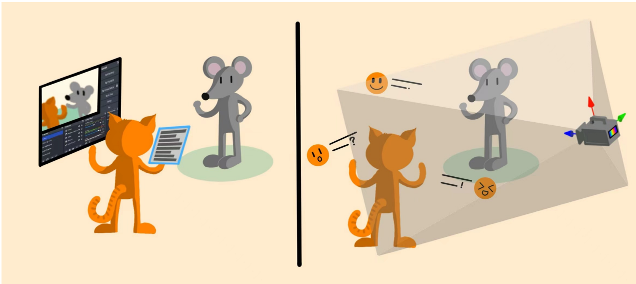

Figure 4: A contrast between the 2D/3D interface awareness for social VR streaming. Left: current 2D interfaces (stream manager and chat window) for audience engagement. Right: Social VR streamers' proposed 3D solutions to localize these 2D interfaces.

surroundings, more flexible solutions allowing smooth transitions between these interfaces should be introduced.

Indeed, recent research has begun to explore more fluid interface transitions between realities [58, 65], such as gradually replacing objects in physical surroundings with virtual environments to enhance users' spatial perception across realities [65]. Building upon these insights, we suggest that future cross-reality systems should implement more responsive interface management strategies that can immediately adjust to users' active attention shift across realities as well as environmental changes in respective realities. For example, future cross-reality systems could implement an attention-aware window system that utilizes sensing technologies (e.g., social VR streamers' eye-tracking) to detect focus patterns and automatically adjust the transparency setting of non-primary interfaces. This could be coupled with automatic spatial interface arrangement that maps different information categories to distinct spatial zones around the user, allowing for more intuitive navigation without manual window arrangement. Additionally, interface elements could be designed to automatically shift their digital representations based on the current ongoing tasks, such as transforming from detailed dashboards during active interactions with people in other realities to ambient indicators during immersive activities in the current reality, which can provide continuous awareness without visual obstruction.

5.2 Rethinking Interactions: Designing for Supporting Natural Cross-Reality Social Interactions Through Everyday Technologies

Another highlight from our findings is social VR streamers' struggles with facilitating natural cross-reality social interactions in a bidirectional and intuitive way. This demand also goes beyond just informing the existence of different realities (e.g., displaying objects and bystanders through visual portals [6, 18, 50, 69, 70]) as described in existing cross-reality object substitution and multi-user interface principles [6]. Indeed, in our study, real-world cross-reality practices, such as social VR streaming, primarily rely on everyday technologies, marking a significant departure from existing research prototypes that focus on specialized software and hardware. In this section, we thus highlight two important considerations for designing more natural cross-reality social interactions beyond lab settings: (1) translating 2D messages into immersive 3D space for intuitive cross-reality social awareness; and (2) supporting immersive video sharing in VR to engage with non-VR audiences.

5.2.1 Translating 2D Messages into Immersive 3D Space for Intuitive Cross-Reality Social Awareness. In our study, social VR streamers often struggle with maintaining social awareness of 2D audience text messages, although many of them have implemented various visibility solutions (e.g., HUD displays) that align with research on body-centric notification approaches for balancing information perception and immersion [53]. As described by our participants, these efforts still cannot compensate for the fundamental imbalance between natural, multi-modal VR interactions and limited single-modal text communications from non-VR audiences, making their presence less salient and potentially be overlooked.

In response, a small body of prior research has begun to explore ways to increase bystander presence through body video streaming [41, 66], and some social VR streamers have also begun to explore immersive notifications through audience redeems. However, how to better support detailed communication through 3D social cues remains unexplored. This reveals a critical design space for developing translation mechanisms that can incorporate basic text messaging within VR while enhancing the social awareness in 3D environments (see Figure 4).

To address this gap, we propose supporting **cross-reality message translation** to substitute incoming 2D texts from non-VR users into VR-specific social cues with tangible effects. For example, future cross-reality systems could implement semantic animation that analyzes message content to generate contextually relevant environmental effects, such as transforming exclamations of excitement into text blocks with particle bursts and converting congratulatory messages into virtual confetti that appears around the streamer. Furthermore, audience presence could be reinforced through ambient representation systems that aggregate viewer activity into dynamic environmental elements that scale with audience size and engagement level, such as a virtual sun that brightens with increased chat activity or spatial audio cues that convey audience reactions. These approaches would maintain the streamer's immersion in VR while providing intuitive awareness of audience engagement outside VR without requiring streamers' explicit attention shifts.

5.2.2 Supporting Immersive Video Sharing in VR to Engage with Non-VR Audiences. Existing design guidelines for cross-reality interactions also acknowledge the importance of shared understanding across realities by leveraging specialized displays (e.g., external screens for co-located viewing [22, 27] or smartphone AR synchronization of certain 3D objects [77]). However, these approaches tend to focus on automated target view synchronization rather than facilitating personal experience sharing that centers emotional delivery and careful viewpoint curation. In contrast, our findings reveal that cross-reality social interactions in social VR streaming tend to resemble everyday social sharing and may even require aesthetic framing and composition of VR views (e.g., streaming as a "cinematographer" in VR).

While this seemingly overlaps with VR filming for professional production [8, 25, 34], in our study, social VR streamers' practices also differ from planned, collaborative shots as their streaming requires more spontaneous and improvisational performances by themselves. This real-time and single-user nature thus requires streamers to constantly divide attention and attend to the virtual camera (e.g., through mirrored camera images), highlighting a critical need for more VR native camera support (e.g., enhancing camera orientation awareness by "flashlight-like" frustum indicator in VR, see Figure 4).

To explore this new design space, we suggest that future cross-reality systems could incorporate immersive camera systems that transcend basic viewpoint sharing by supporting intuitive cinematic creation within 3D spaces. Drawing from our participants' experiences, for example, future cross-reality system could implement camera frustum display methods mentioned by P15 to replace the camera viewpoint windows, such as subtle light cones that represent camera direction and zoom level without requiring attention to a mirrored display. This approach would maintain immersion while providing essential framing awareness. Additionally, intelligent camera assistance features could enable aesthetic content sharing through context-aware following modes and composition aids, such as highlighting potential subjects for following and apply different lighting to 3D objects according to their 2D frame positioning. These enhancements would empower streamers to function as spontaneous cinematographers for their viewers outside VR without undermining their immersive presence in VR.

5.3 Limitations and Future Work

Our study has several limitations. First, while our participants represent a diverse sample of social VR streamers across ethnicity, gender, and sexuality, their experiences primarily centered on VRChat and mainstream streaming platforms (e.g., Twitch, YouTube, and Bilibili in China). This platform-specific focus may constrain our understanding beyond these particular platforms, potentially missing unique challenges and opportunities in other social VR environments. Future work should explore how cross-reality social interactions manifest across a broader range of platforms to develop a more comprehensive understanding of this space.

Second, while our interviews with streamers provided rich insights into their streaming practices and perceived viewer impacts, this single-perspective approach may not fully capture the complete dynamics of cross-reality social interactions. Future research should incorporate multi-stakeholder perspectives, particularly through interviewing or surveying viewers to understand how they perceive and experience streamers' struggles and challenges in social VR streaming. Such research can help inform the design of future cross-reality systems with more in-depth empirical evidence of how both streamers and viewers experience, interpret, and engage with cross-reality social interactions in social VR live streaming.

6 Conclusion

Our investigation of social VR streaming reveals fundamental real-world challenges in cross-reality systems that span both technical and social dimensions. Through analyzing social VR streamers' unique cross-reality practices, we have identified key tensions in social VR streaming emerging in (1) *controlling external 2D software in a 3D VR environment*, (2) *arranging 2D interfaces in a 3D VR environment*, (3) *maintaining 3D social awareness of 2D audience messages*, and (4) *maintaining 3D spatial awareness about the 2D camera system*. These findings extend beyond conventional cross-reality experiments and prototypes to inform the broader design space of such systems, particularly in integrating cross-reality systems into existing 2D software ecosystem and further supporting more natural and intuitive cross-reality social interactions through everyday technologies. We believe that this work will help us understand and design more sophisticated yet accessible cross-reality systems to support people's productivity and various social needs in the future, which will ultimately make cross-reality experiences more inclusive for a broader population and more diverse daily contexts.

Acknowledgments

We thank our participants for sharing their experiences. We also thank the anonymous reviewers for helping us improve this paper. We especially thank Ruchi Panchanadikar and Amelia Hall for helping data collection and Kelsea Schulenberg and Lingyuan Li for providing useful feedback. This work was partially supported by the National Science Foundation under award 2112878 and 2342393.

References

[1] [n. d.]. Apple Vision Pro. https://www.apple.com/apple-vision-pro/ [Accessed 18-01-2025].

[2] [n. d.]. Meta Quest 3. https://www.meta.com/quest/quest-3/ [Accessed 18-01-2025].

[3] [n. d.]. Microsoft 365 productivity apps on Apple Vision Pro. https://support.microsoft.com/en-us/office/microsoft-365-productivity-apps-on-apple-vision-pro-ba077b93-74bf-4fca-a799-1ae801bce0a9. [Accessed 18-01-2025].

[4] Dakyeom Ahn, Seora Park, Seolhee Lee, Jieun Cho, and Hajin Lim. 2025. I Stan Alien Idols and Also the People Behind Them: Understanding How Seams Between Virtual and Real Identities Engage VTuber Fans – A Case Study of PLAVE. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3706598.3714218

[5] Bruno Araujo, Ricardo Jota, Varun Perumal, Jia Xian Yao, Karan Singh, and Daniel Wigdor. 2016. Snake Charmer: Physically Enabling Virtual Objects. In *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction* (Eindhoven, Netherlands) *(TEI '16)*. Association for Computing Machinery, New York, NY, USA, 218–226. doi:10.1145/2839462.2839484

[6] Jonas Auda, Uwe Gruenefeld, Sarah Faltaous, Sven Mayer, and Stefan Schneegass. 2023. A Scoping Survey on Cross-reality Systems. *ACM Comput. Surv.* 56, 4, Article 83 (Oct. 2023), 38 pages. doi:10.1145/3616536

[7] Verena Biener, Daniel Schneider, Travis Gesslein, Alexander Otte, Bastian Kuth, Per Ola Kristensson, Eyal Ofek, Michel Pahud, and Jens Grubert. 2020. Breaking the Screen: Interaction Across Touchscreen Boundaries in Virtual Reality for Mobile Knowledge Workers. *IEEE Transactions on Visualization and Computer Graphics* 26, 12 (2020), 3490–3502. doi:10.1109/TVCG.2020.3023567

[8] Aimone Bodini. 2023. *Filmmaking of the future: A co-design-led investigation into the potential of immersive technologies to enhance the small and medium audiovisual production processes.* Ph. D. Dissertation. Brunel University London.

[9] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[10] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis.* American Psychological Association.

[11] Lung-Pan Cheng, Li Chang, Sebastian Marwecki, and Patrick Baudisch. 2018. iTurk: Turning Passive Haptics into Active Haptics by Making Users Reconfigure Props in Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–10. doi:10.1145/3173574.3173663

[12] EunJeong Cheon and Shengyang Xu. 2024. *Creative Precarity in Motion: Revealing the Hidden Labor Behind Animating Virtual Characters.* Association for Computing Machinery, New York, NY, USA, 3471–3484. https://doi.org/10.1145/3643834.3661545

[13] Tobias Drey, Jan Gugenheimer, Julian Karlbauer, Maximilian Milo, and Enrico Rukzio. 2020. VRSketchIn: Exploring the Design Space of Pen and Tablet Interaction for 3D Sketching in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3313831.3376628

[14] Steven Feiner, Blair MacIntyre, Marcus Haupt, and Eliot Solomon. 1993. Windows on the world: 2D windows for 3D augmented reality. In *Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology* (Atlanta, Georgia, USA) *(UIST '93)*. Association for Computing Machinery, New York, NY, USA, 145–155. doi:10.1145/168642.168657

[15] Guo Freeman, Dane Acena, Nathan J McNeese, and Kelsea Schulenberg. 2022. Working Together Apart through Embodiment: Engaging in Everyday Collaborative Activities in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (2022), 1–25.

[16] Guo Freeman, Yang Hu, Ruchi Panchanadikar, Amelia L Hall, Kelsea Schulenberg, and Lingyuan Li. 2024. "My Audience Gets to Know Me on a More Realistic Level": Exploring Social VR Streamers' Unique Strategies to Engage with Their Audiences. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 6, 7 pages. doi:10.1145/3613905.3651036

[17] Guo Freeman and Divine Maloney. 2021. Body, avatar, and me: The presentation and perception of self in social virtual reality. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–27.

[18] Ceenu George, An Ngo Tien, and Heinrich Hussmann. 2020. Seamless, Bi-directional Transitions along the Reality-Virtuality Continuum: A Conceptualization and Prototype Exploration. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 412–424. doi:10.1109/ISMAR50242.2020.00067

[19] Jens Grubert, Lukas Witzani, Eyal Ofek, Michel Pahud, Matthias Kranz, and Per Ola Kristensson. 2018. Text Entry in Immersive Head-Mounted Display-Based Virtual Reality Using Standard Keyboards. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 159–166. doi:10.1109/VR.2018.8446059

[20] Uwe Gruenefeld, Jonas Auda, Florian Mathis, Stefan Schneegass, Mohamed Khamis, Jan Gugenheimer, and Sven Mayer. 2022. VRception: Rapid Prototyping of Cross-Reality Systems in Virtual Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 611, 15 pages. doi:10.1145/3491102.3501821

[21] Jan Gugenheimer, Evgeny Stemasov, Julian Frommel, and Enrico Rukzio. 2017. ShareVR: Enabling Co-Located Experiences for Virtual Reality between HMD and Non-HMD Users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 4021–4033. doi:10.1145/3025453.3025683

[22] Jan Gugenheimer, Evgeny Stemasov, Harpreet Sareen, and Enrico Rukzio. 2018. FaceDisplay: Towards Asymmetric Multi-User Interaction for Nomadic Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3173574.3173628

[23] William A Hamilton, Oliver Garretson, and Andruid Kerne. 2014. Streaming on twitch: fostering participatory communities of play within live mixed media. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 1315–1324.

[24] Zhenyi He, Christof Lutteroth, and Ken Perlin. 2022. TapGazer: Text Entry with Finger Tapping and Gaze-directed Word Selection. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 337, 16 pages. doi:10.1145/3491102.3501838

[25] Rorik Henrikson, Bruno Araujo, Fanny Chevalier, Karan Singh, and Ravin Balakrishnan. 2016. Multi-Device Storyboards for Cinematic Narratives in VR. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 787–796. doi:10.1145/2984511.2984539

[26] Yang Hu, Guo Freeman, and Ruchi Panchanadikar. 2025. "Grab the Chat and Stick It to My Wall": Understanding How Social VR Streamers Bridge Immersive VR Experiences with Streaming Audiences Outside VR. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 459, 19 pages. doi:10.1145/3706598.3713561

[27] Akira Ishii, Masaya Tsuruta, Ippei Suzuki, Shuta Nakamae, Tatsuya Minagawa, Junichi Suzuki, and Yoichi Ochiai. 2017. ReverseCAVE: providing reverse perspectives for sharing VR experience. In *ACM SIGGRAPH 2017 Posters* (Los Angeles, California) *(SIGGRAPH '17)*. Association for Computing Machinery, New York, NY, USA, Article 28, 2 pages. doi:10.1145/3102163.3102208

[28] Sungchul Jung, Pamela J. Wisniewski, and Charles E. Hughes. 2018. In Limbo: The Effect of Gradual Visual Transition Between Real and Virtual on Virtual Body Ownership Illusion and Presence. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 267–272. doi:10.1109/VR.2018.8447562

[29] Mohamed Kari, Tobias Grosse-Puppendahl, Luis Falconeri Coelho, Andreas Rene Fender, David Bethge, Reinhard Schütte, and Christian Holz. 2021. TransforMR: Pose-Aware Object Substitution for Composing Alternate Mixed Realities. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 69–79. doi:10.1109/ISMAR52148.2021.00021

[30] Daye Kim, Sebin Lee, Yoonseo Jun, Yujin Shin, and Jungjin Lee. 2025. VTuber's Atelier: The Design Space, Challenges, and Opportunities for VTubing. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–23. doi:10.1145/3706598.3714107

[31] Jarrod Knibbe, Jonas Schjerlund, Mathias Petraeus, and Kasper Hornbæk. 2018. The Dream is Collapsing: The Experience of Exiting VR. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3173574.3174057

[32] Pascal Knierim, Valentin Schwind, Anna Maria Feit, Florian Nieuwenhuizen, and Niels Henze. 2018. Physical Keyboards in Virtual Reality: Analysis of Typing Performance and Effects of Avatar Hands. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–9. doi:10.1145/3173574.3173919

[33] Chuan-en Lin, Ta Ying Cheng, and Xiaojuan Ma. 2020. ARchitect: Building Interactive Virtual Experiences from Physical Affordances by Bringing Human-in-the-Loop. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376614

[34] I-Sheng Lin, Tsai-Yen Li, Quentin Galvane, and Marc Christie. 2018. Design and evaluation of multiple role-playing in a virtual film set. In *Proceedings of the 16th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry* (Tokyo, Japan) *(VRCAI '18)*. Association for Computing Machinery, New York, NY, USA, Article 17, 4 pages. doi:10.1145/3284398.3284424

[35] Zhicong Lu, Chenxinran Shen, Jiannan Li, Hong Shen, and Daniel Wigdor. 2021. More Kawaii than a Real-Person Live Streamer: Understanding How the Otaku Community Engages with and Perceives Virtual YouTubers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 137, 14 pages. doi:10.1145/3411764.3445660

[36] Zhicong Lu, Haijun Xia, Seongkook Heo, and Daniel Wigdor. 2018. You watch, you give, and you engage: a study of live streaming practices in China. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

[37] Christian Mai, Lukas Rambold, and Mohamed Khamis. 2017. TransparentHMD: revealing the HMD user's face to bystanders. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia* (Stuttgart, Germany) *(MUM '17)*. Association for Computing Machinery, New York, NY, USA, 515–520. doi:10.1145/3152832.3157813

[38] Stefan Marks, David White, and Manpreet Singh. 2017. Getting up your nose: a virtual reality education tool for nasal cavity anatomy. In *SIGGRAPH Asia 2017 Symposium on Education* (Bangkok, Thailand) *(SA '17)*. Association for Computing Machinery, New York, NY, USA, Article 1, 7 pages. doi:10.1145/3134368.3139218

[39] Joseph A Maxwell. 2010. Using numbers in qualitative research. *Qualitative inquiry* 16, 6 (2010), 475–482.

[40] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–23.

[41] Mark McGill, Daniel Boland, Roderick Murray-Smith, and Stephen Brewster. 2015. A Dose of Reality: Overcoming Usability Challenges in VR Head-Mounted Displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2143–2152. doi:10.1145/2702123.2702382

[42] Joshua McVeigh-Schultz, Anya Kolesnichenko, and Katherine Isbister. 2019. Shaping Pro-Social Interaction in VR: An Emerging Design Framework. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[43] Sharan B Merriam and Elizabeth J Tisdell. 2015. *Qualitative research: A guide to design and implementation*. John Wiley & Sons.

[44] Paul Milgram and Fumio Kishino. 1994. A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information and Systems* 77 (1994), 1321–1329. https://api.semanticscholar.org/CorpusID:17783728

[45] Jens Müller, Johannes Zagermann, Jonathan Wieland, Ulrike Pfeil, and Harald Reiterer. 2019. A Qualitative Comparison Between Augmented and Virtual Reality Collaboration with Handheld Devices. In *Proceedings of Mensch Und Computer 2019* (Hamburg, Germany) *(MuC '19)*. Association for Computing Machinery, New York, NY, USA, 399–410. doi:10.1145/3340764.3340773

[46] Mitchell Norman, Gun A. Lee, Ross T. Smith, and Mark Billingurst. 2019. The Impact of Remote User's Role in a Mixed Reality Mixed Presence System. In *Proceedings of the 17th International Conference on Virtual-Reality Continuum and Its Applications in Industry* (Brisbane, QLD, Australia) *(VRCAI '19)*. Association for Computing Machinery, New York, NY, USA, Article 2, 9 pages. doi:10.1145/

3359997.3365691

[47] Ohan Oda, Carmine Elvezio, Mengu Sukan, Steven Feiner, and Barbara Tversky. 2015. Virtual Replicas for Remote Assistance in Virtual and Augmented Reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 405–415. doi:10.1145/2807442.2807497

[48] Thammathip Piumsomboon, Gun A. Lee, Jonathon D. Hart, Barrett Ens, Robert W. Lindeman, Bruce H. Thomas, and Mark Billinghurst. 2018. Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3173574.3173620

[49] Ronald Poelman, Oytun Akman, Stephan Lukosch, and Pieter Jonker. 2012. As if being there: mediated reality for crime scene investigation. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (Seattle, Washington, USA) *(CSCW '12)*. Association for Computing Machinery, New York, NY, USA, 1267–1276. doi:10.1145/2145204.2145394

[50] Fabian Pointecker, Judith Friedl, Daniel Schwajda, Hans-Christian Jetter, and Christoph Anthes. 2022. Bridging the Gap Across Realities: Visual Transitions Between Virtual and Augmented Reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 827–836. doi:10.1109/ISMAR55827.2022.00101

[51] Joan Sol Roo, Renaud Gervais, Jeremy Frey, and Martin Hachet. 2017. Inner Garden: Connecting Inner States to a Mixed Reality Sandbox for Mindfulness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1459–1470. doi:10.1145/3025453.3025743

[52] Joan Sol Roo and Martin Hachet. 2017. One Reality: Augmenting How the Physical World is Experienced by combining Multiple Mixed Reality Modalities. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, New York, NY, USA, 787–795. doi:10.1145/3126594.3126638

[53] Rufat Rzayev, Sven Mayer, Christian Krauter, and Niels Henze. 2019. Notification in VR: The Effect of Notification Placement, Task and Environment. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Barcelona, Spain) *(CHI PLAY '19)*. Association for Computing Machinery, New York, NY, USA, 199–211. doi:10.1145/3311350.3347190

[54] Leonard Schmidt and Enes Yigitbas. 2024. Development and Usability Evaluation of Transitional Cross-Reality Interfaces. *Proc. ACM Hum.-Comput. Interact.* 8, EICS, Article 263 (June 2024), 32 pages. doi:10.1145/3664637

[55] Daniel Schneider, Alexander Otte, Travis Gesslein, Philipp Gagel, Bastian Kuth, Mohamad Shahm Damlakhi, Oliver Dietz, Eyal Ofek, Michel Pahud, Per Ola Kristensson, Jörg Müller, and Jens Grubert. 2019. ReconViguRation: Reconfiguring Physical Keyboards in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 25, 11 (2019), 3190–3201. doi:10.1109/TVCG.2019.2932239

[56] Jan-Henrik Schröder, Daniel Schacht, Niklas Peper, Anita Marie Hamurculu, and Hans-Christian Jetter. 2023. Collaborating Across Realities: Analytical Lenses for Understanding Dyadic Collaboration in Transitional Interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 97, 16 pages. doi:10.1145/3544548.3580879

[57] Patrick Schuck, Maximilian Altmeyer, Antonio Krüger, and Pascal Lessel. 2022. Viewer types in game live streams: questionnaire development and validation. *User Modeling and User-Adapted Interaction* 32, 3 (June 2022), 417–467. doi:10.1007/s11257-022-09328-9

[58] Adalberto L. Simeone, Mohamed Khamis, Augusto Esteves, Florian Daiber, Matjaž Kljun, Klen Čopič Pucihar, Poika Isokoski, and Jan Gugenheimer. 2020. International Workshop on Cross-Reality (XR) Interaction. In *Companion Proceedings of the 2020 Conference on Interactive Surfaces and Spaces* (Virtual Event, Portugal) *(ISS Companion '20)*. Association for Computing Machinery, New York, NY, USA, 111–114. doi:10.1145/3380867.3424551

[59] Yuta Sugiura, Hikaru Ibayashi, Toby Chong, Daisuke Sakamoto, Natsuki Miyata, Mitsunori Tada, Takashi Okuma, Takeshi Kurata, Takashi Shinmura, Masaaki Mochimaru, and Takeo Igarashi. 2018. An asymmetric collaborative system for architectural-scale space design. In *Proceedings of the 16th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry* (Tokyo, Japan) *(VRCAI '18)*. Association for Computing Machinery, New York, NY, USA, Article 21, 6 pages. doi:10.1145/3284398.3284416

[60] Hemant Bhaskar Surale, Aakar Gupta, Mark Hancock, and Daniel Vogel. 2019. TabletInVR: Exploring the Design Space for Using a Multi-Touch Tablet in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300243

[61] Philipp Sykownik, Linda Graf, Christoph Zils, and Maic Masuch. 2021. The Most Social Platform Ever? A Survey about Activities amp; Motives of Social VR Users. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. 546–554. doi:10.1109/VR50410.2021.00079

[62] Balasaravanan Thoravi Kumaravel, Fraser Anderson, George Fitzmaurice, Bjoern Hartmann, and Tovi Grossman. 2019. Loki: Facilitating Remote Instruction of Physical Tasks Using Bi-Directional Mixed-Reality Telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 161–174. doi:10.1145/3332165.3347872

[63] Kien T.P. Tran, Sungchul Jung, Simon Hoermann, and Robert W. Lindeman. 2019. MDI: A Multi-channel Dynamic Immersion Headset for Seamless Switching between Virtual and Real World Activities. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 350–358. doi:10.1109/VR.2019.8798240

[64] Arda Ege Unlu and Robert Xiao. 2021. PAIR: Phone as an Augmented Immersive Reality Controller. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology* (Osaka, Japan) *(VRST '21)*. Association for Computing Machinery, New York, NY, USA, Article 27, 6 pages. doi:10.1145/3489849.3489878

[65] Dimitar Valkov and Steffen Flagge. 2017. Smooth immersion: the benefits of making the transition to virtual environments a continuous process. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 12–19. doi:10.1145/3131277.3132183

[66] Julius von Willich, Markus Funk, Florian Müller, Karola Marky, Jan Riemann, and Max Mühlhäuser. 2019. You Invaded my Tracking Space! Using Augmented Virtuality for Spotting Passersby in Room-Scale Virtual Reality. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 487–496. doi:10.1145/3322276.3322334

[67] VRChat. 2024. Camera — wiki.vrchat.com. https://wiki.vrchat.com/wiki/Camera. [Accessed 05-01-2025].

[68] Qian Wan and Zhicong Lu. 2024. Investigating VTubing as a Reconstruction of Streamer Self-Presentation: Identity, Performance, and Gender. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 80 (April 2024), 22 pages. doi:10.1145/3637357

[69] Chiu-Hsuan Wang, Bing-Yu Chen, and Liwei Chan. 2022. RealityLens: A User Interface for Blending Customized Physical World View into Virtual Reality. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 49, 11 pages. doi:10.1145/3526113.3545686

[70] Chiu-Hsuan Wang, Chia-En Tsai, Seraphina Yong, and Liwei Chan. 2020. Slice of Light: Transparent and Integrative Transition Among Realities in a Multi-HMD-User Environment. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 805–817. doi:10.1145/3379337.3415868

[71] Chiu-Hsuan Wang, Seraphina Yong, Hsin-Yu Chen, Yuan-Syun Ye, and Liwei Chan. 2020. HMD Light: Sharing In-VR Experience via Head-Mounted Projector for Asymmetric Interaction. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 472–486. doi:10.1145/3379337.3415847

[72] Yuntao Wang, Zichao (Tyson) Chen, Hanchuan Li, Zhengyi Cao, Huiyi Luo, Tengxiang Zhang, Ke Ou, John Raiti, Chun Yu, Shwetak Patel, and Yuanchun Shi. 2020. MoveVR: Enabling Multiform Force Feedback in Virtual Reality using Household Cleaning Robot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3313831.3376286

[73] Amy K Way, Robin Kanak Zwier, and Sarah J Tracy. 2015. Dialogic interviewing and flickers of transformation: An examination and delineation of interactional strategies that promote participant self-reflexivity. *Qualitative Inquiry* 21, 8 (2015), 720–731.

[74] Donghee Yvette Wohn and Guo Freeman. 2020. Audience management practices of live streamers on Twitch. In *ACM International Conference on Interactive Media Experiences*. 106–116.

[75] Donghee Yvette Wohn, Guo Freeman, and Caitlin McLaughlin. 2018. Explaining viewers' emotional, instrumental, and financial support provision for live streamers. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. ACM, 474.

[76] Liwei Wu, Qing Liu, Jian Zhao, and Edward Lank. 2023. Interactions across Displays and Space: A Study of Virtual Reality Streaming Practices on Twitch. *Proceedings of the ACM on Human-Computer Interaction* 7, ISS (Oct. 2023), 242–265. doi:10.1145/3626473

[77] Shihui Xu, Bo Yang, Boyang Liu, Kelvin Cheng, Soh Masuko, and Jiro Tanaka. 2019. Sharing Augmented Reality Experience Between HMD and Non-HMD User. In *Human Interface and the Management of Information. Information in Intelligent Systems*, Sakae Yamamoto and Hirohiko Mori (Eds.). Springer International Publishing, Cham, 187–202.

[78] Xuanhui Xu, Xingyu Pan, and Abraham G. Campbell. 2020. ARLS: An asymmetrical remote learning system for sharing anatomy between an HMD and a light field display. In *Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology* (Virtual Event, Canada) *(VRST '20)*. Association for Computing Machinery, New York, NY, USA, Article 43, 2 pages. doi:10.1145/3385956.3422095

[79] Qian Zhou, George Fitzmaurice, and Fraser Anderson. 2022. In-Depth Mouse: Integrating Desktop Mouse into Virtual Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 354, 17 pages. doi:10.1145/3491102.3501884

[80] Zhenhui Zhu, Zhi Yang, and Yafei Dai. 2017. Understanding the gift-sending interaction on live-streaming video websites. In *International Conference on social computing and social media*. Springer, 274–285.