Content format and quality of experience in virtual reality

Henrique Galvan Debarba[1,2] · Mario Montagud[3,4] · Sylvain Chagué[1] ·
Javier Garcia-Lajara Herrero[5,6] · Ignacio Lacosta[6] · Sergi Fernandez Langa[4] ·
Caecilia Charbonnier[1]

Abstract

In this paper, we investigate three forms of virtual reality (VR) content production and consumption. Namely, pre-rendered 360 stereoscopic video, full real-time rendered 3D scenes, and the combination of a real-time rendered 3D environment with a pre-rendered video billboard used to present the central elements of the scene. We discuss the advantages and disadvantages of these content formats and describe the production of a piece of VR cinematic content for the three formats. The cinematic segment presented the interaction between two actors, which the VR user could watch from the virtual room next-door, separated from the action by a one-way mirror. To compare the three content formats, we carried out an experiment with 24 participants. In the experiment, we evaluated the quality of experience, including presence, simulation sickness and the participants' assessment of content quality, for each of the three versions of the cinematic segment. We found that, in the context of our cinematic segment, combining video and 3D content produced the best experience. We discuss our results, including their limitations and the potential applications.

Keywords Content formats · Quality of experience · Virtual reality

1 Introduction

The recent availability of virtual reality (VR) consumer equipment has boosted the demand for content and experiences for these devices, as well as for adapted content production

✉ Henrique Galvan Debarba
  hgd@create.aau.dk

1   Artanim Foundation, Chemin du Grand-Puits 40, 1217 Meyrin, Switzerland

2   Aalborg University Copenhagen, A. C. Meyers Vænge 15, 2450 Copenhagen, Denmark

3   Universitat de València, Av. de Blasco Ibáñez 13, 46010 Valencia, Spain

4   Fundació i2CAT, Carrer del Gran Capità 2, 08034 Barcelona, Spain

5   Overlat, Calle Merce Rodoreda 2, San Sebastian de los Reyes, 28702 Madrid, Spain

6   The Mo (The Modern Cultural Productions), Espoz y Mina N4, 50003 Zaragoza, Spain

technology, processes and techniques. A range of new and affordable capture hardware, such as 360 degrees (or omnidirectional) cameras and motion tracking equipment (e.g. suits with inertial sensors), became available in the past few years. In fact, large content distribution and sharing platforms now support VR content and experiences, which can be produced and released by professional studios as well as by independent producers. Among these, we have stereoscopic 360 video streaming websites, such as Youtube, and VR games distribution services, such as Steam.

To preserve visual consistency and achieve a high degree of visual fidelity, VR content often consists of stereoscopic 360 videos or full 3D simulations. The production pipeline and the possibilities of interaction with each form of content are vastly different and, as a consequence, the adoption of each method comes with trade-offs. For instance, 360 stereo video is commonly used to present content that is captured from a real world environment. Since it consists of video capture, the result is visually accurate and photorealistic. However, stereo 360 video does not support point of view (POV) translation since the spatial content is captured and recorded as a spherical projection from a single position in space at any given time. The lack of POV translation is a core limitation of 360 stereo video since it blocks the exploration of the space and causes visuo-vestibular sensorimotor conflicts, which is one of the main factors leading to discomfort and simulation sickness in VR experiences [14]. We note that current VR headsets support position tracking and, thus, the stereo 360 video content format also sub-utilizes the capabilities of the hardware. On the other hand, full 3D content is normally used for interactive experiences, such as games. It recreates and stores the geometrical information and properties of the content in a 3D format, and produces 2D projections at the time of content consumption, rendering the images in real-time. Thus, full 3D content supports POV translation and does not incur in the same limitations as 360 stereo video. However, real time photorealistic rendering can only be achieved at high monetary and computational costs. Significant effort is required to create and animate 3D content, and to simulate the physical behavior of objects and their interactions with the light in real-time.

Beyond these two widespread VR content formats, this paper explores the use of video billboards, or impostors, to merge dynamic content captured in video with 3D real-time rendered environment information. In 3D applications, such as games, this technique is often used to represent secondary elements or objects that are too far away from the POV to be properly identified, as a means to reduce computational requirements [10]. A common example is the representation of large crowds in sport games, where a sequence of pre-rendered images can be used to replace hundreds of individual background characters. Although the use of billboards in 3D applications is not new, in this paper we rely on billboards to represent a central, instead of peripheral, aspect of the experience. A major consequence is that the incongruent projection of the video, which does not respond correctly to POV movements, can be easily noticed by the user, as studied by Fourquet et al. [7]. Whilst being aware of the aforementioned limitations, our interest is in better understanding how detrimental these artefacts are to the overall quality of experience when compared to stereo 360 video and full 3D versions of the same content, determining if this is a viable content format for VR experiences. More specifically, in this paper we ask the following Research Questions: Is the combination of 3D environments and video billboards compatible with high quality of experience in VR? How does it compare to traditional full video and 3D environments?

To answer these questions, we designed and conducted an experiment to assess the impact of different VR content formats on the quality of experience. As having appropriate use cases and content is key for the evaluation of a novel technology/medium, we produced

a professional VR content episode, with three versions resembling the conditions to be evaluated, namely: a stereo 360 version (Fig. 1a); a full 3D version (Fig. 1b); and a hybrid version that combines 3D environment and video content as flat billboards added to the scene (Fig. 1c). The content consisted of an interrogation scene, where the VR user could watch two characters discussing while sitting in an adjacent room. This physical setup, separating the user from the content, was preferred because it can accommodate the production and use of all three content formats. In this particular scenario, the frame of the window separating the adjacent rooms will mask the discontinuity between the 3D environment and the edges of the video billboard in the combined condition, reducing the visual artifacts of the billboard to the lack of true perspective and depth. Although constrained, we believe that this configuration can be applied to a variety of VR application scenarios in which relevant information is presented within delimited and know spaces. Some examples include e-learning, virtual events and even shared social experiences.

The experiment included 24 participants, and the results showed that, under the specifications and particularities of our content and experimental design, users often concluded that the combination of video billboards and 3D environment offered the best experience. In summary, our contributions consist of practical considerations about the process of producing VR content in different formats, discussing the intrinsic advantages and disadvantages of each format; and the evaluation of how the different content formats affect the overall user experience.

This paper is organized as follows. In Section 2, we describe the three content formats that we investigate in this paper, discuss their pros and cons, and discuss related work. Section 3 provides an overview of the content that we created and of the production process of the three different versions of that content. Section 4 describes the design as well as the results of an experiment comparing the three versions of our content. Section 5 presents a discussion of our main findings. Finally, Section 6 concludes this paper.

2 Content formats for VR

Up to date, a variety of content formats and rendering technologies for VR scenarios can be adopted, each one with different implications. The most common techniques are 360 degrees video (VR360 from now on) and full 3D environments, but hybrid solutions, like 3D
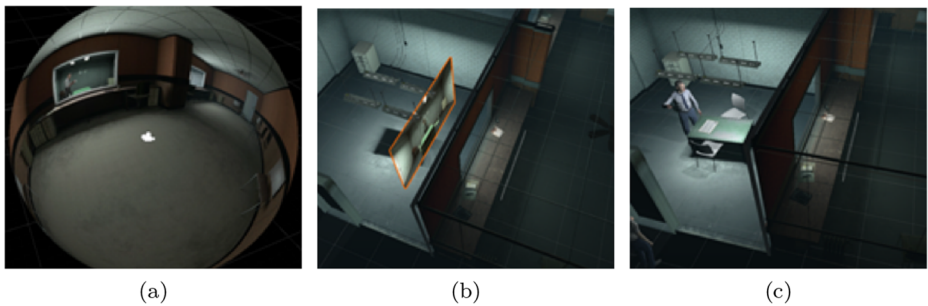

(a)   (b)   (c)

Fig. 1 Overview of the virtual scene in the three different content presentation formats investigated in this paper: **a** stereo VR360 video, with the video of each eye mapped into a sphere that is centered at the virtual camera position of that eye; **b** full 3D environment and cinematic; **c** 3D environment combined with video billboard cinematic, the billboard is drawn behind a window to simulate motion parallax when the user moves

scenarios with inserted video billboards for specific elements of the scene, can also provide satisfactory results. Next, an overview of their a priori main pros and cons is provided, and previous works having adopted these content formats, or comparing are reviewed. The goal of the experiment presented in this work is to confirm, and gain deeper insights about, these assumptions.

2.1 VR360

VR360 videos represent a simple and cheap, yet effective and realistic, way to provide VR experiences. In VR360 videos, a view in every direction is recorded at the same time using an omnidirectional camera or a camera rig that captures overlapping angles simultaneously. The multiple views are then stitched together into a single, high resolution and seamless panoramic video. The camera (rig) represents the center of the omnidirectional scene, and during consumption, the user's viewpoint is also placed at the center of the sphere (see Fig. 2).

Pros of VR360

– VR360 videos are the simplest and cheapest solution in terms of VR content production, when the scenarios to be captured exist and do not need to be created as a model.
– VR360 videos provide high degree of realism, as real scenarios can be captured with high resolution and photographic quality, by using professional cameras, and virtual scenarios can be rendered at very high quality since there are no time constraints. This is especially relevant for dynamic characters.

Cons of VR360

– VR360 videos require a calibration of the employed cameras and the stitching of the captured images. However, existing software tools can effortlessly and successfully provide these features.
– VR360 videos are captured from a single point from where the camera (rig) is physically placed (see Fig. 2). That means that the user's viewpoint is static, matching the camera's position. If the user moves his/her position, then unpleasant parallax effects will soon appear, giving the feeling that the viewpoint also changes and resulting in a perceived deformation of the VR environment. Therefore, POV translation is not supported in VR360 videos, and the point of view has to be defined before recording. Although further cameras at other positions could be used, they could interfere the production and considerably increase the production efforts and costs.
– Beyond the POV translation issues, it should be noted that VR360 videos, even when using stereoscopic recordings, are flat content formats. Therefore, free navigation
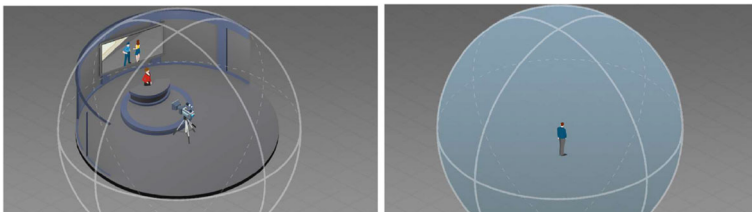

Fig. 2 Capturing and consumption viewpoint in VR360

around the VR environment, which is commonly known as 6 Degrees of Freedom (6DoF), is not supported either.

– In the case that volumetric elements, e.g. users, need to be added in the VR experience, processing and transformation processes are necessary to properly represent them in the VR360 environment. The transformation can be done by applying a 2D mapping of the 3D volumetric data. The volumetric data will be placed in the 3D world and then projected to the 360 sphere where the rest of the video is represented (Fig. 3). This feature is not considered in the presented experiment, but it is important to be kept in mind while deciding on the most appropriate format(s) when producing VR experiences.

VR360 video is often preferred for live broadcasting of events, such as conferences and music concerts. It is cost efficient when it comes to capturing a point of view of real environments, and it generally does not interfere with the execution of the event. In fact, researchers have been working in improving the streaming of VR360 video content over the internet, adapting different strategies for selective transmission of data [16], and assessing the impact of transmission artifacts, such as stalling and bitrate reduction, on the quality of experience of VR360 video [2]. The relative facility of video capture also make the VR360 option interesting for social platforms. For instance, Pece et al. [17] proposed a coherent representation of a meeting room with remote participants by making video inserts of the users in a 360 picture. Their solution stitches the incoming video from each participant into a 360 panorama picture, and anchors them to physical locations represented in the picture. A similar approach is adopted in the Social VR platform proposed by Gunkel et al. [9], where a depth image capture is used to facilitate the real time background extraction and image composition of the users of the system.

2.2 Full 3D

In Full 3D, the whole VR environment, including the characters, is represented in 3D. This content typology is widely used in VR experiences. An example of a Full 3D environment can be seen in Fig. 4, where the building, characters, and end-users are presented in volumetric 3D.

With regard to the content production, the VR environment can be 3D modelled from scratch, and also 3D scanners can be used for existing scenarios. The representation of the characters can be achieved by making use of scanning techniques to create 3D avatars,

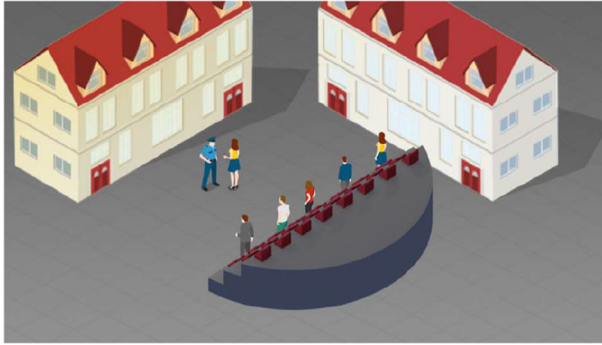**Fig. 3** Projected 3D volumetric elements on a 2D 360 degrees sphere

Fig. 4 Example of a Full 3D environment

which can then be animated using Motion Capture (MoCap) techniques. The use of real-time capturing solutions by using off-the-shelf RGB+D cameras (e.g. Kinect or RealSense sensors) and of volumetric representations as meshes, such as proposed in [1], or point clouds [15] are also possible, but they currently do not provide yet the high resolution that is required for professional and highly immersive VR content.

Pros of Full 3D

– In Full 3D, all elements of the VR environment are three-dimensional. Therefore, the VR environment can be fully explored, supporting POV translation and rotation, without any extra production cost. This also means that the users can freely navigate around the 3D environment, providing 6DoF experiences.
– Full 3D is probably the most immersive content format in terms of geometric reliability and depth estimation.
– Given the absence of video components, Full 3D environments are free from compression artifacts, resulting also in a relatively lightweight option for content distribution.
– Volumetric characters can be seamlessly integrated in Full 3D environments, without needing any specific transformation. In addition, when using the case of pre-rigged 3D characters, the characters can be animated in live scenarios by just sending their data movements, which extremely reduces the transmission and processing load.
– In Full 3D, it is also possible to adapt and amend the cinematic content at a relatively low cost, or even at no extra cost. For example, different animations could be prepared and executed to respond to specific users' actions (e.g., point or gaze at the user, specific answers...).

Cons of Full 3D

– In Full 3D, the main drawback is that it is very challenging and costly — in terms of time and money — to achieve very realistic and natural photorealistic rendering and animations. This is especially true in real-time content and for 3D characters.
– In the case of pre-rigged and animated 3D avatars, a 3D scanner and a MoCap system and room need to be available.
– Meticulous post-production tasks are typically needed to refine the 3D avatars' representations and animations, which have also an impact on the production costs.

– In terms of real-time volumetric users' capturing solutions (e.g. by using meshes or point clouds), they still do not provide the high definition required for professional and realistic VR content, which will impact the users' quality of experience and immersion, especially if used for the 3D actors / characters that integrate the content.

Furthermore, another interesting application where VR360 video and 3D seem to thrive is in remote collaboration. For example, Teo et al. [21] explored the use of live 3D reconstruction and live VR360 stream for remote collaboration in a mixed reality scenario, with an Augmented Reality (AR) user that streamed local information and a VR user that assisted the AR user in a task. A comparison between both content formats actually showed better remote collaboration performance while using the VR360 test condition. However, the advantage seemed to be related to the fact that the VR360 test condition was more competent in conveying the focus of attention of the AR user, which is crucial for carrying collaborative tasks. In addition, the authors conclude that combining both VR360 and 3D reconstruction, so users can alternate between modes, was better overall.

2.3 Hybrid 3D and video billboard

Hybrid 3D and video based solutions for VR content production is also possible (see Fig. 5). The idea consists of validating if an appropriate combination, integration and blending of these content formats can contribute to leveraging the pros of the 3D and video based solutions, while overcoming their cons, at least to a certain extent. This can have an impact on the production costs, but also interestingly on the user experience. Figure 5 shows an example of a 3D scenario, with integrated 3D characters and/or end-users, but also with inserted video billboards for a presenter or instructor captured from a Chroma key room, and a 2D big screen (e.g. displaying TV-related content). In this paper, we refer to this content format as 3D + Billboard.

Pros of 3D + Billboard

– It supports 6DoF for the 3D environment, with high geometry reliability, depth estimation, and without compression artifacts.
– The addition of inserted video billboards can increase the degree of realism, while reducing the production costs for specific, and especially dynamic, scenes.
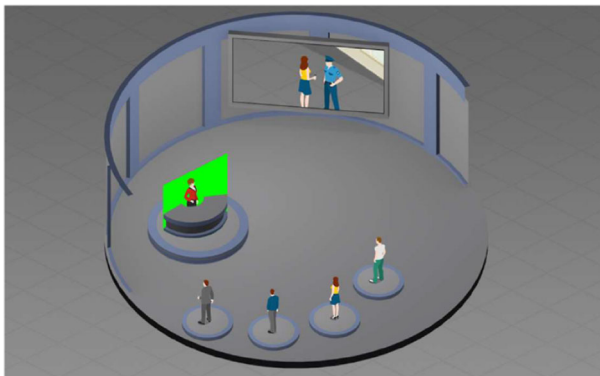

Fig. 5 Example of a 3D environment with a video billboard

– If the video billboards are added at strategic parts of the 3D environment, it can give the feeling that they are an intrinsic part of the three-dimensional environment, and not just an inserted video.

– The addition of inserted video billboards can provide support for POV rotation (the video could always be looking at the user) and limited translation for the 3D environment.

– It can support the addition of volumetric elements without the need to transform it to 2D, unlike in VR360, since these elements can be placed in the three-dimensional part of the VR environment.

Cons of 3D + billboard

– The difference between content formats may be noticeable, which may affect the user experience. Likewise, achieving a seamless integration and blending of heterogeneous content formats may be challenging in specific VR environments.

– POV translation is limited, and it can make the presence of the video billboard evident, resulting in parallax and deformation defects, and in an inconsistent 3D VR environment.

– As for VR360 video, the point of view for the billboard has to be defined before recording, and adding extra point of views considerably increases the recording effort. Likewise, the VR content cannot be easily modified or amended.

In 3D applications, such as games, billboard impostors are used to represent background elements or objects that are too far away from the point of view to be properly identified [10]. They were also popular on early 3D games, when graphics accelerating hardware was not widely available. Video billboards have also been used to approximate 3D reconstruction in video streams, for instance, Hayashi et al. [11] use billboards to represent football players in streams from live sports events as a 3D entity in space, independent of the field plan. The authors take advantage of the stream of video from multiple view sources to produce a billboard per camera point of view, and create a representation that is somewhat consistent from different observation points. More advanced methods [8] explore the problem of reconstructing articulated billboards based on the same kind of video input. It reconstructs football players as an ensemble of video billboards, creating separate billboards for each limb segment of the player.

In contrast to the aforementioned applications, we examine the use of a video billboard to represent central, instead of peripheral, aspects of the content, such as the actions and interactions of the main actors. A major consequence is that, due to the lack of depth, billboards will not respond correctly to translations of the point of view, which are common and expected in VR experiences, making the flatness of the billboard easily noticeable. Whilst being aware of the aforementioned limitations, our interest is in understanding how detrimental these artefacts are to the overall quality of experience when compared to stereo VR360 video and full 3D versions of the same content, determining if this is a viable content format for VR experiences.

2.4 Comparing content formats

Interestingly, there is recent literature comparing Full 3D experiences with VR360 panoramas and videos, but in different applications than the one we present in this paper, as we discuss below.

In [3], a comparison between the visualization of an archaeological site, reproduced in 3D or captured as 360 static pictures, showed an advantage for the 3D experience in terms of presence and fun. However, we note that the implementation of the 360 picture was unusual. The content was presented in a sphere that surrounded users and was stationary in space, while the movement of the virtual camera was allowed. This meant that users could get closer to the surfaces of the sphere as they moved their head, which could cause severe visual distortions of the image, as well as inconsistent representation of scale. VR360 implementations generally assume that the projection sphere is infinitely large, preventing users from ever getting closer to any specific point in its surface. This prevents visual distortions, but also imply that user translation have no effect on the images that are rendered to the VR headset. In this paper we focus on the more typical scenario, where virtual translation is not allowed in VR360, but it is not trivial to understand the impact, either negative or positive, that this could have on the user experience.

In [5], De Simone and colleagues compare the VR360 video social VR system proposed in [9] with Facebook Spaces [6], which implement a 3D abstract user avatar representation combined with a 360 static background image. The authors found a significant advantage for the VR360 video format in terms of social experience, which may be related to the realistic visual appearance of the other participant, in spite of the fact that their eyes were covered by the VR headset in the VR360 video condition. These results seem to relate the visual fidelity of other users to the strength of social response and co-presence.

Moreover, Cai et al. [4] compared the experience of historical heritage represented using passive VR360 video with the experience using interactive Full 3D. The former consisted of a recording of the real location, where an elderly couple were cooking food as part of the experience. The latter consisted of the 3D reconstruction of the real environment, without the presence of the elderly couple, where users could interact with the environment and collect information. Therefore, the two conditions were vastly different in terms of content, the VR360 video. They found that the VR360 video provided stronger sense of reality, most likely due to the audiovisual complexity of the recorded scene, which could provoke a stronger emotional connection. In addition, they have also observed that users were willing to spend more time in the 3D reconstruction experience, which was interactive and had the exploration pace defined by the users. We highlight, however, that the scenarios presented in [4] were not identical across content formats. This makes it hard, if not impossible, to separate the effect of the content format from the effect of the content itself. In contrast to previous work, this paper compares the quality of experience of VR360 video and Full 3D reconstruction for the same piece of content, with the same actors and events.

Finally, we note that there was some previous effort to identify when billboard impostor deformation, as used to replace 3D objects, becomes noticeable to users. This is seen in Hamill et al. [10] and Fourquet et al. [7], where users would try to identify discrepancies in billboard impostor, that replaced 3D models such as inanimate objects, characters and buildings, as seen in a monitor display. However, they did not assess it in a similar setting and application context as we do in this study. Moreover, we are not particularly interested on whether users can notice the use of billboards, but rather on whether, and to what extent, the user experience is affected when video billboards are used.

3 Content production

This section reviews the whole production process for a professional VR content episode, including the pre-production, production and post-production tasks and steps. Note that

the insights from the experiment of this paper will be applied to later evaluate Social VR scenarios as a new communication medium for interaction and communication between remote users [5], such as watching videos together, while apart. Accordingly, this section reports on two different, but connected, scenes about a crime investigation story. These two scenes will be used to evaluate in the future if two remote users feel together when watching related content at the same time, while interacting via audiovisual channels. However, just one of the scenes is used in the experiment presented in this paper.

3.1 Pre-production

Inspired by thriller-like movies, the decision was to create an episode departing from the murder of a celebrity. The plot revolves around the crime investigation, in which two suspects are being interrogated and the two participants are expected to observe the interrogations, playing the role of inspectors.

Iterative design sessions were conducted to assess the most appropriate approach and scene for telling the story and to recreate the shared environment in which two users could virtually "meet". Unlike traditional watching apart together scenarios [5], in which the users watch exactly the same content, it was decided to place the users in a shared observation room, but in front of a different one-way mirror connecting to two separate interrogation rooms (see Figs. 9 and 10). In each of the separate rooms, a different suspect of the same murder is being interrogated by a policer. Therefore, although the users share a common space and can directly see and talk to each other, they can only see and hear one of the two interrogation scenes belonging to the same story.

Based on this concept, the storyboard including the associated spaces, viewpoints and evolution of the story was developed, in order to prepare for the production plan. As an example, Fig. 6 shows the mockups that were generated to recreate the users' viewpoints towards one of the interrogation scenes and the other user.

After the selection of the theme and scenario, the next steps consisted of writing the script, and casting the actors. The story was further developed, revolving around the murder of a fictional celebrity, and the interrogation of a male suspect and a female suspect. The two suspects have a different version about what happened, and the story reflects that they both have things to hide. Details about the casting process to select the actors (two suspects and police inspector) and the two scripts written for each of the interrogations can be found in [18]. The participation of the two actors and the actress were necessary to be able to create the three VR content formats previously introduced:

– *Full 3D* Version: A 3D environment with 3D-riggered characters, combined with MoCap techniques.
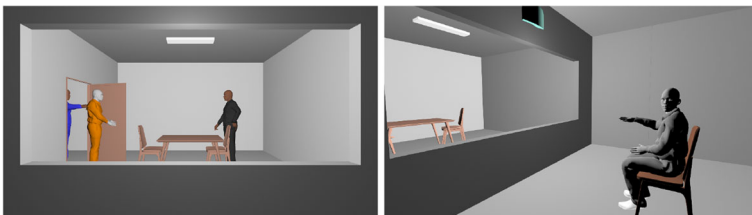

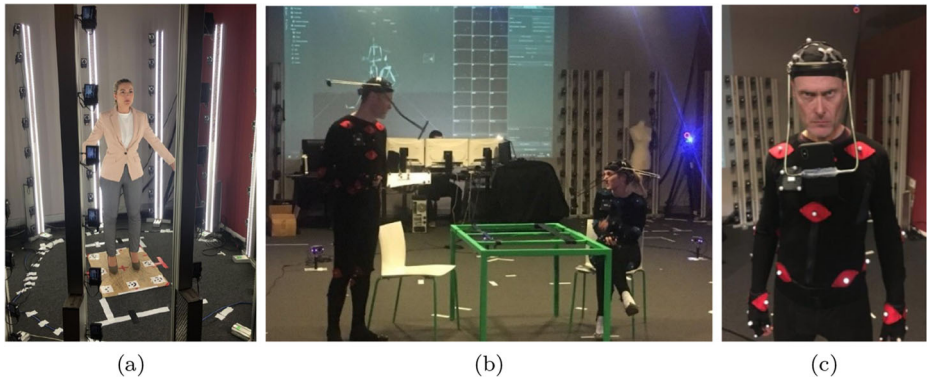**Fig. 6** Users' viewpoint for the produced Social VR scenario

Fig. 7 **a** 3D body model creation, **b** MoCap recording, and **c** facial gestures recording

– *3D + Billboard* Version: A 3D environment with the interrogation scenes captured on video from a Chroma key room, and then rendered as a video billboard within the 3D environment.
– *VR360* Version: The *Full 3D* VR scene rendered as a stereoscopic VR360 video, composing the 3D environment and the masked video of the characters (as the 3D police station environment does not exist in reality).

The experiment presented in this paper makes use of only one of the two interrogation scenes, as its goal was to determine the impact of the content formats on the user experience, and a single person VR experience seemed more appropriate for a controlled experiment on the subject. In total, the VR story presented to the participants in the experiment, using each variant of content formats, has a duration of 6 minutes.

3.2 Production

Next, the processes associated to the production of such content versions are summarized.

3.2.1 MoCap and shooting for the interrogation scenes

In order to create the Full 3D Version, the actors were 3D scanned with a photogrammetric scanner consisting of 96 cameras to obtain the 3D surface of their bodies (see Fig. 7a). The MoCap session was recorded by using a Vicon MXT40S system with 30 cameras (see Fig. 7b), in which each actor wore 59 retro-reflective markers to track their movements. For facial capture, a tool was developed to record the faces by using an iPhone X and the ARKit framework,[1] with the help of an ad-hoc helmet (see Fig. 7c). Then, the captured facial content data were synchronized with the MoCap data, and converted into an appropriate format for further 3D editing and adjustments.

In order to create the 3D + Billboard Version, the same scene, with the actors wearing exactly the same clothing, was shut over a Chroma key room, by using a stereoscopic camera (Canon, with 8-15mm optics sensors, and a separation of 8cm between its lenses). The scene objects, like the table, were also covered in green color. The recording setup can be seen in Fig. 8.

---

[1]https://developer.apple.com/documentation/arkit

Fig. 8 Video shooting over a
Chroma key room

3.2.2 3D environment

The two separate interrogation rooms and the shared space for the two users, together with all associated elements (e.g. chairs, desks, book notes...) were modelled in photorealistic 3D, with the use of optimized geometry, and integrated in a Unity project.

The overall view of the 3D modelled scenario is shown in Fig. 9a. Likewise, Fig. 9b shows the same 3D scenario, but with the rendering 2D video planes for the 3D + Billboard Version, together with the users' positions and the shooting perspective. In both figures, it can also be observed that, for the future Social VR experiment, each user can only see (and only hear) the interrogation scene happening in front of their one-way mirror, but they can see and hear each other through a shared space. The scenes for User 1 were the ones used in the experiment presented in this work.

Realistic lighting conditions were recreated in order to provide a natural integration of the users and characters into the 3D virtual environment, and to provide a thriller-like atmosphere (i.e. direct light in the interrogation rooms and dimmed lights in the dark shared room where the users are placed). Spatial ambient sound was prepared, coming from the direction of the actions. This is the case for the doors opened, actions by the actors, sound from the other user, etc.

An overall view of the recreated 3D environment, resembling a 70s look police station room, is provided in Fig. 10a, while the viewing perspective from one of the users through the shared space in provided in Fig. 10b.
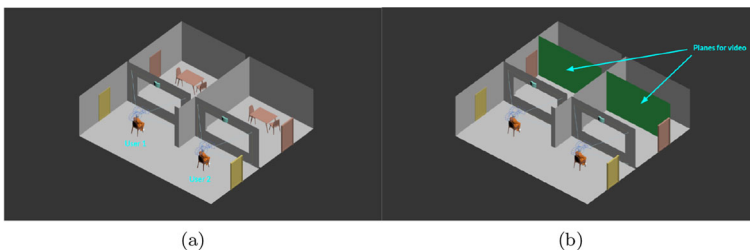

(a)   (b)

Fig. 9 **a** Overall view of the 3D environment to be recreated; **b** 3D scenarios with 2D video billboards for the interrogation scenes
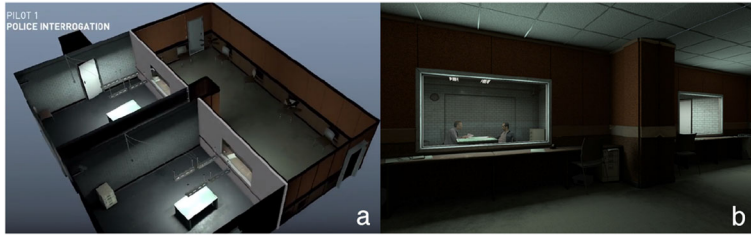
Fig. 10 **a** Overall view of the recreated 3D environment; **b** Users' viewpoint from the 3D shared space

3.3 Post-production

After the recording and modelling of all assets, post-production processes were conducted for all the raw material, including the associated adjustment tasks for an appropriate compositing and seamless blending.

In the case of the 3D + Billboard and VR360 versions, noise reduction and masking processes were initially conducted for the recorded billboards. Masking was especially a time-consuming and laborious process. Figure 11 gives an idea of this process, including the necessary specific treatment for the characters' hair. In addition, color adjustment processes were necessary for an effective removal of the green elements and the replacement with the appropriate color, together with the adjustments to achieve a seamless stereo view.

In the case of the Full 3D Version, the captured 3D surfaces of the actors were initially fit to a template character rig (i.e. humanoid skeleton with joints and bones) used to animate the characters. Then, post-processing techniques were necessary to produce morph targets that comply with the facial capture data and specifications of the ARkit API, to clean the MoCap data (e.g. by resolving occlusions), and to retarget the animations on the character rigs to obtain realistic and natural results.

Screen captures of the final results of the produced VR content scenes, for each described variant, are shown in Fig. 1. All created content assets have been released to the Zenodo open repository [18]. A video describing the created VR content, and summarizing the production process, is available at: https://youtu.be/aHO5M1qNmjY.

3.4 Cost analysis

This subsection provides an estimated analysis of the cost required to produce the assets composing the presented VR story, in each of its variants, in terms of both personnel involvement and duration. Although the costs depend on the involved crew, their expertise
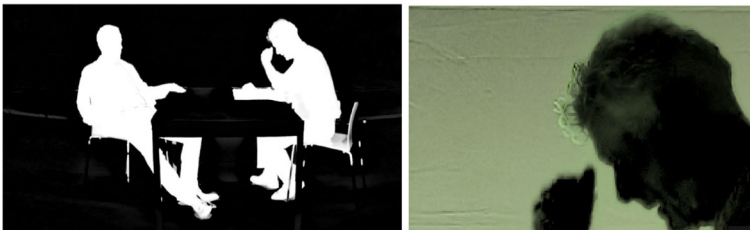

Fig. 11 Examples of the masking process

and skills, and on the used resources (e.g. software, infrastructure), this analysis is meant to provide an approximate idea of the associated requirements and implications for the production of considered modalities for providing appropriate VR experiences. Note that the costs associated to pre-production tasks and to the participation of actors have not been considered in the analysis, as they apply to each considered content format.

Task A: 3D VR Environment (8 weeks, 1500hours, 4 professionals) As the recreated 3D environment does not exist in the reality, it was produced from scratch. This production phase included the 3D modeling of the VR environment and associated elements (chairs, desks, fans...), texturing, illumination and rendering.

Task B: 3D Characters (6 weeks, 800hours, 3 professionals) This phase was required for the integration of the 3D characters in the Full 3D version. It included the bodies scanning, the MoCap sessions for the full performances, including the facial capture, and then all associated post-production and integration tasks. These processes require the availability of a 3D body scanner, a MoCap studio, an iPhone X for the facial capture, and the required software for a seamless integration and refinement.

Task C: Video-based Characters (10 weeks, 1600hours, 4 professionals) Initially, the shootings for the two scenes took two days, and required the availability of a professional (stereoscopic) camera and a Chroma key room. Then, laborious post-production tasks were required for Chroma cleaning, color grading, lighting, compositing and integration. This phase was especially costly for the created scenario due to the dynamism of the actors and to the presence of objects in the surroundings (e.g. tables, chairs) that were replaced by virtual 3D elements to increase the feeling of parallax and immersion.

Task D: Integration in Unity (6 weeks, 700hours, 3 professionals) For all content variants, this phase included the integration of the VR content assets and interaction features in Unity, as well as the required adjustments to provide a refined and smooth experience.

Task E: Direction and Supervision (24 weeks, 200hours, 1 professional) This included all direction and supervision tasks for the complete content production and post-production processes.

In general terms, it can be affirmed that the Hybrid 3D and video billboard version (involving Tasks A+C+D+E) will be cheaper to produce than the Full 3D version (involving Tasks A+B+D+E) most of the times, requiring less costly infrastructure and equipment as well as associated post-production processes. When the objective is to recreate an existing scenario, then the Full VR360 Video Version would be the cheapest content format to produce, because everything can be directly captured with a camera. Likewise, 3D scanners (e.g. a LIDAR) can be used for reconstructing real scenarios in 3D. However, if the scenario does not exist, then the Full VR360 video Version needs to be created by departing from one of the other two versions, by appropriately rendering the VR scenes. This was the case for the presented study.

4 Experiment

This section describes the experiment, including further information on the tested conditions, experiment setup and procedure, instruments used for assessing the experience of the user, and an overview of the results.

4.1 Methodology

4.1.1 Dependent variables

In the experiment, we collected information about simulation sickness, subjective sense of presence, perceived limitations of content format and effectiveness of the experience, comparative post-experiment feedback, and comments on each version of the content condition.

The simulation sickness score was obtained using the Simulator Sickness Questionnaire (SSQ) [13]. The questionnaire was applied before and after each trial (i.e. each participant filled the questionnaire six times), the score for each trial is computed as *score after trial* minus *score before trial*, so that any abnormal state of the participant at the start of a trial is accounted for in the results. For instance, if at the start of a trial the participant presents a symptom that was acquired before the experiment or that has persisted from a previous trial of the experiment, this symptom is recorded and is then subtracted from the SSQ score obtained after the current trial. That is, the symptom is expected to remain through the current trial. Incongruent visuovestibular sensory signals have been identified as a major cause of cybersickness, as discussed in [14]. Therefore, we expect an increase in the total score of simulator sickness symptoms when comparing the pre- and post-exposure SSQ scores in the VR360 condition, where head translation does not result in virtual camera translation and produce a visuovestibular mismatch.

The subjective sense of presence was taken after each trial. It consisted of a modified SUS (Slater, Usoh and Steed) presence questionnaire [22]. The sense of presence is commonly described as the sense of "being there", in the virtual world, or as the feeling of non-technological mediation. That is, when the immersive equipment becomes transparent to the user. The adapted questionnaire is shown in Table 1, answers were provided in a 7 point scale and added to obtain the final presence score. Sensorimotor contingencies have been shown to increase the sense of presence in VR, as discussed in [19]. Therefore, we expect higher sense of presence, as assessed by the SUS questionnaire, in conditions that allow for full control of the point of view, namely Full 3D and 3D + Billboard.

Moreover, we developed a questionnaire to address other four relevant aspects for the user experience, namely the quality of the virtual characters in terms of realistic visual appearance and motion, the visual consistency of the scene (perspective projection and image composition), the feeling of control of the virtual point of view, and the effectiveness of the experience. We refer to these four aspects as *characters quality*, *visual consistency*,

Table 1 Adapted SUS presence questionnaire [22]

Regarding the experimental condition that you have just experienced:
Please rate your sense of being in the virtual environment, on a scale of 1 to 7, where 7 represents your normal experience of being in a place. (Not at all) 1 — 7 (Very much)
To what extent were there times during the experience when the virtual environment was the reality for you? (At no time) 1 — 7 (Almost all the time)
When you think back to the experience, do you think of the virtual environment more as images that you saw or more as somewhere that you visited? (Images that I saw) 1 — 7 (Somewhere that I visited)
During the time of the experience, which was the strongest on the whole, your sense of being in the virtual environment or of being elsewhere? (Being elsewhere) 1 — 7 (Being in the virtual environment)
During the time of your experience, did you often think to yourself that you were actually in the virtual environment? (Not very often) 1 — 7 (Very much so)

Answers were provided in a 7 point scale

viewpoint control, and *effectiveness* for short. The questions are presented in Table 2. Answers were provided in a 5 point scale, ranging from strongly disagree (1) to strongly agree (5).

The questions were formulated taking into consideration the a priori assessment of the advantages and disadvantages of the three content formats, as discussed in Section 2. We identified that the visual appearance and animation of characters could constitute a weakness for the Full 3D condition when compared with the other two experimental conditions (Q1, Q2 and Q3 in Table 2). Similarly, we expected that the visual distortion due to perspective projection errors and the lack of true depth would negatively affect the 3D + Billboard condition (Q4, Q5 and Q6 in Table 2). Finally, we anticipated that the feeling of viewpoint control would be negatively impacted by the lack of POV translation in the VR360 condition (Q7 and Q8 in Table 2), which should also reflect on the aforementioned presence and simulator sickness questionnaires results. The phrasing of these questions were based on the questionnaires used for research in the sense of agency [12].

Finally, the effectiveness of the experience questions (Q9, Q10 and Q11 in Table 2) were designed to assess how successful each content format was in attaining the goal of delivering an involving experience. Unlike the previous questions, these were not pinned to specific a priori observations about the experimental conditions. The response to these questions can be highly subjective, and we expect that they reflect the overall experience of participants.

Finally, we also developed a comparison questionnaire to be applied at the end of the experiment. In this questionnaire, participants had to order the three content presentation conditions from most to least preferred with regard to the same aspects addressed in the post-trial questionnaire above. The comparison questionnaire is presented, together with results, in Fig. 13.

Table 2 Post-trial questionnaire, answers were provided in a 5 point scale ranging from strongly disagree (1) to strongly agree (5)

ID	Question	Aspect of the experience
Q1	The motion of the characters felt expressive.	Quality of characters
Q2	The motion of the characters felt natural.	Quality of characters
Q3	The appearance of the characters is realistic.	Quality of characters
Q4	The perspective of the virtual world and the elements in it looked consistent.	Visual consistency
Q5	The movement of virtual elements, character, objects ... through the space felt consistent with the real world.	Visual consistency
Q6	My movement did not affect the visual consistency of the virtual world and the elements in it.	Visual consistency
Q7	The control of the virtual view felt natural.	Viewpoint control
Q8	When I moved, I expected the virtual view to respond accordingly.	Viewpoint control
Q9	I felt involved in the virtual reality experience.	Effectiveness
Q10	My experiences in the virtual environment seemed consistent with my real world experiences.	Effectiveness
Q11	I felt involved in the interrogation.	Effectiveness

4.1.2 Procedure

In the experiment, participants were asked to read an information sheet and sign an informed consent form. Then, they were asked to fill in a demographics questionnaire asking about their gender, height, age and previous experience with VR and video games, and were presented with an overview of the experiment structure and task. Following the introduction, participants underwent the three experimental trials, one for each content condition. The presentation order of the trials was counterbalanced to control for order effect. Each trial consisted of a pre-trial SSQ questionnaire, the cinematic segment in the current condition, a post-experience SSQ questionnaire, the adapted SUS questionnaire and the post-trial questionnaire.

In preparation for the cinematic segment, participants were positioned sitting in a chair in the center of the capture space and equipped with the Oculus Rift head mounted display. They were also informed that they could stand up during the cinematic content if they wished to do so. This additional degree of control was permitted to help leveraging the advantages of head position tracking in the two conditions where this was allowed (*Full 3D* and *3D + Billboard*). However, participants were asked not to walk. We limited the range of motion to conform to the content scenario (i.e. participants watch the scene from the adjacent room), to prevent accidents in the limited tracking space, and to prevent significant variations of the experience across participants in the experiment.

Lastly, after experiencing the three content conditions, participants were asked to fill in a comparison questionnaire and to provide written feedback on each of the three content condition. The written feedback consisted of listing advantages and disadvantages of each of the three content formats, as assessed by participants themselves.

4.2 Participants

We recruited 24 volunteers, with a total of 16 male and 8 female participants and an average age of 38 years old (standard deviation of 7.6). Five participants were trying a head mounted display for the first time, while nine reported to have worn one a few times in the past, seven reported to wear one every month or week, and three reported to wear one every day. Similarly, five participants reported that this was their first VR experience, while eight had few previous experiences, two often participate in VR experiences, and nine develop a professional competence in the field of VR. The participants were all provided with a description of the experiment and had to sign an informed consent form in order to take part in the experiment.

4.3 Results

4.3.1 Simulation sickness

For each content condition, we tested whether the difference between the SSQ reported after the trial was significantly different than the SSQ reported before the trial. The statistical analysis was carried out using the Wilcoxon signed rank test. We observed a statistically significant increase in the SSQ responses for the VR360 video condition ($p = .002$). The test failed to reject the equality of pre/post trial SSQ responses in the Full 3D and 3D + Billboard conditions, which received similar SSQ scores before and after the trial ($p = .228$ and $p = .671$ respectively). Furthermore, we compared the difference in SSQ scores across content conditions using a Friedman test, results show a statistically significant effect of

condition ($x^2_{(2)} = 7.7$, $p < .022$). Pairwise comparisons with the Wilcoxon signed rank test and Holm-Bonferroni correction show that the VR360 video condition caused significantly more discomfort than the 3D + Billboard condition ($p = .025$). An overview of the pre/post trial differences in SSQ is presented in Fig. 12.

4.3.2 Presence

A Friedman test showed a significant effect of content condition on the subjective presence scores ($x^2_{(2)} = 12.7$, $p < .002$). Pairwise comparisons using the Wilcoxon signed rank test and Holm-Bonferroni correction across the levels of content condition showed that participants reported lower presence in the VR360 video condition than in both the 3D character ($p = .008$) and 3D + Billboard ($p = .005$) conditions. The latter two presented similar presence scores ($p = .749$), as shown in Fig. 12.

4.3.3 Questionnaire and comparison between conditions

For the comparative questionnaire, a chi-squared test was used for each comparison statement to determine whether a statistically significant dependence between the independent variable condition (Full 3D, 3D + Billboard or VR360 video) and the dependent variable classification (1st, 2nd and 3rd – or best, intermediate, worst), as specified by participants, exists. All but one of the tests showed a statistically significant dependence between the two variables (all $p < .001$). The statistical test failed to reject the independence between the variables for the statement "the visual consistency between characters and scenario was more accurate in condition ..." ($p = .112$). A summary of the comparison questionnaire results is presented in Fig. 13.

With regard to the questionnaire, we tested each response variable for statistically significant differences across the levels of content condition for each questions using the Friedman test. Significant results were followed by pairwise comparisons using the Wilcoxon signed
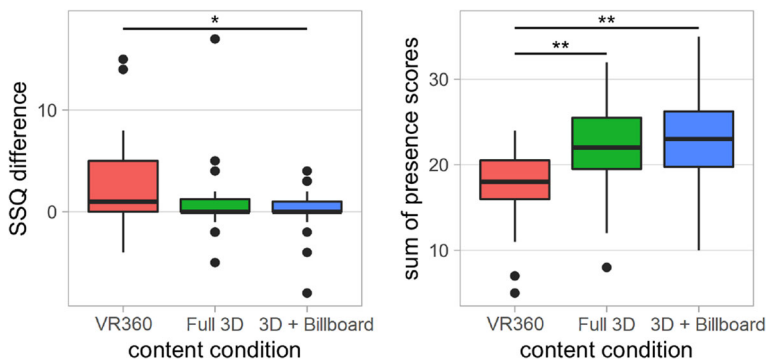

Fig. 12 (Left) Boxplot of the simulation sickness questionnaire (SSQ) score. The score was computed by subtracting the SSQ results obtained before the trial from the SSQ results obtained after the trial. The VR360 video condition presented a statistically significant increase in the SSQ, in addition, its increase was also statistically higher than that of the 3D + Billboard condition. (Right) Boxplot of the subjective sense of presence questionnaire scores. Participants reported significantly higher sense of presence for the Full 3D and 3D + Billboard conditions than for the VR360 video condition. '*' and '**' indicates a significant difference with $p < .05$ and $p < .01$ respectively
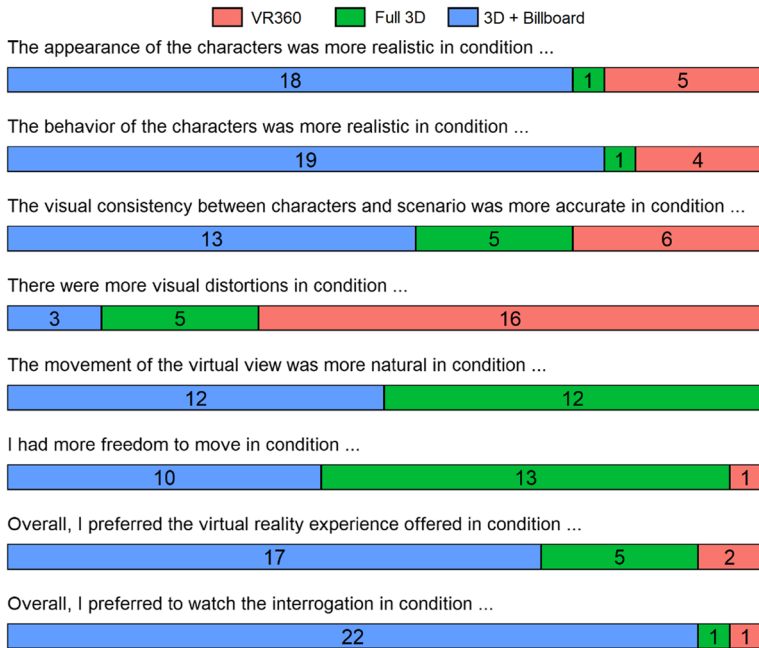
Fig. 13 Comparison questionnaire results. Participants classified each condition from most (1) to least (3) preferred for eight different statements

rank test and Holm-Bonferroni correction. A summary of results is presented in Fig. 14. We summarize the results of the four aspects addressed in this questionnaire below.

The questions concerning expressive motion, natural motion, and realistic appearance of the virtual characters (Q1, Q2 and Q3 respectively) presented similar results, with 3D + Billboard and VR360 characters receiving a higher score than Full 3D (all $p < .01$). In the post-experiment comparative questionnaire, the 3D + Billboard condition was preferred more often when it comes to character appearance and behavior, and was followed by the VR360 condition in second. These results indicate the superiority of recorded video media when it comes to visual representation of character appearance and actions in our scene. Participants have also reported detrimental animation effects in the Full 3D condition, such as unrealistic hand movements and the lack of physical simulation on characters clothing and props.

Considering the perception of correctness of visual perspective (Q4), movement of virtual objects through space (Q5) and visual consistency of scene elements (Q6), the 3D + Billboard condition performed better than the VR360 video condition (all $p < .01$), while Full 3D performed better than VR360 video when we consider the visual perspective and scene elements questions (Q4 and Q6, both $p < .01$), but not in the movement of virtual objects question (Q5, $p = .34$). We did not find a statistically significant difference in Q4, Q5 and Q6 when comparing the 3D + Billboard and Full 3D conditions. The post-experiment comparative questionnaire presented similar results. The VR360 video condition was seen to have more visual distortions, while the Full 3D and 3D + Billboard conditions received similar preference scores.
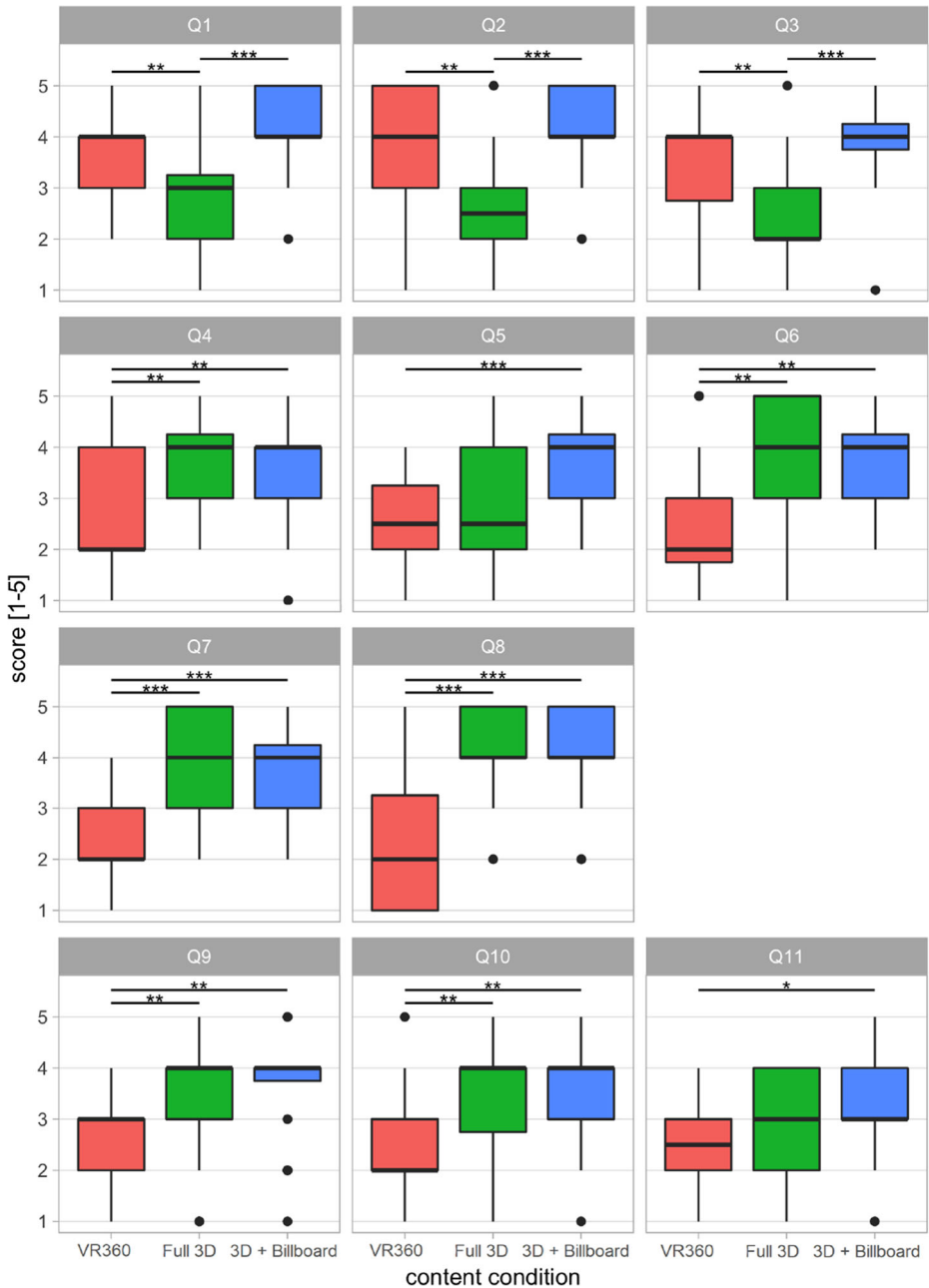
Fig. 14 Boxplots of the post-trial questionnaire. '*', '**' and '***' indicate a statistically significant difference with $p < .05$, $p < .01$ and $p < .001$ respectively

Moreover, the feeling of control of the point of view (Q7 and Q8) was greater in the 3D + Billboard and Full 3D conditions than in the VR360 condition (both $p < .001$), while no significant difference was found between the 3D + Billboard and Full 3D conditions (both

$p > .4$). The comparative questionnaire produced similar results. The movement of the virtual viewpoint was considered the most natural in the Full 3D or 3D + Billboard conditions by nearly the same amount of participants, while VR360 was generally considered inferior. Similar results were also observed for participant's perception of freedom to move, but with a slight advantage for the Full 3D condition over the 3D + Billboard condition. This outcome confirms our assumption, given that the VR360 was the only condition not to allow for POV position updates in the virtual environment. This difference was not observed when comparing the 3D + Billboard and Full 3D conditions.

Lastly, the 3D + Billboard and Full 3D conditions scored higher than the VR360 video condition for the effectiveness of the VR experience and consistency with real word experience (Q9 and Q10, both $p < .01$). Concerning the interrogation experience (Q11), the 3D + Billboard condition ranked higher than the VR360 video condition ($p = .036$), while Full 3D presented a score that could not be differentiated from the other two conditions (both $p = .32$). For all three questions, no statistically significant difference was found between the 3D + Billboard and Full 3D conditions. In the comparative questionnaire, the 3D + Billboard condition was generally preferred for this VR experience. It was then followed by the Full 3D condition, in spite of the perception that the virtual characters appearance and behavior were inferior in this particular condition.

4.3.4 Participants comments

At the end of the experiment, participants were asked to write down the advantages and disadvantages of each of the three content formats, as assessed by themselves. The objective was to identify the most significant and salient characteristics of each experience from the perspective of participants. The feedback was collected as a short statement in a digital form. Observations with equivalent or closely related statements were grouped together. We report on the characteristics that were cited more often for each of the conditions.

The advantages reported for the VR360 video condition were the realism of characters (4 participants) and environment (6 participants) and the consistent quality in terms of character/environment integration (2 participants). However, 7 participants reported that they did not see any advantage in the VR360 video condition. The main disadvantages were the lack of position tracking (12 participants), which led to simulation sickness and is supported by the SSQ results (3 participants), the presence of compression artifacts (6 participants), and decreased immersion (3 participants).

The advantages reported for the Full 3D condition were the consistent visual experience (8 participants), the freedom to move (6 participants), realism of the 3D environment as the adjacent room is not longer contained in an image (4 participants) and increased immersion (4 participants). In addition, 3 participants did not see any advantage for this condition. The most common disadvantages were the characters appearance (15 participants) and acting (12 participants), which did not look as realistic as in the other two conditions.

Finally, the advantages reported for the 3D + Billboard condition were the realism of characters and environment (14 participants), the comfort and/or freedom to move (8 participants) and increased immersion (7 participants). The disadvantages were the visual inconsistency between the 3D environment and the billboard video (4 participants), which was particularly noticeable in the table in the interrogation room, the resolution of the video (2 participants), and the flatness of the billboard (2 participants), which felt like a screen. Notably, 8 participants did not express any disadvantage for the 3D + Billboard condition.

5 Discussion

5.1 Preference toward 3D + Billboard condition

Concerning the question of whether the combination of 3D environments with video content to represent the central elements of the experience (3D + Billboard condition in our study) is compatible with a high quality of experience. Our results demonstrate that, not only this is true, but participants were also more receptive to this specific content format than the other two used for comparison (see comparison results in Fig. 13). The quality of the video characters was pointed as a clear advantage over the Full 3D condition, while the possibility to translate the POV, which results in visuovestibular congruence and motion parallax, was pointed as a clear advantage over the VR360 condition.

Even though this particular experimental condition produces incorrect perspective projection in response to POV translation, this problem was not remarked very often by participants. This can be seen in the responses to questions Q4, Q5 and Q6, which inquired on the presence of perspective distortions, in Fig. 14. Perhaps, these questions were not as representative of these problems as we were anticipating, since the VR360 condition presented the lowest scores, which were significantly below those of the 3D + Billboard condition. However, when participants were asked to spontaneously list advantages and disadvantages for the 3D + Billboard condition, the most common disadvantage was indeed related to inconsistencies between the 3D environment and billboard (remarked by 4 participants) and the "flatness" of the video content (remarked by 2). These are both artifacts that relate to the lack of depth of the video billboard as well as the mismatch between the POV of the user and the POV from which the video content was captured, which results in the aforementioned perspective projection mismatch. This indicates that these artifacts were indeed noticeable. We note, however, that the impact of these inconsistencies to the 3D + Billboard condition experience was not as significant as the impact of the characters quality to the Full 3D condition experience (remarked by 15 participants), or the impact of the lack of POV translation to the VR360 video condition experience (remarked by 12). These results suggest that the compositions problems in the 3D + Billboard condition were not as severe as we were initially anticipating, and participants still preferred to live the experience in this particular condition, as indicated in the comparison results (Fig. 13).

Curiously, the comparison questionnaire also shows a preference for the hybrid 3D + Billboard condition when it comes to character appearance and behavior, even though the same recording was used in this condition and the VR360 one. We believe that there are at least three factors contributing to these results. First, the VR360 condition could contain more noticeable video compression artifacts (as spontaneously remarked by 6 participants). Second, motion parallax and POV translation yielded an increased sense of presence and allowed users to get somewhat closer to the characters, which could improve the illusion of depth. Third, we cannot discard a bias towards favoring the overall preferred condition as a tiebreaking strategy. Finally, we make the remark that, when Q1, Q2 and Q3 answers are compared between VR360 and 3D + billboard, VR360 presented slightly lower score distributions, even if we did not find a statistically significant difference in that comparison.

5.2 Importance of POV translation

In relative terms, the possibility to move POV in the virtual environment seems to have been more important than the visual quality of the content itself. This is reflected on the fact that the quality of the characters was clearly higher in the VR360 and 3D + Billboard conditions

(Q1, Q2 and Q3 in Fig. 14). However, when asked about the effectiveness of the experience (Q9, Q10 and Q11 in Fig. 14) and their preferred condition (Fig. 13), the Full 3D condition had a significant edge over the VR360 condition. In fact, although the preference for the 3D + Billboard condition was clear, we did not find a significant difference between this and the Full 3D conditions. The discrepancy between the two instruments, effectiveness scores and preference, may indicate that this difference might exist, but the total number of participants in the study may have been too small to attain statistical significance.

Furthermore, our results indicate an increase in the simulator sickness score as well as a decrease in the total presence score for the VR360 video condition. We point to two relevant factors that may cause the increase in simulation sickness: (1) the lack of virtual camera translation, that is, the viewpoint position is kept static even if the head of the participant translates, resulting in a visuovestibular sensory mismatch, which is known to induce discomfort and simulation sickness [14]; (2) the use of prerecorded stereo views can cause discomfort because the stereoscopic adjustment used for video recording may not reflect the interpupillary (IPD) distance of the user with precision, and cannot be adjusted by the participant. In fact, recent research suggests that the limited range of IPD adjustments in the current generation of VR headset hardware may be the main reason why women have been reported to suffer more from simulator sickness symptoms in VR [20].

Moreover, we argue that the decrease in the presence score may be explained by the lack of POV translation in the VR360 condition. The sense of presence has been empirically associated to accurate sensorimotor contingencies [19, 23], that is, the coupling of motor commands and appropriate sensory feedback of how these commands affect the environment (the point of view in this case). The VR360 condition is the only to provide incomplete sensory feedback in response to head movement. When participants move their head, the virtual POV cannot be updated to accurately reflect the change in head position due to the lack of proper visual information in the prerecorded video. This causes a mismatch between predicted and collected sensory information, indicating the incompatibility of the observed environment as a real space.

In addition, the insertion of video elements into a 3D virtual environment (3D + Billboard) did not seem to affect the sense of presence when compared to the Full 3D counterpart. Both of these conditions afforded full control of the POV of the participant. However, the tested scenario did not encourage a wide range of movements, unconstrained movement could be detrimental to the *3D + Billboard* condition in terms of billboard image distortion, but beneficial to the Full 3D since our 3D content can be viewed from any POV.

5.3 Limitations and implications

However, we should point out a few limitations concerning the generality of our results.

Notably, our experiment places the user and the content in different rooms. This has two main implications. First, the visual integration of the billboard video with the 3D environment becomes easier, since the billboard is only partially visible through the window opening that connects both rooms. Second, it limits the potential perspective distortion caused by video billboard/3D projection mismatch, since the user can only translate relative to the video billboard by a certain amount, and never get to the point where the billboard is seen as a completely flat object (i.e. when the projection plane of the POV is perpendicular to the billboard plane). Poor integration of 3D content and video borders can produce additional artifacts, such as edges with different lighting, straight lines that do not look straight due to different image projection configurations (between real-time 3D render POV and the

pre-rendered video POV), and objects that seem to float due to the lack of depth in the billboard. As such, the implementation of experiences using the 3D + Billboard configuration has to take these artifacts into account, reducing the range of applications where it can excel.

Furthermore, the Full 3D condition is the only one that can allow the virtual camera to be placed anywhere in the virtual 3D environment without necessarily breaking the experience. For instance, if the point of view is modified to be placed inside the interrogation room, the content would still be meaningful and visually consistent to participants. Full 3D is, therefore, the most generalizable condition since it impose fewer constraints on the type of content and potential applications. This objective advantage of the Full 3D condition was not exploited in our experiment. Nevertheless, the development of high quality Full 3D experiences that take advantage of full freedom of exploration can easily explode in complexity and cost since it has to anticipate different use patterns from users.

To address and gain deeper insight into these limitations, future work could focus on investigating the impact of different combinations of content formats in VR scenarios with 6DoF capabilities, and enabling different forms of interaction with the content and environment.

5.4 Applicability and impact

Despite the discussed limitations, we believe our work is still relevant for many VR scenarios, in which key actions happen within delimited spaces and for which unlimited 6DoF does not become a key requirement (at least within the regions where the video billboards are inserted). Noteworthy examples are Social VR scenarios for shared video watching [5, 9, 17], e-learning scenarios, TV-like shows, as well as events and cultural performances.

It is also beyond doubt that the availability of appropriate virtual scenarios and content is a key aspect to increase the feeling of presence, realism and even togetherness, thus providing satisfactory alternate reality experiences. Thus, the contributions and findings from this work have shed some light on how to efficiently produce and integrate heterogeneous content formats to compose the underlying VR environment and storytelling for a wide set of use cases within the scope of alternate realities.

6 Conclusions

In this paper, we have presented a discussion and experiment on the subject of VR content production and consumption. For this purpose, a professional VR content episode was produced in three content format variants, namely VR360 video, Full 3D, and 3D + Billboard (a combination of video and 3D environment). The production process as well as the intrinsic pros and cons of each content format have been described.

In the experiment, participants experienced and evaluated the three different content format conditions. Under our experimental setting, in which the dynamic actions happened within a delimited space watched through a window, participants were generally more receptive to the content condition that combined video content and 3D environment (3D + Billboard), despite the fact that this particular experimental condition often produces incorrect perspective projection in response to translation of the POV. Overall, our results show that most participants had the best experience in the 3D + Billboard condition. This condition presented subjective presence scores that are similar to the Full 3D condition and higher than the VR360 video, very little variation in simulator sickness due to the VR

experience, and was generally considered the best option on the post experiment comparison questionnaire.

References

1. Alexiadis DS, Chatzitofis A, Zioulis N, Zoidi O, Louizis G, Zarpalas D, Daras P (2016) An integrated platform for live 3d human reconstruction and motion capturing. IEEE Trans Circ Syst Vid Technol 27(4):798–813
2. Anwar MS, Wang J, Ullah A, Khan W, Ahmad S, Fei Z (2020) Measuring quality of experience for 360-degree videos in virtual reality. Sci Chin Inform Sci 63(10):1–15
3. Boukhris M, Paljic A, Lafon-Pham D (2017) 360 versus 3d environments in vr headsets for an exploration task. In: Proceedings of the 27th international conference on artificial reality and telexistence and 22nd eurographics symposium on virtual environments, pp 71–78
4. Cai S, Ch'ng E, Li Y (2018) A comparison of the capacities of vr and 360-degree video for coordinating memory in the experience of cultural heritage. In: 2018 3rd digital heritage international congress (DigitalHERITAGE) held jointly with 2018 24th international conference on virtual systems & multimedia (VSMM 2018), pp 1–4. IEEE
5. De Simone F, Li J, Debarba HG, El Ali A, Gunkel SN, Cesar P (2019) Watching videos together in social virtual reality: an experimental study on user's qoe. In: 2019 IEEE Conference on virtual reality and 3d user interfaces (VR), pp 890–891. IEEE
6. Facebook spaces https://www.facebook.com/spaces. Accessed: 2019-02-19
7. Fourquet E, Cowan W, Mann S (2007) On the empirical limits of billboard rotation. In: Proceedings of the 4th symposium on applied perception in graphics and visualization, pp 49–56. ACM
8. Germann M, Hornung A, Keiser R, Ziegler R, Würmlin S, Gross M (2010) Articulated billboards for video-based rendering. In: Computer graphics forum, vol 29, pp 585–594. Wiley Online Library
9. Gunkel SN, Prins M, Stokking H, Niamut O (2017) Social vr platform: building 360-degree shared vr spaces. In: Adjunct publication of the 2017 ACM international conference on interactive experiences for tv and online video, pp 83–84
10. Hamill J, McDonnell R, Dobbyn S, O'Sullivan C (2005) Perceptual evaluation of impostor representations for virtual humans and buildings. In: Computer graphics forum, vol 24, pp 623–633. Wiley Online Library
11. Hayashi K, Saito H (2006) Synthesizing free-viewpoing images from multiple view videos in soccer stadium. In: International conference on computer graphics, imaging and visualisation (CGIV'06), pp 220–225. IEEE
12. Kalckert A, Ehrsson HH (2012) Moving a rubber hand that feels like your own: a dissociation of ownership and agency. Fronti Human Neurosci 6:40
13. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG (1993) Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 3(3):203–220
14. LaViola JJ Jr (2000) A discussion of cybersickness in virtual environments. ACM Sigchi Bull 32(1):47–56
15. Levoy M, Whitted T (1985) The use of points as display primitives. Technical Report. The University of North Carolina at Chapel Hill, Dept. of Computer Science
16. Mangiante S, Klas G, Navon A, GuanHua Z, Ran J, Silva MD (2017) Vr is on the edge: how to deliver 360 videos in mobile networks. In: Proceedings of the workshop on virtual reality and augmented reality network, pp 30–35
17. Pece F, Steptoe W, Wanner F, Julier S, Weyrich T, Kautz J, Steed A (2013) Panoinserts: mobile spatial teleconferencing. In: Proceedings of the SIGCHI conference on human factors in computing systems, pp 1319–1328
18. Revilla A, Lacosta I, Calahorra G, García-Lajara J (2019) Deliverable D4.1 - first example of content EU H2020 VRTogether project. https://doi.org/10.5281/zenodo.3530496
19. Slater M, Spanlang B, Corominas D (2010) Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. In: ACM Transactions on graphics (TOG), vol 29, p 92. ACM
20. Stanney K, Fidopiastis C, Foster L (2020) Virtual reality is sexist: but it does not have to be. Front Robot AI 7:4

21. Teo T, Lawrence L, Lee GA, Billinghurst M, Adcock M (2019) Mixed reality remote collaboration combining 360 video and 3d reconstruction. In: Proceedings of the 2019 CHI conference on human factors in computing systems, pp 1–14
22. Usoh M, Catena E, Arman S, Slater M (2000) Using presence questionnaires in reality. Presence: Teleoperators & Virt Environ 9(5):497–503
23. Wu TL, Gomes A, Fernandes K, Wang D (2019) The effect of head tracking on the degree of presence in virtual reality. Int J Human–Comput Interact 35(17):1569–1577