Stories for I, They and You: Exploring In-situ Virtual Camera Setups for Live Streaming in Virtual Reality

Liwei Wu
School of Computer Science
University of Waterloo
Waterloo, Ontario, Canada
a92wu@uwaterloo.ca

Zachary McKendrick
School of Computer Science
University of Waterloo
Waterloo, Ontario, Canada
zachary.mckendrick@uwaterloo.ca

Jian Zhao
School of Computer Science
University of Waterloo
Waterloo, Ontario, Canada
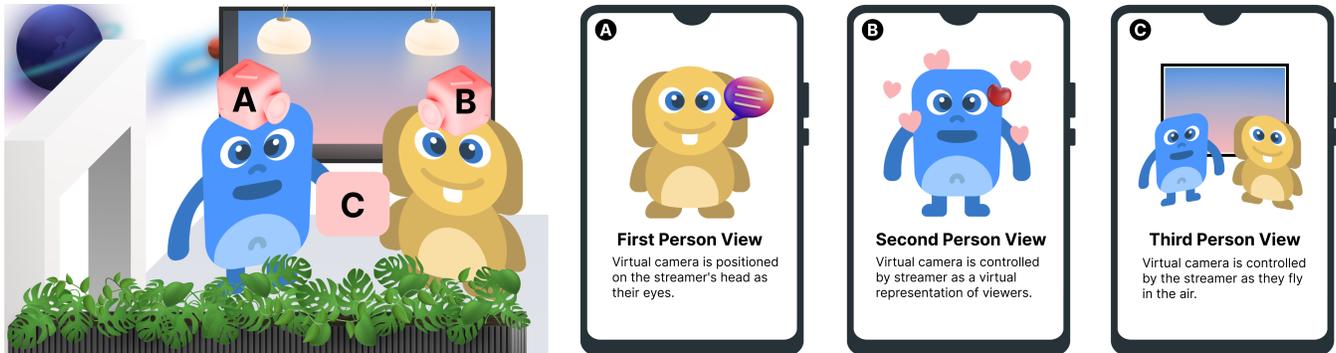jianzhao@uwaterloo.ca

Figure 1: Three fundamental camera views that are widely used in VR streaming and offer distinct ways to engage viewers. First person view (A) immerses users in the streamer's perspective. Second person view (B) fosters a sense of direct interaction and intimacy, which is also noted in prior research [33]. Third person view (C) provides a broader depiction of the streamer's actions and interactions within the virtual environment. Despite their prevalence, there is limited research on the specific patterns and strategies used in each of these perspectives, which motivates this work.

Abstract

Virtual reality live streaming (VR streaming) is becoming increasingly popular as an entertainment medium, but its unique aspects of virtual camera usage remain underexplored. This paper presents a two-phase study investigating how VR streamers currently use virtual cameras to create engaging content for viewers on 2D displays. We first analyzed 8 popular Twitch streamers' videos totaling 2625 minutes to identify common virtual camera usage. Then, we interviewed 10 media experts all with about 10+ years of professional experience to derive design considerations for VR streamers and design implications for future developers. We also proposed the Immersion Triangle, a conceptual framework to analyze and explore the concept of immersion within VR streaming context. Our findings highlight VR streaming as a novel mass media format that can offer new perspectives on both VR and live streaming. This study also suggests opportunities for future research to enhance interactions between streamers and their viewers.

CCS Concepts

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

Keywords

live streaming; virtual reality; cinematography.

1 Introduction

Live streaming has gained significant global popularity as an entertainment medium. Twitch, one of the largest online streaming platforms, has reported over 2.5 million concurrent viewers with a total of more than 1.3 trillion minutes watched in 2023 [1]. Currently, most live streaming occurs in a 2D-2D form, where both streamers and viewers utilize 2D displays. However, virtual reality (VR) streaming, in which streamers use VR headsets and broadcast their immersive experiences to viewers using 2D displays, is gaining popularity in entertainment and research communities [32, 33, 43, 68]. Furthermore, the growing availability of affordable consumer-grade devices, such as the Quest 3 and Pico 4, suggests that interest in VR streaming is expected to persist [2] [3], especially as companies like Meta, Google, and Apple continue to invest and innovate in VR.

---

[1] https://www.twitch.tv/p/press-center
[2] https://mixed-news.com/en/virtual-reality-2025-8-predictions
[3] https://www.researchnester.com/reports/virtual-reality-market/160

Previous studies have highlighted the importance of establishing connections with viewers for the success of live streaming [50, 66, 67]. For VR streamers, virtual cameras, the digital tools that simulate real-world camera functions, allow them to construct live stream narratives to share with their viewers. As such, virtual cameras play a pivotal role in fostering these connections. Compared to cinema and animation, where the principles of manipulating cameras have been well studied, the usage patterns and strategies of virtual cameras in VR streaming practices have received limited attention. Previous studies have observed differences in the viewing experience of VR content when presented on 2D displays through virtual cameras using first person and third person views [17]. Another study has highlighted the creative application of virtual cameras in VR streaming, where the camera is treated as a representation of the viewer's perspective [4] [33]. Despite these insights, there is still limited analysis on how VR streamers commonly use their virtual cameras across these fundamental camera perspectives (Figure 1) in real-life scenarios in detail. We suggest that, despite the similarities to traditional media practices in 2D streaming, in VR streaming, virtual cameras act as active facilitators of the streamer-viewer relationship, extending the streamer to construct both narratives and viewer connections. This aspect has been largely overlooked and is a key motivation behind our work.

In this paper, we present a two-phase study (Figure 2) that investigates two key questions regarding the use of virtual cameras in VR streaming: **RQ1)** how do VR streamers currently utilize virtual cameras during their streams; and **RQ2)** how can traditional media theory and practice better support planning and designing their virtual camera setups? To address RQ1, informed by prior studies examining streaming practices through user-created videos [13, 68], we conducted a thematic analysis [9] of streaming videos from eight popular VR streamers, totalling 2,625 minutes, yielding common camera usage scenarios. Next, we interviewed ten domain experts, each with at least 10 years of professional experience in media-related fields, to gather their critiques and suggestions regarding the identified camera usage scenarios. We synthesized our results from both phases into a set of design considerations to help VR streamers plan their camera setups more intentionally, as well as a set of design implications for future VR streaming tool developers. Based on these findings, we proposed the *Immersion Triangle*, a conceptual framework designed to help analyze and explore the design of VR streaming experiences and to promote thinking about immersion in the context of VR streaming for both streamers and viewers.

The contributions of this paper are three-fold:

- **Empirical knowledge** about common virtual camera usage scenarios in VR streaming;
- A set of **design considerations and implications** derived from our video analysis and expert interviews to guide VR streamers and VR streaming tool developers;
- A **conceptual framework**, Immersion Triangle, to support exploration and understanding of immersion within the context of VR streaming.

---

[4]In this paper, we refer to this special use of the camera as second-person view.

2 Related Work

In this section, we briefly review prior research related to our study in two key areas: 1) sharing VR experience through live streaming, and 2) camera setups for filming and performance.

2.1 Sharing VR Experience through Live Streaming

A combination of accessibility through platforms such as Meerkat, Periscope, Twitch, YouTube Live, and Snapchat Live Stories [26, 55, 58], and variety of topics including gaming [27, 39], education [18, 25], creative arts [19], and even relaxation [1] has contributed to the poplarity of streaming as a medium. Live streaming is unique from other more traditional popular media due to its ability to offer immersion, immediacy, interaction, and social connection between the streamer and audience [26]. Despite being in its early stages, studies have begun exploring sharing VR experiences through live streaming as an online media broadcasting activity. Compared to traditional 2D live streaming, VR streaming presents unique technical and viewing challenges, calling for further exploration to enable more fluent and engaging interactions for streamers and viewers [32, 33, 68].

A common use for streaming techniques in VR is facilitating cross-reality collaboration, where users outside a VR environment can interact with those wearing a headset, enabling a more inclusive, customizable participatory experience [6, 24, 61, 64]. To further enhance the viewer experience, researchers have explored augmenting 2D streaming content from VR streamers to enable viewers to experience it in 3D or even in VR, bridging the gap between traditional 2D streaming and immersive experiences [29, 62]. However, these studies are often limited to one-to-one or one-to-few scenarios, and their direct application to live streaming as a mass medium for a large audience can be challenging. For instance, VR streamers often use virtual avatars, which have been found to play a crucial role in self-identity expression in past research [20–22, 41], particularly in popular social VR games and stream environments/social platforms like VRChat. Still, several domains have benefited from leveraging the combination of VR and live streaming to facilitate and enhance processes and experiences across diverse scenarios, such as education [49, 59, 60] and simulation [57, 65, 69].

Virtual cameras play an important role in shaping the stream-viewing experience, particularly how different viewing perspectives affect users' comprehension of VR content when viewed on traditional 2D screens [17]. Furthermore, effectively and creatively self-managing virtual cameras during live streams poses a significant challenge for the streamer due to the time-consuming and demanding nature of camera management, as well as the risk that improper use of creative camera techniques may negatively impact viewers' experience [32, 33, 68], which echoes early studies on the challenges caused by the inherent asymmetry of control in physical handheld devices [35]. Despite growing recognition of virtual cameras' importance in the streamer-viewer relationship, there remains a lack of practical guidance for VR streamers on how to use them more effectively. Our study aims to address that gap by systematically exploring the role of virtual cameras in VR streaming and providing actionable insights for streamers to enhance their content creation and viewer engagement.
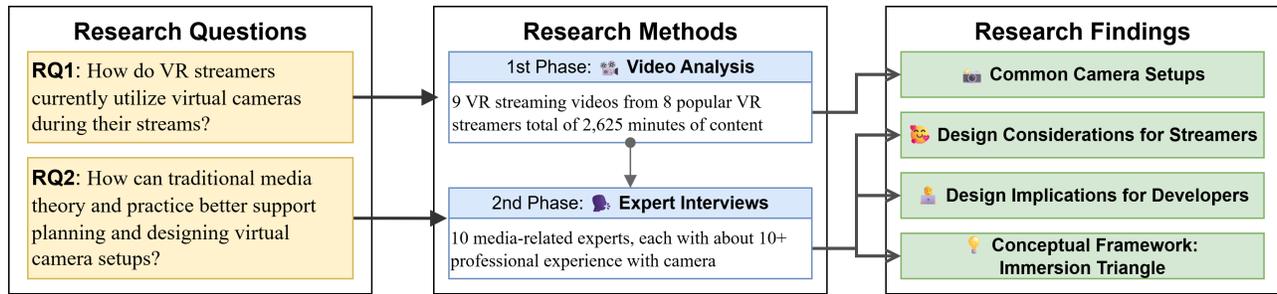
Figure 2: The research questions, methods and findings of this study: 1) we started with a video analysis on 9 streaming videos (2,625 minutes in total) from 8 popular VR streamers to investigate common virtual camera usage in VR streaming; 2) based on the video analysis findings, we conducted interviews with 10 experts in media related domains to collect their recommendations for VR streamers and insights for VR streaming.

From an ethnographic perspective, many streamers record and upload their streaming sessions to platforms like YouTube and Twitch. These user-generated online videos are receiving increasing attention from the research community as an important source of data for observing and analyzing users' behaviours [2, 31, 38], and are increasingly used to understand VR experience and live streaming practices [12, 13, 68]. Thus, drawing upon prior video analysis studies of live streaming practices, our research adopts a similar approach to investigate the common virtual cameras usage in VR streaming. This approach enables us to objectively investigate how virtual cameras are used in real-world streaming practices and explore their potential to broaden our understanding of VR streaming as an emerging mass media format.

2.2 Overview of Camera Setups for Filming and Performance

Camera setups support visual storytelling by strategically organizing shot composition, framing, and camera movement in relation to a subject (e.g., an actor or an object) to convey emotion, narrative depth, and viewer engagement [3, 7]; this concept applies to both physical and virtual cameras. While newer technologies enable greater creative flexibility and fidelity, traditional cinematographic principles such as the rule of thirds, symmetry, and depth of field remain fundamental. The use of leading lines and asymmetrical framing in contemporary cinematography has been studied concerning how visual attention is guided in both linear and interactive storytelling environments [56].

The placement of a camera within a scene influences how audiences engage with the narrative and perceive spatial relationships, thereby shaping their understanding of character dynamics, spatial continuity, and the flow of the story. Cinema scholars and filmmakers have identified several key camera setups that are widely used in both traditional and digital cinematography [7]. One of the most common configurations is the static camera setup, where the camera remains fixed in a specific position, allowing for deliberate control over framing and mise-en-scène (composition) [8]. This approach is frequently used in classical Hollywood cinema and dialogue-heavy scenes to ensure spatial continuity and maintain a stable viewing experience [53]. Another widely used technique is the over-the-shoulder setup, which is particularly effective in

dialogue sequences because it maintains the spatial relationship between characters while directing the audience's attention [36]. In live television, event broadcasting, and real-time production, cinematographers frequently employ multi-camera setups to capture multiple angles simultaneously. This configuration allows for seamless transitions between perspectives, enhancing both efficiency and continuity in storytelling [48]. In addition to static camera setups, cinematographers have developed a range of dynamic camera movement techniques to create a more engaging visual experience. The introduction of Steadicam technology, as popularized by films like Rocky (1976) and The Shining (1980), allowed for smoother tracking shots without the instability of handheld footage [45]. The evolution of gimbal stabilization systems, such as the DJI Ronin [5], has further expanded the capabilities of handheld cinematography, enabling fluid motion with greater flexibility [56]. More recently, drone technology has revolutionized camera motion in filmmaking, offering a cost-effective and accessible way to achieve aerial shots typically done with cranes or helicopters [51].

To achieve effective camera setups for filming and performance, users need intuitive and precise control over their cameras. For virtual cameras, early studies [23, 28, 44, 63] explored direct control techniques, laying the groundwork for future methods. However, direcly controlling a basic virtual camera with six degrees of freedom can still be challenging, often distracting users from their main tasks. As a result, automated camera control became an important research direction [10]. Various studies have investigated ways to interpret high-level camera tasks [14–16, 71] and cinematography idioms [11, 30] to automate virtual cameras. Recent advances in virtual production [6] have led to broader industry use [37], with research combining physical and virtual elements through techniques like example-driven camera control [34, 70], hand-held virtual cameras [40, 54], and automated drone cameras [4, 52]. In VR games like VRChat, virtual cameras offer features similar to professional DSLRs, including aperture control and smooth motion comparable to physical stabilization systems or drones. These capabilities can support a range of camera setups but are underutilized in current VR streaming practices. There is still limited understanding of how virtual cameras are actually used by VR streamers in real-world

---

[5]https://www.dji.com/ca/ronin-4d
[6]https://www.unrealengine.com/en-US/explainers/virtual-production

streaming situations, and many streamers still find it challenging to apply these tools creatively [32, 33]. This motivates our study, which aims to address this gap by investigating common virtual camera setups used in VR streaming and leveraging camera principles in established media to enhance the creative use of virtual cameras.

3 VR Streaming Video Analysis

In the first phase of our study (Figure 2), we used a video analysis method to address RQ1 and examine common virtual camera usage in VR streaming. We followed existing methodologies that leverage user-generated video data to analyze VR and live streaming practices [12, 13, 68], which proved effective in objectively observing and understanding user behaviours in real-world settings. Specifically, we collected and analyzed a corpus of Twitch VR streaming videos. We chose to focus our data collection on VRChat, the most popular social VR game. Compared to other immersive VR platforms, VRChat offers feature-rich virtual cameras and usage scenarios, which are ideal for observing how virtual cameras are used in real-life VR streaming practices. This research step served as an ethnographic exploration into the world of VR streaming to: 1) establish a shared understanding of the topic among our team members, and 2) identify common virtual camera usage patterns through objective observation of VR streaming practices in real-world settings.

3.1 Data Collection and Analysis

To familiarize ourselves with the subject, we started by exploring and watching multiple VR streaming videos on Twitch. In addition, we also watched several live session on other popular VR games in different game genres such as *Beat Saber* (music beat), *Half-life: Alyx* (third-person shooter), *I Expect You to Die* (puzzle) and *Project Cars 2* (racing), and further confirmed our initial observations that VRChat has more scenarios for virtual camera usages than other games. In most games, streamers focus primarily on gameplay, with little attention paid to virtual camera manipulation, as cameras are typically static or minimally controlled. In contrast, streamers in VRChat have greater freedom and actively manipulate their virtual cameras for a variety of expressive purposes.

To gather a diverse and representative dataset of streaming videos, we first filtered streams to English language and ordered them by view count on Twitch. We excluded streams with graphic, physical, or thematic content or streamers who only talk in static positions. We then selected the top 8 streamers (S1-8) from the list and collected their videos, approximately 300 minutes each, to capture the varied styles of VR streaming. In total, our dataset consists of 2,625 minutes of real-life virtual camera footage (Table 1). Motivated by previous studies [13, 68], we adopted a similar thematic analysis method to analyze the collected videos. All members of our research team have extensive experience in VR research, including one member with over two decades of experience in the entertainment industry. Initially, the primary researcher watched all the videos twice to identify key codes with timestamps for moments when VR streamers interact with virtual cameras, such as moving or rotating cameras or switching views. From these initial observations, coding dimensions were developed focusing on common

Table 1: Nine VR streaming videos were collected from eight popular VR streamers, comprising a total of 2,625 minutes of video content, with each streamer contributing approximately 300 minutes on average.

Streamer	Followers	Twitch Video ID	Length	Views
S1	304k	2311801147	183 minutes	12k
S1	304k	2314158681	168 minutes	7.8k
S2	107k	2313981513	362 minutes	5k
S3	75.6k	2315373622	261 minutes	4.7k
S4	16.2k	2312060097	323 minutes	3.9k
S5	22.5k	2313587592	264 minutes	3.7k
S6	25.5k	2316827631	307 minutes	3.7k
S7	49.8k	2314666048	436 minutes	4.8k
S8	12.5k	2316179028	322 minutes	3.3k

camera uses such as self-recording, object capture, and interaction with other players. The research team then reviewed the initial results, refined the coding dimensions based on established filming practices and terminology, and discussed initial assumptions about the potential use patterns of virtual cameras. This led to a focus on essential camera aspects, such as angles and framing, and on how they compared with established filmmaking principles. Using these revised dimensions, the primary researcher revisited the videos to uncover additional codes and organize camera usage into broader themes categorized by three primary camera views. The team continued to refine these themes through ongoing discussions until a consensus on final themes was reached.

3.2 Findings: Camera Usage in VR Streaming

In this section, we organize our findings by camera aspects, highlighting the common use of virtual cameras in VR streaming within the broader context of established camera techniques from traditional media fields, summarized in Table 2. We also briefly discuss the dynamics of streamers' motivations for using various camera techniques and explore the reasons they switch cameras. This analysis is based on observations of their behaviour, chat messages, and their explanations when they switch cameras. Instead of focusing on numerical measures such as time or frequency, our findings emphasize the observed patterns in camera usage to illustrate their roles across different aspects. This follows Maxwell's advocation that reporting qualitative data numerically can mislead readers by suggesting a generalizability that is not supported by the specific contexts in which the data were analyzed [46]. This is also aligned with the reporting style in previous studies on the same topic [32, 33].

3.2.1 🙉 *Camera Views.* The most common camera views across all observed streams are *first-person*, *second-person*, and *third-person* camera views. *A first-person* camera view (Figure 3, A) immerses the audience in the streamer's perspective, allowing viewers to experience the content as if they were present in the scene. Although not commonly used in traditional filming, *second-person* camera views (Figure 3, B) create a unique sense of intimacy by making the viewers feel that the streamers are directly addressing them, effectively breaking the fourth wall and enhancing personal engagement. A *third-person* camera view (Figure 3, C) offers a broader

Table 2: Summary of video analysis findings on common virtual camera usage in VR streaming.

Camera Aspects	Findings
Camera View	- First person view (FPV) immerses the audience in the streamer's perspective - Second person view (SPV) fosters intimacy by making viewers feel as though they are being directly addressed - Third person view (TPV) provides a wider perspective of both the streamer and the virtual environment
Camera Angle	- Traditionally indicates power dynamics relative to the subject or environment - In VR streaming frequently used to generate unique and engaging perspectives for viewers
Camera Frames	- Close-up shot focuses on the facial expressions and emotion - Medium close-up shot covers the area above the shoulder, focusing on upper-body gestures and expressions - Medium-long shots capture the area over the knee, focusing on surrounding environment and streamer's upper body movements - Full-body shot captures the entire body from head to toe to show a complete view of the streamer within their environment
Camera Setups	- Monologue: where the streamer speaks alone and chats with their viewers - Dialogue: shows conversations and interactions between the streamer and another participant for viewers - Multilogue: encompasses interactions and conversations involving the streamer and multiple streaming participants

Figure 3: Examples of three camera views for VR streaming: A) first person view (FPV), allowing viewers to experience the scene as if they were seeing it through their own eyes; B) second person view (SPV), engaging the viewers more intimately while breaking the fourth wall; and C) third person view (TPV), giving viewers a broader context and an overview of the settings and environments.

Figure 4: Examples of three camera heights for VR streaming: A) Eye level view; B) Low angle view; C) High angle view.

visual context, providing viewers with an overview of the setting and interactions within the virtual environment.

Streamers typically choose between first-person view (FPV) and third-person view (TPV) based on their streaming content and design preferences. For streamers primarily using FPV, it is common to include a smaller picture-in-picture TPV of their avatar to help viewers better understand their full-body actions. We observed that TPV streamers occasionally switch between FPV and TPV during their streams to better suit different situations. They used FPV to offer viewers a more engaging and immersive experience when interacting with other players or playing games, while reserving TPV for conversations with their viewers. For example, S2 began with a TPV to present themselves while talking, but switched to an FPV while playing volleyball to give the viewers a sense of participation in the game.

Second-person view (SPV) creates a unique dynamic of direct engagement, positioning the viewer as inside the narrative while being adjacent to the protagonist. In this way, the camera is no longer merely a framing tool but serves as an interactive medium between the actors and their viewers. Echoing the previous study [33], we also observed streamers' creative use of cameras as a SPV. Notably, all TPV streamers in our study have moments of temporarily switching to SPV during their streams to emphasize certain words

or create a sense of intimacy. For example, S2 occasionally kissed the camera as if to kiss the faces of individual viewers, sometimes in response to messages in the chat expressing affection.

3.2.2 Camera Angle. Camera angle refers to the camera's vertical position relative to the subject or environment and often indicates power dynamics between characters or objects. For instance, a higher camera angle, looking down on a subject, can convey a sense of dominance or superiority, while a lower angle, looking up, can suggest vulnerability or submissiveness. In VR streaming, however, these traditional rules do not typically apply to represent power dynamics. Instead, camera heights are frequently used to generate unique and engaging perspectives for viewers that would not be easily observable in real life. The most common camera height for VR streaming is eye level (Figure 4, A), as it provides a natural, immersive perspective that closely aligns with human visual experience, making interactions feel more genuine and relatable to viewers. Eye-level perspective is predominantly used with TPV. For many streamers using FPV, their camera height instead reflects the eye-line of their virtual avatars, which can be relatively short (Figure 4, B) or tall (Figure 4, C) compared to other players.

Streamers often adjust their camera angles in response to changes in their avatars. For FPV streamers, these adjustments are usually intentional, as they experience the change firsthand and use it to offer viewers a unique visual experience. For instance, S1 used a dwarf avatar to create an upward angle, then switched to a giant avatar for a downward view, creating an interesting visual contrast for viewers. In contrast, TPV streamers tend to choose avatars of similar size when switching, thereby minimizing the need for major camera adjustments. For example, when S5 changed to a slightly shorter avatar, they lowered the camera slightly to maintain eye-level interaction with viewers.

3.2.3 Camera Frames. For our purposes, a camera frame refers to the compositional boundary of an image, or scene, i.e., the field of view, used to guide the viewer's eye, focus, and attention.

Here, we focus primarily on TPV scenarios in VR streaming, as in a FPV, the camera frame is directly tied to the streamer's own line of sight and head movement; therefore, the viewer sees what the streamer sees.

We found four commonly used camera frames: *close-up, medium close-up, medium-long, and full-body*. A *close-up shot* (Figure 6, A) focuses on the face, offering detailed facial expressions and emotions. This frame is particularly effective in establishing a strong emotional connection with the viewer, making it ideal for showing intimate or dramatic moments. A *medium close-up shot* (Figure 6, B) covers the area above the shoulder, capturing more of the body than a close-up while still maintaining focus on upper-body gestures and expressions. This type of frame balances facial expression detail with broader body language, making it more suitable for long-term conversations with viewers. *Medium-long shots*, or 'cowboy shots' (Figure 6, C) capture the area over the knee, allowing viewers to see more of the environment around the streamers while still maintaining focus on their upper body movements. This frame is often used to provide a sense of space and movement, making it ideal for scenes where physical actions are important but need to be seen in context with surroundings. Finally, a *full-body shot* (Figure 6, D) includes the entire body from head to toe, giving viewers a complete view of the avatar/streamer's posture and positioning within their environment. This type of frame is typically utilized when depicting larger movements or interactions that require understanding the whole spatial arrangement, such as showing off the virtual avatars and the virtual environment.

In VR streaming, medium shots are frequently used as the default camera frame with auto-focus and auto-follow features for conversations due to their ability to capture both facial expressions and environmental context. Streamers may also switch to other camera frames when specific situations require it, though these changes are usually brief, sometimes even lasting only a few seconds. They may move the camera closer for a close-up shot to emphasize facial expressions or object details, or pull back for medium-long or full-body shots to showcase their environment, avatar, or outfit. For instance, when S6 received chat comments about their avatar's tail, they temporarily switched from a medium to a medium-long shot to better display the tail at the middle of their avatar, and then moved to a full-body shot to showcase their new costume.

3.2.4 🖥️ *Camera Setups.* Camera setups refer to the arrangement and configuration of virtual cameras used in VR streaming, which is composed of the placement, orientation, and movement of these cameras within a three-dimensional space. As a combination of previously discussed camera concepts, camera setups integrate relationships between camera views, heights, and framing techniques to contribute to the narrative structure and emotional engagement for viewers watching on 2D displays. The streaming content we observed focused on streamer-streamer or streamer-viewer conversational scenes with static camera setups.

Much like a film or TV episode, a single VR streaming session can consist of various camera setups to capture the primary action of a scene. We categorize these scenarios into three types: *monologue*, *dialogue*, and *multilogue*. Monologue refers to situations where the streamer speaks alone and chat with their viewers without

direct interaction with other streaming participants or objects. Dialogue involves conversations between the streamer and another participant or object, showcasing their conversations or interactions to viewers. Multilogue encompasses interactions involving the streamer and multiple streaming participants simultaneously, typically through group conversations or dialogues between the streamer and the group. Figure 5 summarizes the most common camera setups we observed in this study for both FPV and TPV, categorized by the three common streaming scenarios. For each setup, there are three key elements: camera, streamer, and other streaming participants. Each element features a specific layout in the plan and the directions it faces.

Streamers often adjust their camera setups in response to changes in the number of people present in their view during interactions. For FPV streamers, who are confined to a fixed camera frame, they often rely on two main strategies: moving back to show the full group or rotating their head to sweep across individuals. For example, when initiating a conversation with a group, S1 would step back to include everyone and then moved closer to focus on one person. As more participants joined, S1 would look around to ensure all active participants were visible to viewers.

TPV streamers, on the other hand, frequently reposition their camera to accommodate new participants, ensuring both the streamer and others are clearly visible. When a player joins temporarily, they might briefly switch focus to show that person's face before returning to show the active participants in the conversation. For instance, streamer S5 initially used a medium shot to include themselves and one friend. As more friends joined, S5 switched to a full-body shot to capture all five participants together. During conversations, S5 then used a medium-long shot to focus on themselves and two other friends active in the conversation to balance their facial expressions and overall context.

4 Expert Interview

As the second phase of this study (Figure 2), to address RQ2, we conducted expert interviews based on insights from the first phase. Following the analysis of video content, we engaged professionals from various media fields with expertise in camera techniques. We also interviewed theater and performance practitioners adept at using narrative and mise-en scene in free-moving 3D (live) stories. These interviews helped us to understand: 1) how established camera principles can provide VR streamers with practical guidance on effectively utilizing virtual cameras; and 2) how the connections between VR streaming and established media potentially uncover new insights and perspectives. We chose not to involve professional VR streamers in the interview process because our focus of RQ2 was on exploring VR streaming through the perspective of established media and on improving current VR streaming virtual camera setups by applying principles used in traditional media.

4.1 Data Collection

We recruited 10 expert participants (Table 3) through a local university research network and recommendations from the recruited participants. Participants were selected based on three criteria: professional experience in media fields, familiarity with VR technology, and a keen interest in exploring VR streaming as an emerging
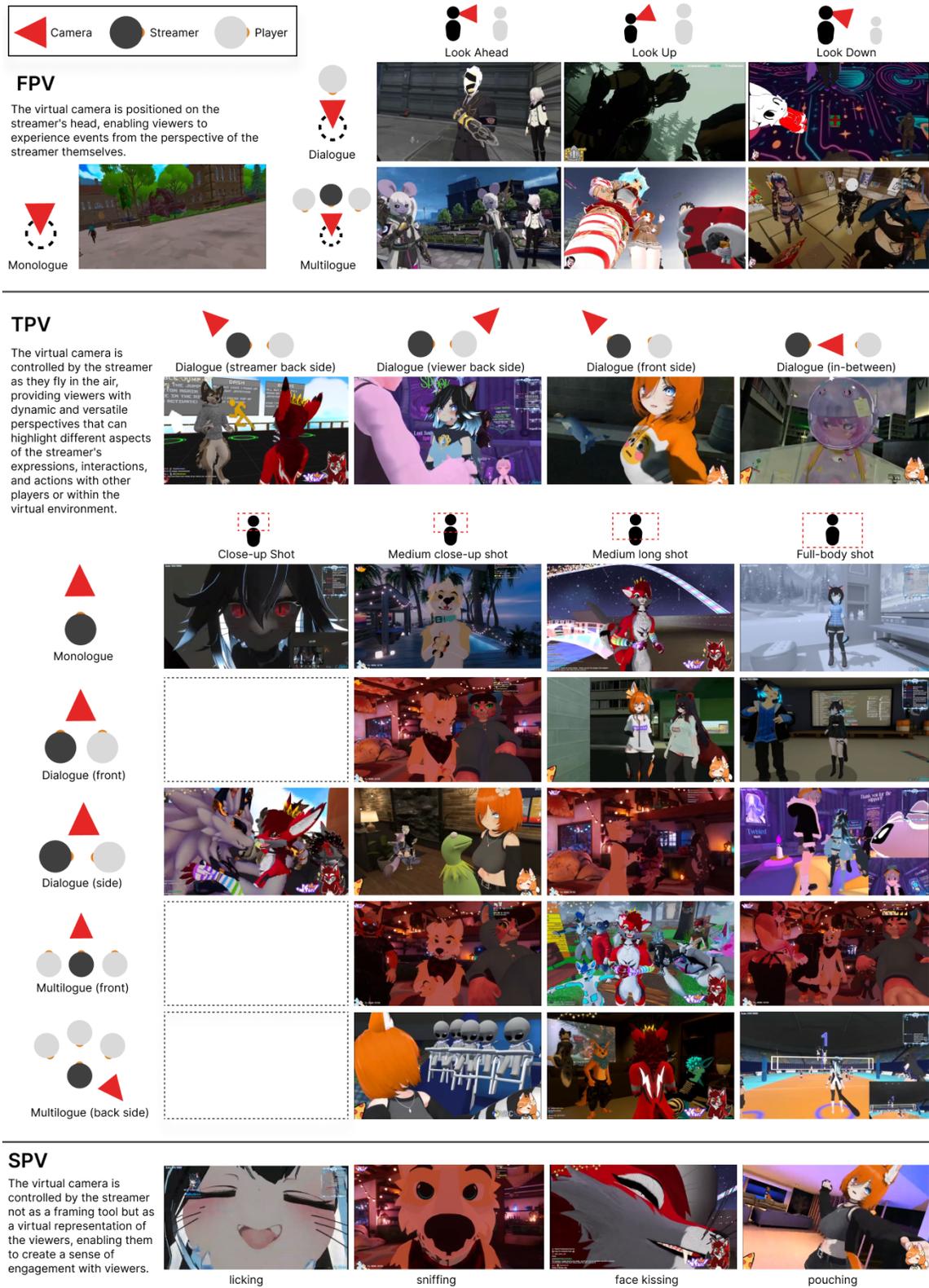
Figure 5: Common camera setups for VR streaming categorized in groups of analogue, dialog and multilog.

Figure 6: Four commonly observed camera frames for VR streaming: A) Close-up shot focusing on the face to offer detailed facial expressions and emotions; B) Medium close-up shot covering the area above the shoulder, achieving a good balance between upper-body gestures and expressions; C) Medium-long shot capturing the area over the knee, allowing viewers to see more of the environment around the streamers; D) Full-body shot including the entire body from head to toe, giving viewers a complete view of the streamer's posture and positioning within their environment.

medium. All participants had at least around 10 years of professional work experience related to camera. Throughout this paper, we will refer to these participants as P#. Each participant received $30 as an appreciation for their time and contribution.

The interview lasted about 60 minutes and were conducted remotely with Zoom through three stages. The initial stage (5 minutes) introduced the interview agenda, providing a brief introduction to VR streaming with two one-minute clips of VR streaming videos, one in FPV and one in TPV. In the second stage (40 minutes), we systematically reviewed all virtual camera usage scenarios identified from our video analysis (Figure 3), collecting participants' thoughts and feedback on each. The focus of this stage was to explore recommended usage methods and potential pitfalls associated with each camera scenario. The final stage (15 minutes) involved a brief discussion on VR streaming. The conversation focused on comparing VR streaming with other well-established media formats and exploring potential design considerations for VR streaming cameras to better support VR streamers.

4.2 Findings: Attitudes and Opinions on VR Streaming

All interviews were recorded and transcribed, with the research team reviewing all materials to identify initial codes and organize them into themes by following Braun and Clarke's thematic analysis framework [9]. We began by categorizing codes into two groups: opinions on VR streaming and feedback on camera techniques. Opinion codes included topics like storytelling in VR and the future potential of the medium, which were later grouped into themes such as debates around VR streaming as an emerging format and anticipated streaming styles (Section 4.2). Feedback codes covering aspects such as medium shots for conversations, poor view composition, lighting issues, and lack of environmental detail, were categorized into design considerations and implications for future VR streaming virtual cameras (Section 5). The identified themes were then shared with the research team for discussion. When there were disagreements or uncertainties regarding specific codes or themes, team members revisited short recording clips using timestamps and engaged in collaborative discussions until a consensus was achieved.

4.2.1 VR streaming can represent a novel form of interactive media. All participants agreed that VR streaming offers an exciting new opportunity for engaging audiences and enhancing storytelling. Compared to movies, which rely entirely on the artist's creative input while the audience remains passive (P8), VR streaming overcomes the physical limitations of traditional media and offers a great potential for inclusive and interactive content, something that traditional media finds difficult to achieve (P7). This unique strength arises from three main factors: the possibility of live interactions with the audience (P1, P3, P5, P7, P8, P10), the diversity of identity embodiment through virtual avatars (P4, P7, P10), and the low-cost and easy-to-explore nature of virtual camera systems (P2, P6, P7, P8).

As pointed out by P10, *"In VR streaming, being able to speak directly to the audience is really really powerful."* This direct communication between streamers and viewers can occur through chat messages and camera manipulation. In particular, streamers can tailor their content based on viewer requests and interact with the camera in a way that *"almost like the streamer can reach out of the camera and directly pull the audience into it."* -P3 The freedom to choose various virtual avatars provides an experience that is difficult to replicate in movies, as P10 noted *"One of the beautiful things about VR is that you don't know who you're going to encounter, or how they want to express themselves as an avatar".* Furthermore, since virtual cameras are easy to create and manipulate, VR streaming enables more creative perspectives, such as bird's-eye views or low-angle shots, which are difficult or costly to achieve in traditional filmmaking. This can also enhance the freedom of expression and exploration in camera usage for VR streaming.

While VR streaming has advantages over traditional media formats, it also faces several unique challenges at current stage. One widely noticed issue is that there is no pre-rehearsing or post-editing involved, making it *"real-time broadcasting, with no chance to change or review later"* -P7. Another challenge is the insufficient visual detail, including environments that lack richness to improve the viewer's experience (P2), facial expressions that fail to convey emotion effectively (P4), and the need for improved lighting to guide attention and sustain engagement (P10).

4.2.2 VR streaming encompasses a diverse range of content across multiple genres. As pointed out by P2, *"for VR streaming, the purpose of storytelling is different, as well as the motivation behind it.",* with VR streaming focusing more on viewer engagement than compelling narratives. In general, most of participants find the current VR streaming more leaning to documentary genres like live broadcasting events and Vlog but believe there is also a huge potential for narrative story telling. Currently, most of the VR streaming is using very limited number of virtual camera setups, making the viewers see the same perspective for a long time. As P8 pointed out, *"the more control he has over the camera, the more narrative he can convey to the audience".* P3 added that being able to provide more camera angles can make the viewers feel the view is more constructed, designed and planned to show or tell a story. In another perspective, as P6 explained, *"VR streaming is more presented in the form of a documentary or vlog. From a cinematic perspective, films like documentaries have less camera movement compared to narrative-style movies".* Most of the current VR streaming cameras

Table 3: Ten professional participants took part in our expert interviews, each with a minimum of around ten years of experience in media and camera-related fields.

ID	Gender	VR Experience	Professional Experience
P1	Male	Played VR Games or Apps	Cinematographer and vice director for 3 theatrical movies (about 10 years)
P2	Male	Developed VR Games or Apps	Cinematographer and digital artist (about 10 years)
P3	Male	Played VR Games or Apps	Critical media and game study researcher (15+ years)
P4	Male	Knowledgable about VR	Filmmaker and cinema study instructor (about 10 years)
P5	Male	Played VR Games or Apps	Director for 2D and 3D animations (15+ years)
P6	Female	Developed VR Games or Apps	Entertainment industry professional and VR filming instructor (about 10 years)
P7	Male	Played VR Games or Apps	Director for 2D animations and advertisements (about 10 years)
P8	Male	Played VR Games or Apps	Filmmaker and virtual production specialist (about 10 years)
P9	Male	Played VR Games or Apps	Theater performance and short film director (20+ years)
P10	Female	Developed VR Games or Apps	Lighting and video designer for theater and VR performance (30+ years)

seldom moves, or just simply following the streamer or auto focusing the streamer, lacking an intention to control the camera movement. Thus, comparing to build narratives and tell a story like films or theater performance, VR streaming currently is more documentary like broadcasting live events in various genres such as personal Vlogging (P2 and P6), TV talk shows (P3 and P9), and sports or events broadcasting (P2, P5, and P7).

However, most participants also believe the potential of narrative story telling in VR streaming, particularly there could be a gradual development from being documentary to narrative. P9 pointed out that *"being in VR and assuming the role of character, being that character, moving through the world and engaging conversation as that character is always fascinating for building narratives."* Being the characters that viewers hardly can be in real life and exploring the world through their perspectives itself provide VR streaming a strong narrative foundation comparing to films that can be hard to achieve these viewing angles. In addition, the game-alike camera angles especially FPV, give the viewers a sense of familiarity of their gaming experience and further promote the sense of interaction and engagement they usually come across in gaming (P2, P3, P8 and P9), which may be uncommon and hard to achieve in the current setting of filming and theater performance. As P8 summarized, *"it's not black and white but it's a continuous spectrum. You can have many different levels of control, corresponding to positions along the spectrum."* VR streaming can be both documentary and narrative, which is versatile in the spectrum. Though the current VR streaming may appear more on the relatively naive and straightforward documenting the streamer activities, VR streaming holds strong potential for engaging and interactive story telling.

4.2.3 VR streaming can draw insights from and offer contributions to other established media formats. Many participants agreed that VR streaming *"has taken a lot of cues from traditional storytelling techniques."* -P3 and *"still relies on the language of traditional filming and performance"* -P4. It is unsurprising that all participants acknowledged VR streaming could be seen as an extension of other traditional media formats and could inherit their well-established principles. Therefore, it can be helpful to enable VR streamers better apply these principles in their streams, as P1 suggested *"First, help VR streamers learn how they can set up their cameras, and then gradually develop patterns or methods."* From another perspective, some participants recognized the uniqueness of VR streaming and

believed that this distinctiveness needs to be more explored. They also thought that VR streaming could potentially contribute back to other well-established media formats. As P4 mentioned, in VR streaming, *"the camera principles can be applied in a different way"*, *"perhaps we should explore more in terms of innovative structures or designs of visual and auditory language"* -P2. P1 summarized *"I don't believe that VR streaming has to be born from traditional filming techniques. In fact, VR streaming needs to find its own unique characteristics."*

5 Design Considerations and Implications

On top of the observed common camera usage patterns (Section 3.2), in this section, we first describe the practical design considerations and recommendations extracted from expert interviews, for VR streamers to plan their streaming camera setups. Then, we present the design implications for VR streaming developers to better support VR streamers to express themselves and enable more genres of VR streaming.

5.1 Design Considerations for VR Streaming Camera Setups

We will first present the design considerations for virtual cameras in FPV and SPV. For TPV, which is more complicated, our discussion will begin with a single static virtual camera, focusing on aspects such as camera height, distance, and view composition. We will then move on to discuss camera movement, followed by an analysis of the functions involved in switching among a set of cameras. See Table 4 for a summary of the design considerations.

5.1.1   First-person View. Participants generally agree that FPV offers an immersive experience, making viewers feel as though they are present in the scene (P2), providing strong involvement (P7), and even allowing them to become fully absorbed in the content (P3). While FPV is commonly used in filming, its application is limited to specific scenes (P4). In contrast, FPV in VR streaming offers greater freedom and creativity in its usage scenarios and is familiar to viewers who also play games, resembling a Vlog captured with a head-mounted Go Pro in real life (P6, P7, and P9). As noted by P1, functionally speaking, this camera angle is more appropriate for tracking or exploration scenarios, such as when streamers navigate a new virtual environment or follow the lead of other players.

Table 4: Summary of design considerations collected through expert interviews for VR streamers.

Camera Aspects	Design Considerations
First Person View	- Pay attention to camera stability and horizontal line - Actively use ambient elements and environment to create visual guidance - Use empty space in the scene as signals for new scenes or transitions - Actively use foreground to establish a reference point for viewers
Second Person View	- May make viewers more aware of their separation from streamer rather than feeling closer - May be perceived as overly intimate or violent
Camera Height	- Use camera height aligned with the eye level of the characters being interacted with - Strategically vary camera heights to enhance engagement and visual interest - Avoid prolonged use of extremely high or low angles to prevent discomfort and maintain natural conversation dynamics - Use one dominant camera height per scene and avoid abrupt changes - Be mindful of perspective distortion at the edges of the view when using very low or high camera heights
Camera Distance	- Use distance to convey intention and guide viewers' attention - Reflect your position within the group or community and attitude towards others - Avoid extreme close-up or long shots to avoid showing lack of details of virtual avatars and environments - Maintain a comfortable social distance during conversations
View Composition	- Actively use foreground and background to create depth and convey spatial information - Intentionally use composition as visual guidance for viewers' attention when interacting with multiple players - Deliberately plan the placement of characters and elements to highlight major roles and avoid information overloading
Camera Movement	- Use larger camera distances during movement to reduce view jitters - Keep the horizontal axis line and region of interest stable during motion - Incorporate subtle jitters for a sense of natural movement and avoid overly rigid stabilization
Switching Camera	- Avoid prolonged use of a single camera view to prevent visual fatigue - Strategically switch cameras to convey information or emotion during conversations - Ensure clear visual distinctions between transitions to avoid confusion

A common issue with FPV is the lack of camera stability, which can lead to sudden shifts and cause visual fatigue over time (P7 and P10). P8 further emphasized that an uneven horizon line may result in viewers feeling dizzy or uncomfortable during movement, addressing the need to minimize such effects. Another frequently mentioned suggestion is using ambient elements and environmental cues to guide viewers' attention toward specific targets or help them anticipate upcoming events in the stream. For example, P1 mentioned that visual guidance can take various forms: narrative-based (such as the streamer explicitly directing where to look), interactive design (like arrows indicating points of interest), or visual structures like corridors and semi-structured spaces that naturally frame the viewer's focus. P2 noted that a clear foreground helps viewers establish reference points and better understand the current scene. Additionally, P5 suggested that an empty FPV view with no characters can signal scene transitions or the start of a new scene.

5.1.2 Second-person View. In our video analysis, we observed that many VR streamers use virtual cameras not merely as a tool for framing videos but rather treat them as representations of their viewers. To capture this unique usage scenario, we introduced the concept of "second-person view" (SPV). However, P1 and P7 expressed caution about this term, suggesting it might be more appropriate to refer to it as an interactive or subject FPV. P7 noted that *"As an actor, when he faces the camera, the director tells him to shoot a subjective shot. He can't treat it just as a camera anymore. He has to imagine it as a little cat or dog, or as his lover."* On the other hand, all other participants found the concept of SPV interesting and viewed it as a powerful tool for streamers to engage directly with viewers.

Although all participants compared SPV to the idea of "breaking the fourth wall" in traditional filmmaking and acknowledged its potential to enhance viewer engagement, some participants raised concerns. P4, P8, and P10 pointed out that *"the fact that viewers cannot really feel the kissing and touching by the streamers, and noticing they are talking to the camera would remind the viewers to draw attention to the mediated nature of the screen"* -P4. This could have an ironic effect, as it may make viewers more aware of their separation from the streamer rather than feeling closer. P3 and P9 also emphasized that certain actions, such as kissing or punching on the face, might be perceived as overly intimate or violent, potentially making them unwelcome to some viewers. P3 added that if a viewer is watching someone they expect to act in a humorous or edgy way, an unexpected intimate gesture might feel out of place and unwelcome. Therefore, the use of SPV should be carefully considered in relation to the target group and the relationship between streamers and viewers.

5.1.3 Camera Height. Echoing the findings in Section 3.2, all participants highlighted the significant impact that camera height has on identity embodiment and the power dynamics between characters. They suggested that streamers should use a camera height aligned with the eye level of the characters they are interacting with, as this can foster a greater sense of equality within the streaming community. However, maintaining an eye-level camera height for extended periods may lead to a monotonous viewing experience,

since it closely resembles real-life perspectives and limits the immersive potential of VR. Participants noted that the unique freedom offered by VR allows for exciting and unconventional angles that are difficult to achieve in traditional media (P1, P6, P9). Therefore, streamers are advised to strategically vary their camera heights from time to time to enhance engagement and visual interest.

Several guidelines were proposed for effective use of camera height. First, streamers should be mindful of how different camera heights affect viewers' emotional responses and comfort levels. While high or low angles can create an engaging experience, prolonged use might cause discomfort or a sense of isolation (P5, P6). P9 also emphasized that such extreme angles may disrupt natural power dynamics during lengthy conversations and recommended combining them with more neutral perspectives to clearly show the faces of all participants. Additionally, P2 and P6 suggested establishing one dominant camera height per scene and avoiding abrupt changes to maintain a smooth and consistent viewing experience. Finally, P8 warned about perspective distortion that occurs when characters are positioned at the edges of the frame using very low or high camera angles, which can negatively affect the overall visual quality and viewing experience.

5.1.4 ↔ *Camera Distance.* Different distances from the camera to the target object being captured can result in various camera frames, which has been discussed in Section 3.2. On top of that, participants highlighted two other important considerations regarding camera distance. First, camera distance serves as a reflection of the streamer's intention, guiding viewers on where to focus their attention. As P3 noted, *"the viewer can notice that there is something significant going on in close-up and even mid-range shots"*. Therefore, VR streamers should carefully choose camera distances based on their communication goals. Given that many current avatars lack detailed facial expressions (P4) and virtual environments often fail to convey ambient context (P3), extreme close-ups or long shots may hinder effective emotional engagement between streamers and viewers.

Camera distance can also influence the sense of involvement, particularly during conversations between streamers and other players. Since the camera represents the viewer's perspective, similar to how we perceive in daily life, P2 suggested that *"camera distance is actually a symbol of our mutual psychological distance or intimacy, or a representation of social relationships"*. This means that if the camera is too far away, viewers may feel like they are not participating in the conversation but rather secretly observing it from a distance (P7). Conversely, if the camera is too close, especially during intimate actions such as streamer kissing with another player, it can make viewers uncomfortable (P4). P1 added that camera distance influences the sense of community, raising questions about one's role and attitude within the group. Participants recommend that VR streamers first reflect their intention when capturing conversations with other players and use camera distance as a means to express their level of engagement and attitudes toward others.

5.1.5 🖼 *Camera View Composition.* Since viewers can only observe streamers' activities in a virtual 3D world through their 2D displays, the arrangement and proportions of elements on these screens determine both the information and intent that streamers

can communicate to their audience. All participants agreed that VR streamers currently do not fully utilize composition in their streaming practices. Three common mistakes were widely noticed by the participants.

First, the current composition often feels flat, with little use of foreground and background to create depth (P3, P4, P7). Although VR streamers are located in a 3D virtual space, most views treat this environment as if it were a 2D background. This approach fails to convey a sense of spatiality to viewers, limiting opportunities for better engagement and understanding of the virtual environment.

Second, during conversations with other players, viewers often lack visual cues such as clear positioning in the frame, suitable camera angles, and adequate lighting to identify who the focus of attention should be. For example, some streamers position their cameras from behind when interacting with others, but *"a big partial back view doesn't provide much information. It is somewhat a waste of (screen) space"* -P8. Additionally, when streamers converse with players who look alike, it can be hard to distinguish between individuals or events due to visual similarity (P5). This is especially challenging for new viewers unfamiliar with the streamer, making it difficult to identify the main character and supporting players (P3 and P9).

Finally, camera views may become overcrowded with many other players or busy background elements, resulting in information overload. This can cause viewers to lose focus and experience discomfort during their viewing session. To address this issue, P2 suggested that streamers should deliberately plan compositions that allow some characters' clothing to blend into the environment while ensuring the streamer or key characters stand out. Additionally, P9 and P10 recommended that streamers avoid cutting off any players at the edges of the view and ensure all faces are visible when a group is present.

5.1.6 ◦➔ *Camera Movement.* Most of the virtual camera usage scenarios identified through our video analysis are static. In contrast, dynamic and complex camera movements seen in filming or theater performances are rarely utilized in current VR streaming setups. Instead, virtual cameras typically move automatically to follow the streamer's movement or focus on them during motion. As P1 noted, *"in VR, movement emphasizes participation. Once something moves, the sense of involvement becomes stronger"*. This suggests a significant opportunity for VR streamers to enhance viewer engagement by making more deliberate use of camera movement. A common challenge across all participants was noticeable jitter and image shaking during camera movement, especially when combined with low video resolution, leading to an uncomfortable viewing experience. To address this problem, two potential strategies were recommended for VR streamers. First, they could use a larger camera distance when moving the camera, as the closer the shot, the more pronounced the jitter and changes will appear (P5). Second, they should attempt to keep the horizontal axis line and region of interest stable during movement, which may help reduce the impact of jitters on the viewing experience (P6).

However, it is important to note that not all camera shaking is undesirable. As P2 mentioned, *"It's actually similar to when a character runs or walks. This kind of subtle jitter is what makes it feel more natural"*. Subtle jitters can contribute to a more natural and

immersive experience. Furthermore, P10 highlighted that virtual cameras, especially those used in FPV, rotate and move faster than real-world counterparts, which can create an unnatural or uncomfortable experience for viewers. Therefore, VR streamers should balance camera movement carefully: avoiding excessive shaking while maintaining enough natural motion to support immersion and viewer comfort.

5.1.7 *Switching Camera.* VR streamers typically use the same camera setup for extended periods. While some streamers do switch between different camera setups, they do so infrequently and with a limited number of options, as noted by all participants. As P10 mentioned, *"I just can't watch 5 min of a single shot"*. This prolonged use of a single camera setup may lead to visual fatigue and reduce viewer engagement. P1, P7, and P9 emphasized that careful planning of a sequence of camera shots is crucial for building narratives. Therefore, the limited frequency of camera switching may contribute to the perception that current VR streaming feels more like a documentary than an engaging story. However, participants were not in favor of frequent camera changes either. They provided three suggestions regarding this issue.

First, while long single-shot sequences can cause visual fatigue, overly frequent transitions may disrupt the viewer's ability to process information and perceive the content effectively. P6 explained that people are accustomed to viewing images in a stable way and find it difficult to follow when camera setups change too often. Second, camera switching can be used strategically to convey information and emotion, especially during conversations. For example, P7 noted that frequent camera switches during a peaceful conversation may inadvertently increase the sense of tension between characters, making the interaction feel less natural and authentic. Finally, camera shots before and after transitions should clearly differ visually to prevent viewer confusion, as P5 emphasized the importance of event boundaries such as changes in movement, lighting, or spatial positioning that can signal the start of a new scene.

5.2 Design Implications for Developing VR Streaming Virtual Camera Systems

Many participants identified a major challenge for VR streamers: they must simultaneously perform three roles (director, videographer, and actor) at the same time. This requires them to "split their brain" while playing the game and engaging with viewers (P10). Operational difficulties are compounded by complex camera interfaces (P1), frequent switching between views (P5), and the lack of post-editing options (P6). To better support streamers in focusing on their actor roles, we distilled participants' suggestions into three essential design implications for future VR streaming camera systems, emphasizing simplifying technical tasks and delegating director and videographer roles to systems or viewers. See Table 5 for a brief summary.

5.2.1 *Simplified and Straightforward Controls.* Virtual cameras can now closely mimic advanced physical ones, offering controls like aperture and focal length. However, much like how many non-professional users rarely use these features in real life, VR streamers may also find them impractical. As a result, displaying

Table 5: Summary of design implications collected through expert interviews for future VR streaming development.

Design Implications	Strategies
🎚️ Simplified Controls	- Highlight only the essential features that are relevant for VR streaming - Always allow VR streamers to see what their viewers are seeing - Embed camera guidelines into its interface to assist streamers during use - Integrate scenario-based camera recommendations within the system
⚙️ Automation Assistance	- Camera can be controlled using voice commands or gestures - Integrate movement-activated auto-stabilization and correction to maintain stable horizontal lines - Adapt autonomously to different streaming scenarios and switch between various camera views
👁️ Viewer Involvement	- Temporarily give camera control to viewers - Provide viewers with their own virtual cameras that orbit around the streamer - Allow users to choose their preferred views from multiple camera angles

all available options might be unnecessary and could complicate the interface. VR streamers face challenges in creatively using their cameras [32, 68], and complex camera operations can distract from their role as actors. An obvious but often overlooked solution is to reduce feature complexity by focusing on essential VR streaming features rather than adding more options. As P2 noted, *"most users are not experts; they're enthusiasts, so the design needs to be practical and straightforward."*

Currently, many VR streamers cannot directly see what their viewers are seeing on their screen captured by the virtual camera. This real-time visibility would help them better assess their performance and quickly address issues (P4). Such feedback allows streamers to make immediate adjustments based on audience perception, much like actors receiving direction during filming. Additionally, many VR streamers may lack knowledge about film-based camera techniques. To support them, future virtual camera systems can integrate cinematography guidelines directly into its interface (P1). For example, they can offer suggestions on framing, composition, lighting, and movement based on best practices, helping to highlight the VR streamer and guide viewer attention. Another approach is to include scenario-based settings within the camera system (P2), allowing streamers to choose from pre-defined configurations tailored for common streaming situations identified in Section 3.2. This would help streamers better focus on their actor roles interacting with other players and viewers rather than managing camera parameters.

5.2.2 *Automation Assistance.* Since the core challenge for VR streamers is managing three roles simultaneously, many participants suggest using automation to delegate camera operations

acting like a virtual production team, thereby allowing streamers to focus more on their performance. Rather than removing the director's role entirely, these automations should assist streamers by managing routine tasks while preserving their creative expression.

First, P4 suggested the virtual camera should be designed to operate similarly to physical drone cameras, which can move around the streamer and be controlled via voice commands or gestures for basic tasks like adjusting settings or changing positions, offering a more intuitive and efficient alternative to manually navigating menus through traditional controllers. Second, an auto-stabilization and correction feature should be integrated into the camera system to assist streamers in maintaining stable horizontal lines and minimizing unwanted shaking (P6). As discussed in Section 5.1.6, the main objective of controlling camera movement is to avoid excessive shaking that may cause viewer discomfort, while also preventing over-correction that can feel unnatural. To achieve precise camera control in physical setups, users typically rely on equipment such as camera stabilizers or cranes, and often require the assistance of a dedicated cameraman to manage the camera movement effectively. In VR streaming, although there is no need for manual operation of virtual cameras, digital stabilization algorithms can provide absolute stability. However, this level of stability may lack the natural, subtle movements that contribute to a more immersive and lifelike visual experience. As a result, VR streamers still need to invest significant effort and focus in managing these details, such as when to enable or disable stabilization and how strong it should be. To address this issue, camera stabilization can activate automatically when motion is detected and apply smooth, dynamic adjustments within a predefined range. This approach would minimize the need for VR streamers to manually control virtual cameras, allowing them to focus more on content delivery. Finally, future camera system can be able to adapt autonomously to different streaming scenarios, enabling automatic switching between various camera views (P10). Unlike manually selecting predefined profiles for specific situations, these automation systems can apply cinema-based principles [11, 16, 30] to provide more refined control over camera parameters than static profiles, completely freeing users from the need to adjust settings manually. This approach can also mimic the role of a professional videographer or even support 360-degree video recording format, allowing streamers to change to any angles effortlessly during live streams (P5, P7), much like editing in post-production.

5.2.3 👁 *Viewer Agency and Involvement.* In addition to delegating camera work to automation systems, another promising approach is to involve viewers in controlling the camera. Several participants suggested that allowing viewers to select content could be more efficient than requiring streamers to create tailored contents. This would free streamers from managing cameras and let them focus on delivering better performances. For viewers, this change can also increase their sense of control and engagement, turning VR streaming into a collaborative experience rather than a one-way performance that can foster stronger emotional connections [33, 68].

In its simplest form, the virtual camera system could temporarily be controlled by the viewers as a means of interaction between streamers and their audience (P6). For instance, viewers might gain control over the camera during specific moments in the stream, such as when they make a donation or participate in an interactive challenge. Such interactions not only provide a sense of participation but also encourage more active participation from viewers. Alternatively, viewers might each have their own virtual cameras that orbit around the streamer like ghosts (P7), allowing them to act as individual directors and customize their viewing experience in their preferred ways, potentially increasing personalization and satisfaction with the content. Another possibility is having multiple cameras streaming from different angles, with viewers able to choose their preferred view while still accessing other perspectives (P8). This could lead to a more flexible and dynamic viewing experience, where the audience has greater control over how they interact with and consume the content.

6 Immersion Triangle: a Conceptual Framework

Through expert interviews, we identified that immersion is a key concept across three domains closely related to VR streaming: VR, streaming, and filming. While the experts noted nuanced differences in how immersion is understood within each domain, they all emphasized its relevance and essential contribution to understanding VR streaming. In this section, we propose *Immersion Triangle* (Figure 7), a conceptual framework developed through our study to help researchers think of VR streaming as an emerging form of mass media, which we believe could benefit future work in the field.

6.1 Three Types of VR Streaming Immersion

For filming experts, immersion tends to lean toward storytelling and *being them*, indicating that viewers can be fully engaged by the story presented on screen and feel connected to its characters and their worlds. We refer to this as *narrative immersion*.

In contrast, participants with more experience in VR emphasized that immersion is more closely related to the feeling of *being there*. This form of immersion is not necessarily tied to storytelling but rather to elements such as the virtual environment, lighting, sound, and interactions that create a sense of presence within the virtual world. We call this *spatial immersion*.

In the context of live streaming, immersion relates more to the feeling of *being together* and community engagement. As one of the major and unique characteristics of live streaming, the ability to feel connected not only to the streamer but also to other viewers and collectively shape a sense of community, has been highlighted in prior research [27, 42, 50, 66] as a vital aspect for the success of live streaming. We refer to this kind of immersion as *social immersion*.

6.2 Immersion for Streamer and Viewer

These three types of immersions together contribute to the overall immersion in a VR streaming experience. They can be considered from both the streamer's perspective and the viewer's perspective. It is possible that enhancing one side's immersion may affect the other, but there are also scenarios where both sides can collaborate to strengthen their respective levels of immersion, which represents an ideal scenario for VR streaming.

Social immersion is particularly interconnected between both the streamer and the viewer. When social immersion is enhanced on

one side, it often has a positive impact on the other. However, as a negative example, if a system encourages the streamer to focus only on chatting with frequent viewers, this may deepen the streamer's sense of connection but leave many non-frequent viewers feeling ignored and less socially immersed.

In terms of narrative immersion, the current setup typically involves the streamer acting as the storyteller while the viewer passively consumes the story which creates a one-direction flow. Ideally, when the streamer becomes more immersed in their own storytelling, this can also enhance the viewers' narrative immersion. On the other hand, if the streamer compromises their genuine interest by tailoring content excessively to audience preferences, it may reduce their own level of narrative immersion, which could negatively affect the viewer's experience as well.

Spatial immersion may look particularly relevant for streamers, as they are present in the virtual environment. However, spatial immersion can also apply to viewers. Even when using 2D displays, viewers can still gain a sense of and feel connected to the 3D virtual world being presented. To improve viewer-side spatial immersion, streamers may need to use more varied camera angles based on filmmaking principles. This, however, could also impact the streamer's experience if they have to focus too much on technical aspects rather than engaging naturally in the environment.

6.3 VR Streaming as a Sequence of Immersion Triangles

In our video analysis, we found that VR streaming sessions can last for hours, but streamers do not repeat the same activity or game content consistently. Instead, these sessions can be divided into blocks, with streamers systematically switching between them to keep their content engaging or take short breaks from the VR experience [68]. We propose that each block has a dominant type of immersion and the overall session consists of a sequence of these dominant immersions. A successful VR streaming session is thus a structured blend of the three types of immersion.

Since streamers have limited energy and focus, concentrating on one immersion type naturally reduces attention to the others. For instance, when prioritizing social interaction through chat and voice, streamers are less able to engage with the virtual environment or continue storytelling. Conversely, when a streamer is deeply focused on the activity they are performing, or on interacting with other players within the virtual environment, they can hardly interact with their viewers, thereby experiencing less social immersion.

From the streamer's perspective, we can categorize current VR streaming session blocks into three distinct types: 1) narrative immersion prioritized, where the streamer is involved in storytelling or role playing for viewers (Figure 7, A); 2) social immersion prioritized, where the streamer focuses primarily on chatting and interacting with their viewers (Figure 7, B); and 3) spatial immersion prioritized, where the streamer is engaged in playing VR games, interacting with other players, or exploring the virtual environment (Figure 7, C).

From the viewer's perspective, we argue that most current VR streaming is social immersion dominated. This reflects both the unique strength of live streaming as a means of fostering community engagement and the fact that chat messages serve as the primary tool for viewers to interact with VR content. As discussed in our design implications (Section 5.2), future systems could offer viewers greater agency, thereby promoting the development of spatial and narrative immersion in VR streaming.

7 Limitations and Future Work

Our work has several limitations. First, we primarily collected camera usage scenarios from VRChat, which is the most popular VR game streamed on Twitch. While we believe it offers the most diverse camera usage patterns, other types of virtual camera usage patterns may exist in different social VR games or VR game genres that were not captured. Future research could explore these differences across various game genres.

Second, our analysis was based solely on video data, which enables an objective investigation of common camera usage patterns in VR streaming. Although incorporating interviews with professional VR streamers could further enrich our study, our current method provides sufficient and reliable data to address our research questions regarding common virtual camera setups within this specific research context. However, this approach may not fully reflect the streamers' subjective motivations and rationales behind their choices. Future studies could prioritize the inclusion of professional VR streamers to gain a more comprehensive understanding of their motivations, decision-making strategies, and inner thought processes when configuring their cameras during streaming. By incorporating in-depth interviews and surveys with these experts, future studies could uncover insights that go beyond observable behaviors, thereby enriching the understanding of how camera setups influence streaming experience and viewer engagement.

In addition, our design considerations and implications were informed by interviews with camera experts across various media fields, providing comprehensive and practical feedback on current virtual camera setups in VR streaming and actionable insights that fully supported investigating our research questions on applying established camera principles to enhance virtual camera setups. Our analysis of expert interviews aimed to extract practical insights that can help VR streamers improve their use of virtual cameras, drawing on established principles from other mature media fields. However, our approach may lack depth in a systematic comparison between VR streaming and traditional cinematography practices. Future research could involve a more thorough and comprehensive comparative analysis to better understand these differences. Additionally, since not all participants were proficient with VR or streaming, our top-down approach may also overlook potential practices and needs within VR streaming communities. Similarly, our focus on film and animation might cause us to overlook other narrative-based formats like live theater, which relies on performer-audience interaction and emphasizes design elements such as lighting to guide attention and experience [47], potentially offering valuable insights for VR streaming design. Involving professional VR streamers might provide additional insights into comparing traditional media and VR practices. Thus, future research could explore other media fields or adopt a bottom-up approach involving professional VR streamers to better compare and understand the relationships and differences between traditional media production and VR streaming.

Figure 7: Immersion Triangle and examples of the three types of immersion in VR streaming: A) narrative prioritized immersion; B) social prioritized immersion; C) spatial prioritized immersion.

Finally, our study mainly concentrated on the technical elements of virtual camera setups when compared to established media such as film and animation. This focus may have led to the omission of some other important non-technical aspects that VR streaming can learn from these mature forms of media, such as narrative techniques, typical camera operations and strategies, as well as standard workflows and related support systems. Future studies could explore these areas to expand our understanding of camera setups in a broader context for VR streaming.

In addition to addressing the above limitations and extending our work, future studies could consider two broader directions to explore VR streaming's potential to deepen our understanding of both VR and live streaming. On one hand, future work could extend existing theories and conceptual frameworks from traditional live streaming [5, 27, 50] to the specific context of VR streaming to better identify its unique strengths and challenges. On the other hand, building on our discussion in both the design implications (Section 5.2) and the Immersion Triangle section, future research could explore systems that delegate streamer roles, empower viewers with greater agency, or enhance immersion types to foster more engaging and interactive VR streaming experiences for both streamers and viewers.

8 Conclusion

In this paper, we present a two-phase study examining virtual camera usage patterns in VR streaming and gathering insights from experts in media-related fields to better understand and improve the VR streaming experience. We have identified common trends in camera views, heights, distances, and setups, along with expert recommendations that can guide future design considerations for VR streamers. Additionally, we have summarized these expert suggestions into design implications aimed at delegating streamers' roles more effectively and empowering viewers with greater agency. We propose the Immersion Triangle as a conceptual framework to encourage reflection on immersion within VR streaming contexts and support the exploration of its design aspects. Our work highlights VR streaming as a distinctive form of mass media with significant potential to deepen our understanding of both VR and live streaming, and it encourages future research to further explore improving engagement and interaction between streamers and viewers.

Acknowledgments

References

[1] Laurensia Anjani, Terrance Mok, Anthony Tang, Lora Oehlberg, and Wooi Boon Goh. 2020. Why do people watch others eat food? An Empirical Study on the Motivations and Practices of Mukbang Viewers. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. http://doi.org/10.1145/3313831.3376567

[2] Lisa Anthony, YooJin Kim, and Leah Findlater. 2013. Analyzing user-generated youtube videos to understand touchscreen use by people with motor impairments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1223–1232. http://doi.org/10.1145/2470654.2466158

[3] Daniel Arijon. 1991. *Grammar of the Film Language*. Silman-James Press. Google-Books-ID: 6bQlAQAAIAAJ.

[4] Amirsaman Ashtari, Stefan Stevšić, Tobias Nägeli, Jean-Charles Bazin, and Otmar Hilliges. 2020. Capturing Subjective First-Person View Shots with Drones for Automated Cinematography. *ACM Trans. Graph.* 39, 5 (Aug. 2020), 159:1–159:14. https://doi.org/10.1145/3378673

[5] Ava Bartolome and Shuo Niu. 2023. A Literature Review of Video-Sharing Platform Research in HCI. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–20. https://doi.org/10.1145/3544548.3581107

[6] 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 (Dec. 2020), 3490–3502. https://doi.org/10.1109/TVCG.2020.3023567 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[7] David Bordwell. 2013. *EBOOK: Film Art: An Introduction*. McGraw Hill. Google-Books-ID: CIZvEAAAQBAJ.

[8] David Bordwell, Janet Staiger, and Kristin Thompson. 2003. *The Classical Hollywood Cinema: Film Style and Mode of Production to 1960*. Routledge, London. https://doi.org/10.4324/9780203358818

[9] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. In *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological*. American Psychological Association, Washington, DC, US, 57–71. https://doi.org/10.1037/13620-004

[10] Jianhui Chen and Peter Carr. 2014. Autonomous Camera Systems: A Survey. https://www.semanticscholar.org/paper/Autonomous-Camera-Systems%3A-A-Survey-Chen-Carr/33bec3da9d1f1eb3b8aa0503ef4f3031a02a145a

[11] David B. Christianson, Sean E. Anderson, Li-wei He, David H. Salesin, Daniel S. Weld, and Michael F. Cohen. 1996. Declarative camera control for automatic cinematography. In *Proceedings of the thirteenth national conference on Artificial intelligence - Volume 1 (AAAI'96)*. AAAI Press, Portland, Oregon, 148–155.

[12] Emily Dao, Andreea Muresan, Kasper Hornbæk, and Jarrod Knibbe. 2021. Bad Breakdowns, Useful Seams, and Face Slapping: Analysis of VR Fails on YouTube. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Number 526. Association for Computing Machinery, New York, NY, USA, 1–14. http://doi.org/10.1145/3411764.3445435

[13] Ian Drosos and Philip J. Guo. 2022. The Design Space of Livestreaming Equipment Setups: Tradeoffs, Challenges, and Opportunities. In *Proceedings of the 2022 ACM*

Designing Interactive Systems Conference (DIS '22). Association for Computing Machinery, New York, NY, USA, 835–848. https://doi.org/10.1145/3532106.3533489

[14] Steven Drucker and David Zeltzer. 1994. Intelligent Camera Control in a Virtual Environment. (May 1994). https://www.microsoft.com/en-us/research/publication/intelligent-camera-control-in-a-virtual-environment/

[15] Steven M. Drucker, Tinsley A. Galyean, and David Zeltzer. 1992. CINEMA: a system for procedural camera movements. In *Proceedings of the 1992 symposium on Interactive 3D graphics (I3D '92)*. Association for Computing Machinery, New York, NY, USA, 67–70. https://doi.org/10.1145/147156.147166

[16] Steven M. Drucker and David Zeltzer. 1995. CamDroid: a system for implementing intelligent camera control. In *Proceedings of the 1995 symposium on Interactive 3D graphics (I3D '95)*. Association for Computing Machinery, New York, NY, USA, 139–144. https://doi.org/10.1145/199404.199428

[17] Katharina Emmerich, Andrey Krekhov, Sebastian Cmentowski, and Jens Krueger. 2021. Streaming VR Games to the Broad Audience: A Comparison of the First-Person and Third-Person Perspectives. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Number 445. Association for Computing Machinery, New York, NY, USA, 1–14. http://doi.org/10.1145/3411764.3445515

[18] Travis Faas, Lynn Dombrowski, Alyson Young, and Andrew D. Miller. 2018. Watch Me Code: Programming Mentorship Communities on Twitch.tv. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (Nov. 2018), 50:1–50:18. https://doi.org/10.1145/3274319

[19] C. Ailie Fraser, Joy O. Kim, Alison Thornsberry, Scott Klemmer, and Mira Dontcheva. 2019. Sharing the Studio: How Creative Livestreaming can Inspire, Educate, and Engage. In *Proceedings of the 2019 on Creativity and Cognition (C&C '19)*. Association for Computing Machinery, New York, NY, USA, 144–155. https://doi.org/10.1145/3325480.3325384

[20] Guo Freeman and Dane Acena. 2022. "Acting Out" Queer Identity: The Embodied Visibility in Social Virtual Reality. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2 (Nov. 2022), 263:1–263:32. https://doi.org/10.1145/3555153

[21] 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 (Jan. 2021), 239:1–239:27. https://doi.org/10.1145/3432938

[22] Guo Freeman, Divine Maloney, Dane Acena, and Catherine Barwulor. 2022. (Re)discovering the Physical Body Online: Strategies and Challenges to Approach Non-Cisgender Identity in Social Virtual Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3491102.3502082

[23] Michael Gleicher and Andrew Witkin. 1992. Through-the-lens camera control. *SIGGRAPH Comput. Graph.* 26, 2 (July 1992), 331–340. https://doi.org/10.1145/142920.134088

[24] 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 (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 4021–4033. https://doi.org/10.1145/3025453.3025683

[25] Lassi Haaranen. 2017. Programming as a Performance: Live-streaming and Its Implications for Computer Science Education. In *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE '17)*. Association for Computing Machinery, New York, NY, USA, 353–358. https://doi.org/10.1145/3059009.3059035

[26] Oliver L. Haimson and John C. Tang. 2017. What Makes Live Events Engaging on Facebook Live, Periscope, and Snapchat. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 48–60. http://doi.org/10.1145/3025453.3025642

[27] 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 SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 1315–1324. https://doi.org/10.1145/2556288.2557048

[28] A.J. Hanson and E.A. Wernert. 1997. Constrained 3D navigation with 2D controllers. In *Proceedings. Visualization '97 (Cat. No. 97CB36155)*. 175–182. https://doi.org/10.1109/VISUAL.1997.663876

[29] Jeremy Hartmann and Daniel Vogel. 2022. Enhanced Videogame Livestreaming by Reconstructing an Interactive 3D Game View for Spectators. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3491102.3517521

[30] Li-wei He, Michael F. Cohen, and David H. Salesin. 1996. The virtual cinematographer: a paradigm for automatic real-time camera control and directing. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques (SIGGRAPH '96)*. Association for Computing Machinery, New York, NY, USA, 217–224. https://doi.org/10.1145/237170.237259

[31] Juan Pablo Hourcade, Sarah L. Mascher, David Wu, and Luiza Pantoja. 2015. Look, My Baby Is Using an iPad! An Analysis of YouTube Videos of Infants and Toddlers Using Tablets. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1915–1924. http://doi.org/10.1145/2702123.2702266

[32] Yang Hu and Guo Freeman. 2025. Understanding Social VR Streamers' Unique Challenges in Managing Cross-Reality Social Interactions Through Multi-dimensional VR Interfaces. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25)*. Association for Computing Machinery, New York, NY, USA, 1725–1739. https://doi.org/10.1145/3715336.3735825

[33] 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, 1–19. https://doi.org/10.1145/3706598.3713561

[34] Hongda Jiang, Bin Wang, Xi Wang, Marc Christie, and Baoquan Chen. 2020. Example-driven virtual cinematography by learning camera behaviors. *ACM Trans. Graph.* 39, 4 (Aug. 2020), 45:45:1–45:45:14. https://doi.org/10.1145/3386569.3392427

[35] Brennan Jones, Anna Witcraft, Scott Bateman, Carman Neustaedter, and Anthony Tang. 2015. Mechanics of Camera Work in Mobile Video Collaboration. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, New York, NY, USA, 957–966. https://doi.org/10.1145/2702123.2702345

[36] Patrick Keating. 2009. *Hollywood Lighting from the Silent Era to Film Noir*. Columbia University Press. Pages: 312 Pages.

[37] Thomas Kilkenny. 2023. Virtual Cinematography with Unreal Engine: Lean how Unreal Engine's virtual cinematography framework has given rise to new and advanced techniques in film and animation studios.. In *ACM SIGGRAPH 2023 Talks (SIGGRAPH '23)*. Association for Computing Machinery, New York, NY, USA, 1–2. https://doi.org/10.1145/3587421.3604476

[38] Aida Komkaite, Liga Lavrinovica, Maria Vraka, and Mikael B. Skov. 2019. Underneath the Skin: An Analysis of YouTube Videos to Understand Insertable Device Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–12. http://doi.org/10.1145/3290605.3300444

[39] Pascal Lessel, Alexander Vielhauer, and Antonio Krüger. 2017. Expanding Video Game Live-Streams with Enhanced Communication Channels: A Case Study. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1571–1576. http://doi.org/10.1145/3025453.3025708

[40] Christophe Lino, Marc Christie, Roberto Ranon, and William Bares. 2011. The director's lens: an intelligent assistant for virtual cinematography. In *Proceedings of the 19th ACM international conference on Multimedia (MM '11)*. Association for Computing Machinery, New York, NY, USA, 323–332. https://doi.org/10.1145/2072298.2072341

[41] 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 (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3411764.3445660

[42] 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*. Association for Computing Machinery, New York, NY, USA, 1–13. http://doi.org/10.1145/3173574.3174040

[43] Deirdre Lyons, Stephen G. Butchko, Jason Moore, Brendan Bradley, and Tanya Leal Soto. 2021. Live Performance in VR: Live performance in virtual reality by creators from different metaverses discuss the challenges and advantages of performance in this new storytelling platform.. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference (SIGGRAPH '21 Panels)*. Association for Computing Machinery, New York, NY, USA, 1–2. https://doi.org/10.1145/3450617.3464495

[44] Jock D. Mackinlay, Stuart K. Card, and George G. Robertson. 1990. Rapid controlled movement through a virtual 3D workspace. *SIGGRAPH Comput. Graph.* 24, 4 (Sept. 1990), 171–176. https://doi.org/10.1145/97880.97898

[45] Kris Malkiewicz. 2012. *Film Lighting: Talks with Hollywood's Cinematographers and Gaffers*. Simon and Schuster. Google-Books-ID: Z6o4bNpbJEsC.

[46] Joseph A. Maxwell. 2010. Using Numbers in Qualitative Research. *Qualitative Inquiry* 16, 6 (July 2010), 475–482. https://doi.org/10.1177/1077800410364740 Publisher: SAGE Publications Inc.

[47] Rachel McDonnell, Bharat Vyas, Uros Sikimic, and Pisut Wisessing. 2025. Feeling Blue or Seeing Red? Investigating the effect of light color, shadow and realism on the perception of emotion of real and virtual humans. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers (SIGGRAPH Conference Papers '25)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3721238.3730728

[48] Gerald Millerson. 2013. *Lighting for TV and Film* (3 ed.). Routledge, New York. https://doi.org/10.4324/9780080573496

[49] Michael Nebeling, Shwetha Rajaram, Liwei Wu, Yifei Cheng, and Jaylin Herskovitz. 2021. XRStudio: A Virtual Production and Live Streaming System for Immersive Instructional Experiences. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Number 107. Association for Computing Machinery, New York, NY, USA, 1–12. http://doi.org/10.1145/3411764.3445323

[50] Anthony J. Pellicone and June Ahn. 2017. The Game of Performing Play: Understanding Streaming as Cultural Production. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 4863–4874. http://doi.org/10.1145/3025453.3025854

[51] Stephen Prince. 2012. *Digital Visual Effects in Cinema: The Seduction of Reality*. Rutgers University Press. https://www.jstor.org/stable/j.ctt5hj5pv

[52] Pablo Pueyo, Juan Dendarieta, Eduardo Montijano, Ana Cristina Murillo, and Mac Schwager. 2024. CineMPC: A Fully Autonomous Drone Cinematography System Incorporating Zoom, Focus, Pose, and Scene Composition. *IEEE Transactions on Robotics* 40 (2024), 1740–1757. https://doi.org/10.1109/TRO.2024.3353550 Conference Name: IEEE Transactions on Robotics.

[53] Barry Salt. 2009. *Film Style and Technology: History and Analysis*. Starword. Google-Books-ID: KCTkQwAACAAJ.

[54] Mohamed Sayed, Robert Cinca, Enrico Costanza, and Gabriel Brostow. 2022. LookOut! Interactive Camera Gimbal Controller for Filming Long Takes. *ACM Trans. Graph.* 41, 3 (March 2022), 30:1–30:16. https://doi.org/10.1145/3506693

[55] Max Sjöblom and Juho Hamari. 2017. Why do people watch others play video games? An empirical study on the motivations of Twitch users. *Computers in Human Behavior* 75 (Oct. 2017), 985–996. https://doi.org/10.1016/j.chb.2016.10.019

[56] David Stump and ASC. 2021. *Digital Cinematography: Fundamentals, Tools, Techniques, and Workflows* (2 ed.). Routledge, New York. https://doi.org/10.4324/9780429468858

[57] Joshua Suttor, Julian Marin, Evan Verbus, and Meng Su. 2019. Implement AI Service into VR Training. In *Proceedings of the 2019 2nd International Conference on Signal Processing and Machine Learning*. ACM, Hangzhou China, 114–121. https://doi.org/10.1145/3372806.3374909

[58] John C. Tang, Gina Venolia, and Kori M. Inkpen. 2016. Meerkat and Periscope: I Stream, You Stream, Apps Stream for Live Streams. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, San Jose, California, USA, 4770–4780. https://doi.org/10.1145/2858036.2858374

[59] 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 (UIST '19)*. Association for Computing Machinery, New York, NY, USA, 161–174. https://doi.org/10.1145/3332165.3347872

[60] Balasaravanan Thoravi Kumaravel, Cuong Nguyen, Stephen DiVerdi, and Björn Hartmann. 2019. TutoriVR: A Video-Based Tutorial System for Design Applications in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–12. http://doi.org/10.1145/3290605.3300514

[61] Balasaravanan Thoravi Kumaravel, Cuong Nguyen, Stephen DiVerdi, and Bjoern Hartmann. 2020. TransceiVR: Bridging Asymmetrical Communication Between VR Users and External Collaborators. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (UIST '20)*. Association for Computing Machinery, New York, NY, USA, 182–195. https://doi.org/10.1145/3379337.3415827

[62] Balasaravanan Thoravi Kumaravel and Andrew D Wilson. 2022. DreamStream: Immersive and Interactive Spectating in VR. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3491102.3517508

[63] Colin Ware and Steven Osborne. 1990. Exploration and virtual camera control in virtual three dimensional environments. In *Proceedings of the 1990 symposium on Interactive 3D graphics (I3D '90)*. Association for Computing Machinery, New York, NY, USA, 175–183. https://doi.org/10.1145/91385.91442

[64] Finn Welsford-Ackroyd, Andrew Chalmers, Rafael Kuffner dos Anjos, Daniel Medeiros, Hyejin Kim, and Taehyun Rhee. 2021. Spectator View: Enabling Asymmetric Interaction between HMD Wearers and Spectators with a Large Display. *Proceedings of the ACM on Human-Computer Interaction* 5, ISS (Nov. 2021), 485:1–485:17. https://doi.org/10.1145/3486951

[65] Frederik Winther, Linoj Ravindran, Kasper Paabol Svendsen, and Tiare Feuchtner. 2020. Design and Evaluation of a VR Training Simulation for Pump Maintenance Based on a Use Case at Grundfos. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Atlanta, GA, USA, 738–746. https://doi.org/10.1109/VR46266.2020.00097

[66] 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 (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174048

[67] Donghee Yvette Wohn, Peter Jough, Peter Eskander, John Scott Siri, Masaho Shimobayashi, and Pradnya Desai. 2019. Understanding Digital Patronage: Why Do People Subscribe to Streamers on Twitch?. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '19)*. Association for Computing Machinery, New York, NY, USA, 99–110. https://doi.org/10.1145/3311350.3347160

[68] 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 (Nov. 2023), 437:242–437:265. https://doi.org/10.1145/3626473

[69] Enes Yigitbas, Ivan Jovanovikj, Janis Scholand, and Gregor Engels. 2020. VR Training for Warehouse Management. In *26th ACM Symposium on Virtual Reality Software and Technology*. ACM, Virtual Event Canada, 1–3. https://doi.org/10.1145/3385956.3422106

[70] Jung Eun Yoo, Kwanggyoon Seo, Sanghun Park, Jaedong Kim, Dawon Lee, and Junyong Noh. 2021. Virtual Camera Layout Generation using a Reference Video. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3411764.3445437

[71] David Zeltzer and Steven Drucker. 1992. A Virtual Environment System for Mission Planning. (July 1992). https://www.microsoft.com/en-us/research/publication/a-virtual-environment-system-for-mission-planning/