Deliverable

Project Acronym:	**VR-Together**
Grant Agreement number:	762111
Project Title:	*An end-to-end system for the production and delivery of photorealistic social immersive virtual reality experiences*

D2.2- User scenarios, requirements and architecture v2

Revision: 1.0

Authors: Gianluca Cernigliaro

Deliverable due date: M15

Deliverable actual submission date: M18

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 762111

Dissemination Level		
P	Public	X
C	Confidential, only for members of the consortium and the Commission Services	

Abstract: This document provides an overview of the user scenarios addressed by the project, compiles the set of requirements that drive the design of the project platform, provides details regarding the pursued system features and provides an initial system architecture design.

REVISION HISTORY

Revision	Date	Author	Organisation	Description
0.1	06-12-2018	V.Lepec	VO	Table of contents updated from D2.1 iteration
0.11	18-12-2018	V.Lepec	VO	New Table of Content
0.12	12-02-2019	V.Lepec	VO	D2.1 v1 section 2 added to D2.2
0.2	21-02-2019	G.Cernigliaro	I2Cat	Document general reorganization and definition of the Pilot 2 sections
0.21	20-03-2019	V.Lepec	VO	Architecture section updates
0.22	26-03-2019	J.Lajara	Entropy	Pilot 2 Production
0.3	26-03-2019	G.Cernigliaro	I2Cat	New Requirement Matrix, Pilot 2 experiments, previous versions review Image captions, Table captions, fixed crossed references, Table content included.
0.31	26-03-2019	R.Bouqueau	MS	Encoding and Encapsulation Component updates
0.32	29-03-2019	G.Cernigliaro	I2Cat	Updates in the Exectutive Summary, Production, Requirements. Plus few corrections of references and tables.
0.33	11/04/2019	G.Cernigliaro	I2Cat	Pilot 2 Production plans developed in detail. Pilot 1 experiments calendar description unified as for Pilot 2.
0.34	15/05/2019	J.Lajara	Entropy	Preproduction content design and Description development content design
0.4	02/07/2019	A.Revilla	Entropy	Preproduction content design and Description development content design
0.41	02/07/2019	G.Cernigliaro	I2Cat	Preproduction content design and Description development content design formatting (fonts and titles)
0.42	02/07/2019	G.Cernigliaro	I2Cat	Requirements section updates

0.5	08/07/2019	G.Cernigliaro	I2Cat	Overall updates after WP3 Leader revision
0.51	13/07/2019	M.Montagud	I2Cat	Updates in the Experiments section
0.52	13/07/2019	G.Calahorra	Entropy	Preproduction content design and Description development content design
1.0	15/07/2019	G.Cernigliaro	I2Cat	Changes after WP2 Leader Revision

Statement of originality:

This document contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

EXECUTIVE SUMMARY

This document represents the reference for the software integration and the experiments performed in the VR-Together project. It aims to describe the requirements, the architecture and how the experimental work envisaged to implement the main paradigm outlined in VR-Together: the creation of a platform, and the corresponding media content, that allows two, or more users to feel as if they were together in a virtual environment. The togetherness feeling is provided delivering photorealistic content, both for media and end-user representations. The planning introduced in this deliverable are then executed in WP4, and thus the results of content creation, experiments, and pilots will be included in D4.3 and D4.4.

In the previous document D2.1, apart from describing the overall characteristics of the VR-Together platform, like requirements and experiments, we described the scenario, architecture and software platform considered for Pilot 1.

This document (D2.2) includes all the aforementioned information and, in addition to that, describes:

- The new scenario considered for Pilot 2: how the plot of the story evolves, with the corresponding production details, together with the technical advances
- The new Requirements Matrix, that contains the requirements needed to reach the goals initially fixed and explaining the methodology used to reach them.
- Architecture and software implementations used for Pilot 2, including the technical details in terms of technology and methodology and the delta between Pilots 1 and 2.
- Experiments and testing: the experiments that will pave the way to the pilots and the plan that the consortium has agreed to perform them.

The document starts with an introduction (Section 1) followed by the description of the high level requirements initially defined for the project including, first, the general ones and, afterwards, the specific requirements for each Pilot (Section 2).

Section 3 describes the several scenarios considered for the different pilots. Starting with Pilot 1 (as in D2.1) followed by the scenario designed for Pilot 2 which has a threefold purpose. The first goal is to improve the feeling of immersion (or togetherness); second, the content of the scene has to be generated and transmitted in real time, adding a live factor to the experience; third, it aims to include 4 or more users (while in Pilot 1 the target was 2). More details in the production process will be provided in deliverable D4.3.

The content produced for Pilot 2 is the natural follow up of the first pilot plot, which was a police interrogation scene and represented the starting point of the thriller like story which will be narrated along the three pilots. More specifically, in pilot 2, the experience starts by placing the user in a TV news studio where the anchor-man gives an overview of the news of the day. At some point, to go more in depth in one of the news items, the end user is holo-ported into a crime scene, or a place that has a direct relationship with the crime, where a journalist describes the scene. Thus, the objective with such transition between sites is twofold: first, to immerse a group of users in an indoor environment, and second, in an outdoor environment.

Section 3, a section depicting the scenarios and analysing the technical implications, explains also the production process. The content of the whole experience is represented as a mixed space where half of the scene is a 3D environment and is merged with a 180º live content. The users will be placed in the 3D environment and will be able to watch the rest of the scene as a half sphere (180º) perceiving it as the rest of the 3D space.

Section 4 describes the actual requirements of VR-Together that represent the evolution of the high level ones mentioned in Section 2. The section begins explaining the scenario of Pilot 1 (use cases, technical analysis and software requirements) as in D2.1, and continues with the specific characteristics of the software platform used for Pilot 2. The section explains in detail the process of definition of the requirements that the VR-Together technology has to meet, in order to provide an experience as mentioned above (togetherness, live content, additional users). The definition process, of the current requirements, has started taking into account the previous requirements matrix (same version as in D2.1) and has gone through three iterations: i) the previous requirements have been object of discussion by all the partners and, when an agreement was reached, each requirement has been confirmed, deprecated or re-defined; ii) the second iteration have been focusing on the clustering of the requirements in order to provide a better grouping and an easier classification for the process; iii) finally, when the new requirements matrix was defined, all the partners have started defining how the requirements are linked to the experiments. This methodology has been complemented with a number of experiments, including large-scale consultation with stakeholders and experts in targeted focus groups. Such mixed-methodology has resulted in a number of requirements that have been provided to the technical team for the implementation and deployment of pilot 2.

Section 5 introduces the overall architecture for VR-Together, how the different components interact and the hardware topology, taking into consideration the software modules from WP3. The specific implementation details are provided in the WP3 deliverables.

Section 6 outlines all the information related to the distributed lab realized within the VR-Together project. It presents the Advisory board, resulting from the large-scale consultation with stakeholders and experts, supporting the project. Then it describes the existing connected user labs in France (web based pipeline) and between Greece and Spain (native pipeline) and the lab nodes at each partner's premises. Finally, it lists the experiments associated and projected within VR-Together Pilot 2, including the particular requirements that each experiment considers, followed by a detailed and updated experiments calendar. The results of the experiments will be reported in D4.4.

Section 7 summarizes the contributions of this deliverable.

Last, Annex I (8) and II (9) provide an example of a questionnaire used for experimentation and a detailed description of the Lab nodes in each one of the partners' locations. Annex III (0) shows the requirements considered in Pilot 1 and can be used as a reference after reading. Finally, Annex IV (11) is used to provide, to the reader, the information about the experiments previously planned for Pilot 1.

CONTRIBUTORS

First Name	Last Name	Company	e-Mail
Javier	**Lajara**	ENTRO	javier.lajara@entropystudio.net
Guillermo	**Calahorra**	ENTRO	guillermo@entropystudio.net
Pablo	**Cesar**	CWI	p.s.cesar@cwi.nl
Omar	**Niamut**	TNO	omar.niamut@tno.nl
Argyris	**Chatzitofis**	CERTH	tofis@iti.gr
Spiros	**Thermos**	CERTH	spthermo@iti.gr
Simon	**Gunkel**	TNO	simon.gunkel@tno.nl
Romain	**Bouqueau**	MS	romain.bouqueau@gpac-licensing.com
Pascal	**Perrot**	VIA	Pascal.PERROT@viaccess-orca.com
Dimitris	**Zarpalas**	CERTH	zarpalas@iti.gr
Vincent	**Lepec**	VIA	v.lepec.ext@viaccess-orca.com
Gianluca	**Cernigliaro**	i2Cat	gianluca.cernigliaro@i2cat.net
Sergi	**Fernandez**	i2Cat	sergi.fernandez@i2cat.net
Mario	**Montagud**	I2CAT	mario.montagud@i2cat.net
Ana	**Revilla**	ENTRO	int@entropystudio.net
Hans	**Stokking**	TNO	Hans.stokking@tno.nl
Tom	**De Koninck**	TNO	Tom.dekoninck@tno.nl

CONTENTS

LIST OF FIGURES

LIST AND TABLES

1. Introduction

1.1. Purpose of this document

The purpose of this public document is to provide the reader with a comprehensive view of the project requirements, the use cases contemplated for the scenarios designed for pilot 1 and pilot 2 and the system architecture envisaged to meet the requirements. The document also gathers information regarding the User lab initiative of the project as well as other feedback methods such as experiments, advisory board and others.

1.2. Scope of this document

This document includes the current status of the review of the project requirements: first for the overall requirements and, then, for the specific requirements of each component. The list of requirements gathered in this document will serve as a basis for discussions towards component implementation and integration of the second version of the VR-Together platform (Pilot 2).

1.3. Status of this document

The D2 deliverables follow the project all along its three iterations. Three different versions will be then formally submitted to the EC and uploaded in the project website. This D2.2 document is an evolution of the D2.1 document and provides the status of the second iteration.

1.4. Relation with other VR-Together activities

This document gathers the outputs of T2.1, T2.2 and T2.3 and serves as input for WP3 and T2.4. It also provides input to WP4 w.r.t experiment definition and evaluation methodology.

2. Project Requirements

In this section we aim to describing the high level requirements of VR-Together platform. The project can be considered as divided in 3 iterations, each one addressing one technical scenario that will be validated by user groups through 3 pilots. In terms of pilot content, the ones initially foreseen to feed public demos and user evaluations was: an intimate concert, a live news format and a fictional story plot. In terms of technical scenario of each pilot, they were classified as offline, live and interactive respectively. This division allows the project to reach intermediate objectives both at creative and technical level, facilitating the consortium to deal with the challenge of delivering a satisfying social VR experience.

The requirement definition has been updated during the project. This section describes the initial status of the requirements in order to provide to the reader, since the beginning, an understanding of the VR-Together platform purpose. A more detailed definition of the requirements gathering process and of the updated status of them can be found in Section 4. The Requirements provided correspond to the status at the time of the delivery of this document and, as a consequence, the most updated version.

This section is structured as follows: first, we introduce the initial set of general requirements (those requirements that have been valid since the beginning of the project and that all versions of the VR-Together system should meet). Second, we show the different ideas, that were included in the project proposal, able to set the differences between pilots, providing a list of the specific requirements of each pilot. Finally, we compile the initial scenario to be addressed in VR-Together.

2.1. General requirements

VR-Together aims at exploring how the combination of various data streams (content, human representations, data) will result in a highly personalized experience that is delivered in an adaptive manner, enabling individuals in different locations participate together in the same experience. The objective is to deliver close to market prototypes and implement an integrated platform to achieve the main project objective: delivering photorealistic immersive virtual reality content which can be experienced together with friends, and demonstrate its use for domestic VR consumption.

VR-Together, as mentioned above, is structured in 3 iterations. After each iteration, the project will deliver a system version that will meet the indicated requirements. Both system and requirements will be validated and the consortium will evaluate if and to what extent the work done meet each of the requirement. The following table gathers the initial list of general requirements considered by the consortium.

CODE	NUM	TITLE	DESCRIPTION
GEN	1	Co-presence	End users should be able to be virtually present in the same virtual space and engage in real-time face-to-face social activities. Co-presence should lead to other-awareness, social behaviour, responsiveness to one another's actions and self-awareness
GEN	2	Distributed experience	End users should be able to access a shared virtual space from different physical locations (equipped with the corresponding capture and visualization systems)

GEN	3	Number of users per physical space	At least one end user should be able to access a shared virtual environment from a specific physical location (equipped with the corresponding capture and visualization systems)
GEN	4	Natural communication	End users should be able to communicate with each other in a natural, fluid, way. This requires real-time interaction (i.e. transmitting/receiving the other user's graphical representation and voice with imperceptible delay)
GEN	5	End user representation	End users inside a virtual space should be able to see other end users body representation
GEN	6	Self-representation	End users inside a virtual space should be able to see their own body representation
GEN	7	Place illusion	End users inside a virtual space should have the feeling of being in the physical space depicted in the VR content
GEN	8	VR content	End users inside a virtual space should be able to see VR content
GEN	9	VR content formats	End users should be able to see different examples of VR content formats
GEN	10	VR content image quality	End users should be able to see photorealistic VR contents
GEN	11	Synchronization	End-users in distributed locations sharing a virtual space should be able to see the same VR content at the same time
GEN	12	End-user image quality	End users should see other users in photorealistic quality
GEN	13	End-user blend	End users should see other users seamlessly blended in the VR content
GEN	14	Perception of VR quality	VR-together should improve the subjective quality of previous Social VR experiences
GEN	15	Comfortability	End users should be comfortable in using the system for at least the duration of the pilot experience
GEN	16	Body language	End users should be able to understand each other’s body language expressions.
GEN	17	Immersive VR audio	The VR audio content should be immersive. That is, when the end user turns the head, audio should change as it does naturally
GEN	18	Audio/Video Synchronization	The VR audio and video content must be synchronized, as in any content experience

GEN	19	End-user audio	The end-user audio for communication should be directional. That is, end-user audio should appear to come from its originating point.
GEN	20	End-user devices	End users should access the experience using commercially available HMDs and capture systems
GEN	21	Data logging	The system has to record end user activity data
GEN	22	Blend of media formats	End users, scene of action and characters should be represented using different media formats. The resulting VR image should be a blend of different formats.
GEN	23	Networks	The VR content and end-user representations need to be delivered over commercial communication and media delivery networks.
GEN	24	Adaptive media delivery	Media streams should provide adaptive quality to network, device and interface capabilities
GEN	25	Web interface	End users should be able to access the experience using a web application.
GEN	26	Native interface	End users should be able to access an experience using a native application

Table 1: General Requirements

2.2. Requirements for Pilot 1

In this subsection we review the initial assumptions to be considered in Pilot 1, as initially planned in the project proposal. As described in the proposal, section 1.3.4.2:

*"Pilot 1. Intimate Concert**.** The goal of the offline pilot is to demonstrate that the innovative media format of VR-Together (orchestrating point clouds, 3DMesh based models and multiple video sources) can produce a more intimate and binding activity than more traditional content production pipelines, based on omnidirectional content. We will compare different capture and production techniques (video, point cloud capture, high-end motion capture) as well as combinations of them to determine quantitative balances among the different formats available (video, point clouds, time-varying meshes, dynamic meshes, motion data). The main variables considered to compare the different means available to deliver such an experience will be:*

· *Production costs, integrating shooting, editing, compositing, post-production, etc.*

· *Bandwidth and computational resources required at the different nodes (capture, encoding, delivery, rendering)*

· *Impact on the subjective social experience among end-users.*

*Typology of contents addressed: An intimate music concert seems an ideal starting point to demonstrate VR-Together's innovative media format. It is a good opportunity to show how the VR-Together works for offline produced content**.** The goal is to demonstrate that the orchestrated delivery of the VR-Together media format, combining several video sources, point cloud and 3D mesh representations will improve closeness with the musicians and with at least 2 distant end-users. Particular care will be taken to integrate facial expression within the production pipeline, i.e. how we will capture the photorealistic 3D actors in costume. For example, uses 108 cameras to capture the actors' performance, costumes, facial expressions and the result is a stream-able 3D model with appropriate facial expressions. This also applies to lighter methods, which are more affordable and portable. For example, uses 4 Kinect sensors and a short automatic calibration process. Industrial methods capturing actor facial MoCap performance using marker-less methods and pre-rigged models will also be considered. Different combination of methodologies and technologies will be studied to deliver the best possible balance between visual quality and cost efficiency in content production."*

As described in the proposal, in T4.1, the task that addresses the prototyping and production of demo content:

"Offline CoVR: The content format that we have pre-selected is an intimate concert, which seems relevant to validate the unique feeling of closeness between the audience and the content that the VR-Together platform will deliver. We will also seek to detect implicit social interaction cues that may improve the connection between the audience and the users, such as real-time retargeting of gaze or pointing gestures in the characters being rendered, in order to further integrate the content consumer's presence."

As described in the proposal, in T4.2, the task that addresses the deployment of demos and pilots, with a more practical (technical deployment) approach:

"Offline CoVR In this first example of content production and delivery, we will focus on validating the staging and capture process to deliver the feeling of co-presence in a shared photo-realistic immersive virtual reality environment. We will study which computer graphics techniques can appropriately blend the representations of end-users, created with real-time constraints, home lightning, affordable cameras and sensors for capture, with the offline produced content. Where possible, we will seek to apply re-illumination techniques to blend end-user representations within the pre-recorded content. "

The following table gathers the subset of high level requirements for pilot 1.

CODE	NUM	TITLE	DESCRIPTION
P1	1	Facial expressions	Some detail to see facial expressions should be available in the end-user and character representations
P1	2	Offline content	The VR content to be displayed must be stored in the end user device
P1	3	Illumination	Illumination should be consistent in the whole experience
P1	4	Gaze	Rendered characters should be able to retarget their gaze according user's viewpoint
P1	5	Pointing gestures	Rendered characters should be able to retarget pointing gestures
P1	6	Rendered Characters	The scene should contain rendered characters
P1	7	Characters' representation	The end-user should perceive the 3D appearance of the characters (some parallax, depth)
P1	8	Basic end user movement	Users can rotate their head and have certain level of translation capacity while seated (3DoF+)

Table 2: Pilot 1 requirements

2.3. Requirements for Pilot 2

In this subsection we review the initial assumptions to be considered in Pilot 2 as initially planned in the project proposal. As described in the proposal, section 1.3.4.2:

"Pilot 2. Live news. We will demonstrate the live production of multi-source immersive content. We will study the conditions which maximize the connection between the audience and the news. Numerous benefits for cost-effective production efficiency will be derived from introducing live processing constraints. Quantitative measures comparing the benefits and costs of introducing offline processing steps will be sought. To realize this scenario, we foresee the creation and demonstration of a hybrid live production that combines omnidirectional cameras and depth sensors and off-the-shelf capture devices targeting consumers (webcam, Kinect) in order to allow several users to feel like being together inside an immersive virtual environment and to increase the feeling of connection with the environment thanks to embodied social interaction. In this scenario, inter-stream synchronisation is critical: this is not a live VR conference, but a production broadcast. Technically speaking, we need clock sync between equipment at both production environments, and insert / correlate timestamps in the recordings. This kind of activity is aligned with current standardization activities in MPEG MORE, to which part of the VR-Together consortium contributes actively.

Typology of contents addressed: *We will demonstrate a novel content format of immersive news consumption, where people can feel like being together where the news actually occurred. For this we will combine more closely the content production expertise (camera placement, social setting between presenters and the audience, how transitions to other settings (for example, a journalist on the field) can be established and delivered comfortably to the audience, etc. The introduction of live delivery for the case of live news will require a production design adapted to the needs and constraints of News Production (Main set with news presenter, live connection with journalist on the field, etc.), but which still allows for a quality of content as close as possible as an offline production."*

As described in the proposal, in T4.1, the task that addresses the prototyping and production of demo content:

"Live CoVR: The content format that we have pre-selected is a broadcasted news, which seems relevant to validate the feeling of immediacy that such techniques can deliver. We will however, study other options if real content production opportunities (events, real concerts, etc) appear, and they seem more appropriate for the validation purpose at hand. "

As described in the proposal, in T4.2, the task that addresses the deployment of demos and pilots, with a more practical (technical deployment) approach:

"Live CoVR: In this second example of content production and delivery, we will focus on validating the real-time processing tooling implemented to deliver, at best as possible, the feeling of co-presence in a shared photo-realistic live immersive virtual reality environment. Building upon the insight of first pilot, we will simply aim at assessing to what extent we can preserve the feeling of closeness and empathic connection between the audience and the content, when real-time constraints are imposed. Imposing real-time processing, with no possible offline manual adjustment and manipulation of the content captured severely limits the range of technical possible options. "

The following table gathers the subset of high level requirements for pilot 2.

CODE	NUM	TITLE	DESCRIPTION
P2	1	Number of users	The system must accept between 2 and 10 end-users (in different rooms/locations)
P2	2	Facial expressions	Sufficient detail to see facial expressions should be available in the end-user and character representations
P2	3	Multi-source	The system must be able to produce multi-source immersive content.
P2	4	Live	The system must be able to deliver a photorealistic live immersive VR environment.

Table 3: Pilot 2 requirements

2.4. Requirements for Pilot 3

In this subsection we review the initial assumptions to be considered in Pilot 3 as initially planned in the project proposal. As described in the proposal, section 1.3.4.2:

"Pilot 3. Interactive Fiction. We will seek to demonstrate how the VR-Together platform, in a custom-designed content production process, can allow for a novel form of content where users meet, and blend within the interactive immersive experience. Thus, consumers can watch passively. However, they are also able to, essentially, become a character within the story plot being rendered. They can have this experience through a more active engagement in the experience, i.e., by moving and talking like one of the characters in the plot, and with these actions change significant aspects of the plot being rendered. This will require the combined delivery of broadcast video, mesh or point-cloud content, together with end-user capture in the form of video, point cloud or interpolated 3dmesh, as well as with event-based synchronization similar to how MMO video-games are synchronized. Regarding the integration of advanced multi-modal pattern recognition, the effort will not be on creating sophisticated multimodal pattern recognition of social actions, which would work for any plot, but rather to demonstrate how readily available pattern recognition tools (speech recognition, existing gesture recognition algorithms) can be used and integrated to convincingly deliver one specific plot. For this matter, the previous work done within the VR-Together project, regarding spontaneous social interaction in SIVE will become essential to guide this process. Regarding the processing of interactive plots in SIVE, we will use tools readily available from previous research initiatives by the partners within the consortium. The main challenge to maintain place illusion and plausibility is to render credible interactivity within the experience. We will address how to integrate the user input with the events being depicted within the immersive virtual environment. The goal will be to show to what extent and how a fiction scenario can be rendered in VR, while still allowing the users immersed in the scene to intervene actively in the scene being broadcasted within the shared virtual reality experience (and thus, preserving the feeling of being there together).

<u>*Typology of contents addressed:*</u> *We will address interactive content rendered in the form of interactive fiction. This will be demonstrated as a story-like plot rendered within the immersive experience. The user will be able to actively intervene and change some aspects of the experience by performing some of the actions (i.e, talking, pointing or performing simple physical actions) that correspond to the character he/she wants to become within the plot."*

As described in the proposal, in T4.1, the task that addresses the prototyping and production of demo content:

"Interactive CoVR. The content format that we have pre-selected is a fiction production, which will allow for additional control in the production process, and will develop a scenario that will be close to a movie script. We will use the insight of subtask T4.3.1 co-presence and social interaction evaluation, in order for the experience of the content to integrate harmonically with possible social interaction occurring, not only among the end-users, but also with the content being rendered. The global aim will be to achieve a qualitatively different level of co-presence, social interaction and place illusion in an aesthetically coherent virtual reality experience."

As described in the proposal, in T4.2, the task that addresses the deployment of demos and pilots, with a more practical (technical deployment) approach:

*"Interactive CoVR**.** In this third example of content production and delivery, we will focus on validating the production of explicitly interactive content to maintain, preserve and if possible reinforce the feeling of co-presence in a shared photo-realistic immersive virtual reality environment. We will seek to detect an expanded range of social and bodily-centred interaction cues (head movements, body movements, peri-personal space, and spoken keywords) to further allow the integration of the end users' actions within the narrative. We will integrate existing innovative*

interactive storytelling engines available within the VR-Together consortium, along with re-illumination, rendering, and interactive character animation techniques. "

The following table gathers the subset of high level requirements for pilot 3.

CODE	NUM	TITLE	DESCRIPTION
P3	1	Facial expressions	Photorealistic detail to see facial expressions should be available in the end-user and character representations
P3	2	Passive watch	End users can watch the content in a passive way
P3	3	Active watch	End users can become a character within the story plot being rendered
P3	4	Movement	End users can move (translation). 6DoF
P3	5	Derived actions	End user actions change significant aspects of the plot being rendered
P3	6	Pattern recognition	The system must demonstrate how multi modal pattern recognition tools can be used and integrated into the plot.
P3	7	Pointing	End users can trigger story actions with pointing gestures
P3	8	Talk	End users can trigger story actions by talking
P3	9	Physical actions (triggering gestures?)	End users can trigger story actions by performing simple physical actions
P3	10	Interactive storytelling	The system will integrate existing interactive storytelling engines
P3	11	Interactive character	The system will integrate interactive character animation techniques

Table 4: Pilot 3 requirements

2.5. Experimental requirements

The evaluation of the VR-Together platform is organised in two different parts. The first part is concerned with validating the different parameters that need to be preserved or improved. This includes aspects such as delays, resolution, etc. These experiments do not imply specific requirements on the overall platform.

The second part is concerned with validating the feeling of being there, in the virtual environment, and of togetherness, i.e., determining under which technical conditions it can be maximized. This presupposes experiments which involve specific requirements on the end-to-end architecture, which we list below.

CODE	NUM	TITLE	DESCRIPTION
EP1	1	Place illusion under bandwidth and delay constraints	one single end-user, through gamepad or wand, can change between different bandwidth and delay constraints, and choose which experience is better, worse, or equal
EP1	2	Place illusion changing content and self-representation formats	one single end-user can change his self-representation (static virtual body, dynamic virtual body, 3d-reconstructed mesh) and the media format (omnidirectional video, 3d geometry + stereo billboards, 3d geometry + 3d virtual characters)
EP1	3	Render the other's virtual body is animated or static	To reproduce the Joint Action Effect On Memory (Wagner et al 2017, Eskenazi et al. 2013), the experimenter needs to be able to show the participant to see the other's virtual body either static or dynamic.
EP1	4	Render the other's virtual body at different distances	To reproduce the Joint Action Effect On Memory (Wagner et al 2017, Eskenazi et al. 2013), the experimenter needs to be able to show the participant to see the other's virtual body at different distances .
EP1	5	capture motion data and speech	To find behavioural measures related with togetherness, we need to be able to record the entire multi-modal data, with good time precision.

Table 5: Experimental requirements

3. Pilot Scenarios

The three pilots of VR-Together address specific goals in terms of technical challenges and evaluation purposes. The Pilots can be considered as project checkpoints to evaluate the creative and technical challenges identified as the ones to overcome toward the creation of a truly realistic social VR experience.

The Pilots were initially planned as individual content capsules addressing completely different content scenarios. The structure and plot complexity of the pilots is linked to a gradually increasing technical difficulty, with the first pilot being the simpler to produce and to accomplish technically and the third the most challenging.

These three pilots were initially planned as follows:

- A first offline pilot, simulating an acoustic music concert able to offer not only the feeling of being together, but also intimacy and closeness, all this through orchestrating clouds of points, 3DMesh models and multiple sources of videos.
- The second pilot was focused on live news, simulating a live production of immersive content from multiple sources that aimed to virtually transfer the user to the location of the news and share the experience with other users.
- The third pilot intended to present a test to users through an interactive and totally immersive experience, with the background of a television series, a movie or simply a scene taken from them, where users can participate in the scene, interact between them, make conclusions, etc.

During the kick-off meeting in Barcelona in mid-October 2017, the artistic partners suggested an alternative approach: creating a coherent storyline that runs across the three pilots. Each pilot represents a scene of an overarching story plot. The hypothesis was that by changing the original concept and plot line of the pilots, in the end we will offer a more attractive and engaging experience to the end-user. Moreover, this will allow the project to provide a concrete and coherent novel "product" that can be showcased in film festivals and other artistic venues. It is expected to draw the attention of the consumers, making them participants of the experience. This will in turn trigger sociological phenomena such as word of mouth, or electronic word of mouth, to play the role of communicators, attracting the interest of general public and media.

So, the three pilots have been connected between each other, telling a story about a murder. The story focuses on the murder of Ms. Armova, a wealthy British lady in unknown circumstances. Two persons were present at the time of the murder: Ryan Zeller and Christine Gérard. Each suspect has his own version of the events that happened that night, and it will be up to the users to draw conclusions.

The 3 Pilots have been planned to be delivered in the periods of July – September 2018, June - August 2019 and May - July 2020 (Figure 1). During those time slots, a number of experiments have been planned to be ran, some in the user labs of some of the partners and others in roadshows demonstrating the status of the project. The results are planned to be presented in technical and creative industrial fairs and events.

The following subsections describe how Pilot 1 and Pilot 2 have been defined, explaining the plot, the storyboard, the pre-production and production activities along with a detailed Pilot Action calendar that lists the number of actions executed by using the content over the VR-Together platform.

Figure 1: VR-Together Pilots Calendar.

3.1. Pilot 1

One key concern of the consortium has been about the new approach for the pilots and if it would fulfil the needs of the project in terms of providing a social VR experience. In particular, it is essential for the pilots to provide an understanding of the technological advances and experiments performed to analyse co-presence, togetherness and immersion.

During several meetings between the creative and the technical partners, the agreement to ensure that the pilots, designed as a story in three episodes, are valid vehicle for the project, has been reached. It was therefore decided, for the first pilot, to focus on the communication between remote participants while performing an activity together; for the second pilot the goal is planned to be the improvement in terms of scalability; for the third, the interactivity with the scene will be the main area of interest.

3.1.1. Plot

The plotline relates to a police theme (police investigation or interrogation), which widely fulfils the requirements of the project. This storyline exploits the unique advantages of the project: a team where technical and artistic experts work together, creating a brand new experience. The final objective is to obtain a new experience that involves the viewers and make them immersed in an uncommon encounter that is different from what they might have previously seen.

One of the questions that often was asked within the consortium was "*What is the target audience of VR-Together*?". As a clear answer was not really evident, we decided to avoid establishing limits and, then, just to follow a generic approach that would serve for any kind of audience, avoiding addressing only specific types of end-users. Being inspired by movies as "*The Usual Suspects"*, the proposal has been of having a thriller-like plot for the three pilots. The viewer is then the one who has the control over the story and who can enjoy the experience not only while experiencing the pilots, but also in between and after the last one. The generic structure of the pilots is presented in Figure 2.

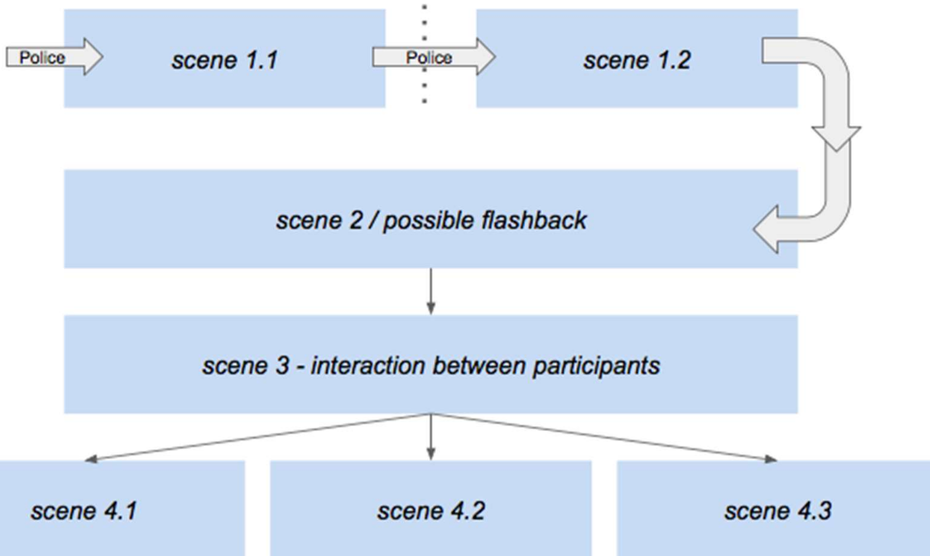

Figure 2: Scenes integrating general story

The creative partners of the project went through a number of iterative design sessions, creating visual representations to facilitate the discussion (see Figure 3 and Figure 4). Coherent plot suggestions and ideas were brought forward, discussed within the consortium and, eventually, the most interesting ones evolved as the final pilot definition.

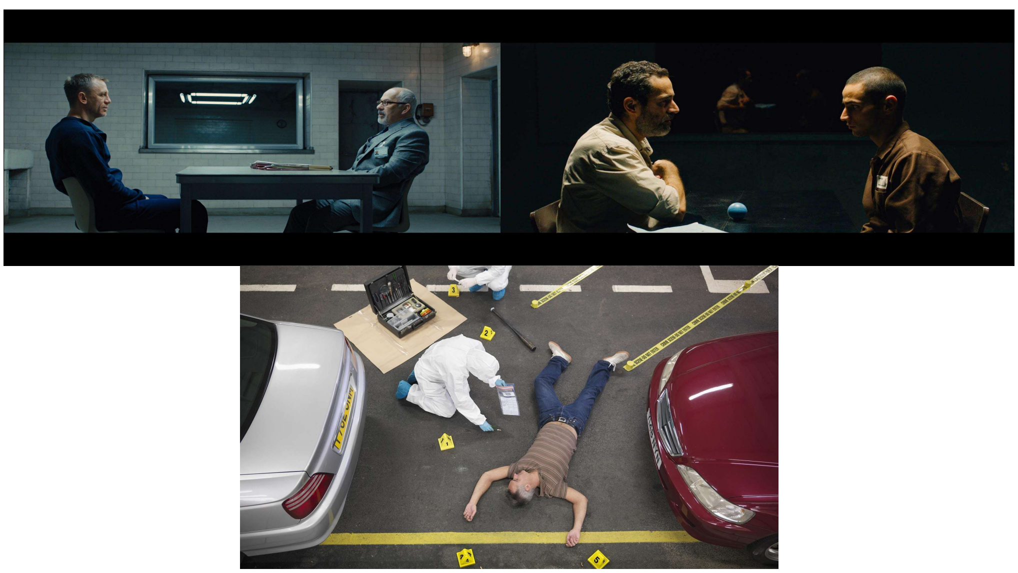

Figure 3:Initial Concepts for the Trial (interrogation, crime scene).

During the face-to-face TCC meeting in Madrid (November 2017), three main ideas for the first pilot were presented: a murder scene (see Figure 4), an interrogation with one-way mirror (see Figure 5), and an interrogation inside a prison (see Figure 6). The initial ideas of each scene can be summarized as follows:

- **Murder scene**: in this scenario both users are found in the same room, where a murder has been committed, and both users are far enough from each other, in order to have different point views that will allow them to see different objects and clues. The collaboration of both users (involving the feeling of togetherness) would be essential to come to conclusions and take a decision about the identity of the murderer.
- **Interrogation room with one-way mirror**: the users are behind a one-way mirror of an interrogation room. Although users are next to each other, each of them is able to see a different interrogation room and they are both aware that the other user is watching a different interrogatory.
- **Interrogatory inside the prison**: In this scenario both users are inside a prison in front of the main suspect. Both users can interact with each other.

Figure 4: Pilot Proposal – murder scene.

Figure 5: Pilot Proposal – interrogation with one-way mirror.

Figure 6: Pilot Proposal – interrogation inside the prison.

After a discussion, the project partners selected the second scenario. Following, Entropy Studio developed the storyboard of the main concept for the pilot and for the production plan.

3.1.2. Storyboard

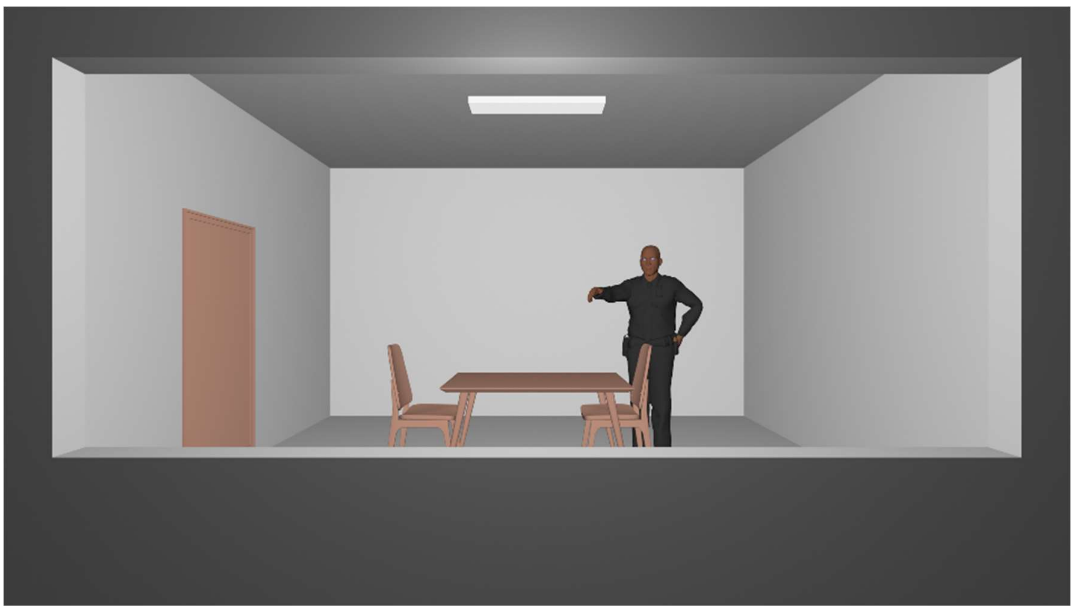

Figure 7: Story-board: Police officer waiting for the suspect (scene 1).

In the intro, we are placed on the dark side of an interrogatory room. A police officer is waiting patiently for the suspect. Beside us, we can see and hear the other user, displayed using a volumetric video technology (as point clouds or meshes), on the other side of the room. After a short time, between 5 and 10 seconds, a second police officer comes in the interrogation room. A suspect is brought in to be questioned and we are witnesses of the conversation. We are supposed to closely pay attention to the discussion and to look for clues that can help to clarify the identity of the criminal.

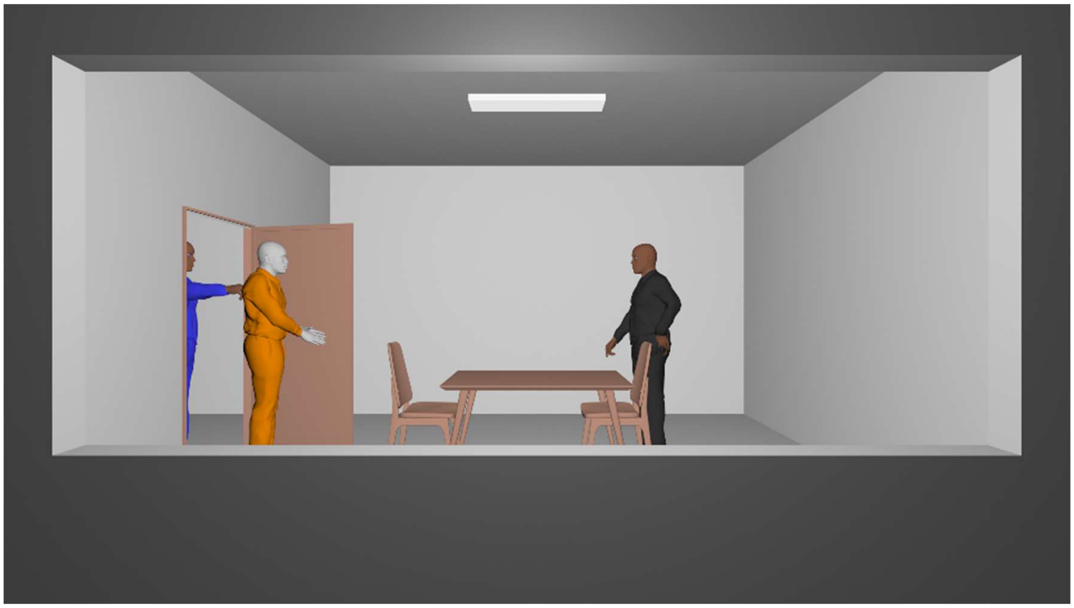

Figure 8: Story-board: Suspect introduction (scene 2).

At the beginning of the interrogation the suspect is placed seating on a chair and handcuffed, having the questioning officer on the other side of the table.

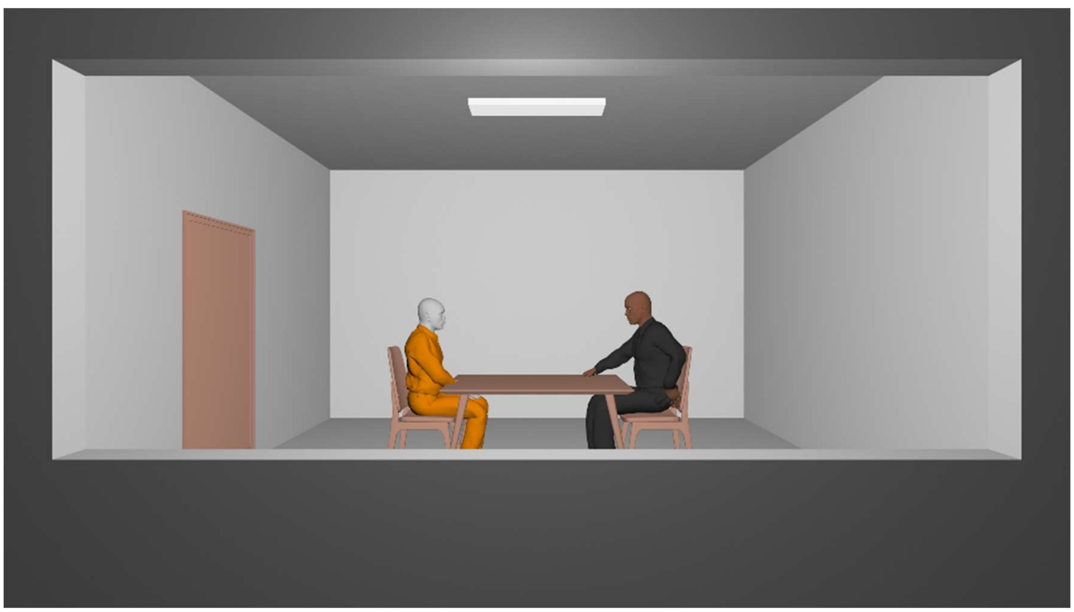

Figure 9: Story-board: Interrogation (scene 3).

In order to find a contradiction in the suspect story, the officer to start asking questions, talking about the crime scene, the alibi of the suspect at the moment of the crime, information about the other suspects, etc.

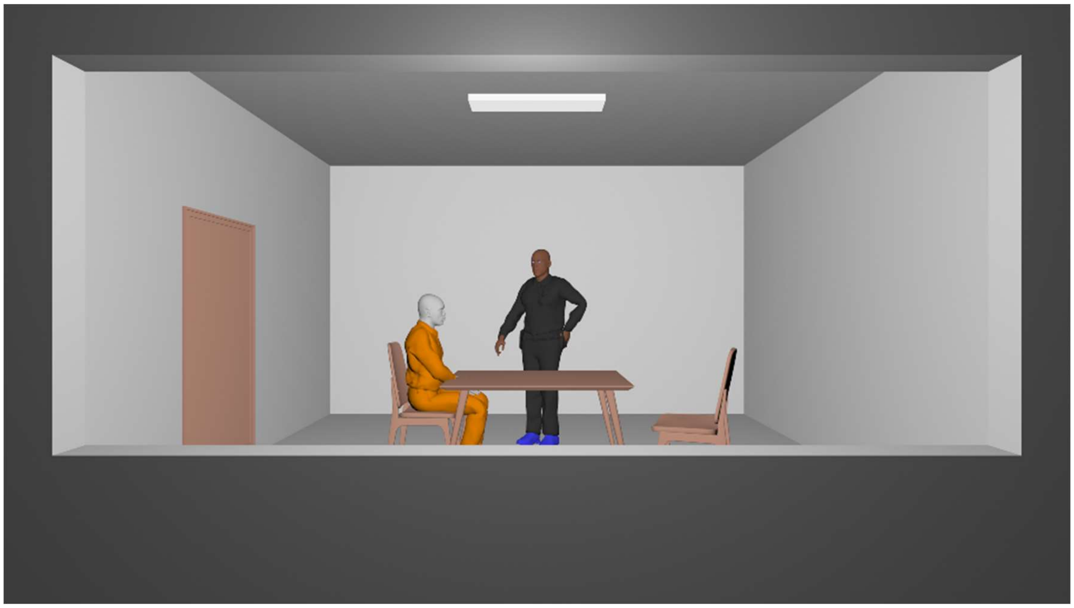

Figure 10: Story-board: Secret revelation (Scene 4).

During the conversation the suspect reveals important and relevant information that can lead us to finally identify the responsible of the murder.

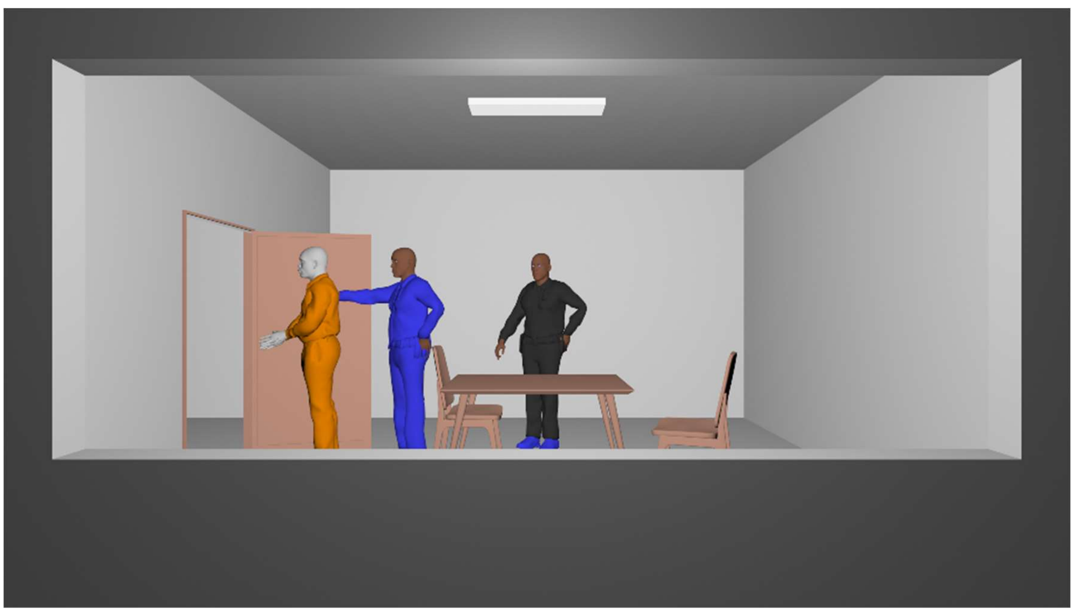

Figure 11: Story-board: Interrogation ends (scene 5).

The denouement of the experience starts when the interrogation ends. The officer is left alone in the room stating the facts and drawing the final conclusions looking at us through the window.

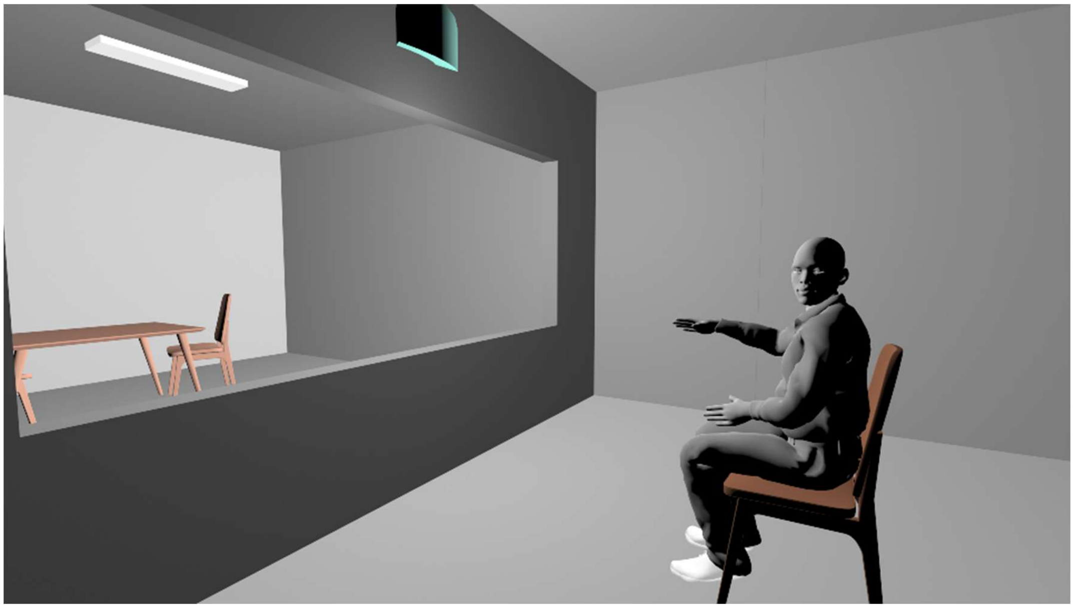

Figure 12: Story-board: User's discussion (scene 6).

At the end of the experience, the participants are expected to have a conversation about the interrogation that they just witnessed. At this point the participants are expected to interact and reach a conclusion about the criminal's identity and/or other details about the committed crime.

3.1.3. Pre-Production

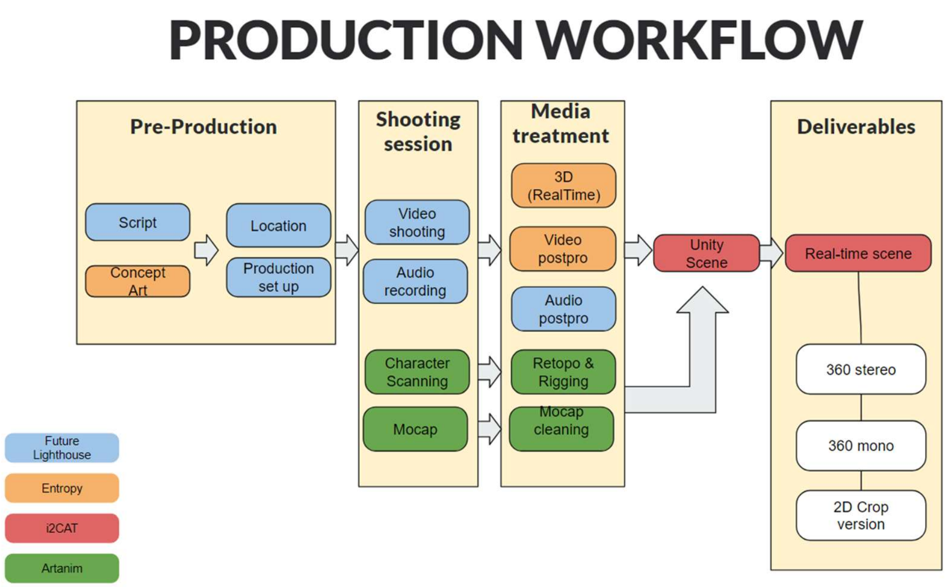

Figure 13: Production workflow.

The diagram on Figure 13 describes the process followed for the production of the Pilots.

The process right after the approval have been:

- Creation of the final script
- Concept art images defining the visual environment
- Casting for actors/actresses
- Dressing selection for actors/actresses
- Location scouting
- Technical team members hiring
    - Director of photography
    - Camera operator
    - Sound team members
- Technical gear rental process
    - Lights
    - Chroma
- Production planning
    - Dates for shooting
    - Travel and accommodation
    - Miscellaneous logistics
- Soundtrack and music
    - Curation of the music and sounds used in the experience.

3.1.4. Production

This section provides a technical breakdown, graphically supported, to describe the production techniques that were planned for Pilot 1 and that, partially, will be exploited also in Pilot 2. All the actions described are part of the WP4 framework.

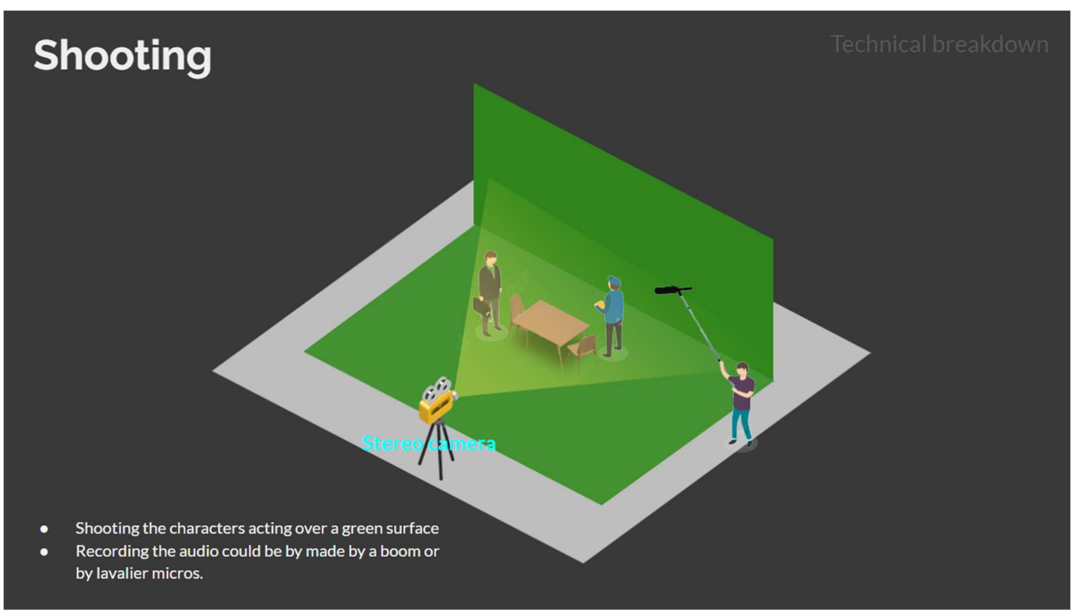

Figure 14: Stereoscopic shooting of character action.

The action for Pilot 1 have been recorded with a stereoscopic rig of two cameras, separated by 67mm, which simulates the distance between human eyes (standard).

The recording has been done in an environment allowing for Chroma-key composition during the post-production phase. In this way it has been possible to remove the background and place the action in any kind of scenario.

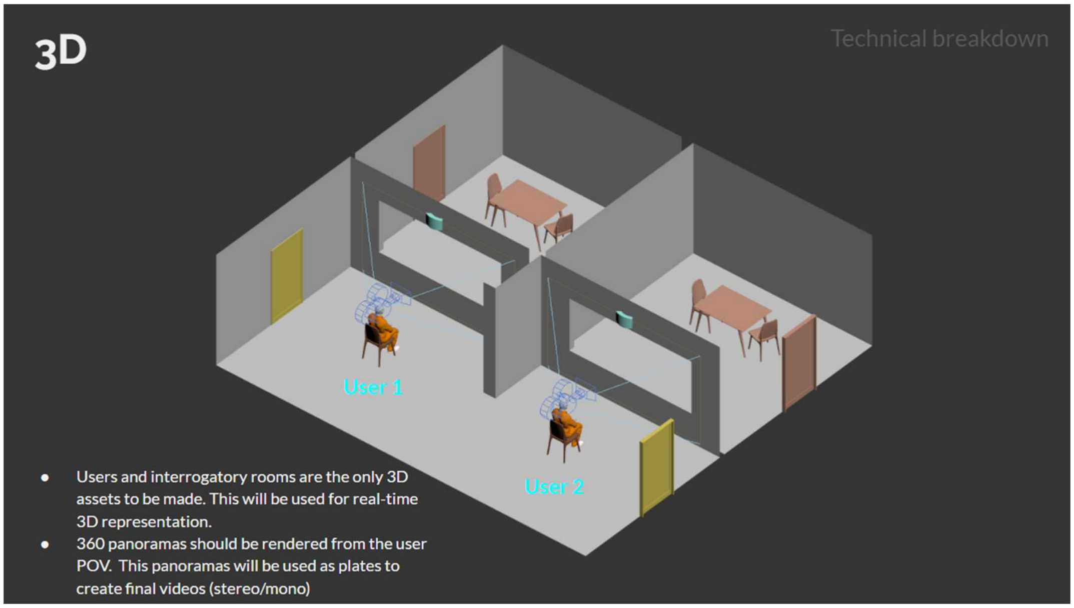

Figure 15: 3D Scene where action takes place.

Afterwards, a room has been modelled to simulate the police station. The participants are able to watch the interrogation scene described in Section 3.1.1.

The ·D environment has been created using the Unity 3D real-time engine.

In the delivered Pilot 1 experience, the users have been rendered as Time Varying Meshes (TVMs). They can see each other and communicate via gestures and voice.

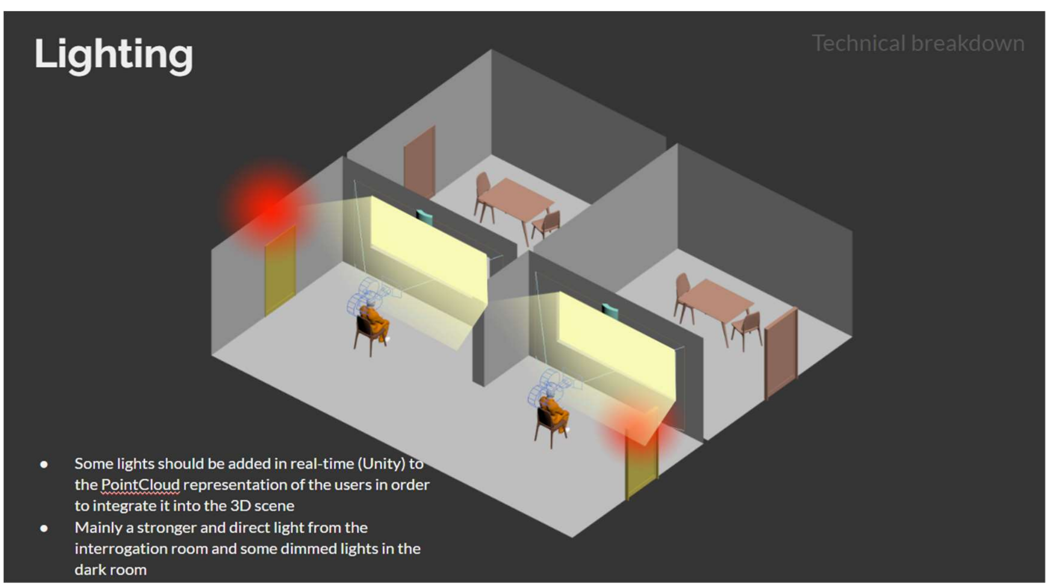

Figure 16: Coherent lighting. Users and scene.

To achieve a high level of realism and to provide to the end-users a feeling of integration in the scene, a simulation of realistic lighting conditions has been implemented in the 3D engine.

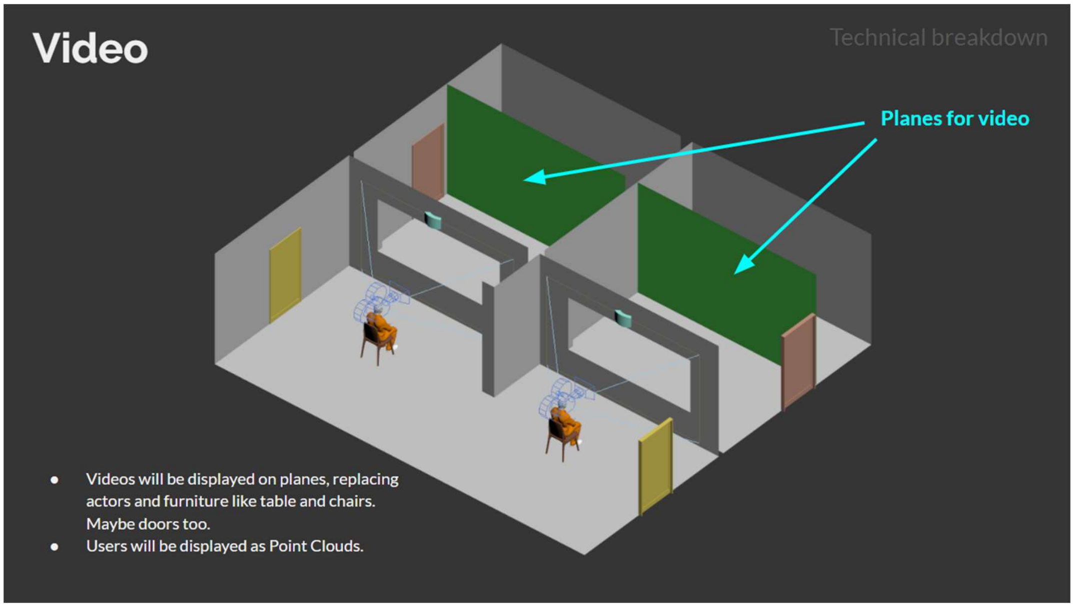

Figure 17: Scene composition (3D Billboards + 3D scene).

The videos recorded with the stereo rig (the interrogation rooms) are placed as a geometric plane inside the scene. The stereo video used has been stored using the Top/Bottom 3D format.

The video is then rendered with the scene.

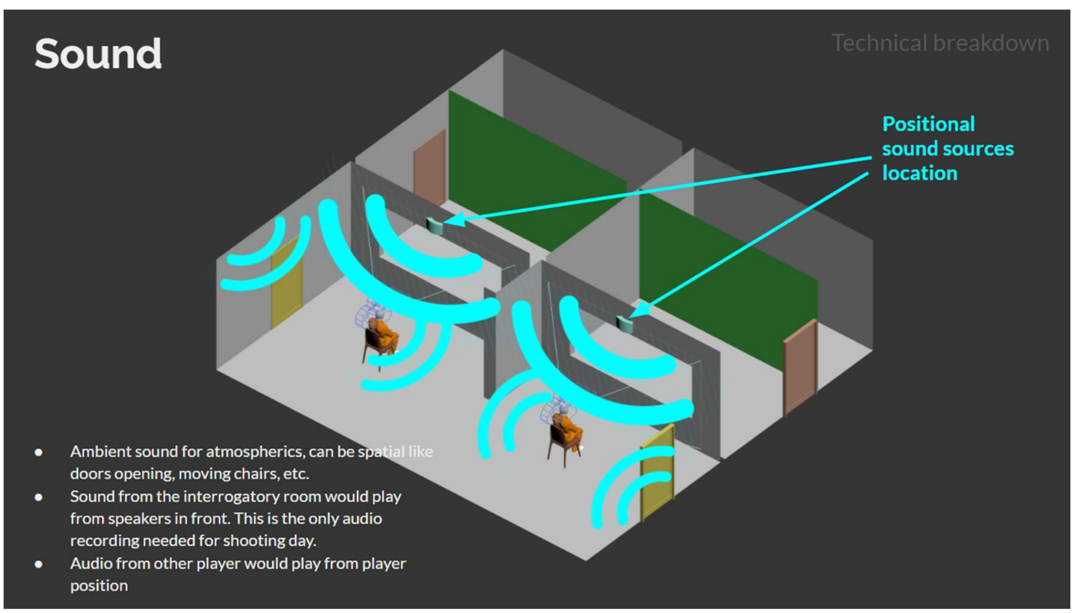

Figure 18: Sound design.

The audio content has been placed in the 3D engine as spatial content. The end-users would then be able to feel a more realistic experience given that the sounds will have a directional component.

For experimentation purposes, the consortium decided to create an additional version using scanned characters, with the purpose of analyse the technical differences, like the streaming performance, but also to have a knowledge about the subjective opinion that the end-users may provide about the quality of the experience.

The scanned characters have been created using a photogrammetric rig of 96 cameras, able to produce a geometric (avatar-like) representation of the actors.

Figure 19: 3D character capture (I).

The recorded 3d representations have been reduced in size and shape to meet the requirements of a real-time production scenario and then rigged for adaptation to the motion capture process.

Figure 20: 3D character capture (II).

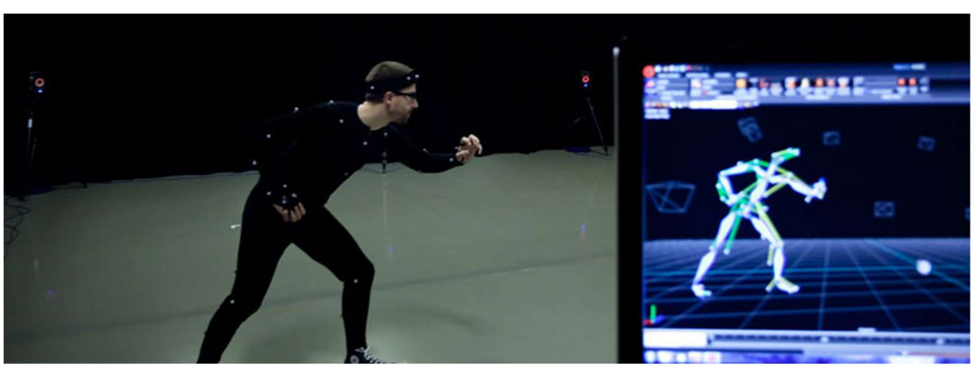

Figure 21: Motion tracking to animate pre-rigged characters.

After the actor's avatar creation, the interrogation scene was acted again with the actors playing wearing motion capture suits, in order to translate their movements to the virtual ones.

Thanks to the process previously described, we obtained several different media formats for the experiments:

- 3D environment and 3D characters
- 3D environment and video billboard representing the characters
- 360 stereoscopic video version of the interrogation scene
- 360 monoscopic video version of the interrogation scene
- Traditional 2D cropped version of the interrogation scene

3.2. Pilot 2

The main goal of Pilot 2 is to improve the immersive experience provided in Pilot 1. The way we aim to meet this goal is by providing a higher feeling of togetherness in the virtual world. In order to do so we have decided to proceed toward 2 directions: i) adding a live factor to the experience and ii) incrementing the number of end-users from 2 (Pilot 1) to 4 or more.

As we decided to consider each pilot as the follow up of the previous one, the story of Pilot 2 continues telling the events started in Pilot 1. Doing so, the experience will be catching the end-users attention and, at the same time, it will be able to show the technical advances reached by the project at the time of Pilot 2 release.

3.2.1. Plot

The story of Pilot 1 started in a crime scene followed by an interrogation room where the user is witnessing an interrogation from the transparent side of a one-way mirror. The interrogation is about a murder and the suspect has to answer to the questions of a policeman.

In Pilot 2, the story continues in a live news studio where a presenter will make a live connection with a reporter in the crime scene. This reporter, whose representation is transmitted live, is reading the news and reporting about the crime that happened in Pilot 1 and will interview the police officer in the murder scenario.

The presenter will have the help of an advisor, displayed in a virtual windows in the TV set.

3.2.2. Virtual Environment Scenario

In order to be able to provide, to the end-users, an immersive experience that accomplishes the goals of Pilot 2, the consortium partners have first suggested a set of proposals for the Virtual Environment Scenario. The proposal, which takes into account several VR rendering technologies, has been used as base for the discussion about which technology will suit better the purpose of the pilot.

The technologies proposed are listed below:

- Full 360
- Full 3D
- Mixed 3D/Billboards
- Mixed 3D/180
- Mixed 3D/180/Billboards

In the following sub-sections, each proposal is presented in detail focusing on pros and cons of each technology.

3.2.2.1. Full 360

The first option considered was a traditional 360 video. The 360 video is one of the easiest solutions in terms of production. The recording is done with a rig of cameras (Figure 22) pointing at several directions and, thanks to mosaicking techniques, the videos, from different cameras, are joint together in a 360 environment that, when seen in a VR HMD, places the end-user in the centre of the 360 world (Figure 23) which corresponds exactly to the position of the rig of cameras (Figure 24).

Figure 22: Example of a 360 camera rig.

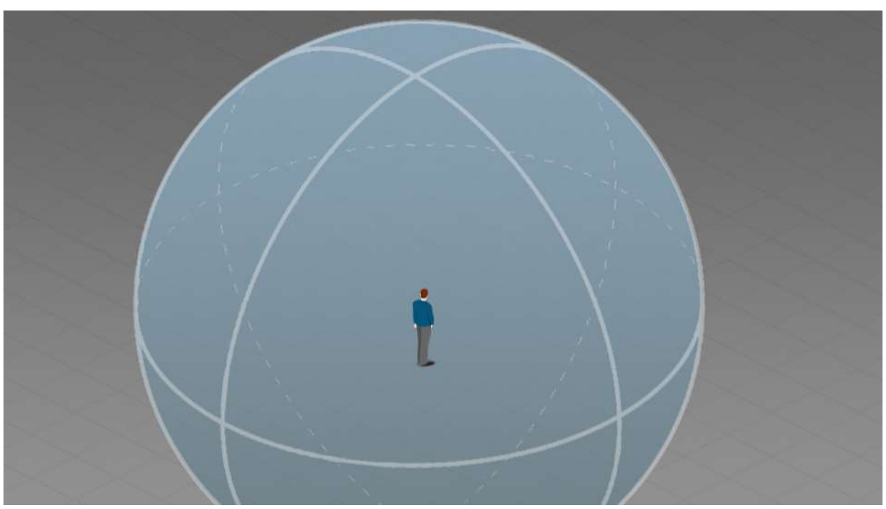

Figure 23: The end-user is the centre of the Virtual Environment.

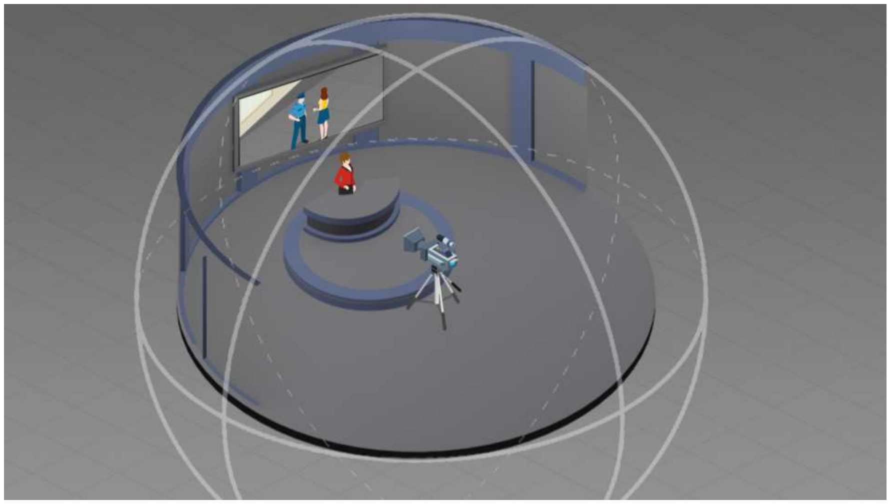

Figure 24: Example of the camera position recording a 360 video.

3.2.2.1.1. **Pros of the 360 video**

As mentioned above, the 360 video is the simplest solution in terms of production. The reason is that the recording is done through traditional 2D videos that are, then, used to compose the scene surrounding the end-user.

Another pro is given by the good image quality that can be reached using this technique. A 360 video is, indeed, the most realistic representation of a 3D content, given that the video is recorded with high definition and photographic quality.

3.2.2.1.2. **Cons of the 360 video**

However, even if the 360 video can be considered an easy solution providing a good quality, there are some technical issues to overcome.

First, the viewpoint is static, because it can be reproduced only where the camera-rig was physically placed. This means that, if the end users tries to move its position, even just slightly, the viewpoint will move as well producing a weird parallax feeling and reducing the quality of the immersive experience.

In addition to that there are several considerations to be done about the representation of the other end-users participating to the experience. The end-users are indeed transmitted as a volumetric 3D reconstruction of their bodies. This means that, in order to be represented in the 360 video of another end-user, their representation needs to be processed and transformed accordingly. The transformation can be done applying a 2D mapping of the 3D volumetric data of the end users. The volumetric data will be placed in the 3D world and then projected to the 360 sphere where the rest of the video is represented (Figure 25). The consortium has considered that the best solution, a part being challenging to achieve, would probably reduce the feeling of immersion when the users become more than 2.

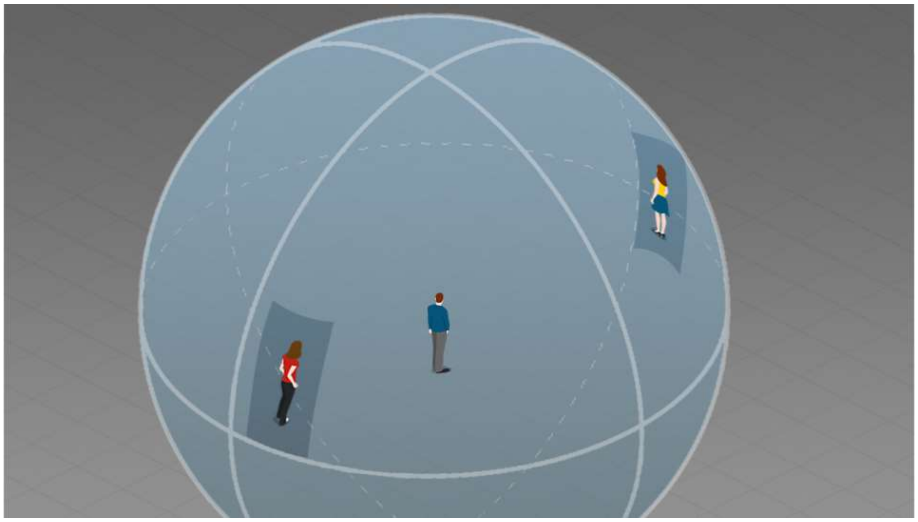

Figure 25: Example of 2D Mapping of the volumetric end-user data.

3.2.2.2. Full 3D

The Full 3D option is probably the most immersive one in terms of geometric reliability. The scenario is fully represented as a 3D environment where the end-users can move and see the scene from any viewpoint.

Figure 26 and Figure 27 show two examples of how the full 3D environment would look like.

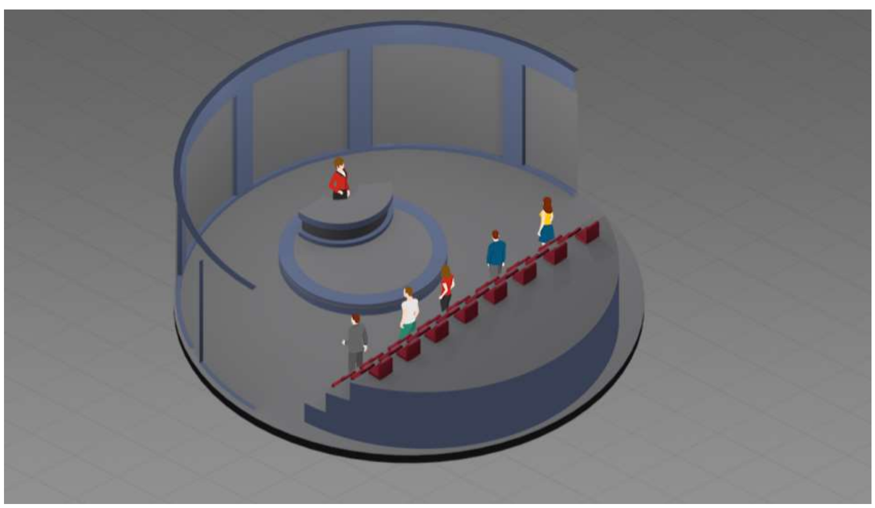

Figure 26: 3D rendered virtual environment. Indoor example.

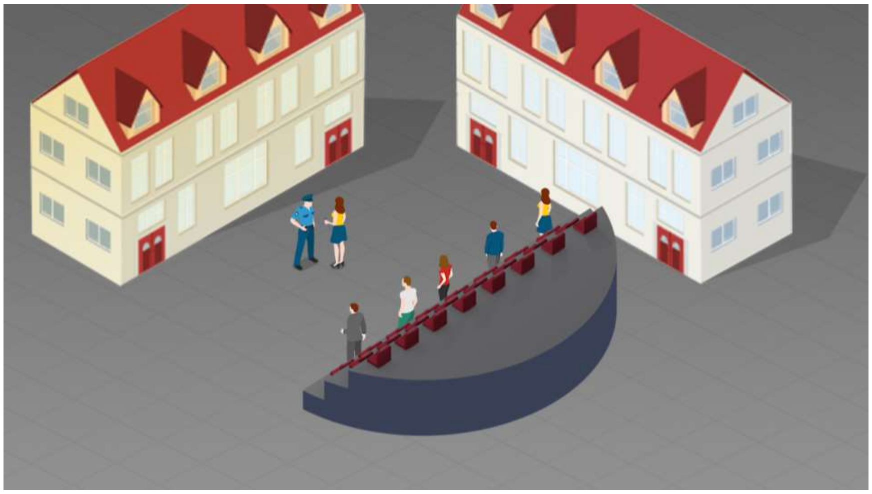

Figure 27: 3D rendered virtual environment. Outdoor example

The representation of the characters, in a full 3D environment, can be done using pre-rigged characters: the actors are initially scanned (Figure 28) to create 3D avatars (Figure 29) which are then animated using motion capture technologies.
This implies the need of a real-time streaming solution which can transmit live data from the motion capture location to the receivers (clients). The responsible partner, Artanim, has functionalities available which fits for the purpose: equipped with a motion capture suit, the actor's movements are captured and a skeleton is reconstructed. This skeleton data is subsequently streamed and used to drive the character's animation on the receiver end. While a dedicated motion capture studio with Vicon cameras and software is available and can be used for the motion capture (as for the characters in Pilot 1), a more flexible alternative, as the Xsens MVN setup, is also available. The latter, based on the inertial sensor technology, allows for more flexibility and higher mobility (which is an advantage when giving demos at fairs or conferences).

In addition to the pre-rigged option, the consortium has also considered the possibility of using, for the characters, the same technology used to represent the end-users (volumetric video as meshes or point clouds).

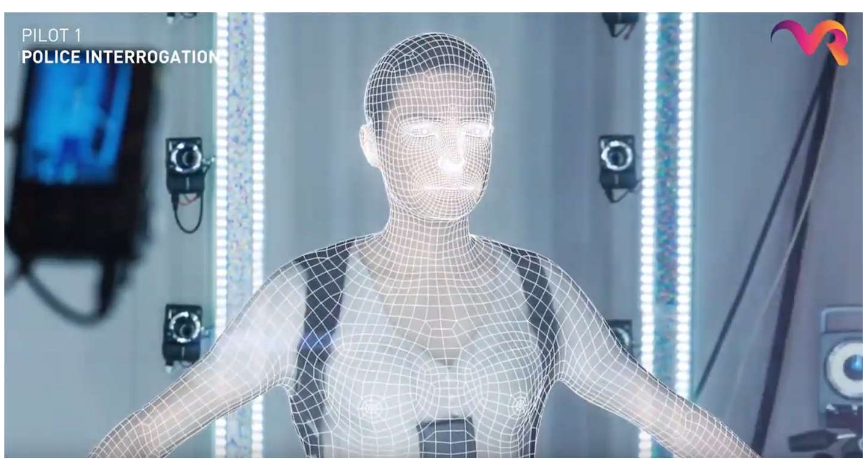

Figure 28: Scanning of pre-rigged characters.

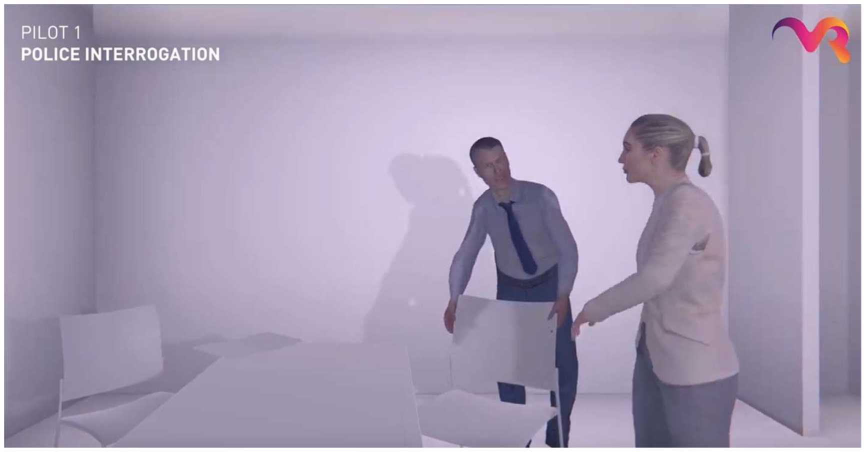

Figure 29: Pre-rigged characters actors animation

3.2.2.2.1. Pros of a Full 3D environment

The main advantage of using a full 3D environment is that the scene, being a 3D world, can be fully explored. In this way the end user would be able to move with 6 DoF in the virtual scene and to experience the full volume of objects, characters and other end-users.

Given the absence of a video component, the 3D environment, would also be completely free from the compression artefacts, typical of a transmitted video stream, that would affect the 360 video mentioned in the previous section.

In addition to the previous statements, the live component would be easily included with the 3D characters’ representation. In the case of the pre-rigged characters, only the data needed to represent the motion would be transmitted, making extremely low the load in terms of bandwidth needed, compared to some of the other capture options. The dataset consists of a set of position and orientation data for the rigged character’s bones. The quality of the end-result mostly relies on the quality of the pre-rigged and skinned character representation as presented under the production implications.

3.2.2.2.2. Cons of a Full 3D environment

In a full 3D environment, the main drawbacks are in the representation of the 3D characters.

Given that the transportation of a Vicon motion capture system is not feasible, the recording and transmission of the main actor (a news Presenter) needs to happen in the Artanim motion capture studio.

This problem can be reduced using an Xsens MVN suit (available at Artanim), based on the inertial sensor technology and that provides good animation results with higher mobility.

But any of those still requires every actor to be scanned, geometry decimated to low poly, textured and rigged, increasing the budget too much.

However, the visual representation of the Presenter is then limited to the 3D representation, given that the avatar has been previously scanned.

On the other hand, the volumetric video option, represented by meshes or point clouds, would have a photographic representation but it is still at an early stage and cannot provide the quality required to represent the actors of the experience.

In addition, given that the actors need to be considered as higher priority content, using one end-user slot for the Presenter would mean to renounce to one end-user for the experience.

3.2.2.3. Mixed Option (3D, Billboards, 180)

Given the several drawbacks of the solutions mentioned in the previous sections, the partners have decided to consider several options where several technical solutions were mixed together, reducing the weaknesses and exploiting the advantages.

The final decision has been made thanks to the participation of several VR industry professionals to some of the experiments and to the demonstration of the Pilot 1 experience, where different options for the representation of the scenario (360 and 3D) and of the characters (pre-rigged, billboards 2D and 3D) were considered (3.1.4). The experiments considered are the ones with the following ID's: i2CAT-2.1, i2CAT-2-2 and Artanim-2.1.

The outcome of the experiments will be reported in detail in the deliverable D4.4. However, in this section, the final decision and the reason behind are explained.

3.2.2.3.1. Scenario

In this subsection, the final decision for the kind of 3D scenario (full 3D or 360) chosen in the experience is presented.

The 360 video, as mentioned in Section 3.2.2.1, is the one providing the best photographic quality in terms of definition and realism of the representation. However, as one of the main issues is the projection of the other end-users (3D to 2D transformation), for the virtual environment it has been decided that the users will be placed in a part of the scenario represented as a Full 3D. In this way the volumetric representation (meshes or point clouds) will be naturally part of a 3D world and disadvantages of the 360 solution are avoided. However, as the representation of the actors in the crime scene does not suffer from the issue of having to be transformed from 3D to 2D, it is possible to represent them as part of a 360 video. The selected option is then a mixed 3D/180 environment (Figure 30).

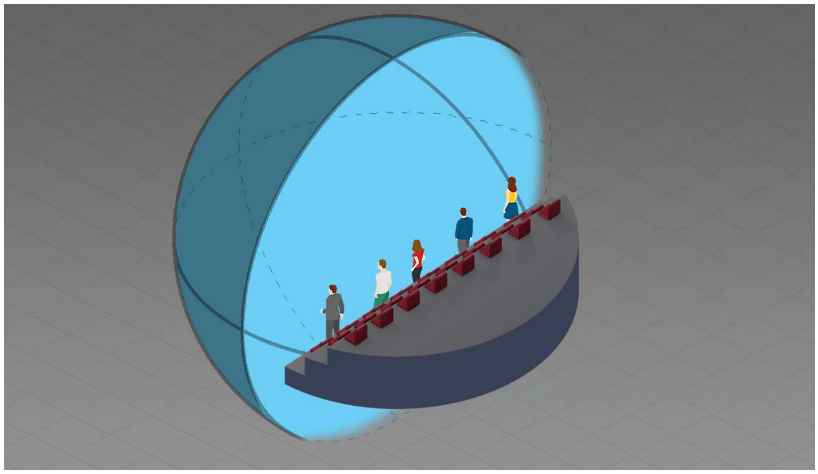

Figure 30: Example of a mixed 3D/180 scenario.

Thanks to this solution, both advantages of 3D and 360 are exploited and the weaknesses are reduced.

3.2.2.3.2. Presenter Representation

In the experiments mentioned at the beginning of this section (3.2.2.3), one of the outcomes has been that the best way to represent the characters in a reduced area of the 3D world is by placing a video billboard in the environment. The production details for the generation of a video billboard are depicted in Section 3.1.4 where it is possible to see that a billboard is a mono or stereoscopic video, placed in the 3D world in a way that, from a certain perspective, the end-user will experience the vision of it as if it is part of the three-dimensional environment. In Figure 31 it is possible to see how the billboard is represented facing the end-users, who will perceive it as part of a 3D scene.

The advantages of having a billboard video to represent the Presenter are several and can be divided into i) quality of the representation and ii) streaming of the content.

About the quality, it is easy to acknowledge that, as the actor is recorded using a traditional camera, the quality reached is comparable to the newest video technologies (such as 4K definition). This allow the end-user to experience a high quality representation of the main part of the experience. Also, given the compression technology available for traditional video, it is possible to deliver the content with a considerably high quality reducing the bandwidth needed to receive it. This affects positively the streaming of the content that can be delivered in real time and it is a crucial advantage given that the Presenter will be acting in a studio and recorded and transmitted in real time during the experience.

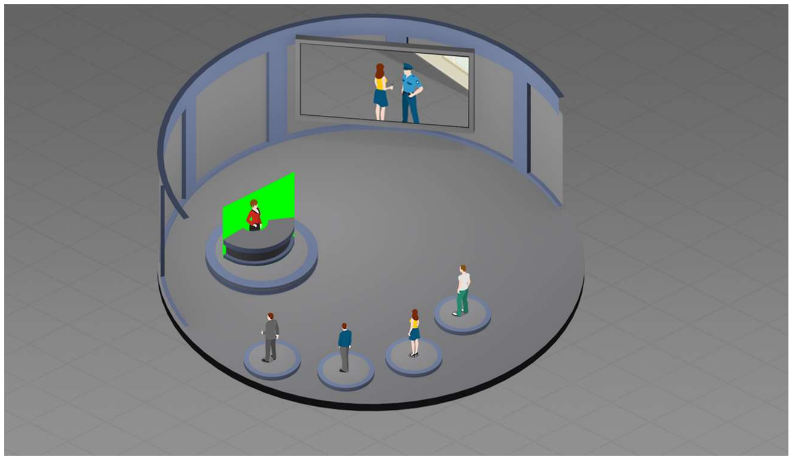

Figure 31: Example of a billboard video placed in a 3D environment.

In the end, as the production plan involves both indoor and outdoor experience, the experience will start in a full 3D environment representing a news studio (as in Figure 31), so that the end-user will start experiencing the immersive environment, the togetherness feeling given by the fact they can see the other end-users as audience of the show, and the live component given by the Presenter.

Afterwards, part of the studio changes into an outdoor scene where the 180 video is represented in order to show the crime scene to the end-users.

Figure 32: Outdoor Scenario Mixed 3D/180

3.2.3. Pre-production content design

The pre-production corresponds to the starting point of an audio-visual shooting. It is a mandatory process in which the production is detailed, taking into account every aspect:

1. Breakdown of shots and Timing
2. Financial budget Design: Study of forecast of expenses and provision of the same.
3. Definition of fiscal, legal and labour needs and services

4. Production design
5. Selection of a professional crew
6. Script
7. Casting
8. PPMS and calendar

Normally, tasks such as the elaboration of the script (first and successive versions of it) and the selection of the actors (the whole casting process) are usually done at the same time in order to manage time and schedules carefully.

VR-Together is a project with a strong post-production and CGI load, so it is very important to define everything in this phase to start works immediately, and take them simultaneously to the rest of the production

3.2.3.1. Strategy and resources identification

The process of production of an audio-visual piece has different and important aspects needed to be considered carefully. It is established, in this whole process:

1. Structure, Design and definition of the shooting: How the shooting is going to be
2. Technical aspects
3. Material, hardware and professional resources
4. Coordination of schedules, etc. This process was programmed to last for two weeks, in which the first stage corresponded to the identification of the tasks needed to be performed. This tasks, among others, are the following:
    a. Script development
    b. Casting selection
    c. Locations
    d. Pre-production meetings
    e. Shooting details
    f. 3D reconstruction

The following paragraphs detail all of the previous aspects.

3.2.3.2. Script development

Alongside bringing up the new technology developments and discovers, it was decided to continue the story of the murder of Ms. Armova, started in the first pilot of the project. On late March, Entropy Studio started to look for writers that could develop and produce a continuation of the story. Kathryn Gould was the writer chosen to explore more about the story of Ms. Armova and her mysterious death.

Kathryn Gould is an actress and writer, known for Gnaw (2017), Web of Lies (2012) and Easy Money (2008).

https://www.imdb.com/name/nm1699361/

The schedule time proposed for this tasks was 4 weeks, starting from late March until late April.

During the entire scheduled time for the development of the script, Ignacio Lacosta (from Entropy Studio) worked alongside Kathryn Gould to complete the task and offer a complete and well-rounded story. At the same time, Ana Revilla (Executive producer of Entropy Studio), Javier García-Lájara (VR Manager of Entropy Studio) and Fernando Pérez (Senior Programmer of Entropy Studio) supervised the process, avoiding unnecessary cost and extra post-production work that would postpone other tasks in the process.

The difficulty of writing and planning a script resides in the inherent characteristics of the project and the pilot, since it was needed to take into account the immersive aspect of the piece, among the following aspects:

1. The interaction between users and fake users: possibility of an alternation between TVM and point cloud. Fake users will take the place of missing users.
2. Presenter is Video Played in 3D environment
3. Presenter is Point Cloud in 3D environment.
4. Possibility of interacting with the presenter. The script must allow the change of content by the presenter in an agile way in case it is transmitted in real time, allowing him to launch or delay the connection with the reporter.
5. Reporter and Police officer are Video Player 180 Stereo

Normally, the script faces some changes depending on final thoughts, technical implications and production needs. Every little change counts as a new version of the script. The final and used version of the script was the fourth.

3.2.3.3. Casting

Ignacio Lacosta and Ana Revilla, from Entropy Studio, were the people responsible for the casting process and crew selection. Since the script had two new characters compared to Pilot 1, new faces were needed to play those new incorporations to the plot. Specifically, the new additions were the Presenter (Gavin Michaels, portrayed finally by Jimmy Shaw) and a news reporter from the crime scene (Maria Espinoza, portrayed by Verónica Polo).

- The character of the TV show is a man in his fifties, approximately. He is quite handsome, a great professional and a funny guy. He has the perfect personality to conduct a TV show that is in the middle of a quiz show and a news channel.
- On the other hand, the news reporter is a younger woman (probably in her forties), very energetic, passionate about her job, responsible and very curious, that's why her job fits her perfectly. The characters have a beef with each other, and the audience can notice some type of live tension.

The casting process was scheduled to last one single week, starting from a call made for experienced actors. A native/proficient knowledge of English was fundamental in order to make it. VR experience was also a positive point in order to consider them for the roles. The actors would have to play and perform a scene from the original script of the pilot in order to get the role. Each take of the candidates would be recorded in order to analyse carefully their acting abilities.

Once the casting team watches every actor perform, they'd make a pre-selection based on criteria made by the Executive Production team. The next step would be detailed analysis of the selected crew by the director of the pilot, to check if the actors fit into the role perfectly.

The resources and personal needed for this process are the following:

- A production team to set everything up
- Camera and lighting set
- Management software to order and classify the recorded files
- Software to edit standard life formats
- Personal Data protection forms to inform the participants about the procedure

3.2.3.4. Locations

Ignacio Lacosta, from Entropy Studio, was in charge of the selection of the locations to shoot the pilot. It was scheduled a single week for the searching process and the selection of places. The production team was also responsible for the selection of the places to record both scenes. As the script narrates, the story takes places in two different locations: a TV set, where Gavin Michaels presents the murder case and news to the audience and the crime scene set, where Maria Espinoza talks to the police in charge of the investigation of the murder.

The TV show needed to be recorded in an interior set with chroma (green background) to, later, edit the set to make it look like a real TV set. Two characters are needed to be recorded in this location: Gavin Michaels (TV host who presents the murder case to the audience) and Howard Chapman (technology experts who explains the use and effectiveness of the new AI technology that allows police to solve crimes). The lighting and chroma crew are crucial in an interior set shooting, since every technical aspect must be perfect in order to edit the footage.

For the second part of the pilot, the production team and the director looked for an exterior that looked like a British environment: industrial and street. Since the shooting is not indoors, an official permit of shooting is mandatory. Shooting outdoors also allows to make the most of the natural sunlight. However, an additional lighting set is needed to shoot every take with the perfect lighting, avoiding the natural movement of the sun. The police officer and the TV news reporter are needed in this scene.

3.2.3.5. Pre-production meetings (PPM)

Every audio-visual piece needs diverse meetings in order to organize and set up every single detail to maximize the results. The production team of Entropy Studio was in charge of this process. There are two type of mandatory meetings in every shooting: PPM script reading meetings and PPM shooting meetings. In some cases, more than one meeting per category is required.

- PPM Script reading: In this type of meeting, the director of the piece, the production team and the selected crew of actors can peacefully read the script together to detect mistakes or unexpected script necessities. For the pilot 2 of the VR-Together project, two of these meetings are required, with the following crew:
    - Actors
    - Director of the piece
    - Executive and line producers
    - VR Manager
- PPM Shooting: In this meeting, previous to the shooting, the whole team gathers to discuss the final decisions about the piece. Details such as the lighting requirements, the type of costumes, etc., are discussed, among other aspects. For the pilot 2 of the VR-Together project, two of these meetings are required, with the following crew:
    - Director of the piece
    - Executive and line producers
    - VR Manager
    - Location supervisor
    - Lighting technicians
    - Chroma technicians
    - Camera operator
    - Sound technician
    - Make-up artist & Stylist

3.2.3.6. Shooting

As it was stated in previous paragraphs, the shooting required two different locations: interior and exterior. However, due to the priority of maximize resources, it was decided to separate the shooting into two time slots on the same day: interior and exterior. The interior part corresponds to the scene where the TV show is presented, and the exterior scene corresponds to the interview of the Police Officer by the TV reporter. Prior every shooting, the production team prepares and sends a shooting order, where a schedule of the recording is explained, where is going to take place, the weather for that specific date, etc. In this order, each worker has a concrete hour to start their activity.

The interior shooting was scheduled to last about 5 hours, recording all the scenes where Gavin Michaels and Howard Chapman have text. Both parts are shot with a green screen background to later edit the background of the footage. The normal schedule in this type of shootings is the following:

- First, the technical team prepares every aspect to make the recording possible (lighting configuration, camera display, costumes preparation, etc.).
- After the actor crew has come, they are dressed up and made up to prepare the shooting.
- Prior to the first official take, a couple of rehearsals are made in order to check that everything is fine, the camera is recording perfectly, the sound is free of external noise and the lighting doesn’t produce any type of shadow.
- After this process, the shooting starts normally, repeating takes until the director considers what a good take is.

Capture devices used:

- Z CAM K1 Pro Cinematic VR180 Camera
    - Image Sensor: 2 x Sony EXMOR™ 4/3” CMOS
    - Lens: 2 x Premium fisheye lens f/2.5
    - Stereo Microphone
    - VR180 Format Certified by Google
    - Recording hardware:
        - Output Resolution: 6K @30fps / 4K @60fps (post stitching output)
        - Individual sensor resolution: 2880x2880 @30fps / 2120x1344 @60fps
        - File format: MOV
        - Video encoder: H264
        - Synchronization: Built-in Z CAM Sync hardware solution
    - Image settings
        - ISO: Auto, Manual (ISO 100~3200)
        - Exposure: Auto, Manual, Coordinated
        - White balance: Auto, Manual, Coordinated
        - Professional Gamma Setting: Z-Log (a proprietary gamma setting for professionals that require more room for colour grading.)
    - Storage
        - Media: 2 x SDXC Class 10 SD card. Support up to 128GB (recommend to use SanDisk Extreme Pro 95MB/s SD card. not included in the package)
        - File System: FAT / exFAT
    - Connectivity

        - Ethernet: 1 x Gigabit Ethernet port for control setting (use Z CAM™ Controller) and live streaming (use Z CAM™ WonderLive)
        - WIFI: 802.11n (equipped with external antenna port)
        - External microphone: 3.5mm jack
    - Power supply: DC 12V 3A, with LEMO connector supporting hot swap
    - Physical specifications
        - Aluminium alloy (full metal housing)
        - 0.7kg (without lens cover)
        - Dimensions (L xW xH): 125mm x 64mm x 87mm

- 1 x MSI Laptop
    - Intel Coffeelake i7-8750H processor
    - 16 GB, DDR4 RAM memory
    - 512GB SSD harddrive
    - Nvidia GeForce GTX 1070, 6GB GDDR5 graphics card
    - Windows 10 Home Advanced 64-bit OS

Capture accessories used:

- **Chroma Set:** Chroma green cyclorama in L: 7 m. x 5 m x 3.5 m (High)
- 3 ARRI spotlights: 2 x 650 W, 1 x 300 W
- Palio 2.44 x 2.44; 2 HMI 1000
- 4 SOFTBOX: 1,20 x 0,80 cm; 900 w.c/u
- Wireless Sound recording: 2 ZAXCOM TRX900 wireless systems with DPA4063 capsules and a Sound Devices 744t recorder, ambience was also recorded with a micro Schoeps CMIT5u

The exterior shooting was scheduled to last about 3 hours, since it is needed a specific time to make the most out of the natural lighting. Prior this time slot, the lighting set is prepared, and some other props to make the scene look natural.

The team needed for the whole shooting (morning and afternoon recording slots) was composed of the following professionals:

    - Director
    - Actors crew
    - Executive and line producers
    - VR Manager
    - Location supervisor
    - Lighting technicians
    - Chroma technicians
    - Camera operator
    - Sound technician
    - Make-up artist & Stylist

Capture devices and accessories used were the same as for the interior shooting.

There's also a catering service offered to every worker on set. This is previously prepared by the production team.

3.2.3.7. 3D set

The schedule for this task is 3.5 weeks. Javier Lajara and Ignacio Lacosta are in charge of the whole process.

3.2.3.7.1. Development of a 3D Set for the Chroma Scenario.

In this pre-production phase, a concept design and the generation of the volumes to be developed in CGI are made to be able to start the hyper-realistic production of the same in the production phase.

3.2.3.8. Live-presenter pre-production

3.2.3.8.1. Stereo Video Billboard representation

In the Sections 3.1.4 and 3.2.2, the several options to represent the actors have been discussed. The billboard solution already adopted for the interrogation scene in Pilot 1, has been the one selected also for the presenter, in Pilot 2.

The Presenter will be rendered using a pre-recorded video if the experience is going to be offline or a using streamed video if the experience is going live.

This video is recorded by a stereo camera, in this case, a 180º camera is sufficient.

We will use the ZCAM K1 Pro model, which allow us to record or stream the video in the case of Live Streaming.

3.2.3.8.2. Production

The recording set will be a standard set used for traditional video recording. It should have green walls and floor for correct chroma treatment and background removal.

The camera will be placed in front of the Presenter, in a middle point of the four users, since a single video feed will be used for every user. Some tests should be made to prove this approach valid.

In the virtual location of each user, we will place a TV screen with the video of each one of them, allowing the presenter to look at their picture and talk to them at the right moment.

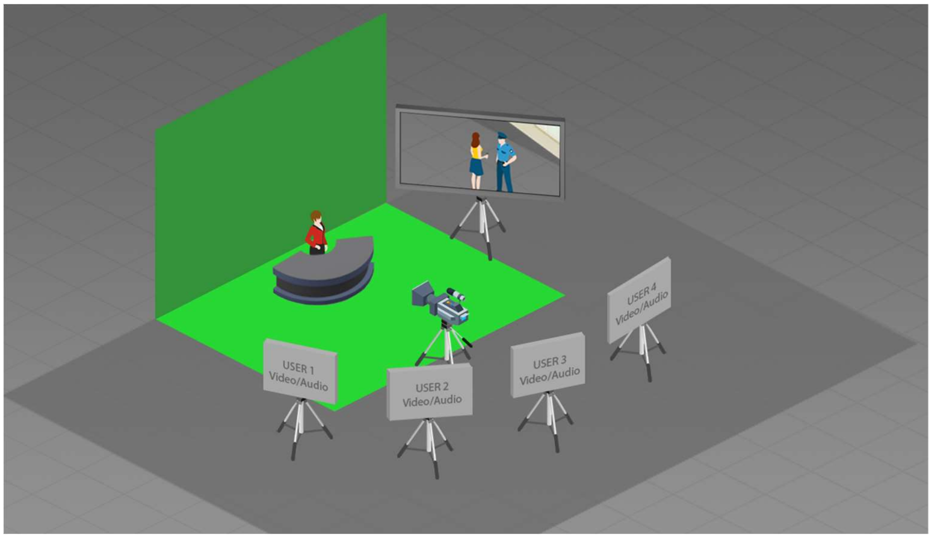

Figure 33:Chroma Set with screens (different users+ action recorded)

At the same time, the other part of the action will be recorded outside the TV set.
The Reporter and the Police Officer will be recorded on a street location, according to the story plot.

Figure 34: Proposal of Outdoor Scenario

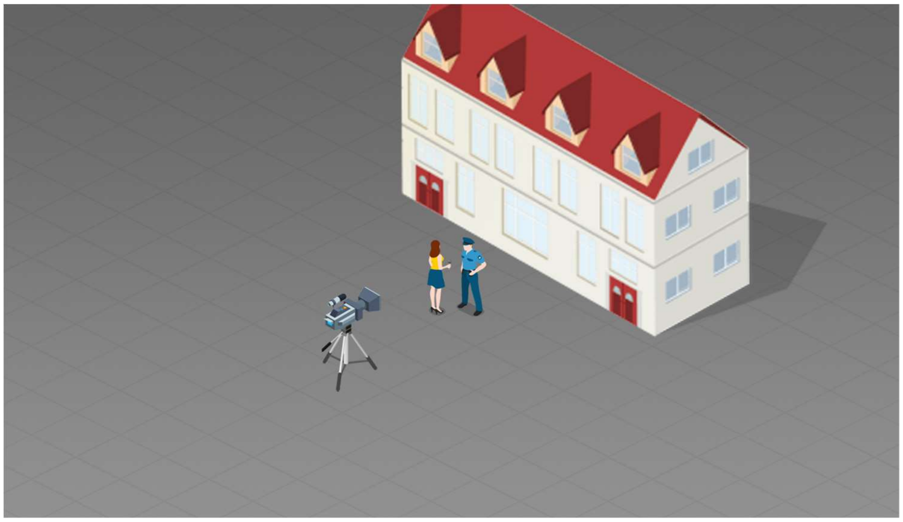

Figure 35: Story-board: Outdoor Scenario.

3.2.3.8.3. Background removal

The background removal process is going to be performed at client side. The user computer will receive a video signal and get this through a shader.
This shader will remove the green colour of the video feed obtaining transparency and opacity, find the silhouette of the subject, upscale the resulting image and smooth the border.
After this, a colour grade process should give us the final image that will be used to represent the Presenter in the virtual set.

3.2.4. Production

All the aforementioned actions have been part of the activities needed to kick off the actual production plan.
The work performed has been needed to create the virtual environment where the pilot experience is set and, in particular, the acting activities for:

- Live presenter
- Anchor journalist
- 180 scene
- Other actors

The detailed explanation of the work performed during the production, can be found in the deliverable D.4.3.

3.2.5. Pilot Action Calendar

Thanks to the lessons learned from the Pilot 1 activities, we have decided on a more defined approach for the planning of Pilot 2. In Table 6 (Pilot 2 Action Calendar) it is possible to notice that the production calendar has been created in a more structured way, dividing the full plan in main tasks and each task in sub-tasks with a certain amount of calendar weeks to be performed. Also, the experiments calendar, that for Pilot 1 was included in the production section, in this document, for Pilot 2, has been moved to the section dedicated to the experiments (see Section 6).

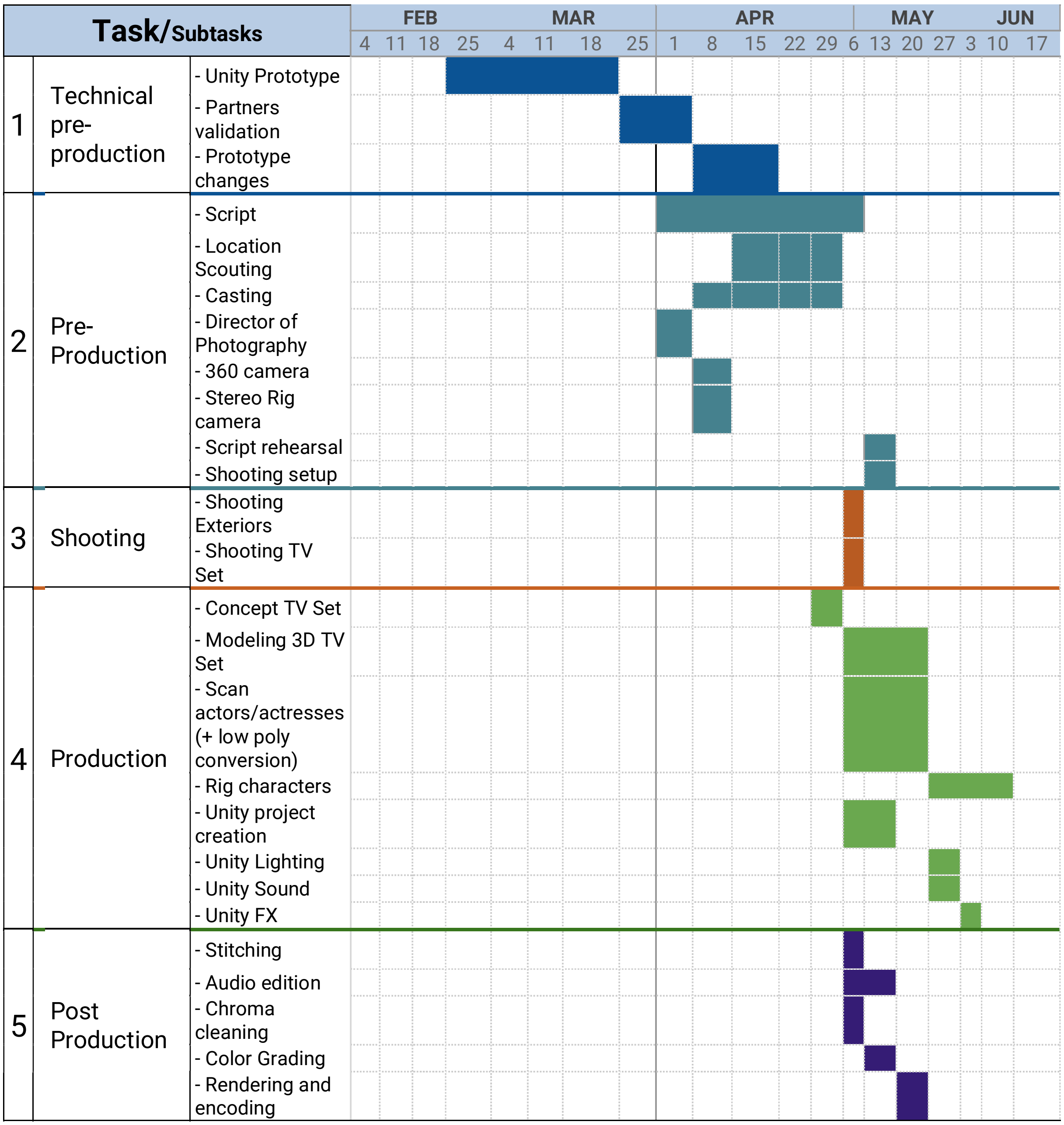

Table 6: Pilot 2 Action Calendar

3.3. Pilot 3

This section is out of the scope of the current document version.

4. Requirements Matrix and Process

This section provides an overview of the goals of the VR-Together project, by the point of view of services provided, technological advance and performance reached; such goals have been written down as requirements to be met by the platform. In the document D2.1 the requirements have been written as *User Scenarios* and *Use Cases* and then recompiled as Functional and Non Functional Requirements; all the information related to the requirements status at the time of the document D2.1 release has been included at the end of this document in the Annex III (0). The current status of the requirements is an updated and refined version of the one mentioned above, according to the outcome of several iterations of refinement and rewriting, in order to provide an accurate description of the status of the VR-Together platform at the time of the release of this document.

4.1. General Requirements Specification

VR-Together is a software platform for an end-to-end communication pipeline between end-users in virtual reality. The software is described by a number of requirements that define its functionalities and characteristics. In the following part we lay out the requirements gathering methodology, describing the attributes of the requirements and their meaning. Next, we present the User profiles, depicting the types of users that can use/participate on the platform and also lining up their characteristics. Last we mention the additional "environment" requirements such as assumptions or Interface requirements.

4.1.1. Requirements gathering methodology

The following part describes the requirements gathering methodology, the attributes of the requirements, how they are prioritized and the distinction regarding the software architecture component that they are referring to.

4.1.1.1. Requirements Gathering Techniques

In the VR-Together project we employ a number of techniques, for the requirements gathering that will define the project's end-results and features of the pilots to be developed. All possible requirements are gathered and classified accordingly by examining the attributes presented in the following sections of this chapter. It is important to mention that all the gathered requirements are judged upon the compliance with the core objectives of VR-Together, as they are described in the Grant Agreement (Section 1.1.2 Part B).

The requirements gathering techniques that we use in VR-Together are:

- Document Analysis: We identify and extract the requirements from documents generated in the VR-Together project, such as deliverables, reports, etc. As a first example of this we identified and extracted the requirements included in the Grant Agreement document. Requirements as "An end user MUST be able to see his own representation in the virtual space of VR-Together" or "An end user MUST be able to see the visual representation of another user in the virtual space of VR Together" are perfect examples of the generic platform requirements defined taking into account the Grant Agreement.

- Focus Groups: Groups of end-users will be asked to perform a specific task on the VR-Together platform. After each focus group gathering, participants are asked to give their feedback on a number of different aspects of the software platform itself as well as the overall experience. The collected feedback is analysed in an effort to determine additional requirements as well as refine and validate the existing ones. The requirements generated or validated by this part of the process are more focused on the way the end-users perceive the experience; examples are requirements like "End users SHOULD feel comfort in being immersed in the virtual space of VR-Together" or "An end-user MUST have a network latency allowing for seamless and natural communication and interaction with other users in the virtual space of VR-Together".
- Interviews: By conducting interviews with end-users and other important stakeholders we identify the expectations that VR-Together should meet. The expectations that align with the objectives stated in the Grant Agreement are translated to requirements and captured in the requirements matrix. An indicative example of the requirements that will be gathered with this process are the same as for the Focus Groups but also more specific requirements addressing the objective performance as "The VR-Together hardware capturing component/system MUST achieve a capture rate of at least 25 fps".
- Surveys/Questionnaires: Carefully designed surveys help in acquiring a large amount of user feedback in a short time as well as in a structured an easily comparable way. The design of the surveys includes questions where options are in the level of agreement/disagreement or rating of an argument. An example questionnaire can be found in Annex I (8).
- Other techniques: Depending on the occasion a number of additional processes could be used in order to generate requirements that would help in developing a higher quality end platform. These additional techniques could be brainstorming sessions, requirements gathering workshops, short interviews and discussions during exhibitions and conferences etc.

The task of requirements gathering is not a finite task with a specifically determined ending point. Thus, by practising the techniques mentioned above we will create new or amend requirements that will be included in future versions of the current document.

4.1.1.2. Types of requirements

VR-Together aims at gathering the software platform requirements from the view point of the end-user. For this we are focusing on two types of requirements:

- Functional requirements (FR): Define what the system must accomplish or must be able to do.
- Non-functional requirements[1] (NFR): The required overall attributes of the system, including portability, reliability, efficiency, human engineering, testability, understanding, and modifiability.

4.1.1.3. Prioritization of requirements

The description of a requirement must contain one of the following terms to define the prioritisation of the requirement: "must", "should", "could" or "won't". The definition of these

[1] A. Davis (1993). Software Requirements: Objects, Functions and States. Prentice Hall.

terms has been adopted from the MoSCoW prioritisation. Negative requirements such as “should not” and “shall not” are omitted, as they are not common in software development.

MoSCoW [1] defines the terms as follows:

MUST	Requirements labelled as ***MUST*** have to be included in the solution to be a success. Think of MUST as a requirement that without it the result is considered a failure.
SHOULD	***SHOULD*** requirements are as important as *MUST*, although *SHOULD* requirements are often not as critical or have workarounds, allowing another way of satisfying the requirement. They are important and of high value to the user but even without them the system could still be considered a success.
COULD	Requirements labelled as ***COULD*** are less critical and often seen as ‘nice to have’.
WON’T	***WON’T*** requirements are either least-critical or not appropriate at that time.

4.1.1.4. System component of requirements

The user requirements are based on the user scenario compilation, separated depending on the component of the VR-Together platform that they are related to.

The system reference for the requirements are categorised following the components as they are seen in the Architecture diagram included in D2.1 (Section 4.1):

- Capturing (CA)
- Encoding & Encapsulation (EE)
- Delivery (DE)
- Orchestration (OR)
- Play-out (PL)
- VRT (VR Together General)

If a requirement refers to a combination of different components within the platform, then the assigned value in the requirements matrix is: VR-Together (VRT)

The requirements matrix can be found in Section 9 of this document.

4.1.2. VR-Together User Profiles

There are three types of users that interact with the system:

- End-users of the native or web player (content consumer),
- Users that can set up, control, monitor and modify the course of the content consumption and social interaction actions (Researcher)
- Administrators

Each of these three types of users has different scope regarding the use of the VR-Together platform and for that reason they have a set of associated requirements as well as available functionalities. Below we give a description for the profile of each one of the different types.

4.1.2.1. End-User

The end user of the VR-Together platform is the content consumer of the VR-Together platform. It can be a person of any age, gender and condition, without acoustic or visual impairment and without any previous known problem while accessing contents using Head-Mounted Displays. End-users can use the web or native players to access the VR-Together contents, consume them, interact with other users participating in the experience, or interact with the content itself (in future versions of the VR-Together platform).

4.1.2.2. Administrator

The administrator of the VR-Together platform is able to create and set up the VR-Together experiences. Typically, the administrator will be able to set different parameters like the content sources, the available media representation formats used in a specific experience session or room, the format used to represent end-users in a specific session, spawn points where end-users are located inside a virtual environment, etc. The administrator will configure most of the previous parameters through a relevant graphical interface.

4.1.2.3. Researcher

The researcher in the VR-Together platform is typically a person who will be able to modify parameters of the experience for comparative research and monitor data collection processes. The researcher will also be able to configure specific instances of the players in lab environments.

4.1.3. Reference documentation

The VR Together experience makes use of the following standards:

- Production audio and video will use standards from MPEG to encode and package the content. [MPEG-4 ISO/IEC 14496, MPEG-H ISO/IEC 23008]
- The delivery of production content will use HTTP. [HTTP 1.1 RFC 2616]
- Audio, video, and depth information might be transported using WebRTC [WebRTC RFC 7478]
- Audio, video and 3D point clouds MPEG-DASH [MPEG DASH ISO/IEC 23009]
- 3D meshes will be used with TCP [TCP RFC 793] or a message broker [https://www.rabbitmq.com/]
- WebVR [Draft: https://w3c.github.io/webvr/], Webaudio [https://www.w3.org/TR/webaudio/], WebGL [https://www.khronos.org/webgl/]

4.1.4. Assumptions and dependencies

All the components the VR-Together platform is based on or is dependent from are properly described in D3.1 and D3.2.

4.1.5. Interface Requirements

Web player interface. Content consumption and social interaction will be accessed through a web application. The objective is to explore social VR cases that are easy to deploy, aiming at a social experience without demanding requirements in terms of equipment.

Native player interface. Content consumption and social interaction will be accessed through a native application based on Unity3D. The objective is to explore social VR cases with specific hardware deployments and higher rendering capabilities.

Admin interface for room configuration. The VR-Together system will be configurable by means of rooms. Here an administrator can define the number of end-users, user spawn points, content sources and other parameters related to the conditions in which the content consumption and social interaction happen. Additionally, the administrator can perform actions such as the initial calibration of capture system of modification of the capture parameters.

Researcher interface for experiment execution. To select a room and start the experience for different players as well as to modify room or player parameters according to the characteristics and configuration of the experiment to be executed.

4.1.6. Conclusions

This current section represents the list of generic rules that have been followed, and are being followed, for the definition of the final requirements of the VR-Together project. The current requirements status represents the application of these rules together with an iterative process of peer reviewing performed by the partners. The peer reviewing process follows an iterative application of the feedback of the experts of each field and it is fully tracked in the corresponding document *VR-Together Requirements Matrix*[2]. The next section (4.2) describes how each iteration is followed and how they generates an outcome that is the starting point for the following iteration.

4.2. Requirements update iterations

As explained in the introduction of this chapter, the requirement status at the time of the writing of this document corresponds to the outcome of several iterations of work done to improve the requirements initially stated and to adapt them better to the actual development of the platform.

In this section, the reader can find the table corresponding to the requirements status corresponding exactly to the time of the release of this document. The updated requirements have been discussed and processed among the partners of the projects following a well-defined process.

First of all, for each requirement, an owner has been identified depending on the area of expertise and responsibility; for example, regarding the production part, Entropy was responsible for most of the requirements. Each owner, or group of owners, is considered responsible also for the decision taken for each requirement, such as update, deprecation etc.

In order to reach an agreement a first iteration has been processed analysing all the requirements and checking if, for each of them, certain parameters were respected. The parameters considered were about the wording of the requirements, how clear they were, if they needed to be split into more requirements, if they were out of scope or not and if they were specific requirements for one pilot or for more. This first iteration ended with a proposed action for each requirement (i.e. deprecation, refinement, etc.).

[2] VR-Together Requirements Matrix
https://docs.google.com/spreadsheets/d/12IF74tYmMmCin5_hAgm0PL4ajia6Ii-WQ3fS-NWS5q0/edit?ts=5c87c8f2#gid=196101635

Afterwards, the second step has been to receive a feedback, about the proposed actions, directly from the partners responsible for each requirement. The process has followed a traditional Requirement peer review process where every partner involved in the decision was in charge to perform one of the following possible actions:

- Confirming the requirement
- Suggest to rephrase the requirement
- If needed, and if a rephrasing was requested, suggest the new requirement
- Suggest if the requirement needed to be split into 2 or more requirements
- If needed, and a splitting was requested, suggest the new requirements
- Suggest deprecation
- Confirm any of the above actions

All the discussion carried over to reach a decision has been fully tracked in the document *VR-Together Requirements Matrix*[3], where it is possible to see, in each tab, the status at the different stages of the requirements process. In particular, in the tab "D2.1 Analysis" we have followed a peer reviewing method tracking the discussion needed to reach an agreement about each single requirement. Those requirements deprecated have been struck-through in order to leave them in the history of the project and to avoid reusing the deleted ID's. For those requirements that have been updated, the peer review reaches an agreement with the definition of the new requirement and of an ID equal to the previous one plus a 1 in the suffix (i.e. FR.42.0 becomes FR.42.1).

In order to show an example of how the discussion have been tracked, the requirement of Table 7 has been provided. It is possible to notice that the discussion has been tracked writing a tag for every comment, indicating the date and the initials of the author of the comment or of the action (i.e. GC 19.03.2019). In addition, in this specific case, the requirement has been deprecated and three new requirements have been created. The old ID of the deprecated requirement has been eliminated and it will never be reused and the three new requirements have three new IDs.

ID	Description	Suggestions
~~NF.71.0~~	~~The VR-Together platform SHOULD allow a low quality point cloud to be decoded from a partial bitstream~~	[GC 19.03.2019 Agreed. NF.71 deprecated. New req: FR.152 The Point Clouds (PC) used to represent the VR together end user content COULD be transmitted using an adaptive bitrate streaming technique. NF.153 The adaptive bitrate streaming technique used should be the Dynamic Adaptive Streaming over HTTP (DASH) NF.154 The DASH adaptation set SHOULD include a low quality version of the PC user representation, for the DASH client to download it when the bandwidth conditions are bad NF.155 The DASH server COULD provide (as an option) a tile-based adaptation set, for the DASH client to download only the portion of the PC that falls in her/his field of view. [FDS 18.03.2019] New proposed reqs: 1. The DASH adaptation set must include a low quality version of the PC user representation, for the DASH client to download it when the bandwidth conditions are bad

[3] VR-Together Requirements Matrix https://docs.google.com/spreadsheets/d/12IF74tYmMmCin5_hAgm0PL4ajia6Ii-WQ3fS-NWS5q0/edit?ts=5c87c8f2#gid=196101635

		2. The DASH server must provide (as an option) a tile-based adaptation set, for the DASH client to download only the portion of the PC that falls in her/his field of view. [GC 14.03.2019] I suggest deprecating and add, if needed, new requirements according to the comments of FDS. CWI please agree and provide new reqs. [FDS 10.03.2019] I guess this applies only to the PC pipeline, thus to the Unity-based solution only (but to be checked with CERTH in case TVM encoder allows scalability). The CWI PC encoder does not have any scalable functionality: so, if this requirement refers to scalable video decoding (as it seems) I do not think this can be achieved within the project duration. On the other hand: a low quality PC version can be downloaded by the DASH client when the adaptation set includes a low quality version of the PC; a portion of the PC can be downloaded by the DASH client when tiling is used to encode the PC. In both these cases, there is no scalability (i.e., the low quality point cloud is NOT decoded from a partial bit stream). [GC 26.02.2019] In addition, I suggest to change to a FR [MM 24.02.19] See comments on the left (refinements)

Table 7: Example of the discussion followed to process a Requirement

When all the parties involved have come up with an agreement about each requirement, the process have been considered closed and the new requirements were included in the new matrix that can be found in the requirement document, in the tab "D2.2 Proposal". Most of the requirements IDs have been maintained as in the previous version (document D2.1 and Annex III (0)). Those requirements that where not considered as perfectly describing the feature mentioned, have been rephrased and updated. Some requirements have been deprecated because redundant or out of scope; in this case the ID have been removed and will never be reused for another requirement in order to provide a perfect traceability of the project evolution. The new requirements included have been defined with new and unique IDs.

The following iteration, has focused on the clustering of the requirements. The fields "Component" and "Title" have been processed with a twofold goal: first, all the partners have been requested to provide a feedback about the correctness of the components the requirements were assigned to, second the titles have been corrected and redefined so that they can represent a second clustering, more specific when compared to the components. In this way, all the tasks related to the requirements analysis will be facilitated by the new classification and the partners will be able to select only the requirements related to specific functions and components. An example of the new clustering process can be found in Table 8. The old Titles and Components, when updated, have been deprecated by a strike-through so that a backward tracking of the previous stage is always feasible. The new clustering can be found in the requirements document in the tab "D2.2 New Clustering" and, in this document, in Table 8.

ID	Component	Title	Component Correct?	Title Correct?	Title and Component Comment
FR.28.0	VRT	~~Facial expressions end users HMD removal~~	Yes	Yes	[FDS 29.04.2019] Title could be "HMD removal" maybe? or if you want to cluster under same title multiple items than you could use "Virtual Experience", "User representation" or

		VR experience			"Virtual content image quality". Component should be CA imo.
FR.30.1	~~VRT~~ PL	~~Offline content~~ VR Content	Yes	Yes	[GC 24.04.2019] Suggest changing component to PL. Suggest changing the Title to "VR Content" as the req mentions exactly that is not only offline

Table 8: Example of the new clustering process

At the moment of the release of this document, the following stage has already started. The current work is focusing on defining the performance values to be reached by the VR-Together technology. For some requirement it was already possible to write a new definition (i.e. NF.72.0 in Table 9) so that there is already a new proposal for many of them. The new proposal is under peer reviewing process similar to the previous iterations.

For other requirements, an experiment is needed before defining the final performance thresholds. So, currently, those requirements have been linked to one, or more, of the experiments planned for the second year of the project (6.4.2). In addition to the process mentioned above, a part from the previous list of responsible for each requirement, that was mentioning only the names of the companies, one, or more, partners have been nominated as responsible and they will be in charge of defining the values based on the outcomes of the experiments. An example of a requirement following this process is NF.75.0 (Table 9).

ID	Description	Update Suggestion	Experiment Linked	Responsible
NF.72.0	The VR-Together platform MUST be able to achieve a compression ratio of up to 1:10 in point cloud streams	The VR-Together platform MUST implement a volumetric video encoding system able to transmit a bit-stream that meets the target scenario bandwidth availability		Jack/Spiros
NF.75.0	The VR-Together platform SHOULD be able to evaluate the expected quality of experience according to the objective metrics TBD	To be written based on the outcome of the experiment	CWI2.1/CWI2.2	Jie

Table 9: Example of the current requirements update end linking to the experiments

In this same iteration, the process to define the final values needed to fully address the requirements is being performed. The partners have analysed the requirements one by one and have assigned, to each of them, one of the experiments planned for the second year of the project. The outcome of this iterations is tracked in the tab "Experiments/Requirements Links" of the *VR-Together Requirements Matrix*[4] document and an example of the linking can be found in Table 10.

[4] VR-Together Requirements Matrix https://docs.google.com/spreadsheets/d/12IF74tYmMmCin5_hAgm0PL4ajia6Ii-WQ3fS-NWS5q0/edit?ts=5c87c8f2#gid=196101635

Experiment Id	Linked Non Functional Requirements	Linked Functional Requirements
i2CAT-2.1	NF.104.1, NF.107.0, NF.109.1, NF.150.0, NF.66.1, NF.96.1, NF.122.0	FR.1.0,FR.2.0,FR.3.0,FR.5.0,FR.15.1, FR.16.0, FR.17.0, FR.27.1, FR.28.0, FR.31.0, FR.37.0

Table 10: Experiments/Requirements Linking example

The current section ends with the current and most updated, at the time of the release of this document, version of the Requirement Matrix. The full list of requirements can be found in Table 11.

ID	Component	Title	Description	Priority	Responsible
FR.1.0	PL	Self-representation	An end user MUST be able to see his own representation in the virtual space of VR-Together	MUST	CERTH, CWI, TNO
FR.2.0	PL	Users audio representation	An end user MUST be able to hear the sounds made by another user in the virtual space of VR Together	MUST	i2CAT/ENTROPY, CERTH, CWI, TNO, MSE
FR.3.0	PL	Users representation	An end user MUST be able to see the visual representation of another user in the virtual space of VR Together	MUST	CERTH, CWI, TNO
FR.4.0	CA	Audio Capturing setup	A location where the VR-Together platform's capturing setup is deployed MUST capture the audio generated by the user	MUST	i2CAT/ENTROPY, CERTH, CWI, TNO
FR.5.0	CA	Visual Capturing setup	A location where VR-Together platform's capturing setup is deployed MUST capture the visual representation of the user	MUST	CERTH, CWI, TNO
FR.8.0	DE	Latency	An end user MUST have a network latency allowing for seamless and natural communication and interaction with other users in the virtual space of VR-Together	MUST	CERTH, CWI, TNO, MSE
FR.9.1	OR	VR Scenario	An end-user client MUST be able to create a reconstruction of the virtual space of VR-Together.	MUST	ENTROPY, i2CAT, TNO
FR.11.0	VRT	VR content formats	End users SHOULD be able to see different examples of VR content formats	SHOULD	ENTROPY, i2CAT, ARTANIM, TNO
FR.12.1	VRT	VR content visual quality	End users MUST be able to see photorealistic VR contents	MUST	ENTROPY, i2CAT, TNO, CWI, CERTH
FR.13.0	VRT	Synchronization	End users in distributed locations sharing a virtual space MUST be able to see the same VR content at the same time	MUST	TNO, i2CAT
FR.15.1	PL	VR content visual quality	End users SHOULD see other users seamlessly blended in the virtual space of VR-Together. The Seamlessness evaluation will be performed by TBD.	SHOULD	ENTROPY, i2CAT

FR.16.0	VRT	VR Experience	End users SHOULD feel comfort in being immersed in the virtual space of VR-Together, at least for the duration of the pilot experience	SHOULD	-
FR.17.0	VRT	VR Experience	An end user SHOULD have an experience that visually and acoustically allows to perceive and understand the other participants' body language expressions.	SHOULD	TNO, CWI, CERTH, i2CAT
FR.18.0	PL	VR Content	The VR audio content MUST be directional giving the perception of point sources within the virtual space of VR-Together.	MUST	TNO, i2CAT
FR.20.0	VRT	End-user devices	End users MUST be able to access the VR-Together platform by using commercially available HMDs and capture systems	MUST	ENTROPY, i2CAT, ARTANIM, TNO, CERTH, CWI
FR.22.0	PL	VR content visual quality	End users, scene of action and characters SHOULD be able to be projected in the virtual space of VR-Together using different media formats. The resulting VR image should be a blend of different formats.	SHOULD	ENTROPY, i2CAT, ARTANIM, TNO, CERTH, CWI
FR.23.1	DE	Networks	The data transmission within VR-Together MUST be using commercial communication (e.g. MPEG-DASH) and media delivery networks (e.g. CDNs)	MUST	MSE
FR.24.0	EE	Networks	Media streams SHOULD provide adaptive quality to network, device and interface capabilities	SHOULD	MSE, CWI, CERTH, TNO?
FR.25.0	VRT	Web interface	End users MUST be able to access the VR-Together platform using a web application.	MUST	TNO
FR.26.0	VRT	Native interface	End users MUST be able to access VR-Together platform using a native application	MUST	i2CAT, ENTROPY
FR.27.1	VRT	VR Content Visual Quality	The virtual character representation MUST be detailed enough to allow for the recognition of facial expressions.	MUST	ENTROPY, ARTANIM, TNO, i2CAT

FR.28.0	VRT	VR experience	The level of detail of end-user representation in the virtual space of VR-Together MUST allow the recognition of facial expressions.	MUST	CERTH, CWI, TNO
FR.30.1	PL	VR Content	The VR-Together platform MUST be able to display the VR content which, depending on the configuration, can be either i) local or ii) stored in a network server	MUST	ENTROPY i2CAT, TNO
FR.31.0	VRT	VR Content	Illumination MUST be consistent in the whole experience	MUST	ENTROPY
FR.32.1	VRT	VR Content	The representations of the rendered characters, inside the virtual space of VR-Together Pilot 3, MUST be able to retarget their gaze according to the end-user's viewpoint	MUST	ARTANIM, ENTROPY, i2CAT, TNO
FR.33.1	VRT	VR Content	The representations of the rendered pre-rigged characters, inside the virtual space of VR-Together Pilot 3, MUST be able to retarget pointing gestures	MUST	ARTANIM, ENTROPY, i2CAT, TNO
FR.35.0	PL	VR content visual quality	The representations of the rendered characters inside the virtual space of VR-Together MUST have parallax and depth to allow for a 3D representation.	MUST	ARTANIM, ENTROPY, i2CAT,TNO
FR.36.0	PL	VR content visual quality	The end-user inside the virtual space of VR-Together MUST be able to perceive the 3D appearance of the characters (parallax, depth)	MUST	ARTANIM, ENTROPY, i2CAT,TNO
FR.37.0	VRT	VR Experience	The end-user inside the virtual space of VR-Together MUST be able to rotate their head and have certain level of translation capacity while seated (3DoF+)	MUST	i2CAT, ENTROPY
FR.41.1	OR	Active watch	The end-user inside the virtual space of VR-Together Pilot 3 MUST be able to become a character within the storyline that is being projected	MUST	
FR.42.1	VRT	Movement	The end-user inside the virtual space of VR-Together Pilot 3 MUST be able to move (translation). 6DoF	MUST	

FR.43.1	OR	Derived actions	The end-user's actions inside the virtual space of VR-Together Pilot 3 MUST lead to changes in the storyline that is being projected	MUST	
FR.44.1	VRT	Pattern recognition interaction	The VR-Together Pilot 3 platform MUST support multi modal pattern recognition mechanics for changing the storyline according to user's choices	MUST	
FR.45.1	VRT	Pointing interaction	The VR-Together Pilot 3 platform MUST be able to recognize pointing gestures of end-users and change the storyline accordingly	MUST	
FR.46.1	VRT	Speech interaction	The VR-Together Pilot 3 platform MUST be able to recognize the speech of end-users and change the storyline accordingly	MUST	
FR.48.1	VRT	Interactive character	The system for Pilot 3 SHOULD integrate and use interactive character animation	SHOULD	ARTANIM
FR.49.0	VRT	Networks	The VR-Together platform MUST support bandwidth configuration options for the end user	MUST	TNO, i2CAT, CWI, CERTH, MSE
FR.50.0	VRT	Networks	The VR-Together platform MUST support delay constraint configuration options for the end user	MUST	TNO, i2CAT, CWI, CERTH, MSE
FR.51.1	VRT	Self-representation	The VR-Together platform MUST support self-representation projection configuration options for the end user.	MUST	TNO, CWI, CERTH
FR.54.0	VRT	VR Scenario	The VR-Together platform MUST allow one end user to see a dynamic projection of another end-user's body representation within the virtual space.	MUST	CWI, CERTH, TNO
FR.55.0	VRT	VR Scenario	The VR-Together platform MUST allow one end user to see the projection of another end-user's body representation positioned at various distances within the virtual space.	MUST	i2CAT, TNO
FR.57.0	CA	**RGB-D capture**	The VR-Together hardware capturing component/system MUST capture RGB-D data from 4 RGB-D devices connected to 4 capturing nodes (RGB-D nodes)	MUST	CERTH
FR.58.0	CA	RGB-D Capture	The VR-Together hardware capturing component/system RGB-D devices SHOULD be automatically calibrated (extrinsic calibration).	SHOULD	CERTH

FR.59.0	CA	RGB-D Capture	The RGB-D frames from the RGB-D nodes MUST be synchronized and grouped in a central node, resulting in a RGB-D group frame.	MUST	CERTH
FR.60.0	CA	VR content formats	The VR-Together platform MUST process end-user's live colored 3D point cloud to reconstruct a 3D time-varying mesh in real-time.	MUST	CERTH
NF.66.1	CA	Latency	The input image captured by the hardware sensors of the capturing component MUST use a framerate of at least 25 fps.	MUST	CERTH, TNO
NF.67.0	CA	Face capture	The VR-Together hardware capturing component/system MUST capture the end-user's face from at least two different sides.	MUST	CERTH, TNO
NF.72.0	EE	Compression	The VR-Together platform MUST be able to achieve a compression ratio of up to 1:10 in point cloud streams	MUST	CWI
NF.73.1		Latency	The VR-Together platform MUST achieve an end to end (capture to projection) latency that is lower than TBD.	MUST	CERTH, CWI
NF.74.1	EE	Compression	The VR-Together platform SHOULD support point cloud compression of arbitrary topology (Topology TBD).	SHOULD	CWI
NF.75.0	VRT	Quality assessment	The VR-Together platform SHOULD be able to evaluate the expected quality of experience according to the objective metrics TBD	SHOULD	CWI, i2CAT
NF.77.0	EE	Compression	The VR-Together platform SHOULD be able to achieve a compression ratio of up to 1:30 for textured mesh (3D geometry and textures) content	SHOULD	CERTH
NF.78.0	EE	Compression	The VR-Together platform MUST support compression for textured 3D time varying mesh content of arbitrary topology.	MUST	CERTH
NF.92.1	OR	Configuration	The VR-Together platform orchestration module MUST be able to configure the native end-user play-out component	MUST	MSE, VO
NF.95.0	PL	VR content visual quality	The VR-Together platform MUST support playback of end-user's representation of at least 960x540 pixels	MUST	CWI, CERTH, i2CAT, ARTANIM, ENTROPY, TNO

NF.96.1	PL	VR Content Visual Quality	The VR-Together platform MUST support playback of end users representation at a framerate of at least TBD fps.	MUST	CWI, CERTH, i2CAT, ARTANIM, ENTROPY, TNO
NF.99.0	PL	WebVR	The VR-Together play-out component's web player MUST operate in a browser that supports WebVR and A-frame.	MUST	TNO
	PL	VR Content Format	The VR-Together play-out component's native player MUST support the reproduction of hybrid VR contents (TBD what is the hybrid VR content) in virtual space.	MUST	ENTROPY, ARTANIM, i2CAT, MSE
NF.107.0	PL	VR Content Visual Quality	The VR-Together play-out component's native player SHOULD be able to alter the lighting of specific objects within the virtual space, on the basis of custom shaders.	SHOULD	i2CAT, ENTROPY
NF.109.1	PL	VR Content Visual Quality	The VR-Together play-out component's native player stereo effective display resolution MUST be up to 4K.	MUST	i2CAT, ENTROPY
NF.110.1	VRT	Latency	The VR-Together play-out component's native player self-representation projection MUST have latency under TBD.	MUST	CERTH, CWI, TNO
NF.111.1	PL	Synchronization	The VR-Together play-out component's native player SHOULD support synchronization between different input formats with less than TBD of delay	SHOULD	CERTH, CWI, i2CAT, ENTROPY, MSE
NF.112.1	PL	Synchronization	The VR-Together play-out component's different players SHOULD support synchronization of frame accurate with a delay lower than TBD.	SHOULD	i2CAT, ENTROPY, MSE
FR.115.0	CA	Quality Assessment	The VR-Together platform capturing component SHOULD record and store the recordings of the HMD for further future analysis purposes.	SHOULD	i2CAT, ENTROPY
FR.117.0	OR	Configuration	The VR-Together platform orchestration component MUST support remote operation.	MUST	MSE, VO

FR.118.0	OR	Configuration	The VR-Together platform orchestration component, for Pilot 1, MUST manage sessions where 2 end-users participate in a virtual space.	MUST	MSE, VO
FR.119.0	OR	Configuration	The VR-Together platform orchestration component, for Pilot 2 and 3, MUST manage sessions where at least 2 end-users participate in a virtual space.	MUST	MSE, VO
FR.120.0	OR	Configuration	The VR-Together platform orchestration component, for Pilot 1, SHOULD support at least one session.	SHOULD	MSE, VO
FR.121.0	OR	Configuration	The VR-Together platform orchestration component, for Pilot 2 and 3, SHOULD support at least two parallel sessions.	SHOULD	MSE, VO
NF.122.0	CA	RGB-D Capture Framerate	The VR-Together hardware capturing component/system MUST achieve a capture rate of at least 25 fps.	MUST	CERTH
NF.123.0	CA	Latency	The VR-Together platform MUST perform the People live 3d reconstruction with a delay lower than 80ms.	MUST	CERTH, TNO
FR.124.0	CA	VR experience	The VR-Together hardware capturing component/system MUST store the captured end-user's face data. The information must be stored (on disk or in memory) and must be accessible in real-time by the face inpainting algorithm.	MUST	CERTH, TNO
FR.125.0	CA	Benchmarking	The VR-Together platform MUST record the position of the end user in the 3D scene at regular time intervals.	MUST	ENTROPY, i2CAT, ARTANIM, TNO, CERTH, CWI, MSE
FR.126.0	CA	Benchmarking	The VR-Together platform MUST record the viewport video visualized by each end user with timestamp.	MUST	ENTROPY, i2CAT, ARTANIM, TNO, CERTH, CWI, MSE
FR.127.0	CA	Benchmarking	The VR-Together platform MUST record the audio information (speech) from the end user with timestamp.	MUST	ENTROPY, i2CAT, ARTANIM, TNO, CERTH, CWI, MSE

FR.128.0	OR	VR Scenario	The VR-Together Pilot 1 platform MUST allow for 2 end users to simultaneously be in the same virtual space.	MUST	ENTROPY, i2CAT, ARTANIM, TNO, CERTH, CWI, MSE
FR.129.0	OR	VR Scenario	The VR-Together Pilot 2 platform MUST allow for 4 end users to simultaneously be in the same virtual space.	MUST	ENTROPY, i2CAT, ARTANIM, TNO, CERTH, CWI, MSE
FR.130.0	OR	VR Scenario	The VR-Together Pilot 3 platform MUST allow for 10 end users to simultaneously be in the same virtual space.	MUST	ENTROPY, i2CAT, ARTANIM, TNO, CERTH, CWI, MSE
FR.132.0	VRT	Compression	The VR-Together platform MUST support TVM compression configuration options for the end-user.	MUST	CERTH
FR.134.0	PL	VR Content Visual Quality	The VR-Together play-out component's native player SHOULD be able to alter the lighting of specific objects within the virtual space, on the basis of custom shaders.	SHOULD	i2CAT, ENTROPY
FR.135.0	PL	VR Experience	The VR-Together play-out component's native player MUST be able to reproduce content adapted to 3DoF or 3DoF+ movements.	SHOULD	i2CAT, ENTROPY
FR.136.0	EE	Compression	The VR-Together platform MUST use typical browser supported audio encoding.	MUST	TNO, MSE
FR.137.0	EE	Compression	The VR-Together platform MUST use typical browser supported video encoding.	MUST	TNO, MSE
FR.138.0	EE	Compression	The VR-Together platform MUST use typical browser supported audio encapsulation.	MUST	TNO, MSE
FR.139.0	EE	Compression	The VR-Together platform MUST use typical browser supported video encapsulation.	MUST	TNO, MSE
FR.140.0	PL	VR Content	The VR-Together platform web player MUST support playback of 2D VR video content.	MUST	TNO
FR.141.0	PL	VR Content	The VR-Together platform web player MUST support playback of 2D end-user representation projection.	MUST	TNO

FR.142.0	PL	VR Content	The VR-Together play-out component platform SHOULD support spatial audio.	SHOULD	i2CAT, ENTROPY, TNO
FR.143.0	PL	VR Content	The VR-Together play-out component MUST support input of separate VR content and end-user representations streams.	MUST	CWI, CERTH, MSE, TNO
FR.144.0	PL	Network	The VR-Together play-out component's web player SHOULD support content bandwidth adaptation.	SHOULD	TNO
FR.145.0	CA	Synchronization	The VR-Together platform capturing component MUST timestamp media content in relation to a platform-wide common clock.	MUST	i2CAT, ENTROPY, MSE, CWI, CERTH
FR.146.0	PL	VR Content	The native player MUST support play-out of content for Point Clouds.	MUST	ENTROPY, ARTANIM, i2CAT
FR.147.0	PL	VR Content	The native player MUST support play-out of content for Static/Dynamic meshes.	MUST	ENTROPY, ARTANIM, i2CAT
FR.148.0	PL	VR Content	The native player MUST support play-out of content for mono/stereo 2d video.	MUST	ENTROPY, ARTANIM, i2CAT
FR.149.0	PL	VR Content	The native player MUST support play-out of content for 360 video.	MUST	ENTROPY, ARTANIM, i2CAT
NF.150.0	PL	VR Content Visual Quality	The VR-Together play-out component's native player mono effective display resolution MUST be up to 2K.	MUST	i2CAT, ENTROPY
NF.151.0	OR	Configuration	The VR-Together platform orchestration module SHOULD be able to configure the web-client end-user play-out component.	MUST	MSE, VO
FR.152.0	PL	Network	The Point Clouds (PC) used to represent the VR together end user content COULD be transmitted using an adaptive bitrate streaming technique.	COULD	CWI
NF.153.0	PL	Network	The adaptive bitrate streaming technique used SHOULD be the Dynamic Adaptive Streaming over HTTP (DASH)	SHOULD	CWI

NF.154.0	PL	Network	NF.154 The DASH adaptation set SHOULD include a low quality version of the PC user representation, for the DASH client to download it when the bandwidth conditions are bad	SHOULD	CWI
NF.155.0	PL	Network	NF.155 The DASH server COULD provide (as an option) a tile-based adaptation set, for the DASH client to download only the portion of the PC that falls in her/his field of view.	COULD	CWI
FR.156.0	VRT	VR content	End users SHOULD be able to see photorealistic Live content (mono or stereoscopic video) in the VR environment	MUST	ENTROPY, i2CAT, TNO, CWI, CERTH

Table 11: Current status of the Requirement

5. ARCHITECTURE

In this section we describe the Architecture of the VR-Together platform from the point of view of both the Software platform (5.1) and the Hardware setup (5.2). After this description we analyse the architecture implications of VR-Together as a product.

5.1. Software architecture

The information presented in this section describes the work done on the platform for Pilot 1 and Pilot 2, highlighting the differences between the two pilots and emphasizing the novelties that the current development introduced with respect to the platform status when the document D2.1 was released.

Figure 36 shows the high-level architecture of the VR-Together system designed for Pilot 2. The VR-Together platform is described in a traditional production to consumption chain: audio-visual information flows from capturing to playout are portrayed together with the additional functional components. The modules and components form the current Software Platform of VR-Together.

In order to provide, to the reader, a follow up of the difference, in terms of development, between Pilot 1 and 2, we have included in this document also the previous architecture, corresponding to Pilot 1 (Figure 37), where it can be noticed that the system has been designed for a simple communication between 2 users. Pilot 2 architecture shows the main differences from Pilot 1, such as the presence of more users, the presence of a live component and, last the need of a server that, together with the new orchestrator, manages the delivery of the several streams.

The delta technology evolution between Pilot 1 and Pilot 2 following the requirements specification is depicted in the Table Table 12.

In the remaining part of this section, we present an extended and detailed software component architecture description of the VR-Together platform, for the release of Pilot 2.

In the following architecture description, we refer to:

- A "**component**" as a conceptual entity related to a general task within the end-to-end communication system. We identify the following five components/general tasks:
  - Capturing
  - Encoding & Encapsulation
  - Delivery
  - Orchestration
  - Play-Out
- A "**module**" as a building block that actually performs a specific technical task. To perform the general task described by a component, multiple modules are needed. The modules within each component are listed and described hereafter.
- A "**user**" as a group of components related to the end-user client definition.
  A user entity instance is considered independent and replicable N times to grant platform scalability. We consider being located on the end-user client side the three components:
  - Capturing
  - Encoding and encapsulation
  - Playout

- A "**server**" as a generic entity group that designates centralized and remotely operational components that interact with each user entity. Although they can be hosted on different server location, we define the two components to be remotely accessible:
    - Delivery
    - Orchestration
- A "**live**" node as a remote entity dedicated to recording and streaming to connected users of the platform a presenter in live.

		PILOT 1		PILOT 2	
Technology Delta	**REQ Title**	**Comment**	**REQ**	**Comment**	**REQ**
User client definition	Configuration / Scalability	Capture, Encoding and Playout components are distinctly separated. Each one maintains a communication with the Orchestration service. Multiple Orchestration APIs required to manage the Orchestrator communication with each component. Here, one session is available for two connected users.	FR. 120.0	Capture, Encoding and Playout components constitute the **User** entity group (the User client). The Orchestrator is connected to the Play-out component with the User-Manager module which manages the other components. One unified Orchestration API allows managing multiple sessions of connected users.	FR. 121.0
Number of connected Users	Configuration	Pilot 1 platform is able to manage statically two connected users in the same time and virtual space.	FR. 118.0	The **User** entity group is considered as a template and duplicable many times. Technically the Pilot 2 platform manages from one to four users connected in the same time and virtual space.	FR. 119.0
Live node definition	VR Content	No live content	-	The **Live** node provides the end-user a live video stream recorded from a specific hardware setup which support: stereoscopic recording and rendering, photorealistic capture (Live-Chroma).	FR. 156.0
Central server definition	Configuration / Scalability	Direct connection between two connected users through a single direct Delivery component.	-	The Delivery of multiple **Users** streams are processed by a central server component (SFU/MCU) managed by the Orchestration component. Advanced forwarding mechanisms allows to adapt streams Delivery according to **User** capabilities (bandwidth, quality, …)	FR.49 FR. 119.0

Table 12: Technologies delta table between Pilot 1 and Pilot 2 platforms

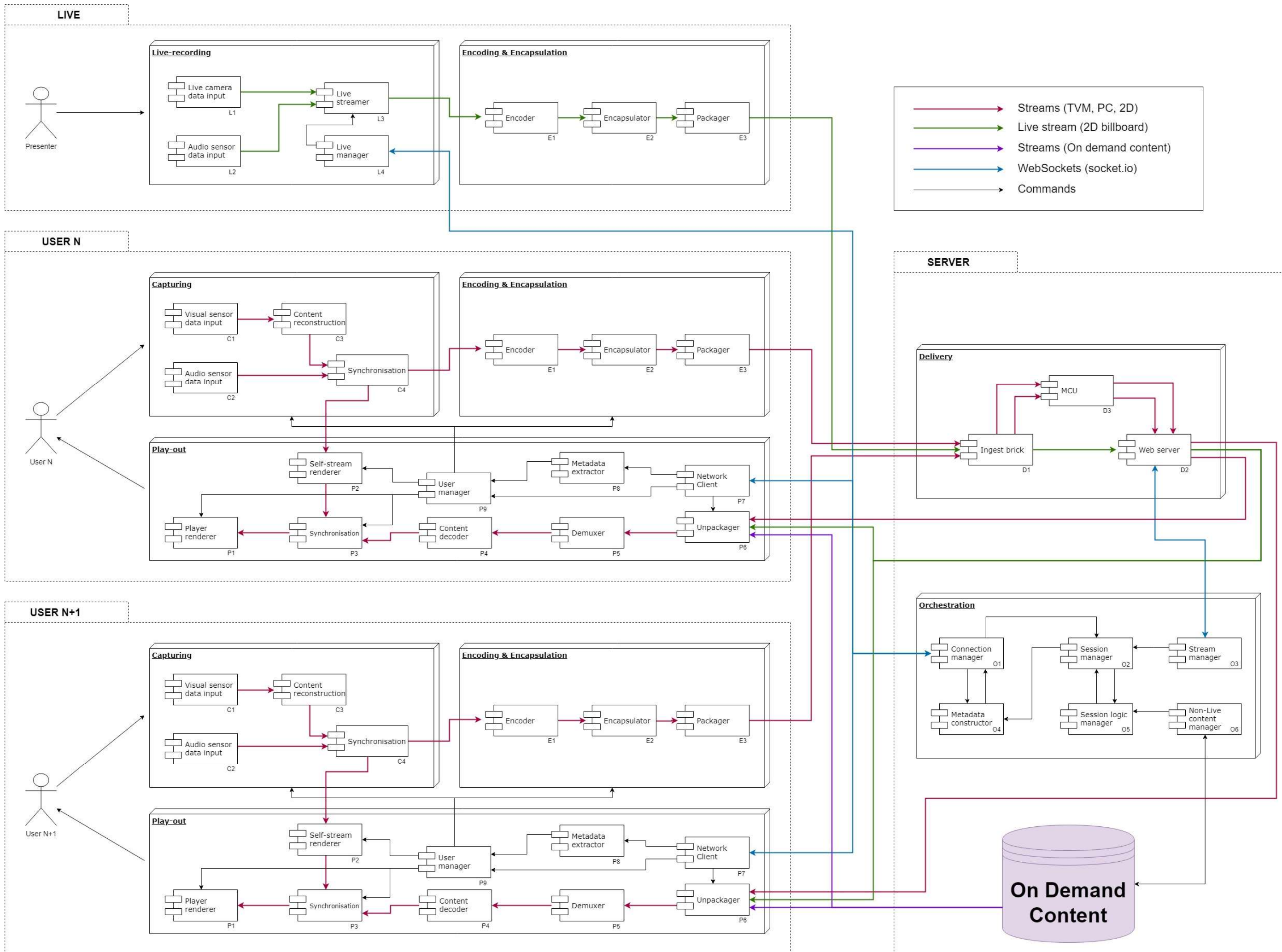

Figure 36: VR-Together architecture (Pilot 2)

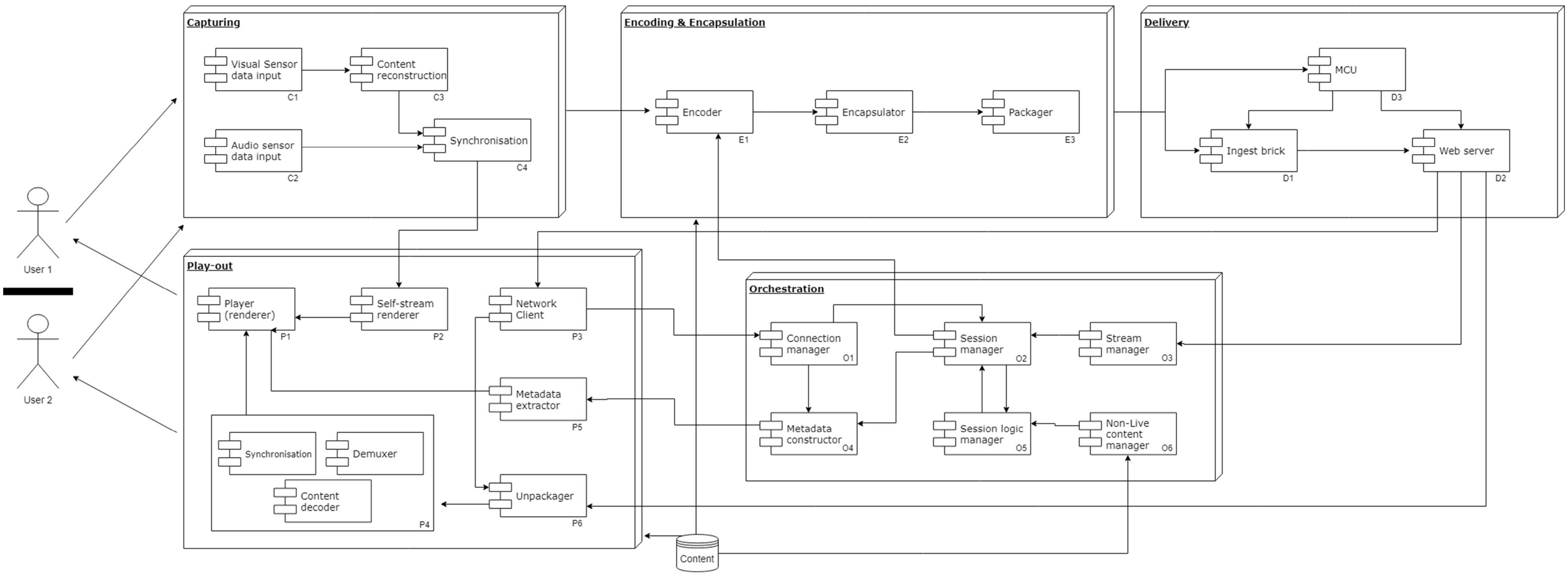

Figure 37: Previous VR-Together Architecture (Pilot 1)

Other terms that are used in the text are defined hereafter:

Term	Description
Frame	An instance of captured content at one specific instant in time
Stream	A collection of consecutive frames
Platform	The end-to-end platform implemented within the VR-Together project
Platform configuration	The platform configuration options defining the operation mode of the components. E.g. video encoder, audio encoder, etc.
Active session	A session in which the platform is used to serve content to 2 or more users immersed in the Virtual Experience. When 2 participants use the platform to interact they are participating in a an interaction session.
Virtual Experience	The virtual world created by the platform and populated with content, for the participants to immerse and interact in
Sensor clock/hardware relative clock	A sensor used in the capturing set-up has an internal clock. Each frame captured by the sensor is timestamped according to this internal clock, the sensor clock or hardware relative clock.
Platform clock	A universal clock used throughout the platform in order to help components synchronise the content. E.g. NTP

5.1.1. Capturing component

The capturing component represents the first block of the VR-Together pipeline. It´s indeed in charge to create the 3D representation of each end-user. While the idea behind the capture technology has not changes from Pilot 1, in Pilot 2 the architecture is designed to accept more than 2 users, so, in Figure 36 you can see that the architecture is designed for an undetermined amount of users.

In the remaining part of this subsection, each sub-block of the capturing component is described.

- **C1 - Visual Sensor data input:** this module receives the data captured by a visual sensor (e.g, a Kinect or RealSense camera) used in one participant's setup. The user is placed in a location where a hardware setup captures his/her motion and texture data. The current setup of VR-Together includes 4 capturing sensors that can be Microsoft Kinect or RealSense Cameras. In any case the input consists of 4 RGBD data streams together with the corresponding texture data. The streams are already time-stamped according to the internal clock of the sensor through which the data was captured.

Input: a user's motion data + texture
Output: raw RGB-D data + visual sensor timestamp

- **C2 - Audio sensor data input:** this module receives the audio sensor signal captured by the microphone used in one participant's setup. The participant's audio is captured in the configured bitrate, channel layout (the direction of the sound is inferred from the HMD direction) and time-stamped according to the audio sensor's internal clock.

Input: user's audio data
Output: user's audio frame + audio sensor relative timestamp

- **C3 - Content reconstruction**: this module receives the data captured from all visual sensors in a participant's setup and merges them into one single visual frame. The content captured from the visual sensors is processed and merged, performing tasks such as background removal, HMD removal or any other additional content reconstruction task that is needed, following the desired experience outcome. Furthermore, this module performs the synchronisation of the separate visual content streams. Each stream is synchronised according to its "creator sensor's" internal clock and all separate streams should be synchronised with each other when merging. The resulting visual data frame represents the RGBD and texture data of one single temporal instance. The resulting visual content stream follows a clock which is relative to the sensors' internal clocks. Note that this clock might drift from the Platform Clock.

Input: raw RGB-D data + sensor timestamp from all sensors in a participant's setup
Output: a visual frame (i.e. fused data, such as TVM or PC, created from data captured by all sensors of one user set-up) + visual frame hardware relative timestamp

- **C4 - Synchronisation**: this module receives the streams to be multiplexed, typically audio and visual components (video, point clouds, TVMs; in one of several layers or tiles), and regulate their speed to send them in a perceived synchronous way.
  Typically the component uses the capture clock times as media timestamps to compare the different media that may come from different sources. The process performed in this module additionally facilitates the synchronisation which will be required later on in the "Play-out" component (between the content streams of different participants) by minimizing the required buffering at the client side.
  As for Pilot 2 this component functionality is disabled because streams are not multiplexed when being sent (which means that synchronization decisions rely on the "Play-out" component - this is what modern international standards such as MPEG-TS, RTP, or MPEG-DASH mandates).

Input: raw Visual stream and raw Audio stream
Output: timestamped visual stream (i.e. TVM or PC raw stream) and time-stamped Audio stream (synchronised but not multiplexed)

5.1.2. Live-recording component

The live-recording component appears in the Pilot 2 definition to allow the live representation of a presenter in the VR-Together experience (live news storyboard). This component takes the same structure as the capturing one except that it uses industrial audio-visual production techniques as the rendering of the live-presenter is based on an advanced panoramic billboard rendering.

- **L1 - Live Camera data input:** this module records the frames captured by a camera (e.g, a Cine stereo camera) installed in the live studio setup. The presenter is placed in front of a green background facing the camera. The incoming video stream is retrieved with a single time-stamp that depends on the clock of the recording device.

Input: presenter frame data
Output: RGB stereoscopic video raw stream data + record device timestamp

- **L2 - Audio sensor data input:** this module receives the audio sensor signal captured by a microphone used in the live studio setup. The presenter's audio is captured in the configured bitrate, channel layout (the direction of the sound is inferred from the HMD direction) and time-stamped according to the audio sensor's internal clock.

Input: presenter's audio data

Output: presenter's audio frame + audio sensor relative timestamp

- **L3 – Live streamer**: This module receives the raw stereoscopic video stream recorded from the panoramic camera in the live studio setup and operates a live stitching[5] operation on the visual frames. Then the module creates a unified RTMP stream constituted of stitched frames and audio frames by processing a synchronisation step based on the recording timestamp. It also provides a set of parameters to configure the output stream according the Live manager requests.

Input: raw panoramic stream + raw audio stream + record timestamp

Output: RTMP stream (including synchronised visual and audio data)

- **L4 - Live manager**: This module handles requests exchanges with the connection manager of the Orchestration component. It is able to manage the live recording component states and provides relevant parameters to the Live streamer regarding stream specifications module according to the running Session specifications.

Input: JSON recording configuration

Output: void

5.1.3. Encoding and Encapsulation component

This subsection describes how the 3D media stream is prepared to be transmitted. In Pilot 1 only one kind of content was involved: animated 3D Time Varying Meshes (TVM), so this component was in charge of processing exclusively a kind of video stream. In Pilot 2 the platform will be able to support both TVM and Point Clouds (PC).

- **E1 - Encoder**: this module receives a visual (or audio) track related to one user (result of the capturing) and encodes it in order to reduce the bitrate needed to represent the visual (audio) signal. The encoding configuration (including for example the target encoding bitrate, the frame rate, etc.) is dictated by the "Session manager" module ("Orchestration" component) that sets the platform configuration for the active session. The visual and audio streams are handled separately, each one by its corresponding encoder module. The result of this process is an encoded visual (or audio) stream.

Input: timestamped visual stream (i.e. TVM or PC raw stream) / timestamped audio stream. Timestamps are set according to the Platform Clock.

Output: encoded visual stream (i.e. TVM or PC encoded stream) / encoded audio stream (Example: .ply file for an encoded PC and .aac file for encoded audio)

- **E2 - Encapsulator:** this module receives an encoded visual stream and an encoded audio stream, which are temporally synchronized (i.e., have timestamps that refer to the platform clock and are aligned), and multiplex them in a single stream. After being encoded the visual stream and audio streams are multiplexed and encapsulated to a media format (e.g., MP4, WebM, other), defined by the "Session manager" module, corresponding to the end-user's playout device, capabilities etc. The input in this process is the encoded and separated visual and audio stream.

Input: encoded visual stream (i.e. TVM or PC encoded stream) + encoded audio stream (Example: .ply file for an encoded PC and .aac file for encoded audio)

Output: audio-visual file (e.g. mp4, webM, etc.) including synchronized audio and visual tracks for one user

[5] https://en.wikipedia.org/wiki/Image_stitching

- **E3 - Packager:** this module receives single audio-visual (i.e., the encapsulated audio-visual content) or multiple audio-visual data corresponding to different users and packages it (them), so that the content can be transmitted by the "Delivery" component. The process can vary depending on the chosen content delivery configuration (e.g., DASH versus WebRTC).

Input: audio-visual file (s) (e.g. mp4, webM, etc.) (i.e., the encapsulated audio-visual content)
Output: packaged content (e.g. MPEG transport stream, SRT, MPEG-DASH, Microsoft Fragmented MP4 Ingest)

5.1.4. Delivery component

The delivery component is the first block of the pipeline on a server/cloud. It is in charge to receive, process and transmit the 3D media stream received by the user. In this component, at the moment of the composition of this document, there are no remarkable changes between Pilot 1 and 2.

- **D1 - Ingest brick:** This module receives the blended/mixed audio-visual stream including the data of multiple users involved in the communication output of the "Encoding and encapsulating" component and adapt them to the format needed for storage in the "Web-server" module that transmits the data on the network.

Input: packaged audio-visual content
Output: audio-visual content ready for transmission (for example, DASH adaptation set)

- **D2 - Web server**: This module makes the content available for consumption and manages the endpoints at which the content is served.

Input: audio-visual content ready for transmission & signalling information (for example, DASH adaptation set, i.e. chunks and .mpd file)
Output: transport protocol messages & packets (for example, HTTP messages and .m4s DASH chunks)

- **D3 - Multi Control Unit (MCU):** Depending on the active session there might be 2 or more participants in one virtual environment. In the case where the participants are more than 2 the "Multi Control Unit" (MCU) module is activated. The MCU is responsible for combining the visual and audio inputs arriving from multiple sources; the MCU allows to reduce the network pressure on the capture system (since only one upload happens instead of number-of-users-minus-one when it is absent) at the cost of an intermediate server. The inputs are blended or re-organized into a common synchronised stream. The result of the MCU is then optionally encoded, encapsulated and packaged before reaching the "Delivery" component.

Input: multiple audio-visual content streams
Output: blended/mixed audio-visual stream including the data of multiple users involved in the communication (i.e., multiple TVMs or PCs)

5.1.5. Orchestrator component

In architecture view, which is shown in Figure 36: VR-Together architecture (Pilot 2), one of the central components is the Orchestrator component, which provides all clients with the information necessary to initiate a communication session of end-users over the VR-Together platform. This includes the management of sessions, instantiation of scenarios, discovery of available rooms, gathering of pointers to content sources and other clients in a session as well as the capture sources.

The Orchestrator is responsible for signalling synchronization data between the different streams consumed by the clients and the MCU. Besides the synchronisation data, other session control data is signalled via the Orchestrator, like content changes, pause/play, scenarios updates, etc.

Contrary to the Pilot 1 definition, the Pilot 2 Orchestration reflexions introduced five conceptual models:

- User: A user who wants to share an immersive social experience with others persons. Also as depicted in Section 5.1, user entity is composed of the three components that orchestrator needs to interact with: capturing, encoding and playout.
- Admin: An administrator can have a global control of the platform and force session control data through a corresponding interface, e.g. in order to facilitate user experiments and demos
- Scenario: This describes the virtual world composed with at least one room. The scenario includes the description of the whole 3D scene and the underlying logic (timeline, interactive event, etc.).
- Session: A session gathers users that want to share an immersive social experience together based on a scenario.
- Room: A room is a virtual space where users are located together which is part of the scenario.

More precisely, we can distinguish a scenario from these two ways:

- A scenario model. It defines a static scenario description; e.g. can't be changed by the VR-Together experience: Room descriptions, Room capabilities and constraints, 3D scene description, and pointers (URIs) to the content sources. This permanent dataset is stored on a dedicated database.
- A scenario instance. It is instantiated from a scenario model when a session is created and will carry on the logics and internal states of the scenario. It temporally stores and maintains dynamic and stateful information related to a VR Together session. The clients in a session need to have a shared state (or view) of the virtual world. This includes: the current time in the world and assets (e.g. videos) in the world, the state of the world, as well as URI's to end-user content streams.

These entities are dependant within each other; we can consider these dependencies with the following UML diagram.

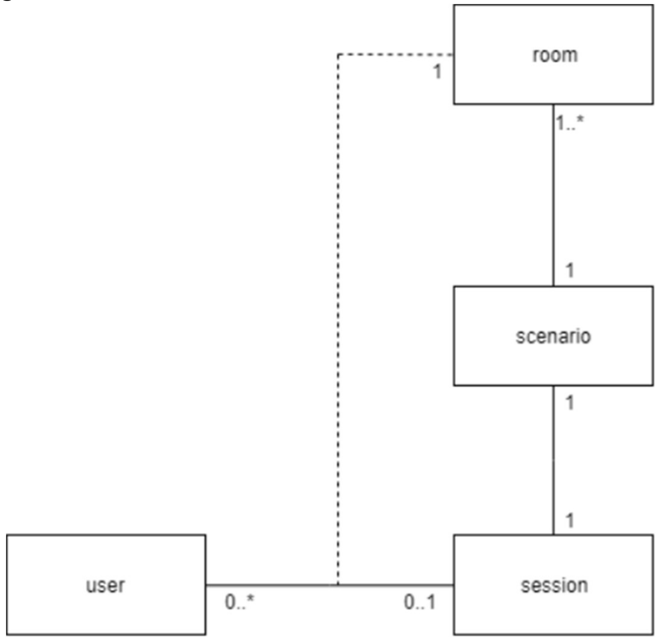

Figure 38: UML representation of the orchestrator entities dependencies

It is important to note that the Orchestrator component controls but does not process media streams. For all control data, all clients (regardless of the type) have a common interface to the orchestrator. However, clients might have different interfaces to content (based on the content type and content server).

In this way a client may retrieve one or multiple media content streams from one or multiple Content Servers. The URLs to the streaming content are provided in the session's metadata as JSON data being sent as parameters through web-requests exchanges. In addition to media content, each client receives streams from other clients that are aggregated by the MCU for audio/visual communication, and transmits streams for other users as well.

Thanks to the user-manager depicted in the following Section 5.1.6 Play-out component description, each client is responsible for its own capture and encoding component integrity both by mater of software integration design and hardware dependency.

As far as the modules functionality that is included in the Orchestrator component we can see the:

- **O1 - Connection Manager**: It is the entry point of the orchestration component. It ensures connections within each node entity of the platform. The users' connections to the VR-Together platform are managed and maintained by this module which then transfers the relevant information to the "metadata constructor" and the "session manager".

Input: user connection information
Output: user connection pointers

- **O2 - Session manager**: This module handles and updates all running sessions in parallel according to users' requests. It operates all sessions commands (session creation,

session connection, session update, etc.) and is aware of all the information regarding a session (e.g., how many users are connected to the session, on which room they are located, rooms attendance, which is the non-live content used, etc.). It finally brings corresponding configuration from the "session logic manager" and "stream manager".
Input: Session information
Output: Session information pointers

- **O3 - Stream manager:** This module is responsible for acquiring the stream information from the Delivery component and transferring it to the "metadata constructor" (through the Session Manager O2) in order to facilitate the stream selection process for the "Play-out" component.

Input: Stream information
Output: Stream information pointers

- **O4 - Metadata constructor**: This module of the Orchestrator constructs the metadata content description file, corresponding to a JSON content, which is used by the "Play-out" component of the VR-Together platform. The responsibility of the constructor is to build the metadata that is necessary for the player in order to project the video stream of the users, pointing to each one's video stream endpoint (originating from the Delivery component), follow the rule set as this is described in the game logic function while using the configuration provided by the "Session manager".

Input: Virtual experience configuration
Output: JSON description files

- **O5 - Session logic manager**: Within the virtual environment a set of rules are defined that form the desired "session logic" to be applied (height of video stream representation, starting position in the room, etc.).

Input: Platform configuration
Output: Configuration pointers

- **O6 - Non-live content manager**: The non-live content, also considered as on-demand-content, that is included in the VR-Together platform (room graphics, stereoscopic videos for billboards, 360 videos, etc.) is managed (host + delivery) by this module. This module also provides the necessary information (URLs) to the metadata constructor related to the non-live content accessibility.

Input: Session configuration
Output: Resource pointers + metadata configuration

5.1.6. Play-out component

The playout component is in charge of delivering, to the end-user, the 3D representation of the virtual world where the experience designed for the platform is set. In Pilot 1 this component had to deal with the virtual scene including only one user because the experience was plan for only 2 users. In Pilot 2 the experience has to be performed for 4 or more users, therefore, an additional sub component has been implemented: the User Manager, described in detail below at the end of this subsection.

- **P1 - Player renderer:** The player is responsible for rendering the content following the metadata pointers in order and project the desired media content.

Input: Metadata descriptor value
Output: Content play-out

- **P2 - Self-stream renderer:** The user's self-created stream (generated from his "own" instance of the capturing component) is "consumed" in this function and passed on to the player for rendering.

Input: void
Output: void

- **P3 - Synchronization**: This module refers to the time-stamp alignment that needs to happen in the received content (decoded stream and self-stream) in order to correctly render the visual and audio streams according to the universal clock's timestamp.

Input: void
Output: void

- **P4 - Content Decoder:** The content arriving from the Delivery component is decoded and transferred to the player for projection. The decoders are different for visual and audio content but are just mentioned here as the same logical module.

Input: encoded and unpacked audio OR visual content
Output: decoded audio or visual content

- **P5 - Demuxer:** This module splits the audio and video of the content received from the web server.

Input: Unpackaged audio-visual content
Output: Unpackaged encoded audio content + unpackaged encoded visual content

- **P6 - Unpackager**: this module un-packages the audio-visual content that is received from the web server through the network client.

Input: void
Output: void

- **P7 - Network client**: This module is responsible for establishing the connection with the web server that provides the content to the "Play-out" component.

Input: void
Output: void

- **P8 - Metadata extractor:** This module parses the metadata file provided by the "Orchestrator" and extracts all the necessary information in order to facilitate rendering of the content.

Input: Metadata descriptor file
Output: Content and configuration values

- **P9 - User manager**: This module has the role of the user components manager and is able to establish local exchanges between Capturing, Encoding and Play-out components. It is also in charge of managing the user states on the Platform by catching and managing orchestrator messages and requests to the Player.

Input: Orchestrator JSON messages
Output: void

5.2. Hardware architecture

This section provides the hardware architecture of the VR-Together platform, as designed for Pilot 1 and as it has been updated for Pilot 2.

In the diagram shown in Figure 39Figure 40: Hardware architecture, we can see the hardware components that form the VR-Together platform when Pilot 1 was released. The HW considered was: two capture rigs; combining 4 RGBD cameras (and a single one in case of the web pipeline) each and a server for capture integration and encoding. Two dedicated servers will take care of content delivery and content orchestration, respectively. Finally, two playout devices will allow end-user content consumption.

Figure 40 gives an updated version of the initial diagram compliant with the Pilot 2 scope. Major change relates to the 5 connected user nodes that constitute the whole platform. We can see that each user node is shown as a user entity depicted according to 3 user node types: web-client node, TVM node, PC node. Additionally, a new node appears to grant the live news feature of the Platform with the Live-recording node. It consists in a dedicated capture setup based on a green background recording of a presenter in live.

For each user node, the HW is constituted of a capture rig (one or four RGBD cameras depending on the node type), an encoding component that processes captured stream data and send it to the delivery server, and the play-out setup that retrieves incoming stream data as well as on-demand content, provided by the content database, to finally project all of it to the user. Then the centralized orchestrator server operates web-requests within all the user nodes and delivery server to manage the whole platform.

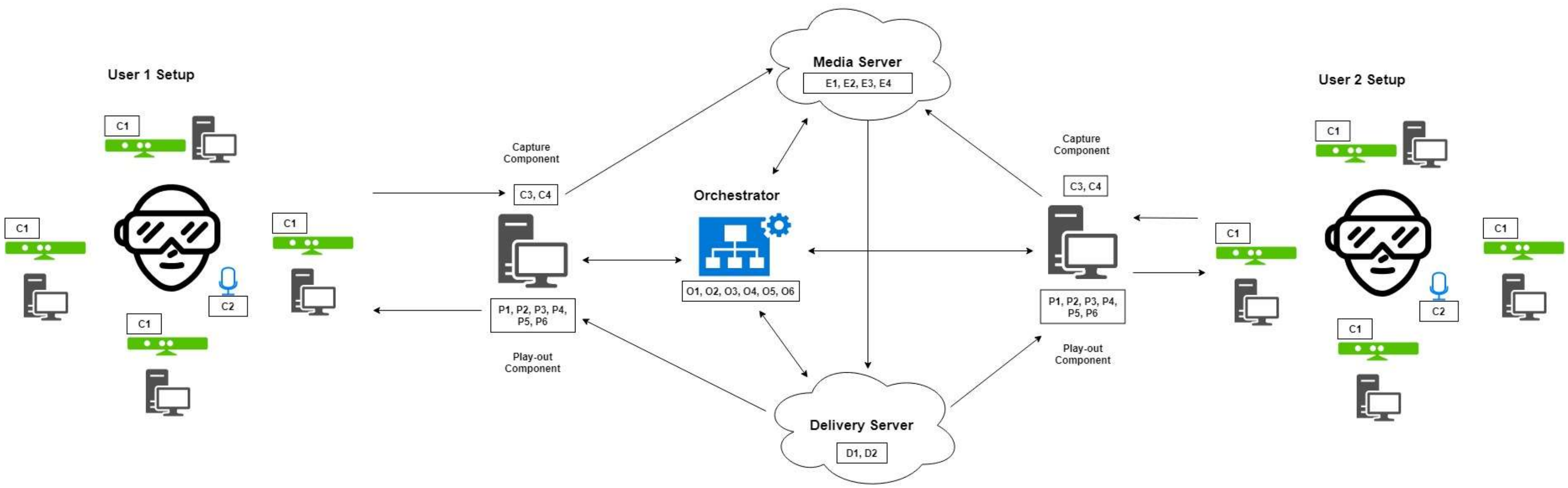

Figure 39: HW Architecture (Pilot 1)

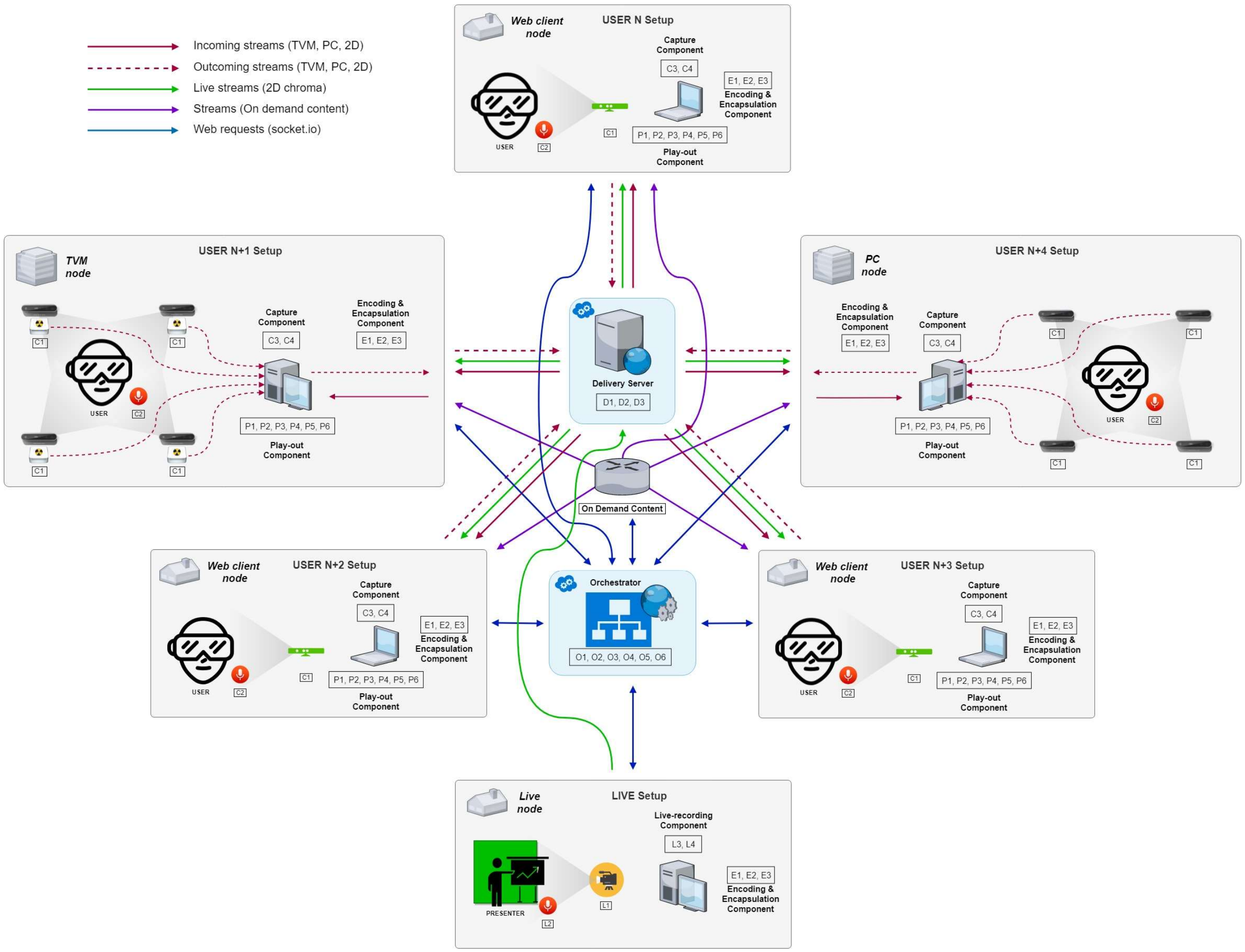

Figure 40: Hardware architecture (Pilot 2)

5.2.1. End-user Set-up

Each end-user needs to be using a capturing setup, an encoding setup and a playout setup in order to be able to use the VR-Together software platform.

First, the capturing setup is composed of:

- Visual sensor capturing: the set-up includes either 1 visual-sensor (for a web-client node) or 4 visual sensors (for a native-client node) responsible for capturing the user's representation that will be encoded and projected in the virtual rooms of the VR-Together platform. Although for the Pilot 1 the retained sensors where the Microsoft Kinect for Xbox One, due to the current discontinuation of the product, the one added for Pilot 2 is the Intel RealSense D415. Nevertheless, the Kinect support remains maintained, particularly for the TVM and web-client pipeline to ensure compatibility with coming Kinect for Azure Microsoft project.
- Audio sensor capturing: for the capturing of the audio of the user the sensor to be used is the microphone that is embedded on the head mounted display that is used by the end-user.
- Local visual sensor processing: For the TVM pipeline, each one of the visual sensors needs to be connected to a separate terminal that processes the raw input stream. In the current set-up of VR-Together the terminals used for this purpose are Intel NUC D54250.
- Local audio-visual stream processing: in a capturing set-up environment a central node/terminal is responsible for collecting the separate visual streams together with the audio stream and processing them accordingly. The hardware device used for this purpose is not specific but the most important minimum hardware requirements that it needs to comply with are:
    - 16GB of Ram
    - I7 or equivalent Central Processing Unit
    - Separate graphics processor with 8GB of memory

Then the encoding setup is the same one as used for the capturing, with the outcoming media stream of the capturing directly injected into the encoding component on the same machine.

At last, the playout setup is composed of:

- Rendering processing: the playout hardware machine is in charge of decoding the incoming streams and rendering the whole scene that aggregates every media (volumetric streams, 3D scene assets, videos …). For all the pipelines the rendering station is the same one as used for capturing except for the TVM pipeline which requires an additional station.
- Head mounted display: the playout devices to be used by the end-users are commercially available products. In our set-up we are using the Oculus Rift Consumer Version 1.

5.2.2. Processing Servers Set-up

All the software components and modules that are not part of the capturing set-up will be hosted in cloud or dedicated servers that will be responsible for performing the required functionalities.

As mentioned in the Software Architecture description, the VR-Together platform will support 3 configurations for capture of the end-users that will be later projected in the in the virtual environment of an active session. The table below describes the requirements of the cloud server nodes supporting the operation of each configuration.

Configuration	Type of server	HW requirements	SW requirements	Storage capacity	Operating System
PC	**Streaming Server / Orchestration Server**	VM instance: CPU: i7 4 cores @3.80 GHz RAM 16 GB	NodeJS Evanescent	128 GB	Windows
TVM	**Streaming Server**	VM instance: CPU: i7 4 cores @3.80 GHz RAM 16 GB	RabbitMQ Server 3.6.15	30 GB	Windows
TVM	**Storing Server**	VM instance: CPU: i5 4 cores @ 2.8+ GHz RAM 8 GB	-	2 TB	Windows
2D	**Other Server**	VM instance: CPU: i7 4 cores @ 3.80 GHz RAM 16 GB	Linux package manager, OS support for Docker	128 GB	Linux
ALL	**Live Streaming Server**	VM instance: CPU: 8 physical cores RAM 8 GB	Linux package manager, OS support for Docker	256 GB SSD	Linux

Table 13: Cloud Server Requirements

The values mentioned above are subject to changes depending on the requirements gathered during the VR-Together platform evaluations and experiments. That is because various aspects of the experience, such as latency or projection quality, might require higher processing capabilities and therefore render the above information out-dated. The currently shown values have been tested against the components as they are delivered in D3.2.

5.3. VR-Together as a Software product

The system to designed for Pilot 1 enables two end-users, located in remote/distributed physical rooms, having available the end-user set-up as it is described in 4.2.1 and access to a high speed internet connection, to access a virtual space where a short scene of VR content can be viewed. Inside the virtual environment, end users are able to see each other’s representation as well as their own representation. They are able to naturally communicate having an acoustic and visual interaction. The content projected in the virtual environment can be generated from a blend of media formats that includes 360-degree Video, Point Clouds and Time Varying Meshes. The end user representations will also be available in these media formats. The audio within the virtual world is immersive by being coherently positioned according to the user position inside the virtual environment and the direction to which they are looking at. The visual and audio content captured and projected in the virtual world will be transmitted in a best effort approach. The play-out of the non-live content in the participants’ location will also follow a best effort synchronisation approach.

6. User Lab

This section describes the VR-Together user lab activities planned in the VR-Together project in order to feed the requirement definition process and to provide a set of objective and subjective results able to give an outcome about the accomplishments reached.

The Requirements Section of this document (Section 4) has introduced the methodology and the process taken into account for the requirements definition and updates. The same section shows how an iterative peer reviewing process, involving the experts from each partner, have contributed to the requirements definition. The outcome of the iterations, involving the requirements definition, deprecation and re-definition, at the moment of the release of this document, have currently been used to define the linking between requirements and experiments. The links can be found in the tab "Experiments/Requirements Links" of the document *VR-Together Requirements Matrix*[6] and in the Section 6.4.2.

This section describes, in detail, the experiments definition process, together with the physical infrastructures used in the experiments, the methodology followed and the planning of the whole process. The section starts with an introduction showing the general guidelines followed for the experiments process (Section 6.1). The following section (Section 6.2) describes the user labs involved in the experiments and, in the last one (Section 6.4), the experiments definition and planning is explained.

6.1. Introduction

The VR-Together experiments process follows a user-centric approach, in which selected user groups are required to provide feedback, in order to evaluate the platform and obtain new relevant requirements leading to further design and implementations. The user groups considered are:

- **Stakeholders**: To identify adequate business models and opportunities. The stakeholders are met in public events (fairs, conferences, congresses) and specific stakeholder workshops.
- **Experts**: To gather requirements about the accomplishments in terms of novelty and about the performance reached. They are consulted internally among the consortium companies and externally at targeted events (fairs, conferences, congresses).
- **End-Users**: To gather a set of requirements able to validate the VR-Together functionalities, the user perception and interaction, and the quality of the representation. The end-users are being consulted during the trials, through user lab experiments, via questionnaires and at open demos.

Both stakeholders and experts will be invited to join the professional Advisory Board, as explained next. The advisory board will include two types of professionals, fulfilling the technical and artistic needs of the project, in the fields of virtual reality and immersive media.

The VR-Together consortium has also built a permanent collaborative distributed user lab with the equipment needed to run demos but also to test the new releases, analysing development and integration. The infrastructure was built in 2018, for Pilot 1, and has been used to run the experiments and evaluations depicted in Section 11. The same infrastructure, according to the

[6] VR-Together Requirements Matrix https://docs.google.com/spreadsheets/d/12IF74tYmMmCin5_hAgm0PL4ajia6Ii-WQ3fS-NWS5q0/edit?ts=5c87c8f2#gid=196101635

plan, will be used as well to run the experiments defined for the second year of the VR-Together project, listed in Section 6.4.

6.2. User Lab Nodes

Through the VR-Together User Lab, the project will run tests and evaluations that will be used to take decisions about the pilots (by the point of view of the production) and about the technical platform. In the project, we implement a distributed User Lab that includes three lab nodes with fully equipped facilities that can support the VR-Together platform. There are, as well, additionally built lab nodes with partially equipped facilities as support for the VR-Together platform.

The user labs provide a complete environment to run the pilot trials. They are equipped with a full media pipeline including capturing and reconstruction, delivery and transmission, and, finally, rendering. They are being used for three main purposes: quantitative evaluation (e.g., performance) of the system, end-users' evaluations (pre-trial and trial), and experts/stakeholders demonstration. Figure 41 shows the basic infrastructure of a hub, including a capture system, several PCs for reconstruction, and a rendering infrastructure based on Head Mounted Displays (e.g., Oculus Rift and HTC Vive). VR-Together has three main lab nodes strategically located in:

- Amsterdam, Netherlands (CWI premises)
- Barcelona, Spain (i2CAT premises)
- Thessaloniki, Greece (CERTH premises)
- Rennes, France (VO premises)

The labs are interconnected and are used to perform demos, experiments and also to run tests in terms of performance as quality, transmission and latency/delays. In particular, the Thessaloniki (CERTH) and the Barcelona (i2Cat) labs currently include two fully equipped nodes for each lab (four in total) and they are in charge of running the VR-Together experience full pipeline using the native pipeline rendered with the Unity application. The Rennes lab (VO) will be the core of the web player based experiments and demos.

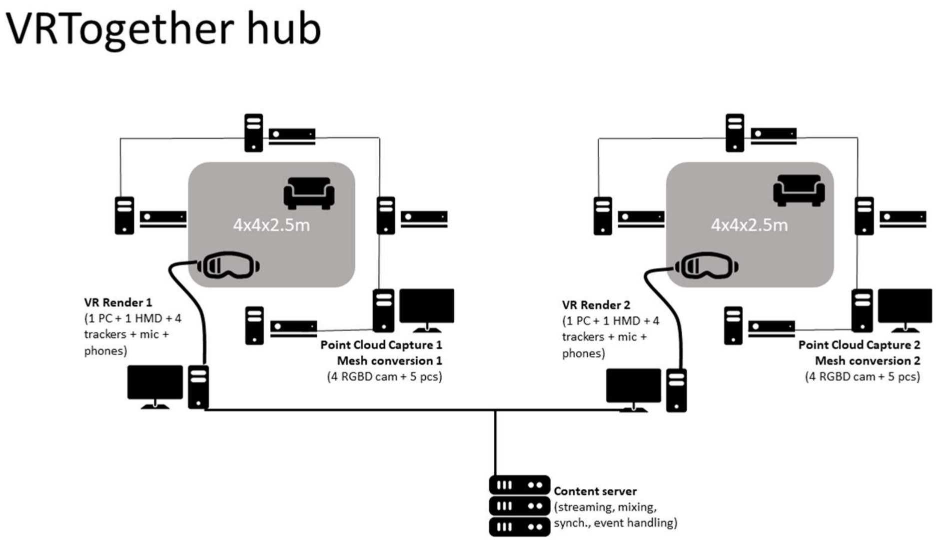

Figure 41: Schematic View of a VR-Together hub

In addition to the hubs, several partners of the project created dedicated user labs with a partial infrastructure of the full VR-Together platform. These labs are being used for targeted experiments that will inform about different aspects of the project: QoE, improved reconstruction, comparison of different media types, and production of media assets. The following partners have provided a lab, intended for different types of experimentation.

- **Artanim**'s user lab primarily focuses on evaluations on the psychological aspects of the project such as "togetherness", "co-presence", and "flow". The initially planned experiments assess the benefit of including different levels of movement fidelity to the tracking of face, hands, full-body and IK extrapolated joints. The goal is to confront these benefits with the costs (monetary and effort) of adoption of these technologies by an end user and define a standard for animation algorithms and hardware that can be adopted in the reminder of the VR-Together project for the alternative of representing user and actors with a rigged mesh of triangles.
- **CERTH**'s user lab primarily focuses on technological evaluations about the visual quality of the real-time 3D reconstruction of people's figures, aiming at both the visual quality and the transmission rate. The experiments performed in this lab also aim to analyse techniques for the removal of the HMD in the 3D reconstructions.
- **CWI**'s user lab primarily focuses on Quality of Experience (QoE), which serves for developing new quality metrics and guidelines for evaluating social VR. Such metrics are being used in the system for optimization purposes and during the trial. CWI already run some initial experiments about the QoE of point cloud compression at the beginning of the project, which has resulted in a new quality metric based on colour information. In addition, the lab expects to run a number of quantitative experiments related system performance at the compression and networking levels.
- **Entropy** 's user lab primarily focuses on production of media assets, in different formats, for the trials. The goal is to better understand the production workflow and cost for creating new social VR experiences, thus gathering requirements regarding content for the trials.
- **I2CAT**'s user lab conducts experiments on both the psychological aspects of the project and on the QoE of the users. For example, whether and in what conditions end-users

feel like being together within the virtual environment or not. Such experiments make use of both questionnaire and behavioural data.

- **TNO**'s user lab primarily focuses on experiments related to the technical functionality. The aim is to run experiments that help the project to improve the quality of experience of the shared space using 360 monoscopic background video in the shared VR platform, to run comparative experiments for better representing users in the shared VR environment by reducing Chroma-keying artefacts, and experiment with methods to improve the feeling of co-presence through shared interaction.

In the Annex II a description for each one of the main Lab nodes can be found.

6.3. Experiments

The partners of VR-Together carry out experiments to inform about the different aspects of the project: technology, pilots, and evaluations. These experiments run either in the hubs (full-fledged infrastructure) or in the lab nodes (partial and targeted infrastructure). In the project we foresee three main categories of experiments, with distinct objectives:

- Assessment of **technology**, such as HMD removal or content distribution. They have a direct influence on the user hubs under development;
- Subjective **quality of experience**[7], mainly based on perception of the medium under different constraints (different compression mechanisms or bandwidth).
- **Psychological** dimension of the VR-Together experiences, evaluating aspects such as the feeling of being there, as well as the feeling of being together.

In all VR-Together experiments we follow informed consent procedures, protect the privacy of personal data and, to the extent that it is possible, make research data publicly accessible to facilitate further experimentation. More information about ethical considerations of the intended experiments, as well as the outline of the different datasets, and the considerations regarding end-user privacy can be found in D1.2.

6.4. List of Experiments

The list of experiments planned for Pilot 1 have been published in the first version of this document (D2.1) and, in order to make such information available also to the readers of this document they have been included at the end of it in the Annex IV (11).

The experiments planned for Pilot 2 will use the same labs and infrastructures designed and organized for Pilot 1 and are divided into 2 sections:

- Requirements gathering and professionals' feedback
- Experiments with end-users

In addition, a detailed description of each experiments, of the process followed to reproduce and validate each of them and of corresponding requirements will be provided also in the deliverable D4.4

[7] https://hal.archives-ouvertes.fr/hal-00977812/document

6.4.1. Requirements gathering and feedback from professionals

i2CAT2.1: Focus groups with VR industry professionals

This consists of a series of focus groups with VR industry professionals to be conducted by all the partners, under the coordination of i2CAT. In particular, the partners will be asked to screen, contact and meet local VR industry professionals, who will be presented the VR-Together platform (in case that a VR-Together demo is available) or a video explicitly created to describe the pilot 1 experience (in case that a VR-Together demo is unavailable), https://www.youtube.com/watch?v=Rel5qnj8rxA. Evaluation materials, such as consent forms, the experience questionnaire used in pilot 1 (in case of having the visit of two representatives of the specific company) and an ad-hoc questionnaire including closed and open questions, will be prepared. This questionnaire is also targeted at driving an interview with the professionals, boosting interesting discussions and interaction.

The objective of this pilot action is to gather feedback from the VR professionals about the topic addressed in the project, its potential impact, the technical contributions of the project (compared to state-of-the-art solutions), the existing limitations and challenges, and getting recommendations about the strategy to follow. Interestingly, the feedback gathered by the professionals / experts will be very valuable to validate / refine the existing requirements, and/or to gather new ones. It is expected to meet with at least 20 companies between all the partners between during the March – June 2019 period.

Likewise, the VR professionals will be asked to become Advisory Board members. It was decided that this strategy would result in a higher success and more fruitful interactions than having an international Advisory Board panel, as meeting with local companies is always less restrictive and involves lower efforts. If the professionals volunteer, they will be asked to fill in a consent form, and to commit to meet with the partners for each one of the subsequent pilots in order to keep track of the evolution of the project and to keep contributing to the validation/identification of requirements. A section on the website will be created to include information about the Advisory Board panel (names, company, short bio and photo), and a quote about their opinion / impression about the project, if they provide their consent for this. It is expected that at least 8 of the professionals will accept the invitation to join the Advisory Board panel.

6.4.1. Experiments with End-Users

Artanim-2.1: Content representation
User's perception and reaction to different forms of content presentation, comparing 360 video, 3D scene + video billboard characters, and 3D scene + 3D characters.

CERTH-2.1: Rendering
Multi-TVM (multiple instance) rendering in Unity - stress test of the system.

CWI-2.1: Human representation

Subjective and objective comparison of mesh-based compression and PC-based compression on same set of dynamic volumetric signals for which both mesh-based and point cloud representations are available.

i2CAT-2.2: Human representation

Dataset of different user capture techniques and social VR activities for TVM/Point Cloud (4 RGB-Depth sensors). RealSense setup will be used.

CERTH-2.2: Human representation

Dataset of different user capture techniques for comparison and validation TVM/Point Cloud (4 RGB-Depth sensors) and a marker based Vicon tracking system. RealSense setup will be used.

CWI-2.2: Delivery and transmission

Experiment in which people are free to move in the virtual space to collect user navigation patterns that can be used to build user navigation prediction models for adaptive streaming strategies

CERTH-2.3: Delivery and transmission

A set of sessions of varying TVM parameterization with respect to the geometry and texture quality will be experimented. Comparing the results given by using the VRTogether Objective Performance Metrics, will allow us to better understand the pros and cons of the distribution options according the different available quality levels of the TVM medium

CWI-2.3: Delivery and transmission

Experiment to test subjective quality achieved by using DASH streaming of point clouds with tiling vs without tiling strategy

i2CAT-2.3: Delivery and transmission

Recreation of a Social TV scenario “watching apart together” (with two users). The goal is to investigate the impact of delays and delay differences, as done for traditional Social TV studies. Delays will be artificially manipulated to the shared TV contents, and possibly to the end-users’ pipeline.

i2CAT-2.4: Delivery and transmission

Recreation of a Social TV scenario “watching apart together” (with two users). The goal is to investigate the impact of the distance between the users. Distance: explore the users’ behaviour, QoE, togetherness, when different distance between the users exist (two users together in a sofa, medium distance in two chairs, longer distance).

VO-2.1: Delivery and transmission

Connected lab of TNO’s system for Pilot 1 content. After successfully installing and configuring the Web pipeline in Rennes, the plan is to deploy it in two different sites: Rennes and Paris on a private network. The objective is to do testing with people on both sites to collect their feedback with the remote setup.

TNO-2.3: Delivery and transmission

To reduce the upload bandwidth and the end-device resources required in current architecture TNO is building a Multipoint Control Unit (MCU). During the development multiple technical experiments will be conducted to steer the team in choosing the right architectural options. The experiment will compare setup with and without CPU in relation to performance (CPU/GPU load, bandwidth and latency). Based on the final design additional tests to validate the good working of the MCU will be defined and performed.

Artanim-2.2: about avatar representation
Evaluate the effect of look-alike virtual representations to the experience of presence, social meaning and quality of interaction in a social VR context

CERTH-2.4: Subjective evaluation of HMD removal and self-representation
Subjective assessment of the new HMD removal method which will be based on generative autoencoders. This method will remove the HMD from the TVM by inpainting both the geometry and the texture. The assessment will be split in two parts. The method will be evaluated on the data acquired during EXP-CERTH-4, and the subjects will assess the quality of the inpainted facial part of the TVM.

TNO-2.1: Subjective evaluation of HMD removal and self-representation
A various number of self-representations have been evaluated with the Kinect v2 and the RealSense D415 using a small user group. Two important aspects are the latency of the self-representation and the quality of the 3D rendering. The latency depends on the amount of 3D points/triangles to render, i.e. the resolution of the image. The quality of the 3D rendering depends on the range for which the sensor is designed and if the sensor is a depth estimator (e.g. stereo) or a depth measure device (e.g. time-of-flight).
These aspects suffer from clear trade-offs in user-setup and sensor choice and are described in the subjective evaluation of February 2019. Towards Pilot 2 we will put this knowledge to use for an optimized self-representation of the user.

TNO-2.2: Subjective evaluation of HMD removal and self-representation
This HMD Removal experiment will be performed during RGB-D user capture on the client site. In an initialisation step; (1) the 3D face (without HMD) is captured, (2) the HMD position and orientation is calibrated with the RGB-D sensor, (3) and the HMD in both the RGB and Depth streams are replaced with the 3D face-capture of the user. The updated RGB-D stream is send to VR for a more personalized user-representation. A subjective evaluation will be performed with users in a VR setup (e.g. the Pilot 1 demo) where they see each other with or without the HMD removed in the 2D or 3D rendering.

CERTH-2.5: about simultaneous delivery + rendering of TVM instances
As discussed in GA 4, continue with CERTH2.1 to also consider the network part.

CWI-2.4: physiological signals for evaluating QoE
Usage of physiological signals for monitoring QoE

CWI-3.1: physiological signals for evaluating QoE (year 3)
Usage of physiological signals for monitoring QoE

CWI-2.4: Experiments for quantifying values of requirements
Impact of end-to-end delay on QoE

CWI-2.5: Experiments for quantifying values of requirements
Impact of framerate on self-representation (point clouds)

CERTH-2.6: Experiments for quantifying values of requirements
Impact of framerate on self-representation (TVMs)

i2CAT-2.6: Experiments for quantifying values of requirements
Impact of synchronization on QoE

6.4.2. Pilot 2 Experiments Calendar

The following table represents a guideline for the experiments planned for Pilot 2.

It includes the experiments Id´s, the title, the responsible partners (with the main one in bold) and the timeline planned at the time of the delivery of this document.

In addition, the table includes also the links with the requirements involved in each experiments. The linking has been the outcome of the latest requirement peer review iteration and they will be validated by the corresponding experiments.

Experiment Id	Area	Involved Partners	When	Linked Non Functional Reqs	Linked Functional Reqs
i2CAT-2.1	Focus groups with VR industry professionals	**i2CAT, TNO, VO, Entropy, CERTH, Artanim**	07/2019	NF.104.1, NF.107.0, NF.109.1, NF.150.0, NF.66.1, NF.96.1, NF.122.0	FR.1.0,FR.2.0,FR.3.0,FR.5.0,FR.15.1, FR.16.0, FR.17.0, FR.27.1, FR.28.0, FR.31.0, FR.37.0
Artanim-2.1	Content representation	**Artanim, i2CAT**	01/2019	NF.96.1, NF.104.1, NF.107.0, NF.109.1, NF.122.0, NF.150.0	FR.3.0, FR.9.1, FR.11.0, FR.12.1, FR.13.0, FR.22.0, FR.27.1, FR.31.0, FR.35.0, FR.36.0, FR.37.0, FR.135.0
CERTH-2.1	Rendering	**CERTH**	01/2019	NF.96.1, NF.122.0 NF.55.0, NF57.0, NF.58.0, NF.59.0, NF.77.0, NF.78.0, NF.123.0	FR5.0. FR12.0, FR15.0, FR.35.0, FR.147.0, FR.119.0, FR129.0
CWI-2.1	Human representation	**CWI**	07/2019	NF.75.0, NF.95.0, NF.72.0, NF.74.1, NF.77.0, NF.78.0, NF.154.0	FR.1.0, FR.3.0, FR.5.0, FR.16.0, FR.20.0
CERTH-2.2	Human representation	**CERTH, Artanim, CWI**	07/2019	NF.66.1, NF.67.0, NF.96.1, NF.122.0	FR.1.0, FR.5.0, FR.12.1, FR.16.0,FR.57.0
CWI-2.2	Delivery and transmission	**CWI, i2CAT, Entropy**	06/2019	NF.75.0, NF.95.0, NF.152.0, NF.153.0, NF.154.0, NF.155.0	FR.3.0, FR.5.0, FR.16.0
i2CAT-2.2	Human representation	**i2CAT, CWI**	09/2019	NF.67.0, NF.96.1, NF.122.0	FR.12.1, FR.57.0
CWI-2.3	Delivery and transmission	**CWI, i2CAT, MS**	09/2019	NF.74.1, NF.78.0, NF.155.0, NF.152.0, NF.153.0, NF.154.0	FR.23.1
CERTH-2.3	Delivery and transmission	**CERTH, i2CAT**	06/2019	NF.75.0, NF.66.0, NF73.0, NF77.0, NF78.0, NF123.0	FR.1.0, FR.5., FR.16.00, FR.60.0
i2CAT-2.3	Delivery and transmission	**i2CAT, CWI**	07/2019	NF.73.1	FR.1.0, FR.3.0, FR.5.0, FR.8.0, FR.12.1, FR.13.0, FR.15.1, FR.16.0, FR.50.0
i2CAT-2.4	Delivery and transmission	**i2CAT**	07/2019	NF.66.1	FR.1.0, FR.3.0, FR.5.0, FR.12.1, FR.13.0, FR.15.1, FR.16.0
VO-2.1	Delivery and transmission	**VO, TNO**	09/2019	NF.66.1, NF.92.1, NF.151.0	FR.1.0, FR.5.0, FR.49.0, FR.3.0, FR.23.1, FR.25.0
TNO-2.3	Delivery and transmission	**TNO**	06/2019	NF.66.1, NF.73.1, NF.123.0, NF.153.0, NF.151.0	FR.8.0, FR.13.0, FR.15.1, FR.49.0, FR.50.0, FR.51.1, FR.54.0 and FR.129.0
Artanim-2.2	Avatar representation	**Artanim**	09/2019		FR.1.0, FR.5.0, FR.16.0

CERTH-2.4	Subjective evaluation of HMD removal and self-representation	**CERTH**	09/2019	NF.67.0, NF.75.0	FR.3.0
TNO-2.1	Subjective evaluation of self-representation	**TNO**	02/2019	NF.66.1, NF.67.0, NF.75.0, NF.96.1, NF.122.0, NF.123.0	FR.1.0, FR.5.0, FR.12.1, FR.50.0, FR.51.1
TNO-2.2:	Subjective evaluation of HMD removal	**TNO**	07/2019	NF.66.1, NF.67.0, NF.73.1, NF.75.0, NF.96.1, NF.110.1, NF.111.1, NF.112.1, NF.122.0, NF.123.0	FR.1.0, FR.5.0, FR.8.0, FR.12.1, FR.13.0, FR.15.1, FR.50.0, FR.51.1
CERTH-2.5	Simultaneous delivery + rendering of TVM instances	**I2CAT, CERTH**	05/2019	NF.55.0, NF57.0, NF.58.0, NF.59.0, NF.77.0, NF.78.0, NF.123.0	FR.13.0, FR5.0, FR12.0, FR15.0, FR.35.0, FR.147.0, FR.119.0, FR129.0
CWI-3.1	Physiological signals for evaluating QoE	**CWI**	10/2019	NF.75.0, NF.95.0, NF.72.0, NF.74.1, NF.77.0, NF.78.0	FR.3.0, FR.5.0, FR.16.0, FR.20.0
CWI-2.4	Experiments for quantifying values of requirements	**CWI**	09/2019	NF.73.1	FR.8.0, FR.13.0, FR.50.0
CWI-2.5	Experiments for quantifying values of requirements	**CWI**	09/2019	NF.96.1	FR.1.0, FR.16.0
i2CAT-2.5	Experiments for quantifying values of requirements	**i2CAT**	09/2019	NF.96.1	FR.1.0, FR.16.0
i2CAT-2.6	Experiments for quantifying values of requirements	**i2CAT**	09/2019	NF.111.1, NF.112.1	FR.1.0, FR.16.0, FR.59.0, FR.145.0

Table 14: Pilot 2 Experiments Calendar and linking to the requirements

7. Conclusions

In this document we have presented an extended description of the software and hardware dimensions of VR-Together, at the point in time prior to the release of Pilot 2. The document presents the Plot, storyboard and production descriptions related to the abovementioned Pilot. More details about the Pilot 2 description can be found in the deliverable D4.3.

After the Pilot description, an extended software platform description is unravelled. It consists of an elaborated analysis of the User scenarios and use cases, a software requirements specification and all related requirements information.

Next, we analysed the Software and Hardware architecture of the VR-Together platform with a view on accurately describing the functionalities that are supported as well as the hardware components that host the functional blocks.

Finally, we have outlined the VR-Together User Labs, and the different experimental work involved in the preparation of the pilots and the validation of the project requirements. The deliverable D4.4 provides a detailed explanation of the experimental process.

This document has therefore provided a global outline of the production and introduced the specific software development and content production efforts needed to deliver it.

Next steps will be focused on implementing these efforts in a concrete calendar, and monitor the appropriate development of the infrastructure, the content production and the validation of the experimental paradigm proposed in VR-Together.

Further versions of this document are under consideration in order to provide more details regarding architecture and user lab actions.

8. Annex I. End User Questionnaire used in VR Days event

2/6/2018 VR Together

VR Together

VRTOGETHER

Our mission is to make VR experiences a social space, where you can share and communicate with your family or friends and to experience VR together.

VR Together is an European research project (funded by the EU). In this project we will create an end-to-end system for the production and delivery of photorealistic and social virtual reality experiences.

With this questionnaire we like to get your feedback about some of the research we like to address in VR Togehter.

Thank you for your time, it will only take a few minutes.

We really appreciate it.

Regarding what you just experienced, how would you rate...

1. ...the video quality?
Mark only one oval.

	1	2	3	4	5	6	7	
Very bad	◯	◯	◯	◯	◯	◯	◯	Very good

2. ...the audio quality?
Mark only one oval.

	1	2	3	4	5	6	7	
Very bad	◯	◯	◯	◯	◯	◯	◯	Very good

3. ...the overall experience?
Mark only one oval.

	1	2	3	4	5	6	7	
Very bad	◯	◯	◯	◯	◯	◯	◯	Very good

https://docs.google.com/forms/d/13G406PFVpGhadeOpyzhsYnwfnsqQolFp5g_jOiFk99Y/edit 1/4

Experiences with VR

4. Have you ever experienced VR before?
Mark only one oval.

- ( ) Yes
- ( ) No

5. Are you interested in Social VR experiences?
Mark only one oval.

	1	2	3	4	5	6	7	
Not at all	( )	( )	( )	( )	( )	( )	( )	Absolutely

Would you like to experience the following topics in Social VR?

6. *Mark only one oval per row.*

	Not at all interested	low interest	Slightly interested	Neutral	Moderately interested	Very interested	Extremely Interested
Sports	( )	( )	( )	( )	( )	( )	( )
Movies	( )	( )	( )	( )	( )	( )	( )
Theatre	( )	( )	( )	( )	( )	( )	( )
Videogames	( )	( )	( )	( )	( )	( )	( )
Education	( )	( )	( )	( )	( )	( )	( )
Music experiences	( )	( )	( )	( )	( )	( )	( )
Live TV shows	( )	( )	( )	( )	( )	( )	( )
Videoconferencing	( )	( )	( )	( )	( )	( )	( )
Dating	( )	( )	( )	( )	( )	( )	( )
Adult entertainment	( )	( )	( )	( )	( )	( )	( )

7. Is there anything else you would like to experience within a VR environment?

____
____
____
____
____

In a VR experience, how important would it be for you to...

8. ...share the experience with someone?
Mark only one oval.

	1	2	3	4	5	6	7	
Not at all	( )	( )	( )	( )	( )	( )	( )	Absolutely

9. ...interact within the experience?
Mark only one oval.

	1	2	3	4	5	6	7	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Absolutely

10. ...enjoy the overall the experience?
Mark only one oval.

	1	2	3	4	5	6	7	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Absolutely

11. ...being able to move within the experience?
Mark only one oval.

	1	2	3	4	5	6	7	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Absolutely

Demographic questions

12. Gender
Mark only one oval.

- ⬭ Female
- ⬭ Male
- ⬭ Other: ____________

13. Age
Mark only one oval.

- ⬭ Less than 18
- ⬭ Between 18 and 30
- ⬭ Between 30 to 45
- ⬭ Between 45 and 65
- ⬭ More than 65
- ⬭ Other: ____________

14. Are you interested in this VR project? If so...
Check all that apply.

- ☐ I would like to receive updates about the project
- ☐ I would like to participate in user studies
- ☐ I would like to give my expert input / feedback
- ☐ Other: ____________

15. Email

____

16. Do you have any other comments or information you like to share with us?

Powered by

9. ANNEX II: VR-TOGETHER LAB NODES

In the following part some of the existing infrastructure for the labs is presented in order to show where different partners will perform targeted evaluations.

9.1. Artanim Lab Node

Figure 42: Artanim's User Lab

Artanim is housed within a facility of over 273 m² with a motion capture studio of the following size: 15 m x 8 m x 3.7 m (see Figure 42 and Figure *43*). The lab is equipped with diverse high and low end motion capture equipment and VR/AR equipment:

- Vicon MXT40S with 24 cameras (up to 515 fps)
- Xsens MVN 17 MTx inertial trackers
- RGB-D cameras
- Variety of head mounted displays (HMD): Oculus CV1, HTC Vive, HoloLens (see-through HMD).
- Set of 6 HTC VIVE trackers

The lab is also equipped with a photogrammetric 3D scanner comprising 96 cameras for polygonal mesh reconstruction of users and objects. For production and VR/AR applications, Artanim uses a full range of software: Vicon Blade, Vicon Tracker, MVN Studio, Autodesk Creation Suite (3ds Max, Maya, MotionBuilder), Adobe Production Premium (After Effects, Premiere, Photoshop), and Unity 3D.

Figure 43: Artanim's User Lab.

9.2. CERTH Lab Node

CERTH has two available rooms (studios) for the user lab, one in Building A of dimensions 4.5m x 4.5m x 2.5m, and one in Building B (see Figure *44*) of dimensions 5m x 5m x 4m. The laboratories are equipped with RGB-D, Motion Capture and VR/AR equipment. In particular:

- Motion capture
- XSens MVN 9 MTx inertial trackers - motion capture suit
- RGB-D cameras for skeleton tracking - 6x Kinect v2, 6 Kinect v1
- Other 3D cameras
- 1x ZED Stereo Camera
- AR/VR HMD
- 1x HTC Vive
- 1x Microsoft HoloLens
- 3x Drones (4K) (to be purchased)

CERTH's software includes MS Visual Studio, Unity 3D, and Photogrammetry Software (to be purchased).

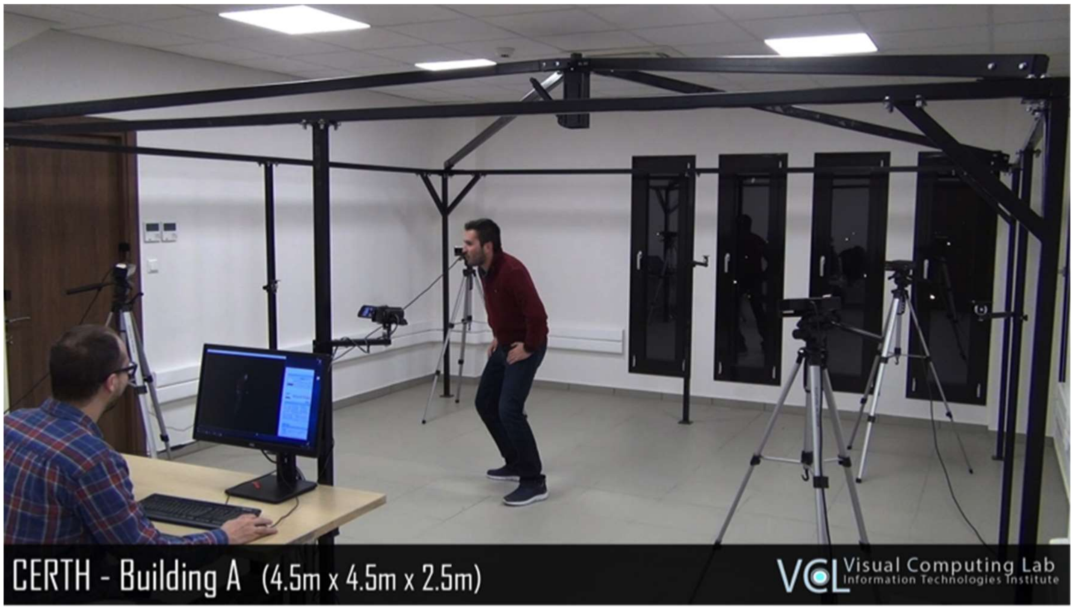

Figure 44: CERTH's User Lab

9.3. CWI Lab node

CWI has two available rooms: Pampus (see Figure 45) and the QoE Lab (see Figure 46). Pampus is a living room like lab, where experiments about user experience can be performed. It includes two sofas, a television, cameras, and a microphone array. The room has as well an interactive table that we don't expect to use during the project. The QoE Lab, under construction, will eventually become a hub for the project. It has been used to run experimentations for MPEG call for proposals in point clouds, and includes accessories, a top quality 55'' TV set (LG OLED 55C7V), and capture and rendering equipment (to be purchased).

Figure 45: CWI's User Lab (Pampus)

Figure 46: CWI's User Lab (QoE Lab)

9.4. TNO Lab node

At TNO premises, we have a media lab of approximately 8mx12m, as well as regular meeting rooms which we can reserve for whole days to run user tests. None of these rooms allow for the setup of a dedicated and (semi-)permanent user lab. The aim is to develop and release a virtual user lab (i.e., a software platform) that can be setup at physical locations for user tests. TNO has equipment for a social VR setup of up to four persons:

- Two VR capable PC systems and three VR capable laptops;
- Four Oculus Rift VR HMDs, including two sets of touch controllers;
- Four Microsoft Kinect RGB+D cameras for user capture;
- Four general-purpose headphones and microphones.

Figure 47: TNO's Media Lab

9.5. i2CAT Lab Node

The Lab node of i2Cat in Barcelona will be composed of:

Space Setup

- A place allowing to run multi-user VR experiences.
- 2 separated rooms that allows two users, each in a minimum clean space of 5x5x2,5m (space for cameras and PCs is extra)
- Furniture (coach, table, chairs, etc.)

Hardware for consumer setup:

- Render:2 VR-ready pcs + 2 HTC vive + 2 headphones + microphones
- Motion trackers
- Communication: LAN connections
- Content streaming: 1 server PC to test services locally
- Mesh capture: 2 Capture rigs based on RGBD cameras (each involves 1 cam for face, 4 Kinect for Xbox one for the body reconstruction, 5 pcs + CERTH software)
- 4 Intel Realsense cameras

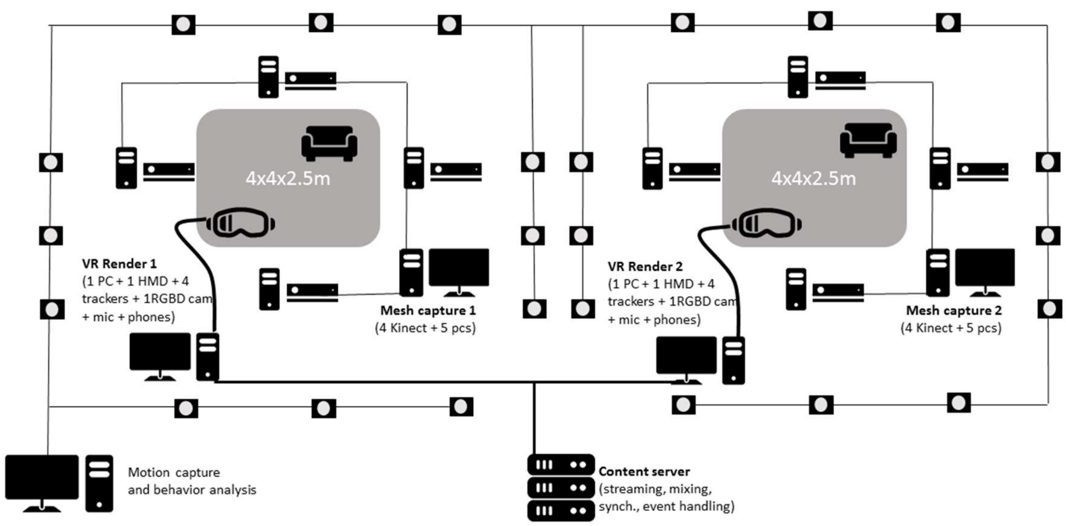

Figure 48: i2CAt lab infrastructure

In conclusion the partners of the projects have adequate facilities for testing and experimentation. The initial six months of the project will be dedicated to one the one hand run some initial experiments in the user labs for gathering requirements and to on the other hand construct the hubs for VR-Together.

10. ANNEX III: PILOT 1 USE CASES AND REQUIREMENTS

This section has been included in this document to show to the reader a comparison between i) the Use Cases and Requirements status when Pilot 1 has been delivered and ii) the current status at the time of the delivery of this document.

In Pilot 1 the end-users participated in a virtual world scenario as described in Section 3 of this document. Below we present a collection of the User Scenarios describing Pilot 1. After that we analyse the user scenarios and define the implicated Use Cases. Next we extract and present the functional and non-functional requirements based on the use cases. Last, we depict the system architecture that will support the functionality realized in Pilot 1.

10.1. User Scenarios

User Scenarios are user level stories describing in simple terms the functionality and what the user wants to do.

For the scenarios that are preceded by an asterisk, it is uncertain whether they will be implemented for Pilot 1 because the development work is focused on creating seamless experiences for each one of the available configurations. However, all of the scenarios will be implemented and presented in Pilot 2. The same rule applies to the connected use case descriptions.

The User Scenarios belonging to Pilot 1 are:

Controls

*As an end-user I want to be able to set my preferences regarding the experience configuration.

*As an end-user I want to be able to create, edit and save my own profile on the VR-Together platform.

*As an end-user I want to be able to create or join an existing virtual room.

As an end-user I want to have the control of the content reproduction (start, pause, exit) when I am the creator of a virtual room.

Experience

As an end-user I want to be able to use one or four Kinect devices in order to capture my body reconstruction and movements.

As an end-user I want to be able to use one or four RealSense devices in order to capture my body reconstruction and movements.

As an end-user I want to be able to use the microphone of my HMD to capture my audio data,

As an end-user I want to be able to view my own representation in the virtual environment of VR-Together.

As an end-user I want to be able to view the other user's representation in the virtual environment of VR-Together.

As an end-user I want to be able to listen the other user's voice in the virtual environment of VR-Together.

10.2. Use Cases

Use cases are extended descriptions of the user's intention, behaviour and interaction with the system. Below we present the general Use Case for Pilot 1 followed by several detailed use cases describing the platform.

For each Use case we analyse the content in the following structure:

- Title: A descriptive title of the use case
- Brief Description: A brief summary of the action described in the use case mentioning the primary actor and the intention upon which the use case is based.
- Actor: The primary actor(s) taking part in the transaction described.
- Precondition: The requirements that need to be fulfilled before the use case can be initiated.
- Post-condition: The condition of the platform after the use case has been executed/completed.
- Primary path: The basic flow of events that leads to a successful use case execution.
- Alternative path: A differentiated flow of events in which the case can be considered complete or incomplete.

10.2.1. General Use Case

Title: Pilot 1

Brief Description: This use case describes the overall Pilot 1 experience for two users with the VR-Together application.

Actors: 2 users located at two distinct geographical locations.

Precondition:

- A shared virtual space has been instantiated and configured to project the end-users' virtual body representation.
- Pilot 1 content has been produced and is available for projection in the virtual room.
- (At least) two users are located at location where the VR-Together set-up has been deployed.
- The locations in which the set-up has been deployed comply with the system requirements required by VR-Together for Pilot 1.

Post-condition: Each end-user managed to:

- Access the virtual room of VR-Together
- See the virtual representation of another end-user
- See the representation of his own body
- Interact visual and acoustically with another user
- View the content produced for Pilot 1 (when in the virtual space of VR-Together)

Primary path:

1. The end-user starts by wearing the HMD in a room where the capturing devices (visual and audio) have been set-up and configured.
2. A Start Menu scene is shown
3. The experience starts
    a. The end-user views content
    b. The end-user interacts with another user
    c. The user can change the viewing content
    d. The user can change the viewer mode (?)
4. Content Playout ends
5. Exit

10.2.2. Detailed Use Cases

In the following part we examine and unwrap the use cases included in the General Use Case described (3.2.1).

Title	A user creates or edits a profile on the platform
Actors	The end-user
Brief Description	The user creates or edits a profile on the VR-Together platform so that his data is saved.
Precondition	• The end-user has accessed the VR-Together platform satisfying all the related requirements (set-up, connection, hardware, etc.) • The end-user wants to create a profile for accessing the experience OR The end-user wants to edit an existing profile.
Post-condition	• The end-user has created/edited a profile that includes information related to his account. The information that is saved had not been decided but will probably include: o Username o Email o User's height o Default self-representation configuration o Default content format o Default viewer mode
Primary Path	1. The end-user has accessed the VR-Together platform 2. End-user selects to create a profile 3. In the "Create Profile" menu the user introduces: a. Username b. Email c. Height d. Other configuration 4. End-user is finished with the data input

	5. Select “Save configuration” 6. Exit
Alternative Path	1. End-user selects to edit a profile 2. In the “Edit Profile” menu the user introduces: a. Username b. Email c. Height d. Other configuration 3. End-user is finished with the data input 4. Select “Save configuration”

Table 15: Use Case: profile creation/edit

Title	Self-representation configuration
Actors	The end-user
Brief Description	The end-user selects the self-representation format in a session of VR-Together
Precondition	• The end-user has accessed the VR-Together platform satisfying all the related requirements (set-up, connection, hardware, etc.) and able to initiate an active session. • The end-user wants to select the configuration of the self-representation format
Post-condition	• The end-user has configured the self-representation format and can view the changes realised within the virtual space of the VR-Together platform.
Primary Path	1. The end-user has accessed the VR-Together platform and is able to initiate a session. 2. End-user selects to configure the self-representation format 3. The end-user is presented with the number of available options a. TVM b. Point Cloud c. 2D 4. End-user selects one of the options 5. Select “Save configuration” 6. Exit
Alternative Path	In this Use case there is no alternative path as the platform cannot operate without a user-representation configuration.

Table 16: Use Case: Self-representation configuration

Title	End-user create/join room
Actors	The end-user
Brief Description	The end-user creates a new room for other end-users to join or the end-user joins an existing room within the VR-Together platform. The two different starting states are grouped together as in this user case the intention and end-result is the same: the end-user is found in an active session within VR-Together.
Precondition	• The end-user has accessed the VR-Together platform satisfying all the related requirements (set-up, connection, hardware, etc.) • The end-user wants to create a new session OR join an existing session
Post-condition	• The end-user is found in an active session within the VR-Together platform
Primary Path	1. The end-user has accessed the VR-Together platform 2. The end-user has completed his profile configuration 3. The end-user views the available active sessions 4. The end-user joins a session: a. The end-user creates a new session and joins it automatically b. The end-user selects an existing session and joins it 5. The end-user representation is found within the virtual space of VR-Together.

Table 17: Use Case: End-user create/join room

Title	End-user session exit
Actors	The end-user
Brief Description	The end-user exits from the active session in which he/she participates
Precondition	• The end-user has accessed the VR-Together platform satisfying all the related requirements (set-up, connection, hardware, etc.) • The end-user is participating on an active session
Post-condition	• The end-user does not participate in any active sessions within VR-Together

Primary Path	1. The end-user is participating in an active session within VR-Together 2. The end-user selects to exit the active session 3. A dialog window confirms asks the user for confirmation to exit the active session. 4. The user exits the active session and is found in the starting menu of the VR-Together platform.
Alternative Path	In this Use case there is no alternative path as the platform always allows a user to exit a session.

Table 18: Use Case: End-user session exit

Title	Pilot 1 content play-out
Actors	The end-users, Non-Live content
Brief Description	The end-users are found in the virtual room created as mentioned in Table 3. And the self-user representation is configured for each one of them, as mentioned in Table 2. The content playout begins for both users to live the experience described in the Pilot 1 plot (2.1.1)
Precondition	• The end-users have accessed the VR-Together platform satisfying all the related requirements (set-up, connection, hardware, etc.) • The end-users are in the same virtual room • The end-users have selected a self-representation configuration
Post-condition	• The end-users have viewed the content of the Pilot 1 plot.
Primary Path	1. The end-users are in the same virtual room 2. The virtual room Session logic manager initiates the content playback. 3. The end-users view the playback of the content and interact with each other 4. The content playback finishes 5. The end-users are free to exit the virtual room or continue the interaction in it.

Table 19: Pilot 1 content play-out

10.3. VR-Together Software Requirements Matrix

In this section we aim at eliciting the requirements and reflecting upon them in a structured and coherent manner. This section considers the requirements that the VR-Together system addressed at its time of ideation, from September 2016 to November 2016. VR-Together is structured in 3 iterations, each one addressing one technical scenario that will be validated with user groups through 3 pilots. In terms of pilot content, the content initially foreseen to be used in public demos and user evaluations were: an intimate concert, a live news format and a fictional interactive story plot. In terms of technical scenario of each pilot, they were classified as offline, live and interactive, respectively. This breakdown allows the project to work with intermediate objectives at both creative and technical levels, facilitating the consortium to deal with the complexity of delivering satisfactory social VR experiences.

The requirements presented in this section describe the software requirements that will set the ground for discussions regarding further refinements and specifications, as well as a guide for the validation of the pilots.

VR-Together aims at exploring how the combination of various data streams (content, human representations, data) will result in a highly personalized experience that is delivered in an adaptive manner, enabling individuals in different locations to participate together in the same experience. The objective is to deliver close to market prototypes and implement an integrated platform to achieve the main project objective: delivering photorealistic immersive virtual reality content which can be experienced together with friends, and demonstrate its use for domestic VR consumption.

VR-Together is structured in 3 iterations, each one addressing one technical scenario that will be validated with user groups in 3 pilots. Out of each one of these iterations, the project will deliver a system version that will meet the indicated requirements. After each iteration, system and requirements will be validated and the consortium will validate if and to what extent the work done meet each of the requirement. The following table gathers the initial list of general requirements considered by the consortium.

In the following part we recapitulate the Pilot descriptions in order to use them as a logical basis for the requirements table presented in 7.5

10.3.1. Requirements for Pilot 1

In this subsection we review the initial assumptions to be considered in Pilot 1, as initially planned in the project proposal. As described in the proposal, Section 1.3.4.2:

*"Pilot 1. Intimate Concert**.** The goal of the offline pilot is to demonstrate that the innovative media format of VR-Together (orchestrating point clouds, 3DMesh based models and multiple video sources) can produce a more intimate and binding activity than more traditional content production pipelines, based on omnidirectional content. We will compare different capture and production techniques (video, point cloud capture, high-end motion capture) as well as combinations of them to determine quantitative balances among the different formats available (video, point clouds, time-varying meshes, dynamic meshes, motion data). The main variables considered to compare the different means available to deliver such an experience will be:*

· *Production costs, integrating shooting, editing, compositing, post-production, etc.*

· *Bandwidth and computational resources required at the different nodes (capture, encoding, delivery, rendering)*

· *Impact on the subjective social experience among end-users.*

<u>Typology of contents addressed:</u> An intimate music concert seems an ideal starting point to demonstrate VR-Together's innovative media format. It is a good opportunity to show how the VR-Together works for offline produced content. The goal is to demonstrate that the orchestrated delivery of the VR-Together media format, combining several video sources, point cloud and 3D mesh representations will improve closeness with the musicians and with at least 2 distant end-users. Particular care will be taken to integrate facial expression within the production pipeline, i.e. how we will capture the photorealistic 3D actors in costume. For example, uses 108 cameras to capture the actors' performance, costumes, facial expressions and the result is a stream-able 3D model with appropriate facial expressions. This also applies to lighter methods, which are more affordable and portable. For example, uses 4 Kinect sensors and a short automatic calibration process. Industrial methods capturing actor facial MoCap performance using marker-less methods and pre-rigged models will also be considered. Different combination of methodologies and technologies will be studied to deliver the best possible balance between visual quality and cost efficiency in content production."

As described in the proposal, in T4.1, the task that addresses the prototyping and production of demo content:

"Offline CoVR: The content format that we have pre-selected is an intimate concert, which seems relevant to validate the unique feeling of closeness between the audience and the content that the VR-Together platform will deliver. We will also seek to detect implicit social interaction cues that may improve the connection between the audience and the users, such as real-time retargeting of gaze or pointing gestures in the characters being rendered, in order to further integrate the content consumer's presence."

As described in the proposal, in T4.2, the task that addresses the deployment of demos and pilots, with a more practical (technical deployment) approach:

"*Offline CoVR in this first example of content production and delivery, we will focus on validating the staging and capture process to deliver the feeling of co-presence in a shared photo-realistic immersive virtual reality environment. We will study which computer graphics techniques can appropriately blend the representations of end-users, created with real-time constraints, home lightning, affordable cameras and sensors for capture, with the offline produced content. Where possible, we will seek to apply re-illumination techniques to blend end-user representations within the pre-recorded content*. "

10.3.2. Requirements for Pilot 2

In this subsection we review the initial assumptions to be considered in Pilot 2 as initially planned in the project proposal. As described in the proposal, section 1.3.4.2:

"Pilot 2. Live news. We will demonstrate the live production of multi-source immersive content. We will study the conditions which maximize the connection between the audience and the news. Numerous benefits for cost-effective production efficiency will be derived from introducing live processing constraints. Quantitative measures comparing the benefits and costs of introducing offline processing steps will be sought. To realize this scenario, we foresee the creation and demonstration of a hybrid live production that combines omnidirectional cameras and depth sensors and off-the-shelf capture devices targeting consumers (webcam, Kinect) in order to allow several users to feel like being together inside an immersive virtual environment and to increase

the feeling of connection with the environment thanks to embodied social interaction. In this scenario, inter-stream synchronisation is critical: this is not a live VR conference, but a production broadcast. Technically speaking, we need clock sync between equipment at both production environments, and insert / correlate timestamps in the recordings. This kind of activity is aligned with current standardization activities in MPEG MORE, to which part of the VR-Together consortium contributes actively.

Typology of contents addressed: *We will demonstrate a novel content format of immersive news consumption, where people can feel like being together where the news actually occurred. For this we will combine more closely the content production expertise (camera placement, social setting between presenters and the audience, how transitions to other settings (for example, a journalist on the field) can be established and delivered comfortably to the audience, etc. The introduction of live delivery for the case of live news will require a production design adapted to the needs and constraints of News Production (Main set with news presenter, live connection with journalist on the field, etc.), but which still allows for a quality of content as close as possible as an offline production."*

As described in the proposal, in T4.1, the task that addresses the prototyping and production of demo content:

"Live CoVR The content format that we have pre-selected is a broadcasted news, which seems relevant to validate the feeling of immediacy that such techniques can deliver. We will however, study other options if real content production opportunities (events, real concerts, etc) appear, and they seem more appropriate for the validation purpose at hand. "

As described in the proposal, in T4.2, the task that addresses the deployment of demos and pilots, with a more practical (technical deployment) approach:

"Live CoVR In this second example of content production and delivery, we will focus on validating the real-time processing tooling implemented to deliver, at best as possible, the feeling of co-presence in a shared photo-realistic live immersive virtual reality environment. Building upon the insight of first pilot, we will simply aim at assessing to what extent we can preserve the feeling of closeness and empathic connection between the audience and the content, when real-time constraints are imposed. Imposing real-time processing, with no possible offline manual adjustment and manipulation of the content captured severely limits the range of technical possible options. "

10.3.3. Requirements for Pilot 3

In this subsection we review the initial assumptions to be considered in Pilot 3 as initially planned in the project proposal. As described in the proposal, section 1.3.4.2:

"Pilot 3. Interactive Fiction. We will seek to demonstrate how the VR-Together platform, in a custom-designed content production process, can allow for a novel form of content where users meet, and blend within the interactive immersive experience. Thus, consumers can watch passively. However, they are also able to, essentially, become a character within the story plot being rendered. They can have this experience through a more active engagement in the experience, i.e., by moving and talking like one of the characters in the plot, and with these actions change significant aspects of the plot being rendered. This will require the combined delivery of broadcast video, mesh or point-cloud content, together with end-user capture in the form of video, point cloud or interpolated 3dmesh, as well as with event-based synchronization similar to how MMO video-games are synchronized. Regarding the integration of advanced multi-modal pattern recognition, the effort will not be on creating sophisticated multimodal

pattern recognition of social actions, which would work for any plot, but rather to demonstrate how readily available pattern recognition tools (speech recognition, existing gesture recognition algorithms) can be used and integrated to convincingly deliver one specific plot. For this matter, the previous work done within the VR-Together project, regarding spontaneous social interaction in SIVE will become essential to guide this process. Regarding the processing of interactive plots in SIVE, we will use tools readily available from previous research initiatives by the partners within the consortium. The main challenge to maintain place illusion and plausibility is to render credible interactivity within the experience. We will address how to integrate the user input with the events being depicted within the immersive virtual environment. The goal will be to show to what extent and how a fiction scenario can be rendered in VR, while still allowing the users immersed in the scene to intervene actively in the scene being broadcasted within the shared virtual reality experience (and thus, preserving the feeling of being there together).

<u>Typology of contents addressed:</u> We will address interactive content rendered in the form of interactive fiction. This will be demonstrated as a story-like plot rendered within the immersive experience. The user will be able to actively intervene and change some aspects of the experience by performing some of the actions (i.e., talking, pointing or performing simple physical actions) that correspond to the character he/she wants to become within the plot."

As described in the proposal, in T4.1, the task that addresses the prototyping and production of demo content:

"Interactive CoVR. The content format that we have pre-selected is a fiction production, which will allow for additional control in the production process, and will develop a scenario that will be close to a movie script. We will use the insight of subtask T4.3.1 co-presence and social interaction evaluation, in order for the experience of the content to integrate harmonically with possible social interaction occurring, not only among the end-users, but also with the content being rendered. The global aim will be to achieve a qualitatively different level of co-presence, social interaction and place illusion in an aesthetically coherent virtual reality experience."

As described in the proposal, in T4.2, the task that addresses the deployment of demos and pilots, with a more practical (technical deployment) approach:

*"Interactive CoVR**.** In this third example of content production and delivery, we will focus on validating the production of explicitly interactive content to maintain, preserve and if possible reinforce the feeling of co-presence in a shared photo-realistic immersive virtual reality environment. We will seek to detect an expanded range of social and bodily-centred interaction cues (head movements, body movements, peri-personal space, and spoken keywords) to further allow the integration of the end users' actions within the narrative. We will integrate existing innovative interactive storytelling engines available within the VR-Together consortium, along with re-illumination, rendering, and interactive character animation techniques. "*

10.3.4. Experimentation requirements

The evaluation of the VR-Together platform is organised in two different parts. The first part is concerned with validating the different parameters that need to be preserved or improved. This includes aspects such as delays, resolution, etc. These experiments do not imply specific requirements on the overall platform.

The second part is concerned with validating the feeling of being there, in the virtual environment, and of togetherness, i.e., determining under which technical conditions it can be maximize

10.3.5. Requirements Specification Table

ID	Type	No	Version	Component	Title	Description	Priority	Origin	Scope
FR.1.0	FR	1	0	PL	Self-representation	An end user MUST be able to see his own representation in the virtual space of VR-Together	MUST	Grant Agreement	Platform
FR.2.0	FR	2	0	PL	Users audio representation	An end use MUST be able to hear the sounds made by another user in the virtual space of VR Together	MUST	Grant Agreement	Platform
FR.3.0	FR	3	0	PL	Users representation	An end user MUST be able to see the visual representation of another user in the virtual space of VR Together	MUST	Grant Agreement	Platform
FR.4.0	FR	4	0	CA	Capturing setup	A location where the VR-Together platform's capturing setup is deployed MUST capture the audio generated by the user	MUST	Grant Agreement	Platform
FR.5.0	FR	5	0	CA	Capturing setup	A location where VR-Together platform's capturing setup is deployed MUST capture the visual representation of the user	MUST	Grant Agreement	Platform
FR.6.0	FR	6	0	CA	Capturing setup	An end user MUST use a location where a capturing setup is deployed, in order to access the virtual space of VR-Together	MUST	Grant Agreement	Platform
FR.7.0	FR	7	0	DE	Connection	An end user MUST be connected to the delivery network used in the project, in order to access the virtual space of VR-Together	MUST	Grant Agreement	Platform
FR.8.0	FR	8	0	DE	Latency	An end user MUST have a network latency allowing for seamless and natural communication and interaction with other users in the virtual space of VR-Together	MUST	Grant Agreement	Platform

FR.9.0	FR	9	0	OR	Place illusion	End users inside the virtual space of VR-Together MUST be able to see a visual representation of the physical space depicted in the VR content	MUST	Grant Agreement	Platform
FR.10.0	FR	10	0	VRT	VR content	End users MUST be able to watch VR content played in the virtual space of VR-Together	MUST	Grant Agreement	Platform
FR.11.0	FR	11	0	VRT	VR content formats	End users SHOULD be able to see different examples of VR content formats	SHOULD	Grant Agreement	Platform
FR.12.0	FR	12	0	VRT	VR content image quality	End users MUST be able to see photorealistic VR contents	MUST	Grant Agreement	Platform
FR.13.0	FR	13	0	VRT	Synchronization	End-users in distributed locations sharing a virtual space MUST be able to see the same VR content at the same time	MUST	Grant Agreement	Platform
FR.14.0	FR	14	0	PL	End-user image quality	End users MUST see other users participating in the virtual space of VR-Together in photorealistic quality	MUST	Grant Agreement	Platform
FR.15.0	FR	15	0	PL	End-user blend	End users SHOULD see other users seamlessly blended in the virtual space of VR-Together	SHOULD	Grant Agreement	Platform
FR.16.0	FR	16	0	VRT	Comfort	End users SHOULD feel comfort in being immersed in the virtual space of VR-Together, at least for the duration of the pilot experience	SHOULD	Grant Agreement	Platform
FR.17.0	FR	17	0	VRT	Body language	An end-user SHOULD have an experience that visually and acoustically allows to perceive and understand the other participants' body language expressions.	SHOULD	Grant Agreement	Platform
FR.18.0	FR	18	0	PL	3D sound	The VR audio content MUST be directional giving the perception of point sources within the virtual space of VR-Together.	MUST	Grant Agreement	Platform
FR.19.0	FR	19	0	VRT	Audio/Video Synchronization	The VR audio and video content projected within the virtual space of VRTogeher MUST be synchronized.	MUST	Grant Agreement	Platform
FR.20.0	FR	20	0	VRT	End-user devices	End users MUST be able to access the VR-Together platform by using commercially available HMDs and capture systems	MUST	Grant Agreement	Platform

FR.21.0	FR	21	0	CA	Data logging	The VR-Together platform SHOULD record all (motion, speech) end-user activity data	SHOULD	Grant Agreement	Platform
FR.22.0	FR	22	0	PL	Blend of media formats	End users, scene of action and characters SHOULD be able to be projected in the virtual space of VR-Together using different media formats. The resulting VR image should be a blend of different formats.	SHOULD	Grant Agreement	Platform
FR.23.0	FR	23	0	DE	Networks	The data transmission within VR-Together MUST be using commercial communication and media delivery networks.	MUST	Grant Agreement	Platform
FR.24.0	FR	24	0	EE	Adaptive media delivery	Media streams SHOULD provide adaptive quality to network, device and interface capabilities	SHOULD	Grant Agreement	Platform
FR.25.0	FR	25	0	VRT	Web interface	End users MUST be able to access the VR-Together platform using a web application.	MUST	Grant Agreement	Platform
FR.26.0	FR	26	0	VRT	Native interface	End users MUST be able to access VR-Together platform using a native application	MUST	Grant Agreement	Platform
FR.27.0	FR	27	0	VRT	Facial expressions characters	The level of detail of character representation in the virtual space of VR-Together MUST allow the recognition of facial expressions.	MUST	Grant Agreement	Platform
FR.28.0	FR	28	0	VRT	Facial expressions end users	The level of detail of end-user representation in the virtual space of VR-Together MUST allow the recognition of facial expressions.	MUST	Grant Agreement	Platform
FR.29.0	FR	29	0	PL	Offline content	The VR content to be displayed COULD be stored in the end user device	COULD	Grant Agreement	Platform
FR.30.0	FR	30	0	VRT	Offline content	The VR content to be displayed COULD be stored in a network server	COULD	Grant Agreement	Platform
FR.31.0	FR	31	0	OR	Illumination	Illumination MUST be consistent in the whole experience	MUST	Grant Agreement	Platform
FR.32.0	FR	32	0	OR	Gaze	The representations of the rendered characters inside the virtual space of VR-Together MUST be able to retarget their gaze according to the end-user's viewpoint	MUST	Grant Agreement	Platform

FR.33.0	FR	33	0	OR	Pointing gestures	The representations of the rendered characters inside the virtual space of VR-Together MUST be able to retarget pointing gestures	MUST	Grant Agreement	Platform
FR.34.0	FR	34	0	PL	Rendered Characters	The virtual space of VR-Together MUST contain rendered characters	MUST	Grant Agreement	Platform
FR.35.0	FR	35	0	OR	Characters' representation	The representations of the rendered characters inside the virtual space of VR-Together MUST have parallax and depth to allow for a 3D representation.	MUST	Grant Agreement	Platform
FR.36.0	FR	36	0	OR	Characters' representation	The end-user inside the virtual space of VR-Together MUST be able to perceive the 3D appearance of the characters (parallax, depth)	MUST	Grant Agreement	Platform
FR.37.0	FR	37	0	OR	Basic end user movement	The end-user inside the virtual space of VR-Together MUST be able to rotate their head and have certain level of translation capacity while seated (3DoF+)	MUST	Grant Agreement	Platform
FR.38.0	FR	38	0	OR	Number of users	The VR-Together platform MUST allow for 2 to 10 end-users to simultaneously be in the same virtual space.	MUST	Grant Agreement	Platform
FR.39.0	FR	39	0	CA	Live	The VR-Together platform MUST be able to capture the live environment of an end-user.	MUST	Grant Agreement	Pilot 2
FR.40.0	FR	40	0	PL	Live	The VR-Together platform MUST be able to project the reconstruction of the live environment of an end-user.	MUST	Grant Agreement	Pilot 2
FR.41.0	FR	41	0	OR	Active watch	The end-user inside the virtual space of VR-Together MUST be able to become a character within the storyline that is being projected	MUST	Grant Agreement	Pilot 3
FR.42.0	FR	42	0	VRT	Movement	The end-user inside the virtual space of VR-Together MUST be able to move (translation). 6DoF	MUST	Grant Agreement	Platform
FR.43.0	FR	43	0	OR	Derived actions	The end-user's actions inside the virtual space of VR-Together MUST lead to changes in the storyline that is being projected	MUST	Grant Agreement	Pilot 3
FR.44.0	FR	44	0	VRT	Pattern recognition interaction	The VR-Together platform MUST support multi modal pattern recognition mechanics for changing the storyline according to user's choices	MUST	Grant Agreement	Pilot 3

FR.45.0	FR	45	0	VRT	Pointing interaction	The VR-Together platform MUST be able to recognize pointing gestures of end-users and change the storyline accordingly	MUST	Grant Agreement	Pilot 3
FR.46.0	FR	46	0	VRT	Speech interaction	The VR-Together platform MUST be able to recognize the speech of end-users and change the storyline accordingly	MUST	Grant Agreement	Platform
FR.47.0	FR	47	0	VRT	Interactive storytelling	The system SHOULD integrate existing interactive storytelling engines	SHOULD	Grant Agreement	Platform
FR.48.0	FR	48	0	VRT	Interactive character	The system SHOULD integrate and use interactive character animation	SHOULD	Grant Agreement	Platform
FR.49.0	FR	49	0	VRT	Bandwidth configuration	The VR-Together platform MUST support bandwidth configuration options for the end-user	MUST	Grant Agreement	Platform
FR.50.0	FR	50	0	VRT	Delay configuration	The VR-Together platform MUST support delay constraint configuration options for the end-user	MUST	Grant Agreement	Platform
FR.51.0	FR	51	0	VRT	Self-representation configuration	The VR-Together platform MUST support self-representation projection configuration options for the end-user.	MUST	Grant Agreement	Platform
FR.52.0	FR	52	0	VRT	VR content projection configuration	The VR-Together platform MUST support VR content projection configuration options for the end-user.	MUST	Grant Agreement	Platform
FR.53.0	FR	53	0	VRT	Static participants' virtual body representation	The VR-Together platftom MUST allow one end-user to see a static projection of another end-user's body representation within the virtual space.	MUST	Grant Agreement	Experiment
FR.54.0	FR	54	0	VRT	Dynamic participants' virtual body representation	The VR-Together platftom MUST allow one end-user to see a dynamic projection of another end-user's body representation within the virtual space.	MUST	Grant Agreement	Experiment
FR.55.0	FR	55	0	VRT	Participants' virtual body representation distance	The VR-Together platftom MUST allow one end-user to see the projection of another end-user's body representation positioned at various distances within the virtual space.	MUST	Grant Agreement	Experiment

FR.56.0	FR	56	0	CA	People RGB-D Capture framerate	The VR-Together hardware capturing component/system MUST achieve a capture rate of at least 25 fps.	MUST	Grant Agreement	Platform
FR.57.0	FR	57	0	CA	People RGB-D Capture image input	The VR-Together hardware capturing component/system MUST capture RGB-D data from 4 RGB-D devices connected to 4 capturing nodes (RGB-D nodes)	MUST	Grant Agreement	Platform
FR.58.0	FR	58	0	CA	People RGB-D Calibration	The VR-Together hardware capturing component/system RGB-D devices SHOULD be automatically calibrated (extrinsic calibration).	SHOULD	Grant Agreement	Platform
FR.59.0	FR	59	0	EE	People RGB-D Synchronization	The RGB-D frames from the RGB-D nodes MUST be synchronized and grouped in a central node, resulting in a RGB-D group frame.	MUST	Grant Agreement	Platform
FR.60.0	FR	60	0	EE	People live 3D reconstruction	The VR-Together platform MUST process end-user's live coloured 3D point cloud to reconstruct a 3D time-varying mesh in real-time.	MUST	Grant Agreement	Platform
FR.61.0	FR	61	0	EE	People live 3D reconstruction speed	The VR-Together platform MUST perform the People live 3d reconstruction with a delay lower than 80ms.	MUST	Grant Agreement	Platform
FR.62.0	FR	62	0	CA	Foreground removal	The VR-Together platform MUST support foreground removal	MUST	Grant Agreement	Platform
FR.63.0	FR	63	0	CA	Background removal	The VR-Together platform MUST support background removal	MUST	Grant Agreement	Platform
FR.64.0	FR	64	0	CA	User distance from capturing sensor	The VR-Together hardware sensors used in the capturing component/system MUST be placed in a distance lower than 5 meters from the end-user.	MUST	Grant Agreement	Platform
NF.65.0	NF	65	0	CA	Image properties for background removal resolution	The input image captured by the hardware sensors of the capturing component MUST have a resolution 960x540 pixels.	MUST	Grant Agreement	Platform

NF.66.0	NF	66	0	CA	Image properties for background removal	The input image captured by the hardware sensors of the capturing component MUST use a framerate of at least 25 fps.	MUST	Grant Agreement	Platform
NF.67.0	NF	67	0	CA	Face capture	The VR-Together hardware capturing component/system MUST capture the end-user's face from at least two different sides.	MUST	Grant Agreement	Platform
NF.68.0	NF	68	0	CA	Captured face storage	The VR-Together hardware capturing component/system MUST store the captured end-user's face data. The information must be stored (on disk or in memory) and must be accessible in real-time by the face inpainting algorithm.	MUST	Grant Agreement	Platform
NF.69.0	NF	69	0	CA	Stored face painting	The VR-Together hardware capturing component/system MUST be able to perform face inpainting using the stored end-user's face data.	MUST	Grant Agreement	Pilot 2, Pilot 3
NF.70.0	NF	70	0	VRT	Real time compression	The VR-Together platform SHOULD have a delay for the encoding and decoding process of less than 200 ms	SHOULD	Grant Agreement	Platform
NF.71.0	NF	71	0	PL	Progressive decoding	The VR-Together platform SHOULD allow a low quality point cloud to be decoded from a partial bit stream	SHOULD	Grant Agreement	Platform
NF.72.0	NF	72	0	CA	Point cloud compression	The VR-Together platform MUST be able to achieve a compression ratio of up to 1:10 in point cloud streams	MUST	Grant Agreement	Platform
NF.73.0	NF	73	0	VRT	Low end to end latency	The VR-Together platform MUST achieve an end to end (capture to projection) latency that is lower than 300ms	MUST	Grant Agreement	Platform
NF.74.0	NF	74	0	CA	Generic compression framework	The VR-Together platform SHOULD support point cloud compression of arbitrary topology.	SHOULD	Grant Agreement	Platform
NF.75.0	NF	75	0	VRT	Quality assessment	The VR-Together platform SHOULD be able to evaluate the expected quality of experience.	SHOULD	Grant Agreement	Platform
NF.76.0	NF	76	0	VRT	Quality assessment information	The VR-Together platform SHOULD informs the end-user about the expected quality of experience.	SHOULD	Grant Agreement	Platform

NF.77.0	NF	77	0	CA	Texture mesh compression	The VR-Together platform SHOULD be able to achieve a compression ratio of up to 1:30 for textured mesh (3D geometry and textures) content	SHOULD	Grant Agreement	Platform
NF.78.0	NF	78	0	CA	3D mesh generic compression framework	The VR-Together platform MUST support compression for textured 3D time varying mesh content of arbitrary topology.	MUST	Grant Agreement	Platform
NF.79.0	NF	79	0	VRT	Delay of time-varying mesh encoding/decoding	The VR-Together platform MUST perform compression and decompression of texture 3D time-varying mesh content achieving a latency of that is lower than 100ms per frame.	MUST	Grant Agreement	Platform
NF.80.0	NF	80	0	VRT	Real-time time-varying mesh encoding parametrization	The VR-Together platform MUST support TVM compression configuration options for the end-user.	MUST	Grant Agreement	Platform
NF.81.0	NF	81	0	VRT	Real-time time-varying mesh encoding parametrization	The VR-Together platform MUST support texture resolution configuration options for the end-user.	MUST	Grant Agreement	Platform
NF.82.0	NF	82	0	VRT	Real-time time-varying mesh encoding parametrization	The VR-Together platform MUST support texture quality configuration options for the end-user.	MUST	Grant Agreement	Platform
NF.83.0	NF	83	0	VRT	Real-time time-varying mesh distribution parametrization	The VR-Together platform SHOULD support TVM frame time life configuration options.	SHOULD	Grant Agreement	Platform
NF.84.0	NF	84	0	VRT	Real-time time-varying mesh distribution parametrization	The VR-Together platform SHOULD support TVM frame queue length configuration options.	SHOULD	Grant Agreement	Platform

NF.85.0	NF	85	0	EE	End-user audio encoding	The VR-Together platform MUST use typical browser supported audio encoding.	MUST	Grant Agreement	Platform
NF.86.0	NF	86	0	EE	End-user video encoding	The VR-Together platform MUST use typical browser supported video encoding.	MUST	Grant Agreement	Platform
NF.87.0	NF	87	0	EE	End-user audio encapsulation	The VR-Together platform MUST use typical browser supported audio encapsulation.	MUST	Grant Agreement	Platform
NF.88.0	NF	88	0	EE	End-user video encapsulation	The VR-Together platform MUST use typical browser supported video encapsulation.	MUST	Grant Agreement	Platform
NF.89.0	NF	89	0	OR	Configuration	The VR-Together platform orchestration module MUST support remote operation	MUST	Grant Agreement	Platform
NF.90.0	NF	90	0	OR	Session management	The VR-Together platform orchestration module MUST manage sessions where 2 or more end-users participate in a virtual space.	MUST	Grant Agreement	Platform
NF.91.0	NF	91	0	OR	Session management	The VR-Together platform orchestration module SHOULD support more than one parallel sessions.	SHOULD	Grant Agreement	Platform
NF.92.0	NF	92	0	OR	Optimization	The VR-Together platform orchestration module MUST be able to configure the end-user play-out component.	MUST	Grant Agreement	Platform
NF.93.0	NF	93	0	PL	Content	The VR-Together platform web player MUST support playback of 2D VR video content.	MUST	Grant Agreement	Platform
NF.94.0	NF	94	0	PL	Content	The VR-Together platform web player MUST support playback of 2D end-user representation projection.	MUST	Grant Agreement	Platform
NF.95.0	NF	95	0	PL	end-user playout resolution	The VR-Together platform MUST support playback of end users representation of at least 960x540 pixels	MUST	Grant Agreement	Platform
NF.96.0	NF	96	0	PL	end-user playout frame rate	The VR-Together platform MUST support playback of end users representation at a framerate of at least 25 fps.	MUST	Grant Agreement	Platform
NF.97.0	NF	97	0	PL	Audio	The VR-Together play-out component platform SHOULD support spatial audio.	SHOULD	Grant Agreement	Platform
NF.98.0	NF	98	0	PL	Streaming	The VR-Together play-out component MUST support input of separate VR content and end user representations streams.	MUST	Grant Agreement	Platform

NF.99.0	NF	99	0	PL	WebVR	The VR-Together play-out component's web player MUST operate in a browser that supports WebVR and A-frame.	MUST	Grant Agreement	Platform
NF.100.0	NF	100	0	PL	Bandwidth	The VR-Together play-out component's web player SHOULD support content bandwidth adaptation.	SHOULD	Grant Agreement	Platform
NF.101.0	NF	101	0	VRT	Stream Latency	The latency between different streams on the VR-Together platform MUST not be higher than 500 ms.	MUST	Grant Agreement	Platform
NF.102.0	NF	102	0	VRT	Scene	The content describing the scene of the VR-Together rooms MUST be static 2D 360 images, having a maximum of 4K pixels, in ERP format.	MUST	Grant Agreement	Platform
NF.103.0	NF	103	0	PL	Multiple format	The native player MUST support play-out of content in different VR formats, like Point Clouds, omnidirectional video, static meshes, dynamic meshes, mono/stereo 2d video.	MUST	Grant Agreement	Platform
NF.104.0	NF	104	0	PL	Hybrid format	The VR-Together play-out component's native player MUST support the reproduction of hybrid VR contents within virtual space.	MUST	Grant Agreement	Platform
NF.105.0	NF	105	0	PL	Audio	The VR-Together play-out component's native player MUST support spatial audio.	MUST	Grant Agreement	Platform
NF.106.0	NF	106	0	PL	Rendering frame rate -native	The VR-Together play-out component's native player SHOULD be able to render combined media formats at 60 fps or more	SHOULD	Grant Agreement	Platform
NF.107.0	NF	107	0	PL	Lighting changes - native	The VR-Together play-out component's native player SHOULD be able to alter the lighting of specific objects within the virtual space, on the basis of custom shaders.	SHOULD	Grant Agreement	Platform
NF.108.0	NF	108	0	PL	DoF	The VR-Together play-out component's native player MUST be able to reproduce content adapted to 3DoF or 3DoF+ movements.	SHOULD	Grant Agreement	Platform

NF.109.0	NF	109	0	PL	Quality of Image	The VR-Together play-out component's native player input/output effective display resolution MUST be up to 4K	MUST	Grant Agreement	Platform
NF.110.0	NF	110	0	PL	Delay on displaying self-representation	The VR-Together play-out component's native player self-representation projection MUST have latency under 20ms.	MUST	Grant Agreement	Platform
NF.111.0	NF	111	0	PL	Sync multiple formats	The VR-Together play-out component's native player SHOULD support synchronization between different input formats with less than 40ms of delay.	SHOULD	Grant Agreement	Platform
NF.112.0	NF	112	0	PL	Sync inter device	The VR-Together play-out component's different players SHOULD support synchronization of frame accurate with a delay lower than 20ms.	SHOULD	Grant Agreement	Platform
NF.113.0	NF	113	0	VRT	Sync control	The VR-Together platform SHOULD support manual synchronization control options for the end-user.	SHOULD	Grant Agreement	Platform
NF.114.0	NF	114	0	VRT	Timestamping	The VR-Together platform capturing component MUST timestamp media content in relation to a platform-wide common clock.	MUST	Grant Agreement	Platform
FR.115.0	FR	115	o	CA	Sound Recording	The VR-Together platform capturing component SHOULD record and store the recordings of the HMD for further future analysis purposes.	SHOULD	Experiments Analysis	Experiments
FR.116.0	FR	116	0	CA	Field of View Recording	The VR-Together platform capturing component SHOULD record and store the field of view of the HMD at each moment for further future analysis purposes.	SHOULD	Experiments Analysis	Experiments

Table 20: D2.1 Requirements Matrix

11. ANNEX IV: LIST OF EXPERIMENTS (PILOT 1)

This Annex provides an overview of the experiments initially considered in the project, and serves as a plan for project activities in terms of piloting and evaluation, as part of WP4 tasks. Further information regarding experiments will be provided in future versions of this document and WP4 documents. *All the experiments planned for Pilot one have been performed and can be found in the Deliverable D4.2.*

11.1.1. Technology Evaluation

These types of evaluations have a technical value for the project, as they allow for further development of the technology, or profiles the technical performance. In particular, we have run the following studies:

- CWI-1: with the objective of defining a quality metric for evaluating point clouds. This is ongoing work that will feed standardisation activities and will help on the optimisation of the system
- CERTH-1 and CERTH-2: with the objective of evaluate and assess the technical performance of the system
- CERTH-3 and CERTH-4: with the objective of helping the development related to HMDs and their removal

11.1.1.1. CWI-1

Point cloud is a good alternative for representing 3D objects and scenes in immersive systems. This study explores the objective and subjective quality assessment of point cloud compression. Existing work on point cloud quality assessment has mainly focused on point cloud geometry, and demonstrated that state-of-the-art objective quality metrics poorly correlate with human subjects' assessments. Not much attention has been given to point cloud quality evaluation based on its colour, even though real world applications utilize colour point clouds, and colour artifacts may be introduced during compression due to different colour coding schemes. As for point cloud subjective quality assessment, limited insight has been presented on how users evaluate and perceive the quality of compressed point clouds. Through our experiments, we propose objective quality metrics for point cloud compression based on colour distribution, and provide a comparison of its performance with the commonly used geometry-based metrics.

11.1.1.2. CERTH-1

A very early technical experiment was conducted in order to assess the total distribution performance of time-varying mesh (TVM) pipeline, allowing us for better understanding the required improvements.

Offline TVM data were used and transmitted in real-time, enabling the evaluation of the real-time distribution of TVMs. Two RabbitMQ server instances were used, one in i2Cat (Spain) and one in CERTH (Greece) allowing the evaluation of different networking topology for the RabbitMQ servers.

11.1.1.3. CERTH-2

In this experiment, a technical evaluation was conducted in order to assess the per-module distribution performance of time-varying mesh (TVM) pipeline, allowing us for better understanding the required improvements.

Users in Greece (Thessaloniki, CERTH) will be captured and reconstructed, while the data will be transmitted in real-time, enabling the evaluation of the real-time distribution of TVMs. One user lab node (CERTH - 5 PCs and 4 RGB-D sensors) and two RabbitMQ server instances were used, allowing the evaluation of local and remote RabbiMQ server usage.

11.1.1.4. CERTH-3

The CERTH-3 experiment considered and ran a comparator analysis in order to draw a conclusion on the most appropriate hardware devices to be used for the TVM configuration of the VR-Together project for Pilot 1.

On the side of the capturing camera hardware components, the compared items were:

- Kinect for Xbox One
- Intel RealSense D415

On the side of the head mounted displays to be used as the hardware rendering devices, the compared items were:

- HTC Vive
- Oculus Rift DK2

Given the promising results of Oculus Rift DK2, we decided to conduct some further experiments using Oculus Rift HMD. As assumed, the Oculus Rift HMD worked properly with both RGB-D devices, thus, it has been considered the appropriate device for Pilot 1.

11.1.1.5. CERTH-4

When a user is immersed in VRT, he/she wears a VR HMD, thus, the face of the full body 3D user representation (Time-Varying Mesh) is occluded, leading to major loss of discriminant facial information. The presence of the HMD during multi-user communication in the virtual environment weakens the feeling of co-presence and prevents the user from being fully immersed.

The main goal of this experiment was to create a dataset in order to develop, train and evaluate an algorithm that will perform efficient and real-time HMD removal, exploiting the full information medium (i.e., colour (RGB) and depth data). A special data capturing system was designed to acquire RGB-D faces with and without HMDs. The dataset will be publicly available and will be utilized for the HMD removal task, in the context of VR-Together.

11.1.2. User Experience Evaluation

These types of evaluations have the objective to better understand the user experience. In particular, during this year we have been able to develop a new protocol for evaluating social VR, tested in two different settings. Such protocol is the one that has been used for evaluating the pilot content. In particular, we report:

- Artanim-1 and Artanim-2: initial experimentations with avatar representations and the impact of different levels of body animation fidelity, paving the path towards pilots 2 and 3
- CWI-2 and CWI-3: user experience evaluations used for the development of a protocol for social VR, including both subjective and objective methodologies. The experiments include comparisons between different level of representations (avatars, 2D)

11.1.2.1. ARTANIM-1 and ARTANIM-2

We present two experiments to assess the relative impact of different levels of body animation fidelity of a user controlled virtual avatar (ARTANIM-1) and of a virtual character that is not controlled by the user (ARTANIM-2) to plausibility illusion (Psi). Psi concerns the feeling that events in a virtual environment may be really happening and is part of Slater's proposition of two orthogonal components of presence in virtual reality (VR). We emphasize that these experiments only address self and others representation based on 3D rigged meshes, which will be used as a baseline for experiments in the case of self-representation, and part of the content in the pilots for pre-recorded as well as live actors interacting with users.

In the first experiment (ARTANIM-1) we address the question: to what extend the self-avatar animation fidelity affects Psi? In addition, we also asked users to rate whether each animation feature had a positive effect on the sense of control of their self-representing avatar. The sense of control relates to the concepts of agency and embodiment, where the perception of sensorimotor contingencies can affect the experience of agency, the sense that one has motor control over the avatar, that one develops with the virtual representation of oneself. By improving our understanding of how users perceive the animation features of a self-avatar we can propose a baseline self-representation that other partners can use as a parameter to measure how and whether and to what extend the photorealistic (lookalike) self-representation technologies proposed in VR Together improves the experience of the user.

In the second experiment (ARTANIM-2) we address the question: to what extend the animation fidelity of a character that is not controlled by the user affects Psi? By improving our understanding of users' perception of pre-recorded or live actor character animation.

11.1.2.2. CWI-2

The goal of the experiment is to understand the user experience of photo sharing in social VR, comparing with face-to-face photo sharing and Skype photo sharing.

Research questions

1. "Compared with Face-to-face condition and Skype condition, how is the user experience of digital photo sharing in social VR."
2. "What are the advantages and disadvantages of social VR?"

This experiment applied a within-subjects research method [1].

This method helps to better compare the three conditions. Each pair of subjects was asked to share photos with each other in three conditions:

Condition 1 (A): Face-to-face
Condition 2 (B): Skype
Condition 3 (C): Facebook space social VR

The Face-to-face condition was selected because it serves as a standard condition. The Skype condition was selected because it is one of the traditional mediated social communication tools, and the way people interact in Skype is close to real life.

The results of this experiment and an extensive description can be found in D4.2

11.1.2.3. CWI-3

This experiment aimed at developing and testing the subjective and objective methodologies to evaluate and compare social VR systems to be used during pilot 1.

We considered the scenario of two users sitting in the same Virtual Environment (VE), where they can interact with each other by audio and visual interaction, and watching movie trailers together on a virtual screen.

The results of this experiment and an extensive description can be found in D4.2

11.1.3. Feedback from Professionals

These experiments are intended for gathering feedback from professionals and experts, at fairs and exhibitions.

In particular, we report:

- TNO-1: at VRDays 2017 in Amsterdam
- TNO-2: exploration of the system in other case studies (work meetings)
- TNO-3: at MMSys 2018

11.1.3.1. TNO-1: Initial use-case study

TNO's first experiment was conducted at VR days 2017 in Amsterdam. With the components and platform available at that moment, feedback was collected from users about relevance and importance of Social VR in general and most important use cases in Social VR in particular.

Research questions:

- RQ1: Is Social VR relevant for people?
- RQ2: What are the most important social VR use cases?
- RQ3: How do you measure the user experience in Social VR?

Hypothesis:

- H1: People are interested in being together in immersive VR while being able to communicate with each other.
- H2: People are interested in Social VR.
- H3: Social VR gives people a better experience then VR or traditional mediated communication.

The results of this experiment and an extensive description can be found in D4.2

11.1.3.2. TNO-2: Try-out of VR stand-up

The aim of this experiment was to determine to what extent the current video-based Social VR system is suitable for doing field trials for stand-ups in VR in an enterprise setting. The company in question is doing IT development according to Scrum, and is a global company with many teams being distributed across countries. Currently, its developers are not satisfied with their current video conferencing capabilities at hand. Partly because of this, a lot of developers travel back and forth a lot, e.g. on a weekly basis, to keep the contact within the team optimal.

The main goal of the experiment is to determine if a field trial would be suitable, and if so, under what conditions.

A secondary goal of the experiment is to gather feedback on our system from a market party, in this case a potential buyer of such a Social VR system. The experiment is thus also about requirements gathering.

The expectation was that the audio and video quality of the system will be sufficient for interaction/communication purposes.

Also, it was expected that the system as offered, would not be sufficient for a field trial. During an intake with the company, we discussed various issues with the current setup:

- Maximum number of participants of 4, while most teams within the company are between 6 and 8 persons.
- The HMD is visible during communication, which prevents eye-contact. The expectation was that HMD removal would be needed.
- Many teams use some kind of Kanban board, the company also often uses whiteboards and markers during these sessions. For a field trial, it is expected that some additional functionality (i.e. shared interactivity) is needed.

11.1.3.3. TNO-3: Representing the environment and users in either 2D or 3D

In this experiment the goals were the following:

- Goal 1: Test the technical feasibility of representing both users and the environment in 3D using the web player.
- Goal 2: Test the technical feasibility of utilizing RGB-D data for constructing 3D user representations using the web player.
- Goal 3: Compare the new 3D representation with the 2D monoscopic 360-degree web version.

This formed the following research questions:

- RQ1: What is the performance of the 3D user approach (bandwidth, CPU) and 3D room environment (CPU/GPU/Memory)?
- RQ2: Which room representation is better 2D or 3D?
- RQ3: Which user representation is better 2D or 3D?

The results of this experiment and an extensive description can be found in D4.2

11.1.3.4. TNO-4: Input from professionals based on the pilot 1 experience

The goal of experiment is to gather input from industry professionals on the VR-Together project in the broader sense. As the VR-Together project aims at delivering components that can actually see use in the industry, it is important to get feedback from the industry on expected timelines and on which aspects are more important than others.

Research questions

- RQ1: When is VR expected to take off?
- RQ2: What are the most important VR applications?
- RQ3: Which content is suitable for VR?

- RQ4: Which content is suitable for experiencing it together in VR?
- RQ5: Which aspects are important for shared VR experiences?

Hypothesis

No specific hypotheses were developed for this experiment, as the goal was to collect open input of industry professionals.

11.1.1. Calendar of experiments

The list and calendar followed for the Pilot 1 experiments can be found in the table below.

Experiment Id	Area	Involved Partners	When
CWI-1	QoE objective metrics of point cloud compression (24 Users)	**CWI**	10/2017
-	Questionnaire on demo platform	**TNO**	11/2017
CERTH-1	Real-Time distribution of time varying meshes (no users involved)	**CERTH**	02/2018
CERTH-3	interference between HMD and multiple depth-sensing (2 users)	**CERTH**	02/2018
-	Questionnaire on demo platform	**TNO**	04/2018
CWI-2	Design guidelines for Social VR (10 + 52 users)	**CWI**	05/2018
-	Focus Group	**CWI**	06/2018
CERTH-2	Real-Time distribution of time varying meshes part 2 (5 users)	**CERTH**	07/2018
CWI-3	social VR Ground Truth (32 users expected)	**CWI**	07/2018
Artanim-1	Impact of movement animation of the virtual body parts to presence	**Artanim**	07/2018
Artanim-2	Impact of movement animation of the virtual body parts to co-presence	**Artanim**	07/2018
CERTH-4	HMD removal part 2	**CERTH**	07/2018
-	User tests on Pilot 1	**I2Cat**	09/2018
-	Interview to the advisory board members about Pilot 1	**CWI/TNO**	01/2019
-	Questionnaire to end-users at the Sundance Film Festival 2019	**TNO**	01/2019

Table 21: Pilot 1 Experiments Calendar