Toward Hyper-realistic and Interactive Social VR Experiences in Live TV Scenarios

Sergi Fernández, Mario Montagud, Gianluca Cernigliaro, David Rincón

Abstract— **Social Virtual Reality (VR) allows multiple distributed users getting together in shared virtual environments to socially interact and collaborate. This article explores the applicability and potential of Social VR in the broadcast sector, focusing on a live TV show use case, by providing three main contributions: 1) a novel and lightweight social VR platform; 2) a professional piece of VR content to recreate an interactive live TV show; and 3) an analysis of the performance and user experience.**

The Social VR platform includes different innovative and outstanding features compared to state-of-the-art solutions. It allows a real-time integration of remote users in shared virtual environments, using realistic volumetric representations and affordable capturing systems, thus not relying on the use of synthetic avatars. It supports a seamless and rich integration of heterogeneous media formats, including 3D scenarios, dynamic volumetric representation of users and (live/stored) stereoscopic 2D and 180º/360º videos. In addition, it enables low-latency interaction between volumetric users and a video-based presenter (Chroma keying) and a dynamic control of the media playout to adapt to the session's evolution. The article also describes the production process of an immersive an interactive TV show to demonstrate the platform's capabilities and its potential benefits. On the one hand, the results from objective tests show the satisfactory performance of the platform. On the other hand, the promising results from user tests support the potential impact of the presented platform, opening up new opportunities in the broadcast sector.

Index Terms—**Broadband, Broadcast, Immersive Media, Immersive TV, Interactive Media, Social TV, Social VR, Virtual Reality, Volumetric Media, VR TV.**

I. INTRODUCTION

INTERACTING around media content has been traditionally a social habit. A relevant and widely common example is a group of users gathering at a common location for watching TV content (e.g. sports events, shows) together. In the last two decades, huge efforts have been devoted to achieving a seamless convergence between broadcast and broadband, opening the door to new interactive services thanks to the availability of IP-enabled consumption devices. In this context, two TV-related scenarios can be highlighted. The first one relates to the massive usage of companion screens (e.g. tablets, smartphones) while watching TV content (e.g. [1, 2]), which allows being provided with extra content or engaged with Social Media interactions, among other rich features. The second one relates to the usage of technological solutions to allow the concurrent consumption of the same content by remote users, while being able to socially interact, e.g. via text and/or audiovisual chat channels (e.g. [3, 4]). These latter scenarios and related technology, combined with the former ones, are typically embraced within the Social TV concept [5]. Social TV scenarios have massively awakened the interest of consumers [5, 6], and currently many Video-on-Demand (VoD) platforms (e.g. Youtube), Social Networking platforms (e.g. Facebook), and even platforms by the research community (e.g. [4]), offer these kinds of services.

With the proliferation of immersive technologies in the last years, this connected hybrid ecosystem can go even further. On the one hand, it is now possible to integrate Virtual Reality (VR) content, live VR360 videos, and consumption displays, like Head Mounted Displays (HMDs), in hybrid broadcast scenarios (e.g. [7], [8]). On the other hand, social interaction between remote users can now be enabled through shared immersive virtual environments, bringing up a new communication medium termed as Social VR [9]. Social VR rapidly attracted a high interest, magnified with the social distancing measures brought by the worldwide pandemic. Many Social VR platforms are currently available[1], being Facebook Horizon (formerly Facebook Spaces) and AltspaceVR (by Microsoft) two relevant examples. The existing Social VR platforms can be categorized based on the media formats used for the representation of the shared virtual environment (e.g. 360º scenes [10] or 3D environments [11]) and of the users (e.g. avatars [12, 13], video-based representations [10, 14] or 3D volumetric representations [15]). Likewise, although virtual meetings and gaming-like scenarios could be seen as the main Social VR use cases, this novel medium can also bring significant added value to the broadcast sector. A proof of evidence is Oculus Venues[2], a worldwide-adopted Social VR platform that aims at virtually bringing crowds to live broadcasted events, like concerts and sports. This is also the

[1] Comparison of Social VR platforms, https://ryanschultz.com/2019/11/12/an-updated-comparison-chart-of-sixteen-social-vr-platforms-first-draft-november-2019/ Last Access in September 2021.

[2] Oculus Venues https://www.oculus.com/experiences/go/1555304044520126/?locale=en Last Access in September 2021.

case for Fox Sports VR[3]. This paper focuses on this research opportunity with market potential, addressing key technological limitations and challenges, as well as exploring key research questions, to boost and confirm the applicability of Social VR in the broadcast arena. In this context, the paper provides three main contributions, including innovative technological enablers, the production of a VR content experience to explore the potential of Social VR in the broadcast sector, focusing on a live TV show, and the evaluation of the potential impact of Social VR for such scenarios, in terms of performance results, user experience aspects and awakened interest.

The first and main contribution of the paper is the design and implementation of an innovative and lightweight Social VR platform that enables interactive and hyper-realistic experiences, including a live and low-latency ingest of (broadcasted) heterogeneous video content and a real-time volumetric video capturing and integration of users in the virtual scenario. The presented Social VR platform incorporates key technological enablers for an effective applicability in the broadcast sector, making possible an interactive participation of the audience in a live TV show. Concretely, there are five aspects that make the presented Social VR platform outstanding compared to the state-of-the-art ones (reviewed in Section II) and can potentially bring relevant benefits, thus becoming our (implicit) research hypotheses:

- The platform enables photo-realistic volumetric user representations, unlike most of the existing solutions in which users are represented as avatars (e.g. AltspaceVR, Facebook Horizon, etc.). This allows richer identification of the self and others' representations, as well as richer interaction between the users and the VR environment and among themselves.
- The platform is lightweight and low-cost. It uses off-the-shelf hardware unlike other platforms that require high-end hardware and a fast Internet connection to achieve high quality real-time 3D reconstructions and to provide realistic representation of users (e.g. [13, 15]). In particular, it makes use of affordable RGB-D cameras (i.e. Kinect or Intel RealSense sensors) for the volumetric capture, by adopting from 1 frontal to N surrounding (being $N=4$ a typical value) sensors to capture the human body, and requiring reasonable processing and bandwidth requirements.
- The platform supports a rich combination of media formats to compose the shared virtual environment, like 3D scenarios, 3D reconstructed users and 2D/360º/180º videos, unlike other existing platforms that support specific content types or limited combinations between them. Recent studies (e.g. [16]) have proved the potential benefits of an adequate integration of media formats on the production costs and user experience in VR consumption scenarios, which can also be reflected in Social VR with integrated users.
- The platform supports a live and low-latency ingest of broadcast audiovisual content, including stereoscopic 180º/360º videos and video billboards from a Chroma key room with appropriate background removal features, unlike other existing platforms just support the integration of live (monoscopic) 2D video content via third-party players, which add large delays. This feature not only enables a live interaction between the audience members, integrated as volumetric representations, but also between the audience members and a remote broadcasted show presenter, integrated as a stereoscopic video billboard with Chroma keying.
- The platform allows to control the presentation of specific contents (e.g. related videos, live connection with reporters…) based on the evolution of the session, unlike other existing platforms limited to launching a predefined set of contents according to a timeline and to enabling interaction between the users and with the virtual space (e.g. manipulating objects, controlling the playout of additional media). This enables the presenter to control the experience and launch, or not, contents previously generated or live ingests, increasing the feeling of realism of the show.

The second contribution of the paper is the design and production of a VR content piece in which a live TV show is recreated. This includes the TV set, a live presenter, the placeholders for the integration of audience members, and a set of content pieces (i.e. interviews with various experts, connection with a remote reporter) and realistic animations. The availability of this Social VR environment and content not only serves to prove the performance and benefits of all the features integrated in the platform, but also to provide a proof of concept of how Social VR can open the door to new experiences and business models in the broadcast sector. Fig. 1 shows an example of the created 3D Social VR scenario, and of users interacting with the Social VR experience in the lab.

The third contribution of the paper is the evaluation of both the technological and creative components of the newly envisioned Social VR experience for the broadcast sector via objective and subjective testing, shedding some light on the associated research hypotheses. The results from the objective evaluation demonstrate the satisfactory performance achieved and give an idea of the (reasonable) requirements to effectively run these experiences using conventional Internet connections and low-cost off-the-shelf hardware. The results from the subjective tests ($N=40$) show that end-users rated positively the developed Social VR experience, which provides a satisfactory quality of interaction, immersion and togetherness [17], and confirm the expected benefits provided by each one of the key innovative technological contributions adopted in the Social VR platform. In addition, the insights from semi-structured interviews confirm the potential of the presented contributions, the high interest they awake, and identify aspects to be improved in future releases to boost the adoption in the broadcast domain.

[3] Fox Sports VR https://www.foxsports.com/virtual-reality Last Access in September 2021.

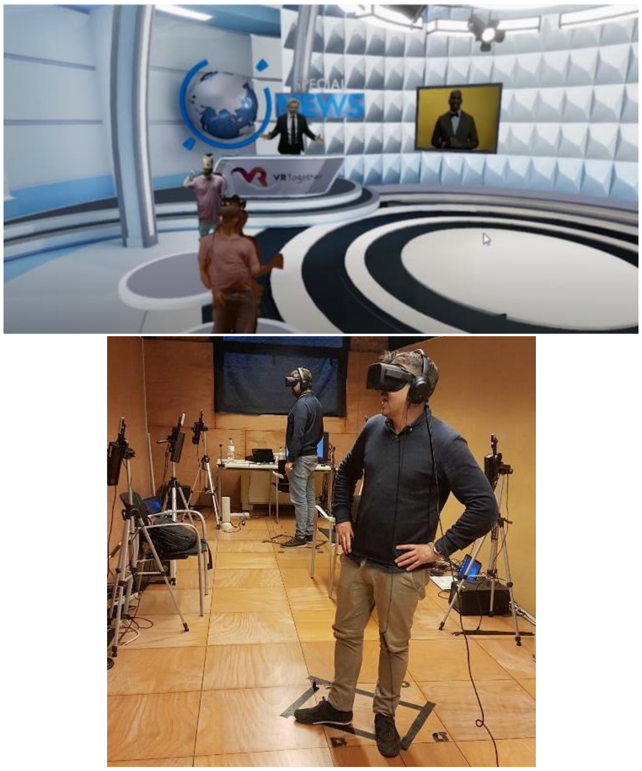

Fig. 1. Up: Two users and a live presenter integrated in the Social VR scenario presented in this work. Down: Two users experiencing the Social VR platform in the lab.

In summary, the paper presents a full-fledged and outstanding Social VR solution for live broadcast events and reports a holistic evaluation of the full workflow for Social VR including technological, content production and user experience aspects.

The structure of the paper is as follows. Section II reviews the state-of-the-art, from Social TV contributions to more recent Social VR contributions, addressing associated technological enablers. Section III presents the technological components and aspects of the innovative Social VR platform that has been developed, while Section IV reports on the production of a Social VR story and content to provide an outstanding Social VR experience using the developed platform. Section V reports on the objective and subjective evaluation of the platform and the experience, respectively. Section VI provides a discussion about the obtained results and lessons learned. Finally, Section VII outlines our conclusions and suggests ideas for future work.

II. Related Work

This section reviews the state-of-the-art in this field, starting with an overview of relevant contributions focused on Social TV, and then continuing with technological solutions and studies toward or on Social VR, including open-source and commercial platforms. While the first sub-section motivates the relevance of the topic, reflects on lessons learned and potential benefits, as well as justifies the evolution toward Social VR, the second sub-section reviews related contributions in the Social VR field, identifying the limitations of existing solutions and the advantages and benefits of the presented solution.

A. From Social TV to Social VR

Research on Social TV (a.k.a. social viewing) has attracted attention in the last decade. Some example works focused on: analyzing the advances in Social TV and categorizing the existing developments [5, 18]; studying the appropriateness of different chat modalities [4, 19, 20]; determining the impact of delays [3, 20]; and assessing the interest in these scenarios [6]. For instance, the survey in [6] reflected the high interest of consumers in enjoying Social TV like scenarios, but also the need for better technological solutions and interaction modalities to support them. Likewise, many lab-controlled [4, 20] and in-home [21] studies have shown the benefits provided by Social TV mainly in terms of engagement, togetherness (i.e. feeling of being together), intimacy and improved relationships. In addition, recent works have revealed a high interest in Social TV platforms, not just in the entertainment sector, but also for training, education and collaboration [4].

B. Social VR: technology, user studies and market solutions

Given the benefits and high potential of Social TV, both the research community and industry started to explore how to support these scenarios through VR technology and formats with the goal of increasing the feeling of engagement, immersion, and togetherness. State-of-the-art contributions for these aspects are reviewed next.

1) Works from the research community

Many research works have provided valuable contributions and insights in the area of Social VR, with different application contexts, including broadcast environments as the key focus. These include Virtual/Augmented Reality (VR/AR) meeting systems integrating Computer Generated Imagery (CGI) and 3D content for the shared environments, as reviewed in [22].

First, some relevant works have focused on enabling telepresence and social interaction for collaborative and training scenarios, which are relevant use cases of Social VR. The work in [11] presents a multi-party telepresence system based on the use of color and depth sensors, like Kinect [23], for the end-users' reconstruction and their integration in 3D environments. The Social VR scenario in [11] was based on the use of projection-based displays, not HMDs, and was evaluated for the use case of virtual tourism. The work in [14] presents a similar telepresence system, but using virtual avatars and video-based reconstructions techniques, like free-view point video, for the end-users' representations. The target scenario in that case was collaborative training and exploration spaces. An evolved version of the system in [14] was then prepared in [24] for its application in Mixed Reality (MR) environments.

Second, some other relevant works have focused on supporting shared media consumption with Social VR platforms. The work in [21] highlighted that the adoption of HMDs in conjunction with RGB-D cameras for the end-users' capturing and representation can lead to an increased engagement, feeling of immersion and enjoyable embodied telepresence compared to traditional 2D social viewing tools. The work in [25] analyzed the requirements and challenges to efficiently support shared media consumption of 360º videos using HMDs, and proposed guiding and interaction strategies to contribute to this. The work in [10] presented a web-based and video-based Social VR platform mainly focused on shared media consumption of stored content. In that platform, users are photo-realistically captured by a single RGB-D camera (Kinect), and the shared VR scenario is represented as a 360º static image. Finally, the work in [17] proposed an experimental protocol and a questionnaire for evaluating Social VR experiences. By adopting a photo sharing use case, the experiment consisted of comparing the quality of interaction, social meaning and presence/immersion levels in three scenarios: face-to-face, Skype, and Social VR (using the platform from [10]). The results of the experiment not only proved that the proposed evaluation methodology was appropriate (i.e. the designed VR questionnaire was reliable), but also that Social VR provides an enhanced user experience compared to traditional conferencing tools, like Skype.

Third, other works have investigated on the optimization of pipelines for the live delivery of immersive media (e.g. [26]), and proved the benefits of inserting video billboards for not only representing peripheral (e.g. simulation of crowds) but also central parts (e.g. representation of key characters) of 3D VR experiences [16].

2) Works from industry: existing Social VR platforms

The industry is also devoting efforts to the development and deployment of Social VR, telepresence and collaborative virtual environments.

With regard to collaborative workspaces, IBM recently presented DataSpace [13], a re-configurable hybrid reality system supporting both AR/VR scenarios. Even though the key focus of DataSpace relies on the (re-)configuration and combination of physical and digital resources for supporting next-generation workspaces, including the interactive presentation of media content, it also supports the interaction and collaboration between remote users, represented as 3D avatars.

With regard to realistic representations of users, two solutions from the industry can be highlighted. The first one is the Microsoft Holoportation system for HoloLens [15]. This system is however mostly focused on AR scenarios and requires a complex and expensive capturing setup, with eight custom camera pods. The second one is the solution by Mimesys (a Belgian startup acquired by Magic Leap in 2019[4]) that developed an AR telepresence system based on holographic representations of end-users, by using Kinect and Intel RealSense sensors for the capture and Magic Leap headsets for the visualization. However, no references about the usage of these solutions for multi-party scenarios have been found.

The availability of many platforms in the market is a proof of the high interest Social VR is awakening. Many commercial and open-source Social VR platforms have appeared in the last few years, and even qualitative comparisons among them have been conducted, like the one by Ryan Schultz[5], which pays particular attention to high-level and commercial aspects. Table I provides a categorization of key Social VR platforms and takes into account key features that support the use case explored in this paper, mainly: types of end-users' representations, supported media types, and integration of live broadcasted ingest (including Chroma keying capabilities). The comparison serves to confirm and highlight the value of the features provided by the platform presented in the paper, compared to other existing ones. As it can be seen in the table, almost every platform supports 3D environments, provides support for desktop and VR modes, allows for (traditional) media sharing, and allows a live broadcasting of running VR sessions to other 2D video platforms, like Youtube and Twitch. The presented Social VR platform also provides these widely supported features. Interestingly, all these platforms rely on the use of (either cartoon-like or human-like, even customizable) 3D avatars for the end-users' representations, and few of them (e.g. Mozilla Hubs and Spatial.io) also support the integration of live 2D windowed videos from the webcam. The presented Social VR platform not only supports these features, but it additionally supports realistic volumetric end-users' representations using Time Varying Meshes (TVM). Although previous works have provided technology for the reconstruction of users and their integration in 3D virtual environments by using single RGB-D sensors (e.g. [10, 11]) or expensive setups (e.g. [15]), the presented platform not only integrates a multi-sensor but still low-cost capturing setup [27, 28] to provide full volumetric realistic representations of the involved users, but also an end-to-end pipeline to enable low-latency audiovisual interaction between many of them. In addition, unlike existing Social VR platforms that integrate third-party streaming solutions, like Youtube or Twitch, for the addition of live sources for traditional media consumption, the presented Social VR platform supports the integration of live broadcasted (stereo and mono) streams, including 180º/360º feeds and background removal (Chroma keying), though a custom and low-latency standard-compliant pipeline. This does not only allow a significant reduction of the delay compared to the mentioned third-party solutions, but even to enable a live interaction with media content delivered via such a pipeline (e.g. a remote presenter from a Chroma key room). This is not supported by any of the existing Social VR platforms.

[4] Mimesys Joins The Magic Leap Family (May 2019) https://www.magicleap.com/en-us/news/press-release/mimesys-joins-magic-leap Last Access in September 2021.

[5] Comparison of Social VR platforms, https://ryanschultz.com/2019/11/12/an-updated-comparison-chart-of-sixteen-social-vr-platforms-first-draft-november-2019/ Last Access in September 2021

TABLE I
COMPARISON OF STATE-OF-THE-ART SOCIAL VR PLATFORMS

Platform	VR / Desktop Support	End-users' Representation	3D Environment	Integration of Live Broadcasted Video	Chroma Keying for Live Videos	Live 180º / 360º video	Broadcast Live Sessions
AltspaceVR[a]	Y / Y	Human-like avatars (customizable clothes, but no faces)	Y	Partially (integration of YouTube player)	N	N	Y (e.g. on Twitch)
BigScreen[b]	Y / N	Cartoon-like avatars (customizable)	Y	Yes (but integrating player of third-party platforms and TV channels)	N	N (only static 360º scenes for the environment)	Y (e.g. on Twitch)
Mozilla Hubs[c]	Y / Y	Cartoon-like avatars (customizable) and live 2D video from the webcam	Y	Partially (integration of YouTube and Twitch players)	N	N (only static 360º scenes for the environment)	Y (e.g. on Twitch)
NeosVR[d]	Y / Y	Cartoon-like avatars (customizable)	Y	Partially (integration of Twitch player)	N	N	Y (e.g. on Twitch)
Spatial.io[e]	Y / Y	Human-like avatars and 2D videos from the webcam	Y	Partially (integration of video players and screen sharing feature)	N	N	-
Virbela[f]	Y / Y	Human-like avatars (customizable)	Y	Partially (integration of YouTube player)	N	N	Y (e.g. on Twitch and YouTube)
Vive Sync[g]	Y / Y	Human-like avatars (customizable)	Y	-	N	-	Y (e.g. on YouTube)
Presented Social VR platform[h]	Y/ Y	Realistic volumetric representation (TVM), 3D avatars, live 2D video from the webcam, just audio communication, or no audio and video but just presence (i.e. ghost user)	Y	Y+ (Own live broadcasting pipeline)	Y	Y	Y (on YouTube)

[a] https://altvr.com/ [b] https://www.bigscreenvr.com/ [c] https://hubs.mozilla.com/ [d] https://neos.com/ [e] https://spatial.io/ [f] https://www.virbela.com/ [g] https://sync.vive.com/ [h] https://vrtogether.eu/ Last Access for all URLs: September 2021

Finally, the presented Social VR platform not only supports a seamless integration of a wider variety of media formats than other available Social VR platforms (Table I) and proposed solutions, like DataSpace [13], but enables a user-level interactive presentation of the available media assets and live ingests based on the session evolution, which becomes very valuable for live interactive sessions that may not be linear or may not follow a predefined timeline.

All these innovative and enhanced features provided by the presented platform enable richer and more interactive experiences within the broadcast sector. This work not only integrates all these features into a single and modular platform, but demonstrates their associated requirements and performance, as well as the potential benefits for the broadcast sector, by focusing on a live virtual TV show.

A more detailed comparative analysis and benchmarking for all these Social VR platforms can be found in [29].

III. SOCIAL VR PLATFORM FOR BROADCAST CONTENT

Based on the insights from the review of existing Social VR solutions (Section II), it was decided to develop a new platform overcoming the limitations of existing ones in terms of support for: 1) ingest of live content with low-latency; 2) photo-realistic volumetric representations of users, instead of avatars, captured in real-time, even including self-representations for each user; 3) blending of heterogeneous content in virtual immersive environments, including live 2D video, stereoscopic 180º/360º video and 3D content; and 4) interactivity features, between the real-time integrated presenter and users and for the presentation of different media sources. In addition, the development of a new platform allows having higher control over the technological components (e.g. fine tuning settings to maximize performance and quality) and the experience.

This section presents an overview of the novel, lightweight and hyper-realistic Social VR platform that has been developed and used for evaluating the Social VR experience, by describing its main parts and components, including technical and implementation details. A high-level architecture of the platform, depicting the integrated components as well as the streams exchanged among them, is shown in Fig. 2. This architecture is aimed at effectively supporting all developed innovative components and features in a standard-compliant manner, and thus can potentially become a reference concept implementation for Social VR in live broadcast scenarios.

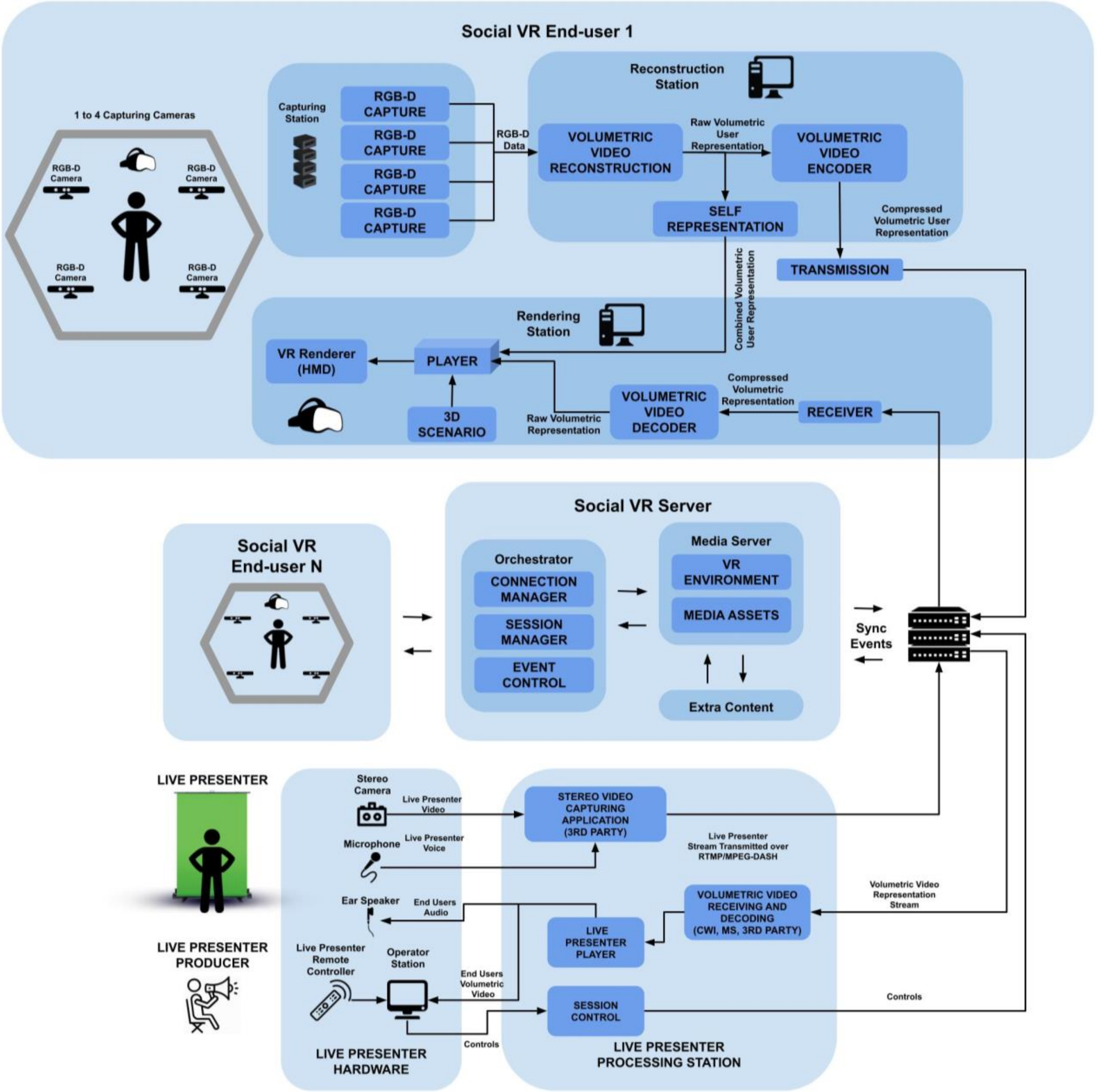

Fig. 2. High-level architecture and flow diagram of the presented Social VR platform

A. *Pipeline for (Live) Volumetric Media (TVMs)*

This component allows the integration of end-users, captured in real-time and represented in 3D as full volumes, in shared VR environments. Its integration allows realistic volumetric representation of end-users, without having to rely on the use of avatars, as in most of the existing Social VR platforms (Section II). Next, the key sub-components and involved processes for this component are depicted.

1) *Capturing & Reconstruction*

To enable photo-realistic and fluid volumetric representations of users in the Social VR experience, a real-time video capturing and reconstruction sub-system has been integrated, based on the work in [27] and [28], which is publicly available as open-source[6]. In that sub-system, the video capturing is performed by using multiple RGB-D sensors [29], like Kinect [23, 30, 31] and Intel RealSense [32], which both capture the color and depth information.

Theoretically, there is no limitation with regard to the number of RGB-D sensors to be used in the capturing and reconstruction sub-system. However, limitations like the physical space, computational resources and interference between sensors need to be considered. To keep the costs and computational load low, the setup considered in this work is based on a capturing sub-system using four RGB-D sensors, concretely Intel RealSense D415 cameras[7] [32], placed, calibrated and synchronized according to the specifications described in [29]. The four RGB-D sensors are connected to four capturing stations, with no particular requirements beyond being able to receive the data from the sensors (e.g. mini PCs). These stations are connected via a Local Area Network (LAN)

[6] Open-Source Portable, Flexible and Facile Volumetric Capture System: https://github.com/VCL3D/VolumetricCapture Last Access in September 2021

[7] Intel RealSense D415 sensor: https://www.intelrealsense.com/depth-camera-d415/ Last Access in September 2021.

to a Reconstruction Station with a graphical board supporting Graphics Processing Unit (GPU) operations. In this work, PCs with an Intel Core i7 processor, 32 GB of RAM and a GeForce 1080 Ti board, have been used.

The effective capturing area is approximately a circle with a 3m radius. The RGB-D sensors are placed around the circle and are all pointing towards the action area in the center of the circle. The reconstruction is performed by merging the captured RGB-D frames from each sensor and extracting their 3D geometry. Then, the data from all sensors are synchronized to achieve a coherent volumetric capturing. Next, a background removal process is performed to isolate the geometry from the color information that is needed for the user's 3D representation. The sensors' color information is mapped into voxels and filtered to remove noise. A volumetric point cloud is then created and the voxels are projected onto meshes to be delivered as volumetric video. An absolute timestamp, by means of Network Time Protocol (NTP), is inserted into each TVM frame to allow their in-sync presentation at the client side. Details for all these previous steps are provided in [27] and [28].

2) Encoding & Transmission

The reconstructed volumetric sequences need be encoded and encapsulated for an appropriate real-time distribution via IP networks. Among the formats supported by the presented Social VR platform, this work is based on the use of dynamic meshes (i.e., TVMs) for which many compression methods have been proposed (e.g., [33, 34]), and open-source compression software solutions are available. In particular, the presented (version of the) platform has adopted the open source Draco library[8] for the compression of TVMs, and the open source RabbitMQ tool[9] for the delivery of the compressed TVMs data. Every node generating a TVM stream uploads that stream to a local RabbitMQ server, and the interested recipients retrieve the stream from that server, by getting the connection information from the Orchestrator (introduced in Section III.C).

Apart from the visual communication channel, the platform integrates an audio communication pipeline relying on the use of socket connections for the data exchange. In particular, the open source Socket.io[10] library has been adopted for such a purpose in the presented implementation.

With these two technological components and processes, the Social VR platform is able to integrate in real-time volumetric and realistic representations of end-users, also enabling a low-latency audiovisual interaction among them.

B. Pipeline for (Live) 2D Media

In addition to the pipelines for volumetric media integration, pipelines for the integration of non-volumetric audiovisual formats have been also integrated. This allows the interactive presentation of media assets stored on a (cloud) Media Server (see Fig. 2), but most importantly, the integration of live media sources for the interactive presentation of broadcasted content, like video feeds (e.g. for interviews, scenes from a remote event or location, etc.) and even (billboards of) remote presenters from Chroma key rooms.

The pipeline for live (and on demand) traditional media sources includes support for audio and video systems (including 180º/360º stereoscopic formats), the preferred encoding/transcoding settings, and a low-latency (multiplexed) transmission via Real-Time Messaging Protocol (RTMP), or alternatively its conversion into Dynamic Adaptive Streaming over HTTP (DASH), if required for deployment in large and open environments. This end-to-end pipeline for ingesting non-volumetric media has been implemented by using the open source GStreamer framework[11].

Beyond being able to effectively integrate live 2D and 180º/360º (stereoscopic) media sources in an interactive manner, the availability of a low-latency pipeline allows a real-time bidirectional communication between remote people, like a presenter and the audience, augmenting the interaction possibilities compared to other existing Social VR platforms.

C. Orchestration and Interactive Session Control

1) Orchestration

Orchestration components (i.e. Orchestrators) are commonly used in video conferencing systems to handle the set of audiovisual and control streams [35]. In the presented Social VR platform, an Orchestrator has been developed and integrated to deal with session and stream management tasks. The Orchestrator handles the remote networking information (e.g., IP addresses, ports, protocols), accommodates all remote users in a shared virtual environment, manages the real-time interaction channels, acts as a relay server for media streams, and ensures a consistent synchronized experience (by informing about an NTP clock reference to synchronize with).

In addition, the Orchestrator informs about potential errors or unexpected behavior in the distributed shared sessions and can potentially perform a set of recovery actions in case of connection problems.

2) Interactive Session Control

The Social VR platform targets at enabling highly interactive sessions, in which remote audience members and presenter(s) can communicate and exchange impressions within the context of TV-related content consumption, and different media assets can be dynamically shown and hidden. Therefore, it becomes essential to be able to adapt the presentation of content based on the evolution of the session, which is hardly supported in existing Social VR platforms (Section II). To enable this, a Unity app that runs on both mobile devices and PCs has been developed. The app includes a Graphical User Interface (GUI) that allows watching the representations of the end-users, a timeline, a panel with all available content assets, and GUI elements to interactively control their presentation (Fig. 3). Thanks to the availability of this app, an operator / realizer, or even the same presenter (with the mobile version of the app), can trigger and control the presentation of different media

[8] Draco: https://google.github.io/draco/ Last Access in September 2021.

[9] RabbitMQ: https://www.rabbitmq.com/ Last Access in September 2021.

[10] Socket.io: https://socket.io/ Last Access in September 2021.

[11] GStreamer media framework, https://gstreamer.freedesktop.org/ Last Access in September 2021.

content (e.g., related videos, live connections, etc.), according to the ongoing interactions, which becomes essential in live and dynamic sessions, like in TV shows.

D. Playout

The final stage of the end-to-end pipelines in the presented Social VR platform consists of the interactive presentation of the media content at the client side, and the integration of all considered interaction modalities for the Social VR experience.

A Unity-based player has been developed to properly receive, integrate and present all available streams for the shared VR scenes, the end-users' representations (as TVMs), the 2D live media sources (i.e., presenter billboard, traditional and 180º video feeds), and all other stored assets that will enrich the experience (e.g., recorded videos, graphics, visual effects, etc.).

The player includes different components and engines to provide the following features:

- Connection to, and interaction with, the Orchestrator. With regard to the communication with the Orchestrator, the user, through the player interface, needs first to log in, and then create and/or join a shared Social VR session, by selecting the desired VR scenario among the available ones. During the session, the necessary information will be exchanged to enable interactive and coherent experiences. Finally, at the end of the experience, the session will be terminated by freeing all associated resources.
- Loading or receiving the 3D virtual scenario (and its associated media assets) where the end-users will be teleported, placing initially each user in an appropriate position and orientation within the virtual scenario.
- Receiving the data streams for the self and others' representations, as TVMs.
- Receiving the data streams from the live and stored media sources, including traditional 2D and stereoscopic 180º/360º video. For such a purpose, an open-source software component[12] that connects the GStreamer media pipelines with Unity has been adopted. It is called Gstreamer Unity Bridge (GUB), and is able to transmit and play any media Uniform Resource Identifier (URI) provided by GStreamer pipelines into Unity 3D textures, with low latency.
- Eliminating in real-time the green background from the incoming live video stream captured in Chroma key rooms, thanks to an ad-hoc Unity shader. This is very useful to just display the (stereoscopic) silhouette of remote participants (i.e., presenter billboard) in the VR environment, thus increasing the perceived realism.
- Seamlessly blending all content formats and streams that constitute the Social VR experience.
- Ensuring intra-media and inter-media synchronization [36], as well as a timely and synchronized presentation of events and each selected media stream (e.g. launched through the "*Interactive Session Control*" app), in coordination with the Orchestrator. The in-sync presentation for each and between each involved media stream and event is achieved by interpreting the timestamps added at the origin side, thanks to the use of NTP as the global clock sync mechanism in the shared session, retrieved from the Orchestrator.

By combining all these features, the developed player allows running more complete and innovative experiences than those offered by state-of-the-art Social VR platforms.

The player can run on the Reconstruction Station, or on a different station with similar characteristics (see Fig. 2). The same station has been used in the setup of this work.

IV. Social VR Content Production

A professional VR content piece has been produced to be able of not only assessing the potential of Social VR to provide satisfactory shared immersive experiences while apart, but to demonstrate the benefits of all innovative components and features added to the developed Social VR platform, when specifically applied to a live TV (broadcast) show integrating a remote virtual presenter (stereo billboard), (volumetric) audience members, and interactive media ingests.

As immersive stories tend to limit synchronous interaction [37], an innovative narrative was ideated to include interaction elements with the media content and between audience members. The developed story revolves around a last minute piece of news announced in a live TV show, and includes immersive and interactive elements to ensure that the participants get the necessary insights while they have space to converse. Participants are virtually placed in the TV studio and interact with the presenter (Fig. 1) who controls the production of the show, which media elements appear and when.

This sub-section provides details about the chosen scenario and the production process of the Social VR content and scenario to evaluate this new type of interactive and immersive experiences where VR and TV converge.

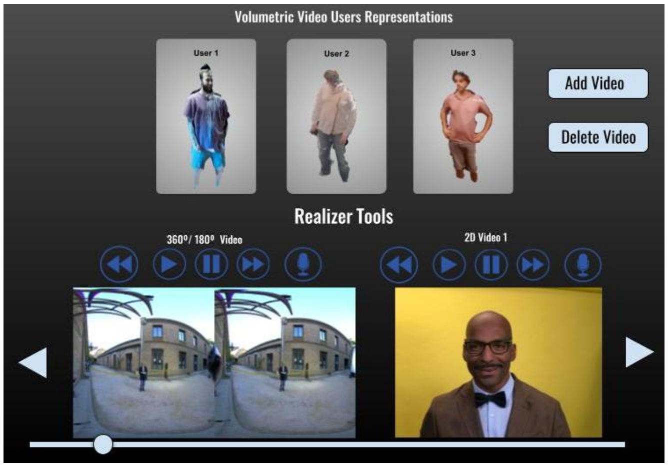

Fig. 3. GUI Mockup of the developed interactive session control app

[12] Gtreamer Unity Bridge (GUB), GitHub repo: https://github.com/ua-i2cat/gst-unity-bridge Last Access in September 2021.

A. *Pre-Production*

The created content in this study belongs to the second episode of a three-episode thriller-like plot which theme is an investigation about the murder of a celebrity. In the first episode, two suspects are interrogated by a police officer in a police station [16], and two users are tele-ported to the virtual scenario to attend the interrogation scenes, playing the role of inspectors. That first episode consisted of an offline content experience (i.e., all content pieces where pre-produced, stored and linearly presented in the client application) and focused technically on enabling the interaction between users, exchanging information extracted from the interrogations. The presented second episode contains live and interactive elements, and integrates heterogeneous video and 3D formats. On the one hand, the presented scenario includes the integration of a remote presenter, who is live captured from a Chroma key room and inserted (i.e. somehow tele-ported) into the virtual TV set as a video billboard. The presenter is the one conducting the TV shown and informs about the last minute murder. Participating users are also real time captured as volumetric 3D video and placed in different positions of the virtual environment, thus experiencing the same story from different viewpoints. During the episode, different interactive content pieces are presented and live connections with experts and protagonists at the crime location are made, aiming at increasing the feeling of realism and immediacy. With all these features, the users not just feel as passive and remote audience members, but as active participants inside the live show, even requiring their cooperation at some points to better understand what could happen between the murdered and the suspects. This opens up with new possibilities in the broadcast media sector.

1) *Scripting and Casting*

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 wealthy celebrity at the peak of her career. Two persons are the main suspects: the lover of the victim; and her assistant. In the first episode, the two suspects were interrogated by a police inspector, revealing that both of them have a different version about what happened and have things to hide.

This second episode informs the audience about the murder and gives some information about what happened and the investigation procedure, also making a live connection to the crime location where the investigations are in place. Therefore, a script for this second VR episode was written [38]. Likewise, a casting process was conducted to select the actors representing the roles of the presenter, experts to be interviewed, reporter and investigators. The actors playing the role of the inspector and suspects were already selected for the first episode, and thus kept for a consistent evolution of the VR story. In that sense, the participation of professional actors contributes to making the experience more credible, and thus immersive.

More details about the developed story, the pre-production tasks and the casting processes are provided in [38].

B. *Production and Post-Production*

With respect to content production and consumption, the media formats to use can have direct implications on the required infrastructure, complexity, costs and on the user experience. The conducted tests in [16] showed that a strategic combination of 3D and video-based content does not just contribute to a reduction of production efforts and costs, but also provides a very satisfactory Quality of Experience (QoE) in terms of feeling of realism, presence and simulation sickness, as well as certain levels of motion parallax if video planes are appropriately placed, when providing unlimited 6 Degrees of Freedom (6DoF) – i.e. freedom to explore and navigate around a 3D virtual environment – is not necessary. By leveraging the insights from that study, it was decided to place the users within specific sub-areas of the virtual environment delimited with circles, as in the *Still Standing* TV game show broadcasted in many countries (see Fig. 4). This gives a mixture between classical TV news and contest sets. The whole virtual TV set becomes then the shared VR environment for all participants, having all of them an appropriate viewing perspective between themselves, with the host and with the media projection spaces using a semi-sphere layout (see Figures 1 and 5).

Key aspects about its production and post-production processes are provided next, but readers can refer to [38, 39] for a detailed description and getting open access links to the created media assets.

With regard to the shared VR environment, the TV set was modelled and recreated in realistic 3D using with Autodesk Maya and then exported into FBX format and integrated in a Unity project (see Fig. 6). The 3D modelled environment took into consideration the appropriate layout and distribution of elements, with space for up to four users, the live presenter, and a semi-sphere projection space for projecting additional videos (e.g. stereoscopic 180º video connecting with the crime location and onsite reporter, or 2D videos for videoconferencing connections with experts), as shown in Fig. 5.

With regard to the dynamic video-based elements, they were shot using a Z CAM K1 Pro Camera (stereoscopic 180º video, 2880p30 resolution). All video scenes were recorded to be able of showcasing demos of the VR episode without requiring the real-time participation of the actors. The scenes for the interventions of the presenter and a technology expert being interviewed were shot in a Chroma key room (see Fig. 7), while the ones for the connections with the reporter and investigators were shot in the exterior of a building where the crime was supposed to happen (see Fig. 8). This transition between a fully indoor 3D environment and an outdoor stereoscopic 180º video environment when the live connection with the reporter is made was targeted at achieving an appropriate omnidirectional scene blending, potentially increasing the feeling of immersion.

The story was designed to be very dynamic, with a quickly changing environment with the presentation of different interactive content pieces, and moving the users' attention from one location to another. In particular, it develops as follows:

- *Phase 1)* The TV show starts with some interactive visual effects and immersive music.
- *Phase 2)* The presenter welcomes the users, and quickly informs about the last minute murder.

Fig. 4. Typical distribution of *Still Standing* TV game shows (Spanish version, retrieved from https://www.antena3.com/, Last Access in August 2021).

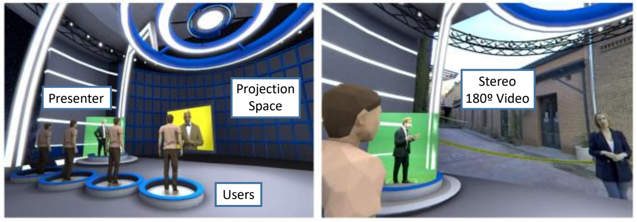

Fig. 5. 3D distribution of the virtual TV set for the designed Social VR scenario, with the different media sources and viewing perspectives.

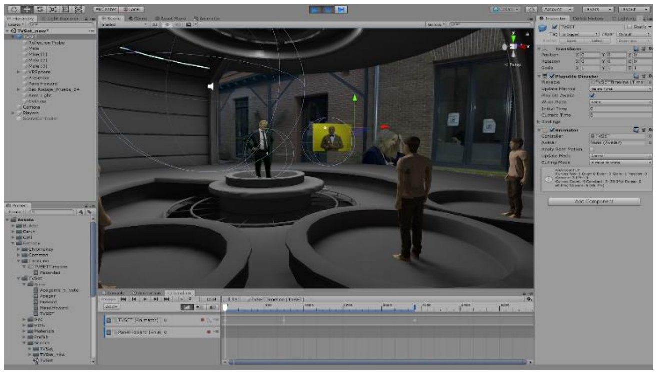

Fig. 6. 3D model of the virtual TV set (up) and Unity project for the designed Social VR scenario (down).

Fig. 7. Shooting to the presenter from a Chroma key room.

Fig. 8. Shooting in the exterior of the building to simulate the live connection with the crime scene location.

- *Phase 3)* The presenter makes some questions to fake remote users to provide a higher feeling of realism and immersion. This allows increasing the attention of the real participants, boosting interaction, and giving credibility in case that a recorded version of the experience is used (e.g. when an actor is not available for a demo). In the case of having a live connection with the presenter, he/she can directly talk to the real-time captured users.
- *Phase 4)* A connection with a remote expert is made via 2D videoconferencing to describe a disrupting technology that will be used to help solving the crime.
- *Phase 5)* A connection with the crime location scene is made via a stereoscopic 180º video scene shown on the projection space, giving the feeling of being tele-ported there. During that connection, a reporter at the remote location will interact with the presenter of the show and will also make questions to a police inspector to get further information about what happened and what is ongoing.
- *Phase 6)* After the connection is closed, further discussions between the participants and with the presenter can happen.

After the recording and modelling of all assets, post-production processes were conducted for all the raw material, including the required adjustment tasks for an appropriate compositing and seamless blending. Finally, realistic lighting conditions were recreated in order to provide a natural integration of the users and characters into the 3D virtual environment. A variety of post-processing effects were also applied to increase the realism. Some examples include: ambient occlusion, addition of dark corners, addition of vintage effects, correction and equalization of colors, etc.

The final result for the 3D virtual TV set is shown in Fig. 9. A demo video of the developed Social VR platform and the produced content experience for the live TV show can be watched at: https://www.youtube.com/watch?v=KfpTIyS5cA0 The overall VR episode lasts around 7 minutes.

V. Evaluation

This section firstly describes the adopted evaluation methodology together with the evaluation setup and scenario. Then, it presents the obtained results, both from objective and subjective testing. Regarding the objective (performance) evaluation, we report the consumption of computational and network resources on the client side as well as end-to-end delays for the involved media pipelines. Regarding the subjective evaluation, we report on the perceived quality of interaction, togetherness and immersion, as well as on the answers to conducted interviews.

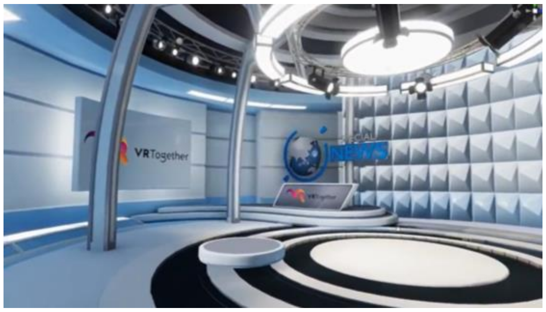

Fig. 9. Overview of the final 3D modelled Social VR scenario.

A. *Methodology*

The Social VR experience was evaluated in sessions of two participants, plus a live presenter from a Chroma key room to increase the interaction possibilities. Although the VR scenario and the platform themselves support up to four participants, it was decided to proceed with sessions of two participants to first assess the user experience in simpler sessions, with a number with high applicability that can actually boost interaction [6].

On the one hand, the evaluation included objective performance tests to gain insights about the computational and bandwidth requirements of the experience, as well as about the delays for the exchanged live streams. On the other hand, the evaluation included user tests by making use of questionnaires and semi-structured interviews.

B. *Evaluation Setup and Scenario*

The experiments were conducted in a Social VR lab in Barcelona (Spain), which facilities are shown in Fig. 1. The lab room included the necessary equipment for the TVM-based users' reconstruction, including four RGB-D cameras (Intel RealSense) and five PCs (one per camera plus one controller, Fig. 10). With regard to the TVM streams, they were set with a resolution of 12k vertices and a capturing frame rate of 22 fps, which was dropped to 14fps for an effective real time encoding and transmission. As parametrization, we adopted the outcome of the subjective study on mesh compression performance performed in [40]. For the reconstruction and rendering stations, a PC with an Intel Core i7 processor, 32 GB of RAM and a GeForce 1080 Ti board, was used for each involved user. Although the two participants were in fact placed in the same physical room (see Fig. 1), they were interconnected through an Orchestrator deployed in Rennes (France), thus recreating an inter-country Social VR session.

The room had no background or surrounding noise. Each user was equipped with an Oculus Rift, with an integrated microphone for the audio interaction, and noise-cancelling headphones to isolate external noise and perceive better the spatial audio provided in the experience. Thus, the users were able to interact through (spatial) audio and (volumetric) visual channels. The users were standing at the center of the effective capturing region during the experience (see Fig. 1) and had limited 6DoF (although were instructed to not move too much during the experience, especially because of the cables). In addition, a laptop was used to record the audio and video from each participant via its integrated webcam and microphone.

The live presenter was captured from a Chroma key room located in an upper level of the same building (see Fig. 11). Its audiovisual stream was delivered through the Orchestrator via RTMP, which is the output provided by the 180º camera (Z CAM K1 Pro Cinematic VR180 Camera). This allows minimizing the latency, avoiding the conversion into DASH. For a more pleasant experience, an experiment facilitator located at the same Chroma key room controlled the interactive session app for managing the presentation of content. That way, the participants see the presenter without any device on hand, and the presenter can actually focus on the play and interaction with the audience. The same PC used to run the interactive session app was also used to run the software that manages the live streaming session from the camera (Z Cam Wonderlive software that comes with the camera).

With this setup, audiovisual communication channels were available between the two participants and the presenter. Apart from the Operator at the Chroma key room, an experiment facilitator was present in the Social VR lab room to assist the users and to control the test. Chat tools were used to enable communication between the experiment facilitators. The Orchestrator was used to synchronously launch the shared VR experience for each involved participant, by choosing the second chapter of the VR story presented in Section IV as stimuli for the Social VR experience.

C. *Objective Testing: Performance Metrics*

This sub-section reports on objective performance metrics on the client application (i.e., the Unity-based player) measured when running the produced Social VR experience in this study (second episode). In particular, it reports on:

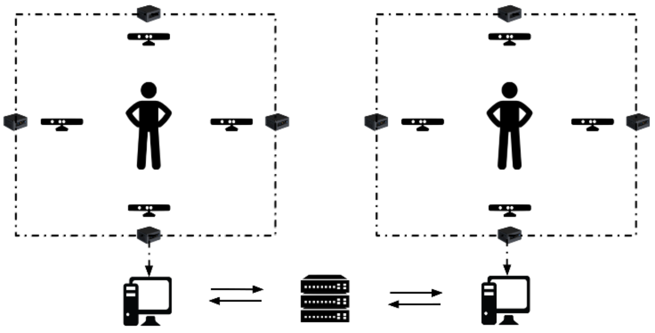

Fig. 10. Social VR lab setup.

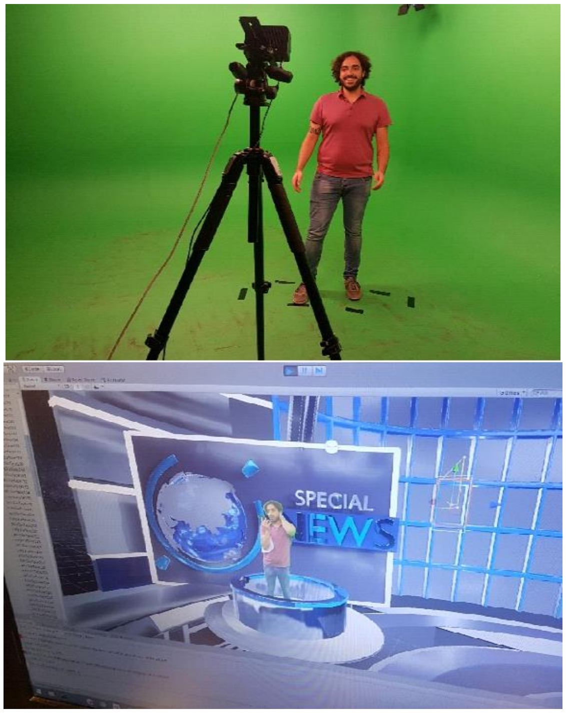

Fig. 11. Live capture of the presenter and its integration in the virtual scenario.

- Computational Resources metrics: CPU load (%), GPU load (%) and RAM usage (MB), by using the tool from [41].
- Bandwidth consumption (Mbps), as reported by Wireshark[13].
- End-to-end delays, by comparing the capturing and rendering timestamps (explained later).

The metrics were measured on a PC with the following characteristics:

- CPU: Intel(R) Core(TM) i7-10750H @ 2.60GHz 2.59 GHz
- GPU: NVIDIA GeForce RTX 2070
- RAM: 16 GB.

The metrics were sampled along the duration of the whole session, and the reported values refer to the mean values from 5 repetitions for each assessed test condition.

1) Computational Resources Usage

The usage of computational resources was measured for different iterative test conditions with increased complexity to gain insights about the computation costs of adding the different visual elements, content formats and streams in the Social VR experience. These test conditions, along with the obtained values, are summarized in Table II.

As expected, the iterative addition of extra content elements and streams in the session (live presenter, TVM streams) resulted in a higher consumption of CPU, GPU and RAM resources. The overall usage for the full evaluated experience was not that high, resulting in a smooth performance, and still providing some margin to add at least one extra TVM stream for a third user using the same (affordable) PC.

2) Bandwidth consumption

On the one hand, the Z CAM K1 Pro camera used to capture the live presenter can provide an output stream of 4K resolution at 60fps (or alternatively 6K resolution at 30 fps), using H.264 video encoder, with input bitrates up to 30 Mbps, and Advanced Audio Coding (AAC). By setting a resolution of 4K@30fps, an input bitrate of 30 Mbps and an output bitrate (after encoding) of 5 Mbps, the average bandwidth consumption for the incoming RTMP stream at the client side was very close to that latter setting, as expected.

On the other hand, the average bandwidth consumption for each TVM stream providing the end-users' representations (with the settings detailed in the previous sub-section) was 8.83 Mbps (stdv=0.87 Mbps).

These bandwidth consumption values per stream are a bit higher than typical bitrates targets and requirements in High Definition (HD) multi-party 2D videoconferencing (around 4 Mbps, according to [42]), but comparable to the ones in HD video streaming platforms (from 2 Mbps up to 15Mbps, according to [43]), when delivering high definition (HD) resolution video (1920 × 1080) in such scenarios, using similar encoding settings. Taking into account that the streams analyzed in this work carry out immersive media content (stereoscopic and volumetric video), these are reasonable and satisfactory bandwidth requirements.

3) End-to-End delays for the live streams

On the one hand, as a third-party software was used to broadcast the RTMP stream from the live presenter camera, the delays were measured by visually comparing timestamps captured by the camera (pointing at a visual clock counter) with the same ones being displayed at the player side, by placing the clock counter and the player screen side-by-side, and recording a video showing their evolution. That way, by pausing the recorded video at some instants, the end-to-end delay (which in this case is actually the glass-to-glass delay) can be determined by calculating the differences between the timestamps. This method has been used in related works (e.g. [44]). The delays, including the Orchestrator (deployed in another country) as a relay server, the connection between GStreamer and Unity and the background removal process, were in the order of 1.5s, with very low variance.

On the other hand, the delays for the TVM streams were measured by inserting absolute timestamps for each captured frame at the origin side, extracting them prior rendering the frames at the destination side, and synchronizing the involved machines by using Network Time Protocol (NTP) to be able to accurately compute the difference between the rendering and capturing instants. By doing so, the average end-to-end delay for the TVM streams (in this case capture-to-render delay) was 751.57ms (stdv=140.45ms). Unlike for the delays for the live presenter stream, the delays for TVM streams do not include the acquisition and rendering delays (as it is very challenging to visually compare clock counters for this media format and resolution).

Although there was a delay difference between both types of streams (i.e. RTMP stream from the live presenter's representation and TVM stream for the end-users' representations), no inter-media synchronization mechanism was adopted, as it would had involved to delay the TVM streams, and providing highly interactive sessions between the participants was a key goal for the experience. Also, no significant delay differences between the two TVM streams were detected, although neither inter-source nor inter-client media synchronization solutions [36] were adopted in this work to compensate for such potential differences.

TABLE II
COMPUTATIONAL RESOURCES USAGE

Test Condition (TC)	CPU (%)	GPU (%)	RAM (MB)
TC1: 3D VR Scenario + Recorded (2D + 180º) Videos	12.51	15.22	375.24
TC2: TC1 + Live Presenter (with Chroma Keying)	17.86	24.53	550.11
TC3: TC2 + 1 TVM (1 user)	29.21	45.2	850.3
TC4 (Full Experience): TC2 + 2 TVMs (2 users)	42.33	70.5	1170.82

[13] Wireshark, https://www.wireshark.org/ Last Access in September 2021.

However, the delay differences between the audiovisual streams from each participant in the shared session are within the tolerable limits to the human perception for the lack of inter-media and inter-client synchronization, as reported for different media scenarios in [3], [36] and [45], and as confirmed in the user tests, which alleviates the needs for such media synchronization solutions. For instance, the study in [45] investigated the impact of the magnitudes of end-to-end delays and of the delays offsets between participants in multi-party 2D video conferencing services. On the one hand, that study shows that the existence of symmetric end-to-end delays (i.e. same order of magnitudes of delays for all participants) in the order of 1s or 2s for all participants results in a lower QoE compared to when the delays are in the order of 500ms or close to zero ms, but the overall perceived experience is still acceptable and allows for effective and rich communications, especially when one of the participants plays the role of moderator (as it is the case in our study, with the live presenter). On the other hand, that study also shows that the addition of an extra delay of 500ms or 1000ms to one of the participants in a multi-party video conferencing session (i.e. asymmetric delay group setting, with larger delays for one or a sub-group of the participants) may be noticed in some situations, but does not result in severe QoE degradations, neither for moderated sessions nor for sessions with highly active interaction patterns. This reflects the case of interaction between the live presenter (RTMP stream) and the participants (TVM streams) in the analyzed Social VR scenario.

D. Subjective Evaluation: Protocol and Procedure

The evaluation protocol and procedure for the user tests are summarized next.

First, the participants were recruited based on the following three criteria:

- They had to be older than 18 years old.
- They needed to have a good English level (to be able to understand the story).
- They had to know each other (to ensure a fluid and natural social interaction).

Second, the user tests underwent the next steps:

- *Step 1 (~10min)*. The facilitators welcome the participants, and briefly describe them the tests, with the necessary context and its procedure. The participants are also informed that their participation is totally voluntary, and that they can leave the experiment at any time, if they would like to do so for whatever reason.
- *Step 2 (~5min)*. The participants fill in a consent form, a demographic and background information form, and the Simulation Sickness Questionnaire (SSQ) [46].
- *Step 3 (~5min)*. The participants are brought to the lab room. Once arriving there, they are equipped with the HMD and audio headset, with the help of the facilitator(s) if needed. Even though the participants were wearing an HMD during the VR experience, the experiment room had controlled lighting conditions, as recommended in ITU-R BT. 500-13 [47].
- *Step 4 (~10min)*. When all the involved participants and the presenter are ready, the facilitators launch the experience via the interface with the Orchestrator, and the produced Social VR content piece (second episode) is presented to them. Although the duration of VR content episode is about 7min, the participants were instructed to feel free to interact and talk to each other before, during and after the live TV show, and even to explore the designed VR environment at the end of it. The consumption of the produced stimuli together with the interactions between participants thus took around 10-15min, which is aligned with recommended durations of VR experiences to induce adequate immersion, while still avoiding the appearance of simulation sickness and other related symptoms [48-52], including the recommendations from ITU-T-P.809 [53]. The participants were standing during the experience (Fig. 1).
- *Step 5 (~3min)*. With the help of the facilitator(s), the participants step out of VR, and are brought to a meeting room with a round table.
- *Step 6 (~15min)*. The participants will fill in the SSQ questionnaire and the Experience questionnaire for Social VR designed in [17], slightly adapted by re-phrasing the question items according to the evaluated experience (see Tables III-VII). That questionnaire was constructed due to the nature of this novel medium (Social VR) and the lack of resources to properly evaluate it, and it addresses three main dimensions: quality of interaction; social meaning; and presence & immersion. The associated question items for each dimension have been selected and adapted from a wide variety of state-of-the-art questionnaires designed for both traditional and immersive media applications and services [17], which in turn also take into account two main ITU recommendations: 1) ITU-T P.911 “Subjective audiovisual quality assessment methods for multimedia applications” [54], including the evaluation procedure, methods and scales as well as the number of participants in the test); and 2) ITU-T P.1305 “Effect of delays on telemeeting quality” [55], including the analysis of the impact of the delays on the perceived experience.
- *Step 7 (~15min)*. The facilitators drive a semi-structured interview to discuss about the Social VR technology, the experience itself and other potential applicability scenarios with the participants of each session.
- *Step 8 (~2min)*. Participants are thanked, given a voucher of 30 euros, and said goodbye.

Overall, each test session took between 60 and 75 minutes.

E. Subjective Evaluation: Sample of Participants

Overall, 40 participants took part in the tests, which is aligned with the recommendations from ITU-T P.911 [54]. Next, background information about them is provided:

- Aged between 18 and 60 years (average = 31, standard deviation = 11.61).
- 28 males and 12 females.
- 1 participant was left handed, 38 were right handed, and 1 was ambidextrous.
- None of the participants expressed to have audio-visual impairments.

The participants were also asked about their skills using computers and their previous experience in VR:

TABLE III

SOCIAL VR EXPERIENCE QUESTIONNAIRE – “QUALITY OF INTERACTION (QI)” PART

Question	Totally Disagree	Partially Disagree	Neutral	Partially Agree	Totally Agree
QI1. “I was able to feel the other users’ emotions in the virtual shared experience.”	0	1 (2.5%)	13 (32.5%)	**22 (55%)**	4 (10%)
QI2. “I was sure that the other users often felt my emotion.”	0	1 (2.5%)	**23 (57.5%)**	11 (27.5%)	5 (12.5%)
QI3. “The virtual experience with the other users seemed natural.”	0	3 (7.5%)	14 (35%)	**20 (50%)**	3 (7.5%)
QI4. “The actions used to interact with the other users were similar to the ones in the real world.”	0	5 (12.5%)	10 (25%)	**18 (45%)**	7 (17.5%)
QI5. “It was easy for me to contribute to the conversation.”	0	0	4 (10%)	16 (40%)	**20 (50%)**
QI6. “The conversation with the other users seemed highly interactive.”	0	0	6 (15%)	**22 (55%)**	11 (27.5%)
QI7. “I could readily tell when the other users were listening to me.”	16 (40%)	**18 (42.5%)**	3 (12.5%)	1 (2.5%)	0
QI8. “I found it difficult to keep track of the conversation.”	16 (40%)	**18 (42.5%)**	3 (12.5%)	1 (2.5%)	0
QI9. “I felt completely absorbed in the conversation.”	0	0	9 (22.5%)	**21 (52.5%)**	10 (25%)
QI10. “I could fully understand what the other users were talking about.”	0	0	1 (2.5%)	**20 (50%)**	19 (47.5%)
QI11. “I was very sure that the other users understood what I was talking about.”	0	0	3 (7.5%)	**24 (62.5%)**	12 (30%)
QI12. “I often felt as if I was all alone in the virtual shared experience.”	17 (42.5%)	**22 (45%)**	1 (2.5 %)	0	0
QI13. “I think the other users often felt alone in the virtual shared experience.”	17 (42.5%)	**20 (50 %)**	3 (7.5%)	0	0

TABLE IV

SOCIAL VR EXPERIENCE QUESTIONNAIRE – “SOCIAL CONNECTEDNESS (SC)” PART

Question	Totally Disagree	Partially Disagree	Neutral	Partially Agree	Totally Agree
SC1. “I often felt that the other users and I were together in the same space.”	0	0	2 (5%)	**25 (62.5%)**	13 (32.5%)
SC2. “I paid close attention to the other users.”	0	2 (5%)	12 (30%)	**19 (47.5%)**	7 (17.5%)
SC3. “The other user was easily distracted when other things were going on around us.”	0	4 (10%)	11 (27.5%)	**19 (47.5%)**	6 (15%)
SC4. “I felt that the having the VR experience together enhanced our closeness.”	0	2 (5%)	7 (17.5%)	**25 (62.5%)**	6 (15%)
SC5. “Having the VR experience together created a good shared memory between us.”	0	1 (2.5%)	6 (15%)	**25 (62.5%)**	8 (20%)
SC6. “I derived little satisfaction from the virtual shared experience.”	4 (10%)	**16 (40%)**	**16 (40%)**	4 (10%)	0
SC7. “The virtual shared experience with my partner felt superficial.”	5 (12.5%)	**18 (45%)**	16 (40%)	1 (2.5%)	0
SC8. “I really enjoyed the time spent with the other users.”	0	**0**	1 (2.5%)	**24 (60%)**	15 (37.5%)
SC9. “In the virtual world I had a sense of ‘being there’.”	0	0	5 (12.5%)	**24 (60%)**	11 (27.5%)
SC10. “Somehow I felt that the virtual world was surrounding me and my partner.”	0	0	4 (10%)	**27 (67.5%)**	9 (22.5%)
SC11. “I had a sense of acting in the virtual space, rather than operating something from outside.”	0	1 (2.5%)	11 (27.5%)	**22 (55%)**	7 (17.5%)
SC12 “My virtual shared experience seemed consistent with a real world experience.”	0	0	15 (37.5%)	**20 (50%)**	5 (12.5%)
SC13. “I did not notice what was happening around me in the real world.”	0	2 (5%)	10 (25%)	**16 (40%)**	12 (30%)

TABLE V
SOCIAL VR EXPERIENCE QUESTIONNAIRE – "PRESENCE / IMMERSION (PI)" PART

Question	Totally Disagree	Partially Disagree	Neutral	Partially Agree	Totally Agree
PI1. "I felt detached from the outside world while having the VR experience."	0	2 (5%)	9 (22.5%)	**19** **(47.5%)**	10 (25%)
PI2. "At the time, the shared VR experience with the other users was my only concern."	0	3 (7.5%)	13 (32.5%)	**14** **(35%)**	10 (25%)
PI3. "Everyday thoughts and concerns were still very much on my mind."	5 (12.5%)	11 (27.5%)	**17** **(42.5%)**	6 (15%)	1 (2.5%)
PI4 "It felt like the VR shared experience took shorter time than it really was."	0	1 (2.5%)	4 (10%)	**22** **(55%)**	13 (32.5%)
PI5. "When having the VR experience together, time appeared to go by very slowly."	10 (25%)	**17** **(42.5%)**	11 (27.5%)	2 (5%)	0

TABLE VI
SOCIAL VR EXPERIENCE QUESTIONNAIRE – EXTRA AD-HOC QUESTIONS (AQ)

Question	Totally Disagree	Partially Disagree	Neutral	Partially Agree	Totally Agree
AQ1. "I liked the created VR content and scenario."	0	0	1 (2.5%)	17 (40%)	**23** **(57.5%)**
AQ2. "The created VR content and scenario are realistic."	0	0	2 (5%)	**29** **(72.5%)**	9 (22.5%)
AQ3. "The spatiality in the VR scenario (i.e. perceived distances and sizes of elements, including the participants' bodies) is consistent with a real-life scenario."	0	0	8 (20%)	**22** **(55%)**	10 (25%)
AQ4. "Having more than 2 users in a shared virtual environment can provide added-value to the social VR experience"	0	0	3 (7.5%)	**22** **(55%)**	15 (37.5%)
AQ5. "Having a remote presenter / actor in real-time provides added-value to the social VR experience"	0	0	8 (20%)	**22** **(55%)**	10 (25%)

- 1 participant stated to be novice, 16 intermediate and 23 experts regarding the use of computers.
- 10 participants stated not having previous experience in VR, 25 affirmed to have some experience, and 5 of them expressed to be very experienced.

F. Subjective Evaluation: Results from Questionnaires

In this sub-section, the results from the used questionnaires are presented.

1) SSQ Questionnaire

With regard to the results from SSQ, no significant effects / symptoms were noticed to be caused by the VR experience.

2) Social VR Experience Questionnaire

The Social VR experience questionnaire includes question items categorized to assess four relevant aspects to be answered using a 5-level likert scale [17] (Tables III-VI), with the potential answers detailed in Tables III-VI):

- Quality of interaction (Table III): including emotional experience, quality of the communication, and naturalness of the communication.
  From the results of Table III, it can be affirmed that the presented Social VR platform and experience provided a satisfactory quality of interaction to the participants. This is mainly supported by the highly positive scores for the items related to the naturalness and understanding of the conversations, and to the feeling of not being alone in the VR environment. Likewise, participants stated that the conversations were highly interactive and that they could contribute to such conversions effortlessly. This reflects that the magnitudes of the end-to-end delays for the involved streams are satisfactory (Section V.D).
- Social connectedness (Table IV): including feeling of togetherness, emotional closeness, and enjoyment of the relationship.
  From the results of Table IV, it can be affirmed that the presented Social VR platform and experience provided a satisfactory social connectedness to the participants. This is mainly supported by the highly positive scores for the items related to the feeling of being together in the same space, low level of distraction by "real world" issues, and having enjoyed the shared experience.
- Presence / Immersion (Table V): including mainly plausibility and place illusion.
  From the results of Table V, it can be affirmed that the presented Social VR platform and experience provided a satisfactory level of immersion / presence, with most of the participants stating to having felt detached from the real world, engaged with the VR story, and declaring to have had the feeling that the experience took shorter than its real duration.
- Additional ad-hoc aspects about the experience (Table VI): including level of realism, how much the content likes to the users, etc.
  From the results of Table VI, it can be affirmed that the participants liked the created content and the whole experience very much, and rated the experience as realistic and immersive. Interestingly, participants were especially surprised and satisfied with the ability to interact with other realistic volumetric users and a video-based presenter,

which are two of the key innovations of the presented platform.

Given the non-appearance of simulation sickness effects (through SSQ), the reported satisfactory immersion levels (Table V, especially question item PI5 with regard to the perceived evolution of time) and the explicit answers indicating that the experienced liked very much the participants (Table VI), combined with the fact that none of them shown concerns about the duration of the VR experience, it can be concluded that the duration was appropriate, which is in line with the existing recommendations in literature [48-52].

G. Subjective Evaluation: Results from Interviews

Finally, the pairs from each session participated in a semi-structured interview with the experiment facilitators. The interview was driven by a series of questions, whose answers are detailed next, although participants were also encouraged to express other impressions and/or concerns caused by the Social VR experience. The interviews took around 15-20min per each pair. The audio recordings of the interviews were transcribed and coded, following an open coding approach [55]. Since the interviews were conducted with the two participants for each pair together, their answers were transcribed and coded as a participant pair, not as individual participants. Therefore, the 20 participant pairs are hereafter labelled as P1-P20. From the coded transcripts, relevant aspects and insights were observed, which are further elaborated next.

1) Benefits and Potential of Social VR

All participants thought that the Social VR platform enabled them to experience *social presence*. First, they felt identified with the end-users' representations, both with their own and the other's representations. "*The quality is not great, but it is impressive to see yourself and your partner as part of the VR environment, in a volumetric representation*", P12 said. "*I could even see my watch / the pictures on my T-shirt*", participants from P3 and P11 stated. A few participants also pointed out that although the end-users' reconstructions provide natural interactions (50%), the facial expressions were partially blocked by the visual quality and the HMD occlusion (30%).

The participants generally felt *being together* with the other participant, which enriched the overall experience. P2 and P4 stated "*We felt together, sharing an experience, and this is really an added value to VR!*". P7 mentioned: "*We were aware of the activities and feelings of the other participant*". The fact of being standing and close to each other was well received by participants, as explicitly stated by P3 and P14. However, the short distance between participants also influenced the noticeability of the visual artefacts for the end-users' representations. This was pointed out by the majority of participants (70%). Three pairs (P5, P6, and P16) claimed: "*Having your colleague closer is great, but then it is easier to realize of the limitations in the visual quality of her/his representation*". P18 said: "*When your partner is closer, it also becomes clearer that the she/he is wearing the HMD, and thus that you cannot see her/his face*". Participants generally expressed that having eye contact is important, but that the lack of it – because of the HMD blocking – is not a major barrier for a rich interaction and enjoyable experience (50%).

The participants also found the VR environment and the created content immersive and realistic. P1, P3, P7, P19 said "*The TV set was realistic and consistent with the real world*". "*The high quality and realism of the VR environment help you to feel immersed in the experience, and part of the story*", P3 and P19 added. "*The presenter was talking to and pointing at you. This makes you feeling part of the story*", stated by many pairs. "*This is like being inside and being part of a TV program!*", P9 and P15 highlighted. "*The presenter and reporter looked very well integrated in the TV set. You felt like if you were where the news are actually happening*", stated by P3 and P13.

The participants in general felt comfortable in the virtual environment. "*As the experience is not too long, a standing posture gives the feeling of higher freedom and that you can move around*", P5 and P14 said. A few participants (10%) mentioned having felt a bit tense at the first contact with the Social VR platform, because of the uncertainty, but then they rapidly felt more relaxed.

Besides the feeling of immersion and social presence, the quality of communication was found satisfactory in general. Even though the visual quality for the end-users' representation has room for improvement, being able to see themselves in VR was a fascinating feature for the participants. "*The quality of visual communication between us was not high, but it was a fascinating feature to see my full body and clothing, as well as my pair inside the virtual world*", P4, P9 and P20 stated. "*Despite of noticeable artefacts and not so fluent movements, we could fully and easily recognize ourselves*", P8 and P10 mentioned. "*The quality of my partner's representation seemed better than mine*", stated by P2 and P10. "*The delays for the end-users' reconstruction was noticeable for some gestures, but it was not a barrier for an effective communication*", stated P3. The quality of the audio communication and the spatial audio effects were perceived as satisfactory by the participants, and good enough to feel immersed. "*You could perceive the spatial audio effects, especially when different speakers from different positions were active at different times*", P2 and P13 stated. In general, the interactions between the participants were perceived as natural. "*The interaction was natural, but it is not identical as in real life scenarios: you're wearing an HMD with cables, and you're experiencing a novel medium, not so common for us yet*", P15 stated. Around 90% of the participants stated that the audio-visual interactions enabled them to sense the emotions of their partners to a certain degree. "*We were able to feel the emotions and our excitement*", P6 stated. "*You don't have a full sensing of the emotions, but you can infer them from the audio communications and visual gestures*", stated by P9, P11 and P18. "*It is not always possible to tell the emotions from the expressions, especially when you cannot see the faces*", stated by P1 and P16.

All participants believed that the photo-realistic representations for the end users can help maintain, strength, and even create new, relationships in life. P3, P7 and P12 stated "*It is a very innovative and useful solution. We have friends and* *family members living apart. This would enable us to meet and share experiences, overcoming distance barriers, and saving time*". In general, participants believe that these systems can be applied to interact with both known people and new contacts, although the use of avatars was also considered convenient for the latter cases, especially when personal relationships are not so important, to overcome shyness, and/or to provide a higher privacy. Suggested applicability use cases for this Social VR technology are enumerated later.

Many participants (35%) affirmed it was an amazing experience for them, and that Social VR can be a powerful tool to evade from the real world in certain situations (20%).

2) Missing aspects / Weaknesses in Social VR

Most participants (90%) would like to be provided with an improved visual quality for the end-users' representations. Having more fluid movements (i.e. higher frame rates) was mentioned by 50%, and having faster reactions (i.e. lower delays) was mentioned by 35% of the participants, as aspects to be improved in the future. The limitations related to the visual quality of the end-users' representations have been already mentioned, so the lack of higher quality for this was also identified as a missing aspect. "*I felt identified with my self-representation, and also could easily recognize my partner. But I know him. This level of quality might not suffice when using the platform to meet with unknown people or for professional use cases*", as stated by P4. "*The quality of the end-users' representation should improve in the future*", declared by P5, P11, P13 and P18.

Some participants pointed out that the integration of multi-sensory stimuli, like scents (10%) and especially haptic feedback (75%), was a missing aspect in the presented Social VR platform. P4, P10 and P13 "*It would be great if you could touch things, and if the haptic interactions indeed have an effect on the VR environment or story*".

80% of participants would like to move freely in VR (e.g., 6DoF). "*It would be great if you could move around, get closer to other elements and participants in the shared environment*", stated by P2 and P11. "*If you can move close to each other, then the interactions could be richer; you could e.g. see more details of the emotions and gestures*", P18 stated.

With the combinations of haptic feedback and 6DoF features, participants mainly pursue enjoying more interactive and active experiences. "*If you can actively explore things and complete tasks together, as well as influence the VR environment, then you would be able to really enjoy an interactive and collaborative experience*", P3 remarked. "*The possibility to explore the environment and interact with it would largely increase the immersion*", mentioned P6 and P20.

3) Potential Use Cases

In general, the participants foresee a big impact of Social VR. When asked about the most interesting use cases for Social VR according to them, the answers were: virtual meetings and consultation (85%), training (65%), virtual events (60%) - like conferences, fairs and religious events -, gaming (60%), shared video watching (30%), co-creation spaces (30%) and dating (20%). In the case of virtual events, 20% of participants remarked that Social VR can become a powerful tool and medium to plan these events, to experience with the organization and distribution of spaces, furniture, presentation rooms, etc. In these kinds of events, participants highlighted that Social VR can contribute to increase the audiences, because there is no need to travel, thus also contributing to accessibility, to reduce pollution, and to save time and costs. Some participants (15%) also identified Social VR as an ideal tool for migrants and to connect with known people living far away (30%), while others (15%) showed concerns about the duration of the Social VR experiences. "*If the experience is not too long, then Social VR can work. But for longer experiences, you may get tired and dizzy. HMDs should become more lightweight and comfortable*".

In general, participants believed that Social VR is a powerful medium to meet with known users, but also to meet new contacts. Most of the participants (90%) declared their willingness to use Social VR in the future. "*I want this at home!*" stated by P8. "*This can be seen as the next generation Skype*", stated by P11. Many participants (25%) stated that the virtual interactions can be very intense and effective and that they are a good alternative especially for first contacts. A few participants (10%) thought that Social VR is more adequate in corporate environments, and not yet for domestic environments. Other ones (10%) shown concerns about Social VR contributing to sedentariness.

All participants agreed that being able to interact with elements of the VR environment, like the live presenter, provides added value. "*You can actually interact with a presenter, or alternatively an instructor, and your conversation influences the evolution of the session. It really provides added value, as you are not just a passive watcher*", stated by P7 and P12. Most of them (90%) also think that supporting more than 2 participants is beneficial and interesting. The rest affirmed that two-person meetings could be just enough in specific use cases, and provide rich interactions.

4) Next Generation of Social VR

Finally, participants were asked about their vision for next generation Social VR systems. Most of them envisioned Social VR as futuristic environments where the boundaries between the real and the virtual worlds are blurred (P2, P4, P9, P12 and P17), under the umbrella of eXtended Reality (XR). P9 and P17 envisioned: "*A hybrid space where the real and virtual worlds are seamlessly mixed, with virtual elements augmenting the reality and detailed information about certain real elements, as well as multi-sensory stimuli, are provided*". P3 and P20 stated "*Virtual worlds where you can freely move around, and be tele-transported to the places of your choice or need*". P5 stated "*multi-user gatherings with real and virtual users, where you can hardly distinguish between the virtual and real ones, or that at least the quality of the virtual users does not impact the overall experience*". 10% of participants also pointed out that the application of Artificial Intelligence (AI) techniques can bring added value to next generation Social VR systems, emphasizing its potential adoption in training scenarios.

VI. Discussion

This paper has presented an innovative Social VR platform that is able to seamlessly present and blend heterogeneous media formats and to integrate in real-time remote participants in shared virtual environments, both represented as volumetric TVMs and as video billboards (Chroma keying). The platform provides many outstanding and more complete features compared to state-of-the-art solutions, in terms of media and interaction capabilities. The paper has also described a professionally produced TV show-like VR story that has been used to demonstrate the platform's capabilities and to assess both its performance and user experience related aspects through an experiment involving 20 pairs of users.

On the one hand, the obtained results from objective tests reveal that the platform performs satisfactorily for sessions integrating various content modalities, a pair of participants and one live presenter, when using off-the-shelf hardware components. The magnitudes of bandwidth requirements are reasonable for current-day Internet connections and result in acceptable audiovisual quality levels, while the magnitudes of delays for the exchanged streams are comparable to those in HD multi-part scenarios, and still provide high quality of interaction and satisfactory QoE levels. Likewise, the experience runs smoothly in VR ready laptops, with are becoming affordable in these days. These results are valuable for use cases in which no more than 2-3 users are required (e.g., watching TV in VR together, one-to-one meetings, gaming, etc.). On the other hand, the obtained results from the user tests have proved that the Social VR experience (platform plus the produced content) provides satisfactory quality of interaction, immersion and togetherness levels, and that these experiences awake high interest. These results confirm the potential impact that both the proposed innovative features and the ideated Social VR scenarios can bring to the media and broadcast sector(s), thus also validating and/or shedding key light on the research hypotheses about the expected benefits to be provided by the developed and integrated features in terms of immersion, realism, quality of interaction, togetherness, and interaction capabilities, overcome key limitations of state-of-the-art contributions, by using a lightweight and low-cost platform.

With regard to its applicability, the paper has conceptualized how certain future TV and broadcast services could look like, integrating immersive and traditional formats and enabling new forms of interactions, going a step beyond currently existing Social VR platforms and commercial experiences (e.g. Fox Sports). By using this novel technology and medium, the remote audience can become active participants inside TV events, being no more outside passive spectators. They can also feel together and interact with the usual participants of the TV event, like the presenter(s), who can also join the shared experience from remote locations. Thus, the proposed experience goes one-step beyond current watch-together TV scenarios, enabling new be-together-in TV scenarios where there is still an unlocked potential in terms of technological, creative and commercial levels. Besides, the demonstrated use case has awakened a high interest to the participants, anticipating a potential positive impact of this technology in the broadcast and media ecosystems. Even though the tests were conducted in February 2020, before the COVID-19 resulting in a lockdown in Spain (and worldwide), the participants already foresaw many other user cases in which Social VR can provide valuable benefits, like training, virtual meetings and consultation, and virtual events. Although having obtained very satisfactory and promising results, it is firmly believed that the ratings related to user experience aspects, provided benefits and potential impact in other use cases would had been even more positive if the tests had been conducted after the COVID-19 out there, when the use of digital communication tools has been magnified, as well as their limitations when it comes to a natural and realistic communication, interaction and collaboration.

Certainly, the current platform and the provided experience have limitations in terms of both technological and creative aspects. When it comes to technical aspects, additional work is necessary to scale up the number of live video feeds and volumetric users to recreate more massive TV show scenarios in a more realistic manner. Besides, the quality of the volumetric end-users' representations needs to improve (higher resolutions and frame rates) to provide commercially acceptable solutions. So far, the current bottleneck to scale up in terms of number of participants is on the computational needs to render each volumetric user representation at the client side. Additionally, although the server based components in the presented platform mainly perform orchestration, session management and stream relay features, such components or additional ones could also contribute to enhanced the scalability of the system, and reduce the burden at the client side [56]. Due to the implementation of non-resource intensive functionalities in the current version of the platform, no performance indicators have been reported for the server components in this paper.

When it comes to production and scenario-related aspects, the addition of extra interaction features would provide added value. This includes the availability of higher degrees of freedom (e.g. 6DoF), the chance of manipulating the virtual environment and influence the storyline via user's actions and behaviors, and the integration of multi-sensory stimuli, like haptic feedback.

All these limitations are however an opportunity to perform further research in the field of Social VR, which has been proven to offer a new way of telling stories and to bring up distributed users together in an immersive and interactive manner. Among others, this can open new opportunities in the media broadcast and Over-the-Top (OTT) sectors.

VII. Conclusions and Future Work

Social VR is expected to have a big impact in the near future. This work has presented an innovative and lightweight platform that provides key outstanding features. First, it allows a real-time integration of remote users in shared virtual environments, by using (photo-)realistic volumetric representations and affordable capturing systems, and thus having the chance of avoiding the use of synthetic avatars, and by using video-based billboard representations. Second, it support a seamless integration of heterogeneous immersive media formats, including 3D scenarios, dynamic volumetric representation of users and (stored and live) stereoscopic 2D traditional and 180º/360º videos. Third, it provides two main types of interaction features, like low-latency interaction channels between the users and with the presenter, and a dynamic control of the media playout to adapt to the session's evolution.

The Social VR platform has been evaluated for a live broadcast use case, by having recreated a TV show experience, and having obtained very satisfactory results, in terms of performance, computational and bandwidth requirements, user experience, and awakened interest. The evaluations have also shed some light on aspects to improve and on next steps to maximize the impact.

In particular, future work will be focused on four key aspects. First, the system's performance, including the delays and the visual resolution of the volumetric user's representations, will be continuously improved. Second, it is planned to perform a comparison between: i) the presented platform and other existing ones and with baseline conditions; ii) different type of capturing sensors (e.g. RealSense vs Kinect) and setups (e.g. single-sensor vs multi-sensor); and iii) TVMs and other representation formats, like Point Clouds [57]. Third, it is planned to investigate the impact of the number of users in terms of performance and scalability issues, but also on the perceived experience. Finally, the platform will be evaluated for other use cases, including the ones suggested by the users in the interviews, like multi-party conferencing / meetings.

ACKNOWLEDGMENT

This work has been partially funded by the European Union's Horizon 2020 program, under agreement nº 762111 (VR-Together project), and partially by ACCIÓ, under agreement COMRDI18-1-0008 (ViVIM project). Work by Mario Montagud has been additionally funded by the Spanish Ministry of Science, Innovation and Universities with a Juan de la Cierva – Incorporación grant (Ref. IJCI-2017-34611). Work by David Rincón has been additionally supported by the Agencia Estatal de Investigación of Spain of Ministerio de Ciencia e Innovación of Spain under project PID2019-108713RB-C51 MCIN/ AEI /10.13039/501100011033. The authors would also like to thank the EU H2020 VR-Together project consortium for their relevant and valuable contributions.

REFERENCES

[1] P. Cesar, D. Bulterman, J. Jansen, "Usages of the Secondary Screen in an Interactive Television Environment: Control, Enrich, Share, and Transfer Television Content", In: Tscheligi M., Obrist M., Lugmayr A. (eds) Changing Television Environments. EuroITV 2008. Lecture Notes in Computer Science, vol 5066. Springer, Berlin, Heidelberg, 2008.

[2] F. Boronat, D. Marfil, M. Montagud, J. Pastor, "HbbTV-Compliant Platform for Hybrid Media Delivery and Synchronization on Single and Multi-Device Scenarios", IEEE Transactions on Broadcasting, 64(3), pp. 721-746, Sept. 2018.

[3] D. Marfil, F. Boronat, M. Montagud, A. Sapena, "IDMS Solution for Hybrid Broadcast Broadband Delivery within the context of HbbTV standard", IEEE Transactions on Broadcasting, 65(4), pp. 645-663, December 2019.

[4] F. Boronat, M. Montagud, P. Salvador, J. Pastor, "Wersync: a web platform for synchronized social viewing enabling interaction and collaboration", JNCA, 2020

[5] P. Cesar, D. Geerts, "Past, present, and future of social TV: A categorization", IEEE Consumer Communications and Networking Conference (CCNC), Las Vegas (USA), pp. 347-351, 2011

[6] F. Boronat, D. Marfil, M. Montagud, J. Pastor, "Hybrid Broadcast/Broadband TV Services and Media Synchronization. Demands, Preferences and Expectations of Spanish Consumers", IEEE Transactions on Broadcasting, 64(3), pp. 52-69, August 2017

[7] J. A. Núñez, M. Montagud, I. Fraile, D. Gómez, S. Fernández, "ImmersiaTV: an end-to-end toolset to enable customizable and immersive multi-screen TV experiences", Workshop on Virtual Reality, co-located with ACM TVX 2018, Seoul (South Korea), June 2018

[8] M. Montagud, O. Soler, I. Fraile, S. Fernández, "VR360 Subtitling: Requirements, Technology and User Experience", IEEE ACCESS, 2020.

[9] J. Li, V. Vinayagamoorthy, R. Schwartz, W. IJsselsteijn, D.A. Shamma, P. Cesar, "Social VR: A New Medium for Remote Communication & Collaboration," in Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems (ACM CHI 2020), Honolulu (USA), April 2020

[10] S. Gunkel, M. Prins, H. Stokking, O. Niamut, "Social VR Platform: Building 360-degree Shared VR Spaces", ACM TVX 2017, Hilversum (The Netherlands), June 2017

[11] S. Beck, A. Kunert, A. Kulik, B. Froehlich, "Immersive group-to-group telepresence", IEEE Transactions on Visualization and Computer Graphics, 19, 4, 616–625, 2013

[12] P. Heidicker, E. Langbehn, F. Steinicke, "Influence of avatar appearance on presence in social VR", IEEE Symposium on 3D User Interfaces (3DUI) 2017, 233–234, Los Angeles (CA, USA), March 2017

[13] M. Cavallo, M. Dholakia, M. Havlena, K. Ocheltree, M. Podlaseck, "Dataspace: A Reconfigurable Hybrid Reality Environment for Collaborative Information Analysis", IEEE Conference on Virtual Reality and 3D User Interfaces (VR) 2019, Osaka (Japan), March 2019.

[14] D.J. Roberts et al., "withyou—An Experimental End-to-End Telepresence System Using Video-Based Reconstruction", IEEE Journal of Selected Topics in Signal Processing, vol. 9, no. 3, pp. 562-574, April 2015

[15] S. Orts-Escolano, et al., "Holoportation: Virtual 3d teleportation in real-time", 29th ACM Annual Symposium on User Interface Software and Technology (UIST'16), 741–754, Tokyo (Japan), October 2016

[16] H. Galvan Debarba, M. Montagud, S. Chagué, J. Lajara, I. Lacosta, S. Fernandez, C. Charbonnier, "Content Format and Quality of Experience in Virtual Reality", (Under Review in Multimedia Tools and Applications Journal), May 2020, arXiv:cs.MM/2008.04511

[17] J. Li, et al., "Measuring and Understanding Photo Sharing Experiences in Social Virtual Reality", ACM CHI 2019, Glasgow (UK), May 2019

[18] P. Cesar, D. Geerts, "Understanding Social TV: a survey", Networked and Electronic Media Summit 2011, 27–29, Turin (Italy), September 2011.

[19] E. M. Huang, el al. "Of social television comes home: a field study of communication choices and practices in tv-based text and voice chat", ACM CHI 2009, 585–594, Boston MA (USA), April 2009.

[20] D. Geerts, et al., "Are we in sync?: synchronization requirements for watching online video together", ACM CHI 2011, 311–314, Vancouver (Canada), May 2011.

[21] M. McGill, J. H. Williamson, S. Brewster, "Examining the role of smart TVs and VR HMDs in synchronous at-a-distance media consumption", ACM Transactions on Computer-Human Interaction (TOCHI), 23, 5, 33, 2016.

[22] R. A. J. de Belen, et al., "A systematic review of the current state of collaborative mixed reality technologies: 2013–2018", AIMS Electronics and Electrical Engineering, 3(2), 181-223, May 2019

[23] Z. Zhang. Microsoft Kinect sensor and its effect. IEEE MultiMedia, 19(2):4–10, Apr. 2012.

[24] A. J. Fairchild, et al, "A mixed reality telepresence system for collaborative space operation", IEEE Transactions on Circuits and Systems for Video Technology, 27, 4, 814–827, 2016.

[25] S. Rothe, A. Schmidt, M. Montagud, D. Buschek, H. Hußman, "Social viewing in cinematic virtual reality: a design space for social movie applications", Virtual Reality, October2020

[26] A. Yaqoob, T. Bi, G. M. Muntean, "A Survey on Adaptive 360° Video Streaming: Solutions, Challenges and Opportunities", IEEE Communications Surveys & Tutorials, vol. 22, no. 4, pp. 2801-2838, 2020

[27] D. S. Alexiadis, et al., "An integrated platform for live 3D human reconstruction and motion capturing", IEEE Transactions on Circuits and Systems for Video Technology, 27, 4, 798–813, 2016.

[28] V. Sterzentsenko, et al., "A low-cost, flexible and portable volumetric capturing system". 14th International Conference on Signal-Image Technology & Internet-Based Systems (IEEE SITIS 2018), pp. 200-207, Las Palmas de Gran Canaria (Spain), November 2018

[29] M. Montagud, P. Cesar, "Deliverable D4.6: Technical Report on Third Pilot", EU H2020 VR-Together project, https://vrtogether.eu/, December 2020.

[30] F. Endres, et al, "3-D mapping with an RGB-D camera", IEEE transactions on robotics, 30, 1, 177-187, 2013.

[31] E. Lachat, et al., "First experiences with Kinect v2 sensor for close range 3D modelling", The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 40, 5, 93, 2015.

[32] L. Keselman, J. I. Woodfill, A. Grunnet-Jepsen, A. Bhowmik, "Intel realsense stereoscopic depth cameras", IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1-10, 2017

[33] Z. Karni, C. Gotsman, "Spectral compression of mesh geometry", 27th annual conference on Computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co., 279–286, 2000

[34] J. Peng, C-S. Kim, C-C Jay Kuo, "Technologies for 3D mesh compression: A survey", Journal of Visual Communication and Image Representation, 16, 6, 688–733, 2005.

[35] M. F. Ursu, et al., "Orchestration: TV-like mixing grammars applied to video-communication for social groups", 21st ACM international conference on Multimedia (ACM MM 2013), 333–342, Barcelona (Spain), October 2013.

[36] M. Montagud, P. Cesar, J. Jansen, F. Boronat (Eds.), "Handbook on Multimedia Synchronization" (23 chapters), Springer-Verlag, ISBN 978-3-319-65840-7, 2018.

[37] D. Geerts, P. Cesar, D. Bulterman, "The Implications of Program Genres for the Design of Social Television Systems", 1st International Conference on Designing Interactive User Experiences for TV and Video (UXTV '08), 71-80, Silicon Valley (California, USA), 2008

[38] A. Revilla, I. Lacosta, G. Calahorra, "Deliverable D4.3: Second example of content", EU H2020 VR-Together project, https://vrtogether.eu/, July 2019.

[39] M. Montagud, P. Cesar, "Deliverable D4.4: Technical Report on Second Pilot", EU H2020 VR-Together project, https://vrtogether.eu/, July 2020.

[40] K. Christaki, et al., "Subjective Visual Quality Assessment of Immersive 3D Media Compressed by Open-Source Static 3D Mesh Codecs", 25th International Conference on Multimedia Modeling (MMM 2019), Thessaloniki (Greece), January 2019.

[41] M. Montagud, J. Antonio De Rus, R. Fayos-Jordán, M. Garcia-Pineda J. Segura-Garcia, "Open-Source Software Tools for Measuring Resources Consumption and DASH Metrics", ACM MMSYS 2020, Istanbul (Turkey), June 2020.

[42] M. Schmitt, J. Redi, P. Cesar, D. Bulterman, "1Mbps is enough: Video quality and individual idiosyncrasies in multiparty HD video-conferencing", Eighth International Conference on Quality of Multimedia Experience (QoMEX), 2016

[43] C. Lee, S. Woo, S. Baek, "Bitrate and Transmission Resolution Determination Based on Perceptual Video Quality", 10th International Conference on Information, Intelligence, Systems and Applications (IISA), 2019

[44] M. Montagud, F. Boronat, P. Cesar, "A customizable open-source framework for measuring and equalizing e2e delays in shared video watching", ACM TVX 2014, Newcastle (UK), June 2014.

[45] G. Berndtsson, et al. "Methods for Human-Centered Evaluation of MediaSync in Real-Time Communication", In: M. Montagud, et al. (eds) "MediaSync: Handbook on Media Synchronization", Springer, 2018

[46] R. S. Kennedy, N. E. Lane, K. S. Berbaum, Mi. G. Lilienthal, "Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness", The International Journal of Aviation Psychology, 3:3, 203-220, 1993.

[47] ITU Recommendation ITU-R BT. 500-13, "Methodology for the subjective assessment of the quality of television pictures", 2012

[48] S. Sharples, S. Cobb, A. Moody, J.R. Wilson, "Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems", Displays, 29, 58–69, 2008

[49] M. Melo, J. Vasconcelos-Raposo, M. Bessa, "Presence and cybersickness in immersive content: Effects of content type, exposure time and gender", Computers & Graphics, Volume 71, 159-165, 2018

[50] C. Zhang, A. S. Hoel, A. Perkis, S. Zadtootaghaj, "How Long is Long Enough to Induce Immersion?", 10th Int. Conf. Quality Multimedia Exp. (QoMEX 2018), Sardinia (Italy), May 2018,

[51] P. Kourtesis, S. Collina, L.A.A. Doumas, S.E. MacPherson, "Validation of the Virtual Reality Neuroscience Questionnaire: Maximum Duration of Immersive Virtual Reality Sessions Without the Presence of Pertinent Adverse Symptomatology", Front. Hum. Neurosci, 13:417, 2019.

[52] ITU-T Recommendation P.809, "Subjective Evaluation Methods for Gaming Quality", International Telecommunication Union, Geneva, June 2018.

[53] ITU-R Recommendation P.911, "Subjective audiovisual quality assessment methods for multimedia applications", 1998.

[54] ITU-T Recommendation P.1305 "Effect of delays on telemeeting quality", 2016.

[55] D. R. Thomas. 2006. A general inductive approach for analyzing qualitative evaluation data. American journal of evaluation 27, 2 (2006), 237–246.

[56] G. Cernigliaro, M. Martos, M. Montagud, A. Ansari, S. Fernández, "PC-MCU: Point Cloud Multipoint Control Unit for Multi-user Holoconferencing Systems", ACM NOSSDAV 2020, Istanbul (Turkey), June 2020

[57] J. Jansen, S. Subramanyam , R. Bouqueau, G. Cernigliaro, M. Martos, F. Pérez, P. Cesar, "A Pipeline for Multiparty Volumetric Video Conferencing: Transmission of Point Clouds over Low Latency DASH", ACM Multimedia Systems Conference (MMSys) 2020, Istanbul (Turkey), June 2020

Sergi Fernández Langa is a computer scientist and holds a MsC degree on Artificial Intelligence by the Universitat Politècnica de Catalunya (Spain). Since May 2008, he has given support to the coordination of i2CAT's Media & Internet Unit and led i2CAT's involvement in national and European projects. Since July 2012, he is the director of that Unit. He has co-authored several articles in the field of Interactive and Immersive TV, and he has coordinated different European projects, like TV-Ring (2013-2016) ImmersiaTV (2016-2018), VR-Together (2017-2020), and ImAc (2017-2020)

Mario Montagud Climent was born in Montitxelvo, Spain. He received a BsC in Telecommunications Engineering in 2011, an MsC degree in "Telecommunication Technologies, Systems and Networks" in 2012 and a PhD degree in Telecommunications (Cum Laude Distinction) in 2015, all of them at the Polytechnic University of Valencia (UPV). He has experience as a postdoc researcher at UPV and at CWI (The National Research Institute for Mathematics and Computer Science in the Netherlands). He is currently a postdoc researcher at i2CAT Foundation (Spain) and Part-Time Professor at University of Valencia. His topics of interest include Computer Networks, Interactive and Immersive Media, Media Synchronization and QoE (Quality of Experience).

Dr. Montagud is (co-)author of over 100 scientific and teaching publications and has contributed to standardization. He is member of the Organization and Technical Committee of many international conferences, and of the Editorial Board of international journals. He is currently involved in many regional, national and European projects. Webpage: https://sites.google.com/site/mamontor/

Gianluca Cernigliaro is a Senior R&D Engineer at i2Cat Foundation (Barcelona, Spain). He received his B.S. and M.E. degrees in telecommunication engineering from the Politecnico di Torino (Turin, Italy) and his Postgraduate Master Degree and PhD in Telecommunication Systems and Technologies from the Universidad Politécnica de Madrid (UPM, Madrid, Spain). Between 2008 and 2013, he has been a member of the Grupo de Tratamiento de Imágenes, Image Processing Group (UPM). In 2013, he joined Valeo (Tuam, Ireland), working on Computer Vision based Advanced driver-assistance systems (ADAS) until February 2016. Then, between February 2016 and February 2018, he has worked as Senior Researcher for 8i (www.8i.com, Wellington, New Zealand), focusing on Point Cloud Compression. His research interests include VR, AR, multi-view video coding, 3D video coding, Point Cloud Compression and Computer Vision. He joined i2CAT Foundation in February 2019 to lead the technical advances in the development of video technologies.

David Rincón Rivera received a M.Sc. in Telecommunication engineering and a Ph.D. in Computer Networks from Universitat Politècnica de Catalunya - BarcelonaTech (UPC). In 1998 he joined the Department of Network Engineering (Telematics) at UPC, where he is currently an Associate Professor. He has been a visiting researcher at the Teletraffic Research Centre (University of Adelaide, Australia, 2007) and at the Institute of Pure and Applied Mathematics (IPAM) at UCLA (2008). His interests include network softwarization, audiovisual services over IP, and energy consumption in computer networks.