European Commission | Horizon 2020 European Union funding for Research & Innovation

Deliverable

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

D2.1- User scenarios, requirements and architecture v1

Revision: 1.4

Authors: Themistoklis Karavellas (i2CAT), Sergi Fernandez (i2CAT), Javier Lajara (Entropy), Pablo Cesar (CWI), Pascal Perrot (VO), J.A. Núñez (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.Prins, 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.
1.1	26-6-2018	T. Karavellas	I2CAT	Introduction of Updated Requirements matrix and Use cases and refinement of the whole document content.
1.2	20-7-2018	T. Karavellas	I2CAT	Incorporation of comments on calendars and matrix
1.3	31-10-2018	T. Karavellas	I2CAT	Included the updated list of experiments in chapter 5
1.4	13-11-2018	T. Karavellas	I2CAT	Final updated revision

Disclaimer

The information, documentation and figures available in this deliverable, is written by the VR-Together – project consortium under EC grant agreement H2020-ICT-2016-2 762111 and does not necessarily reflect the views of the European Commission. The European Commission is not liable for any use that may be made of the information contained herein.

Statement of originality:

This document contains original unpublished work except where clearly indicated otherwise.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

EXECUTIVE SUMMARY

The present document is the reference document for the software integration and experimental work executed within VR-Together. It describes the requirements, architecture and experimental work envisaged to implement the main paradigm outlined in VR-Together: 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 within a virtual environment.

To clarify the integration of these aspects, after a brief introduction, we start by drafting presenting the scenario for pilot 1. It features 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 content similar to a police interrogation scene. This will allow introducing a concrete overarching story across the three pilots, and is perceived to establish balance between a commercially-relevant content format, close to the typical thriller-like content, and one that enables exploring different possibilities regarding experimental validation. We have also decided to produce two versions of the same content: one where actors are captured on video and rendered as stereoscopic video billboards within a 3D environment, and another one with 3D-rigged characters, combined with motion capture techniques. In addition, we will also create versions of this content rendered as omnidirectional video. All production and post-production efforts aim at facilitating direct comparison of media formats enabling a rich 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 3 outlines in further detail the software platform of VR-Together, particularly for the scenario of pilot 1. First, it specifies the characteristics of the two main use cases –i.e., content consumption and social interaction. Then analyses the implicated user scenarios in use cases. Next presents the software requirements specification for the platform by describing the requirements gathering methodology, the user profiles, the assumptions and other factors defining the elicitation and capture of new requirements. At the end of the section the requirements description, as they are presented in the grant agreement, are recapitulated and an extensive requirements Specification table is presented in 3.4.5

Section 4 introduces the overall architecture for VR-Together, how the different components interact and the hardware topology, taking into consideration the software modules from WP3.

Section 5 outlines all the information related to the distributed lab realized within the VR-Together project. It presents the Advisory board, supporting the project. Then describes the lab nodes, in the partners' premises. Next lists the experiments associated and projected within VR-Together, followed by a detailed and updated experiments calendar.

Section 6 summarizes the contributions of the deliverable.

Last, annex 1 and 2 provide an example of a questionnaire used for experimentation and a detailed description of the Lab nodes in each one of the partners' locations.

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

Revision History 2
.... 3
Executive Summary 4
Contributors 5
Table of Figures and tables 9
1. Introduction 10
1.1. Purpose of this document 10
1.2. Scope of this document 10
1.3. Status of this document 10
1.4. Relation with other VR-Together activities 10
2. Pilot Scenarios 10
2.1. Pilot 1 11
2.1.1. Plot 11
2.1.2. Storyboard 14
2.1.3. Pre-Production 17
2.1.4. Production 18
2.1.5. Pilot Action Calendar 22
2.2. Pilot 2 25
2.3. Pilot 3 25
3. Software Platform Description 25
3.1. User Scenarios 25
3.2. Use Cases 26
3.2.1. General Use Case 26
3.2.2. Detailed Use Cases 27
3.3. Requirements Specification 31
3.3.1. Requirements gathering methodology 31
3.3.1.1. Requirements Gathering Techniques 31
3.3.1.2. Types of requirements 32
3.3.1.3. Prioritization of requirements 32
3.3.1.4. System component of requirements 33
3.3.2. VR-Together User Profiles 33
3.3.2.1. End-User 33
3.3.2.2. Administrator 33

3.3.2.3. Researcher 34
3.3.3. Reference documentation....34
3.3.4. Assumptions and dependencies....34
3.3.5. Interface Requirements 34
3.4. VR-Together Software Requirements Matrix....34
3.4.1. Requirements for Pilot 1 35
3.4.2. Requirements for Pilot 2 36
3.4.3. Requirements for Pilot 3 37
3.4.4. Experimentation requirements 38
3.4.5. Requirements Specification Table 39
4. Architecture 50
4.1. Software architecture 50
4.1.1. Capturing component 52
4.1.2. Encoding and Encapsulation component....53
4.1.3. Delivery component....54
4.1.4. Orchestrator component 54
4.1.5. Play-out component 56
4.2. Hardware architecture....57
4.2.1. End-user Set-up 57
4.2.2. Processing Servers Set-up 59
4.3. VR-Together as a Software product 59
5. User Lab....59
5.1. Advisory Board 60
5.2. User Lab Nodes....61
5.3. Experiments 63
5.4. List of experiments 63
5.4.1. Technology Evaluation 64
5.4.1.1. CWI-1....64
5.4.1.2. CERTH-1 64
5.4.1.3. CERTH-2 64
5.4.1.4. CERTH-3 65
5.4.1.5. CERTH-4 65
5.4.2. User Experience Evaluation....66
5.4.2.1. ARTANIM-1 and ARTANIM-2....66
5.4.2.2. CWI-2....66

5.4.2.3. CWI-3 67
5.4.3. Feedback from Professionals 68
5.4.3.1. TNO-1: Initial use-case study 68
5.4.3.2. TNO-2: Try-out of VR stand-up 68
5.4.3.3. TNO-3: Representing the environment and users in either 2D or 3D 69
5.4.3.4. TNO-4: Input from professionals based on the pilot 1 experience 69
6. Conclusions 70
7. Annex I. End User Questionnaire used in VR Days event 71
8. Annex II: VR-Together Lab Nodes 75
8.1. Artanim Lab Node 75
8.2. CERTH Lab Node 76
8.3. CWI Lab node 77
8.4. i2CAT Lab Node 78

TABLE OF FIGURES AND TABLES

Figure 1. Scenes integrating general story 17
Figure 2. Initial Concepts for the Trial (interrogation, crime scene). 18
Figure 3. Pilot Proposal – murder scene. 19
Figure 4. Pilot Proposal – interrogation with one-way mirror. 19
Figure 5. Pilot Proposal – interrogation inside the prison. 19
Figure 6. Police officer waiting for the suspect (scene 1) 20
Figure 7. Suspect introduction (scene 2). 20
Figure 8. Interrogatory (scene 3). 21
Figure 9. Secret revelation (Scene 4). 21
Figure 10. Interrogatory ends (scene 5). 22
Figure 11. User's discussion (scene 6). 22
Figure 12. Production workflow. 23
Figure 13. Stereoscopic shooting of character action. 24
Figure 14. 3D Scene where action takes place. 24
Figure 15. Coherent lighting. Users and scene. 25
Figure 16. Scene composition (3D Billboards + 3D scene). 25
Figure 17. Sound design 26
Figure 18. 3D character capture. 26
Figure 19. Motion tracking to animate pre-rigged characters 27
Figure 20. Component diagram for platform v1. 41
Figure 21. Hardware architecture. 42
Figure 22. Schematic View of a VR-Together hub. 45
Figure 23. Artanim’s User Lab. 46
Figure 24. Artanim’s User Lab. 47
Figure 25. CERTH’s User Lab. 48
Figure 26. CWI’s User Lab (Pampus) 48
Figure 27. CWI’s User Lab (QoE Lab) 49

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. PILOT SCENARIOS

The three planed 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 identified towards the creation of truly realistic social VR experiences.

The Pilots were initially planned as individual content capsules addressing completely different content scenarios. The structure and plot complexity of the pilots is linked to a gradually increasing technical difficulty, with the first pilot being the simpler to produce and 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 focused on live news, simulating a live production of immersive content from multiple sources that aimed to virtually transfer the user to the location of the news and share the experience with other users.
- the third pilot intended to present a test to users through an interactive and totally immersive experience, with the background of a television series, a movie or simply a scene taken from them, where users can participate in the scene, interact between them, make conclusions, etc.

Pilots are planned to be executed in the periods of July – September 2018, June - August 2019 and May - July 2020. During those periods, a number of experiments will run in the project user labs and project roadshows demonstrating the status of results are planned to be launched in technical and creative industrial fairs and events.

In the following section we describe the current version of Pilot 1, its related plot, the storyboard, the pre-production and production activities along with a detailed Pilot Action calendar that lists the number of actions executed by using the Pilot 1 content over the VR-Together platform. An extended description regarding Pilot 1 and all relevant actions performed in VR-Together can be found in D4.1 and D4.2.

2.1.Pilot 1

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

One key concern of the consortium was if such a new approach in the pilot scenarios 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 for the pilots to serve as a carrier of technological advances and experimentation (co-presence, togetherness, immersion). During subsequent meetings between the creative and the technical partners, specific agreements have been made in order to ensure that the series of the pilots is a valid vehicle for the project. It was, therefore, decided for the first pilot to remain focused on communication between remote participants while performing an activity together, the second on the scalability of the platform, and the third one on the interactivity with the content.

2.1.1. Plot

The chosen 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 unique advantages of the project, a team composed of technical and artistic experts, by creating a brand new experience that makes thrives on the artistic creation. Thereby, the final objective is to obtain a new conceptual experience that involves the viewers and immerses them in an uncommon encounter that is different from what they might have previously seen.

One of the questions that often was posed within the consortium was "*What is the target group for a product such as VR-Together*?" As an answer was not profoundly evident, it was decided to not set limits and therefore follow generic point of view that would serve for all the various candidate audiences, avoiding addressing only specific types of end-users. Being inspired by movies such as "*The Usual Suspects"*, the proposal is to have a thriller-like plot as the theme for the three pilots. This way, the viewer, who will have control over the pilot, can enjoy the experience not only during the course of the pilot, but also after that. Overall the structure of the pilots is presented in Figure 1.

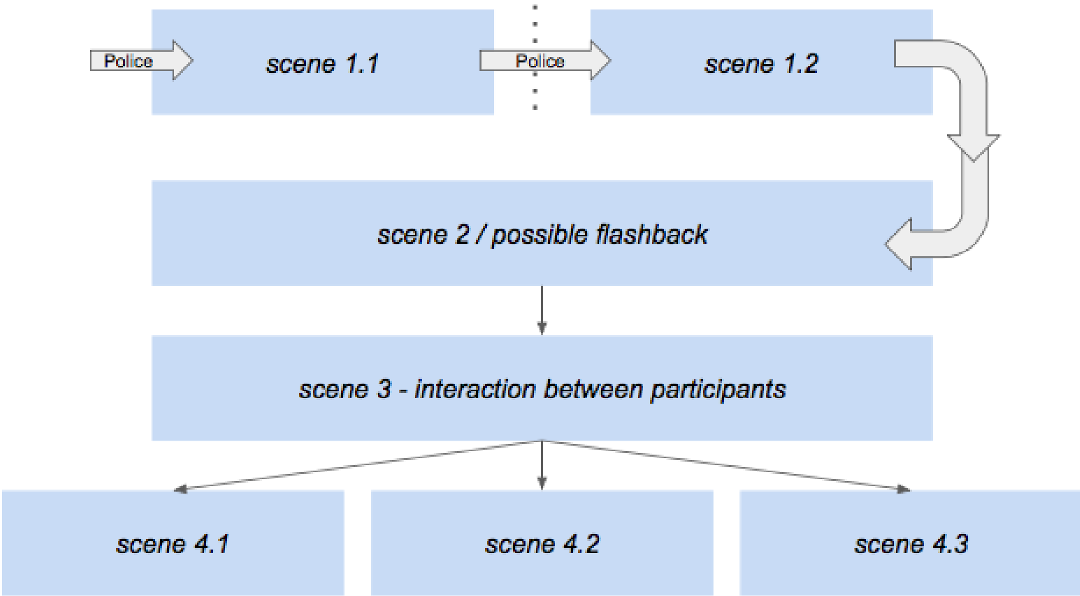

Figure 1. Scenes integrating general story

The creative partners of the project went through a number of iterative design sessions, creating visual representations that facilitated discussion (see Figures 2. and 3). Coherent plot suggestions and ideas were brought forward, discussed and argumented within the consortium eventually evolving into a complete a pilot definition.

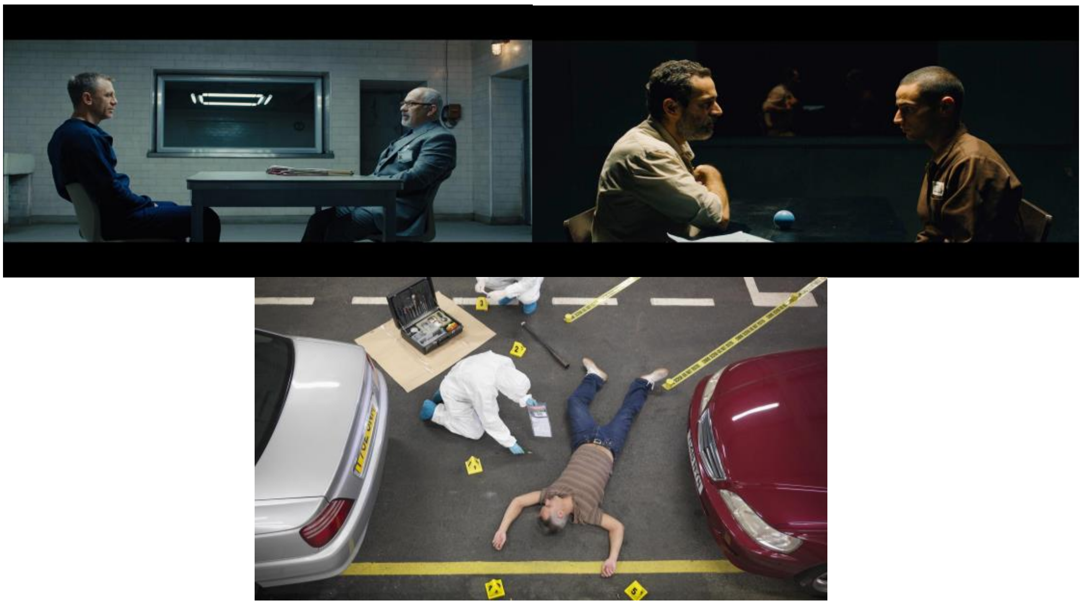

Figure 2. Initial Concepts for the Trial (interrogation, crime scene)

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 3), interrogation with one-way mirror (see Figure 4), and interrogation inside a prison (see Figure 5). The general ideas of each type of scene can be summarized as follows:

- Murder scene: In this scenario both users are found in the same room, where a murder has been committed, and both users are sufficiently distant from each other in order 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 reach into conclusions and decide on the identity of the murderer.
- Interrogation room with one-way mirror: the users are behind a one-way mirror of an interrogation room. Although users are next to each other, each user is able to see his/her own interrogation room, both being aware that the other user is having a different variation of the story.
- Interrogatory inside the prison: In this scenario both users are inside a 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 scenario. Following, Entropy Studio is developing the storyboard of the general concept of the pilot and the production plan.

2.1.2. Storyboard

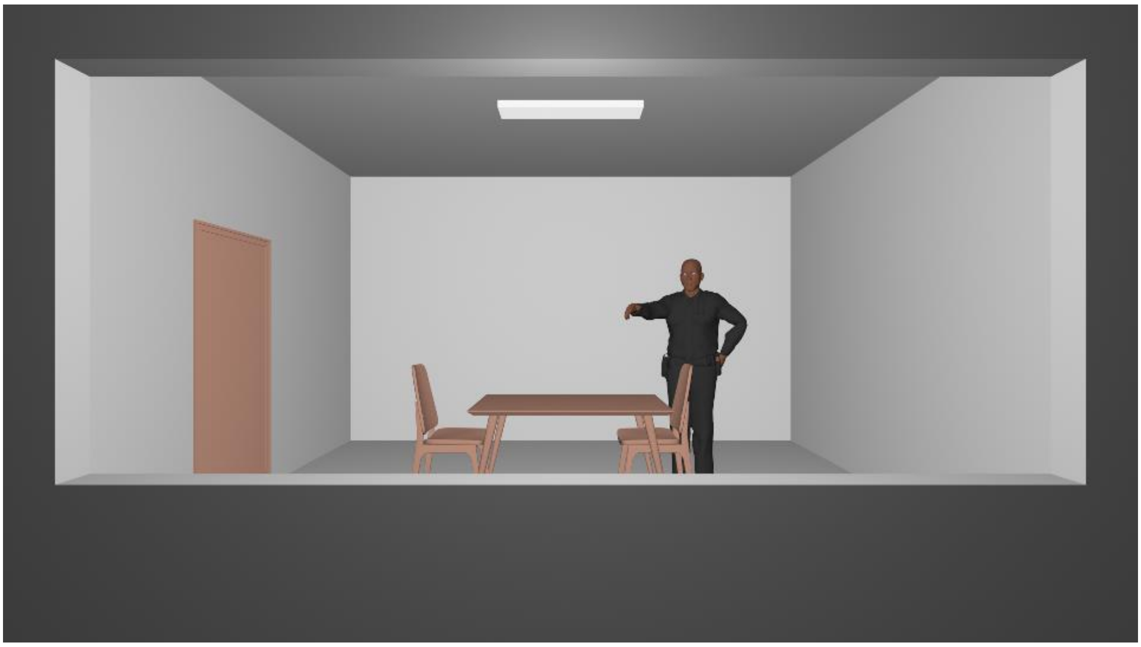

Figure 6. Police officer waiting for the suspect (scene 1)

In the intro, we are found 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, a second police officer enters in the room. A suspect is brought in to be interrogated and we are the witness of the interrogation. We are supposed to closely pay attention to the discussion and looks for clues that can help in clarify the identity of the criminal.

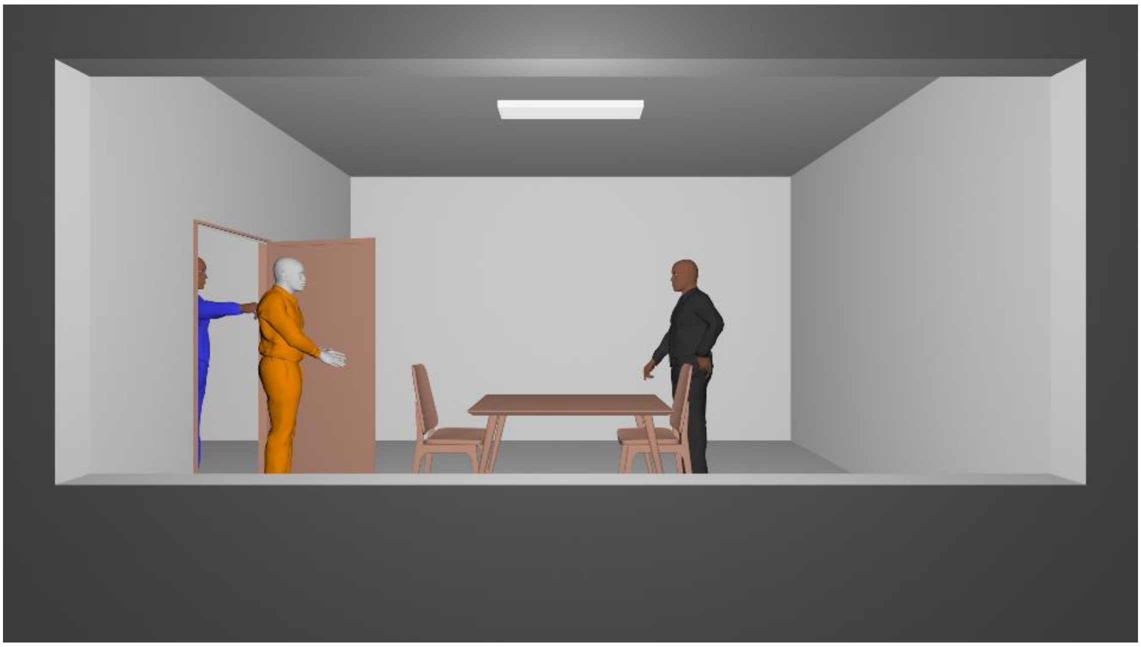

Figure 7. Suspect introduction (scene 2)

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

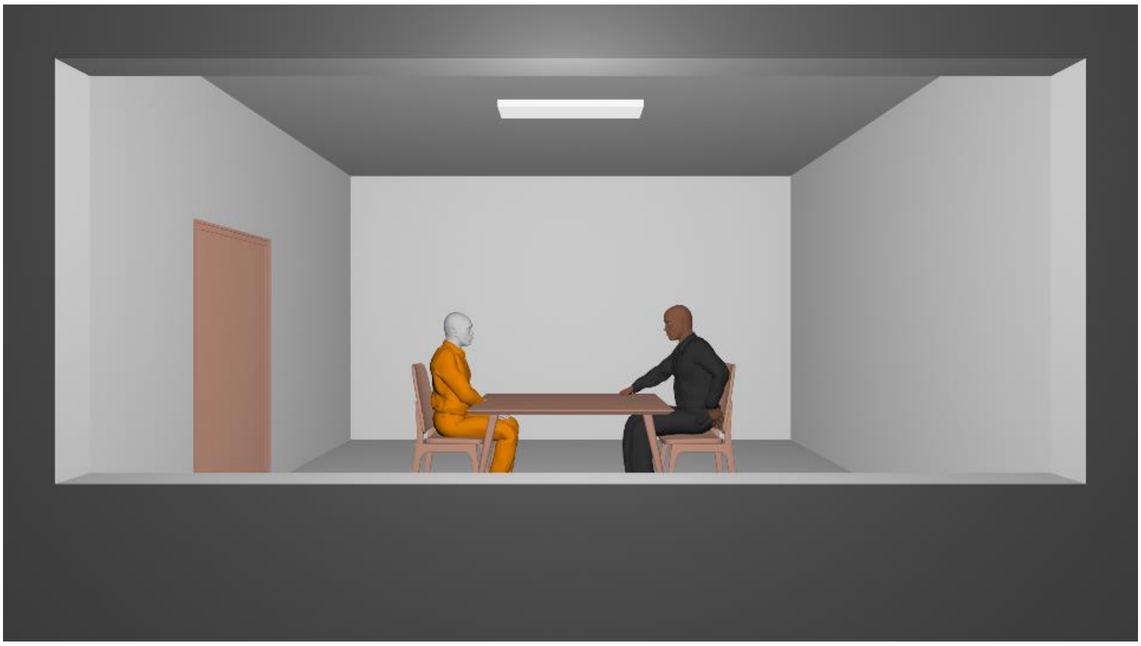

Figure 8. Interrogation (scene 3)

To create a conflict in the plot,the officer to start asking questions, talking about the situation of the crime scene, the location of the suspect at the moment of the crime, the location of the other suspects, etc.

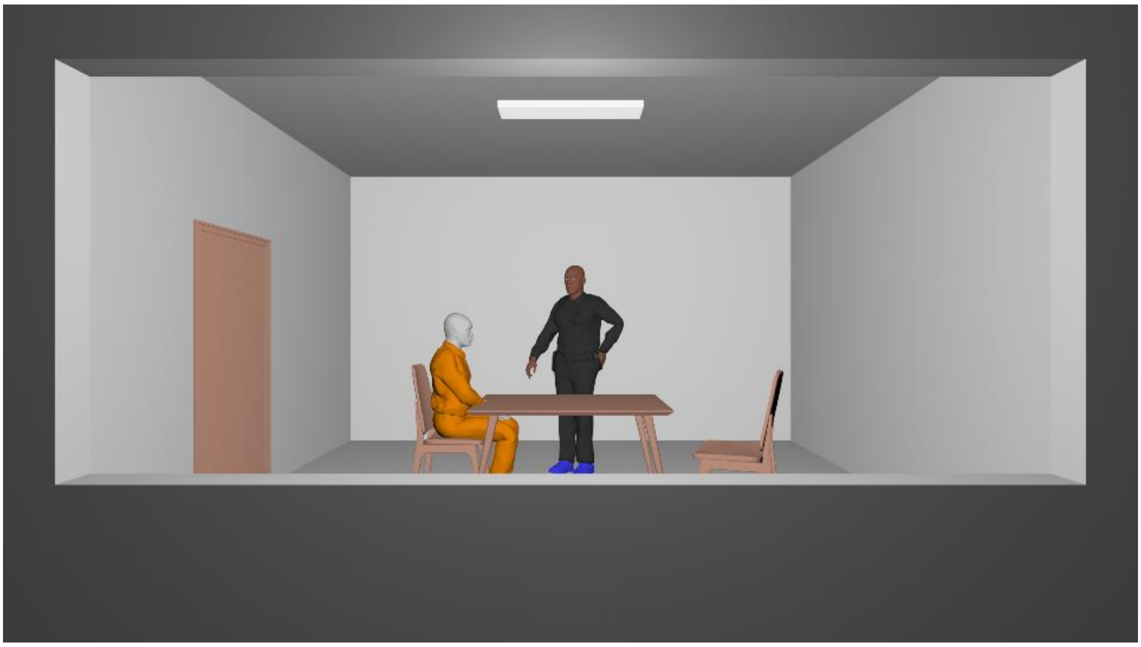

Figure 9. Secret revelation (Scene 4)

During the conversation that takes place at this point the suspect reveals important and relevant information that can lead us in identifying the criminal.

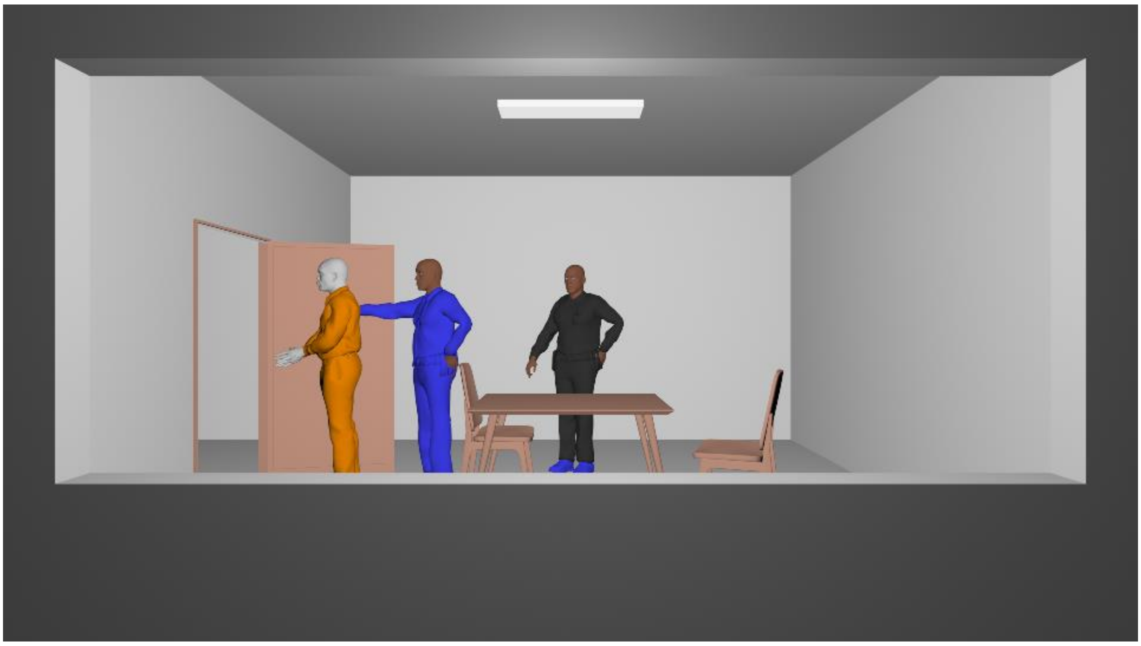

Figure 10. Interrogation ends (scene 5)

For denouement, the interrogation ends, leaving the officer alone in the room stating the facts and drawing the final conclusions looking at us through the window.

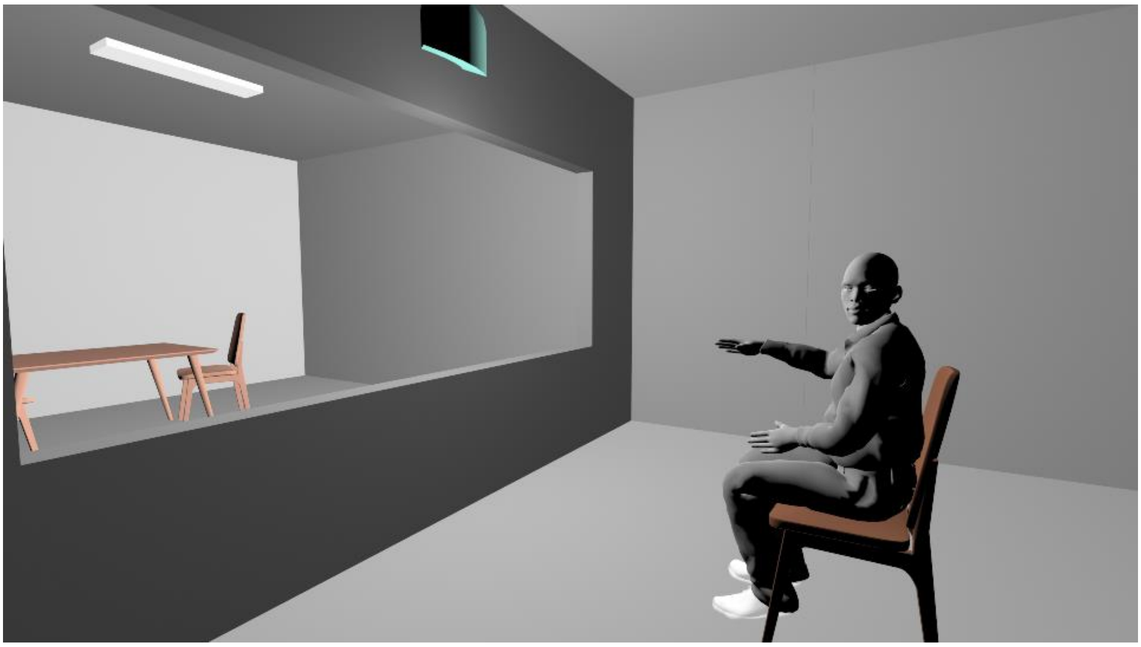

Figure 11. User's discussion (scene 6)

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

2.1.3. Pre-Production

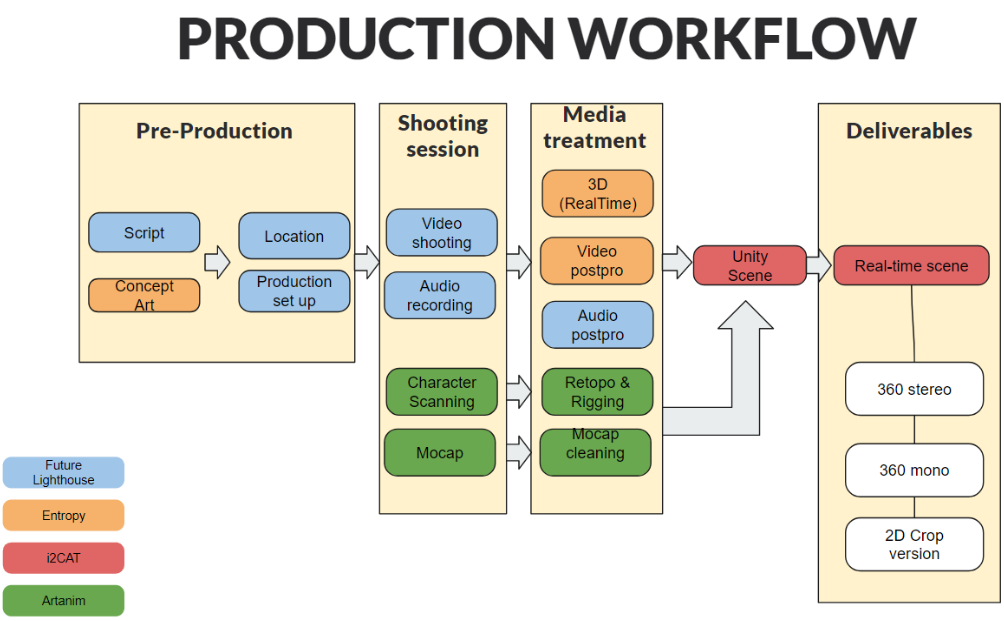

Figure 12. Production workflow

The diagram on Figure 12 describes the process through which the partners will the 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
  - Miscellaneous logistics
- Soundtrack and music
  - Curation of the music and sounds that will be used in the experience. Preferably of CC0 or Pd licence.

2.1.4. Production

This section provides a technical breakdown, graphically supported, to clearly describe 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

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

This will take place in an environment allowing for chroma key compositing during the post-production phase. This way it will be possible to remove the background and place the action in any kind of setting required by the scenario.

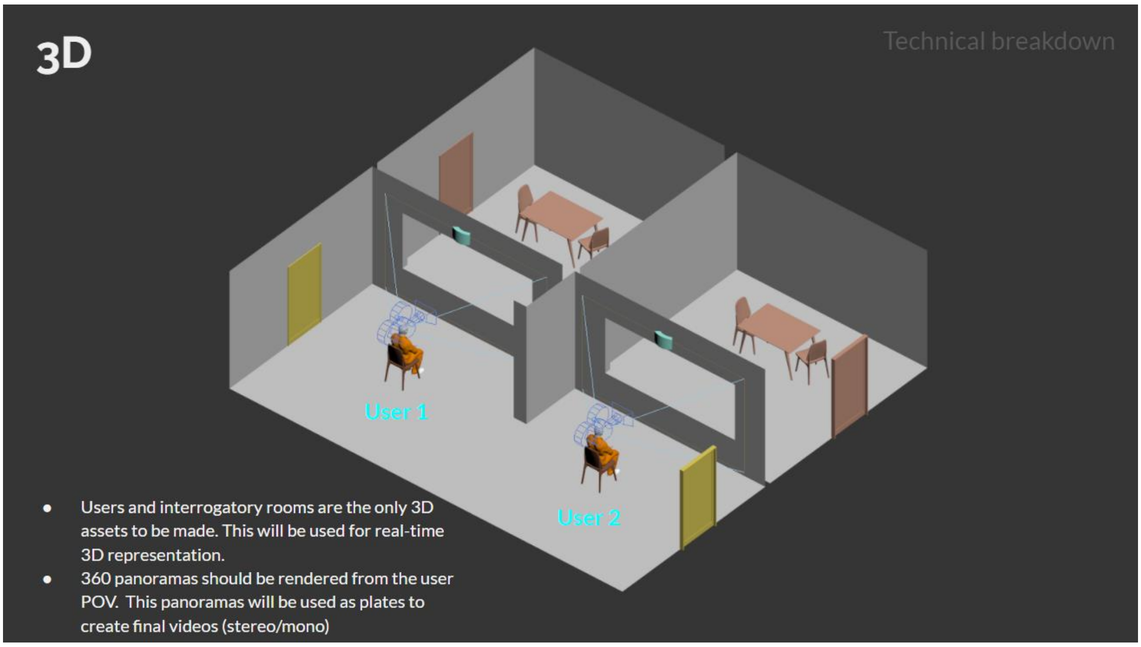

Figure 14. 3D Scene where action takes place.

Next, a room will be modeled , to simulate the Police station, where each participant will watch the interrogation scene described in 2.1.1.2 . 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.

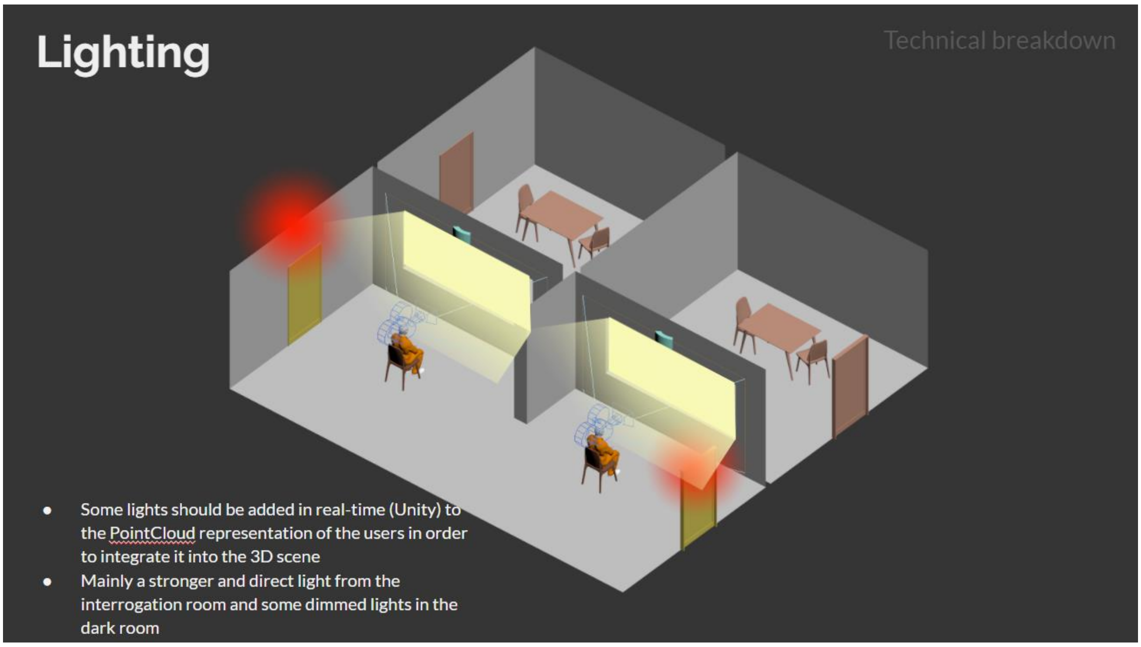

Figure 15. Coherent lighting. Users and scene.

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

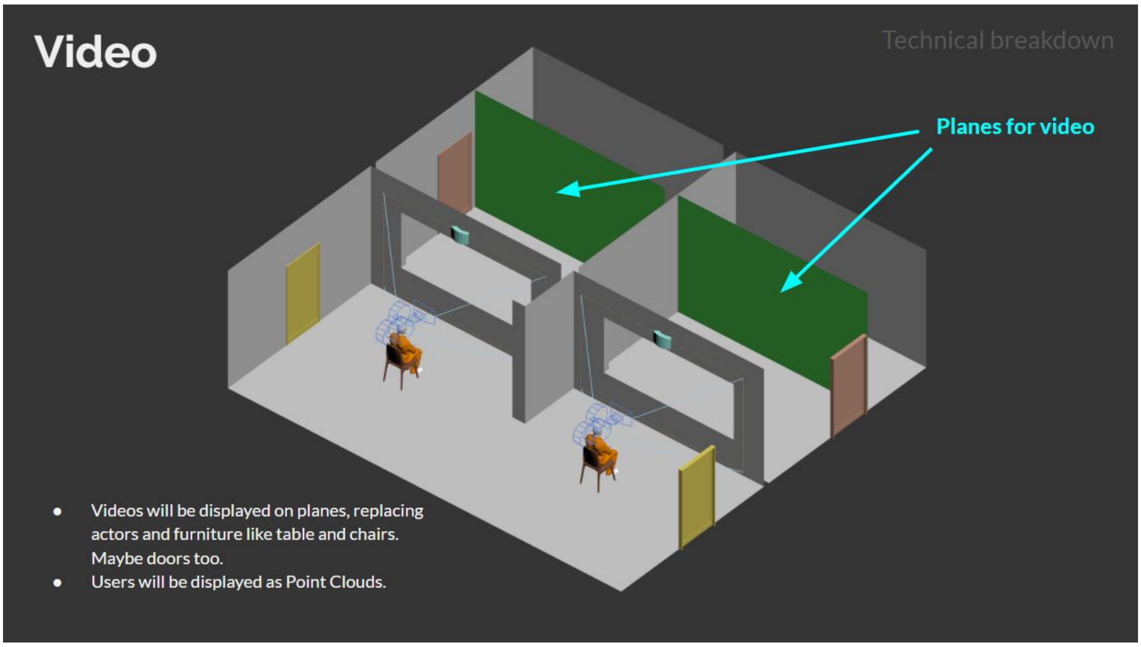

Figure 16. Scene composition (3D Billboards + 3D scene)

The video recorded with the stereo rig will be used as content within the virtual room and placed in a geometric plane inside the Unity scene. This video will 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.

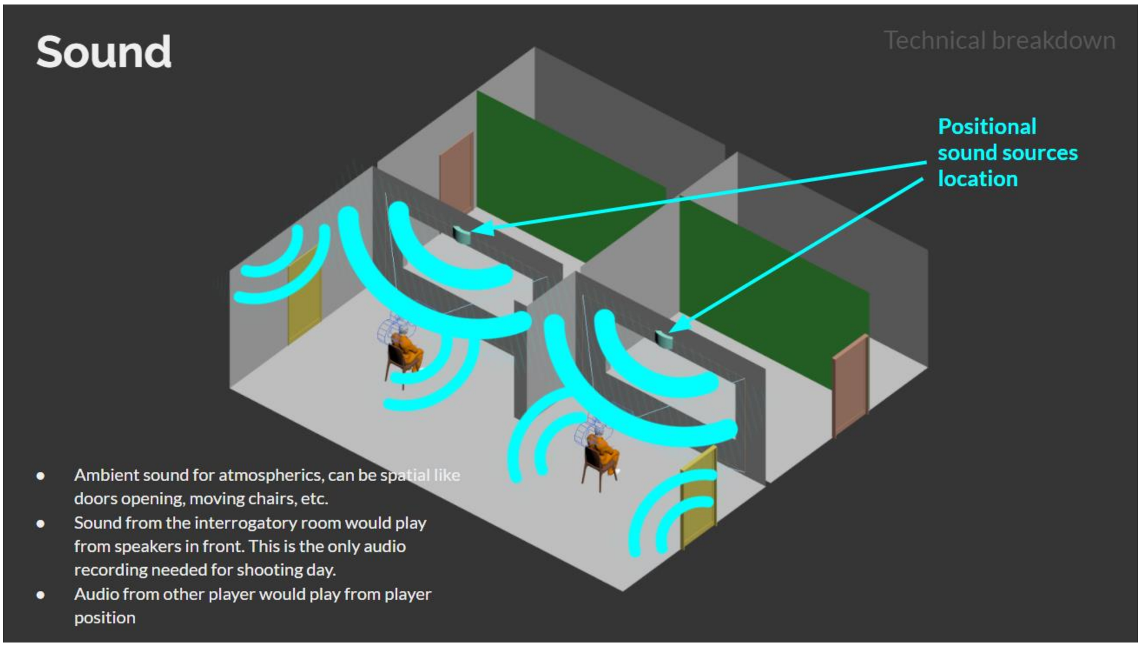

Figure 17. Sound design

The sound of the recorded scene and additional audio content will be placed in the Unity scene as objects, giving the sensation of spatiality, despite the fact that fore the recording there were used with traditional stereo microphones.

The audio content created at this step will be added to the video.

For experimentation purposes, the consortium decided to create an additional version using scanned characters, with the purpose of comparing streaming and psychological experience differences between the different media formats.

This version will be achieved by using a photogrammetric rig of 96 cameras, which are able to produce a geometric representation of the characters.

Figure 18. 3D character capture.

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

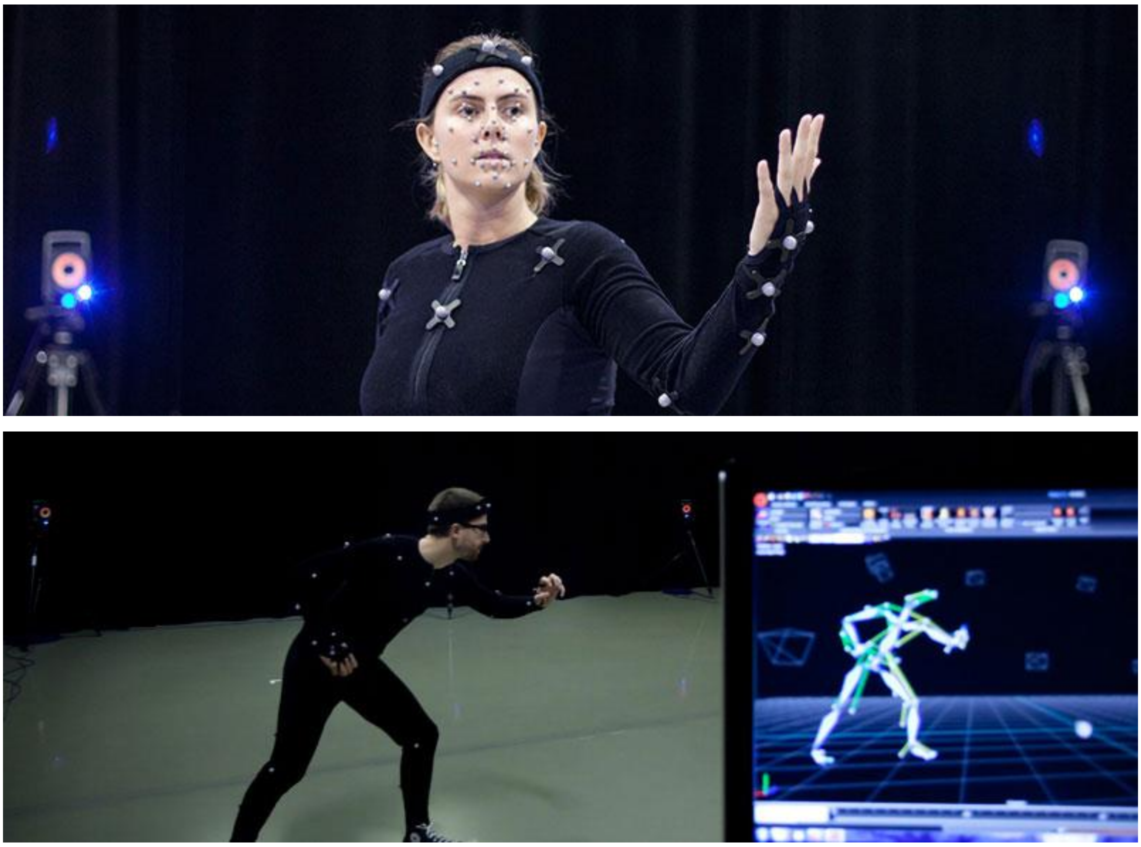

Figure 19. Motion tracking to animate pre-rigged characters

Therefore, after the interrogation scene is recorded in video, the actors will replay it wearing motion capture suits, in order to translate their movements to the virtual ones.

From the process previously described, we will obtain various media formats for the experiments and dissemination (if needed):

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

2.1.5. Pilot Action Calendar

Together with the development of the Pilot 1 content a number of Pilot actions and experiments are implemented in order to examine different aspects of the platform such as the user satisfaction, the feeling of togetherness, etc. An initial calendar of such activities can be found below. There might be additional actions taking place which are not foreseen at the moment, but will be reported in future versions of this document.

Table 1Pilot Actions Calendar

No	Date	Partner	Technology	Methodology	Network Configuration	Pilot Action	Type of Content	User profile	Location	Description
1	October 2017	TNO	WebRTC	Questionnaire	Offline	Questionnaire on demo platform	Demo Content	End-user	VR Days 2017	A number of questionnaires was handed out to end-users who tried the socialVR experience platform developed by TNO
2	November 2017	CWI	PCC	Objective metrics	-	Experiment CWI-1	-	End-user	Amsterdam Lab Node	QoE objective metrics of point cloud compression (24 Users)
3	April 2018	TNO	WebRTC	Questionnaire	Offline	Questionnaire on demo platform	Demo Content	End-user	Dutch Financial institution	A number of questionnaires was handed out to end-users who tried the socialVR experience platform developed by TNO
4	February 2018	CERTH	TVM	-	Online	Experiment CERTH-1	-	-	Thessaloniki Lab Node	Real-Time distribution of time varying meshes (no users involved)
5	February 2018	CERTH	TVM	-	-	Experiment CERTH-3	-	End-user	Thessaloniki Lab Node	interference between HMD and multiple depth-sensing (2 users)
6	May 2018	CWI	PCC	-	-	Experiment CWI-2	-	End-user	Amsterdam Lab Node	Design guidelines for Social VR (10 + 52 users)
7	June 2018	TNO	WebRTC	added value questionnaire	-	Questionnaire on demo platform	Demo Content	End-user	MMSys 2018	A number of questionnaires was handed out to end-users who tried the socialVR experience platform developed by TNO (25 users)
8	June 2018	CWI	PCC	Questionnaire	Offline	Focus group	Demo Content	End-user	Amsterdam Lab Node	A focus group on social VR experience
9	July 2018	CERTH	TVM	-	Online	Experiment CERTH-2	-	End-user	Thessaloniki Lab Node	Real-Time distribution of time varying meshes part 2 (5 users)
10	July 2018	CWI	PCC	-	-	Experiment CWI-3	-	End-user	Amsterdam Lab Node	social VR Ground Truth (32 users expected)
11	July 2018	Artanim	-	-	-	Experiment Artanim-1	-	-	Artanim Lab node	Impact of movement animation of the virtual body parts to presence

12	July 2018	Artanim	-	-	-	Experiment Artanim-2	-	-	Artanim Lab node	Impact of movement animation of the virtual body parts to co-presence
13	July 2018	CERTH	TVM	-	-	Experiment CERTH-4	-	-	Thessaloniki Lab Node	HMD removal part 2
14	September 2018	i2CAT	TVM	Subjective+objective metrics	Online	User tests on Pilot 1	Pilot content	End-user	Barcelona Lab Node	User tests will be run with end-users participating in the Pilot 1 experience.
15	September 2018	TNO	WebRTC	demonstration	Offline	Pilot 1 demonstration	Pilot content	Professional-Users	IBC2018	Demonstration of the Pilot 1 to the exhibition visitors of IBC 2018
16	October 2018	TNO	WebRTC	demonstration	Offline	Pilot 1 demonstration	Pilot content	Professional-Users	VR Days 2018	Demonstration of the Pilot 1 to the exhibition visitors of VR Days 2018
17	December 2018	TNO	WebRTC	demonstration	Offline	Pilot 1 demonstration	Pilot content	Professional-Users	ICT 2018	Demonstration of the Pilot 1 to the exhibition visitors of ICT 2018
18	January/ February 2019	CWI/TNO	TVM/PCC/WebRTC	Professionals questionnaire	-	Interview on Pilot 1	Pilot content	Advisory board member	N/A	Interview the advisory board members who regarding Pilot 1
19	January 2019	TNO	WebRTC	demonstration	Offline	Pilot 1 demonstration	Pilot content	End-user	Sundance Film Festival 2019	Demonstration of the Pilot 1 to the exhibition visitors of Sundance Film Festival 2019
20	January 2019	TNO	WebRTC	Questionnaire	Offline	Pilot 1 Questionnaires and interviews	Pilot content	End-user	Sundance Film Festival 2019	Questionnaire to end-users participating in the demonstration of the Pilot 1 to the exhibition visitors of Sundance Film Festival 2019

2.2.Pilot 2

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

2.3.Pilot 3

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

3. Software Platform Description

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

In Pilot 1 the end-users will participate in a virtual world scenario as this is described in section 2.1.1. of this document. Below we present a collection of the User Scenarios describing Pilot 1. After that we analyse the user scenarios and define the implicated Use Cases. Next we extract and present the functional and non-functional requirements based on the use cases. Last, we depict the system architecture that will support the functionality realized in Pilot 1.

3.1.User Scenarios

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

For the scenarios that are preceded by an asterisk, it is uncertain whether they will be implemented for Pilot 1 because the development work is focused on creating seamless experiences for each one of the available configurations. However, all of the scenarios will be implemented and presented in Pilot 2. The same rule applies to the connected use case descriptions (Section 3.2.2)

The User Scenarios forming Pilot 1 are:

Controls

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

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

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

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

Experience

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

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

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

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

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

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

3.2. Use Cases

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

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

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

3.2.1. General Use Case

Title: Pilot 1

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

Actors: 2 users located at two distinct geographical locations.

Precondition:

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

- The locations in which the set-up has been deployed comply with the system requirements required by VR-Together for Pilot 1.

Post-condition: Each end-user managed to:

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

Primary path:

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

3.2.2. Detailed Use Cases

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

Table 2 Use Case: profile creation/edit

Title	A user creates or edits a profile on the platform
Actors	The end-user
Brief Description	The user creates or edits a profile on the VR-Together platform so that his data is saved.
Precondition	• The end-user has accessed the VR-Together platform satisfying all the related requirements (set-up, connection, hardware, etc.) • The end-user wants to create a profile for accessing the experience OR The end-user wants to edit an existing profile.
Post-condition	• The end-user has created/edited a profile that includes information related to his account. The information that is saved had not been decided but will probably include: o Username o Email o User's height

	o Default self-representation configuration o Default content format o Default viewer mode
Primary Path	1. The end-user has accessed the VR-Together platform 2. End-user selects to create a profile 3. In the "Create Profile" menu the user introduces: a. Username b. Email c. Height d. Other configuration 4. End-user is finished with the data input 5. Select "Save configuration" 6. Exit
Alternative Path	1. End-user selects to edit a profile 2. In the "Edit Profile" menu the user introduces: a. Username b. Email c. Height d. Other configuration 3. End-user is finished with the data input 4. Select "Save configuration"

Table 3 Use Case: Self-representation configuration

Title	Self-representation configuration
Actors	The end-user
Brief Description	The end-user selects the self-representation format in a session of VR-Together
Precondition	• The end-user has accessed the VR-Together platform satisfying all the related requirements (set-up, connection, hardware, etc.) and able to initiate an active session. • The end-user wants to select the configuration of the self-representation format
Post-condition	• The end-user has configured the self-representation format and can view the changes realised within the virtual space of the VR-Together platform.
Primary Path	1. The end-user has accessed the VR-Together platform and is able to initiate a session. 2. End-user selects to configure the self-representation format

	3. The end-user is presented with the number of available options a. TVM b. Point Cloud c. 2D 4. End-user selects one of the options 5. Select “Save configuration” 6. Exit
Alternative Path	In this Use case there is no alternative path as the platform cannot operate without a user-representation configuration.

Table 4 Use Case: End-user create/join room

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

Table 5 Use Case: End-user session exit

Title	End-user session exit
Actors	The end-user
Brief Description	The end-user exits from the active session in which he/she participates
Precondition	• The end-user has accessed the VR-Together platform satisfying all the related requirements (set-up, connection, hardware, etc.) • The end-user is participating on an active sssion
Post-condition	• The end-user does not participate in any active sessions within VR-Together
Primary Path	1. The end-user is participating in an active session within VR-Together 2. The end-user selects to exit the active session 3. A dialog window confirms asks the user for confirmation to exit the active session. 4. The user exits the active session and is found in the starting menu of the VR-Together platform.
Alternative Path	In this Use case there is no alternative path as the platform always allows a user to exit a session.

Table 6 Pilot 1 content play-out

Title	Pilot 1 content play-out
Actors	The end-users, Non-Live content
Brief Description	The end-users are found in the virtual room created as mentioned in Table 3. And the self-user representation is configured for each one of them, as mentioned in Table 2. The content playout begins for both users to live the experience described in the Pilot 1 plot (2.1.1)
Precondition	• The end-users have accessed the VR-Together platform satisfying all the related requirements (set-up, connection, hardware, etc.) • The end-users are in the same virtual room • The end-users have selected a self-representation configuration

Post-condition	• The end-users have viewed the content of the Pilot 1 plot.
Primary Path	1. The end-users are in the same virtual room 2. The virtual room Session logic manager initiates the content playback. 3. The end-users view the playback of the content and interact with each other 4. The content playback finishes 5. The end-users are free to exit the virtual room or continue the interaction in it.

3.3.Requirements Specification

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

3.3.1. Requirements gathering methodology

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

3.3.1.1. Requirements Gathering Techniques

In the VR-Together project we employ a number of techniques in gathering the requirements that will define the project's end-results and features of the pilots to be developed. All possible requirements are gathered and classified accordingly by examining the attributes presented in the following sections of this chapter (3.3.1). It is important to mention that all the gathered requirements are judged upon the compliance with the core objectives of VR-Together, as they are described in the Grant Agreement (Section 1.1.2 Part B).

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

- <u>Document Analysis</u>: We identify and extract the requirements from generated in the VR-Together project, such as deliverables, reports, etc. As a first example of this we identified and extracted the requirements included in the Grange Agreement document.
- <u>Focus Groups:</u> Group gatherings of potential end-users who will be asked to perform a specific task are expected to generate requirements for the VR-Together platform. After each focus group gathering, participants are asked to give their feedback in a number of different aspects of the software platform itself as well as the experience. The collected feedback is analysed in an effort to determine additional requirements as well as refine and validate the existing ones.
- <u>Interviews:</u> By conducting interviews of end-users and other important stakeholders we identify the expectations that VR-Together should meet. The expectations that align

with the objectives stated in the grant agreement are translated to requirements and captured in the software requirements matrix. An indicative example of the requirements that will be gathered with this technique are the interviews to be conducted with the Advisory Board members.

- Surveys/Questionnaires: Carefully designed surveys help in acquiring a large number of user feedback in a short time as well as in a structured an easily comparable way. The design of the surveys includes questions where options are in the level of agreement/disagreement or rating of an argument. An example questionnaire can be found in Annex I.
- Other techniques: Depending on the occasion a number of additional techniques could be used in order to generate requirements that would help in developing a higher quality end platform. These additional techniques could be brainstorming sessions, requirements gathering workshops, etc.

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

3.3.1.2. Types of requirements

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

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

3.3.1.3. Prioritization of requirements

The description of a requirement must contain one of the following terms to define the prioritisation of the requirement: “must”, “should”, “could” or “won’t”. The definition of these terms has been adopted from the MoSCoW prioritisation. Negative requirements such as “should not” and “shall not” are omitted, as they are not common in software development.

MoSCoW [1] defines the terms as follows:

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

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

WON'T	***WON'T*** requirements are either least-critical or not appropriate at that time.

3.3.1.4. System component of requirements

The user requirements are based on the user scenario compilation described in 3.2, separated depending on the component of the VR-Together platform that they are related to.

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

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

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

The requirements matrix can be found in Section 3.4 of this document.

3.3.2. VR-Together User Profiles

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

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

Each of these three types of users has different scope regarding the use of the VR-Together platform and for that reason they have a set of associated requirements as well as available functionalities. Below we give a description for the profile of each one of the different types. It is important to notice that in the requirements matric presented in section 3.4 the User profile related to each requirement is not specified due to the flexibility of VR-Together in supporting different modes of functionality depending on the user's intention.

3.3.2.1. End-User

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

3.3.2.2. Administrator

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

3.3.2.3. Researcher

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

3.3.3. Reference documentation

The VR Together experience makes use of the following standards:

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

3.3.4. Assumptions and dependencies

All the components in which VR-Together platform Pilot1 is based or is dependent from are properly described in D3.1 and D3.2.

3.3.5. Interface Requirements

Web player interface. Content consumption and social interaction will be accessed through a web application. The objective is to explore social VR cases 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 hardware deployments and higher rendering capabilities.

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

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

3.4. VR-Together Software Requirements Matrix

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

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

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

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

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

3.4.1. Requirements for Pilot 1

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

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

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

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

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

*<u>Typology of contents addressed:</u> An intimate music concert seems an ideal starting point to demonstrate VR-Together's innovative media format. It is a good opportunity to show how the VR-Together works for offline produced content**.** The goal is to demonstrate that the orchestrated delivery of the VR-Together media format, combining several video sources, point*

cloud and 3D mesh representations will improve closeness with the musicians and with at least 2 distant end-users. Particular care will be taken to integrate facial expression within the production pipeline, i.e. how we will capture the photorealistic 3D actors in costume. For example, uses 108 cameras to capture the actors' performance, costumes, facial expressions and the result is a stream-able 3D model with appropriate facial expressions. This also applies to lighter methods, which are more affordable and portable. For example, uses 4 Kinect sensors and a short automatic calibration process. Industrial methods capturing actor facial MoCap performance using marker-less methods and pre-rigged models will also be considered. Different combination of methodologies and technologies will be studied to deliver the best possible balance between visual quality and cost efficiency in content production."

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

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

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

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

3.4.2. Requirements for Pilot 2

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

"*Pilot 2. Live news. We will demonstrate the live production of multi-source immersive content. We will study the conditions which maximize the connection between the audience and the news. Numerous benefits for cost-effective production efficiency will be derived from introducing live processing constraints. Quantitative measures comparing the benefits and costs of introducing offline processing steps will be sought. To realize this scenario, we foresee the creation and demonstration of 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. "

3.4.3. Requirements for Pilot 3

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

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

in SIVE will become essential to guide this process. Regarding the processing of interactive plots in SIVE, we will use tools readily available from previous research initiatives by the partners within the consortium. The main challenge to maintain place illusion and plausibility is to render credible interactivity within the experience. We will address how to integrate the user input with the events being depicted within the immersive virtual environment. The goal will be to show to what extent and how a fiction scenario can be rendered in VR, while still allowing the users immersed in the scene to intervene actively in the scene being broadcasted within the shared virtual reality experience (and thus, preserving the feeling of being there together).

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. "*

3.4.4. Experimentation requirements

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

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

3.4.5. Requirements Specification Table

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

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

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

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

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

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

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

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

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

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

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

4. Architecture

In the following part we describe the Architecture of the VR-Together platform both from the view of the Software platform (4.1) as well as that of the Hardware setup (4.2). Last we examine the architecture implications in viewing VR-Together as a complete software product.

4.1. Software architecture

This section gives an initial architecture description of the different software components used in the VR-Together platform as well as the interaction between them.

Figure 20 depicts the high-level architecture of the VR-Together system. The VR-Together platform is described in a traditional production to consumption chain: audio-visual information flows from capturing to playout are portrayed together with the additional functional components. The modules and components, as seen in the Figure 20, form the Software Platform of VR-Together.

Next, we present an extended and detailed software component architecture description referring to Figure 20.

In the following architecture description, we refer to:

- A "**component**" as *a conceptual entity related to a general task within the end-to-end communication system*. We identify the following five components/general tasks:
  - Capturing
  - Encoding and encapsulation
  - Delivery
  - Orchestration
  - Play-out
- A "**module**" as *a building block that actually performs a specific technical task*. To perform the general task described by a component, multiple modules are needed. The modules within each component are listed and described hereafter

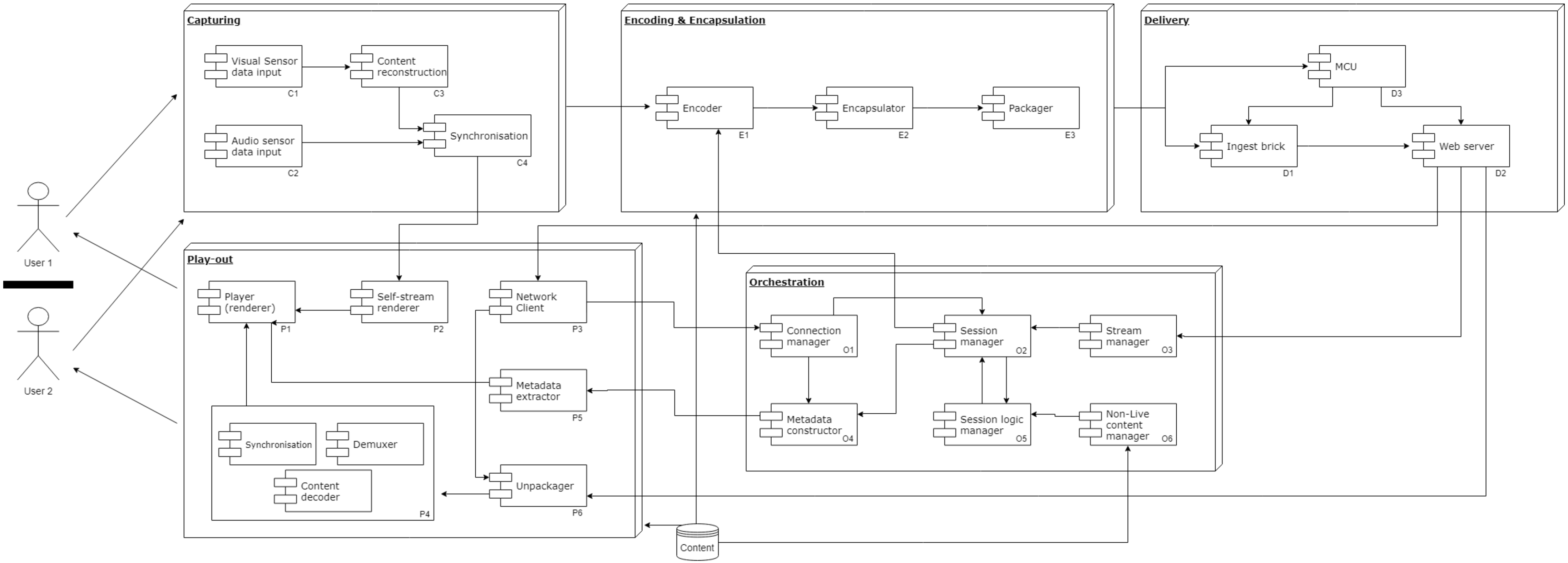

Figure 20. Component diagram for platform

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

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

4.1.1. Capturing component

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

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

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

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

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

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

- **C4 - Synchronisation**: this module receives the audio and video frames and their timestamps and aligns them temporally, in order to output temporally synchronised audio-visual tracks. The process performed in this module additionally facilitates the synchronisation which will be required later on in the "Play-out" component (between the content streams of different participants).

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

4.1.2. Encoding and Encapsulation component

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

Input: timestamped visual stream (i.e. TVM or PC raw stream) / timestamped audio stream. Timestamps are set according to the Platform Clock.
Output: encoded visual stream (i.e. TVM or PC encoded stream) / encoded audio stream (Example: .ply file for an encoded PC and .aac file for encoded audio)

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

Input: encoded visual stream (i.e. TVM or PC encoded stream) + encoded audio stream (Example: .ply file for an encoded PC and .aac file for encoded audio)
Output: audio-visual file (e.g. mp4, webM, etc.) including synchronized audio and visual tracks for one user

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

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

4.1.3. Delivery component

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

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

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

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

- **D3 - Multi Control Unit (MCU):** Depending on the active session there might be 2 or more participants in one virtual environment. In the case where the participants are more than 2 the "Multi Control Unit" (MCU) module is activated. The MCU is responsible for combining the visual and audio inputs arriving from multiple sources. The inputs are blended into a common synchronised stream. The result of the MCU is then encoded(??), encapsulated and packaged before reaching the "Delivery" component.

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

4.1.4. Orchestrator component

In architecture view, that is show in Figure 20, one of the central components is the Orchestrator component, which provides all clients with the information necessary to initiate a communication session of end-users over the VR-Together platform. This includes the discovery of available rooms, VR room configurations, pointers to content sources, other clients in a session and the capture sources. The Orchestrator is responsible for signalling synchronization data between 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 VR-Together 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 content streams.

An administrator can control and force session control data through a corresponding interface, e.g. in order to facilitate user experiments and demos. It is important to note that the Orchestrator component controls but does not process media streams. For all control data all clients (regardless of the type) have a common interface to the orchestrator. However, clients might have different interfaces to content (based on the content type and content server). In this way a client may retrieve one or multiple media content streams from one or multiple Content Servers. The URIs to the streamable content are provided in the session's metadata. 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 its own capture component integrity both by mater of hardware as well as software. In content stream transmission, from client to client, a processing nodes will be used.

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

- **O1 - Connection Manager**: The users' connections to the VR-Together platform are managed and maintained by this component which then transfers the relevant information to the "metadata constructor".

<u>Input</u>: user connection information
<u>Output</u>: user connection pointers

- **O2 - Session manager**: This module controls and is aware of all the information regarding a session (e.g., how many users are joining the session, which is the non-live content used, etc.) and its corresponding configuration.

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

<u>Input</u>: Stream information
<u>Output</u>: Stream information pointers

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

<u>Input</u>: Virtual experience configuration

Output: Metadata descriptor file

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

Input: Platform configuration
Output: Configuration pointers

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

Input: Session configuration
Output: Resource pointers + metadata configuration

4.1.5. Play-out component

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

Input: Metadata descriptor value
Output: Content play-out

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

- **P3 - Network client**: This module is responsible for establishing the connection with the web server that provides the content to the "Play-out" component.

- **P4 - Demuxer:** This module splits the audio and video of the content received from the web server.

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

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

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

- **P4 - Synchronisation**: This module refers to the time-stamp alignment that needs to happen in the received content in order to correctly render the visual and audio streams according to the universal clock's timestamp.

- **P5 - Metadata extractor:** This module parses the metadata file provided by the "Orchestrator" and extracts all the necessary information in order to facilitate rendering of the content.

Input: Metadata descriptor file
Output: Content and configuration values

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

4.2.Hardware architecture

This section provides the hardware architecture of the VR-Together platform, as designed for the case of Pilot 1. In the diagram show in Figure 21 we can see the hardware components that form the VR-Together platform, as well as the expected location of the software components and modules, as they are described in 4.1 In the next part we examine each hardware component separately and provide a relevant description.

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

4.2.1. End-user Set-up

Each end-use needs to be using a capturing setup and a playout setup in order to be able to use the VR-Together software platform. The capturing setup is composed of:

- Visual sensor capturing: the set-up includes 4 visual sensors responsible for capturing the user's representation that will be encoded and projected in the virtual rooms of the VR-Together platform. The sensors are the *Microsoft Kinect for Xbox One*. Due to the announced discontinuation of the product it is currently being examined the possibility of a switch to *Intel RealSense D415*.
- Audio sensor capturing: for the capturing of the audio of the user the sensor to be used is the microphone that is embedded on the head mounted display that is used by the end-user.
- Local visual sensor processing: each one of the visual sensors needs to be connected to a separate terminal that processes the raw input stream. In the current set-up of VR-Together the terminals used for this purpose are *Intel NUC D54250.*
- Local audio-visual stream processing: in a capturing set-up environment a central node/terminal is responsible for collecting the separate visual streams together with the audio stream and processing them accordingly. The hardware device used for this purpose is not specific but the most important minimum hardware requirements that it needs to comply with are:
    - 16GB of Ram
    - I7 or equivalent Central Processing Unit
    - Separate graphics processor with 8GB of memory
- Head mounted display: the playout devices to be used by the end-users are commercially available products. In our set-up we are using the *Oculus Rift Consumer Version 1*.

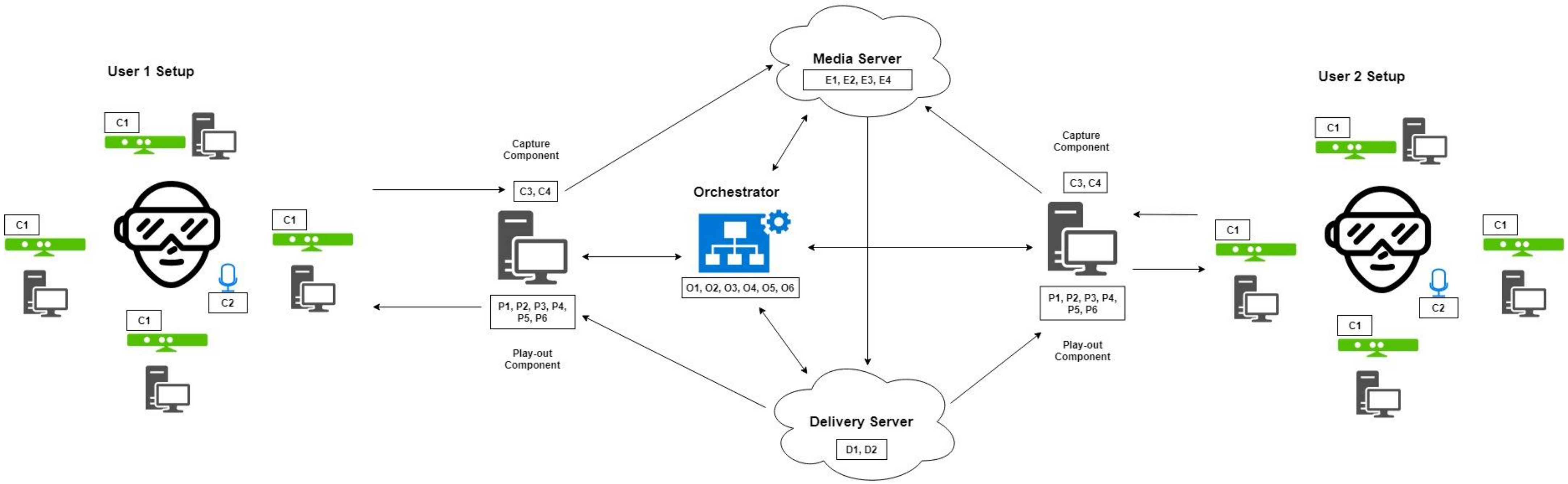

Figure 21. Hardware architecture

4.2.2. Processing Servers Set-up

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

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

Table 7 - Cloud Server Requirements

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

The values mentioned above are subject to changes depending on the requirements gathered during the VR-Together platform evaluations and experiments. That is because various aspects of the experience, such as latency or projection quality, might require higher processing capabilities and therefore render the above information out-dated. The currently shown values have been tested against the components as they are delivered in D3.1.

4.3. VR-Together as a Software product

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

5. User Lab

VR-Together 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.

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. The calendar of the actions included can be found in the Pilot Action Calendar (Section 2.1.5)

5.1.Advisory Board

The advisory board of VR-Together 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 full list of the advisory board members can be found at http://VR-Together.eu/advisory-board/

The current list of the advisory board members includes:

- Morgan Bouchet from Orange
  - https://www.linkedin.com/in/morganbouchet/
- Travis Rice from Lens Immersive
  - https://lens-immersive.com/about
- Nils Duval, VR Consultant
  - https://www.linkedin.com/in/nils-duval/
- Ricard Gras, Co-founder Timepath S.L.
  - http://VR-Together.eu/ricard-grass/
- Scott Ross, Digital Media Pioneer (latent)
  - https://www.linkedin.com/in/scottross

- Dolf Schinkel from KPN (latent)
  - https://www.linkedin.com/in/dolfschinkel/
- Sebastian Sylwan from Felix & Paul Studios
  - http://www.imdb.com/name/nm4492489/

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 or other technical conferences around Europe in which VR-Together is present.
- Artistic members in a film or other artistic festival around Europe where VR-Together is present.

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.

5.2.User Lab Nodes

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 implement a distributed User Lab that includes three lab nodes with fully equipped facilities that can support the VR-Together platform. There will additionally be built lab nodes with partially equipped facilities what will support part of the VR-Together platform.

The user labs will provide a full environment to run the field trials of the pilots. They will have 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, 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 VR-Together will have three main lab nodes strategically located in:

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

VRTogether hub

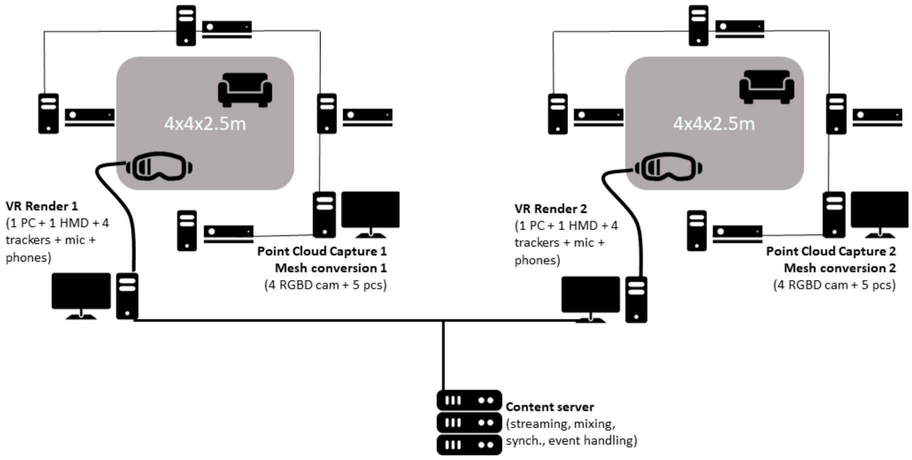

Figure 22. Schematic View of a VR-Together hub.

In addition to the hubs, several partners of the project will create dedicated user labs with a partial infrastructure of the full fledge VR-Together 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 VR-Together 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 Annex II a description for each one of the main Lab nodes can be found.

5.3. Experiments

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

- Assessment of **technology**, such as HMD removal or content distribution: 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**[2], 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;
- **Psychological** dimension of the VR-Together 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.

5.4. List of experiments

This sub-section provides an overview of the experiments initially considered in the project, and serves as a plan for project activities in terms of piloting and evaluation, as part of WP4 tasks. Further information regarding experiments will be provided in future versions of this document and WP4 documents. The following experiments list accompanied by the complete description of each experiment can be found in Deliverable 4.2

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

5.4.1. Technology Evaluation

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

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

5.4.1.1. CWI-1

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

5.4.1.2. CERTH-1

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

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

5.4.1.3. CERTH-2

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

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

5.4.1.4. CERTH-3

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

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

- Kinect for Xbox One
- Intel RealSense D415

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

- HTC Vive
- Oculus Rift DK2

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

5.4.1.5. CERTH-4

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

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

5.4.2. User Experience Evaluation

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

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

5.4.2.1. ARTANIM-1 and ARTANIM-2

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

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

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

5.4.2.2. CWI-2

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

Research questions

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

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

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

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

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

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

5.4.2.3. CWI-3

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

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

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

5.4.3. Feedback from Professionals

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

In particular, we report:

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

5.4.3.1. TNO-1: Initial use-case study

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

Research questions:

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

Hypothesis:

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

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

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

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

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

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

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

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

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

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

In this experiment the goals were the following:

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

This formed the following research questions:

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

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

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

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

Research questions

- RQ1: When is VR expected to take off?
- RQ2: What are the most important VR applications?
- RQ3: Which content is suitable for VR?
- RQ4: Which content is suitable for experiencing it together in VR?
- RQ5: Which aspects are important for shared VR experiences?

Hypothesis

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

6. Conclusions

In this document we have presented an extended description of the software and hardware dimensions of VR-Together, at the point in time prior to the release of Pilot 1. The document presents the Plot, storyboard and production descriptions related to Pilot 1. Having this as a basis, an extended software platform description is unravelled. It consists of an elaborated analysis of the User scenarios and use cases, a software requirements specification and all related requirements information.

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

Finally, we have outlined the VR-Together User Labs, and the different experimental work involved in the preparation of the pilots and the validation of the project requirements. 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.

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

2/6/2018 VR Together

VR Together

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

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

Experiences with VR

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

- ◯ Yes
- ◯ No

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

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

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

6. *Mark only one oval per row.*

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

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

____
____
____
____
____

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

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

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

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

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

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

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

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

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

Demographic questions

12. Gender
Mark only one oval.

◯ Female
◯ Male
◯ Other: ______________________

13. Age
Mark only one oval.

◯ Less than 18
◯ Between 18 and 30
◯ Between 30 to 45
◯ Between 45 and 65
◯ More than 65
◯ Other: ______________________

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

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

15. Email

____

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

Powered by

8. ANNEX II: VR-TOGETHER LAB NODES

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

8.1. Artanim Lab Node

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.

8.2.CERTH Lab Node

CERTH has two available rooms (studios) for the user lab, one in Building A of dimensions 4.5m x 4.5m x 2.5m, and one in Building B (see Figure 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).

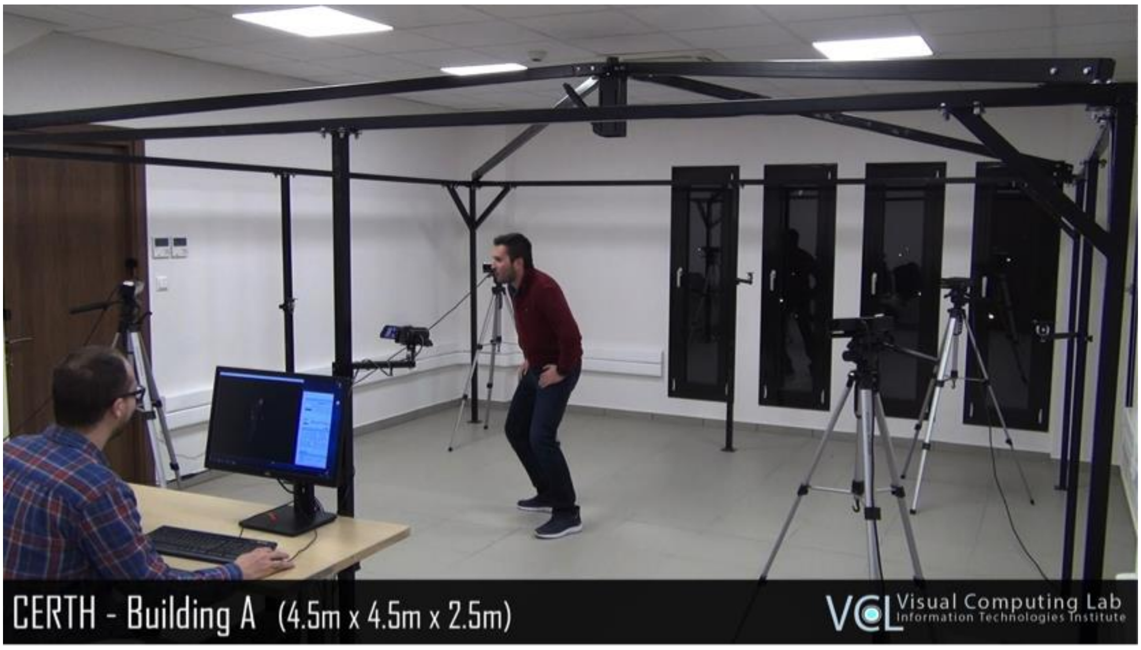

Figure 25. CERTH's User Lab.

8.3.CWI Lab node

CWI has two available rooms: Pampus (see Figure 26) and the QoE Lab (see Figure 27). 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 VR-Together.

8.4.i2CAT Lab Node

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

Space Setup

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

Hardware for consumer setup:

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

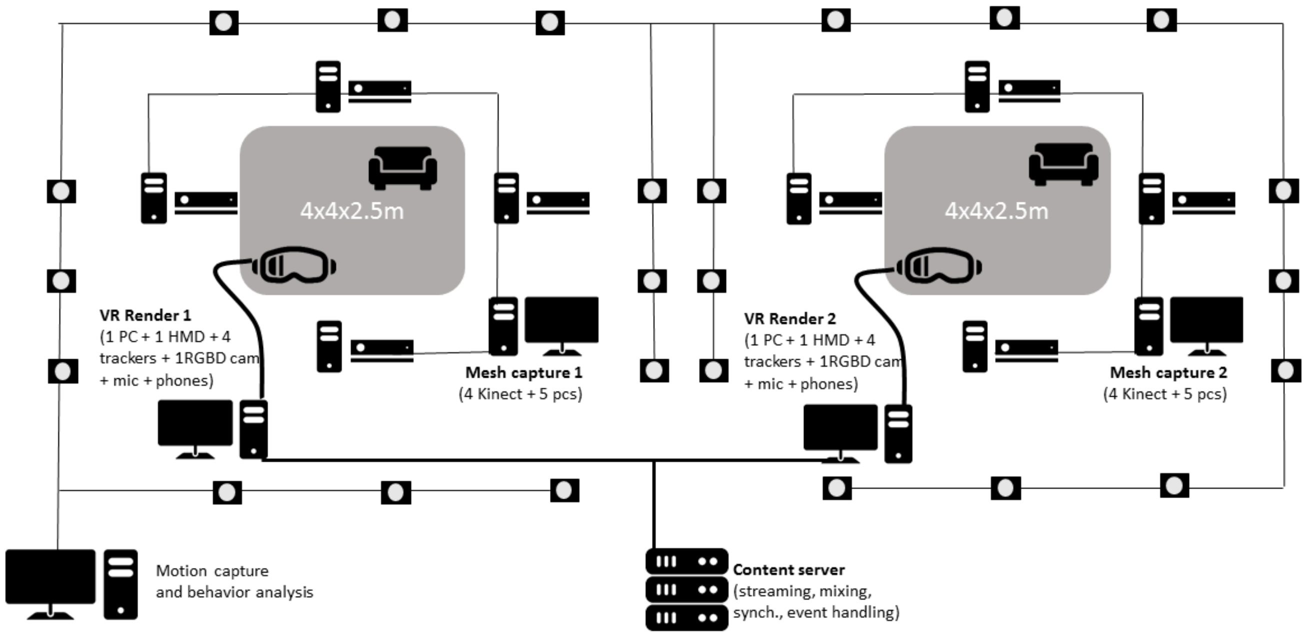

Figure 28 i2CAt lab infrastructure