Deliverable

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

D2.1- User scenarios, requirements and architecture v1

Revision: 1.0

Authors: Sergi Fernandez (i2CAT)

Delivery date: M5 (12-02-2018)

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

Dissemination Level		
P	Public	
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	10-12-2017	J.Llobera, J.A.Núñez, P.Cesar	I2CAT,CWI	Table of contents, detailed work breakdown and use cases
0.2	17-12-2017	J.Lajara	FLH, ENT	Content pilot designs and plans
0.3	22-12-2017	P.Cesar	CWI	Drafting and consolidation
0.4	15-12-2017	Sergi Fernandez	I2CAT	Modification of ToC structure, Initial set of use cases, high level requirements
0.5	24-12-2017	ALL	ALL	Product Functions
0.6	31-01-2018	ALL	ALL	Specific Requirements
0.7	01-02-2018	M.Prims, J.Llobera, P.Perrot	TNO, i2CAT, VIA	Alternative UCs & SW/HW Architecture Diagrams
0.8	05-02-2018	S.Fernandez	I2CAT	Use Cases
0.9	07-02-2018	J.Llobera	I2CAT	Executive summary & Conclusions
1.0	12-2-2018	S.Fernandez	I2CAT	Compilation & final review.

EXECUTIVE SUMMARY

The present document is the reference document for the software integration and experimental work to be developed within VRTogether. It outlines the requirements, architecture and experimental work envisaged to implement the main paradigm outlined in VRTogether: the creation of a platform and media content that allows two users to feel as if they were together, on the basis of delivering photorealistic media both for content and for end-user representation.

To clarify the integration of these aspects, after a brief introduction, we start outlining the project requirements, as defined in the annex of the grant agreement (section 2), for the three pilots. It begins by outlining general requirements of the project, and then specific requirements for each of the three pilots.

Section 3 details the scenario for pilot 1. It includes a technical breakdown in the form of a visual storyboard, aimed at facilitating the understanding of the pilot for the reader. We have chosen to produce a content similar to a police interrogation scene. This will allow introducing a good over-arching story across the three pilots, and seems a good balance between a commercially-relevant content format, close to typical thriller-like content, and one that enables exploring different possibilities regarding experimental validation. We have also chosen to produce two versions of the same content, one where actors are captured on video and rendered as stereo video billboards within a 3D environment, and another with 3D rigged characters combined with motion capture techniques. In addition, we will also create versions of these rendered as omnidirectional video. All these production and post-production efforts will greatly facilitate direct comparison of media formats and enable a much richer evaluation of the experience, and a better understanding of how the feeling of being together in virtual reality is shaped by technical and psychological factors.

Section 4 outlines in further detail the software requirements for the VRTogether platform, particularly in the scenario outlined for pilot 1. First, it outlines the characteristics of the two main use cases –i.e., Content consumption and Social interaction. Then, we analyse it from a product perspective, and outline further requirements for different components: the native player, the web player, and the orchestration components. Furthermore, we outline the user characteristics of the expected content consumer, but also of the platform administrator, in charge of managing and deploying the overall platform, and the experimenter, who is in charge of using this platform to assess togetherness. Finally, we introduce the reference documentation and additional technical considerations.

Section 5 introduces the overall architecture for pilot 1, and how the different components interact, taking into consideration the software modules from WP3.

Section 6 outlines all the experimental and validation work to be performed in the end-user labs. It details the advisory board, and it includes a detailed layout of the experimental platforms available or installed for VRTogether, and a planning of the different experiments involved in these. Finally, section 7 summarizes the contributions of the deliverable.

Additional annex 1 and 2 provide the complete list of requirements and an example of a questionnaire used for experimentation.

CONTRIBUTORS

First Name	Last Name	Company	e-Mail
Javier	**Lajara**	FLH	javier.lajara@futurelighthouse.com
Guillermo	**Calahorra**	ENTRO	guillermo@entropystudio.net
Pablo	**Cesar**	CWI	p.s.cesar@cwi.nl
Joan	**Llobera**	I2CAT	joan.llobera@i2cat.net
Omar	**Niamut**	TNO	omar.niamut@tno.nl
Juan	**Nuñez**	I2CAT	juan.antonio.nunez@i2cat.net
Martin	**Prins**	TNO	martin.prins@tno.nl
Argyris	**Chatzitofis**	CWI	tofis@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

CONTENTS

TABLES OF FIGURES 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 on the initial project requirements, the use cases contemplated for the pilot 1 scenario and the system architecture envisaged to meet the initial 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 an initial review of project requirements and specific requirements per project component. The list of requirements gathered in this document will serve as a basis for discussions towards component implementation and integration of the first version of the VR-Together platform.

1.3. Status of this document

This document will be alive during the whole project period, that is, during the 3 iterations of the project. Three different versions will be formally submitted to the EC and uploaded in the project website.

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 at collecting and reflecting in a structured manner the high level 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 contents, contents initially foreseen to feed public demos and user evaluations were: an intimate concert, 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 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 gathered in this section describe the initial project requirements which will be ground of discussion for further refinement and specification, as well as a guide for the validation of the pilots. The section is structured as follows: first, we introduce the initial set of general requirements (those requirements that will or should be valid during the overall execution of the project and that all versions of the VR-Together system should meet). Second, we recapitulate the ideas included in the project proposal that differentiate each one of the pilots and we provide a list of specific requirements per pilot. Forth, we compiled the initial scenarios 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 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.

CODE	NUM	TITLE	DESCRIPTION
GEN	1	Copresence	End users should be able to be virtually present in the same virtual space and engage in real-time face-to-face social activities. Copresence 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 others 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

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+)

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 an 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.

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).

Typology of contents addressed: 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

2.5. 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 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 behavioral measures related with togetherness, we need to be able to record the entire multi-modal data, with good time precision.

3. PILOT SCENARIOS

The three pilots of VR-Together address specific objectives in terms of technical challenges and evaluation purposes. Pilots are project checkpoints to evaluate the creative and technical challenges addressed towards the creation of truly realistic social VR experiences.

Pilots are initially planned as individual content capsules addressing completely different content scenarios. The structure and plot complexity of the pilots is linked to an increased technical difficulty, being the first pilot the simpler to produce and technically elaborate and the third the most complex. These three pilots were initially planned as follows:

-A first offline pilot, simulating an acoustic music concert in which it is intended 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 supposed to focus on live news, pretending to be a live production of immersive content from multiple sources that aimed to transfer the user to the location of the news and share the experience with other users.

-Finally, the third pilot intended to present a test to users to 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.

Pilots are planned to be executed in July –September2018, June-August 2019 and May-July 2020. During this period, a number of experiments will run in the project user lab and project roadshows demonstrating the status of results are planned to be launched in technical and creative industrial fairs and events)

3.1. Scenario Pilot 1

During the kick-off meeting in Barcelona in mid-October 2017, the artistic partners proposed a different approach, by creating a coherent storyline across the three pilots (each one of them being one scene in an overall plot). The hypothesis is that by changing the original concept and plot line of the pilots, we will offer in the end a more attractive and engaging experience to the user. Moreover, this will allow the project to provide a finished coherent novel "product" that can be showcased in cinema festivals and other venues. It will capture the attention of the consumers, making them participant in the show thanks to an elaborated plot and interaction

between players. This will in turn help 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.

One key concern from the consortium was if such a new approach would fulfil the needs of the project in terms of providing a representative social VR experience, which can be supported by our novel platform. In particular, it is essential that the pilots serve as a vehicle for technological advances and for experimentation (co-presence, togetherness, immersion). During subsequent meetings between the creative partners and the technical partners, specific agreements have been reached in order to ensure that the set of pilots are a valid vehicle for the project. The first trial will still focus on communication between remote participants while doing an activity together, the second one will focus on the scalability of the platform, and the third one on the interactivity with the content.

3.1.1. Plot Pilot 1

3.1.1.1. Initial plot ideas

The selected plotline relates to a police theme (police investigation or interrogation), which will still fulfil the basic requirements of the project. This new storyline will exploit the uniqueness of the project, a team composed on technical and artistic experts, by creating a brand new experience that makes the most of the writing possibilities. The final objective thus is to obtain a new concept experience that involve the viewers and make them live an experience totally unique and different from what they had previously experienced.

One of the questions we were wondering was “how is the target of this type of product like?” We prefer not to limit the product, so we propose a general point of view for all kind of audience, making it more general and not limited to a specific type of public. Having as inspiration movies like The Usual Suspects, the proposal is to have a thriller plot as a theme of the pilots. This way, the viewer (who will have a say) can enjoy the experience not only during the pilot, but also after that. Overall the structure of the pilots is shown in Figure 2.7.

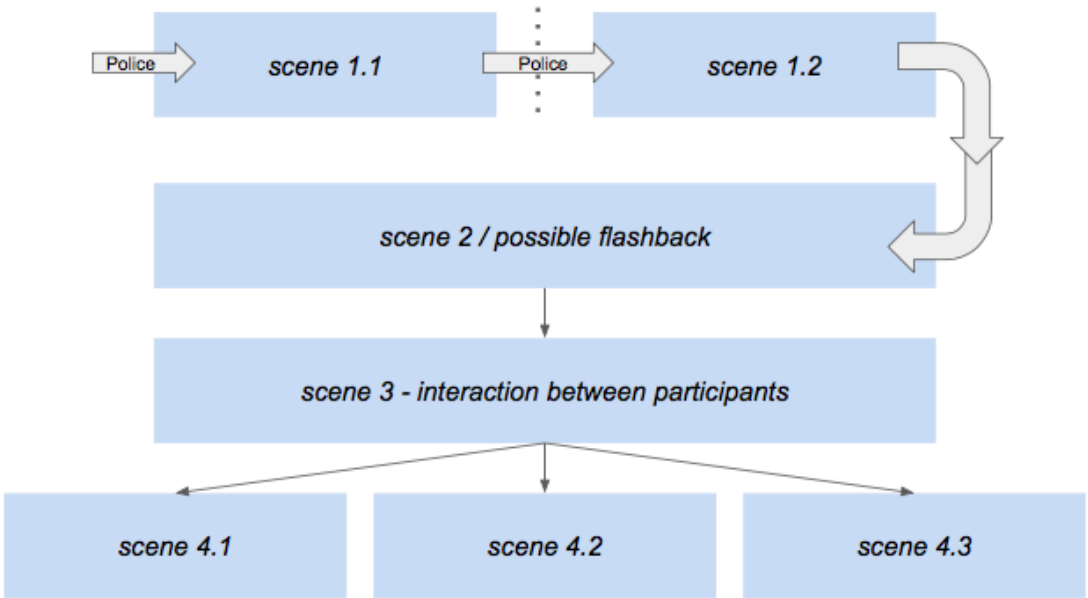

Figure 1. Scenes integrating general story

The creative partners of the project went through a number of iterations, creating visual representations that facilitated discussion (see Figures 2.8 and 2.9). More concrete ideas were put on the table, approved by the partners, and step by step evolved towards a trial definition.

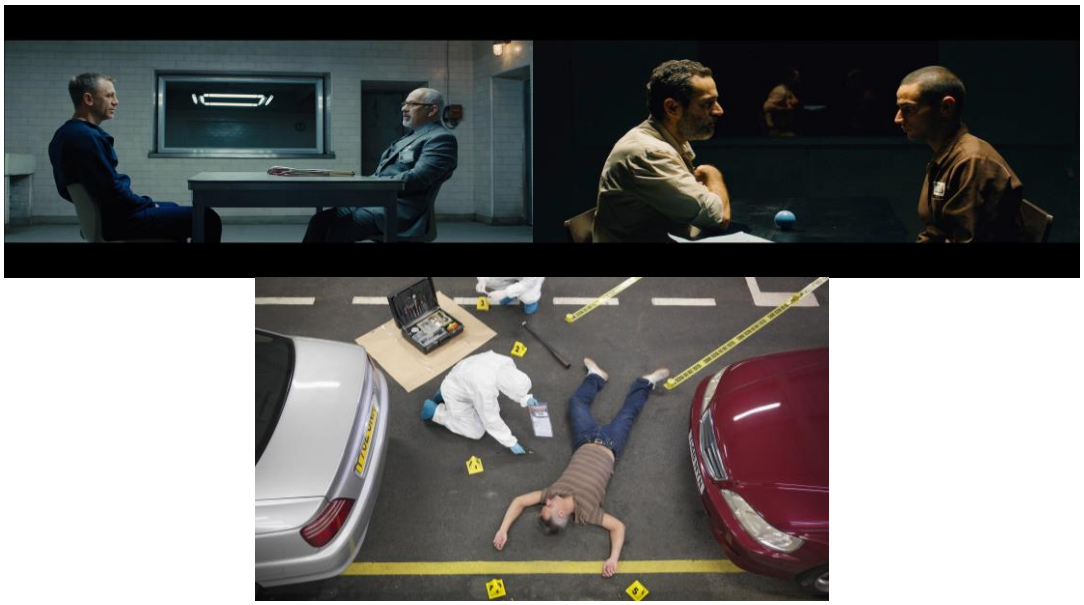

Figure 2. Initial Concepts for the Trial (interrogation, crime scene).

During the face-to-face TCC meeting of the project partners in Madrid (November 2017), three main ideas for pilot one were presented: murder scene (see Figure 2.10), interrogation with one way mirror (see Figure 2.11), and interrogation inside the prison (see Figure 2.12). The general ideas of each type of scene can be summarized as follows:

- Murder scene: it was intended that both users were in the same room where a murder had been committed, and that both users were sufficiently separated from each other to have different viewing angles and, therefore, visual contact with different objects and tracks. The collaboration of both users (involving the feeling of togetherness) would be essential to make conclusions.
- Interrogatory room with one-way mirror: the users would be behind a one-way mirror of an interrogation room. Although users would be next to each other, each user would see their own interrogation room, both being aware that the other user is having their own story.
- Interrogatory inside the prison: In this third version, both users are inside the prison in front of the accused. The interaction between both users within the scene is possible.

Figure 3. Pilot Proposal – murder scene.

Figure 4. Pilot Proposal – interrogation with one-way mirror.

Figure 5. Pilot Proposal – interrogation inside the prison.

After deliberation, the project partners selected the second one. In the next months, Entropy Studio and Future Lighthouse will develop the storyboard of the general concept of the pilot and plan the production.

3.1.1.2. Storyboard

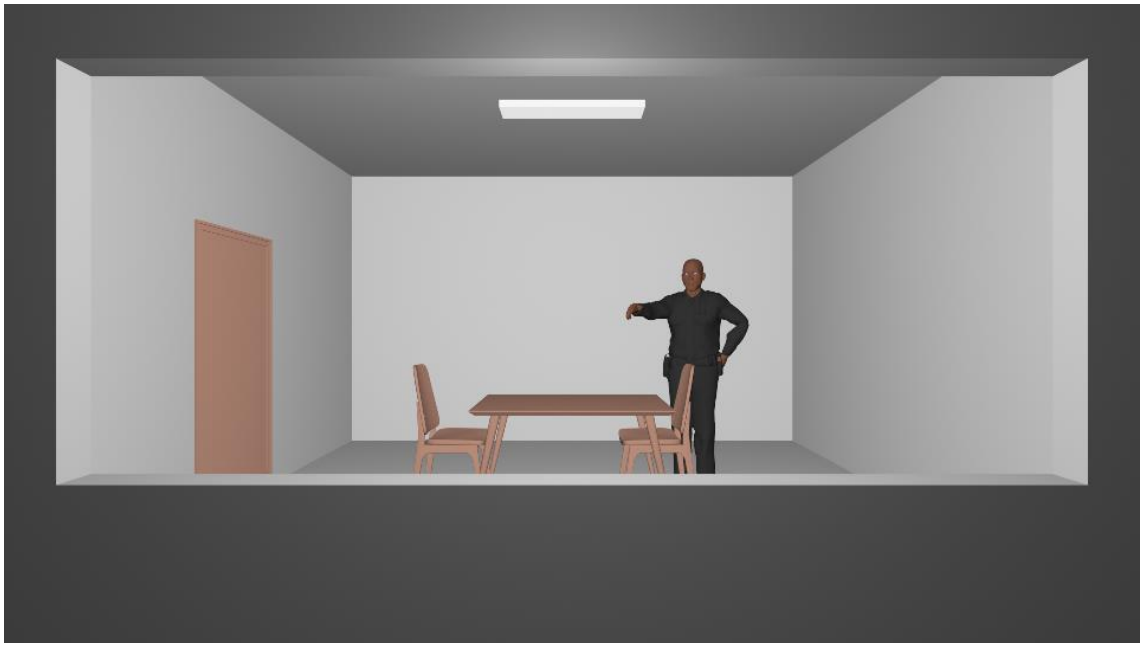

Figure 6. Police officer waiting for the suspect (scene 1)

In the intro, we are on the dark side of a interrogatory room. A police officer is waiting patiently for the suspect. Beside us, we can see and hear our friend, displayed as a point cloud, in a room like ours. After a short time, between 5 and 10 seconds, the police officer makes the intro to the plot. A suspect is about to be interrogated by him and we are the witness of the interrogation. Some clues should be gathered by us in order to clarify the authority of the crime.

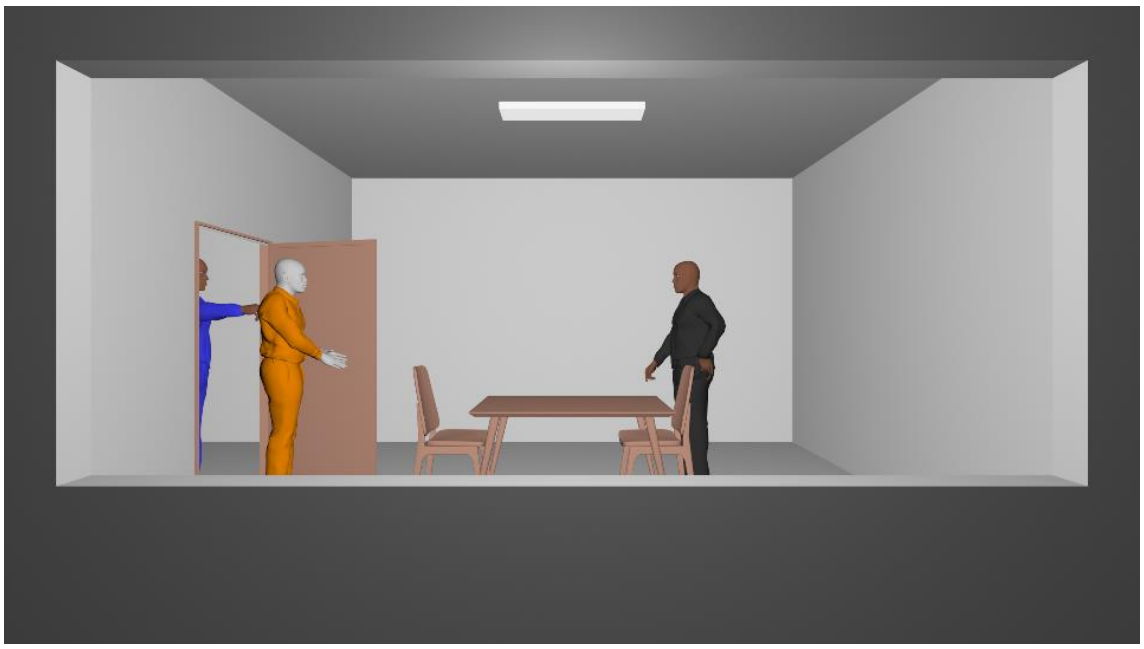

Figure 7. Suspect introduction (scene 2).

For first suspect introduction, he enters the room and sits on a chair. He is handcuffed

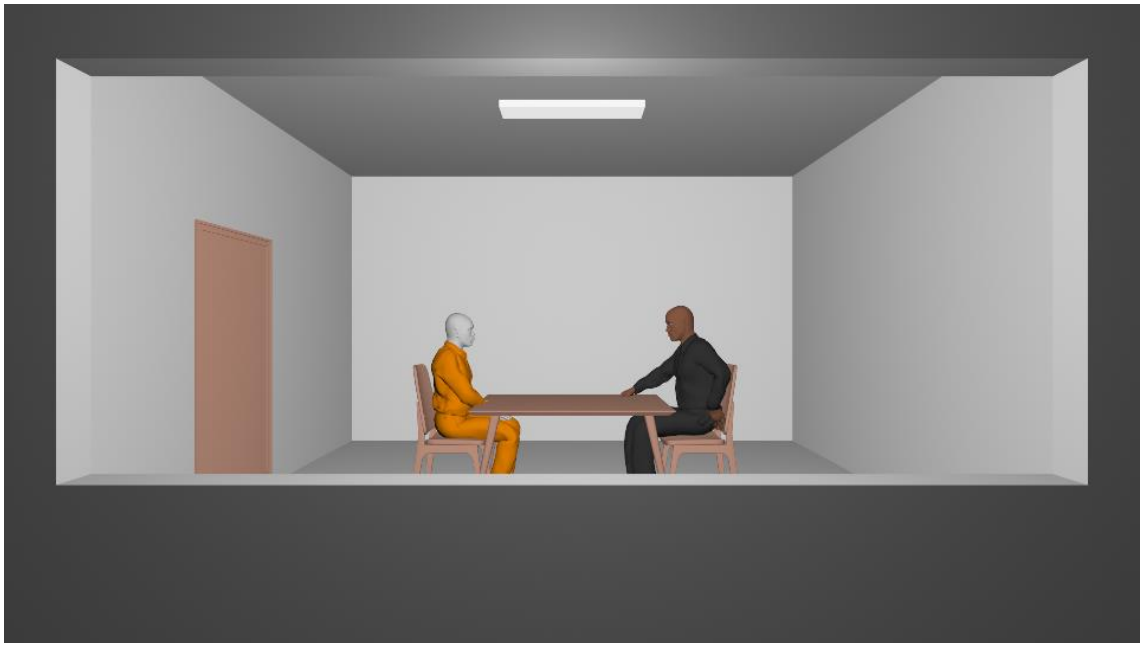

Figure 8. Interrogatory (scene 3).

To create a conflict in the plot, the script makes the officer to start making questions, talking about the situation of the crime scene, where were him at that moment, where were the other suspect, etc.

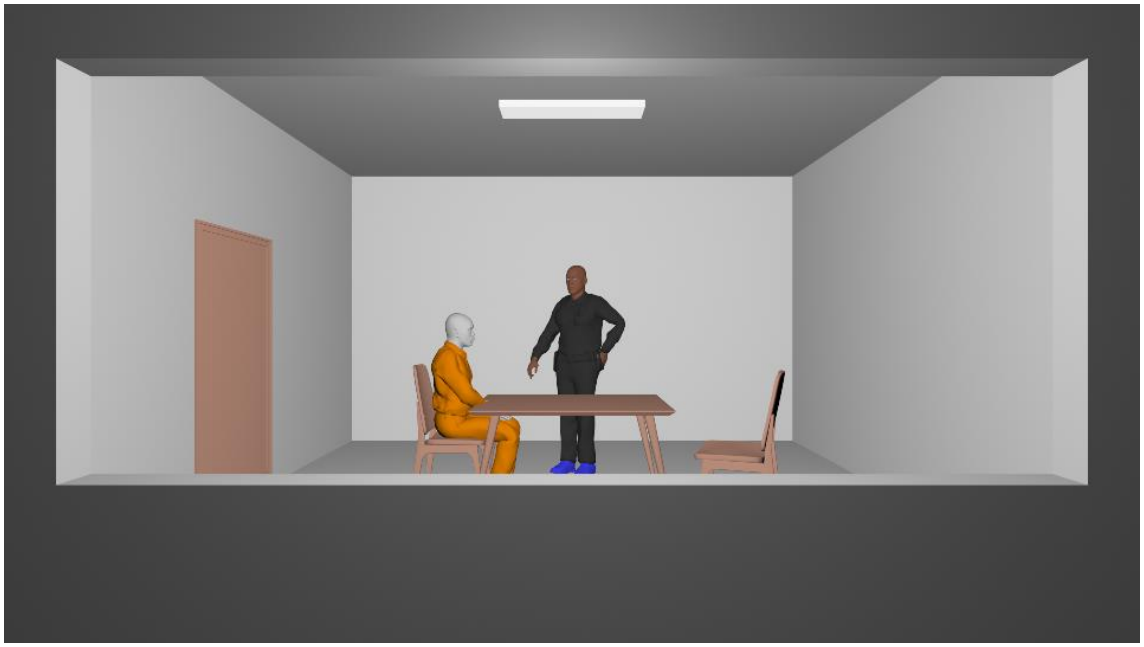

Figure 9. Secret revelation (Scene 4).

At some point, things get serious and relevant information is revealed by the suspect.

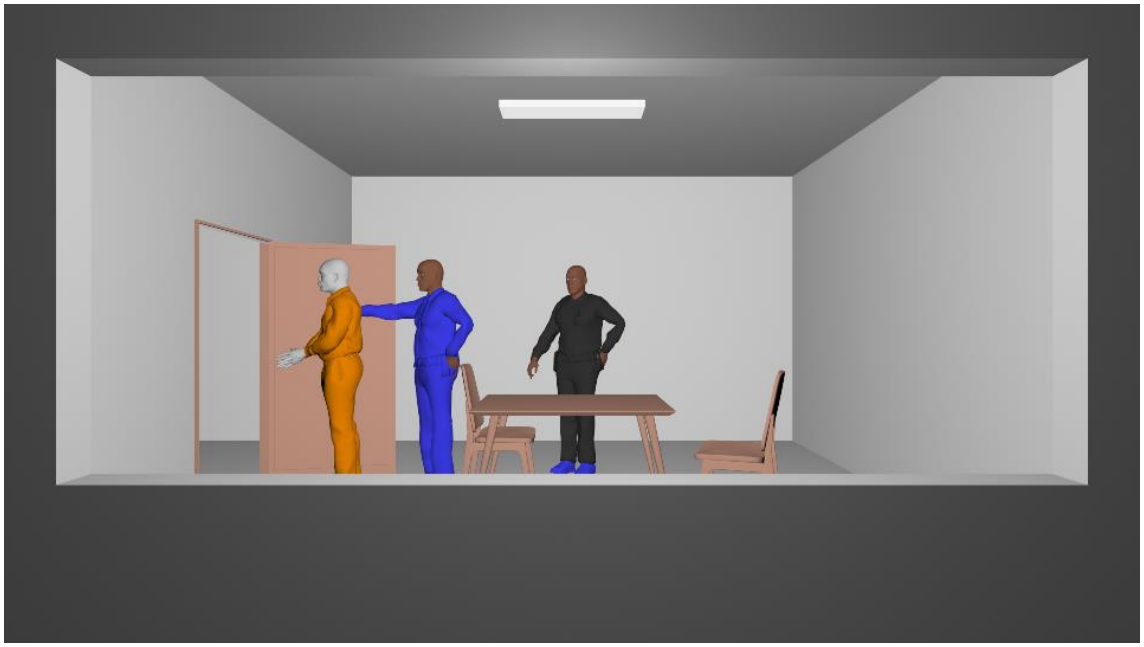

Figure 10. Interrogatory ends (scene 5).

For denouement, the interrogatory ends, leaving the officer alone in the room and making the final conclusions looking at us through the window.

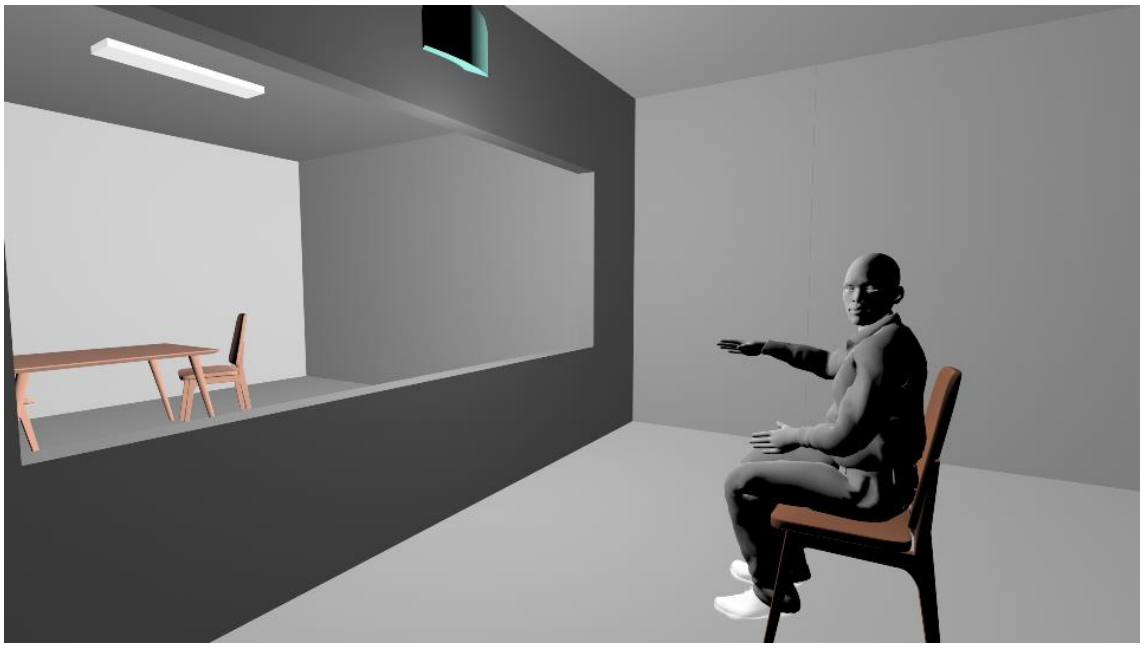

Figure 11. User's discussion (scene 6).

To finish, participants have a conversation about their impressions on the scenes they have just experienced, leaving them some time for interaction.

3.1.1.3. Pre-Production

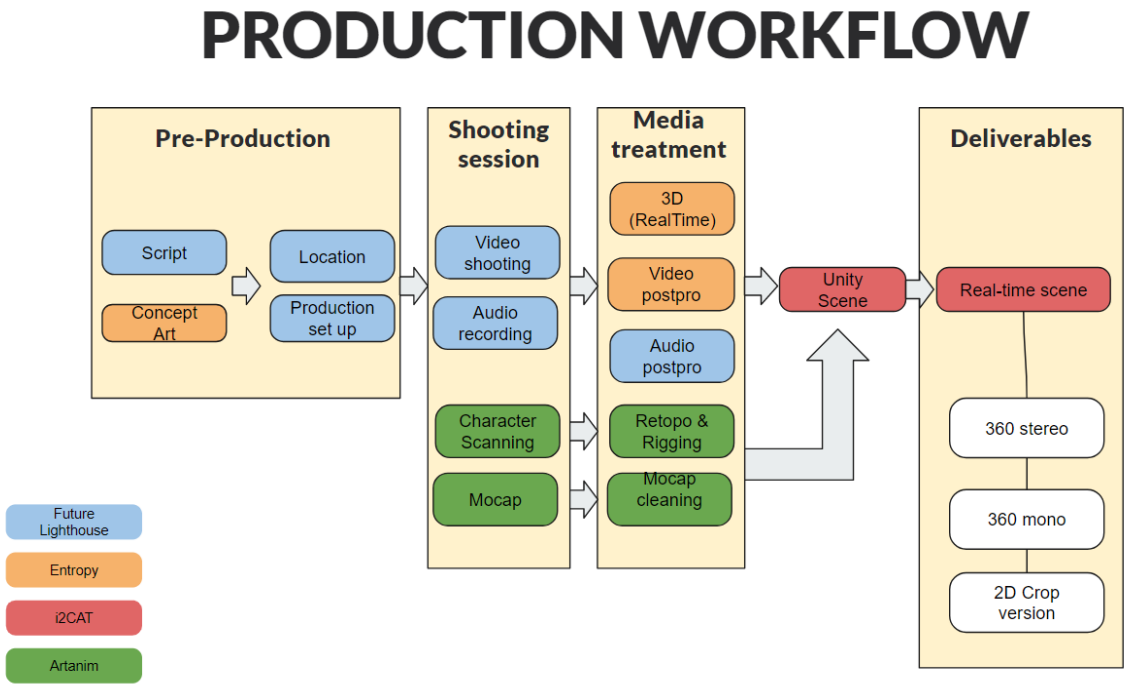

Figure 12. Production workflow.

This diagram describes the process through which the partners may provide media assets for the experiments.

The process right after approval of the pilot will be:

- Creation of the final script
- Concept art images defining the visual aspect of the 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
    - Insurances dealings
- Soundtrack and music
    - We'll try to find music with free use and distribution model

3.1.1.4. Production

This section provides a technical breakdown graphically supported to allow the reader a better understanding of the production techniques that have been planned for pilot 1. All the depicted and described actions will take place in the framework of WP4.

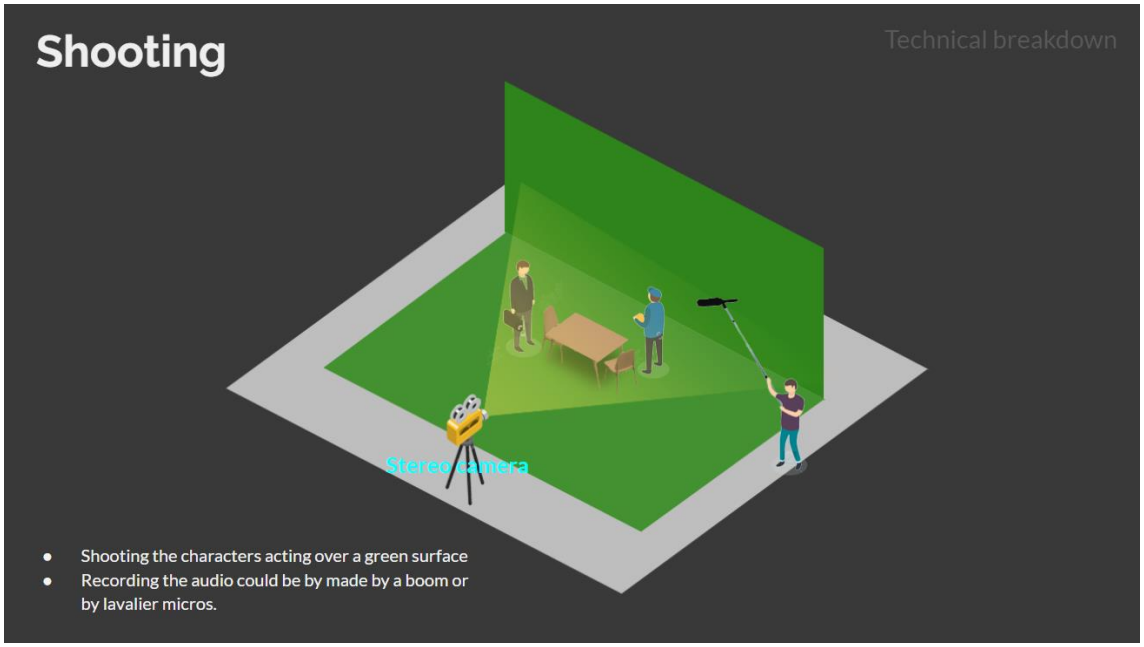

Figure 13. Stereoscopic shooting of character action.

Action for this first pilot should be recorded with a stereo rig of two cameras, separated by 67mm, which simulates the distance between human eyes (standard).

This will take place in a chroma environment, allowing us to remove the background and place the action wherever we want.

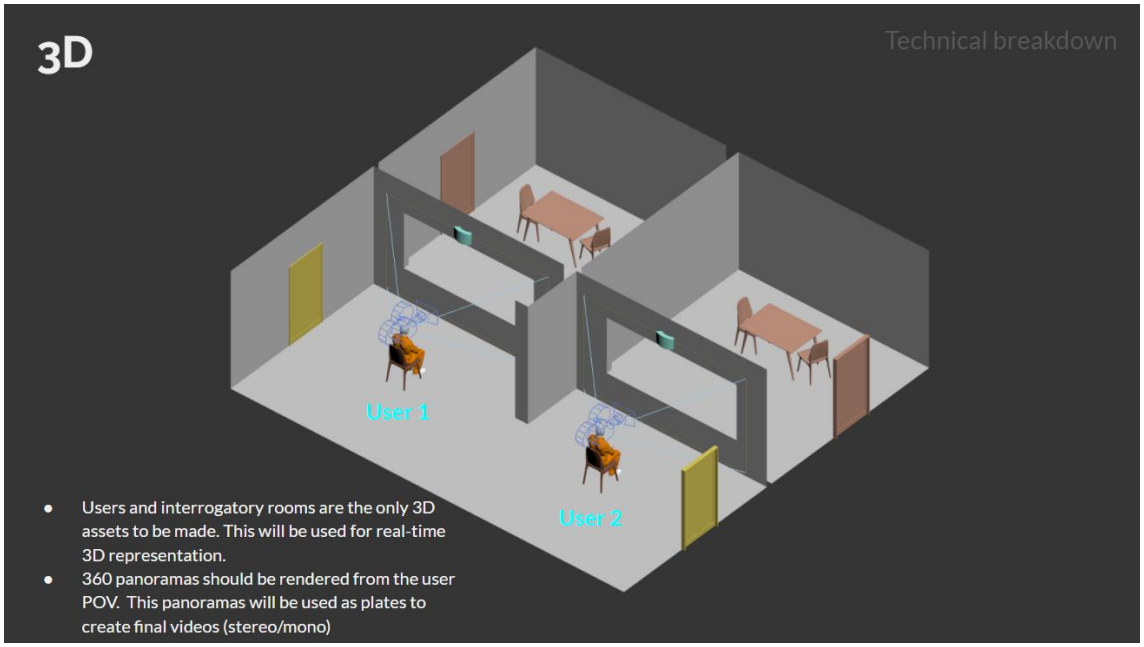

Figure 14. 3D Scene where action takes place.

Then we will model a room, simulating the Police office, where each user could watch the action taking place in the other side of the window. This environment will be used in a 3D real-time Unity scene, allowing us to move with 6 DOF (Degrees Of Freedom).

Users will be rendered with Point Clouds or TVMs in real-time, which gives them the ability to see each other and communicate via gestures and voice.

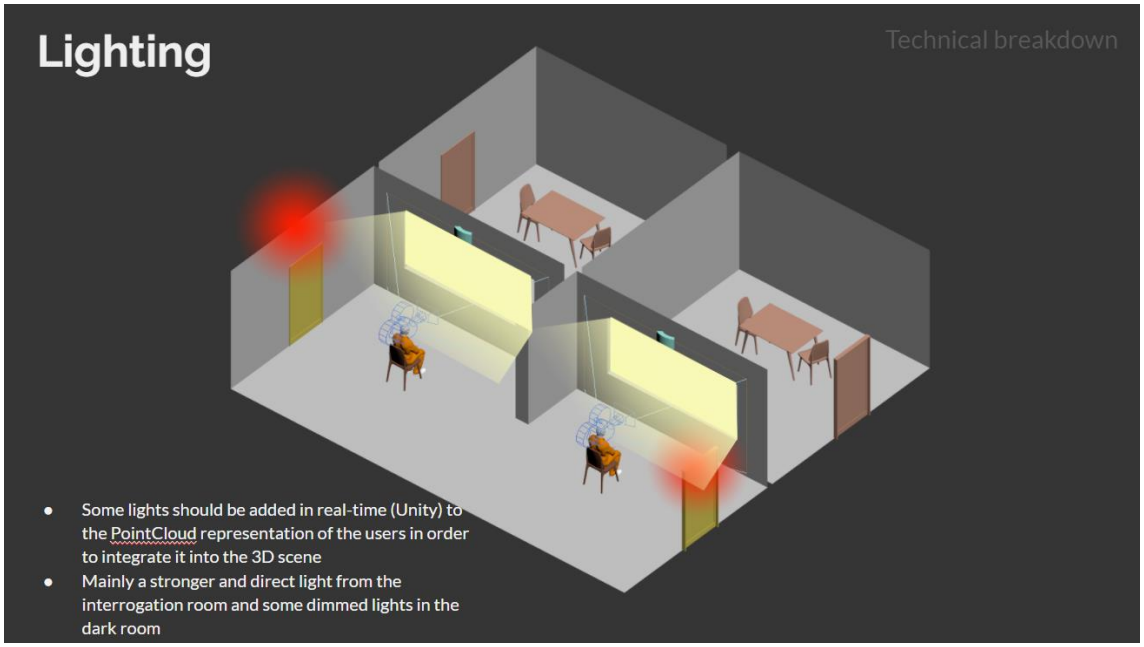

Figure 15. Coherent lighting. Users and scene.

To achieve a good level of visual credibility and integration between the users and the environment, light conditions of the scene has to be simulated inside the player scene to visually match users and rooms.

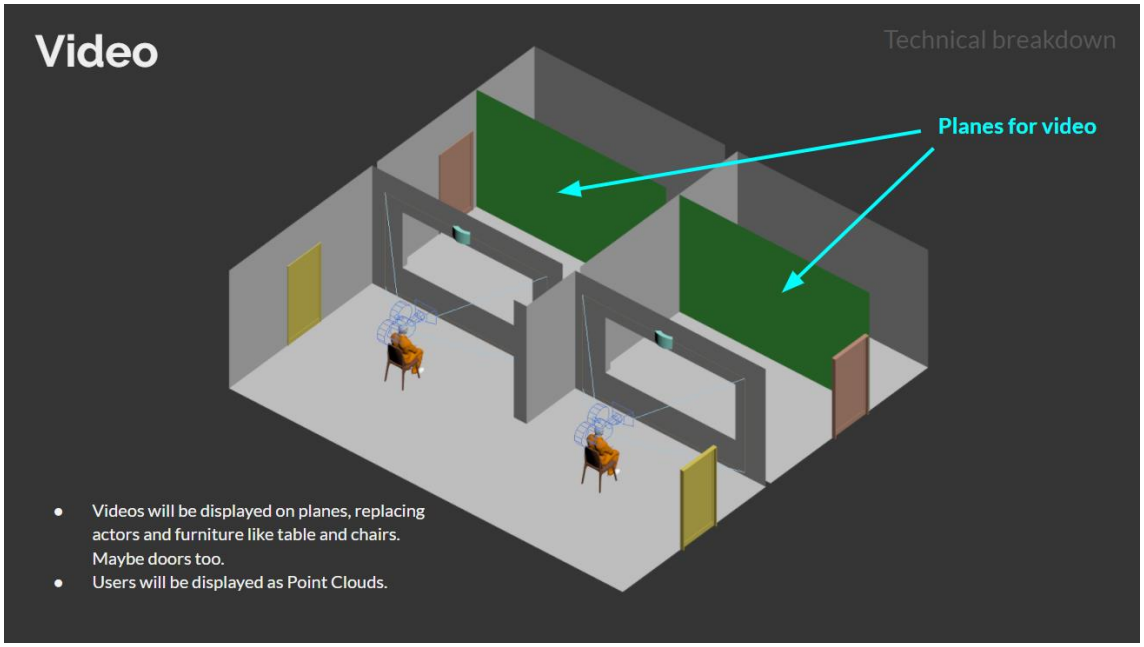

Figure 16. Scene composition (3D Billboards + 3D scene).

The video previously shooted with the stereo rig will be played in a geometric plane inside the Unity scene. This video would have stereo format, this means putting together the frames from the left camera and the right camera, filling the frame from the left camera the upper part of the new video and the frames from the right camera the lower part.

This is the Top/Bottom format.

Then, this video will be rendered from this scene to generate the video (stereo/mono) version of the experience.

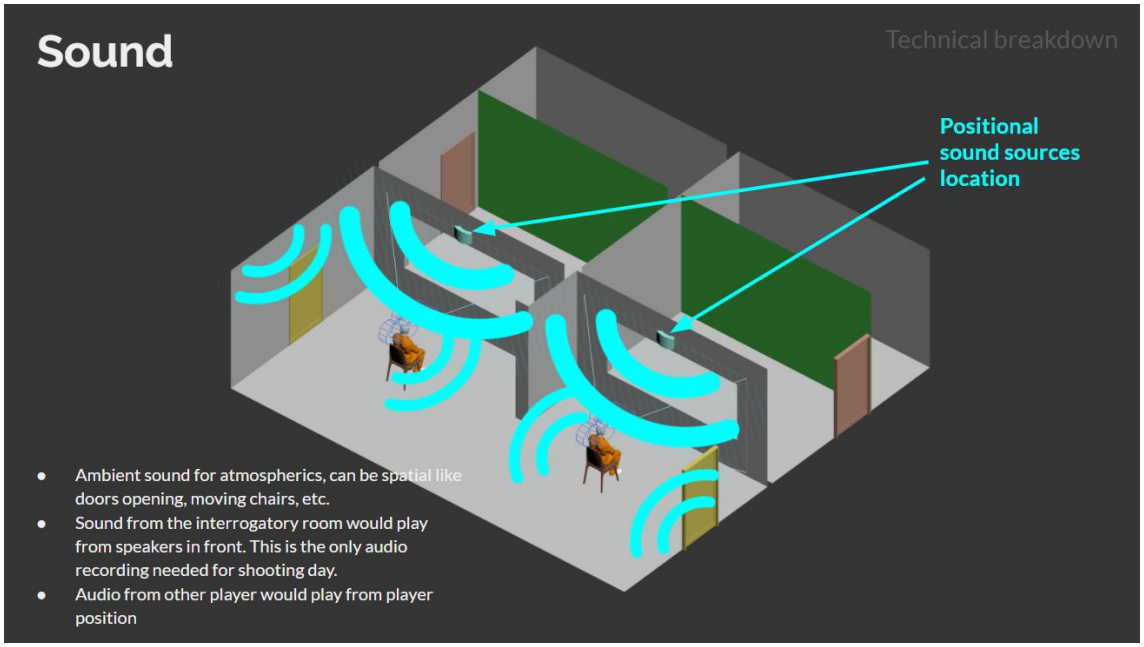

Figure 17. Sound design

The sound of the recorded scene and some extra sounds will be placed in the Unity scene as objects, giving us the sensation of spatiality, despite have being recorded with traditional stereo mics.

This same sounds should be added to the video version.

For experimentation purposes, the consortium decided to create a different version with scanned characters as well, with the purpose of comparing streaming and psychological differences between different media formats.

This action will be made with a photogrammetric rig of 96 cameras, which produce a geometric representation of the characters.

Figure 18. 3D character capture.

These meshes will be reduced in size and shape to meet the requirements of a real-time production and then rigged for adaptation to motion capture process.

Figure 19. Motion tracking to animate pre-rigged characters

Once we have recorded the scene in video, the actors will replay action with motion capture suits in order to translate their movements to the virtual ones.

From all of the process previously described, we will obtain different media formats for the experiments and dissemination if wanted:

- Unity scene with 3D environment and 3D characters
- Unity scene with 3D environment and video billboard representing the characters
- 360 stereo video version of the action
- 360 mono video version of the action
- Traditional 2D cropped version of the action (it has been considered although it is still under consideration, since comparing traditional and vr content is not one of the aims of the project. This would be made from the 360 mono version, taking only a canvas of 1920x1080 pixels. Due to the absence of a user moving the camera, this has to be an edited version.

3.2. Scenario Pilot 2

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

3.3. Scenario Pilot 3

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

4. SOFTWARE REQUIREMENTS SPECIFICATION

4.1. Platform Scenario 1

4.1.1. Use cases

This section provides the use cases view of the VR-Together platform for pilot 1. In later stages of project execution, mainly design and final implementation activities, these use cases may vary slightly, although they provide an accurate view of what is expected to be done and how the end user will interact with the system.

4.1.1.1. UC1-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 different locations.

Precondition: A shared room has been created with a specific configuration by the administrator. Pilot 1 content has been produced and is available for consumption in the shared room. Two users are located at separate VR-Together user capture and consumption systems, at different locations.

Post-condition: Each user has accessed the VR-Together application in an HMD, consumed Pilot 1 content and has interacted with the other user.

Step by step description:

1. Each user accesses the VR-Together application, views the available rooms, selects a room to join and is ready to consume content.
2. User 1 is joining the room and sees the room, but with the interactive content at pause (alternatively with the addition of a waiting for user 2 indication)
3. User 2 is joining the room and sees the room, but with the interactive content at pause
4. The administrator is starting the immersive experience
5. An image is shown in front of the users (image plane) with the text “look here to start the experience”, after gazing at this surface for several seconds the experience is started
6. Both users enjoy the interactive and synchronized experience (see UC-1.1)
7. In the time of the experience users can still see/hear each other and communicate (see UC1.2)
8. After the experience finishes, any interactive media content is stopped and users can communicate until closing the application

4.1.1.2. UC-1.1 Content consumption

Brief Description: This use case describes how a user consumes content with the VR-Together application.

Actors: 1 end-user.

Pre-condition: The user has accessed the application and selected a room.

Post-condition: The user has experienced the immersive content.

Step by step description:

1- User watches a mono 360 video, turns the head in any direction and the content image reacts accordingly. The audio also reacts accordingly.

2- User translates the head slightly and content does not reacts accordingly. (Alternative: Pause Menu)

3- User arrives to the end of the content, menu appears, user press exit and goes back to the list of rooms.

Alternative courses:

UC 1.1 - A.C.1 Pause Menu

1- Clicking anywhere the content pauses and the image/audio react accordingly, paused, content paused (visible or not), lights go a bit down and menu appears.

2- The user can change the scene format. *(UC 1.1 - A.C. 1.1 Change Scene Format)*

3- The user can change the end user format. *(UC 1.1 - A.C. 1.2 Change User Format)*

4- The user can also seek to a specific moment of the experience.

5- User's communication is stopped or alternatively is kept while content paused.

6- User press resume, menu disappears, lights go up, and content continues reproduction according actions taken if any.

UC 1.1 – A.C.1.1 Change Scene Format

1- The user sees a list of available scene formats.

2- The user selects the scene format. Different options are offered: Mono 360, Stereo 360, Stereo 360 + stereo billboards, 3D Room + stereo billboards, 3D Room + 3D Characters.

3- User press resume, menu disappears, lights go up, and content reproduces.

4- User rotates head and content acts accordingly (all options).

5- User translates head and content acts accordingly (only in case of formats 3D+stereo, where some parallax is seen, and 3D+3D where scene and actors are well combined in the same 3D scene).

6- User translates full body and content acts accordingly (only in case of format 3D+3D).

UC 1.1 – A.C.1.2 Change User Format

1- The user sees a list of available user representation formats.

2- The user selects the user representation format. Different options are offered for both users: 2D, Point clouds, TVMs.

3- User press resume, menu disappears, lights go up, and content reproduces.

4- A user can see the other user (friend) represented in the selected format (2D, PC and TVMs) consistently blended/placed in the surrounding immersive environment.

5- Users watches their own hands and body and they can see them represented in the selected format without minor delay between real movement and digital representation.

6- Users rotate and translate their head and the friend representation reacts accordingly

7- Users can see themselves and their friend well blended with the reproducing scene (in case of real time captured PCs and TVMs, colours are modified to simulate the lighting conditions of the scene).

4.1.1.3. UC-1.2 Social interaction

Brief Description: This use case describes how two users can communicate and interact in the VR-Together application.

Actors: Two users, in different locations.

Pre-condition: 2 users have accessed the application, selected a specific room, and content has started.

Post-condition: The user has interacted with the other person in the virtual room

Step by step description:

1- Both users appear one next to each other (specific positions have been previously set in the room).

2- Both users see each other represented according the room configuration or the user selection (depending on the user representation, different details can be seen. See UC 1.1 – A.C.1.2).

3- Both users can talk naturally, delay in the communication is imperceptible.

4- Friend's voice is well positioned in the space, that is, in the same position where the other user is located.

5- Users can see each other facial expressions (i.e. HMD has been digitally removed) and voice and lips are synchronized.

6- Both users watch the content and if both point an agreed element in the scene, both pointing gestures are coherent.

4.1.2. Product perspective

The system to be designed for pilot 1 must enable two end users located in remote/distributed physical rooms, equipped with commercial HW for body capture and VR headsets and a commercial network connection, to enter into a virtual space or virtual world where a short scene of VR content can be visualized. Inside the virtual environment, end users see each other and themselves, and they can naturally communicate (i.e. talk and look to each other as if they were communicating in a common physical location next to each other). The content represented in the virtual environment will be generated from a blend of media formats (i.e. video directive and 360 and mono and stereo, point clouds and 3D meshes) and the end user representations will also be based in these different media formats. Audio will be immersive, that is, audio be coherently positioned according users and their position inside the visual content being rendered and displayed around them. As in any audio-visual experience, visual and audio contents will be synchronized for both VR scene and end-user representations. In addition to that, the display of the content in both ends of the system should happen also in a synchronized manner.

4.1.3. Product functions

This section summarizes the major functions the VR-Together platform must perform or must let the user perform. The platform functions are described in a non-technical way, understandable by any reader. In future sections, these functions are analysed and decomposed in different specific requirements.

4.1.3.1. Capture and reconstruction components

People 3D Capture and reconstruction. The system, using a multi-RGBD sensor setup, will be able to capture in real time color and depth data from multiple points of view around the user. It will be developed to provide in real-time information including group frame index, color data, depth data and proper timestamps. It will also be able to extract the foreground information of the user in real-time, allowing for optimized handling of the collected data. Moreover, it will extract the coloured 3D point cloud of the user body in real-time, allowing for point-cloud based 3D reconstruction, which the main objective of the system. In particular, the 3D point cloud extraction will be achieved by projecting the RGB-D data from each view using the intrinsic and extrinsic parameters of the cameras. Thus, the user body along time will be reconstructed by the extraction of the user's 3D geometry and appearance on a per-frame basis, i.e. for each time instance. Therefore, given multiple captured depth-maps at a specific time instance, along with the corresponding RGB images, the objective will be the fast 3D reconstruction in the form of a single textured triangular mesh.

CODE	NUM	TITLE	DESCRIPTION
CRR	1	People RGB-D Capture	The people capture component/system will capture RGB-D data from 4 RGB-D devices connected to 4 capturing nodes (RGB-D nodes). The capture rate will be at least at 25 fps.
CRR	2	People RGB-D Calibration	The RGB-D devices of the system will be automatically calibrated (extrinsic calibration).
CRR	3	People RGB-D Synchronization	The RGB-D frames from the RGB-D nodes will be synchronized and grouped in a central node, resulting in a RGB-D group frame.
CRR	4	People live 3D reconstruction	The people live 3D reconstruction component will process user's coloured 3D point cloud to reconstruct a 3D time-varying mesh in real-time (under 80ms per frame).
CRR	5	People live 3D reconstruction visual quality	The people live 3D reconstruction method will be extended, resulting in higher visual quality level than the initial version.

Video background removal component. The system will use a single RGBD sensor to capture each user in real time, both color and depth data. The sensor will capture the user from the direction of the virtual other user, thus providing a viewpoint for the other user. Based on the depth data, the foreground will be segmented from the background, and the user-only color video will be provided.

CODE	NUM	TITLE	DESCRIPTION
CRR	6	static background removal	Users must be standing or sitting. The FGBG module requires a static placement of the Kinect V2 sensor with a background and foreground object within 5 meters distance. After a replacement of the sensor, a new background image must be captured.
CRR	7	Image properties for background removal	Input image properties should be 960x540 pixels at a framerate of 25 fps; the FGBG then drops approx. 1 frame per second on a low (i.e. below 30%) CPU load.

HMD removal component. In an initialization phase, the system will use a single RGBD sensor to capture the face of each user from multiple sides without the HMD on. The color and depth information of the face is combined into a face model. With a HMD on, the HMD is detected in the RGBD stream and replaced with the correct viewpoint of the face model. The upgraded RGBD stream will be provided. Also the current viewpoint of the user is provided to the VRT platform to enable avataring in VR.

CODE	NUM	TITLE	DESCRIPTION
CRR	8	Face capture	The user's face needs to be captured from at least two different poses.
CRR	9	Captured face storage	The captured user face needs to be stored (on disk or in memory) and must be accessible in real-time by the face inpainting algorithm.
CRR	10	Face inpainting	The HMD removal process through face inpainting must work offline, i.e. non real time. The process may work in real time.

4.1.3.2. Encoding and Distribution

Point Cloud encoding and distribution. The point cloud compression system offers three codecs for the compression of dynamic sequences of point clouds. The codecs included feature lossy colour attribute coding using JPEG compression, inter prediction to exploit temporal redundancies, progressive decoding and a parallelized implementation.

The transmission system leverages the state-of-the art technology to deliver point clouds at the best possible quality to all users depending on their capabilities. The system is able to adapt depending on the available device, network bandwidth or any additional constraint. The system may or may not have a limitation on the number of clients. Motion Spell leverages its Signals platform to offer a all-in-one product for muxing, signalling, and transport. This product allow to experience video teleportation of 3D objects with additional media synchronized for astounding social experiences.

CODE	NUM	TITLE	DESCRIPTION
EDR	1	Real time compression	The framework has low delay encoding and decoding of less than 200 ms
EDR	2	Progressive decoding	The framework allows a lower quality point cloud to be decoded from a partial bitstream
EDR	3	Efficient compression	The framework can achieve a compression ratio of up to 1:10 to in order to stream point clouds
EDR	4	Low end to end latency	The framework has an end to end latency below 300ms similar to video conferencing requirements
EDR	5	No a priori information	No geometric properties of the objects are assumed
EDR	6	Generic compression framework	The framework can be used to compress point clouds of arbitrary topology (independent of capture device, point precision, file format etc.)
EDR	7	Quality assesment	The framework will make use of quality metrics (to be decided) which informs about the expected quality of experience from the user perspective

End-user real-time mesh encoding and distribution. The system will be able to efficiently compress user 3D meshes, i.e., 3D geometry and RGB textures, in real time. Furthermore, the system will transmit the compressed data using 3D time-varying mesh live streaming techniques from a capture location to a display component. Thus, the main objective is to achieve real-time time-varying mesh encoding and transmission, which is challenging and not standardized, over commercial communication and media delivery networks. It will also allow local display components to access the reconstructed 3D meshes without delay, also reducing the bandwidth needed for sharing the data between the remote display components.

CODE	NUM	TITLE	DESCRIPTION
EDR	8	Real time compression	The framework has low total delay, encoding and decoding in less than 80 ms per frame

EDR	9	Efficient compression	The framework can achieve textured mesh (3D geometry and textures) compression ratio of up to 1:30
EDR	10	Generic compression framework	The framework can be used to compress textured 3D mesh of arbitrary topology (independent of size, number of surface triangles, etc.)
EDR	11	Real-time time-varying mesh encoding	This component must compress and decompress 3D time-varying meshes and the corresponding RGB textures in real time. (under 80ms per frame)
EDR	12	Real-time time-varying mesh encoding parametrization	The component will allow the selection of different TVM compression and texture resolution and quality level.
EDR	13	Real-time time-varying mesh distribution	The component will transmit 3D time-varying meshes in real time (under 100ms per frame), depending on the network type and quality.
EDR	14	Real-time time-varying mesh distribution parametrization	The component will enable the tweaking of TVM frame time life and the frame queue length.

Traditional audio and video encoding and delivery. Audio is encoded and sent synchronously with video, point cloud and meshes.

CODE	NUM	TITLE	DESCRIPTION
EDR	15	End-user audio encoding	End-user audio encoding should be done with formats supported by the browsers in which the web player run, at bitrates comparable to state-of-the-art video communication applications.
EDR	16	End-user video encoding	End-user video encoding should be done with formats supported by the browsers in which the web player run, at bitrates comparable to state-of-the-art video communication applications.
EDR	17	Content audio encoding	Content audio encoding should be done with formats supported by the browsers in which the web player run, at bitrates comparable to state-of-the-art video communication applications.
EDR	18	Content audio encoding	Content video encoding should be done with formats supported by the browsers in which the web player run, at bitrates comparable to state-of-the-art video communication applications.
EDR	19	End-user audio distribution	End-user audio distribution should be done with formats

			supported by the browsers in which the web player run, at bitrates comparable to state-of-the-art audio communication applications.
EDR	20	End-user video distribution	End-user video distribution should be done with formats supported by the browsers in which the web player run, at bitrates comparable to state-of-the-art video communication applications and similar latency requirements.
EDR	21	Content audio distribution	Content audio distribution should be done with formats supported by the browsers in which the web player run, at bitrates comparable to state-of-the-art audio communication applications and similar latency requirements.
EDR	22	Content video distribution	Content video distribution should be done with formats supported by the browsers in which the web player run, at bitrates comparable to state-of-the-art video communication applications.

4.1.3.3. Orchestration

The system must be able to organize and orchestrate all the different media sources that compose a VR-Together experience (content representations, user representations) as well as manage and provide specific details of a specific experience "session" (for instance, session and content discovery, room description and time of reproduction, position of and timeslots for end users). The orchestration of the system may also include optimizations in order to reduce the computing requirements at players side (for instance, mixing videos that can be delivered as a single video to a player, blending user representations in order to avoid different parallel decoding process at player side, etc.). Also, the orchestration should manage session synchronization.

CODE	NUM	TITLE	DESCRIPTION
OR	1	Configuration	The orchestration modules must support remote configuration through an administrative interface.
OR	2	Content discovery	The orchestration modules must be able to facilitate discovery of both content streams and files, and at least two end-user representation streams.
OR	3	Session management	The orchestration modules must manage sessions and facilitate the discovery, setup, control and synchronization of a session between a least two end-users.
OR	4	Session management	The orchestration modules should support at least two parallel sessions.

OR	5	Database storage and access	The orchestration modules should store and access data in a shared database. This database should contain the current time in the room, the number and state of users in the room, the state of the content being consumed, and the state of the shared environment.
OR	6	Optimization	The orchestration modules must be able to reduce the requirements on the end-user players by processing at least two end user representations coming from the 3D reconstruction system.

4.1.3.4. Rendering and display

The system allows playback of the orchestrated scene (content plus end user representations) on both native and web application platforms.

Web player. The system will allow end users to access the VR-Together experience through a web browser. This web browser will allow (a) playback of video content in a VR environment composed by WebVR enabled clients, (b) A/V communication with another user in the VR space blended in the VR environment, as if both end users are sitting next to each other, and (c) spatially aligned audio rendering.

CODE	NUM	TITLE	DESCRIPTION
WPR	1	Content	The web player must support playback of the content as 2D video.
WPR	2	End user representation	The web player must support playback of end users represented as 2D video; the video format of end user representations should be at least 960x540 pixels at a framerate of 25 fps.
WPR	3	Audio	The web player should support spatial audio, to align the content audio with the end user audio.
WPR	4	Communication	The end user representation audio and video should be played back by the web player in real time. This means played back within a latency common for real time communication (capture to display delay of under 400ms).
WPR	5	Streaming	The web player must receive the streaming of both content and end user representations streams.
WPR	6	WebVR	The web player must operate in a browser that supports WebVR and A-frame.
WPR	7	Bandwidth	The web player should support content bandwidth

			adaptation. The web player must rely on the mechanisms supported by the browser.
WPR	8	Latency	The latency between different streams should not be higher than 500 ms.
WPR	9	Scene	The scene must be a static 2D 360 image of at most 4K pixels, in ERP format; Other scene formats (like 2D 360-degree ERP video) may be supported, depending on the performance restrictions of the PC and browser.
WPR	10	End user placement	The web player must allow for manual configuration of the end user location; the web player should allow for end user placement control by the orchestration modules.

Native player. The system will allow end users to access and share the VR-Together experience through a native application. This application will allow the user to (a) access the VR-Together contents, (b) synchronize playback of VR content across different native players, (c) render different media formats in parallel (video 360/mono/stereo, 3D mesh, point clouds) creating an hybrid media representation, (d) render self representation and remote user representation in different media formats (3D mesh, point clouds), (e) render immersive audio.

CODE	NUM	TITLE	DESCRIPTION
NPR	1	Multiple format	The native player supports consuming contents in different VR formats, like Point Clouds, omnidirectional video, static meshes, dynamic meshes, mono/stereo 2d video.
NPR	2	Hybrid format	The player reproduces hybrid VR contents in the same scene, that is, compositions made of blending different Media formats
NPR	3	Audio	The player reproduces immersive audio tracks, the point of view of the user affects the perceived sound.
NPR	4	Rendering 1	The player is able to reproduce smoothly relevant combinations of available formats
NPR	5	Rendering 2	The player is able to render at least some of the combined media formats at 60 fps or more
NPR	6	Rendering 3	To better integrate different media formats and sources, the player can alter the lighting of specific objects within the scene, on the basis of custom shaders.
NPR	7	DoF	The client can consume content adapted to 3DoF or 3DoF+ movements (i.e., head position and rotation). The

			media player integrates within the rendering pipeline user movements (3DoF+)
NPR	8	Quality of Image	input/output effective display resolution will support up to 4K
NPR	9	Delay on displaying self representation	Self representation impose a latency constraint under 20ms.

Synchronization. The system should further make sure that playback of the content that the users are watching in a room is sufficiently synchronized across the devices, such that the users do see and hear the same content at the same time. This system will make sure that the playback of voice is sufficiently aligned with the rendered visual representation of the user such that lip-sync is established.

CODE	NUM	TITLE	DESCRIPTION
PSR	1	Sync audio and video (lips-mouth)	The delay between audio and video has to be acceptable in the range of +90 ms (audio first) to –185 ms (video first).
PSR	2	Sync multiple formats	The player must support synchronization between different formats with less than 40ms of delay), but the acceptability, would be under 200ms delay.
PSR	2	Sync inter device	The synchronization between different players should be frame accurate (less than 20ms of delay), but the acceptability, would be under 100ms delay.
PSR	3	Sync control	Players should be able report their playback status and be able to adjust playback to resynchronize with other players.
PSR	4	Timestamping	Media streams must be timestamped at the source in relation to a common timeline, such that the player is able to establish synchronized playback of the audio and visual components.

4.1.4. User characteristics

There are two types of users that interact with the system: 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 (experimenter), and administrators. Each of these three types of users has different use of the system so each of them has their own requirements.

4.1.4.1. Content consumer

The end user of the VR-Together platform is the content consumer, that is, a person of any age, genre and condition, without any hear or visual impairment and without any previous known problem while accessing contents using Head-Mounted Displays. Content consumers 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 platform.

4.1.4.2. 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 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 an orchestration admin interface, although at the moment of writing the present text other configuration capabilities are also being considered: the calibration of a specific instance of a capture system, or even a specific instance of a player.

4.1.4.3. Experimenter

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

4.1.5. 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/]

4.1.6. Assumptions and dependencies

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

4.1.7. 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 easy to deploy, aiming at a social experience without exigent requirements in terms of equipment.

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

Admin interface for room configuration. The VR-Together system will be configurable by means of rooms. Here an admin can define number of users, user drop points, content sources and other parameters related to the conditions in which the content consumption and social interaction happen.

Admin interface for capture calibration. To enable initial calibration of capture system and to modify capture parameters.

Admin interface for experiment running. To select a room and start the experience for different players as well as to modify room or player parameters.

4.2. Platform Scenario 2

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

4.3. Platform Scenario 3

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

5. ARCHITECTURE

5.1. System architecture for pilot 1

5.1.1. Software architecture

This section gives an initial architecture description of the different software components in the system and how these components are orchestrated.

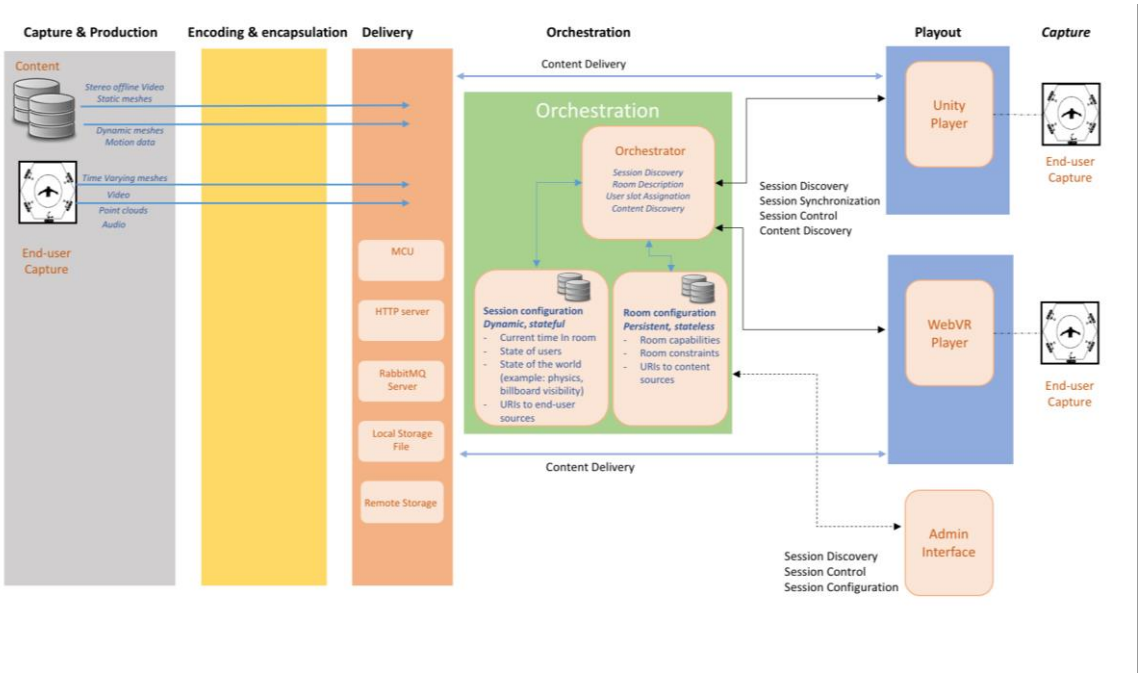

Figure 20. Component diagram for platform v1.

Figure 20 shows the initial high-level architecture of the VRTogether system, with a focus on the components integral for the first Pilot. The VRTogether system is described in a traditional production to consumption chain: audio-visual information flows from production (on the left) towards consumption (on the right). For the sake of clarification, the End-User capture is also shown at the Playout, to reflect that the End-user is also part of the Capture of content.

In this architecture one of the central components is the Orchestrator which provides all clients with the information necessary to start VRTogether Experiences. This includes the discovery of available sessions, VR room configurations, pointers to content sources, other clients in a session and the capture sources. The Orchestrator is responsible to signal synchronization data for the different streams consumed by the clients. Besides the synchronisation data other session control data is signalled via the Orchestrator, like content changes, pause/play, VR room configurations, etc.

The Orchestrator is connected to two databases:

- A Room configuration database. This database stores information related to a room that is Persistent, e.g does not change during a VRTogether experience: Room descriptions, Room capabilities and constraints, and pointers (URIs) to the Content sources.
- A Session configuration database. This database 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 capture streams.

Via an Admin Interface an administrator can control and force such session control data, for example to facilitate user experiments and demos. Important to note is that the Orchestrator controls but does not process media streams. For all control data all clients (regardless of 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 URIs to this content is provided in the room description. In addition to media content each client receives streams from other clients for audio/visual communication, and transmits streams for other

users as well. Each client is responsible for it's capture device(s) and notifying (via the Orchestrator) where and how they can retrieve streams (regardless of protocol or format). For such stream transmission, from client to client, a processing node may be used , such as a RabbitMQ server, a HTTP Server, or MCU as shown in the diagram, or a STUN/TURN server or media converter (not shown).

5.1.2. Hardware architecture

This section provides the initial hardware architecture required to deploy pilot 1.

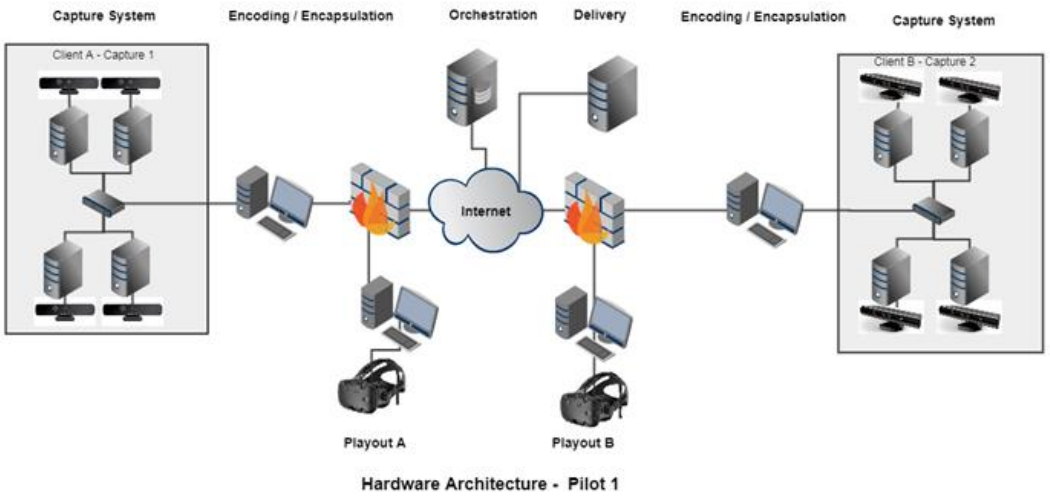

Figure 21. Hardware architecture.

Figure 21 lays out the hardware infrastructure for pilot 1. It will involve two capture rigs, combining 4RGBD cameras 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.

5.2. System architecture for pilot 2

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

5.3. System architecture for pilot 3

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

6. USER LAB

VRTogether user lab activities include a set of known methodologies to gather requirements about the platform and the use cases. The principle is to follow a user-centric approach, in which the right user groups are consulted in order to evaluate the platform, obtain relevant new requirements, which then will lead to further designs and implementations. In particular, the following user groups are considered:

- **Stakeholders**: they will help the project to identify adequate business models and exploitation opportunities. We consult stakeholders in public project events (fairs, conferences, congresses) and specific stakeholder workshops. The project also counts on an advisory board composed by relevant professionals in the field of virtual reality and immersive media. The advisory board includes two types of professionals, fulfilling the needs by the project: technical and artistic.
- **Experts**: they will help the project to gather requirements about the pilots, in order to demonstrate the novelties introduced by the project and about the technology support (e.g., architecture, performance) for making the pilots work. We consult them internally within the companies forming the consortium and externally at targeted events (fairs, conferences, congresses).
- **End-Users**: we consult end-users of the systems to gather a variety of requirements in terms of functionality, perception and interaction, and aesthetics. This will happen during the trials of the system, as well as through user lab experiments, and via questionnaires and open demos.

The consortium will employ a variety of methodologies for gathering the requirements from large-scale questionnaires to targeted experiments to field trials. Some examples include:

- **Questionnaires** at fairs and congresses: they will provide us large quantity of data. The project plans to run questionnaires at events where our system is showcased in order to increase the relevance of the responses (primarily targeting experts). For example, we have done so during the VR Days in Amsterdam from 24th until 26th of October 2017, using the questionnaire shown in Annex II;
- **Focus groups and interviews**: they will provide us highly quality data, given the pre-selection of the interviewees. The project plans to run focus groups and interviews at special events organised by us such as workshops, showcases and meetings (primarily targeting experts and stakeholders). For example, the experts meeting which happens once a year or the yearly “CWI in Bedrijf” event;
- **Experiments in the user labs**: they will provide us high quality data about specific aspects of the project: technology, co-presence, immersion. The project plans to run a number of experiments at the user labs, as shown in Section 6.3.1, with the intention of informing decisions about the technology and artistic choices (primarily targeting end-users). For example, CWI run an experiment (EXP-CWI-1) in their user lab in November to identify new quality metrics for evaluating point cloud compression;
- **Field trials**: they will provide us large quantity of data both quantitative and qualitative. The data will be captured using questionnaires (e.g., co-presence and social interaction evaluation) and logging (performance and interactivity). The project plans to run three field trials, evaluating different aspects of the system during the project, with the intention of assessing the pilots and the platform (primarily targeting end-users). More information about experiments run in the user hubs and the trials can be found in Section 6.3.2.

To vertebrate project user actions, VR-Together consortium has planned to build a permanent collaborative distributed user lab with the necessary equipment to run as a demonstrator and a fast track to evaluate new developments or integrations in controlled environments. It is expected that once this infrastructure is built, and an initial version of the platform has been deployed, by September2018, Pilot 1 will start a series of more or less periodic experiments and evaluations that should involve a relevant number of end users.

6.1. Advisory Board

The advisory board of VRTogether has the role to advise on IP, scientific direction and on business opportunities. The committee reviews on a regular (yearly) basis the progress made and primarily advises on the business aspects of the IP. Some examples include new academic or technological achievements the consortium should consider, new important trends, societal developments the project should take into account, concrete proposals hoe new business may be generated and how exploitation should be organised from the project results.

In particular, the consortium proposed a number of candidates of areas related to the project, both artistic and technical, from which seven were initially selected and contacted. The current list of the advisory board members include:

- Morgan Bouchet from Orange
  - https://www.linkedin.com/in/morganbouchet/
- Yan Chen from Lens Immersive
  - https://lens-immersive.com/about
- Nils Duval, VR Consultant
  - https://www.linkedin.com/in/nils-duval/
- Wijnand Ijsselstein from TU Eindhoven
  - https://scholar.google.com/citations?user=0JNTD4cAAAAJ
- Dolf Schinkel from KPN
  - https://www.linkedin.com/in/dolfschinkel/
- Sebastian Sylwan from Felix & Paul Studios
  - http://www.imdb.com/name/nm4492489/
- Graham Thomas from the BBC
  - http://www.bbc.co.uk/rd/people/g-a-thomas

The Advisory Board meets twice a year together with the technical coordinator and the work package leaders:

- Technical members in September in Amsterdam during IBC
- Artistic members in a festival to be decided

During the advisory board meeting, the consortium presents the project and updates the board, showcasing the current status and requests for feedback in the form of a focus group. The consortium may as well individually contact members of the board for running structured interviews via phone or Skype twice a year about specific topics.

6.2. User Hubs and User Labs

Through the VR-Together User Lab, the project will run tests and evaluations that will be used for taking decisions about the pilots (artistic side) and about the platform (technical side). In the project, we differentiate between two experimental facilities that integrate the VR-Together User Lab, hubs and labs.

- The first one, a hub, is a node with the full capturing and rendering infrastructure of the VR-Together platform;
- The second one, a lab, is a facility with partial functionality of the VR-Together platform.

The user hubs will provide a full environment to run the field trials of the pilots. They will include the complete media pipeline including capturing and reconstruction, delivery and transmission, and rendering. The expectation is that they will be 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 22 shows the basic infrastructure of a hub, including a capture system (based on Intel RealSense Technology), several PCs for reconstruction, and a rendering infrastructure based on Head Mounted Displays (e.g., Oculus Rift and HTC Vive). It has been decided that VRTogether will have three main hubs strategically located in:

- Amsterdam (CWI premises)
- Barcelona (i2CAT premises)
- Thessaloniki (CERTH premises)

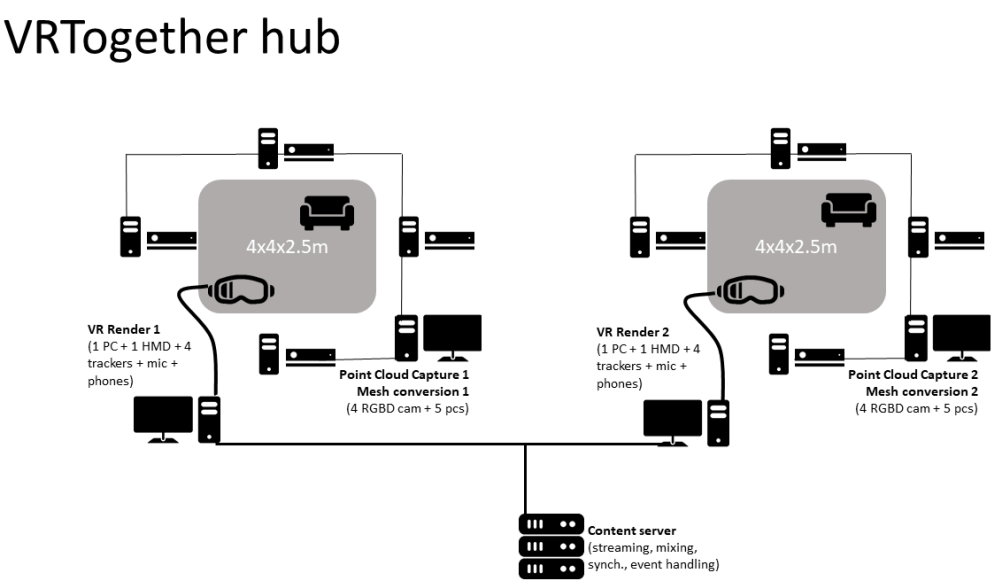

Figure 22. Schematic View of a VRTogether hub.

In addition to the hubs, several partners of the project will create dedicated user labs with a partial infrastructure of the full fledge VRTogether platform. These labs will be 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 agreed to provide a lab, intended for different types of experimentation.

- **Artanim**'s user lab will primarily focus on evaluations on the psychological aspects of the project such as "togetherness", "co-presence", and "flow". Currently planned experiments will 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 VRtogether project for the alternative of representing user and actors with a rigged mesh of triangles.
- **CERTH**'s user lab will primarily focus 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 production rate. Some initial experiments on removal of the HMD of provided 3D reconstructions are as well expected. Such experiments will inform the capturing side of the platform.
- **CWI**'s user lab will primarily focus on Quality of Experience (QoE), which will in turn serve for developing new quality metrics and guidelines for evaluating social VR. Such metrics will be used in the system for optimization purposes and will be used during the

trial. CWI already run some initial experiments about the QoE of point cloud compression in the beginning of the project (October), 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.

- **FLH and Entropy** 's user lab will primarily focus 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 will conduct 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 will make use of both questionnaire and behavioral data, and will inform the use cases and the definition of the trial.
- **TNO**'s user lab will primarily focus 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 artifacts, and experiment with methods to improve the feeling of co-presence through shared interaction.

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

Figure 23. 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 Figures 2.2 and 2.3). 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 24. Artanim's User Lab.

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 2.4) 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).

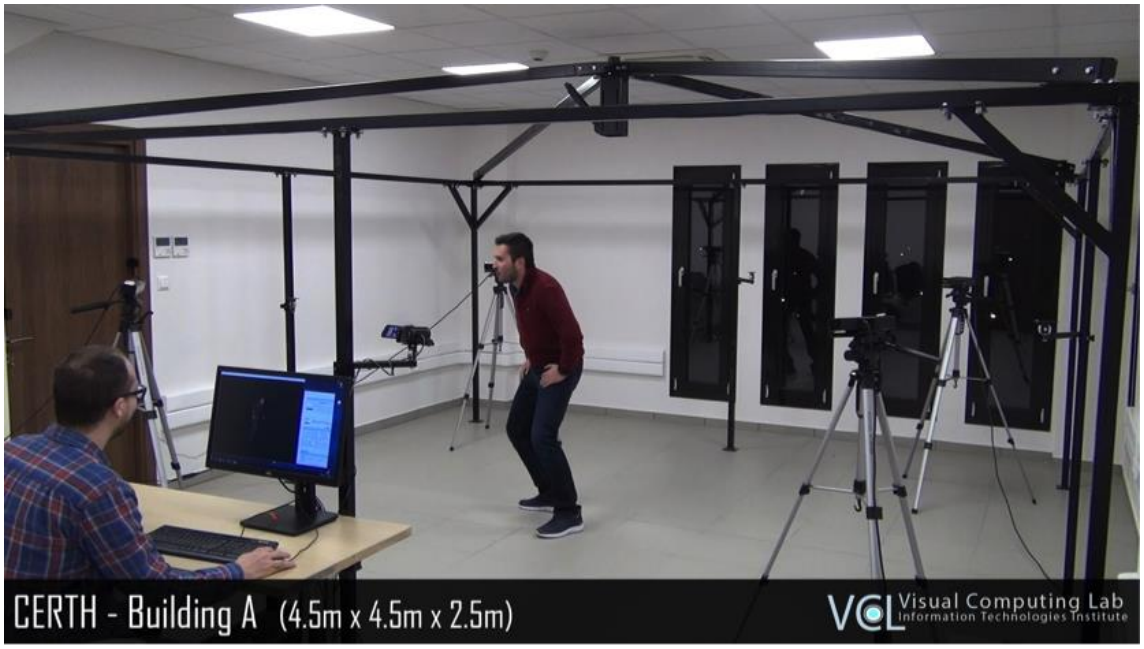

Figure 25. CERTH's User Lab.

CWI has two available rooms: Pampus (see Figure 2.5) and the QoE Lab (see Figure 2.6). 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 26. CWI's User Lab (Pampus)

Figure 27. CWI's User Lab (QoE Lab)

Finally, 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.

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 VRTogether.

6.3. Experiments

The partners of VRTogether will carry out experiments to inform about different aspects of the project: technology, pilots, and evaluations. These experiments will run either in the hubs (full-fledged infrastructure) or in the labs (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: EXP-CERTH-1, EXP-CERTH-2, EXP-CERTH-3, and EXP-CERTH-4. They have a direct influence on the user hubs under development;
- Subjective **quality of experience**[1], mainly based on perception of the medium under different constraints (different compression mechanisms or bandwidth): EXP-CWI-1, EXP-CWI-2, EXP-CWI-3, EXP-i2CAT1, and EXP-i2CAT2;

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

- **Psychological** dimension of the VRTogether experiences, evaluating aspects such as the feeling of being there, as well as the feeling of being together. These include: EXP-Artanim-1, EXP-Artanim-2, EXP-i2CAT-3, and EXP-i2CAT-4

In all VR-Together experiments we will 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.

6.3.1. Initial List of experiments

This sub-section provides an overview of the experiments initially considered in the project, and serves as an initial 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.

Code: EXP-Artanim-1 Title: Impact of movement animation of the virtual body parts to presence Type: Psychology Date: April 2018 (scheduled) Number of users: 24	
Description	Experiment to assess the relative impact of different levels of movement animation fidelity of the virtual body parts to presence (place and plausibility illusion). The following virtual body parts will be manipulated: face (eyes/gaze and mouth); fingers; avatar legs; and inverse kinematics driven joints. These body parts could be set to three different animation conditions: no animation; procedural animation; and motion capture. HTC tracker will be used for head, pelvis, feet and hands, manusVR of leap motion for fingers, facial features tracking has to be defined.
Objective and expected results	We want to confront the relative importance of animation features with the costs of adoption (monetary and effort) to provide software and use guidelines for live 3D rigged character mesh animation based on affordable hardware. This outcome is necessary for real-time animation of rigged character meshes by actors/users, which will have different uses in the project: baseline for comparison and validation of the less orthodox mediums for visual representation of the user (i.e. point clouds and real-time mesh reconstruction); live animation of actors that will be used in pilot 2.
Methodology	Users start with the simplest setting, and improve features of the tracking until they are satisfied with the experience. Using the order in which users improve the movement features we can assert on the most valuable animation features to the users, and statistically model these transitions in a Markov chain.

Data set description	Preferred improvement path per subject, i.e. what was the order with which each user improved animation features to transition from the least realistic to the most realistic animation scenario. Motion data, position and rotation of the motion tracking sensors and of the user-controlled avatar joints.

Code: EXP-Artanim-2 Title: Impact of movement animation of the virtual body parts to co-presence Type: Psychology Date: June 2018 (scheduled) Number of users: 24	
Description	Experiment to assess the relative impact of different levels of movement animation fidelity of the virtual body parts to co-presence (the subjective experience of being in the presence of other users). Will follow a similar setup and protocol to EXP-Artanim-1, but users will participate in pairs. Each user will be able to control animation features of the other user's avatar instead of features of their own avatar.
Objective and expected results	Improve the guidelines obtained in experiment EXP-Artanim-1, considering a shared VR experience. Software implementation for two networked users that will be used in the joint experiment of i2cat and Artanim during the second half of 2018 (see EXP-i2CAT-3 and EXP-i2CAT-4).
Methodology	Users start in the simplest setting, and improve features of the tracking until they are satisfied with the experience. Using the order in which users improve the movement features we can assert on the most valuable co-presence animation features to the users, and statistically model these transitions in a Markov chain.
Data set description	Preferred improvement path per pair of subjects, i.e. what was the order with which each user improved animation features of the other user to transition from the least realistic to the most realistic animation scenario. Motion data, position and rotation of the motion tracking sensors and of the avatar joints controlled by each pair of users. This dataset will be used for early analysis of potential co-presence features that can be identified through movement behavior, thus helping in the design of the joint i2cat and Artanim experiment on the subject (see EXP-i2CAT-4).

EXP-CWI-1 Title: QoE of point cloud compression Type: QoE Date: November 2017 (completed) Number of users: 24	
Description	Existing work on point cloud quality assessment has mainly focused on point cloud geometry, and demonstrated that state-of-the-art objective quality metrics correlate poorly with human subjects' assessments. Not much attention has been given to point cloud quality evaluation based on its color, even though real world applications utilize color point clouds, and color artifacts may be introduced during compression due to different color 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. We propose to use a mixed methodology to look into this, and explore what users' perceive as annoying in compressed point cloud sequences.
Objective and expected results	To understand user perception of point cloud quality – what attributes of point cloud do users find critical in evaluating quality. ● New objective quality metrics for point cloud compression ● Analysis of the reliability and accuracy of the Double Stimulus Impairment Scale for evaluating point cloud quality ● Some key insights from our subjective experiments that would be useful in designing subjective experiments and quality metrics to evaluate user Quality of Experience (QoE) in immersive systems
Methodology	We first perform a subjective experiment using a mixed methodology. Users are asked to rate the quality of a set of point cloud stimuli using the Double Stimulus Impairment Scale. Next, they are asked to perform a qualitative study using the Sorted Napping method.
Data set description	We use 6 frames from 6 different sequences of full-body humans from the 8i Voxelixed Full Bodies (8iVFB v2) dataset. Each frame of point clouds from the dataset is compressed into 4 different levels of compression. The compression parameters used are Level of Detail (LoD) 10, LoD 9, LoD 8, and LoD 7. LoD 10 means that the compression uses a 10-b octree setting, i.e., 10 quantization bits per direction. We obtain a total of 24 point cloud sequences to be rated by our participants. During the rendering of the point clouds, we assign different point sizes for each compression level, such that each object can be seen in full (i.e., no hollow parts due to missing points can be seen).

Code: EXP-CWI-2 Title: QoE Metrics

Type: QoE Date: February 2018 (scheduled) Number of users: 30 (5 of them experts)	
Description	To design better metrics and experiments for Social VR experiences, we need to understand which factors or features in a Social VR experience that actually matters to users or experts. This experiment is a step to understand this.
Objective and expected results	• The objective of this experiment is to find factors that users and experts consider important in a social VR experience. • Moreover, we hope to compare the preferences of users and experts.
Methodology	We would distribute a questionnaire to naïve users and experts, asking them to point out which factors do they think are important for an excellent social VR experience. The advisory board of this project would be the experts to whom we send our questionnaires. The naïve users would be a subset of the general population.
Data set description	Based on the Quality of Experience (QoE) framework proposed in (Redi, 2013), and the initial questionnaire data obtained from VR Days 2017, we would build a questionnaire showing all factors the possible factors to influence a social VR experience. Users would then be asked to indicate or rank the factors that they think are important for an excellent social VR experience.

Code: EXP-CWI-3 Title: Visual perception of mixed media Type: QoE Date: March 2018 (scheduled) Number of users: 24	
Description	In the VRTogether project, we may use point clouds to represent humans or other objects while having a 360 video to represent the surrounding environment. In real cases, it may not always be possible to provide the same level of quality for both the point cloud object and 360 video. This raises the issue of how to optimize the quality of our mixed reality to users. Should we prioritize on providing more bandwidth for the 360 video, or for the point cloud representation? We, humans, use the world as our reference point, when we judge something. And so, we would expect users to be much more critical of the 360 video's signal fidelity. On the other hand, research on visual perception has shown that humans are very sensitive to depictions of human or animal faces. Thus, our first

	assumption of users being more critical of the 360 video's quality may not hold. For the above reason, we propose to perform an experiment on quality evaluations of mixed reality stimuli.
Objective and expected results	This experiment aims to understand users preference of visual quality in a mixed reality setup, i.e. display of point cloud object(s) together with a 360 or regular natural video. The expected results are: ● A new dataset of mixed reality videos, along with subjective quality scores ● User scores of the overall quality of a mixed reality video, user scores of the point cloud object quality, and user scores of the 360 video quality. ● A model of user preference of mixed reality (point cloud object vs. 360 video background) quality.
Methodology	We would ask a number of users to evaluate the quality of a set of mixed reality stimuli. Using a single stimulus methodology, we would ask them to rate the quality of the point cloud object and 360 background of a video separately. We would then ask them to rate the overall quality of the video. The following are the independent variables of the experiment: quality levels of the point cloud object and 360 videos, different positions of the point cloud object, and possibly different types of content for the 360 video (indoor vs outdoor?) The dependent variables of the experiment would be the point cloud quality score, 360 video quality score, and overall quality score.
Data set description	We would need to create a dataset of mixed reality stimuli. In each video, there would be a point cloud object in the foreground, with a 360 video as background. The levels of quality for the point cloud and for the 360 videos will be varied.

Code: EXP-i2CAT-1 Title: Effect of network on experience Type: QoE Date: September-October 2018 Users: 5	
Description	Place illusion of one user using different media formats, under realistic streaming constraints.

Objective and expected results	Determine the limits in bandwidth and delays that are acceptable for each format
Methodology	Assess place illusion (feeling of being there) while manipulating bandwidth and delays. Manipulate delays, or effective bandwidth and ask systematically about subjective quality metrics. We will set up a simple questionnaire-based test with groups of users per format, 3 conditions for bandwidths and 3 more for delays.
Data set description	Questionnaire data

Code: EXP-i2CAT-2 Title: Comparison of different types of human representation Type: QoE Date: June 2018 (scheduled) Number of users: 24	
Description	Individual assessment of media format preference and its contribution to plausibility and place illusion.
Objective and expected results	In this experiment, we want to: 1. Determine whether self-representation as traditional virtual body or as 3D reconstructed body is better for a good subjective perception of a virtual reality experience (Place illusion and Plausibility) 2. Determine which kind of content format is better, according to these measurement tests, and/or whether things like retargeting of gaze play a significant role in Place illusion and Plausibility
Methodology	One user experiencing content from scene 1, with the ability to switch between media formats (both for self and for the content). Build a graph of transitions of subjective between different representations
Data set description	A Markov model of subjective preferences between different experimental conditions and content from pilot 1

Code: EXP-i2CAT-3 Title: Impact of virtual body representation on togetherness (part 1) Type: Psychology Date: July 2018 (scheduled) Number of users: 24

Description	This experiments will evaluate the impact of virtual body representation on togetherness
Objective and expected results	In this experiment, we want to: 1. Determine whether self-representation as traditional virtual body or as 3D reconstructed body is better for a good subjective perception of a virtual reality experience (Place illusion and Plausibility) 2. Determine which kind of content format is better, according to these measurement tests, and/or whether things like retargeting of gaze play a significant role in Place illusion and Plausibility
Methodology	Two users watching pilot 1 content together. To validate the feeling of togetherness in VR. 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. The experimenter also needs to be able to show the participant to see the other's virtual body at different distances. And to give them instructions, either together or separately.
Data set description	Questionnaire data on a memory recall task and media content of pilot 1

Code: EXP-i2CAT-4 Title: Impact of virtual body representation on togetherness (part 2) Type: Psychology Date: October 2018 (scheduled) Number of users: 24	
Description	This experiments will evaluate the impact of virtual body representation on togetherness, when two users watching pilot 1 content together.
Objective and expected results	In this experiment, we want to: 1. Determine whether self-representation as traditional virtual body or as 3D reconstructed body is better for a good subjective perception of a virtual reality experience (Place illusion and Plausibility) 2. Determine which kind of content format is better, according to these measurement tests, and/or whether things like retargeting of gaze play a significant role in Place illusion and Plausibility
Methodology	VCapture motion data, speech. Undetermined predictive factors in such data (this is what the experiment is for: to have a dataset for exploratory data analysis)

Data set description	Behavioral data from 20 end-users, plus post-experimental questionnaires, and media content of pilot 1.
Code: EXP-CERTH-1 Title: Real-Time Distribution of time-varying-mesh (part 1) Type: Technology Date: February 2018 (scheduled) Number of users: No users involved	
Description	Time-varying-mesh Rabbit-MQ server simulations (dockers) will be established in Greece (Thessaloniki, CERTH) and Spain (Barcelona, i2Cat) to evaluate the real-time distribution of TVM between distant countries
Objective and expected results	To evaluate the distribution of time-varying mesh, allowing for better analyzing the required improvements of the related VR-Together modules.
Methodology	Uploading and downloading frame rate / 3D reconstruction production rate
Data set description	In this experiment, data will be collected for assessing the distribution rate of TVM between remote users using different parameterization of TVM geometry generation and texture quality usage.

Code: EXP-CERTH-2 Title: Real-Time Distribution of time-varying-mesh (part 2) Type: Technology Date: February 2018 (scheduled) Number of users: 5	
Description	Users in Greece (Thessaloniki, CERTH) and Spain (Barcelona, i2Cat) will be captured/reconstructed and the data will be transmitted in real-time, allowing us to evaluate the real-time distribution of TVMs between distant countries. Two people 3D capture and reconstruction setups (5 PCs per setup and 4 RGB-D sensors) will be used, one per user.
Objective and expected results	To evaluate the distribution of time-varying mesh, allowing for better understanding the required improvements of the related VR-Together modules
Methodology	Uploading and downloading frame rate / 3D reconstruction production rate

Data set description	TVM sequences and metadata will be stored and considered as baseline for the next/improved versions of the modules used in the pipeline.

Code: EXP-CERTH-3 Title: Assessment of the interference between HMD and multiple depth-sensing Type: Technology Date: End of January 2018 (scheduled) Number of users: 3	
Description	Users will be captured using 3D capture and reconstruction setup, wearing HMDs. In particular, the people 3D capture and reconstruction setup will be used when users have VR experience using HMDs. In these experiments, both Kinect for Xbox One and Intel RealSense D400-Series will be utilized.
Objective and expected results	To evaluate the interference between HMD and multiple depth-sensing devices in different sensor placements and configurations, allowing us to better understand the problems and investigate potential solutions.
Methodology	HMD Flickering rate, Functioning, QoE
Data set description	This is a hardware evaluation experiment, allowing us to assess the compatibility between the people 3D capture and reconstruction setup and of-the-shelf head mounted displays (Oculus and HTC Vive). Details regarding the results of the experiment will be reported.

Code: EXP-CERTH-4 Title: HMD Removal Type: Technology Date: May 2018 (scheduled) Number of users: 35	
Description	For HMD removal (open research topic) from real-time people 3D reconstruction, we will record a dataset of users wearing and not wearing HMDs in order to research and develop solutions for both 3D mesh geometry approximation as well as face texture image synthesis.
Objective and expected results	The RGB-D captures will be used to generate sufficient data for training the developed approaches towards artificially removing the HMD from the user 3D reconstruction. This will contribute to the QoE of the VR-Together platform, as it will boost the “reality” experience of being together in VR environments, as users will be able to see each other’s faces.

Methodology	The quality of face synthesis from HMD removal falls into a semantic concept which can not be accurately quantified. The use of human annotators or automatic methods to evaluate the performance will be employed for measuring the quality.
Data set description	RGB-D videos of approximately 40 subjects posing a variety of facial expressions, with and without HMD.

7. CONCLUSIONS

In this document we have introduced the general and the specific requirements for the first pilot. We have also introduced the content scenario for pilot 1, as well as specified software requirements for the different modules involved in pilot 1, and how the different components fit together in a global architecture.

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. 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.

Annex I. Table of requirements

CODE	NUM	Title	Description
General project requirements			
GEN	1	Copresence	End users should be able to be virtually present in the same virtual space and engage in real-time face-to-face social activities. Copresence 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 others 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
Specific requirements pilot 1			

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+)
Specific requirements pilot 2			
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.
Specific requirements pilot 3			
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
Evaluation Pilot 1 requirements			
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 behavioral measures related with togetherness, we need to be able to record the entire multi-modal data, with good time precision.
Capture and Reconstruction requirements			
CRR	1	People RGB-D Capture	The people capture component/system will capture RGB-D data from 4 RGB-D devices connected to 4 capturing nodes (RGB-D nodes). The capture rate will be at least at 25 fps.
CRR	2	People RGB-D Calibration	The RGB-D devices of the system will be automatically calibrated (extrinsic calibration).
CRR	3	People RGB-D Synchronization	The RGB-D frames from the RGB-D nodes will be synchronized and grouped in a central node, resulting in a RGB-D group frame. The delta time used to group the frames will be configurable.
CRR	4	People live 3D reconstruction	The people live 3D reconstruction component will process user's coloured 3D point cloud to reconstruct a 3D time-varying mesh in real-time (under 80ms per frame). The resolution of the voxel grid used for 3D reconstruction will be configurable, allowing for different mesh interpolation for TVM production.
CRR	5	People live 3D reconstruction visual quality	The people live 3D reconstruction method will be extended, resulting in higher visual quality level than the initial version.
CRR	6	static background removal	Users must be standing or sitting. The FGBG module requires a static placement of the Kinect V2 sensor with a background and foreground object within 5 meters distance. After a replacement of the sensor, a new background image must be captured.
CRR	7	Image properties for background removal	Input image properties should be 960x540 pixels at a framerate of 25 fps; the FGBG then drops approx. 1 frame per second on a low (i.e. below 30%) CPU load.
CRR	8	Face capture	The user's face needs to be captured from at least two different poses.
CRR	9	Captured face storage	The captured user face needs to be stored (on disk or in memory) and must be accessible in real-time by the face inpainting algorithm.
CRR	10	Face inpainting	The HMD removal process through face inpainting must work offline, i.e. non real time. The process may work in real time.
Encoding and Decoding requirements			
EDR	1	Real time compression	The framework has low delay encoding and decoding of less than 200 ms
EDR	2	Progressive decoding	The framework allows a lower quality point cloud to be decoded from a partial bitstream
EDR	3	Efficient compression	The framework can achieve a compression ratio of up to 1:10 to in order to stream point clouds
EDR	4	Low end to end latency	The framework has an end to end latency below 300ms similar to video conferencing requirements
EDR	5	No a priori information	No geometric properties of the objects are assumed
EDR	6	Generic compression framework	The framework can be used to compress point clouds of arbitrary topology (independent of capture device, point precision, file format etc.)
EDR	7	Quality assesment	The framework will make use of quality metrics (to be decided) which informs about the expected quality of experience frorm the user perspective
EDR	8	Real time compression	The framework has low total delay, encoding and decoding in less than 80 ms per frame
EDR	9	Efficient compression	The framework can achieve textured mesh (3D geometry and textures) compression ratio of up to 1:30
EDR	10	Generic compression framework	The framework can be used to compress textured 3D mesh of arbitrary topology (independent of size, number of surface triangles, etc.)
EDR	11	Real-time time-varying mesh encoding	This component must compress and decompress 3D time-varying meshes and the corresponding RGB textures in real time. (under 80ms per frame)
EDR	12	Real-time time-varying mesh encoding parametrization	The component will allow the selection of different TVM compression and texture resolution and quality level.
EDR	13	Real-time time-varying mesh distribution	The component will transmit 3D time-varying meshes in real time (under 100ms per frame), depending on the network type and quality.
EDR	14	Real-time time-varying mesh distribution parametrization	The component will enable the tweaking of TVM frame time life and the frame queue length.

EDR	15	End-user audio encoding	End-user audio encoding should be done with formats supported by the browsers in which the web player run, at bitrates comparable to state-of-the-art video communication applications.
EDR	16	End-user video encoding	End-user video encoding should be done with formats supported by the browsers in which the web player run, at bitrates comparable to state-of-the-art video communication applications.
EDR	17	Content audio encoding	Content audio encoding should be done with formats supported by the browsers in which the web player run, at bitrates comparable to state-of-the-art video communication applications.
EDR	18	Content audio encoding	Content video encoding should be done with formats supported by the browsers in which the web player run, at bitrates comparable to state-of-the-art video communication applications.
EDR	19	End-user audio distribution	End-user audio distribution should be done with formats supported by the browsers in which the web player run, at bitrates comparable to state-of-the-art audio communication applications.
EDR	20	End-user video distribution	End-user video distribution should be done with formats supported by the browsers in which the web player run, at bitrates comparable to state-of-the-art video communication applications and similar latency requirements.
EDR	21	Content audio distribution	Content audio distribution should be done with formats supported by the browsers in which the web player run, at bitrates comparable to state-of-the-art audio communication applications and similar latency requirements.
EDR	22	Content video distribution	Content video distribution should be done with formats supported by the browsers in which the web player run, at bitrates comparable to state-of-the-art video communication applications.
Orchestration requirements			
OR	1	Configuration	The orchestration modules must support remote configuration through an administrative interface.
OR	2	Content discovery	The orchestration modules must be able to facilitate discover both content streams and files, and at least two end-user representation streams.
OR	3	Session management	The orchestration modules must manage mangage sessions and facilitate the discovery, setup, control and synchronization of a session between a least two end-users.
OR	4	Session management	The orchestration modules should support at least two parallel sessions.
OR	5	Database storage and access	The orchestration modules should store and access data in a shared database. This database should contain the current time in the room, the number and state of users in the room, the state of the content being consumed, and the state of the shared environment.
OR	6	Optimization	The orchestration modules must be able to reduce the requirements on the end-user players by processing at least two end user representations coming from the 3D reconstruction system.
Web player requirements			
WPR	1	Content	The web player must support playback of the content as 2D video.
WPR	2	End user representation	The web player must support playback of end users represented as 2D video; the video format of end user representations should be at least 960x540 pixels at a framerate of 25 fps.
WPR	3	Audio	The web player should support spatial audio, to align the content audio with the end user audio.
WPR	4	Communication	The end user representation audio and video should be played back by the web player in real time. This means played back within a latency common for real time communication (capture to display delay of under 400ms).
WPR	5	Streaming	The web player must receive the streaming of both content and end user representations streams.
WPR	6	WebVR	The web player must operate in a browser that supports WebVR and A-frame.
WPR	7	Bandwidth	The web player should support content bandwidth adaptation. The web player must rely on the mechanisms supported by the browser.
WPR	8	Latency	The latency between different streams should not be higher than 500 ms.
WPR	9	Scene	The scene must be a static 2D 360 image of at most 4K pixels, in ERP format; Other scene formats (like 2D 360-degree ERP video) may be supported, depending on the performance restrictions of the PC and browser.

WPR	10	End user placement	The web player must allow for manual configuration of the end user location; the web player should allow for end user placement control by the orchestration modules.
Native player requirements			
NPR	1	Multiple format	The native player supports consuming contents in different VR formats, like Point Clouds, omnidirectional video, static meshes, dynamic meshes, mono/stereo 2d video.
NPR	2	Hybrid format	The player reproduces hybrid VR contents in the same scene, that is, compositions made of blending different Media formats
NPR	3	Audio	The player reproduces immersive audio tracks, the point of view of the user affects the perceived sound.
NPR	4	Rendering 1	The player is able to reproduce smoothly relevant combinations of available formats
NPR	5	Rendering 2	The player is able to render at least some of the combined media formats at 60 fps or more
NPR	6	Rendering 3	To better integrate different media formats and sources, the player can alter the lighting of specific objects within the scene, on the basis of custom shaders.
NPR	7	DoF	The client can consume content adapted to 3DoF or 3DoF+ movements (i.e., head position and rotation). The media player integrates within the rendering pipeline user movements (3DoF+)
NPR	8	Quality of Image	input/output effective display resolution will support up to 4K
NPR	9	Delay on displaying self representation	Self representation impose a latency constraint under 20ms.
Playback synchronizaton requirements			
PSR	1	Sync audio and video (lips- mouth)	The delay between audio and video has to be acceptable in the range of +90 ms (audio first) to –185 ms (video first).
PSR	2	Sync multiple formats	The player must support synchronization between different formats with less than 40ms of delay), but the acceptability, would be under 200ms delay.
PSR	2	Sync inter device	The synchronization between different players should be frame accurate (less than 20ms of delay), but the acceptability, would be under 100ms delay.
PSR	3	Sync control	Players should be able report their playback status and be able to adjust playback to resynchronize with other players.
PSR	4	Timestamping	Media streams must be timestamped at the source in relation to a common timeline, such that the player is able to establish synchronized playback of the audio and visual components.

Annex II. End User Questionnaire used in VR Days event

VR Together

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

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
Google Forms