Virtual Reality & Intelligent Hardware

• Article •

December 2024 Vol. 6 No. 6: 527–549
10.1016/j.vrih.2024.12.001

Previs-Real: Interactive virtual previsualization system for news shooting rehearsal and evaluation

Che QU[1,2†], Shaocong WANG[2†], Chao ZHOU[2], Tongchen ZHAO[2], Rui GUO[2], Cheng Wa WONG[3], Chi DENG[4], Bin JI[4], Yuhui WEN[5,6], Yuanchun SHI[2], Yong-Jin LIU[2*]

1. *Department of Film and Television Technology, Beijing Film Academy, Beijing* 100088*, China;*
2. *Department of Computer Science and Technology, MOE-Key Laboratory of Pervasive Computing, Tsinghua University, Beijing* 100084*, China;*
3. *School of Social Sciences, Tsinghua University, Beijing* 100084*, China;*
4. *News Production and Broadcasting First Division, China Media Group, Beijing* 100038*, China;*
5. *School of Computer and Information Technology, Beijing Jiaotong University, Beijing* 100044*, China;*
6. *Beijing Key Laboratory of Traffic Data Analysis and Mining, Beijing Jiaotong University, Beijing* 100044*, China*

Received 28 October 2023; **Revised** 7 July 2024; **Accepted** 10 December 2024

Abstract: Background In the demanding field of live news broadcasting, the intricate studio production procedures and tight schedules pose significant challenges for physical rehearsals by cameramen. This paper explores the design and implementation of a lightweight virtual news previsualization system, leveraging virtual production technology and interaction design methods to address the lack of fidelity in presentations and manipulations, and the quantitative feedback of rehearsal effects in previous virtual approaches. **Methods** Our system, Previs-Real, is informed by user investigation with professional cameramen and studio technicians, and adheres to principles of high fidelity, accurate replication of actual hardware operations, and real-time feedback on rehearsal results. The system's software and hardware development are implemented based on Unreal Engine and accompanying toolsets, incorporating cutting-edge modeling and camera calibration methods. **Results** We validated Previs-Real through a user study, demonstrating superior performance in previsualization shooting tasks using the virtual system compared to traditional camera setups. The findings, supported by both objective performance metrics and subjective responses, underline Previs-Real's effectiveness and potential in transforming news broadcasting rehearsals. **Conclusions** Previs-Real eliminates the requirement for complex equipment interconnections and team coordination inherent in a physical studio by implementing methodologies complying the above principles, objectively resulting in a lightweight design of applicable version of virtual news previsualization system. It offers a novel solution to the challenges in news studio previsualization by focusing on key operational features rather than full environment replication. This design approach is equally effective in the process of designing lightweight systems in other fields.

Keywords: News broadcast; Previsualization; Human computer interaction design; User-centered design; Interface evaluation

Supported by Research Project of the State Key Laboratory of Ultra HD Video and Audio Production and Broadcasting Presentation of China Media Group (CMGSKL2021KF015); the Natural Science Foundation of China (62332019).

[†]**Contributed equally**

[*]**Corresponding author**, liuyongjin@tsinghua.edu.cn

Peer review under the responsibility of Beijing Zhongke Journal Publishing Co. Ltd.

Citation: Che QU, Shaocong WANG, Chao ZHOU, Tongchen ZHAO, Rui GUO, Cheng Wa WONG, Chi DENG, Bin JI, Yuhui WEN, Yuanchun SHI, Yong-Jin LIU. Previs-Real:Interactive virtual previsualization system for news shooting rehearsal and evaluation. Virtual Reality & Intelligent Hardware, 2024, 6(6): 527–549.

1 Introduction

News studio programs, live or not, are produced under tight schedules, and the efficient realization of precise framing and camera movement in the program relies on the proficient operation of the cameraman. Typically, the cameraman is expected to accurately capture multi-camera keyframes within the shooting scene in strict accordance with the timeline, which necessitates comprehensive knowledge of the studio and cameras to execute reliable and precise operations, thus requiring specific pre-practice. Nevertheless, the conventional rehearsal approach presents a challenge as it depends on the original studio setup and requires coordinated presence of lighting, packaging, and other technicians. This often proves difficult when conflicting with formal programming schedules[1]. Due to the intricate and distinctive nature of news broadcast studios, establishing an additional site for rehearsals is prohibitively expensive. Therefore, developing a cost-effective, high-fidelity virtual previsualization environment utilizing computer-generated imagery (CGI) techniques is of great significance for news studio cameramen, akin to various professional skill simulation systems employed in the absence of physical conditions[2–4].

However, in comparison to these increasingly sophisticated systems meeting urgent needs in specialized domains[5,6], current virtual previsualization systems for newscasts lack a targeted approach for camera rehearsal tasks and exhibit incomplete functionality[7]. Our collaboration and continuous communication with news studio professionals have identified that the problems need to be addressed lie in three main aspects: fidelity, hardware interaction, and rehearsal feedback. Fidelity is a comprehensive requirement that demands explicit methodologies to ensure the coincidence of geometric accuracy, rendering, controllability, and framing with the real-world circumstance. Nevertheless, these aspects are often not well addressed simultaneously in existing systems, leading to the inability of rehearsals conducted in virtual environments to serve as a substitute for those held in physical studios. Second, current previsualization systems lack dedicated hardware interaction devices for shooting rehearsals, resulting in reduced physical realism of the operations. Finally, the evaluation of framing and camera movement rehearsals typically relies solely on the subjective judgment of experienced technicians and lacks descriptive, objective criteria to follow. And subjective feedback sometimes relies on a submission and checking process that is not real-time and lacks an approach of automated evaluation feedback to ensure the efficiency of rehearsals.

In this paper, building upon relevant research and our formative investigations, we introduce Previs-Real (Figure 1), a virtual news previsualization system designed to elevate the realism and self-supporting capabilities for rehearsals in news studio shooting. When creating 3D scene assets, Previs-Real incorporates the hierarchical reconstruction workflows we established that respectively support high-quality previsualization and quick preview modes. And the scene fidelity can be achieved combined with the use of Unreal Engine's real-time rendering capabilities (Figure 1a,b). The user interface (UI) facilitates the simulation of studio running status, including screen content and lighting conditions. Utilizing the Unreal Engine (UE) plug-in development toolset, Previs-Real attains direct control of asset events in PIE (Play In Editor) mode within UE's timeline (Sequencer module) through our UI, including the activation of lens files, and thus offers concurrent fidelity in scene presentation, element controllability, and lens distortion (Figure 1a, c). The controller of Previs-Real is capable of performing the main daily operational functions for cameramen and faithfully replicates their operational habits, thereby enhancing rehearsal realism (Figure 1d). Based on the image similarity algorithms, the feedback module of Previs-Real can provide an automatically calculated evaluation of the quality of framing and camera movement operations (Figure 1e).

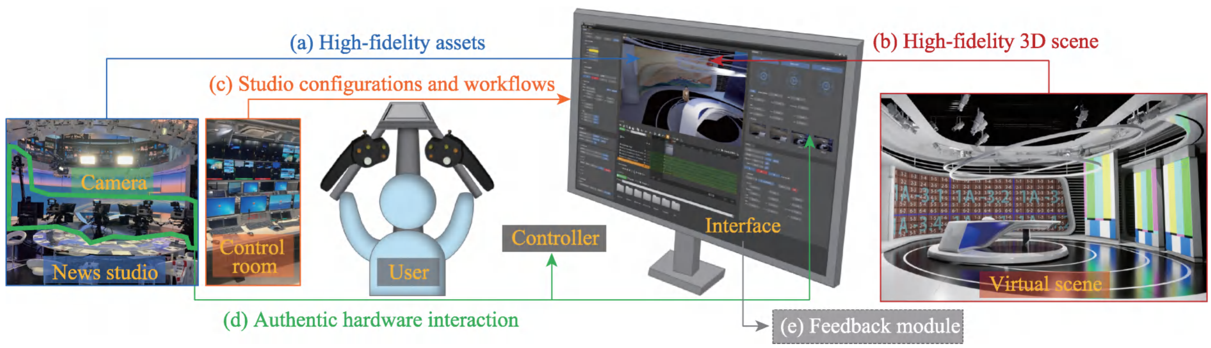


Figure 1 The framework of the virtual previsualization system Previs-Real, which supports high-fidelity of (a) virtual assets including (b) 3D scene, and (c) news studio configurations and workflows, (d) authentic hardware interaction, and (e) a feedback module for virtual rehearsals.

This research follows a user-centered design (UCD) cycle, with distinct phases of user research (Section 3), system development (Section 4), and user testing (Section 5). Briefly, in response to the three challenging aspects mentioned above, we make the following contributions:

(1) We implemented the formative interviews guided by the concept of UCD to identify design elements and principles tailored to the existing challenges, which informed the design of the system's software UI and hardware interaction controller.

(2) We presented Previs-Real, which integrates UE with the frontier virtual production and CGI technologies, and interaction design methods to address the limitations of previous virtual rehearsal approaches. This system enables users to perform virtual rehearsals as an alternative to physical ones, thereby compensating for resource constraints.

(3) Employing the UI, controller and feedback module of Previs-Real, we conducted user experiments and a questionnaire survey involving 27 participants with cinematography backgrounds. The results confirmed the effectiveness and positive interaction experience of Previs-Real in camera rehearsal tasks, and thus the validity of the design principles.

2 Related work

In the context of virtual technology in the film and television industry, it is crucial to distinguish between two sets of concepts: virtual production/virtual studio, and virtual previsualization. The former emphasizes the composition of fake backgrounds (including studio backgrounds) as well as virtual foreground objects with live-action characters or props to achieve formal programming, while the latter emphasizes the implementation of realistic previsualization of concerned aspects of the shoot[8]. The two are distinct yet interconnected; certain techniques of the former can also be applied to virtual previsualization, albeit with limitations, while the latter originates from film virtual previsualization and has a divergent focus from that of news broadcasting virtual previsualization. Additionally, they both benefit from the substantial advancements in game engine technology driven by the evolution of graphics technology in recent years. Mainstream engines such as UE and Unity provide high-quality real-time rendering effects and extensive interface development flexibility[9,10], offering a promising outlook for the application of virtual simulation technology. Regarding hardware interaction, many of the concepts encapsulated in prior research provide valuable inspiration for the design of dedicated hardware controllers.

2.1 Virtual news studio based on virtual production

The prototype of virtual news broadcasting studios emerged as early as the 1980s with the maturation of blue and green screen keying techniques[11,12]. These techniques allowed real subjects to be composited onto virtual backgrounds, providing viewers with novel and enriched experiences, representing an early version of virtual production technologies. In today's news broadcasting domain, virtual broadcasting based on augmented reality (AR) technology can also implant virtual effects into real footage[13,14]. Scenes where news anchors interact realistically with virtual objects enhance the audiences' experience. This common form of online packaging in the broadcasting field integrates technologies such as motion capture[15], camera tracking[16,17], and workflow coordination management. It has become an indispensable interactive technology across a spectrum of television programs, including news and variety shows, to enhance expressiveness.

Online packaging systems like Vizrt, Aximmetry, Hirender and ViBox VS for instance, are mainly used to achieve precise composition of virtual elements with real footage in terms of spatial positioning and timing during online broadcasting control[18]. However, when utilized for previsualization, some of these systems lack corresponding scene models, and even if they include model assets, these models may lack specific controllability designed for previsualization. Additionally, the rendering prioritizes the computer-generated (CG) feel of special effects over realism. Although virtual previsualization is feasible through the secondary development of the online packaging system[7], this system is integrated with the studio and requires connection to the local area network (LAN) of the entire broadcasting control environment. As some controlled objects, such as screens, lights, sound equipment, and robotic arms are physical devices, it is not straightforward to generate corresponding controllable virtual assets. Furthermore, the majority of these systems lack adequate camera calibration functionality. For instance, Vizrt's calibration process is laborious and limited, while Aximmetry's calibration procedure is intricate and appears to lack an explicit distortion model. In conclusion, these systems are not tailored for previsualization purposes; they offer extensive functionality but require substantial hardware and software components while being costly and lacking cost-effectiveness for previsualization. Moreover, none of them incorporates a suitable hardware controller or automatic effect evaluation.

2.2 Virtual previsualization in film production and news broadcasting

Virtual previsualization technology is commonly employed in the film industry to plan and rehearse the shooting process using digital graphic technology, thereby avoiding the high costs of trial and error during formal production[19]. This includes creating CG animated previsualizations based on storyboards[19–21], constructing virtual cinematic environments for trial previsualizations[22], and using virtual reality (VR) for collaborative creative processes[23–26]. However, it's important to note that the requirements for fidelity[27] in these activities can vary depending on the specific stages of the production. In many cases, the emphasis is not on achieving high fidelity but rather on providing basic creative references for directors and producers or verifying the technical feasibility of crucial sequences. Hence, there is often no necessity for an all-encompassing, full-process universal previsualization workflow. In summary, film previsualizations vary across projects, with each previz requiring a specific production time for a single program. This process involves the use of diverse digital content creation (DCC) software, including game engines. However, these software's native interfaces are inadequate for meeting the rehearsal needs of cameramen in terms of practice and feedback due to their inability to align with cameramen's operational habits and limited interactivity. Therefore, the virtual previsualization methods employed in the film industry are not directly applicable to the field of news broadcasting.

Unlike films, news broadcast operations are standardized and routinized. We can utilize UE's plugin development toolset to establish a universal interface that integrates UE's robust asset blueprinting capabilities and high-quality rendering performance for diverse virtual broadcasting camera rehearsals. However, it is crucial to address the compatibility issue of controlling assets in PIE mode directly from the plugin interface due to the involvement of calling asset blueprints in the timeline. Furthermore, a set of comprehensive and user-friendly 3D scene reconstruction methodologies is imperative.

2.3 Hardware interaction based on the UCD concept

Virtual productions, virtual studios, and traditional virtual previsualization prioritize the production of video clips while neglecting the development of dedicated hardware controllers. In the realm of virtual previsualization for film and television, some studies have employed improved VR forms or VR-based natural interaction interfaces[23,26], aiming to achieve both portability and immersion simultaneously, but not quite matching the needs of camera rehearsal in virtual studios. According to the experience in the field of UCD, the difficulty of design sometimes lies not in the relentless pursuit of technological advancement, but in the accurate grasp of users' needs[28]. By conducting structured or semi-structured interviews with the target users, potential user requirements can be more thoroughly explored, so that the design can be better adapted to the application[29,30].

Some studies on physical interaction interfaces have shown that customized adaptations tailored to specific application scenarios, achieved through various technological means, can simulate and align better with the characteristics of the scenario, resulting in improved interaction effects[2,31–34]. Inspired by these studies, for cameramen's usage, adopting a hardware interaction interface that simulates real-world equipment might be a suitable choice. Therefore, we have developed a controller specifically designed to the operational habits of cameramen, incorporating an inertial measurement unit (IMU) and other electronic components to enhance the physical interaction aspect of our previsualization system.

3 Formative interview

The purpose of this formative interview is to gain a deeper understanding of users' perspectives on the requirements for the virtual news previsualization system, so as to identify design elements and principles that can address the present challenges.

3.1 Participants

Seven participants (two females and five males, age=31.1 ± 4.3) were recruited for this interview. They are all senior studio staff members, 5/7 have over 10 years' experience and 2/7 have over 5 years'. They are tasked with responsibilities such as virtual asset production and management, and camera adjustment and maintenance, and very familiar with the studio production and broadcasting procedures and program requirements. Two of them are familiar with camera operation, and two of them have preliminary experience in the use of UE. They form a complete team for running a news studio and are suitable for this formative interview. The interviews were conducted face-to-face, and each lasted about 40 minutes.

3.2 Interview findings

The main content of the interviews centered around the functions and requirements that would be covered by the target system. At the beginning of the thematic analysis, we organized our interview results on the level of coded sub-themes according to a direct division of labor categories of the development process. The division into three groups encompasses the software interface that implements comprehensive functionality, the hardware controller associated with camera operations, and the auxiliary functions that fulfill additional functionality and feature requirements, which correspond to the subsequent sections below. Furthermore, we discovered that after a higher level of coding, these specific requirements could ultimately be encapsulated into three design principles aimed at addressing the three challenges outlined at the beginning. Hereafter, in the thematic analysis framework, P stands for principle (theme) and S for sub-theme. For sub-themes, we summarized the perspectives put forward by participants in the interviews.

3.2.1 Functional requirements for software interface

S1: The studio's running mode (7/7). The camera shooting and directorial control processes take place in the studio and the broadcast control room, respectively. The cameraman can view the composited shooting frames with virtual assets through a monitor, which are transmitted to the control room via the camera control unit (CCU). Based on this visual feed, the director coordinates camera movements, video switching, audio arrangements, AR and subtitle packaging, as well as supervises studio lighting adjustments and display content on each screen (Figure 1). These operations are facilitated through communication interfaces between the studio and the control room.

S2: Self-supporting rehearsal preparation. Before the camera rehearsal, the studio cameras, screens, lighting, packaging effects and other relevant aspects can be configured (5/7). This process can be carried out independently, and if it necessitates cooperation with other team members, it compromises convenience (2/7). The packaging effects align consistently with reality (6/7).

S3: User friendliness. Native UE functionalities are decentralized and project-specific (2/7); it is desirable to have a generalized interface (5/7) that is user-friendly and easy to use (7/7).

3.2.2 Functional requirements for hardware controller

S4: The cameras' operating sets (Figure 2). Camera types in a studio conventionally encompass jib and stationary cameras equipped with various broadcasting lenses (5/7), and ideally a free-positioned camera is integrated into the virtual system (2/7). The controller is adaptable to all types through multiple configuration schemes (4/7). Operations involve pan, tilt, zoom, focus, boom pan, and boom tilt, and the cameraman employs a combination of these operations to achieve framing and camera movement (7/7). For

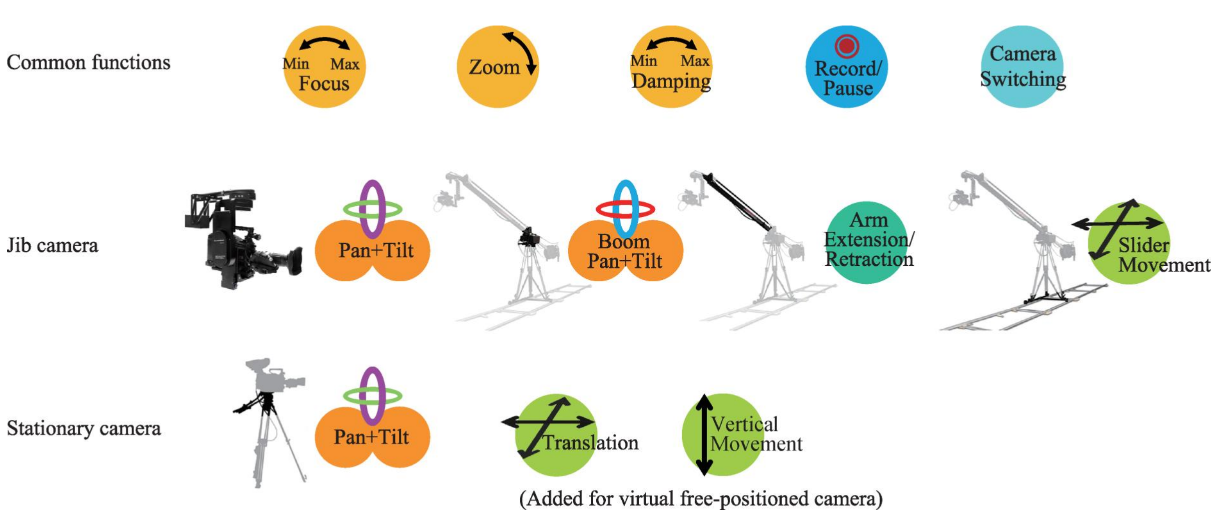


Figure 2 The operating sets of studio cameras.

specific programs, aperture adjustments are not required (1/7).

S5: Operation experience. The controller is designed to emulate the operational characteristics of real broadcast cameras, including gimbal control modes (e.g. pan, tilt) and key control modes (e.g. zoom) (3/7), while reserving some keys to support customized functions (2/7).

3.2.3 Auxiliary function requirements

S6: Verisimilitude level. The original virtual assets within the online packaging system, primarily AR assets, do not support export. Therefore, the previsualization system needs to reconstruct scene assets and achieve the controllability of lightboxes, screens, and other components (7/7). Unlike film previsualization's focus on realistic rendering or animation tests, it is essential to prioritize geometric accuracy in scene reconstruction to avoid deviation from actual framing in the previsualization (5/7). The rendering focus of the online packaging system is on special effects performance, while the previsualization system can draw on realism rendering from film previsualization (3/7).

S7: Lens simulation. The virtual camera enables focusing, zooming, and distortion correction (3/7). The manual camera calibration procedure in the adopted online packaging system is overly complex, and it is desirable for the virtual system to streamline the procedure (2/7).

S8: Objective evaluation models and metrics for rehearsals. Traditional evaluations rely on subjective opinions of experts, which may not be easily accessible. Although the operation will not be concerned with precise parameter settings, is it possible to perform automatic computational evaluations by other metrics? If automatic evaluation can be achieved, in conjunction with self-supporting preparation (S2), the previsualization system can serve multiple purposes including program ideation, rehearsal facilitation, and novice training (3/7).

3.3 Design principles

Based on the coded sub-themes, we further coded to obtain three themes as the design principles to guide the development of the virtual news previsualization system.

P1: Fidelity. As is explained, this principle incorporates the geometry precision and rendering realism (S6) of scene model assets, consistent controllability of studio scene elements in line with their real-world counterparts (S1, S2, S6) through a universal interface (S3), and the alignment between virtual and live images through the often overlooked yet indispensable virtual camera calibration (S7).

P2: Hardware interaction that aligns with operational habits. While adhering to authentic operating modes can be considered a form of fidelity, we regard the controller as an independent creative control system, encompassing multi-type camera compatibility (S4), gimbal and key control mode replication, and key function flexibility (S5).

P3: Automatic rehearsal feedback. Designing a standardized and extensible automated evaluation module, utilizing techniques such as image computing (S8).

4 Virtual news previsualization system design and implementation

The news previsualization system Previs-Real aims to offer a virtual platform that realistically reproduces on-site conditions in a broadcasting studio, encompassing the scene, shooting, packaging control, cameramen's operational habits, and rehearsal feedback. The system architecture guided by the design principles is illustrated in Figure 3. The system's hardware components include a workstation that runs the UE-based system, a video capture card for camera calibration, display devices, and the hardware controller for physical interactions. The hardware and software configurations for asset generation tools are also labeled in Figure 3.

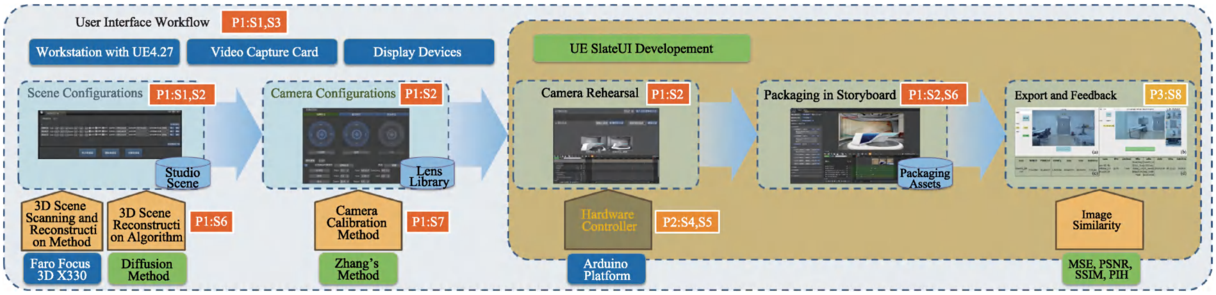


Figure 3 The architecture diagram of Previs-Real labeled with design principles, hardware configurations, and software methods.

4.1 Software interface of Previs-Real

The software interface as shown in Figure 4 is implemented as a plugin based on UE 4.27 and is primarily divided into three main parts: (1) digital assets, (2) virtual cameras and (3) the production system and workflow, in accordance with the fidelity design principle (P1). The interface comprises multiple functional modules including the scene control module, camera control module, storyboard module, standard viewport, track interface, and resource browser (Figure 4b). These modules serve to facilitate various aspects of the production process such as packaging control, camera switching, shot generation control, storyboard arrangement and management.

Figure 4 The software interface of Previs-Real. (a) Editor mode prior to the launch of the previsualization environment; (b) PIE mode subsequent to the launch of the previsualization environment.

4.1.1 Studio controllability

Interface integration through SlateUI is not simply a direct assembly of existing functions. When controlling Actor components with blueprints using a non-native interface, it is necessary to manage data transfer between the Editor and PIE modes. In the context of supporting non-linear editing of tracks, via UE's reflection system, we can dynamically retrieve instance types and invoke their attributes and functions to realize the generation and destruction of various virtual assets, as well as the display of specific effects. The functionality for storyboard and shot track control is achieved by deconstructing Sequencer's Track-Section-Channel-Key objects to realize the automatic running of the broadcast timeline.

By utilizing the integrated interface (P1: S3), the following workflows can be accomplished, with the former two in Editor mode (Figure 4a) and the latter two in PIE mode (Figure 4b):

(1) Selecting the broadcast studio, deploying cameras, attaching lens files, and presetting lens parameters in accordance with on-set conditions (P1: S1, S2, S6).

(2) Adjusting scene parameters and previewing digital packaging assets, such as content displayed on

screens, lighting boxes, AR assets, and subtitles (P1: S2, S6).

(3) Operating virtual cameras to generate reusable key frames for automatically creating shot frames by inserting them into the track. Alternatively, shot frames can be manually created through continuous recording. Regardless of their generation method, the frames can be previewed in the standard viewport (P1: S2, S3).

(4) In the storyboard control module, arranging shot frames, applying packaging animations to the timeline track, and generating a complete previz for export (P1: S2, S3, S6).

4.1.2 3D scene reconstruction

We employ the following methodology to create 3D scene assets according to the hierarchical requirements:

Laser scanning approach: A Faro Focus 3D X330 laser scanner[35] was used to capture precise point cloud data of the broadcasting studio scene as shown in Figure 5. The studio, covering an area of approximately 400 square meters, underwent scanning at 22 stations over a total duration of around 5 hours. Additionally, comprehensive photographs were taken to document the overall structure and detailed elements on-site, aiding in establishing spatial relationships within 3D DCC software. This dataset was also leveraged for generating rendering materials and textures. The outcome yielded a millimeter-level accurate point cloud, while maintaining the error in the 3D scene within 5 centimeters after completion of the DCC software modeling.

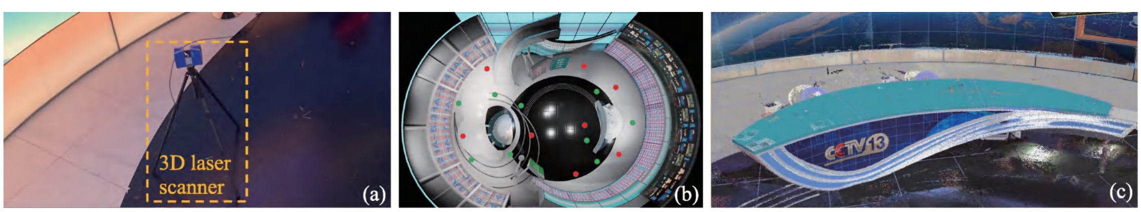


Figure 5 3D point cloud laser scanning and schematic of the reconstruction precision. (a) Working state of the laser scanner; (b) schematic of the scanning positions on the ground floor of the studio, red for normal height (approx. 1.7m) stations, green for lower height (approx. 0.5m) stations; (c) alignment between the point cloud and the model represented by the mesh lines.

Computer vision approach: An RGB-D data-based method[36] was utilized to swiftly reconstruct a preliminary scene model. This technique was primarily employed to provide rapid overviews of the scene.

The aforementioned two methodologies synergistically integrate to establish a comprehensive tool system that leverages laser scanning techniques for acquiring intricate models to facilitate precise program previsualization, and employs computer vision algorithms for generating rough models to enable creative preview. Upon integration of the scene model into the software interface, realistic visualization can be achieved through UE's materials and real-time rendering capability (P1: S6), while also allowing for easy configuration of component controllability (P1: S2, S3, S6).

4.1.3 Lens parameters calibration

In addition to achieving precise scene reconstruction, another crucial aspect in aligning virtual and real shooting images is the calibration of virtual lens parameters. This primarily involves determining distortion parameters for specific focusing distances and focal lengths. By leveraging Zhang's calibration method[37], absolute focal length values and their corresponding distortion parameters can be inversely computed in UE for specific focusing distances.

Using the identical method as described in Section 4.1.1, we transfer the lens file instance data within the LiveLink component from Editor mode to PIE mode to enable the lens distortion to take effect during track playback. In Previs-Real, the accuracy of camera calibration and 3D reconstruction can be directly visualized by comparing the deviation between scenes in virtual and real captured frames (Figure 6) to mutually verify the accuracy of camera calibration and 3D reconstruction (P1: S7).

Figure 6 Comparison of co-registered virtual and real frames captured with identical camera and lens parameters. (a) Pure real frame; (b) both real and virtual frames set at 50% transparency; (c) pure virtual frame.

4.2 Hardware controller of Previs-Real

We have developed a specialized controller as depicted in Figure 7, meticulously crafted based on the Arduino platform using insights obtained from the formative interview in Section 3.2.2 and specific subsequent research. The design features of the controller are manifested in the following three areas.

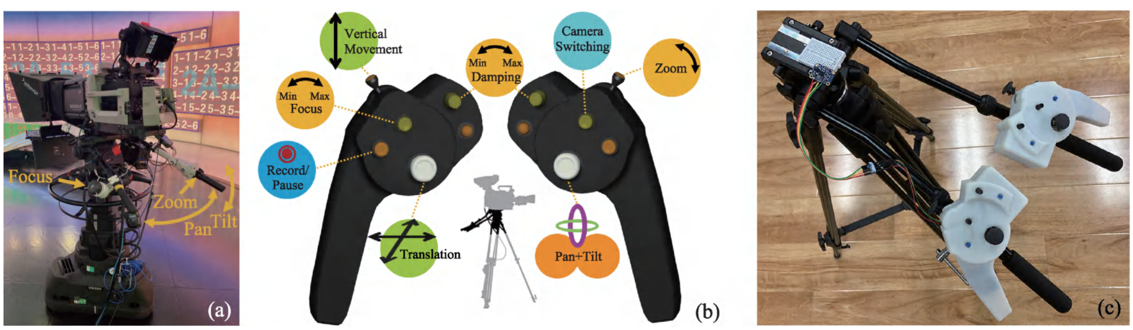


Figure 7 Hardware controller in the real and virtual systems. (a) The operating set of an actual stationary studio camera; (b) the controller design scheme for the stationary camera; (c) the implemented controller.

(1) Covering the cameras' operating sets. The controller design requires comprehensive coverage of the entire operating sets for all types of studio cameras (P2: S4). This involves addressing control over five spatial dimensions (translation along the *xyz* axes and yaw (pan) and pitch (tilt) rotations, mapped to five potentiometers), focus, zoom, and aperture adjustments (mapped to three potentiometers), damping control (mapped to two potentiometers), as well as mode switching and customized functionalities (comprising no fewer than five switches).

(2) Covering all camera types with multiple configuration schemes. In addition, potentiometers and switches are thoughtfully grouped into three distinct sets, each tailored to meet the control requirements associated with one of the three camera types (P2: S4).

(3) Replicating cameramen's operational habits.

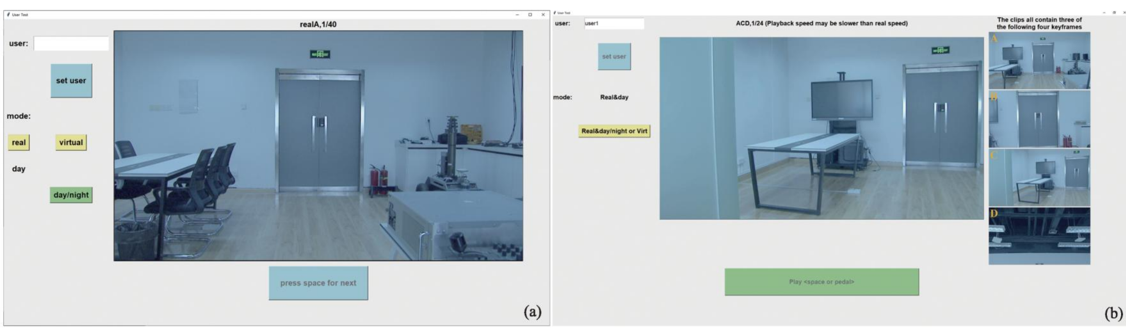
Potentiometer mapping method: To flatten the learning curve of the controller and thus ensure the previsualization efficiency, we mapped the functions of potentiometers and switches according to cameramen's regular operational habits (P2: S5). Therefore, although some functions may use the same electronic components, such as focus and zoom adjustments that employ potentiometers to convert rotational angles into output values, it is crucial to map the focus value to the absolute lever angle and the zoom value to the relative angle of the unidirectional auto-return lever in order to replicate the operational logic of an authentic camera's servo lens controller.

Replication of the gimbal control modes: The controller needs to replicate the gimbal control modes of multiple camera types (P2: S5). For instance, the cameraman typically needs to perform pan or boom pan, and tilt or boom tilt operations for stationary or jib cameras. We emulate the control mode utilizing an activatable IMU motion sensor as shown in Figure 7c.

Ergonomic design: The controller's model structure takes into consideration the ergonomics and mounting dimensions of the potentiometers, battery, and ESP32 board to establish a stable and reliable internal framework while ensuring a comfortable grip. Figure 7c illustrates the controller mounted on a tripod to simulate a stationary camera. The positioning of the core function keys (Figure 7b) and the height and spacing of the two fixed handles are tailored to align with actual cameraman operations (P2: S5). During the previsualization control, the controller communicates bi-directionally with the software interface over UDP.

4.3 Automatic rehearsal feedback module

In the context of traditional studio camera rehearsals, precise framing according to the envisioned vision is paramount. The broadcast footage is composed of several key frames, and cameramen manipulate camera movement to facilitate transitions between these key frames. To fully align Previs-Real with rehearsal requirements, we developed two switchable feedback interfaces (Figure 8a, b) for key frame framing and camera movement (P3: S8).



case	MSE/0	PSNR/inf	SSIM/1	date	time	datetime
User02_B6virt	23.52067	34.41631	0.928242	20240315	16:00:05	110742 (c)

case	PIH	simBool	tflie	ufile	date	time	datetime
user02_Real&day_CABfps10.53	(0.95	True)	Real/day/0316_Real&day_CABrfps9.34.mp4	results/user02/user02_Real&day_CABfps10.mp4	20240316	16:15:13	342878 (d)

Figure 8 Evaluation feedback interfaces and examples of similarity results. (a), (c) for framing tasks; (b), (d) for camera movement tasks.

For the still and video images of the two tasks, each is preloaded with an expandable set of commonly used standard frames to assess the quality of the captured frames, with distinct calculation methods employed for each accuracy evaluation. For still images, framing quality is assessed by comparing the similarity between the standard image as shown in Figure 8a and the captured image obtained by taking screenshots in the system backstage using three computational metrics: mean square error (MSE), peak signal to noise ratio (PSNR), and structural similarity (SSIM)[38], which are commonly used to assess image quality, with greater similarity to the standard image representing higher quality. For video frames, camera movement accuracy is measured by comparing the video similarity between the standard clip as shown on the left of Figure 8b and the shot clip obtained by recording in the backstage, with perceptual image hashing (PIH) selected as the computational metric[39] to make a rapid and robust assessment. Examples of the feedback results for both tasks are illustrated in Figure 8c and d, and the calculation principles for each metric are detailed in Appendix A.

5 User study

To assess the performance of Previs-Real and users' experience, we conducted a user study comprising two experimental tasks and a subjective scale evaluation.

5.1 User experience study design

Based on the two modules designed in Section 4.3 for real-time assessment of user performance, we established the framing shooting task (Task 1) and camera movement shooting task (Task 2) respectively. The procedures for these tasks will be detailed in Sections 5.4 and 5.5, correspondingly. Within the experimental task sections, all participants were designated as camera operators and tasked with completing two phases of shooting assignments: firstly, the previsualization shooting tasks in Phase 1, followed by the formal shooting tasks in Phase 2. Participants were divided into two groups; during Phase 1, the control group utilized the actual shooting system while the experimental group employed Previs-Real. Both groups used the actual shooting system for their formal tasks in Phase 2, which had an identical format to those in Phase 1. The formal task performance of users who rehearsed the same task using different systems will reveal the role played by the rehearsal system. If Previs-Real successfully fulfills the user requirements for previsualization outlined in Section 3, it is anticipated that the virtual group of participants will perform no less effectively than the real group in Phase 2 live-action tasks.

In the subjective evaluation section, participants who have experienced both previsualization systems (i.e., those from the control group who completed the experimental tasks were subsequently invited to experience Previs-Real) will be asked to assess their perceptions of the fidelity of Previs-Real and the consistency of operational experience between Previs-Real and the real shooting system by completing several questionnaires. These questionnaires include the Igroup Presence Questionnaire (IPQ), which assesses the virtual reality experience and serves as a subjective evaluation metric for the realism of the virtual system. Additionally, the overall usability will be assessed using the third edition of the Post-Study System Usability Questionnaire (PSSUQ), the System Usability Scale (SUS) and the Usefulness, Satisfaction, Ease of Use and Ease of Learning (USE) scale to evaluate user experience and satisfaction.

5.2 Participants

This ongoing study recruited 27 participants (14 males and 13 females, age=23.9 ± 2.5), all with normal or corrected vision, right-handedness, and a background in photography but no studio experience to eliminate the influence of professional experience on experimental results. Among them, 9 (6 males and 3 females) were assigned to framing shooting rehearsals, while the remaining 18 (8 males and 10 females) were assigned to camera movement shooting rehearsals. Participants were recruited independently for the two experimental tasks, and because of the simplicity of Task 1, the results reached significance more quickly, thus Task 1 had half the number of participants as Task 2. The experimental and control groups were balanced as closely as possible in terms of gender, comprising 3 males and 2 females in the experimental group for Task 1, and 4 males and 6 females in the experimental group for Task 2. Prior to the experiment, participants received a comprehensive orientation session covering Previs-Real's construction background, experimental objectives, and procedures. This included an elucidation of the basic operating methods of virtual or real shooting systems and provided opportunities for users to acquaint themselves with the system. Subsequent to obtaining informed consent from each participant, the study lasted approximately 30-40 minutes. The research was conducted in compliance with the principles outlined in the Declaration of Helsinki and approved by our affiliate's Human Research Ethics Committee.

5.3 Experimental laboratory environment and instruments

The layout of the laboratory setup is depicted in Figure 9a from a top-down perspective. To ensure full control over environmental variables such as furniture positioning and to secure adequate experimental time, the user experience evaluation of Previs-Real was conducted in a temperature-controlled and well-lit laboratory using the same broadcast camera employed in the actual studio, and other conditions were carefully regulated to maintain stability, thus enabling replication of the same experimental effects experienced in the real studio during specific camera rehearsal tasks. A virtual scene based on laser scanning 3D reconstruction method described in Section 4.1.2 was created to replicate the spatial dimensions and furnishings of the laboratory prototype as accurately as possible.

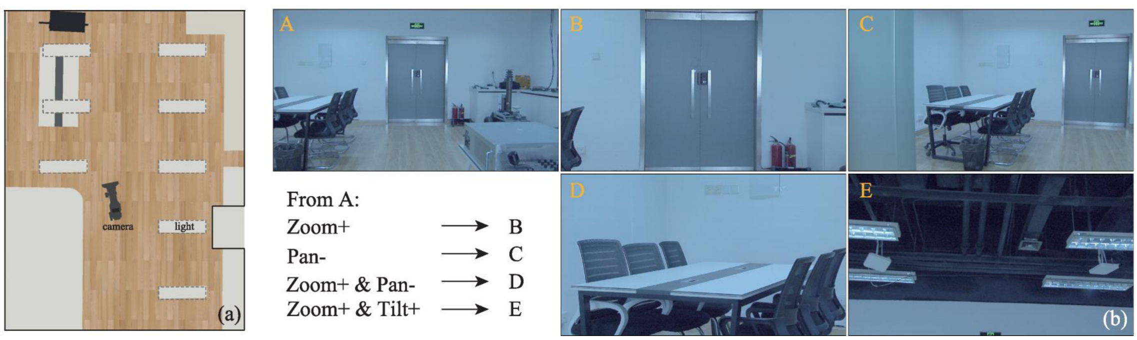


Figure 9 Experimental laboratory and stimulus. (a) Top-down layout of the laboratory setup; (b) the five frames A, B, C, D and E correspond to specific camera positions and three focal lengths (A, C-7.6mm, B-14.6mm, D-18.9mm, and E-11.4mm) at a focusing distance of 3m. Operations from A to other frames are indicated.

The real shooting system utilizes a Sony HDC1580 broadcast camera, equipped with a Fujinon HA18× 7.6BERD-S48 lens, connected to a CCU, and transmits footage to the workstation via a DeckLink Studio 4K video capture card. The software interface of Previs-Real operates on the workstation featuring an Intel Core i9-12900K CPU and an ASUS ROG RTX 3090Ti O24G Gaming graphics card.

5.4 Task 1: Framing shooting rehearsal experiment

5.4.1 Experimental stimulus

In Task 1, we initially selected five standard static frames (A, B, C, D, E) for evaluation as shown in Figure 9b, where each of them represents a specific camera position and posture, and one of four different focal lengths at a 3-meter focusing distance. These frames were pre-recorded in the laboratory under the guidance of professional studio technicians and loaded into Previs-Real. Subsequently, we recorded the virtual counterparts of these frames for the virtual previsualization group. Typically, news shots are standardized and patterned. Therefore, we can create several keyframes to represent the typical framing based on the laboratory layout. The library of standard keyframes can be expanded to accommodate specific scene and program requirements.

Then we generate a 40-frame sequence for each user as stimulus materials, ensuring that the order of presentation of the 5 keyframes is equilibrated using the Latin Square design. The stimulus material consisted of 5 blocks, each containing 10 stimuli in a fixed order (2 occurrences of each of the 5 keyframes). Each participant reproduced the stimulus sequence composed of 4 randomly selected blocks, with 20 stimuli from 2 blocks used in Phase 1 and 20 stimuli from the remaining 2 blocks used in Phase 2. The random selection and ordering of the blocks were equilibrated across participants.

5.4.2 Experimental procedure

The experimental task involves replicating the 40-frame sequence through panning, tilting, and zooming operations with a stationary camera. There is a 2-second refresh interval between each two frames with an empty display area. The shooting effect is measured by comparing the similarity between the shot and the standard frames (refer to Section 4.3). The rehearsal feedback module interface has been shown in Figure 8a, allowing users to preselect options for real (day or night) and virtual conditions in order to match different standard frames in the similarity calculation.

Figure 10 illustrates a participant engaging in both virtual and live shooting tasks. Subsequent to the completion of each frame shooting, the participant step on a foot pedal, and the evaluation module will automatically pair corresponding standard frames for calculating image similarity. This involves comparing frames captured by participants using a physical camera with real (day or night) standard frames, and comparing those captured through the virtual system with virtual standard frames. Additionally, completion time will be recorded. Task completion time refers to the duration from presentation of the stimulus frame until the participant steps on the pedal after completing the frame shooting. There is a slight interval between Phases 1 and 2.

Figure 10 Two-phase images of a participant in the experimental group. (a) Manipulating Previs-Real's controller; (b) operating the actual studio camera.

5.4.3 Experimental results

The descriptive statistics of the similarity metrics are shown in Table 1. To explore whether there exist differences between the virtual and real systems in both Phase 1 and 2 for users, a 2 (phase: previsualization or formal shooting) ×2 (group: virtual or real) repeated measures analysis of variance (ANOVA) was performed on the three image similarity metrics PSE, PSNR, SSIM, and the task completion time. The independent variables included phases and groups; the phase served as an intra-subject factor with two levels: previsualization phase and formal shooting phase, while the group functioned as an inter-subject factor with two levels: virtual previsualization group and real previsualization group. The ANOVA results are presented in Table 2, indicating a significant main effect of the experimental phase, thus confirming the utility of rehearsal. However, the main effect of the experimental group was not consistently significant. Given the presence of an interaction effect, further simple effects analyses were conducted, revealing no significant differences between the real and virtual groups in Phase 2 for any of the three metrics. This demonstrates that Previs-Real exhibits the same level of effectiveness in rehearsal

Table 1 Descriptive statistics of the similarity metrics in Phase 2 (*M*±*SD*)

Metrics	Control Group	Experimental Group
MSE	84.99±2.30	85.39±1.61
PSNR	29.01±1.38	28.92±0.95
SSIM	0.62±0.03	0.64±0.02
Completion time/s	48.9±9.33	42.4±16.7

Table 2 ANOVA results for MSE, PSNR, SSIM

Metrics	MSE		PSNR		SSIM	
	$F_{(1,7)}$	η_p^2	$F_{(1,7)}$	η_p^2	$F_{(1,7)}$	η_p^2
Phase	17.47**	0.71	17.45**	0.71	1 100.94***	0.99
Group	2.95	0.30	4.76	0.41	53.51***	0.88
Phase×Group	9.04*	0.56	10.92*	0.61	1 342.82***	0.99
Simple effect	*MD* = 0.40, *p*=0.98		*MD* = 0.08, *p*=0.92		*MD* = 0.03, *p*=0.18	

* indicates $p < 0.05$, ** indicates $p < 0.01$, *** indicates $p < 0.001$.

as the actual system.

Statistical analyses of completion time per task did not reveal significant main effects for experimental phase and group, nor a significant interaction effect for phase×group (*p*-values>0.05). This indicates that there were no discernible differences in the time spent by users across different groups during each task in Phase 2, thereby confirming that Previs-Real exhibits a rehearsal effect comparable to the real system again.

5.5 Task 2: Camera movement shooting rehearsal experiment

5.5.1 Experimental stimulus

In Task 2, 24 real camera-captured clips of 10−20 seconds in length were used as both stimulus materials and standard frames for evaluation. These clips were pre-recorded in the laboratory by professional studio technicians and loaded into the system. Subsequently, we recorded the virtual counterparts of these clips for the virtual previsualization group.

Each video clip was generated with three still key frames, which were arbitrarily selected from A, B, C, and D (corresponding to A, B, C, and E in Task 1) on the right side of the task evaluation interface as illustrated in Figure 8b. These frames were then randomly ordered to form the key frame sequences of the video clip in a continuous transition, serving as "start frame-middle frame-end frame". To minimize the influence of extraneous variables, we equilibrated the frequency and sequential order of keyframes. So that, we ensured that each video task is a randomized three-keyframe sequence, and the start frame of each subsequent video matched the end frame of the preceding one to mitigate potential uncertainties arising from redundant adjustment operations. Each user's set of 24 video clips was distinct, with 12 stimulus clips allocated to Phase 1 and the remaining 12 to Phase 2.

5.5.2 Experimental procedure

The experimental task involves replicating 24 standard real shot clips by panning, tilting, and zooming operations with a stationary camera. Following the playback of each stimulus video clip, participants are required to initiate recording by stepping on the foot pedal. Upon completion of their shooting operation, participants must step on the pedal again to cease recording. A 2-second blank screen precedes the commencement of the next task. Subsequent to the completion of each clip shooting, the software will automatically pair corresponding real or virtual standard videos for calculating video similarity. The comparison is also based on predetermined options for real (day or night) and virtual conditions on the evaluation interface.

Additionally, completion time will be recorded. Task completion time refers to the duration from presentation of the stimulus clip until when participants activate the pedal trigger after finishing recording tasks. At the conclusion of each video clip, the playback area transitions to a solid gray background. Participants are then required to complete rehearsal video recordings based on their memory of stimulus videos and keyframe sequence prompts displayed above the playback area. There is a slight interval between Phases 1 and 2 as well.

5.5.3 Experimental results

As in Section 5.4.3, a 2 (phase: previsualization or formal shooting) × 2 (group: virtual or real) repeated-measures ANOVA was performed on PIH as described in Section 4.3. The statistical analyses revealed a significant main effect of previsualization on PIH ($F_{(1, 16)} = 6.02$, $p < 0.05$, $\eta_p^2 = 0.27$), indicating that the PIH score of the virtual previsualization group was significantly higher than that of the real previsualization group (Figure 11). No significant effects from phases ($F_{(1, 16)} = 1.30$, $p = 0.27$) or interaction ($F_{(1, 16)} = 0.01$, $p = 0.927$) were observed on the PIH score.

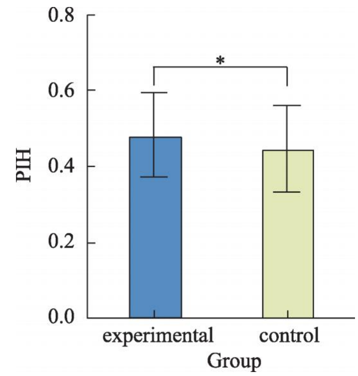


Figure 11 The result of ANOVA indicates that the experimental group outperformed the control group.

5.6 Subjective scale evaluation of Previs-Real

5.6.1 Realism experience evaluation

The immersion of participants using Previs-Real is measured through the IPQ scale, which consists of three-dimensional components (refer to Appendix B for details): spatial presence (SP), experienced realism (REAL), and involvement (INV). The scores of the participants in the three dimensions are shown in Table 3. The results indicate that the participants performed well in the SP dimension, and their experience was consistent with a sense of spatial immersion (a score approaching and exceeding 1), suggesting that the level of spatial perception induced by the virtual scene enabled them to feel present within the virtual space visually and bodily. The results from the REAL dimension indicate the challenge in assessing whether participants perceived a high level of realism, implying a need for improvement in the artificial scene presented in the system to mitigate discrepancies from its real-world counterpart. This may be partly attributed to insufficient refinement in laboratory modeling. The underperformance in the INV dimension indicates that, with regard to interactive perception, users' attention is not effectively focused on the events designed within the virtual scene, resulting in a low degree of engagement. It is hypothesized that this may be attributed to the need for participants to shift their gaze between the system's evaluation interface and standard viewport when completing tasks, in order to view the stimulus materials and virtual previsualization effects, respectively. This divided attention prevented them from maintaining exclusive focus on the virtual environment for prolonged periods, leading to reduced involvement. In future research, potential alternatives such as directorial voice commands and markers within the scene could be explored as instructions for shooting operations instead of standard visual materials.

Table 3 The scores of IPQ

Observation indicators	N	Mean	Variance
SP	25	0.84	0.72
REAL	25	0.16	0.24
INV	25	−0.2	0.99

5.6.2 Usability evaluation

The third version of PSSUQ was utilized for measurement. Table 4 presents the descriptive statistical findings of participants' scores across multiple metrics, including overall usability, system usefulness, information quality, and

Table 4 The scores of PSSUQ

Observation indicators	Reference score	Score of system	Result
Overall usability	3.02	5.86	high
System usefulness	3.02	6.08	high
Information quality	3.24	5.60	high
Interface quality	2.71	5.92	high

interface quality. The results demonstrate that Previs-Real outperforms norm reference scores[40] in all metrics, indicating its exceptional usability.

Furthermore, the SUS was employed to validate the usability of Previs-Real. The descriptive statistical results of SUS scores are presented in the initial two rows of Table 5. The total score is converted into a percentage and averaged, yielding a score of 70.85 (refer to Appendix C for detailed evaluation methodology). The overall usability of Previs-Real complies with established standards, garnering an acceptability rating of "acceptable" and an adjective rating of "good", signifying strong user acceptance. Moreover, the SUS learnability score stands at 5.82 out of 7, indicating that mastering the usage of Previs-Real is highly accessible.

Table 5 The scores of SUS and USE scale

Observation indicators	N	Mean	Variance
Usability	25	4.96	0.17
Easy-to-learn	25	5.82	0.75
Satisfaction	25	5.87	0.77

5.6.3 Satisfaction evaluation

User satisfaction with the utilization of Previs-Real was assessed using a USE scale. As indicated in the third row of Table 5, the findings demonstrate that participants expressed favorable overall satisfaction with the system usage process.

6 Discussion and prospects

6.1 Limitations of the experiments

The results indicate that, whether in the previsualization or the formal shooting phase, the similarity of frames captured by users utilizing the Previs-Real system for rehearsal is comparable to those consistently using the real shooting system. This preliminary finding suggests the effectiveness of Previs-Real to a certain extent. However, it is important to note that our experimental results regarding visual content similarity do not conclusively establish the superiority of Previs-Real over a real system. This is due to the simplified studio setup in our laboratory for strict experimental control, which lacks complexity compared to an actual news studio environment. Furthermore, since participants were only engaged in experimental tasks, subpar performance did not lead to direct consequences, unlike in news programs where low tolerance for errors can cause significant mental stress among staff. Conversely, it is crucial to recognize that these limitations are inherent in the experimental scene and conditions rather than deficiencies of Previs-Real itself. In our subsequent experiments, we aim to better replicate physical and psychological settings resembling real studio scenarios.

In addition, upon checking the results of similarity metrics, we observed that while the overall trend indicated consistency for the camera movement tasks in Section 5.5.3, there were instances of significant computational errors when the camera movement routes and operational durations deviated substantially from the standard samples. Therefore, further exploration of refined directorial and computational methods is necessary to improve the robustness of feedback results.

6.2 Post-experimental interviews

After the user experiments, we conducted interviews with three of the participants in virtual group to gather their feedback regarding their experience using the virtual system compared to the real shooting system. All three participants unanimously agreed that the virtual Previs-Real system was more flexible and easier to use. They noted that the real system's controller was noticeably bulkier (One of them figured that a bulkier controller for Previs-Real can provide a sense of operation more closely resembling that of a real camera, which is valuable input for future iterations). They also mentioned that Previs-Real's interface was clearer, allowing them to focus more on the experimental task. In contrast, while using the real system, they found that other information and functions on the camera body and screen could easily cause distraction or mis-touching, making the learning curve steeper for novices. This feedback provides some evidence for the comparable and even superior performance of the virtual previsualization group over the real shooting group in the experiments.

6.3 Advantages of our hardware controller

The current virtual previsualization technology focuses more on enhancing visual effects and less on the interaction interface, particularly the design of the controller to preserve the original operational habits of studio cameramen. Previously, only a universal gamepad was available as an interaction medium for virtual previsualization. However, using a universal gamepad as the physical interactive interface for the news studio previsualization system may not provide enough potentiometers and switches to cover the operating sets of the virtual camera. While some joysticks designed for specific functions, such as flight gamepads, can meet the basic component requirements, but there exists a significant disparity in shape, structure, and operational feel when compared to the studio camera servo lens controller, resulting in a higher learning cost for the cameraman's operation.

Therefore, we developed the specialized controller. It should be admitted that the controller does exhibit certain disparities in construction when compared to the real camera servo lens controllers. This is primarily attributed to cost considerations. While it meets essential operational requirements, it does not precisely mimic the appearance of specialized real camera servo lens controllers due to its utilization of easily accessible and standard components (which needs to be counterbalanced by the weight simulation expectation outlined in Section 6.2). Additionally, in scenarios where tripods are not utilized, such as in VR applications, the controller can be employed in a lightweight configuration and effectively function as a conventional gamepad.

6.4 Lightweight features of Previs-Real

As described in Section 4.1.1, Previs-Real enables the cameraman to independently executing all supporting processes for camera shooting previsualization. This is because the interface has modularly integrated the functions of a real system, made common assets reusable, and automated the standardized workflows. Although this results in the interface appearance and operational pipeline differing from the actual situation, the subjective evaluation results indicate that the interface functions are easy to learn and have been well-received by users. Besides, information obtained from post-experimental interviews also suggests that the simplicity of Previs-Real can reduce visual cognitive burden and shield users from being overwhelmed by excessive information unrelated to their specific tasks, making the system novice-friendly. Hence, Previs-Real effectively replicates the functions achieved through intricate equipment interconnections and team coordination in a real studio system, in a straightforward and self-supporting manner.

Similarly, the virtual scene models deployed in the system are not infinitely detailed but strike a balance between operational efficiency and realistic viewing. Nevertheless, the dimensions and geometric relationships of the scene must be accurate, as cameramen often rely on visual elements within the studio's layout and decoration as markers and boundary references. For Previs-Real, the high geometric accuracy of the virtual scene combined with precise virtual camera calibration ensures alignment between the virtual and real framing, while UE's rendering capabilities and performance optimization methods for diverse materials and textures contribute to the lifelike appearance of the scene.

For virtual rehearsal tasks equipped with adequate hard disk space and a minimum of 32MB RAM preloaded with the necessary assets, the system runs in real time, with processing duration primarily determined by the ideation of shooting schemes and the length of virtual shots. This duration is not comparable to that required for live collaboration within a conventional studio environment.

Collectively, these findings support the notion that Previs-Real objectively achieves a balance between effectiveness and lightweight design. The basis for this point can also be found in the theory of mental simulation, where active cognitive construction can occur through pre-existing cognition when information is lacking[41]. In conclusion, Previs-Real effectively encapsulates the fundamental aspects of a news studio system, facilitating cost-effective and efficient construction of a virtual previsualization system for camera rehearsal while upholding fidelity.

6.5 Potential applications of Previs-Real

In this study, we exclusively conducted user experiments focusing on the fundamental framing and camera movement rehearsals of a stationary camera. For indoor shooting with a stationary camera, the subject is typically fixed at an appropriate focusing distance, and the follow-focus operation—often challenging for novices—was not included in this investigation. Our future work will encompass more comprehensive system function verification. It is anticipated that the Previs-Real system will cater to all categories of technicians in news studios.

Additionally, Previs-Real could find application in the field of film and television technology education. Similar to TV program previsualizations, conventional teaching methods in this domain encounter challenges related to high venue and equipment costs, potentially limiting their effectiveness. Previs-Real holds promise for enhancing teaching practices and is anticipated to address these limitations. Consequently, our future research will involve a comprehensive assessment of the instructional efficacy facilitated by our interaction system to enhance classroom efficiency and learning experiences.

7 Conclusion

The traditional news studio production with intricate and fast-paced nature often relies heavily on physical locations, equipment, and manpower, which incurs substantial time and cost, and thus presents challenges for on-set camera rehearsals. Previous virtual rehearsal approaches have also encountered difficulties related to fidelity, hardware interaction, and rehearsal feedback. The primary research question addressed in this paper is whether the utilization of virtual previsualization technology and interactive design methodologies can effectively facilitate the development of a virtual previsualization system to respond these challenges. Following the UCD philosophy, we have identified and implemented key features essential for the virtual news previsualization system through the formative interviews with professional studio technicians, including the fidelity demonstrated in the appearance of virtual assets and the functionalities of the software interface (P1: S1, S2, S6, S7), the hardware controller's operating sets and control modes that replicate actual scenarios (P2: S4, S5), and several auxiliary conveniences explored through engine development and computational methodologies (P1: S3, P3: S8). These key elements guarantee that virtual previsualization processes performed through our Previs-Real system effectively align with the real industry workflow, despite being detached from the physical environment. As a bonus, we objectively achieve a lightweight system design. For the natural consideration of development efficiency and cost feasibility, Previs-Real virtualizes numerous devices from the actual broadcasting system while integrating complex pipelines into the software modules, thereby automating and simplifying the overall workflow. Although the studio environment parameter settings and online production functions may not be identical to those of the real system, this actually emphasizes the central importance of Previs-Real's core function—camera shooting previsualization and rehearsal—with all other functions serving to optimize its efficiency.

We conducted framing and camera movement user experiments utilizing the system's feedback module and a subjective assessment. The experimental results confirm Previs-Real's effectiveness in assisting cameramen during previsualizations. Additionally, questionnaire scores related to immersive experience, usability, and user satisfaction support the fidelity and hardware controller interaction experience of Previs-Real. These findings lead us to believe that this design process is equally applicable for developing lightweight simulation systems in other fields—a research area of increasing industrial significance yet not fully explored.

Declaration of competing interest

We declare that we have no conflict of interest.

CRediT authorship contributions statement

Che Qu: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing-original draft, Writing-review & editing; **Shaocong Wang:** Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing-original draft; **Chao Zhou:** Data curation, Formal analysis, Investigation, Methodology, Validation, Writing-original draft; **Tongchen Zhao:** Conceptualization, Investigation, Methodology, Software, Visualization; **Rui Guo:** Conceptualization, Investigation, Project administration, Resources, Supervision; **Cheng Wa Wong:** Data curation, Investigation; **Chi Deng:** Conceptualization, Investigation, Methodology, Project administration, Resources, Software; **Bin Ji:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision; **Yuhui Wen:** Investigation, Methodology, Project administration, Software, Supervision; **Yuanchun Shi:** Resources, Supervision, Writing-review & editing; **Yong-Jin Liu:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing-review & editing.

References

1 Thomas G. Virtual graphics for broadcast production. In: Advances in Computers. Amsterdam: Elsevier, 2011, 165–216
DOI: 10.1016/b978-0-12-385512-1.00005-0

2 Bouyer G, Chellali A, Lécuyer A. Inducing self-motion sensations in driving simulators using force-feedback and haptic motion. In: 2017 IEEE Virtual Reality (VR). Los Angeles, CA, USA, IEEE, 2017, 84–90
DOI: 10.1109/vr.2017.7892234

3 Costello D M, Huntington I, Burke G, Farrugia B, O'Connor A J, Costello A J, Thomas B C, Dundee P, Ghazi A, Corcoran N. A review of simulation training and new 3D computer-generated synthetic organs for robotic surgery education. Journal of Robotic Surgery, 2022, 16 (4): 749–763
DOI: 10.1007/s11701-021-01302-8

4 Mukhopadhyay A, Reddy G R, Saluja K S, Ghosh S, Peña-Rios A, Gopal G, Biswas P. Virtual-reality-based digital twin of office spaces with social distance measurement feature. Virtual Reality & Intelligent Hardware, 2022, 4(1): 55–75
DOI: 10.1016/j.vrih.2022.01.004

5 Braun P, Grafelmann M, Gill F, Stolz H, Hinckeldeyn J, Lange A K. Virtual reality for immersive multi-user firefighter-training scenarios. Virtual Reality & Intelligent Hardware, 2022, 4(5): 406–417
DOI: 10.1016/j.vrih.2022.08.006

6 Yang L R, Feng Z Q, Meng J H. Design of a new kind of chemical experiment container with virtual reality fusion. Virtual Reality & Intelligent Hardware, 2023, 5(4): 317–337
DOI: 10.1016/j.vrih.2022.07.008

7 Deng C. Research and realization of virtual simulation pre-visualization system for news studio. Advanced Television Engineering, 2021 (7): 88–93 (in Chinese)

8 Muender T, Volkmar G, Wenig D, Malaka R. Analysis of previsualization tasks for animation, film and theater. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. Glasgow Scotland Uk, ACM, 2019, 1–6
DOI: 10.1145/3290607.3312953

9 Alyousify A L, Mstafa R J. AR-assisted children book for smart teaching and learning of Turkish alphabets. Virtual Reality & Intelligent Hardware, 2022, 4(3): 263–277
DOI: 10.1016/j.vrih.2022.05.002

10 Zhao Y X, An X W, Sun N L. Virtual simulation experiment of the design and manufacture of a beer bottle-defect detection system. Virtual Reality & Intelligent Hardware, 2020, 2(4): 354–367
DOI: 10.1016/j.vrih.2020.07.002

11 Smith A R, Blinn J F. Blue screen matting. In: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques. ACM, 1996, 259–268
DOI: 10.1145/237170.237263

12 Gibbs S, Arapis C, Breiteneder C, Lalioti V, Mostafawy S, Speier J. Virtual studios: an overview. IEEE MultiMedia, 1998, 5(1): 18–35
DOI: 10.1109/93.664740

13 De Gaspari T, Sementille A C, Vielmas D Z, Aguilar I A, Marar J F. ARSTUDIO: a virtual studio system with augmented reality features. In: Proceedings of the 13th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry. Shenzhen China, ACM, 2014, 17–25
DOI: 10.1145/2670473.2670491

14 Aguilar I A, Sementille A C, Sanches S R R. ARStudio. Multimedia Tools and Applications, 2019, 78(23): 33899–33920
DOI: 10.1007/s11042-019-08064-4

15 Jiao Y N, Liu G S. Design and implementation of real-time interactive pre-visualization for motion capture characters and actors. Advanced Motion Picture Technology, 2019(7): 13–18 (in Chinese)
DOI: 10.3969/j.issn.1673-3215.2019.07.002

16 Rahbar K, Pourreza H R. Inside looking out camera pose estimation for virtual studio. Graphical Models, 2008, 70(4): 57–75
DOI: 10.1016/j.gmod.2008.01.001

17 Yang P, Wu W Y, Moniri M, Chibelushi C C. A sensor-based SLAM algorithm for camera tracking in virtual studio. International Journal of Automation and Computing, 2008, 5(2): 152–162
DOI: 10.1007/s11633-008-0152-6

18 Feng J F. Vizrt+Kinect: an exploration of interactive online packaging. Advanced Television Engineering, 2015(8): 110–114 (in Chinese)

19 Wang J Y, Juhlin O, Johnson E C B. Previsualization with computer animation (previs): communicating research to interaction design practice. In: Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design. Sydney New South Wales Australia, ACM, 2014, 11–20
DOI: 10.1145/2686612.2686616

20 Choi H K, Cho S H. Development of effective pre-visualization authoring tool using conversion technologies—based on film storyboard application. Cluster Computing, 2014, 17(2): 585–591
DOI: 10.1007/s10586-013-0303-6

21 Henrikson R, De Araujo B, Chevalier F, Singh K, Balakrishnan R. Storeoboard: sketching stereoscopic storyboards. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. San Jose California USA, ACM, 2016, 4587–4598
DOI: 10.1145/2858036.2858079

22 Lin I S, Li T Y, Galvane Q, Christie M. Design and evaluation of multiple role-playing in a virtual film set. In: Proceedings of the 16th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry. Tokyo Japan, ACM, 2018, 1–4
DOI: 10.1145/3284398.3284424

23 Muender T, Fröhlich T, Malaka R. Empowering creative people: virtual reality for previsualization. In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. Montreal QC Canada, ACM, 2018, 1–6
DOI: 10.1145/3170427.3188612

24 Ardal D, Alexandersson S, Lempert M, Abelho Pereira A T. A collaborative previsualization tool for filmmaking in virtual reality. In: European Conference on Visual Media Production. London United Kingdom, ACM, 2019, 1–10
DOI: 10.1145/3359998.3369404

25 Muender T, Reinschluessel A V, Drewes S, Wenig D, Döring T, Malaka R. Does it feel real? : using tangibles with different fidelities to build and explore scenes in virtual reality. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. Glasgow Scotland Uk, ACM, 2019, 1–12
DOI: 10.1145/3290605.3300903

26 Volkmar G, Muender T, Wenig D, Malaka R. Evaluation of natural user interfaces in the creative industries. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. Honolulu HI USA, ACM, 2020, 1–8
DOI: 10.1145/3334480.3375201

27 Ivanov A, Au Yeung T, Blair K, Danyluk K, Freeman G, Friedel M, Hull C, Hung M Y S, Pratte S, Willett W. One week in the future: previs design futuring for HCI research. In: CHI Conference on Human Factors in Computing Systems. New Orleans LA USA, ACM, 2022, 1–15
DOI: 10.1145/3491102.3517584

28 Chen L Q, Xiao S H, Chen Y N, Song Y X, Wu R Y, Sun L Y. ChatScratch: an AI-augmented system toward autonomous visual programming learning for children aged 6-12. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. Honolulu HI USA, ACM, 2024, 1–19

DOI: 10.1145/3613904.3642229

29 Ren J, Yu C, Weng Y T, Zhou C C, Shi Y C. Design and evaluation of window management operations in AR headset+smartphone interface. Virtual Reality & Intelligent Hardware, 2022, 4(2): 115–131
DOI: 10.1016/j.vrih.2021.12.002

30 Liu W. Designing generation Y interactions: The case of YPhone. Virtual Reality & Intelligent Hardware, 2022, 4(2): 132–152
DOI: 10.1016/j.vrih.2021.12.005

31 Diaz R, Prieto J, Pardo J, Zambrano C, Uribe-Quevedo A, Godoy E, Perez-Gutierrez B. Development of a first person shooter game controller. In: 2015 IEEE Games Entertainment Media Conference (GEM). Toronto, ON, Canada, IEEE, 2015, 1–2
DOI: 10.1109/gem.2015.7377253

32 Alves G F, Souza E V, Trevisan D G, Montenegro A A, de Castro Salgado L C, Clua E W G. Applying design thinking for prototyping a game controller. In: Clua E, Roque L, Lugmayr A, Tuomi P, eds. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018, 16–27
DOI: 10.1007/978-3-319-99426-0_2

33 Ali M, Cardona-Rivera R E. Comparing gamepad and naturally-mapped controller effects on perceived virtual reality experiences. In: ACM Symposium on Applied Perception 2020. Virtual EventUSA, ACM, 2020, 1–10
DOI: 10.1145/3385955.3407923

34 Borsum F, Pascher M, Auda J, Schneegass S, Lux G, Gerken J. Stay on course in VR: comparing the precision of movement between gamepad, armswinger, and treadmill. In: Proceedings of Mensch Und Computer 2021 MuC '21. NewYork, NY, USA, ACM, 2021, 354–365
DOI:10.1145/3473856.3473880

35 Wang J, Zhang J, Xu Q T. Research on 3D laser scanning technology based on point cloud data acquisition. In: 2014 International Conference on Audio, Language and Image Processing. Shanghai, China, IEEE, 2014, 631–634
DOI: 10.1109/icalip.2014.7009871

36 Hu Y B, Ye S, Zhao W, Lin M, He Y Z, Wen Y H, He Y, Liu Y J. O^2-recon: completing 3D reconstruction of occluded objects in the scene with a pre-trained 2D diffusion model. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(3): 2285–2293
DOI: 10.1609/aaai.v38i3.28002

37 Zhang Z. A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11): 1330–1334
DOI: 10.1109/34.888718

38 Wang Z, Bovik A C, Sheikh H R, Simoncelli E P. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 2004, 13(4): 600–612
DOI: 10.1109/tip.2003.819861

39 Monga V, Evans B L. Perceptual image hashing *via* feature points: performance evaluation and tradeoffs. IEEE Transactions on Image Processing, 2006, 15(11): 3452–3465
DOI: 10.1109/tip.2006.881948

40 Sauro J, Lewis J R. Quantifying the user experience: practical statistics for user research. Amsterdam: Elsevier/Morgan Kaufmann, 2012

41 Taylor S E, Schneider S K. Coping and the simulation of events. Social Cognition, 1989, 7(2): 174–194
DOI: 10.1521/soco.1989.7.2.174

Appendix A: Principles for calculating image similarity metrics

The formula for MSE is

$$\mathrm{MSE}(F,\ f)=\frac{1}{nm}\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}\left[F(i,j)-f(i,j)\right]^2, \tag{A1}$$

where the functions $F(i, j)$ and $f(i, j)$ denote the pixel values of two frames (virtually or real) captured by the user and pre-stored as the detection standard, respectively. The variables n and m represent the dimensions of the frame in terms of pixel width and height. The MSE is computed as the mean of the squared differences between corresponding pixels in the user-captured image and the standard image (for color images, three-channel averages are calculated). A smaller MSE value indicates a closer match between the user-captured image and the standard image.

The PSNR is a widely used method for evaluating image quality, which quantifies the extent to which peak details have been reproduced based on MSE, using the following formula:

$$\mathrm{PSNR}=10\log_{10}\frac{\mathrm{MAX}_1^2}{\mathrm{MSE}}. \tag{A2}$$

In Equation (A2), MAX_I represents the maximum pixel value of an image, i. e., when the image is represented by n-bit binary numbers, $MAX_I = 2^n - 1$. A higher value obtained through the PSNR evaluation method indicates a smaller disparity between the user-recorded image and the standard image.

The SSIM assesses image similarity by considering brightness, contrast, and structure. It formulates the similarity measures for brightness, contrast, and structure by computing the mean and standard deviation of pixel values in each 11×11 pixel block (using a Gaussian weighting function with $\sigma = 1.5$) between two images. The three types of similarity are then multiplied together with an exponent ratio (all set to 1 in this case) to obtain the overall structural similarity of the two images. A result value of 1 indicates identical images.

The PIH detection takes video length into account, using the shorter video as the baseline, and measures similarity by computing the Hamming distance between each standard and shot frames (both compressed to 32 × 32 pixels for enhanced computational efficiency), and counts all similar frames to obtain a total similarity value. A value of 1 indicates perfect identity between the two video clips.

Apperdix B: Functional description of the three dimensions of the IPQ

(1) Spatial presence (SP): Assessing the degree of spatial perception evoked by the virtual reality environment.

(2) Experienced realism (REAL): Measuring the extent to which an artificial environment avoids phenomena that deviate from real-world principles. This questionnaire comprised 13 items, with responses rated on a Likert 7-point scale spanning from −3 to 3, and scores were computed as averages.

(3) Involvement (INV): Evaluating whether users authentically directed their attention to events within the virtual scene, as opposed to passively observing the world from an external standpoint.

Apperdix C: Methodology for evaluating SUS scores

The relationship between acceptability ranges, grade scale, adjective ratings, and SUS score, as proposed by Bangor et al., is depicted in Figure 4 of their paper[C1].

Reference

C1 Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. Journal of Usability Studies, 2009, 4(3): 114−123