Deliverable

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

D4.6-Technical Report on Third Pilot.

Revision: 3.0
Authors: Mario Montagud (i2CAT) and Pablo Cesar (CWI)
Delivery date: M39

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

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

Abstract: This document provides a comprehensive overview on the activities around the third Pilot of the project. It provides information about the evaluation methodologies and metrics developed by the project, the preparation work in the form of experiments (technological and with users), and a detailed report on the Pilot. Moreover, information about relevant activities such as the involvement of the local advisory board and the establishment of connected user labs is detailed.

REVISION HISTORY

Revision	Date	Author	Organisation	Description
0.1	24-11-2020	Pablo Cesar	CWI	Structure and Table of Contents
0.2	9-12-2020	Nacho Reimat	CWI	Section 6.4.3
0.3	10-12-2020	Various authors	CWI	Section 6.4 completed
0.4	11-12-2020	Various authors	CWI, TNO	Sections 2.2 and 4.2
0.5	14-12-2020	Varios authors	CWI, theMo	Sections 2.1, 6.1
0.6	15-12-2020	Various authors	CWI	Sections 2.2 completed, 2.4, 3, 5.3.1, 6.2, 6.3
0.7	16-12-2020	Various authors	I2CAT, CERTH, TNO	Sections 5.3.2, 5.3.3, 5.3.4, 5.3.5
0.8	17-12-2020	Various authors	CWI, CERTH, i2CAT	Sections 2.3, 5.2.1
0.9	21-12-2020	Various authors	CWI, CERTH, i2CAT	Sections 1, 4.1, 6.3.6, 7
1.0	22-12-2020	Pablo Cesar	CWI	Consolidated draft version for review
1.5	23-12-2020	Tom De Koninck	TNO	Internal review
2.0	23-12-2020	Pablo Cesar	CWI	Consolidated version after the review
2.5	27-12-2020	Mario Montagud	I2CAT	Sections 5.1, 5.2.2, 5.2.3
3.0	31-12-2020	Pablo Cesar and Mario Montagud	CWI, i2CAT	Final version, ready to be submitted

EXECUTIVE SUMMARY

The project has successfully accomplished the planned tasks towards the execution of the third Pilot. The coordination between WP2 and WP4 facilitated this process. In particular, the activities that have taken place in the third year of the project include:

- Focus groups and interviews with the local advisory board, including 9 companies and stakeholders
- Definition, development, and capturing of objective and subjective metrics for social VR scenarios
- Extensive testing of the two existing connected labs between France and the Netherlands (web client pipeline), and between Spain, Greece and the Netherlands (native pipeline)
- Execution of nine pilot actions paving the way towards Pilot 3, including benchmarking of existing social VR solutions and exploration of other case studies
- Full content creation, production and post-production, of high-quality assets for Pilot 3 (see D4.5 and D4.7)
- Pilot 3 deployment and evaluation with 48 end users in Amsterdam
- Content deployment and demonstration to professionals in the 2020 IMX and MMsys conferences (with demonstration awards) and in 2020 VRDays (showcasing the first volumetric video conferencing over a public 5G network)

Section 2 focuses on the metrics and evaluation methods. During the first year, the project created a new set of metrics and protocols for evaluating social VR. The impact of them may go beyond the project, since it can be become the de-facto standardized manner for evaluating a new genre of experiences: social VR. The protocol and metrics include both quantitative and qualitative aspects; a set of objective metrics based on the behaviour of the user, focusing on neck rotation, body movement, etc.; and performance metrics for profiling the system aspects. In year 2, these evaluation methodologies were refined based on deeper analysis and extended. In year 3, the metrics have been consolidated and used across pilot actions and the pilot. In particular, the project contributes a validated protocol to evaluate social VR, including a questionnaire, objective metrics regarding user behaviour and performance metrics. Moreover, in this year new quality metrics for evaluating point clouds (full reference and reduced reference) have been proposed, developed, and published, leading the path in this new area of research.

Section 3 provides information about the local Advisory Board. In the second year, the project took the decision of involving local professionals, closer to each partner, instead of having a global committee. This local activity complements the effort on the Joint Business Clinics followed in WP5 for better defining the value proposition and the business plan. As a result, 14 professionals representing nine companies participated in a focus group in Amsterdam, providing valuable insights about the potential of the technology and enabling the creation of a hub in Amsterdam around social VR.

Section 4 reports the extensive evaluations that have been done between the connected user labs for both the native and the web-based pipelines. The native pipeline connects Spain, Greece, and the Netherlands and incorporates several types of volumetric representations (TVMs and point clouds) over a number of transmission protocols (RabbitMQ, socket.io, DASH). The web pipeline connects France and the Netherlands.

Section 5 reports on nine experiments, or pilot actions, that have taken place in the third year of the project in order to help the project constructing the Pilot and testing other platforms and use

cases. In particular, three types of experiments have taken place: benchmarking of social VR platform, experiments for evaluating the end-users and the environment representation, and assessment of the underlying technology.

Finally, Section 6 details the pilot evaluations with both end users and professionals. Pilot 3 run in Amsterdam with 48 end users, where information about their experience (subjective and objective) was gathered. The chapter details the Pilot and reports the results. In addition, the project was showcased in three major events: ACM IMX 2020 conference, ACM MMSys 2020 conference and VRDays 2020. The last one premiered the usage of volumetric video over a public 5G network, exploring the area of healthcare and remote consultation.

The project has achieved the objectives foreseen, even during a global health pandemic, exploring how volumetric video can be used for communication and collaboration, providing new protocols and metrics, creating connected user labs between different locations, running the three pilots detailed in the project proposal, and validating all the work with end users and professionals. The expected KPIs have been adequately addressed.

			Year 1	**Year 2**	**Year 3**	**Total**
Action	**Description**	**KPI**				
Focus group with professionals	Requirements gathering	20-30	0	31	13	44
Focus group with end users	Requirements gathering	20-30	0	15	0	15
Lab validation with professionals	Continuous evaluation of functionality	100-150	181	199	390	770
Lab validation with end users	Continuous evaluation of functionality	100-150	216	166	56	438
Industrial fair evaluations	Showcase of results in industrial events	1000-2000	58000	61000	1800	120800
Pilots with end users	Formal evaluation of Pilots 1, 2, and 3	150-300	30	44	100	174
Advisory board	Market evolution, competitors, prosumer feedback…	5-10	----	----	----	11

The table shows:

- Focus group with professionals: invited to the user labs of the partners, showcasing the technology
- Focus group with end users: invited to the user labs of the partners, showcasing the technology
- Lab validation with professionals: showcase of the technology and results in scientific events; the provided number is the total number of participants, where maybe a 40% of them experienced the demonstration
- Lab validation with end users: pilot activities, evaluating the technology and the system
- Industrial fair evaluations: showcase of the technology and results in large fairs like IBC or VRDays. The provided number is the total number of participants, where maybe a 10% of them experienced the demonstration
- Pilots with end users: the formal evaluation of the Pilots 1 (D4.2), 2 (D4.4), and 3 (D4.6)
- Advisory board: the advisory board of the project

CONTRIBUTORS

First Name	Last Name	Company	e-Mail
Bart	Kevelham	Artanim	bart.kevelham@artanim.ch
Henrique	Galvan Debarba	Artanim	henrique@artanim.ch
Prodromos	Boutis	CERTH	prod@iti.gr
Kyriaki	Christaki	CERTH	kchristaki@iti.gr
Petros	Drakoulis	CERTH	petros.drakoulis@iti.gr
Paschalina	Medentzidou	CERTH	medentzidoup@iti.gr
Argyris	Chatzitofis	CERTH	tofis@iti.gr
Pablo	Cesar	CWI	P.S.Cesar@cwi.nl
Jack	Jansen	CWI	Jack.Jansen@cwi.nl
Kinga	Ławicka	CWI	kinga.j.lawicka@gmail.com
Jie	Li	CWI	Jie.Li@cwi.nl
Yanni	Mei	CWI	Yanni.Mei@cwi.nl
Nacho	Reimat	CWI	nacho.reimat@cwi.nl
Shishir	Subramanyam	CWI	s.subramanyam@cwi.nl
Irene	Viola	CWI	irene.viola@cwi.nl
Ana	Revilla	The Modern Cultural	anarevilla@themoderncultural.com
Guillermo	Calahorra	The Modern Cultural	guillermo@themoderncultural.com
Ignacio	Lacosta	The Modern Cultural	lacosta@themoderncultural.com
Javier	Lajara	The Modern Cultural	javier.lajara@futurelighthouse.com
Mario	Montagud	I2CAT	mario.montagud@i2cat.net
Gianluca	Cernigliaro	I2CAT	gianluca.cernigliaro@i2cat.net
Sergi	Fernández	I2CAT	sergi.fernández@i2cat.net
Tom	De Koninck	TNO	Tom.dekoninck@tno.nl
Loic	Landais	VO	loic.landais@viaccess-orca.com
Vincent	Lepec	VO	vincent.lepec@viaccess-orca.com

CONTENTS

TABLES OF FIGURES AND TABLES

LIST OF ACRONYMS

Acronym	Description
6DoF	6 Degrees of Freedom
AI	Artificial Intelligence
CGI	Computer Generated Imagery
EC	European Commission
F2F	Face To Face
FoV	Field of View
GDPR	General Data Protection Regulation
HMD	Head Mounted Display
IPQ	Igroup Presence Questionnaire
MOS	Mean Opinion Score
PI	Presence/Immersion (PI)
PLCC	Pearson Linear Correlation Coefficient
QoE	Quality of Experience
QoI	Quality of Interaction
RGB	Red, Green and Blue
RGB-D	Red, Green, Blue and Depth
SRCC	Spearman Rank Correlation Coefficient
SSIM	Structural Similarity Index Measure
SSQ	Simulation Sickness Questionnaire
TVM	Time Varying Mesh
UX	User Experience
VE	Virtual Environment
VR	Virtual Reality

1 INTRODUCTION

1.1 Purpose of this document

The purpose of this deliverable is to provide the reader with a comprehensive overview on the activities around the third Pilot of the project. The document provides information about the metrics and methodologies developed by the project, the preparation work in the form of experiments (technological and with users), and a detailed report on the Pilot. Moreover, information about relevant activities such as the involvement of the local advisory board and the establishment of connected user labs is detailed. This is the third version of the deliverable (the first one is D4.2 and the second one is D4.4). Overall, the objectives of this WP have been met, with high-quality production of content (see D4.5) and successful pilot evaluations both with end-users and professionals.

1.2 Scope of this document

This document reports all the activities leading towards the third Pilot of the project, and the Pilot itself. This includes the metrics and methods that have been developed by the project to conduct experiments and evaluate the results, the experiments that have paved the way towards the Pilot, and the Pilot status.

1.3 Status of this document

This document is complete and complements the previous reports for Pilot 1 (D4.2) and Pilot 2 (D4.4)

1.4 Relation with other VRTogether activities

This document gathers the outputs of all the activities of WP4 during the third year (T4.1 to T4.3). D4.5 and D4.7, from the same WP, further detail the content production process. The work is as well closely related to WP2, responsible for requirements gathering, the user labs (and experiments), and the technical integration for the pilot infrastructure.

2 METRICS AND METHODOLOGIES

This section details the metrics and methodologies created during the project and used during the third year, applied to the different pilot actions and user evaluations. During the last year, all the metrics needed for evaluating social VR have been defined, developed and used. In particular, the project proposes the following ones (with more information in the subsequent sub-sections):

- User experience (subjective): questionnaires, interviews, observations for gathering end-user data
- User experience (objective): gaze, head direction, speech for gathering end-user data
- Technical performance (objective): bandwidth, jitter, frame-rates for profiling the system
- Added value (objective/subjective): protocols for identifying the added value of the proposed solutions

2.1 Subjective Evaluation

In the project, we have used a number of subjective questionnaires to measure the user experiences. These questionnaires include:

- Simulator Sickness Questionnaire (SSQ): in order to evaluate the effect of the experience on the user's wellbeing
- Social VR questionnaire: novel questionnaire developed and validated during this project for evaluating social VR in terms of presence, immersion, and connectedness
- Presence questionnaire: in order to better understand the sense of presence
- NASA Task Load Index: in order to evaluate the cognitive load
- Visual Quality Questionnaire: in order to assess the visual quality of the volumetric user representations and the virtual characters.

These questionnaires are presented next.

The **Simulator Sickness Questionnaire** was used in order to monitor simulator performance and measure users' levels of sickness symptoms. It was administered before the experimental condition to measure the baseline state of the participants and after to check their reactions to the simulation.

SIMULATOR SICKNESS QUESTIONNAIRE

Kennedy, Lane, Berbaum, & Lilienthal (1993)***

Instructions : Circle how much each symptom below is affecting you right now.

1. General discomfort	None	Slight	Moderate	Severe
2. Fatigue	None	Slight	Moderate	Severe
3. Headache	None	Slight	Moderate	Severe
4. Eye strain	None	Slight	Moderate	Severe
5. Difficulty focusing	None	Slight	Moderate	Severe
6. Salivation increasing	None	Slight	Moderate	Severe
7. Sweating	None	Slight	Moderate	Severe
8. Nausea	None	Slight	Moderate	Severe
9. Difficulty concentrating	None	Slight	Moderate	Severe
10. « Fullness of the Head »	None	Slight	Moderate	Severe
11. Blurred vision	None	Slight	Moderate	Severe
12. Dizziness with eyes open	None	Slight	Moderate	Severe
13. Dizziness with eyes closed	None	Slight	Moderate	Severe
14. *Vertigo	None	Slight	Moderate	Severe
15. **Stomach awareness	None	Slight	Moderate	Severe
16. Burping	None	Slight	Moderate	Severe

* Vertigo is experienced as loss of orientation with respect to vertical upright.

** Stomach awareness is usually used to indicate a feeling of discomfort which is just short of nausea.

Last version : March 2013

***Original version : Kennedy, R.S., Lane, N.E., Berbaum, K.S., & Lilienthal, M.G. (1993). Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology*, *3*(3), 203-220.

The aim of the **Social VR Questionnaire** was to collect users' subjective experiences regarding the quality of interactions (Q1-Q11), social connectedness (Q12-Q22), sense of presence and immersion (Q23-Q32) in the VR experience. At the end of the Social VR Questionnaire, we added three extra questions, asking users to rate the Pilot 3 content (Q33-Q35). In the project, we have use two types of them (see below): one for multiple user experience and one for two users experience.

Social VR Questionnaire (multiple users)

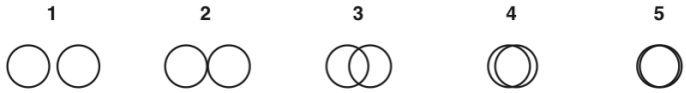
Please answer the questions, according to **your virtual movie experience before going to the virtual bedroom/kitchen.**

The scale of the following questions are from 1 to 5, representing the following meanings:
1 Strongly disagree 2 Disagree 3 Neutral 4 Agree 5 Strongly agree

Strongly disagree 1 2 3 4 5 Strongly agree	1	2	3	4	5
1. "I was able to feel other users' emotion while in the virtual movie."	☐	☐	☐	☐	☐
2. "I was sure that other users' often felt my emotion."	☐	☐	☐	☐	☐
3. "The experience of solving a crime in the virtual movie with other users seemed natural."	☐	☐	☐	☐	☐
4. "The actions used to interact with other users were similar to the ones in the real world."	☐	☐	☐	☐	☐
5. "It was easy for me to contribute to the conversation with other users."	☐	☐	☐	☐	☐
6. "The conversation with other users seemed highly interactive."	☐	☐	☐	☐	☐
7. "I could readily tell when other users were listening to me."	☐	☐	☐	☐	☐
8. "I found it difficult to keep track of the conversation."	☐	☐	☐	☐	☐
9. "I felt completely absorbed in the conversation."	☐	☐	☐	☐	☐
10. "I could fully understand what other users were talking about."	☐	☐	☐	☐	☐
11. "I was very sure that other users understood what I was talking about."	☐	☐	☐	☐	☐

Strongly disagree 1 2 3 4 5 Strongly agree	1	2	3	4	5
12. "I often felt as if I was all alone while in the virtual movie."	☐	☐	☐	☐	☐
13. "I think other users often felt alone while watching the content."	☐	☐	☐	☐	☐
14. "I often felt that other users and I were staying together in the same space."	☐	☐	☐	☐	☐
15. "I paid close attention to other users."	☐	☐	☐	☐	☐
16. "Other users were easily distracted when other things were going on in the real world."	☐	☐	☐	☐	☐
17. "I felt that solving the virtual crime together enhanced the closeness with other users."	☐	☐	☐	☐	☐
18. "Solving the virtual crime together created a good shared memory between me and other users."	☐	☐	☐	☐	☐
19. "I derived little satisfaction from the virtual movie experience with other users."	☐	☐	☐	☐	☐
20. "The virtual movie experience with other users felt superficial."	☐	☐	☐	☐	☐
21. "I really enjoyed the time spent with other users in the virtual movie."	☐	☐	☐	☐	☐

See graphs below to indicate the emotional closeness	1	2	3	4	5
22. How emotionally close to other users do you feel now?	☐	☐	☐	☐	☐



Strongly disagree 1 2 3 4 5 Strongly agree	1	2	3	4	5
23. "In the virtual world I had a sense of 'being there'."	☐	☐	☐	☐	☐
24. "Somehow I felt that the virtual world was surrounding me and other users."	☐	☐	☐	☐	☐
25. "I had a sense of acting in the virtual space, rather than operating something from outside."	☐	☐	☐	☐	☐
26. "My virtual experience seemed consistent with my real world experience."	☐	☐	☐	☐	☐
27. "I did not notice what was happening around me in the real world."	☐	☐	☐	☐	☐
28. "I felt detached from the outside world while experiencing the virtual movie."	☐	☐	☐	☐	☐
29. "I focused on solving the virtual crime with other users."	☐	☐	☐	☐	☐
30. "Everyday thoughts and concerns were still very much on my mind."	☐	☐	☐	☐	☐
31. "It felt like the virtual movie experience took shorter time than it really was."	☐	☐	☐	☐	☐
32. "When experiencing the virtual movie with other users, time appeared to go by very slowly."	☐	☐	☐	☐	☐

Extra questions	1	2	3	4	5
33. "I liked the virtual movie."	☐	☐	☐	☐	☐
34. "The virtual movie is realistic (i.e. resemble a real scenario)."	☐	☐	☐	☐	☐
35. "The spatiality in the virtual movie (i.e., perceived distances and sizes of elements) is consistent with a real-life scenario."	☐	☐	☐	☐	☐

Thank you for your feedback!

Social VR Questionnaire (2 users)

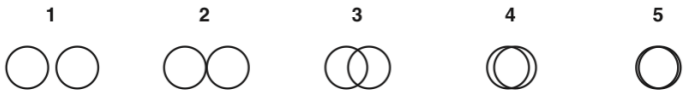
Please answer the questions, according to **your virtual movie experience in the virtual bedroom/kitchen.**

The scale of the following questions are from 1 to 5, representing the following meanings:
1 Strongly disagree 2 Disagree 3 Neutral 4 Agree 5 Strongly agree

Strongly disagree 1 2 3 4 5 Strongly agree	1	2	3	4	5
1. "I was able to feel the other user's emotion while in the virtual movie."	☐	☐	☐	☐	☐
2. "I was sure that the other user often felt my emotion."	☐	☐	☐	☐	☐
3. "The experience of solving a crime in the virtual movie with the other user seemed natural."	☐	☐	☐	☐	☐
4. "The actions used to interact with the other user were similar to the ones in the real world."	☐	☐	☐	☐	☐
5. "It was easy for me to contribute to the conversation with the other user."	☐	☐	☐	☐	☐
6. "The conversation with the other user seemed highly interactive."	☐	☐	☐	☐	☐
7. "I could readily tell when the other user were listening to me."	☐	☐	☐	☐	☐
8. "I found it difficult to keep track of the conversation."	☐	☐	☐	☐	☐
9. "I felt completely absorbed in the conversation."	☐	☐	☐	☐	☐
10. "I could fully understand what the other user was talking about."	☐	☐	☐	☐	☐
11. "I was very sure that the other user understood what I was talking about."	☐	☐	☐	☐	☐

Strongly disagree 1 2 3 4 5 Strongly agree	1	2	3	4	5
12. "I often felt as if I was all alone while in the virtual movie."	☐	☐	☐	☐	☐
13. "I think the other user often felt alone while watching the content."	☐	☐	☐	☐	☐
14. "I often felt that the other user and I were staying together in the same space."	☐	☐	☐	☐	☐
15. "I paid close attention to the other user."	☐	☐	☐	☐	☐
16. "The other user was easily distracted when other things were going on in the real world."	☐	☐	☐	☐	☐
17. "I felt that solving the virtual crime together enhanced the closeness with the other user."	☐	☐	☐	☐	☐
18. "Solving the virtual crime together created a good shared memory between me and the other user."	☐	☐	☐	☐	☐
19. "I derived little satisfaction from the virtual movie experience with the other user."	☐	☐	☐	☐	☐
20. "The virtual movie experience with the other user felt superficial."	☐	☐	☐	☐	☐
21. "I really enjoyed the time spent with the other user in the virtual movie."	☐	☐	☐	☐	☐

See graphs below to indicate the emotional closeness	1	2	3	4	5
22. How emotionally close to the other user do you feel now?	☐	☐	☐	☐	☐



Strongly disagree 1 2 3 4 5 Strongly agree	1	2	3	4	5
23. "In the virtual world I had a sense of 'being there'."	☐	☐	☐	☐	☐
24. "Somehow I felt that the virtual world was surrounding me and the other user."	☐	☐	☐	☐	☐
25. "I had a sense of acting in the virtual space, rather than operating something from outside."	☐	☐	☐	☐	☐
26. "My virtual experience seemed consistent with my real world experience."	☐	☐	☐	☐	☐
27. "I did not notice what was happening around me in the real world."	☐	☐	☐	☐	☐
28. "I felt detached from the outside world while experiencing the virtual movie."	☐	☐	☐	☐	☐
29. "I focused on solving the virtual crime with the other user."	☐	☐	☐	☐	☐
30. "Everyday thoughts and concerns were still very much on my mind."	☐	☐	☐	☐	☐
31. "It felt like the virtual movie experience took shorter time than it really was."	☐	☐	☐	☐	☐
32. "When experiencing the virtual movie with the other user, time appeared to go by very slowly."	☐	☐	☐	☐	☐

Extra questions	1	2	3	4	5
33. "I liked the virtual movie."	☐	☐	☐	☐	☐
34. "The virtual movie is realistic (i.e. resemble a real scenario)."	☐	☐	☐	☐	☐
35. "The spatiality in the virtual movie (i.e., perceived distances and sizes of elements) is consistent with a real-life scenario."	☐	☐	☐	☐	☐

Thank you for your feedback!

The **Presence Questionnaire** was used to measure the participants' sense of being in the virtual environment. We removed the Item 23 and 24 from the questionnaire, since Pilot 3 didn't include any haptics experiences. The construct of the presence questionnaire includes (Witmer et al., 2005): Realism (Items 3 + 4 + 5 + 6 + 7 + 10 + 13), Possibility to act (Items 1 + 2 + 8 + 9), Quality of interface (Items (all reversed) 14 + 17 + 18), Possibility to examine (Items 11 + 12 + 19), Self-evaluation of performance (Items 15 + 16), Sounds (Items 20 + 21 + 22).

PRESENCE QUESTIONNAIRE

(Witmer & Singer, Vs. 3.0, Nov. 1994)*
Revised by the UQO Cyberpsychology Lab (2004)

Characterize your experience in the environment, by marking an "X" in the appropriate box of the 7-point scale, in accordance with the question content and descriptive labels. Please consider the entire scale when making your responses, as the intermediate levels may apply. Answer the questions independently in the order that they appear. Do not skip questions or return to a previous question to change your answer.

WITH REGARD TO THE EXPERIENCED ENVIRONMENT

1. How much were you able to control events?

|_____|_____|_____|_____|_____|_____|_____|
NOT AT ALL SOMEWHAT COMPLETELY

2. How responsive was the environment to actions that you initiated (or performed)?

|_____|_____|_____|_____|_____|_____|_____|
NOT MODERATELY COMPLETELY
RESPONSIVE RESPONSIVE RESPONSIVE

3. How natural did your interactions with the environment seem?

|_____|_____|_____|_____|_____|_____|_____|
EXTREMELY BORDERLINE COMPLETELY
ARTIFICIAL NATURAL

4. How much did the visual aspects of the environment involve you?

|_____|_____|_____|_____|_____|_____|_____|
NOT AT ALL SOMEWHAT COMPLETELY

5. How natural was the mechanism which controlled movement through the environment?

|_____|_____|_____|_____|_____|_____|_____|
EXTREMELY BORDERLINE COMPLETELY
ARTIFICIAL NATURAL

6. How compelling was your sense of objects moving through space?

|_____|_____|_____|_____|_____|_____|_____|
NOT AT ALL MODERATELY COMPELLING VERY COMPELLING

7. How much did your experiences in the virtual environment seem consistent with your real world experiences?

|_____|_____|_____|_____|_____|_____|_____|
NOT CONSISTENT MODERATELY CONSISTENT VERY CONSISTENT

8. Were you able to anticipate what would happen next in response to the actions that you performed?

|_____|_____|_____|_____|_____|_____|_____|
NOT AT ALL SOMEWHAT COMPLETELY

9. How completely were you able to actively survey or search the environment using vision?

|_____|_____|_____|_____|_____|_____|_____|
NOT AT ALL SOMEWHAT COMPLETELY

10. How compelling was your sense of moving around inside the virtual environment?

|_____|_____|_____|_____|_____|_____|_____|
NOT COMPELLING MODERATELY COMPELLING VERY COMPELLING

11. How closely were you able to examine objects?

|_____|_____|_____|_____|_____|_____|_____|
NOT AT ALL PRETTY CLOSELY VERY CLOSELY

12. How well could you examine objects from multiple viewpoints?

|_____|_____|_____|_____|_____|_____|_____|
NOT AT ALL SOMEWHAT EXTENSIVELY

13. How involved were you in the virtual environment experience?

|_____|_____|_____|_____|_____|_____|_____|
NOT INVOLVED — MILDLY INVOLVED — COMPLETELY ENGROSSED

14. How much delay did you experience between your actions and expected outcomes?

|_____|_____|_____|_____|_____|_____|_____|
NO DELAYS — MODERATE DELAYS — LONG DELAYS

15. How quickly did you adjust to the virtual environment experience?

|_____|_____|_____|_____|_____|_____|_____|
NOT AT ALL — SLOWLY — LESS THAN ONE MINUTE

16. How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?

|_____|_____|_____|_____|_____|_____|_____|
NOT PROFICIENT — REASONABLY PROFICIENT — VERY PROFICIENT

17. How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?

|_____|_____|_____|_____|_____|_____|_____|
NOT AT ALL — INTERFERED SOMEWHAT — PREVENTED TASK PERFORMANCE

18. How much did the control devices interfere with the performance of assigned tasks or with other activities?

|_____|_____|_____|_____|_____|_____|_____|
NOT AT ALL — INTERFERED SOMEWHAT — INTERFERED GREATLY

19. How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?

|_____|_____|_____|_____|_____|_____|_____|
NOT AT ALL — SOMEWHAT — COMPLETELY

IF THE VIRTUAL ENVIRONMENT INCLUDED SOUNDS:

20. How much did the auditory aspects of the environment involve you?

|_____|_____|_____|_____|_____|_____|_____|
NOT AT ALL SOMEWHAT COMPLETELY

21. How well could you identify sounds?

|_____|_____|_____|_____|_____|_____|_____|
NOT AT ALL SOMEWHAT COMPLETELY

22. How well could you localize sounds?

|_____|_____|_____|_____|_____|_____|_____|
NOT AT ALL SOMEWHAT COMPLETELY

IF THE VIRTUAL ENVIRONMENT INCLUDED HAPTIC (SENSE OF TOUCH):

23. How well could you actively survey or search the virtual environment using touch?

|_____|_____|_____|_____|_____|_____|_____|
NOT AT ALL SOMEWHAT COMPLETELY

24. How well could you move or manipulate objects in the virtual environment?

|_____|_____|_____|_____|_____|_____|_____|
NOT AT ALL SOMEWHAT EXTENSIVELY

Last version : March 2013

*Original version : Witmer, B.G. & Singer. M.J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence : Teleoperators and Virtual Environments*, *7*(3), 225-240. Revised factor structure: Witmer, B.J., Jerome, C.J., & Singer, M.J. (2005). The factor structure of the Presence Questionnaire. *Presence, 14*(3) 298-312.

The **NASA Task Load Index** is used to assess the workload of the assigned tasks.

NASA TASK LOAD INDEX

Date: _____-_____-_____ Participant ID: _____

Condition: Desktop | HMD

Please rate your work load of the tasks and overall experiences in the virtual movie on a 7-point scale. Increments of high, medium and low estimates for each point result in 21 gradations on the scale.

1. Mental Demand: How mentally demanding was the experience/task in the virtual movie?

Very low Very high

2. Physical Demand: How physically demanding was the experience/task in the virtual movie?

Very low Very high

3. Temporal Demand: How hurried or rushed was the pace in the virtual movie?

Very low Very high

4. Performance: How successful were you in accomplishing what you were asked to do?

Very low Very high

5. Effort: How hard did you have to work to accomplish your level of performance?

Very low Very high

6. Frustration: How insecure, discouraged, irritated, stressed and annoyed were you?

Very low Very high

The aim of the **Visual Quality Questionnaire** was to measure participants' perception of the quality of the virtual characters appearance, as well as the quality of self-representation and the representation of other users. There are four versions. Depending on the users' positions in the virtual world, they use the corresponding version of the questionnaire (User 1, introduced in the scene at Position 1, will use the Version 1).

Visual Quality Questionnaire (Version 1)

Date: _____-_____-_____ Participant ID: _____

Condition: HMD User (Position 1, next to the light switcher)

Please rate the visual quality of your own representation, the representations of the other users, and the movie characters, on a 5-point scale.

1=bad, 2=poor, 3=fair, 4=good, 5=excellent

	1	2	3	4	5
1. Your own representation	☐	☐	☐	☐	☐
2. The representation of the user at Position 3	☐	☐	☐	☐	☐
3. The representation of the user next to the phone finder (Position 2)	☐	☐	☐	☐	☐
4. The representation of the user at Position 4	☐	☐	☐	☐	☐
5. The virtual characters	☐	☐	☐	☐	☐

Visual Quality Questionnaire (Version 2)

Date: _____-_____-_____ Participant ID: _____

Condition: HMD User (Position 2, next to the phone finder)

Please rate the visual quality of your own representation, the representations of the other users, and the movie characters, on a 5-point scale.

1=bad, 2=poor, 3=fair, 4=good, 5=excellent

	1	2	3	4	5
1. Your own representation	☐	☐	☐	☐	☐
2. The representation of the user next to the light switch (Position 1)	☐	☐	☐	☐	☐
3. The representation of the user at Position 3	☐	☐	☐	☐	☐
4. The representation of the user at Position 4	☐	☐	☐	☐	☐
5. The virtual characters	☐	☐	☐	☐	☐

Visual Quality Questionnaire (Version 3)

Date: _____-_____-_____ Participant ID: _____

Condition: Desktop User (Position 3)

Please rate the visual quality of your own representation, the representations of the other users, and the movie characters, on a 5-point scale.

1=bad, 2=poor, 3=fair, 4=good, 5=excellent

	1	2	3	4	5
1. Your own representation	☐	☐	☐	☐	☐
2. The representation of the user next to the light switch (Position 1)	☐	☐	☐	☐	☐
3. The representation of the user next to the phone finder (Position 2)	☐	☐	☐	☐	☐
4. The representation of the user at Position 4	☐	☐	☐	☐	☐
5. The virtual characters	☐	☐	☐	☐	☐

Visual Quality Questionnaire (Version 4)

Date: _____-_____-_____ Participant ID: _____

Condition: Desktop User (Position 4)

Please rate the visual quality of your own representation, the representations of the other users, and the movie characters, on a 5-point scale.

1=bad, 2=poor, 3=fair, 4=good, 5=excellent

	1	2	3	4	5
1. Your own representation	☐	☐	☐	☐	☐
2. The representation of the user next to the light switch (Position 1)	☐	☐	☐	☐	☐
3. The representation of the user next to the phone finder (Position 2)	☐	☐	☐	☐	☐
4. The representation of the user at Position 3	☐	☐	☐	☐	☐
5. The virtual characters	☐	☐	☐	☐	☐

2.2 Objective Evaluation

The project has defined as well a number of objective metrics for the evaluation of the experience (user behaviour) and the quality of the volumetric video (perceptual quality metrics for point clouds).

2.2.1 User behaviour

The user navigation data can be collected, recorded by tracking each user's camera position and orientation at 30Hz for the entire duration of the experience. The position is logged using X,Y and Z coordinates in the global coordinate system while the orientation is logged using three Euler angles about the principal axes (Rx,Ry,Rz). Additionally, for HMD users, we also log the position and orientation of each of the Oculus hand controllers.

During the experience all the events (of the game in Pilot 3) are logged separately for each user along with the associated timestamp. This includes different sections of the story, users moving to different rooms in the apartment, users being prompted to interact as well as user interactions with objects and virtual characters in the scene.

In addition, we also recorded general performance statistics for components in the pipeline for each user. This includes the framerate of each component, the average number of points in the captured self-view point cloud, the average latency of point clouds received from other users and the average points per decoded cloud for the other users in the session.

A summary of the metrics collected in the project can be found in Table 1

Objective	Metrics
Viewport (30Hz)	X, Y, Z position coordinates, Rx,Ry,Rz orientation in Euler angles about the principal axes
Controllers (30Hz)	Position and orientation of Oculus hand controllers for HMD users
Captured self-view point clouds (1Hz)	Average points per cloud and framerate
Remote user point clouds (1Hz)	Average points per cloud, latency, decode latency and framerate
Game events	Timestamps for session start, interaction prompts, user interaction triggers with objects and virtual characters, movement to different rooms and session end

Table 1 Objective metrics (user behavior) captured in social VR experiences.

2.2.2 Objective Perceptual Quality of the Signal: Volumetric Video

In order to predict how the signal will be perceived by the end users, we designed and tested two objective quality models for point cloud contents. The first belongs to the category of full reference metrics, that is, metrics that perform a direct algorithmic comparison between a distorted content and its pristine counterpart, whereas the second is a reduced reference metric, which signifies that only a small subset of features is needed to compute the prediction. The first is more suitable to drive compression engines, whereas the other can be used to optimize at the receiver's side.

2.2.2.1 Full reference: A color-based objective quality metric for point cloud contents

In order to measure distortions introduced by compression on the colour domain, 2D image-based metrics, such as PSNR, have been applied on the colour information of couples of points taken from the reference and distorted point cloud. The metrics thus capture variations in colour on a point-to-point basis. However, these metrics fail to capture the broad variations in texture, often resulting in poor prediction capabilities. The intuition behind this work resides in trying to capture the difference between coloured point clouds from a global perspective. We use colour statistics, such as colour histogram, to analyse how the general distribution of the colours changes when artefacts are introduced. In image processing, colour histograms represent the probability distribution of pixel values for the entire image. Distortions applied on the colour channel, such as compression artefacts, tend to modify the statistical distribution of the colours, for example by applying quantization. Thus, computing the distance between the colour histogram of a distorted model with respect to the one obtained from its unaltered reference, will give us an idea of the level of distortion. We propose the use of histogram distance as a measure of distortion of a test point cloud with respect to a reference. We compute the colour histogram on the luminance channel, which has been shown to better correlate with human perception of colour. Moreover, we apply a weighted average between the distances associated with the YCbCr channels, in order to have a distance metric that takes chrominance into account:

$$dist_{YCbCr} = \frac{6\, dist_Y + dist_{Cb} + dist_{Cr}}{8}$$

Several distances can be defined between two histograms. In our work, we select the L2 (Euclidean) distance, as it showed better performance.

To test the validity of our metric, we used a publicly available dataset, which consists of subjective and objective quality scores given to 8 point cloud contents, of which 4 depicting human bodies, under compression distortions, for a total of 232 stimuli. We then computed several performance indexes on the data, to assess the prediction power of our metrics with respect to the subjective ground truth, expressed as Mean Opinion Score (MOS). In particular, Pearson Linear Correlation Coefficient (PLCC), Spearman Rank Correlation Coefficient (SRCC), Root Mean Square Error (RMSE) and Outlier Ratio (OR) were chosen to account for linearity, monotonicity, accuracy and consistency, respectively, following ITU-T Recommendations P.1401. Before computing the indexes, linear, cubic and logistic fitting was applied on the results of the objective metrics.

Table 2 depicts the results of the performance indexes for the histogram-based objective quality metrics. Results are shown separately for the luminance channel and for the YCbCr weighted average. Results obtained by performing a weighted average on the YCbCr color space show that aberrations in the chrominance space do not correlate with human perception of distortions. In fact, incorporating chrominance distance information leads to generally poorer results in terms of performance indexes.

		Linear			Cubic			Logistic		
	SRCC	PLCC	RMSE	OR	PLCC	RMSE	OR	PLCC	RMSE	OR
Y	0.8841	0.5318	1.1521	0.8707	0.8195	0.7796	0.8362	0.8532	0.7096	0.7802
YCbCr	0.7020	0.5161	1.1651	0.8836	0.6698	1.0101	0.8491	0.6763	1.0021	0.8319

Table 2 Performance indexes for histogram-based objective quality metrics, separately for luminance channel and weighted luminance and chroma channels.

Point-based metrics are traditionally used to assess the quality of geometry-only or colour-only distortions. The histogram-based metric, as it is, can reflect geometry distortions, if they result in a

loss of points. In that case, the statistics of the colour distribution would likely change between the reference and distorted contents, even if the colour information itself is not altered. On the other hand, geometry-only metrics are not able to assess the perceived quality of point cloud contents, if the distortions are introduced only on the colour values without altering the topology.

To overcome the limitations of both approaches, we propose to combine geometry- and colour-based metrics in an individual metric which can capture both deviations. In particular, we define a new metric d_{gc} as a linear combination of the geometry-only metric d_g and the colour-only metric d_c:

$$d_{gc} = \alpha \cdot d_g + (1 - \alpha) \cdot d_c$$

in which α is a real number between 0 and 1. In our case, the p2plane (point to plane) metric with MSE error is selected as d_g whereas the colour histogram metric on the Y channel is selected as d_c

We performed a grid search on values of α, to understand whether an improvement in performance could be found by linearly combining the two metrics. Figure 1 displays the performances indexes PLCC and SRCC obtained for different values of α. Logistic fitting was applied on the metric d_{gc} before computing the indexes. As can be seen, combining the two metrics always leads to an increase in performance. In particular, the lowest indexes are obtained for $\alpha = 0$, which corresponds to using only the histogram-based colour metric, and α = 1 for which only the geometry metric is used. The highest value of linear correlation is reached for $\alpha = 0.6597$, for which PLCC=0.9037. In terms of SRCC, the highest value of 0.9205 is reached for α = 0.6364. Figure 2 shows the scatterplot of our proposed metric against the MOS, for $\alpha = 0.6597$. For the best-performing α the null hypothesis of equivalence between the PLCC associated with d_{gc} and the PLCC obtained with d_g and d_c is rejected.

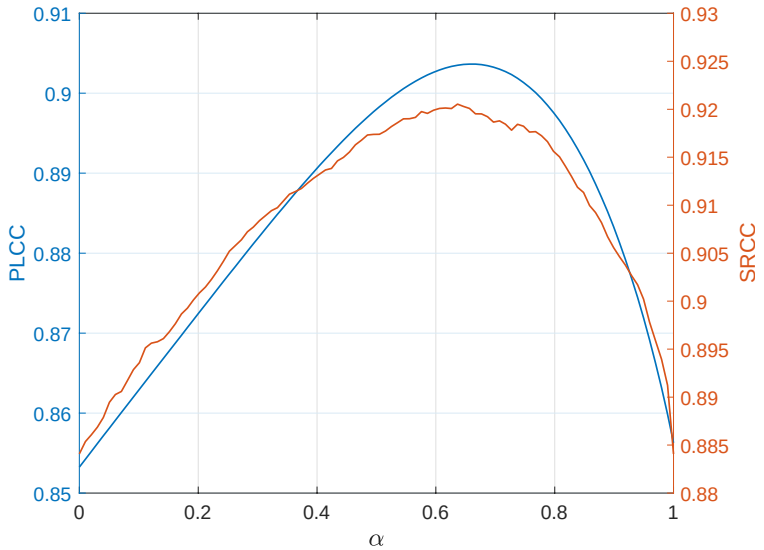


Figure 1 Performance indexes PLCC and SRCC for values of alpha. For alpha = 0, only the color-based metric is used; for alpha = 1, only the geometry metric is selected.

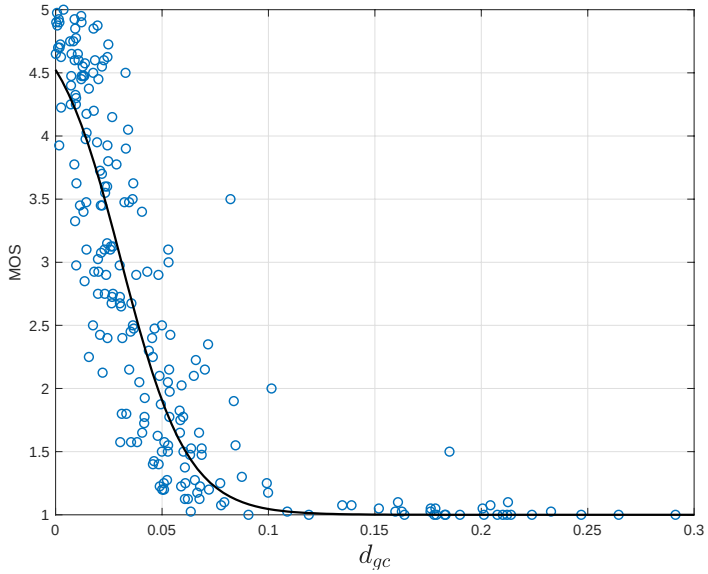


Figure 2 Proposed metric against MOS values. To improve readability, results are shown for lower levels only.

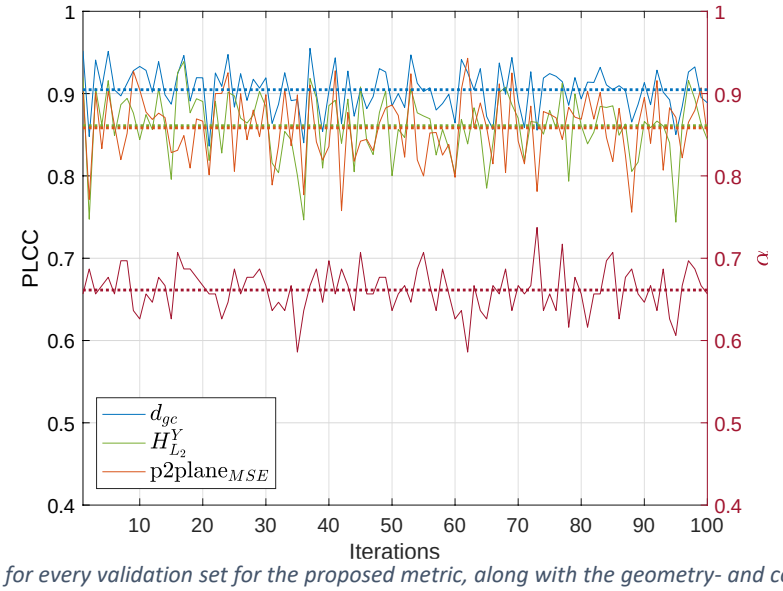


Figure 3 PLCC index for every validation set for the proposed metric, along with the geometry- and colour-based metrics, and chosen alpha.

In order to validate that the improvement in performance was not caused by overfitting, we performed 100 random splits on the data (80% training, 20% test set), and we computed the optimal level of α in terms of PLCC. We then used the α value we defined on the training set to compute d_{gc} on the test data. We also computed the PLCC obtained on the test set for the geometry- and colour-based metrics alone ($d_g$and d_c, respectively), to assess whether an increase in performance could be observed. Results of the iterations are shown in Figure 3. The mean value over the iterations is shown with a dotted line. Small variations can be observed in the performance index, depending on the composition of the test set. Nonetheless, the combined metric d_{gc}achieves better results in term of PLCC with respect to the colour-only metric $d_c$and geometry-only metric $d_g$on 95 of the 100 iterations under consideration.

2.2.2.2 Reduced reference: A reduced reference metric for visual quality evaluation of point cloud contents

Reduced reference metrics need to rely on extracting a set of features from a reference content in order to predict the level of distortion in the content under assessment. As the set of features needs

to be transmitted alongside the content, it needs to be as informative as possible while maintaining a low cardinality. They have been adopted in the image and video community in order to produce a real-time estimation of visual quality at the receiving side of the transmission. However, adapting a reduced reference framework to point cloud contents requires rethinking in terms of what dimensions will be affected by compression and transmission distortions. Traditional static 2D contents lie on a regular grid, which is unlikely to be tampered with. Hence, distortions will likely be present in the luminance or colour domain. On the other hand, static point cloud contents can be distorted in the geometrical domain, along with the point attributes domain.

In our work, we propose to use statistical features computed on the geometry information, luminance channel, and normal vectors, in order to measure the level of distortion of a degraded point cloud content. In particular, we extract three feature sets Φ^G, Φ^Y, Φ^N, comprised of 21 features extracted from a given distorted point cloud content. We then compare them to the features extracted from the corresponding reference content, by computing the absolute difference of every single feature. We then run a linear optimization algorithm to find the best weight to give to each feature.

To train and evaluate our metric, we use the same dataset as before, consisting of subjective and objective quality scores assigned to 8 point cloud contents (4 human bodies, 4 inanimate objects) under compression distortions, resulting in 232 stimuli. We extract the features from the reference and distorted point clouds. Features are computed and stored in single float precision, requiring 84 bytes to be transmitted.

To obtain the weights, we run a linear optimization algorithm, which aims at maximizing the Pearson Linear Correlation Coefficient (PLCC) between our metric and the corresponding subjective scores, after logistic fitting. To see how the metric generalizes to previously unseen contents, we perform Leave p Out cross-validation (LpOCV) by selecting 4 contents out of the 8 provided to be used for testing, and training on the remaining 4. We repeat the procedure for all 70 pairs, and we report the average performance. Additionally, we perform Monte Carlo cross-validation (MCCV) with 100 random splits on our dataset (80% training, 20% test). Following ITU-T Recommendations P.1401, the performance of our metric is assessed using the Spearman Rank Correlation Coefficient (SRCC), along with the aforementioned PLCC, to account for monotonicity and linearity, respectively, after logistic fitting.

Table 3 reports the mean correlation coefficient, along with the corresponding standard deviation, obtained through cross-validation. To offer a comparison with widely-used metrics in the state of the art, we also report the results of metrics D1 (point-to-point) and D2 (point to plane), as defined and employed in the MPEG standardization effort of the point cloud ad-hoc group. As no training of parameters is involved, correlation results are reported for the entire dataset. However, it should be noted that those metrics are full reference, thus including information from the full point cloud content, and only assess distortion in the geometrical domain. It can be observed that our metric is outperforming the aforementioned FR solutions both in terms of PLCC and SRCC, for both cross-validation methods.

	SRCC	PLCC
PCMRR (LpOCV)	0.826 ± 0.102	0.798 ± 0.111
PCMRR (MCCV)	0.907 ± 0.028	0.901 ± 0.029
D1	0.759	0.720
D2	0.807	0.756

Table 3 Performance results of the proposed metric in the cross-validation.

Figure 4 depicts the optimal weight for each feature, averaged across the 70 pairs in the LpOCV, with relative confidence intervals. Features are grouped per feature set to facilitate comprehension. It can be observed that the two largest weights (0.285 and 0.141) are assigned to features in set $\Phi^Y$and Φ^G, which corresponds to the energy of the luminance histogram, and the mean in the geometry domain, respectively. Generally, the weights appear to be balanced between structure and colour information, although less weight is given to normal vector features: set Φ^Y accounts for 48.47% of the total weights, whereas sets Φ^G and Φ^Ncomprise 51.53% (41.54% and 9.99%, respectively).

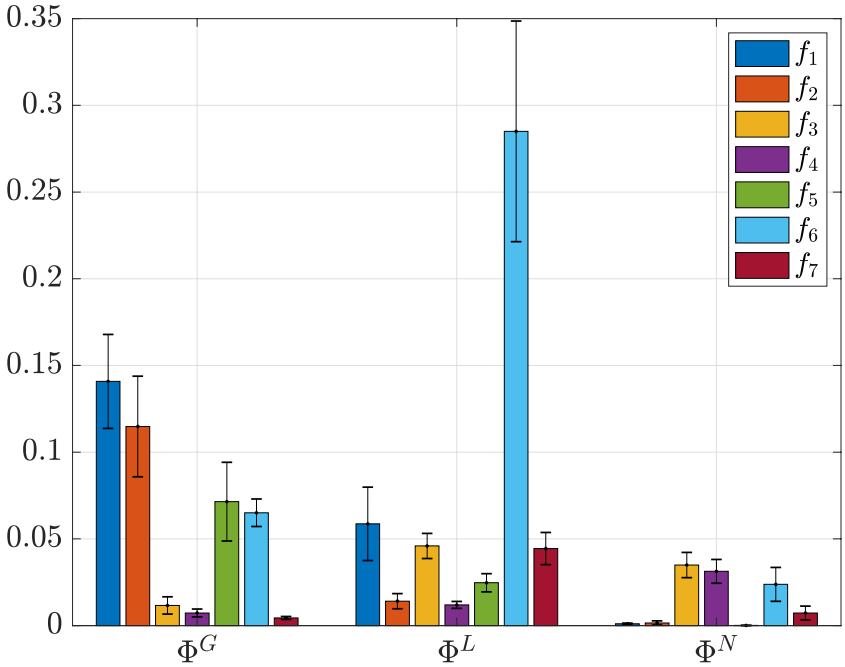


Figure 4 Optimal weights for each feature, averaged across the LpOCV splits, with relative 95% confidence intervals.

2.3 Objective Performance

Objective performance metrics enable to assess and profile the performance of the VRTogether system and of its components, when using different variants, in different scenarios and/or under different situations. They are also essential to determine limits (e.g., in terms of scalability) and obtain thresholds / recommendations for specific requirements (e.g. frame rates, delays, support of number of users given specific network and processing capabilities, etc.).

The objective metrics can be processed locally, by running specific modules within the players, and/or by using external tools. They can also be processed in the servers (e.g. MCUs), taking advantage of the higher processing power of such servers and allowing a centralized system. All captured data are processed and stored locally, for efficiency and to ensure privacy.

In D4.2 (year 1 of the project), an initial analysis was made to provide a list of objective performance metrics to be considered in the project (*what*), together with the reason behind measuring them (*why*), the system components in which the metrics need to be measured (*where*) and the measurement method / tool (*how*). A subset of these metrics was already measured in some year 1 experiments reported in D4.2: CWI-1, CERTH-1, CERTH-2, CERTH-3, CERTH-4 and Pilot 1.

In D4.4 (year 2 of the project), some progress was made with regard to the implementation and measurement of metrics for each of the pipelines considered in the project: web platform, and native platform, including both the TVM and Point Clouds pipelines and the player. These metrics

were measured in many of the year 2 experiments reported in D4.4: connected labs for the web and native platforms, CERTH 2.1, CERTH-2.3, CERTH-2.6 and TNO-2.3.

During the third year, the implementation and measurement of objective performance metrics has been consolidated, for each of the platforms and pipelines considered in the project. Next, the available metrics are listed.

Native Platform: TVM pipeline

With regard to the TVM pipeline, the next metrics (average and standard deviation values are also computed) have been implemented and integrated as part of the logging system of the VRTogether platform:

- Frames per second
- Missed / Skipped frames
- Frametime per TVM (ms)
- End-to-end delay (ms)
- Number of vertices per TVM and standard deviation
- Size of received compressed TVM (MB)
- Size of decompressed TVM (MB)
- Total deserialization-decompression time per TVM (ms)
- Deserialization-decompression function execution time per TVM (ms)
- Marshalling time for the texture data per TVM (ms)
- Marshalling time for the geometry data per TVM (ms)
- Marshalling time for the extra parameters per TVM (ms)
- Rendering time per TVM (ms)

In addition, an external tool has been integrated to be able to collect extra RabbitMQ (RMQ) related metrics:

- Message rate in per second of RMQ exchange for the TVM data
- Message rate out per second of RMQ exchange for the TVM data
- Receiving rate (MB/s) of connections to RMQ server to the respective IPs
- Sending rate (MB/s) of connections to RMQ server to the respective IPs
- Many metrics related to the use of memory (MB) of RMQ server's nodes

Native Platform: Point Cloud pipeline

With regard to the Point Cloud pipeline, the next metrics have been implemented and integrated as part of the logging system of the VRTogether platform:

- Decoded frames per second (fps).
- Number of Decoded points per Point Cloud frame.
- Latency (ms): latency since capturing to rendering, assuming clock synchronization between the involved devices, or by relying on a mapping between local and absolute timestamps provided by the Orchestrator at the beginning of the session (assuming negligible Round Trip Times).

Web Platform:

D4.4 reported on the ability to gather the metrics provided by the WebRTC monitoring APIs[1], when using the web platform. These metrics include:

- Video upload delay
- Video upload jitter
- Audio upload delay
- Audio upload jitter

As reported in Section 4.2, these metrics have been extended by incorporating relevant WebGL/Aframe Rendering stats[2]:

- fps: frames per second.
- requestAnimationFrame Latency
- Textures: number of three.js textures in the scene. A lower count means the scene is using less memory and sending less data to the GPU
- Programs: number of OpenGL Shading Language (GLSL) in the scene
- Geometries: number of three.js geometries in the scene. A lower count means the scene is using less memory
- Vertices: number of vertices in the scene
- Faces: number of faces in the scene
- Calls: number of draw calls on each frame
- Load Time (ms): how long it took for the scene to start rendering
- Entities: number of A-Frame entities

Native and Web Platforms: Resources Consumption Metrics Measurement

The project has contributed with the development of a tool for the measurement of resources consumption metrics for any either single process or multi-process Windows program, leveraging the PowerShell features (https://docs.microsoft.com/en-us/powershell/). Given the target Windows process name, measurement period, and target destination for logs (input arguments), the tools is able to retrieve the following metrics:

- CPU usage: The measurement of the CPU usage when performing data-intensive tasks is important to know about the requirements of specific components and the resources that are needed to meet them. The CPU usage is typically measured in percentage. However, the percentage usage (i.e., CPU load in %) is highly influenced by the computational resources of the involved processors/machines. Therefore, the CPU usage in terms of amount of memory (bytes) can be also measured.
- RAM usage: The same rationale for measuring this metric than for the CPU usage applies. It is measured in amount of memory (bytes).
- GPU usage: The same rationale for measuring this metric than for the CPU usage applies. It is especially relevant when performing video- or graphic-related processing tasks. It can be measured in terms of percentage and bytes, and gives information about the number of engines used.

[1] https://w3c.github.io/webrtc-stats/

[2] https://github.com/aframevr/aframe/blob/master/docs/components/stats.md

That tool has been published at ACM MMSYS 2020 conference and obtained the ACM Reproducibility badge (*Functional Artefact*) from ACM. It is publicly available on Zenodo, as open-source, and in a Github repo (https://github.com/ETSE-UV/RCM-UV), together with the instructions to appropriately run and customize it.

[Mon20a] M. Montagud, J. Antonio De Rus, R. Fayos-Jordán, M. Garcia-Pineda, and J. Segura-Garcia, "Open-Source Software Tools for Measuring Resources Consumption and DASH Metrics", ACM MMSYS 2020, Istanbul (Turkey), June 2020.

That tool has been used for the evaluation of many software components of the project, like the web and native player, the depth sensor and capture process modules, and the native MCU, both in pre-validation steps and in the execution of experiments (e.g. CERTH-3.1, VO-3.1, i2CAT-3.3).

2.4 Added-Value

The goal of the professional evaluation is to gain insights from the industry about the potential of the VRTogether social VR platform and challenges that the project is facing. We have strictly followed the Covid-19 prevention protocol at CWI. The elevation protocol is presented as follows:

Step 1 [15 mins]. Welcome and Introduction of the VRTogether Project
Pablo Cesar gave a presentation about the VRTogether project.

Key messages of the presentation:

- What is SocialVR? Why is it important?
- Enable social interaction (and even collaboration) between remote users
- Photo-Realistic Representation of Users, using state-of-the-art technologies and off-the-shelf components and low-cost equipment
- Introduce the three pilots and achievements of the project. Reflect on the challenges and goals of each pilot
- Describe the goal and procedure of the professional evaluation.

Step 2 [30 mins]. Showcase Pilot 3 and the virtual meeting room
Bring the professionals to the labs, make sure they have experienced both Pilot 3 and the virtual meeting room and tried both the HMD and the Desktop versions.

Step 3 [30 mins]. Group Discussion
We prepared a list of questions for the group discussion. The questions cover seven aspects, namely potential in the market, social VR experiences, technological components, quality of interaction, immersion and togetherness, missing aspects, industry links.

2.5 Use of Informed Consent Forms & Ethics Aspects

In all pilot actions in which data from the users are collected, Informed Consent Forms are provided to them before starting the experiment, explaining the data to be collected and the purposes. Data are only collected if the users give their consent. A GDPR compliant Consent Form template has been produced to ease this task to the partners conducting user experiments.

3 FOCUS GROUPS WITH PROFESSIONALS

CWI invited, on December 8th 2020, the local advisory board to experience the VRTogether platform and to play with Pilot 3 (see Figure 5). In addition to the valuable insights about the potential of the technology, this enabled the creation in the Netherlands of a network of interested parties on social VR. The local advisory board was formed by 14 professionals, representing nine companies/institutes, including Sound and Vision Institute, Medical VR, The Virtual Dutch Men, NEMO Science Museum, Erasmus University Medical Center, Sensiks, PostNL, BuitenboordMotor, and Interface. The event strictly followed the Covid-prevention protocols of CWI.

Figure 5 Focus group with the local advisory board in Amsterdam at CWI.

3.1 Methodology

We have set up three labs at CWI for the professional evaluations (see Figure 6). Two labs were equipped with an Oculus Rifts HMD and controllers, and one lab was equipped with a desktop computer and a game joystick. Each user was captured by three Kinect depth cameras. The 14 professionals came in two groups. The first group started at 13:00 and the other group started at 15:00. The first group includes professionals from the Sound and Vision Institute, Medical VR, The Virtual Dutch Men and NEMO Science Museum and the Erasmus University Medical Center. Professionals from the other companies came as the second group.

The professional evaluation consists of three parts. We started with a presentation introducing the goal, vision and accomplishment of the VRTogether project. Then, the professionals went to labs to experience Pilot 3 and the social VR meeting. All professionals have experienced both the HMD and the desktop version of the pilots. After the demos, the professionals were gathered in a large meeting room for a half-hour discussion about their overall social VR experiences of the pilots, the potentials and challenges of the VRTogether project.

Figure 6 The three connected rooms at CWI that showcased the VRTogether technology to the local advisory board.

3.2 Participants

The local advisory board in the Netherlands participated in the event. Table 4 summarizes the roles, the company/institute of the professionals who participated in the evaluation at CWI on December 8, 2020.

Name of the professional	Role of the professional	Name of the company/Institute
Chris Hordijk	Founder and CEO	Medical VR
Kristina Petrasova	Project lead digital heritage & public media	Sound and Vision Institute

Philo van Kemenade	User Interface Engineer	Sound and Vision Institute
Roelof Terpstra	Founder	The Virtual Dutch Men
Mart Vogel	Program maker	NEMO Science Museum
Eveline Corten	Principle investigator, Plastic surgeon for after-cancer body reconstruction	Erasmus University Medical Center
Gijs Hijmans	High-tech manager/physicist	Sensiks
Fred Galstaun	Founder & Director	Sensiks
Jurre van Ruth	Digital strategy consultant	PostNL
Niels van den Helder	Digital learning & Development specialist	PostNL
Madelon Barens	consultant	BuitenboordMotor
Sanne Akkersdijk	Project manager	BuitenboordMotor
Daan van der Woude	IT manager	Interface
Eline Oudenbroek	VP Operations	Interface

Table 4. Local Advisory Board in the Netherlands.

3.3 Results

All professionals were impressed by the simple setup of the "hologram" capturing system, and were excited to see each other in realistic representations. All of them had a strong presence experience, feeling that they were actually co-present in the same room. They were waving at each other, and talking about the texture of the clothes, the virtual environment and the possible scenarios for applying the VRTogether platform. The content of the pilot was quite simple, but still attracted all the professionals to spend a considerable amount of time in the virtual environment. They see the full potential of the VRTogether social VR platform in all market sectors, including medical care, education, immersive meetings, family reunion, virtual dating.

Fred from Sensiks:
"It's definitely, I mean, it doesn't feel like an avatar. It feels like the real person. That's the key thing. The fact is, it's so real, despite the quality, obviously, still needs a lot of improvement, but you go beyond that, that uncanny valley thing, and it's really the person there. That's amazing. That's more than I expected. I expected it to be a nice virtual environment with a Skype-like call, but you go way beyond that."

The professionals also gave constructive recommendations on how to further improve the experiences.

Philo from Sound and Vision Institute:
"Instead of fully replicating the real-world interactions, encoding the interactions in another way, like highlighting the representation of the speaker, visualizing the users' emotions on the 'hologram'."

"To increase the accessibility of social VR, to make it pervasive, we need to centralise it. We need to build on the fact that a medical centre, a festival and museum, these are centralised hubs where people come to use social VR. It is more efficient in a way to make the investment."

Eveline from Erasmus University Medical Center:
"The realism level of the virtual objects needs to be much higher in clinical context. Suppose we are going to reconstruct the breast of a cancer patient, not only the visual quality needs to be fully realistic, the haptic feelings of the 3D reconstruction also need to be realistic. In this way, patients can have a realistic expectation towards the surgery."

Roelof from The Virtual Dutch Men:
"When you want to move around the space, you have to teleport with a controller. It's easy if you know how to do it. But if you need your own hands to make some strange gestures to control the interface, it is really strange, and takes time to memorize the gestures. So, we have to think about this."

Madelon from BuitenboordMotor:
"The HMD was very immersive, but on the other hand, the screen with a game controller was more practical, perhaps also more addictive. I am curious to see the effects on 3D screens."

In summary, the professionals are very positive towards the social VR platform. They see its full potential not only in supporting remote experiences, but also augmenting the co-present experiences (hyper-realistic medical objects). The improvement suggestions mainly cover three aspects: (1) the visual quality of the volumetric representations (e.g., less noisy, higher resolution), (2) the possibility to interact (e.g., develop a new set of interactions/social cues dedicated to virtual world), and (3) the accessibility of the social VR platform (e.g., decentralized at home or centralized at public hubs).

4 USER LABS: CONNECTED NODES

This section reports the work on the connected user labs, detailing results from both of the pipelines: native (Spain, Greece, the Netherlands) and web (France and the Netherlands).

4.1 Native Pipeline: CERTH-3.1

4.1.1 Objective

This section reports on a series of experiments to assess the performance of the VRT Native pipeline across 3 different, remote, connected nodes: Thessaloniki (CERTH), Barcelona (i2CAT) and Amsterdam (CWI). Beyond demonstrating the connectivity and communication capabilities in a cross-country scenario, even when using/mixing different combination of media representation formats in specific test conditions, the objective also relies on reporting on "resource consumption", "network demands" and "capturing capabilities" metrics. With this assessment, we offer a benchmark of VRTogether based on the latest version of the VRTogether native platform and components.

4.1.1 Methodology

The following VRTogether nodes were used for conducting the experiment, each one using a different representation format / variant:

- **CERTH (Thessaloniki):**
  - 1 TVM node
- **CWI (Amsterdam):**
  - 1 Full-capture Point Cloud
  - 1 Single-Cam Point Cloud
- **i2CAT (Vilanova i la Geltrú, Barcelona):**
  - 1 Single-Cam Point Cloud

More details about the available lab's infrastructure and equipment are provided on the project website: https://vrtogether.eu/about-vr-together/user-lab/

Figure 7. Connected nodes location map

In each of the lab nodes, the client PCs used to run the sessions have the following characteristics:

- **CERTH (Thessaloniki):**
  - CPU: Intel(R) Core(TM) i7-8700K @ 3.70GHz
  - GPU: Nvidia GeForce GTX 1070
  - RAM: 32 GB
- **CWI (Amsterdam):**
  - 12-core i9, 2.9GHz
  - 128GB
  - 2x Nvidia GeForce GTX 1080 Ti
- **i2CAT (Vilanova i la Geltrú, Barcelona):**
  - CPU: Core i7 10750H @CPU 2.60GHz 2.59 Ghz (6 cores) 12 CPUs

It is worth noting that all tests were conducted with the simplest 3D scenario possible to add minimum interference to the connection, streaming and rendering performance. In all tests, real persons were captured while naturally talking to recreate real-life social VR sessions. The duration of each test was approximately 5min to also recreate real-life social VR scenarios.

4.1.2 Experiment sessions

We conducted 6 sessions in the following conditions:

Session	Participants	Description	Details
Test 1	CERTH, CWI	A TVM with RabbitMQ	- CERTH sends 1 TVM - CWI sends an avatar for minimum interference
Test 2	CWI, I2CAT	Single-Cam Point Cloud with socket.io	- CWI sends 1 Single-Cam Point Cloud - i2CAT sends 1 Single-Cam Point Cloud
Test 3	CWI, I2CAT	Full-Capture Point Cloud with socket.io	**Test 3A:** - CWI sends 1 Full-Capture Point Cloud - i2CAT sends an avatar for minimum interference
			Test 3B: - CWI sends 1 Full-Capture Point Cloud - i2CAT sends 1 Single-Cam Point Cloud
Test 4	CWI, I2CAT	Full-Capture Point Cloud with DASH	**Test 4A:** - CWI sends 1 Full-Capture Point Cloud - i2CAT sends an avatar for minimum interference
			Test 4B: - CWI sends 1 Full-Capture Point Cloud - i2CAT sends 1 Single-Cam Point Cloud
Test 5	CWI, I2CAT, CERTH	TVM and 2 Point Clouds with socket.io	- CERTH sends 1 TVM - CWI sends 1 Full-Capture Point Cloud - i2CAT sends 1 Single-Cam Point Cloud - Delivery Protocol: socket.io
Test 6	CWI, I2CAT, CERTH	TVM and 2 Point Clouds with DASH	- CERTH sends 1 TVM - CWI sends 1 Full-Capture Point Cloud - i2CAT sends 1 Single-Cam Point Cloud - Delivery Protocol: DASH

Table 5 Experiment sessions table

4.1.3 Metrics

- **Metrics from the TVM stream**
  - Frame rate
  - Average end to end delay (ms) and standard deviation

- Average size of received compressed TVM (MB) and standard deviation
- Average size of decompressed TVM (MB) and standard deviation
- Average total deserialization-decompression time per TVM (ms) and standard deviation
- Average rendering time per TVM (ms) and standard deviation
- Average of Message rate in per second of RMQ exchange for the TVM data and standard deviation
- Average of Message rate out per second of RMQ exchange for the TVM data and standard deviation
- Average of Receiving rate (MB / second) of connections to RMQ server (from sessions' users) to the respective IPs and standard deviation
- Average of Sending rate (MB / second) of connections to RMQ server (from sessions' users) to the respective IPs and standard deviation

- **Metrics from the Point Cloud stream**
  Per node:
  - Average decoded frames per second (FPS)
  - Average decoded #points/cloud
  - Average decoded latency (ms)

- **Resources Consumption Metrics**
  Per node:
  - RAM usage (MB)
  - CPU usage (%)
  - GPU usage (%)

- **Metrics extracted by analysing Wireshark captures (.pcap),**
  - Round Trip Times for TCP segments (ms)
  - TCP errors
  - TCP throughput per stream (bps)

4.1.4 Results

In the following, we present the main results of the experiments

Test 1

Resources Consumption Metrics (Averages):

Node / Metrics	Avg. CPU (%)	Avg. GPU (%)	Avg. RAM (MB)
CERTH	15.48	12.09	600.44
CWI	2.8	31.22	612.14

Table 6: Test 1 Resources consumption metrics

Wireshark Analysis of the incoming TVM stream (*Advanced Message Queueing Protocol (AMQP)* over TCP, segments size = 1514 bytes)

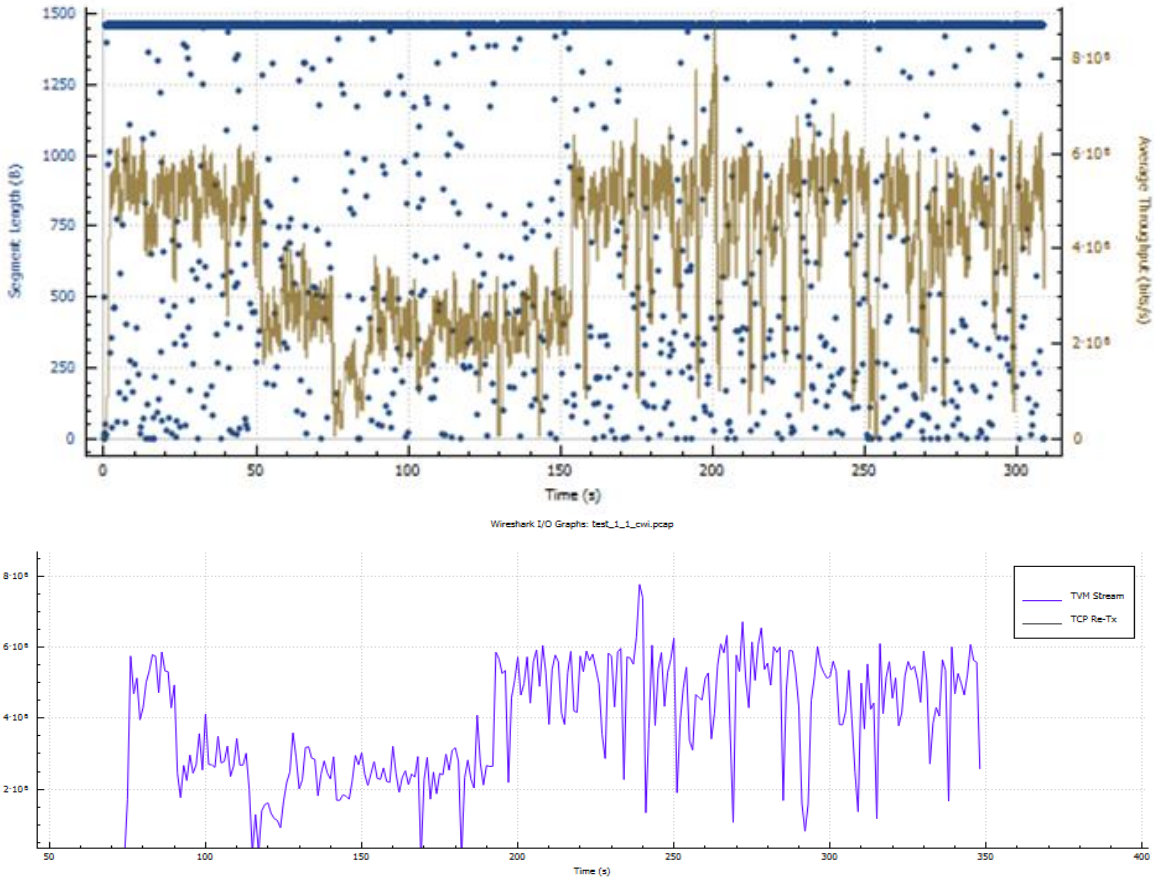


Figure 8 Test 1: Average Throughput (bps) and TCP segment length related to the TVM stream

As can be seen in the previous figure, the TCP throughput for the TVM stream suffered some fluctuations, but was in the order of 6Mbps. No TCP errors, re-transmissions and out of order TCP segments occurred during the session.

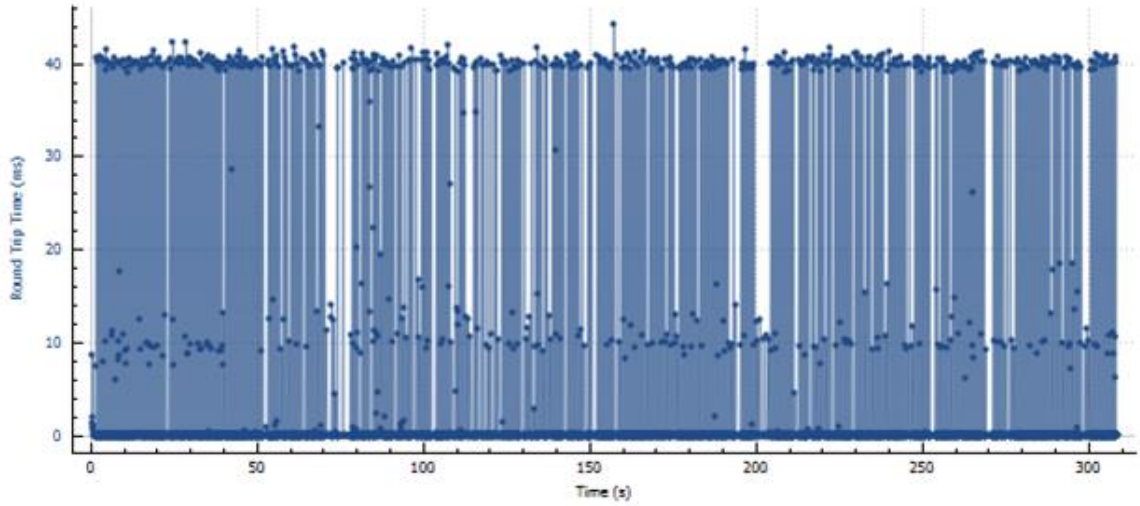


Figure 9 Test 1: Round Trip Times for TCP segments related to the TVM stream

Unity pipeline related metrics		Transmission related metrics	
Frames per second	4.97	**Average of message rate in / second of RMQ exchange for the TVM data**	14.64
Average missed/skipped frames	0	**StD of message rate in / second of RMQ exchange for the TVM data**	8.4
Average end-to-end delay (ms)	1361.64	**Average of message rate out / second of RMQ exchange for the TVM data**	28.2
StD of end-to-end delay (ms)	216.7	**StD of message rate out / second of RMQ exchange for the TVM data**	18.84
Average size of received compressed TVM (MB)	0.1	**Average of receiving rate (MB / second) of connection to distant RMQ server**	6.22
StD of size of received compressed TVM (MB)	0.011	**StD of receiving rate (MB / second) of connection to distant RMQ server**	5.3
Average size of decompressed TVM (MB)	3.173		
StD of size of decompressed TVM (MB)	0.148		
Average number of vertices per TVM	14089		
StD of number of vertices per TVM	3943.5		
Average total deserialization-decompression time per TVM (ms)	10.1		
StD of total deserialization-decompression time per TVM (ms)	1.52		
Average rendering time per TVM (ms)	5.01		
StD of rendering time per TVM (ms)	0.52		

Table 7: Test 1: Unity pipeline average metrics from and between VRT nodes.

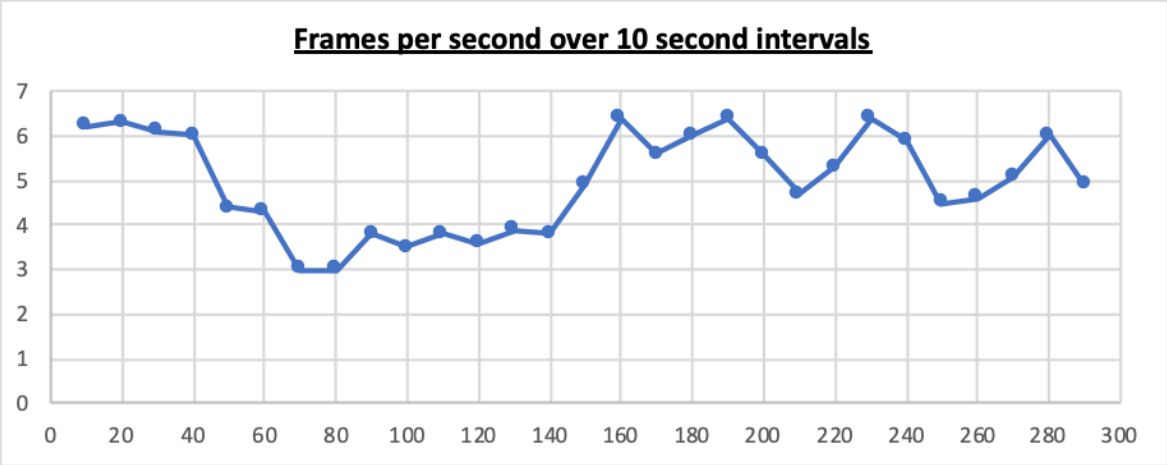


Figure 10 Test 1: FPS during experiment duration graph

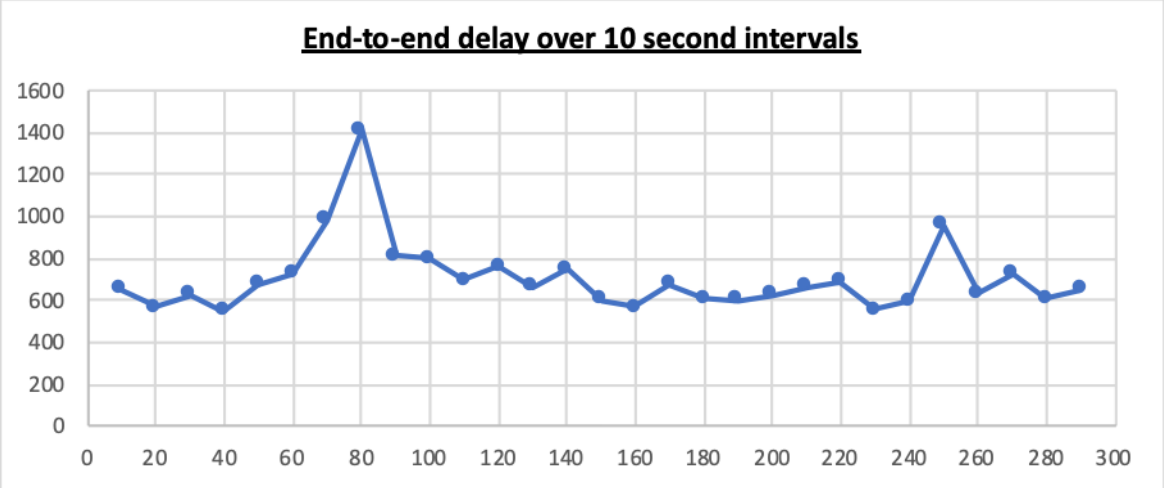


Figure 11 Test 1: End-to-end delay (in milliseconds) during experiment duration graph

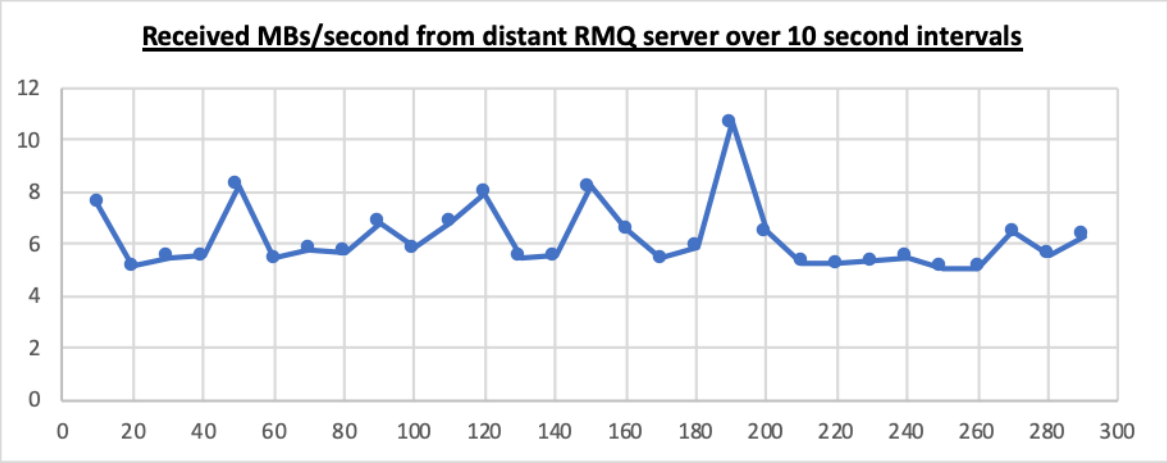


Figure 12 Test 1: Received MB/sec from RMQ server graph

Test 2

Resources Consumption Metrics:

Node / Metrics	Avg. CPU (%)	Avg. GPU (%)	Avg. RAM (MB)

i2CAT	18.97	20.51	1048.62
CWI	8.21	32.85	817.2

Table 8 Test 2: Resources consumption metrics

Point cloud metrics (averages)

Node/Metrics	**Avg. decoded fps**	**Avg. dec #points/cloud**	**Avg latency (ms)**
i2CAT	8.1	31393.55	Measurement error
CWI	14.99	36748.875	556.675

Table 9 Test 2: Point cloud metrics

Test 3A

Resources Consumption Metrics (Averages):

Node / Metrics	**Avg. CPU (%)**	**Avg. GPU (%)**	**Avg. RAM (MB)**
i2CAT	5.63	9.96	878.28
CWI	12.32	37.4	1076.33

Table 10 Test 3A: Resources consumption metrics

Point cloud metrics:

Node/Metrics	**Avg. decoded fps**	**Avg. dec #points/cloud**	**Avg latency (ms)**
i2CAT	5.28125	68941.15625	Measurement error
CWI	avatar		

Table 11 Test 3A: Point cloud metrics

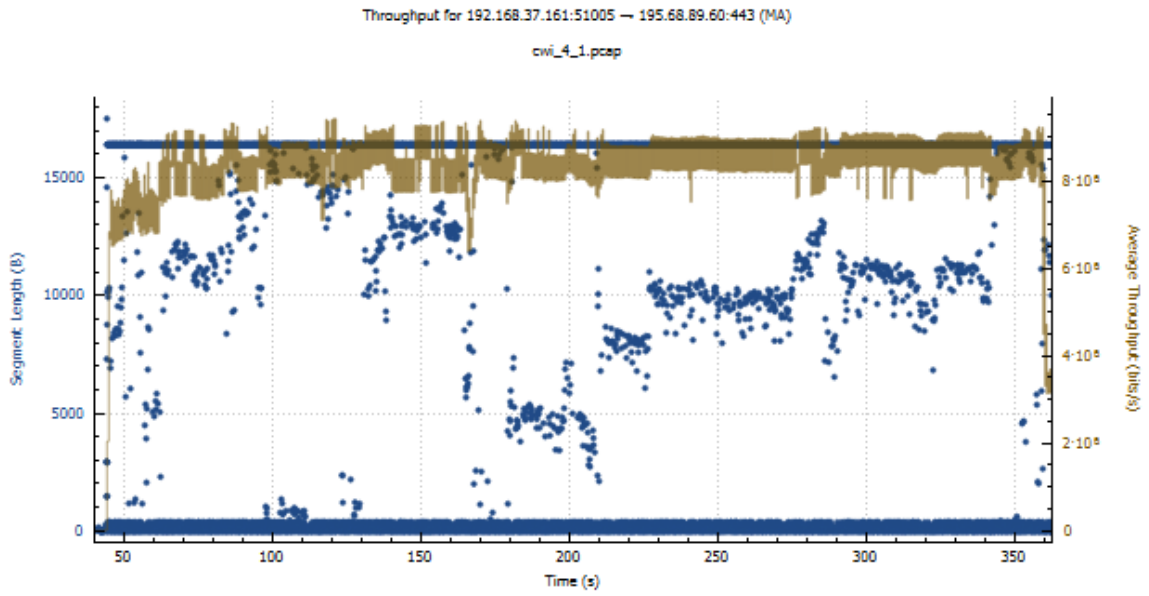


Figure 13 Test 3A: Average Throughput (bps) and TCP segment length related to the Full-capture Point Cloud stream

As can be seen in the previous figure, the TCP throughput for the Full-Capture Point Cloud stream was quite stable, in the order of 8-9Mbps. No TCP errors, re-transmissions and out of order TCP segments occurred during the session.

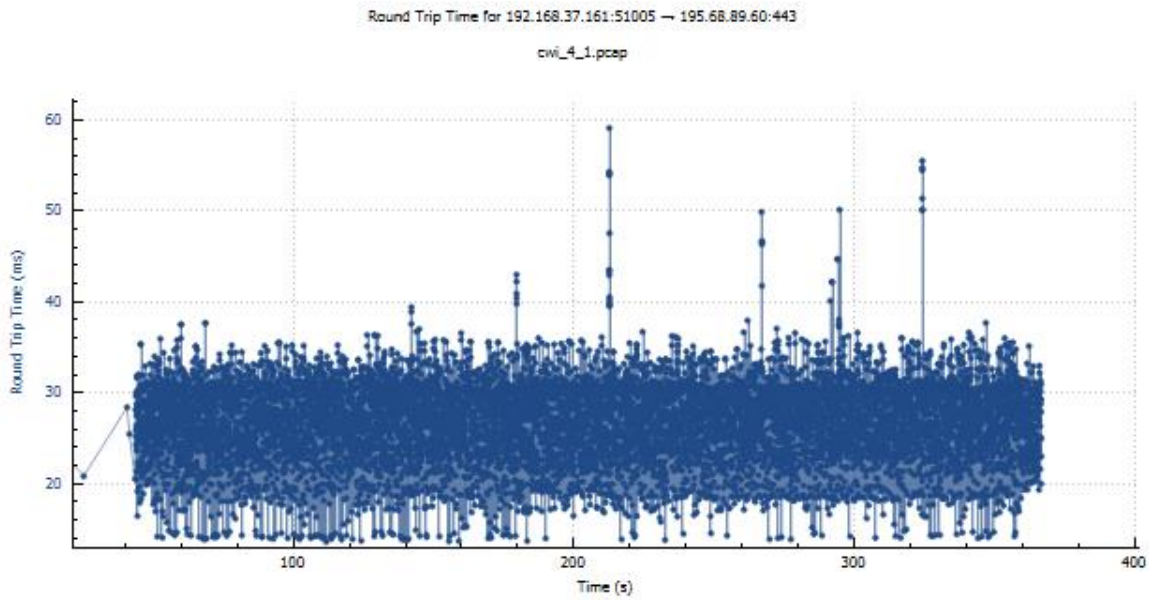


Figure 14 Test 3A: Round Trip Times for TCP segments related to the Full-Capture Point Cloud stream

Test 3B

Resources Consumption Metrics:

Node / Metrics	**Avg. CPU (%)**	**Avg. GPU (%)**	**Avg. RAM (MB)**
i2CAT	17.57	19.48	1038.92
CWI	8.05	36.63	904.95

Table 12 Test 3B: Resources consumption metrics

Point cloud metrics:

Node/Metrics	**Avg. decoded fps**	**Avg. dec #points/cloud**	**Avg latency (ms)**
i2CAT	7.793103448	46397.96552	Measurement Error
CWI	14.73793103	33746.62069	557.9310345

Table 13 Test 3B: Point cloud metrics

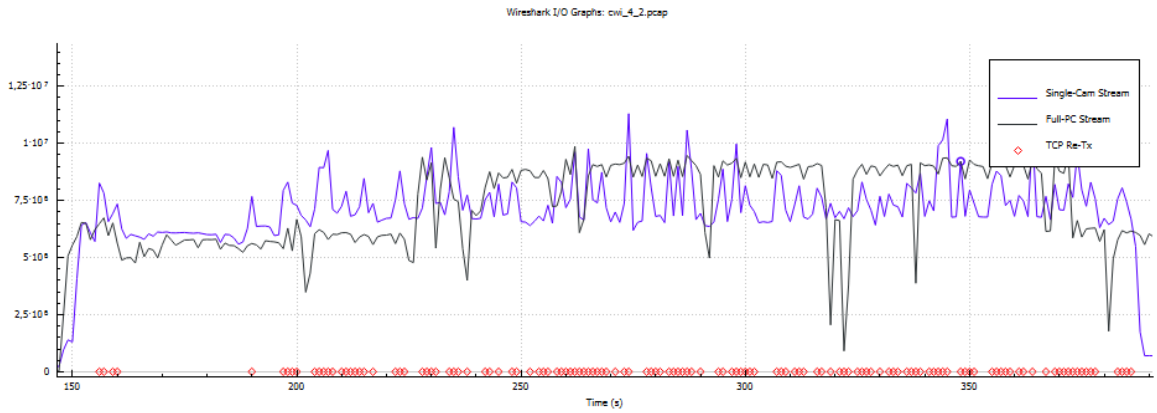


Figure 15 Test 3B: Average Throughput (bps) related to the Full-capture and Single-Camera Point Cloud streams

As can be seen in the previous figure, the TCP throughput evolution for both the Full-Capture and Single-Camera Point Cloud streams was quite stable, ranging mostly between 8-10Mbps, being a bit higher the throughput for the former due to the larger amount of captured points.

Test 4A

Resources Consumption Metrics:

Node / Metrics	Avg. CPU (%)	Avg. GPU (%)	Avg. RAM (MB)
i2CAT	8.13	20.06	757.35
CWI	14.35	42.53	1200.72

Table 14 Test 4A: Resources consumption metrics

Point cloud metrics:

Node/Metrics	Avg. decoded fps	Avg. dec #points/cloud	Avg latency (ms)
i2CAT	5.3125	71701.21875	435.65625
CWI	14.578125	21569.8125	695.375

Table 15 Test 4B: Point cloud metrics

Test 5

Resources Consumption Metrics:

Node / Metrics	Avg. CPU (%)	Avg. GPU (%)	Avg. RAM (MB)
i2CAT	18.75	25.13	964.72
CWI	11.81	47.81	1060.33

Table 16 Test 5: Resources consumption metrics

Point cloud metrics:

Node/Metrics	Avg. decoded fps	Avg. dec #points/cloud	Avg latency (ms)
i2CAT	5.454545455	72585.90909	527

CWI	14.99393939	33278.0303	554.25

Table 17 Test 5: Point cloud metrics

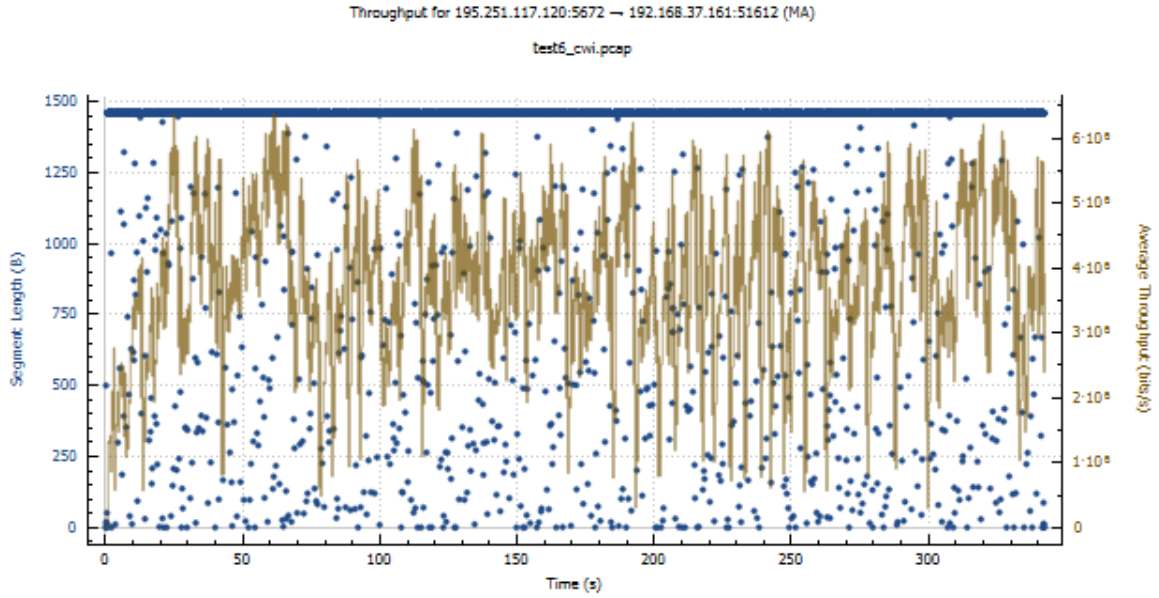


Figure 16 Test 5: TCP Throughput and segment length for the TVM stream

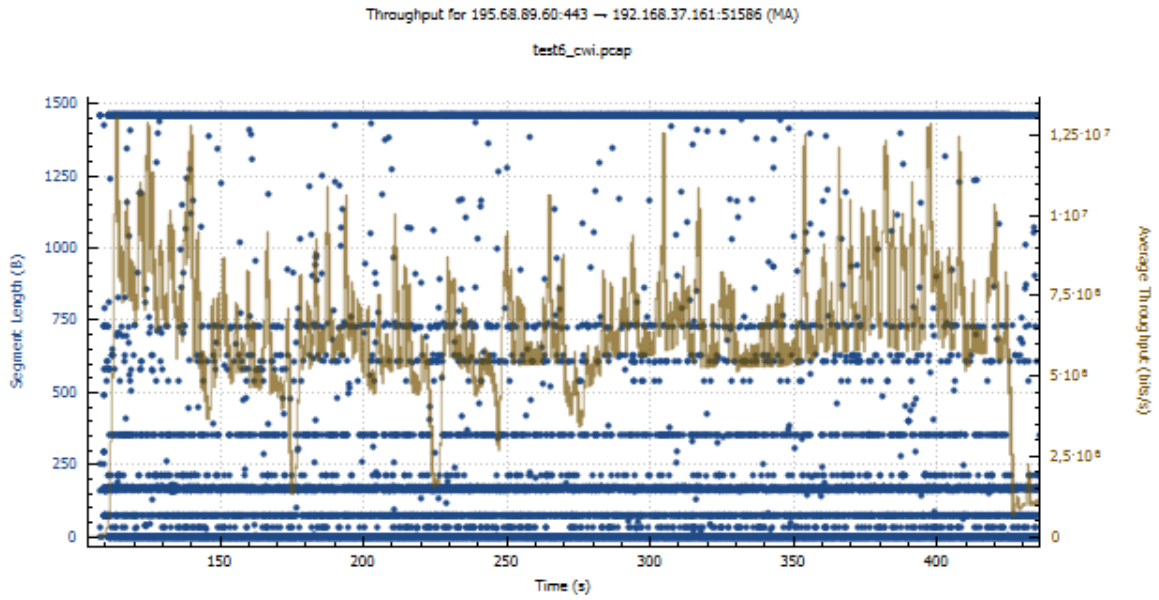


Figure 7 Test 5: TCP Throughput and segment length for the Single Camera Point Cloud stream

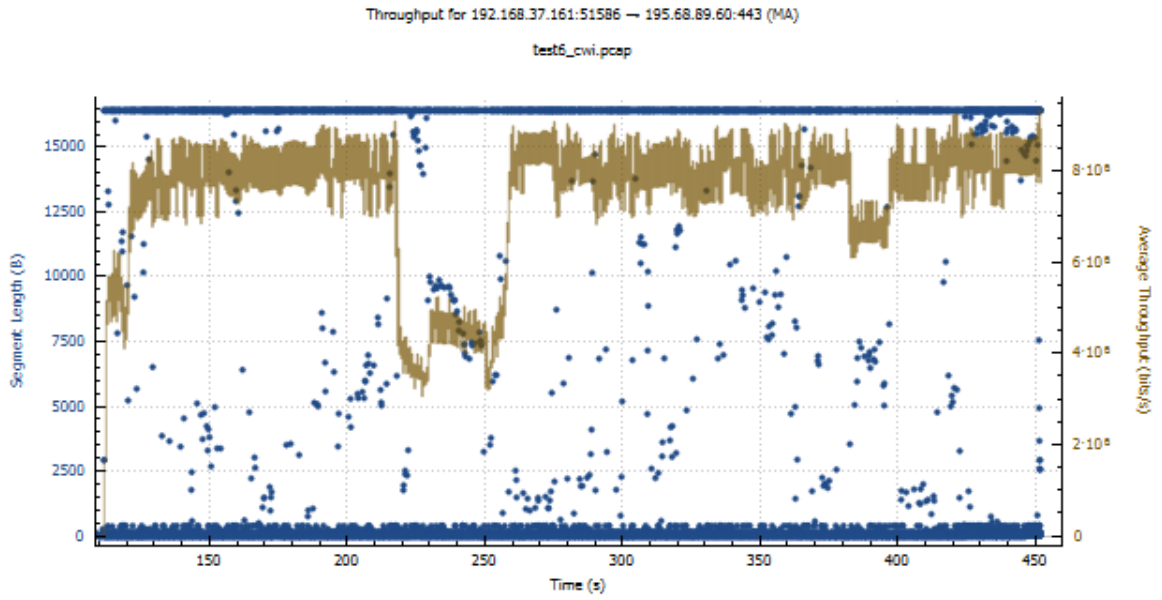


Figure 8 Test 5: TCP Throughput and segment length for the Full-Capture Point Cloud stream

As can be seen from the previous figures, the throughput evolution for each one of the TCP streams was of a similar magnitude than in previous simpler test conditions, so this confirms the successful connection and communication between the three remote nodes, each one using a different representation format. A slight exception occurred though for the Full-Capture Point Cloud stream, for which a decrease of the throughput momentarily happened, as it can also be observed from the graph showing the rate of the TCP sequence numbers increase. However, it is an illustrative case that even in such conditions, the connection is not lost, but even recovered with the expected throughput.

The last figure compares the throughput for the three streams, showing that in this case many TCP re-transmissions happened due to the large amount of exchanged real-time data.

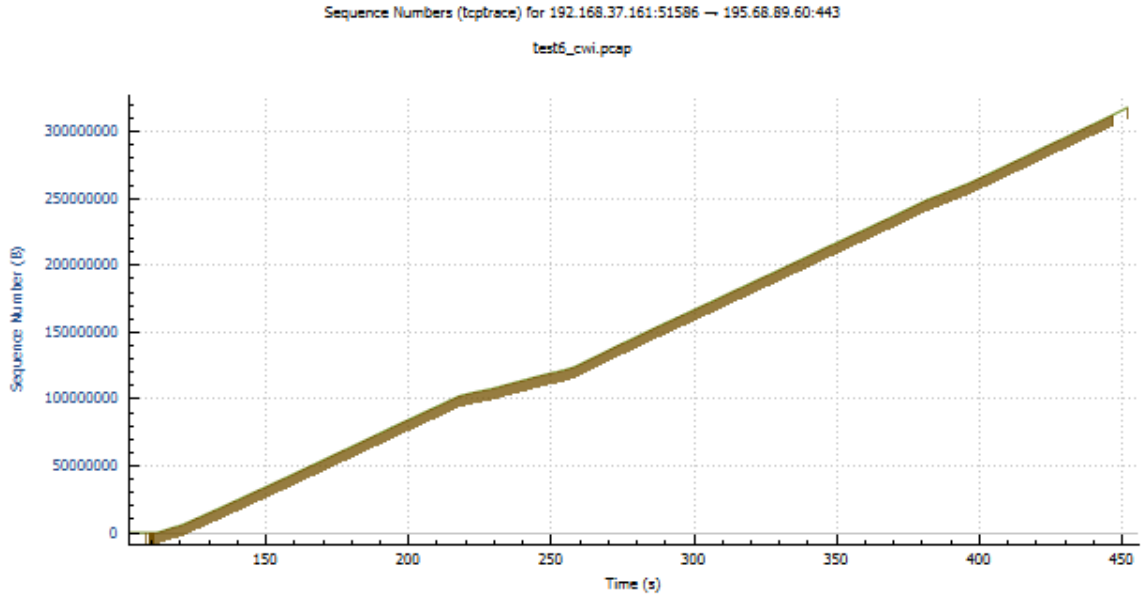


Figure 10 Test 5: TCP Sequence Number Evolution for the Full Capture Point Cloud Stream

Figure 11 Test 5: TCP Throughput for TVM, Full-PC, Single-Cam PC Streams – socket.io (capture at CWI)

Unity pipeline related metrics		Transmission related metrics	
Frames per second	4.8	**Average of message rate in / second of RMQ exchange for the TVM data**	14.84
Average missed/skipped frames	0	**StD of message rate in / second of RMQ exchange for the TVM data**	8.33
Average end-to-end delay (ms)	751.57	**Average of message rate out / second of RMQ exchange for the TVM data**	39.3
StD of end-to-end delay (ms)	187.52	**StD of message rate out / second of RMQ exchange for the TVM data**	30.54
Average size of received compressed TVM (MB)	0.1	**Average of receiving rate (MB / second) of connection to distant RMQ server**	5.84
StD of size of received compressed TVM (MB)	0.01	**StD of receiving rate (MB / second) of connection to distant RMQ server**	5.44
Average size of decompressed TVM (MB)	3.17		
StD of size of decompressed TVM (MB)	0.131		
Average number of vertices per TVM	14103.2		
StD of number of vertices per TVM	3607.4		
Average total deserialization-decompression time per TVM (ms)	10.97		
StD of total deserialization-decompression time per TVM (ms)	2.49		
Average rendering time per TVM (ms)	3.16		
StD of rendering time per TVM (ms)	1.975		

Table 18 Test 5: Unity pipeline average metrics from and between VRT nodes.

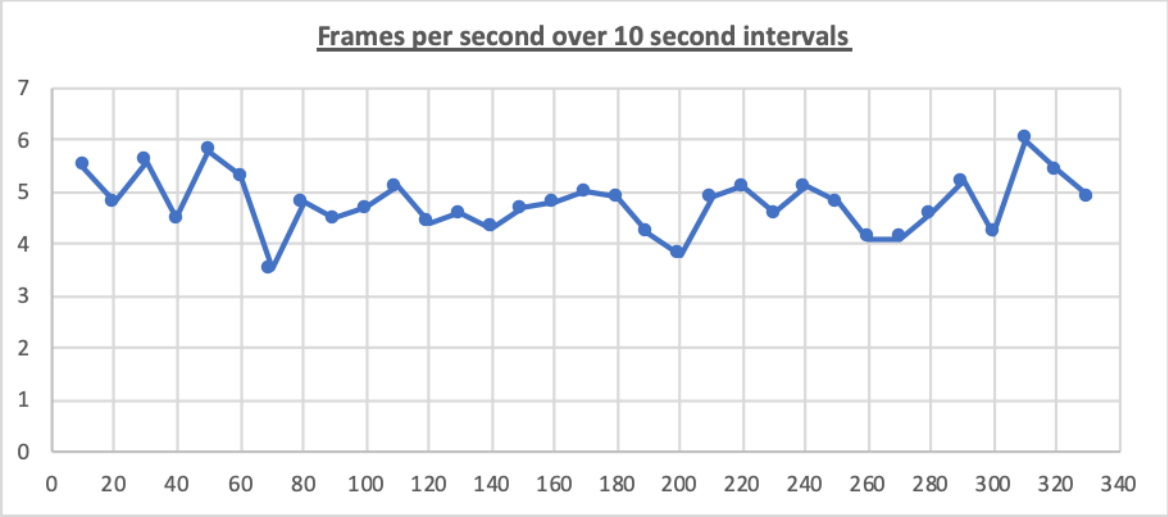


Figure 17 Test 5: FPS during experiment duration

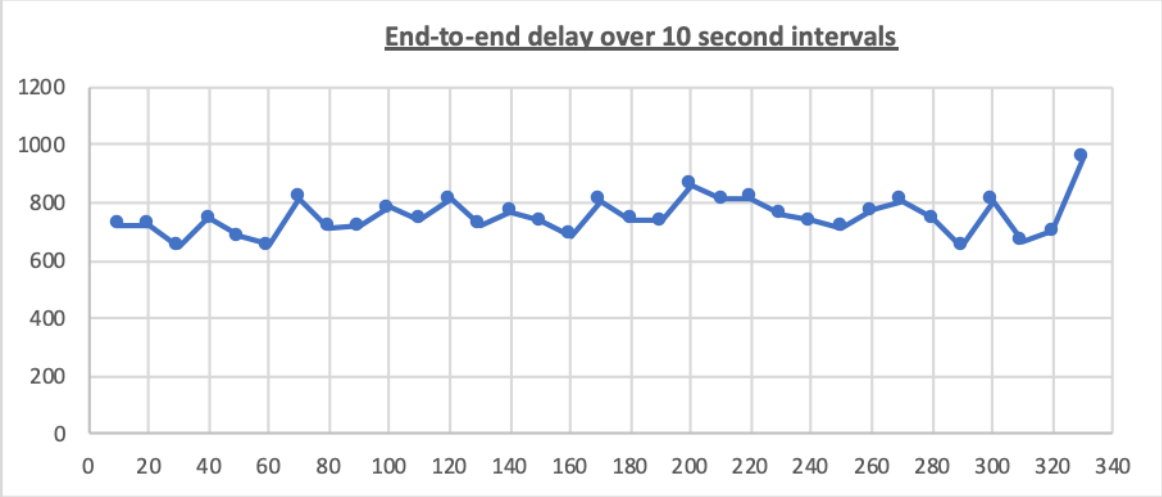


Figure 18 Test 5: End-to-end delay during experiment duration (in ms)

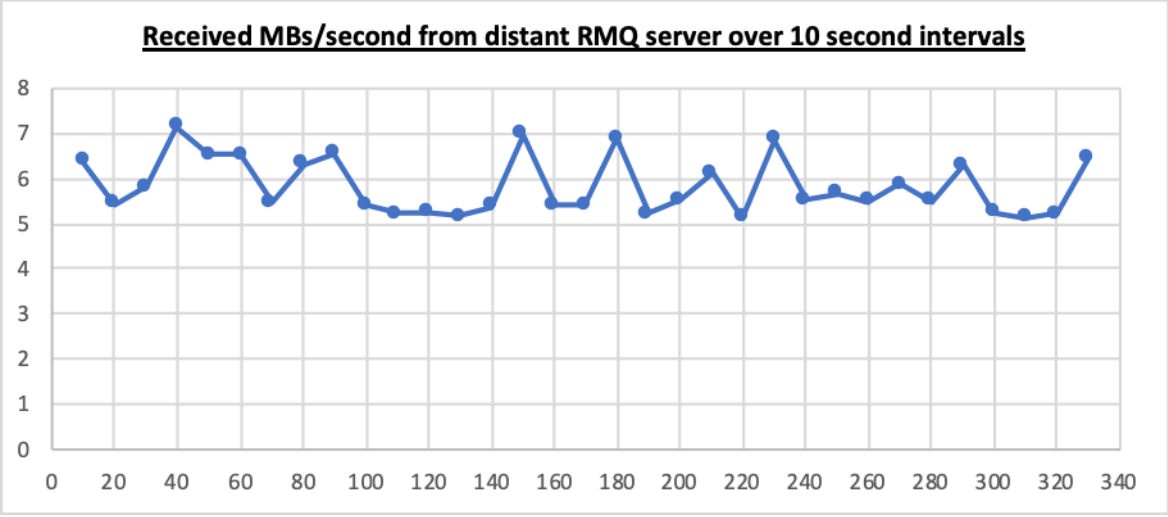


Figure 19 Test 5: Received MB/sec from RMQ server during experiment duration

Test 6

Resources Consumption Metrics:

Node / Metrics	Average CPU (%)	Average GPU (%)	Average RAM (MB)
i2CAT	16.77	24.58	1072.74
CWI	14.52	49.43	1218.65

Table 19 Test 6: Resources consumption metrics

Point cloud metrics:

Node/Metrics	Avg. decoded fps	Avg. dec #points/cloud	Avg latency (ms)
i2CAT	6.3125	66789.96875	650.5
CWI	14.528125	27256.6875	772.15625

Table 20 Test 6: Point cloud metrics

Unity pipeline related metrics		Transmission related metrics	
Frames per second	5.14	**Average of message rate in / second of RMQ exchange for the TVM data**	15.19
Average missed/skipped frames	0	**StD of message rate in / second of RMQ exchange for the TVM data**	8.94
Average end-to-end delay (ms)	714.75	**Average of message rate out / second of RMQ exchange for the TVM data**	37.49
StD of end-to-end delay (ms)	140.45	**StD of message rate out / second of RMQ exchange for the TVM data**	32.48
Average size of received compressed TVM (MB)	0.095	**Average of receiving rate (MB / second) of connection to distant RMQ server**	5.76
StD of size of received compressed TVM (MB)	0.0075	**StD of receiving rate (MB / second) of connection to distant RMQ server**	5.41
Average size of decompressed TVM (MB)	3.12		
StD of size of decompressed TVM (MB)	0.095		
Average number of vertices per TVM	12763.7		
StD of number of vertices per TVM	2691.1		
Average total deserialization-decompression time per TVM (ms)	11.33		

StD of total deserialization-decompression time per TVM (ms)	2.1	
Average rendering time per TVM (ms)	3.21	
StD of rendering time per TVM (ms)	1.94	

Table 21 Test 6: Unity pipeline average metrics from and between VRT nodes.

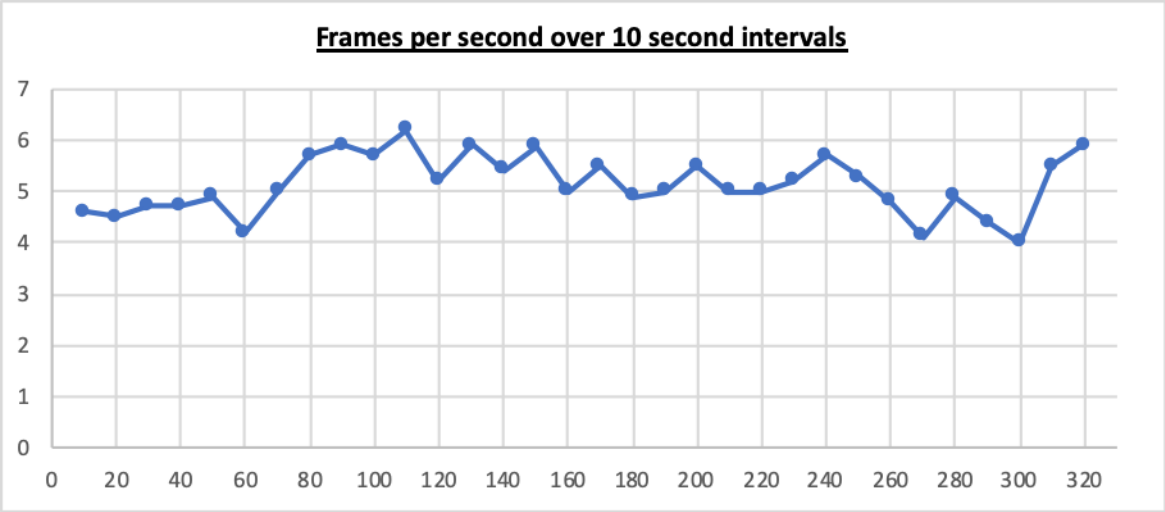


Figure 20 Test 6: FPS during experiment duration

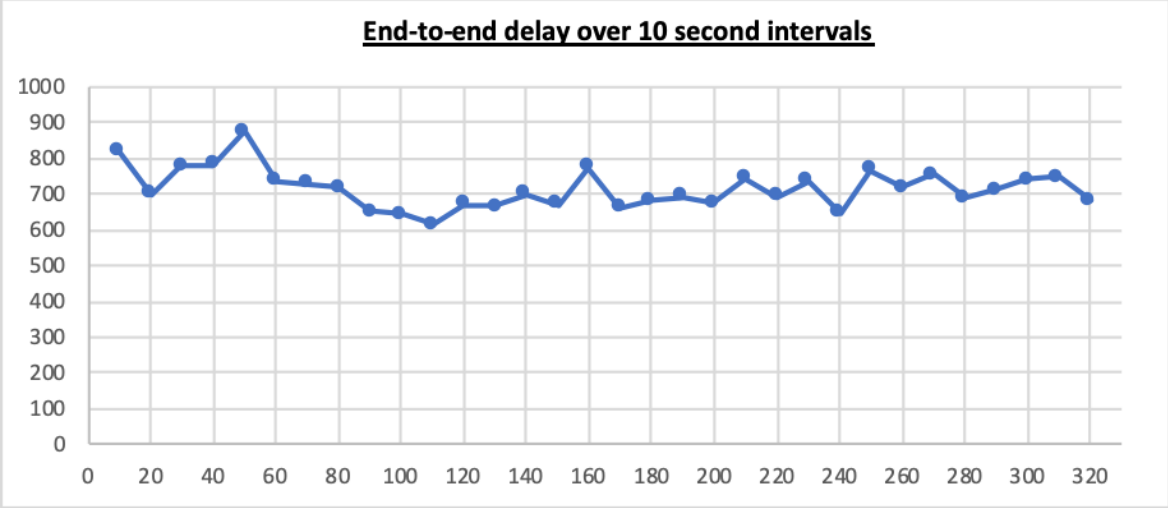


Figure 21 Test 6: End-to-end delay during experiment duration (in ms)

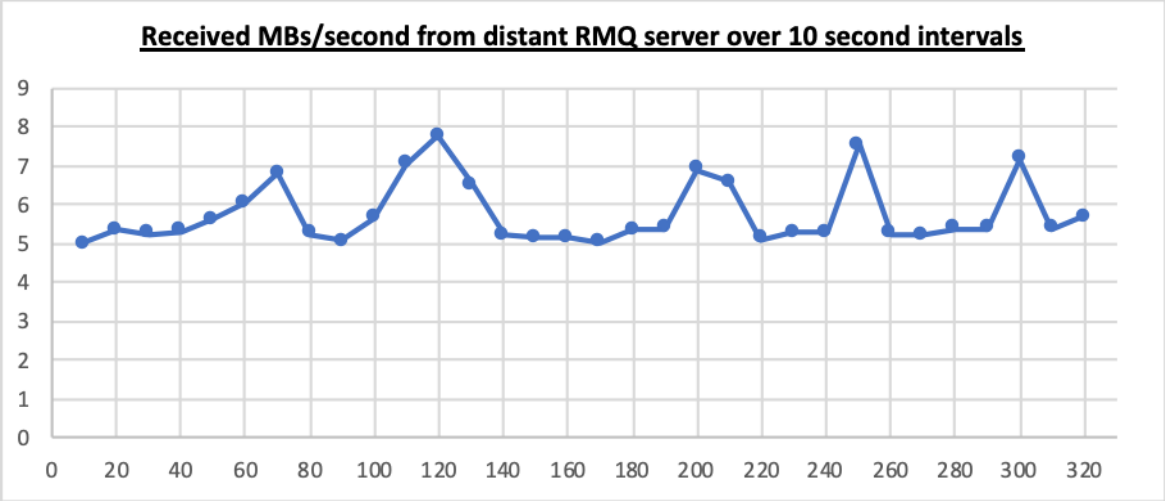


Figure 22 Test 6: Received MB/sec from RMQ server during experiment duration

4.1.5 Conclusion

Concluding, the native client connected user lab measurement experiments proves that the affordability of the VRT pipeline even in multi-player scenarios with heavy loading (Test 5 and 6). The main bottleneck seems to be the transmission part and the network bandwidth limitations due to the massive amount of data, however with reasonable performance under realistic user tests in terms of resource consumption. Native pipeline can offer a ground-breaking social VR experience between remote and multiple users, with low-latency communication, involving cutting-edge media such as point- and mesh-based Volumetric Video. On top of that, observing the results in depth, we deduct the following observations per representation:

Point-cloud specific conclusions:
With respect to the inclusion of Point Cloud representations, it can be concluded that the performance was quite smooth and successful, in term of decoded fps (especially at the CWI node), number of decoded points, delays and throughput. This is true both when using DASH and socket.io as the delivery protocols, and in the different test conditions, from the simpler to the more complex. Despite that some variations in the throughput evolution occurred, the connection was not interrupted and the original expected throughput was recovered. It was also observed that a slight larger throughput was used for the Full-Capture Point Cloud streams than for the Single-Camera ones, due to the large amount of point in the former, as also expected.

TVM specific conclusions:
In these experiments, specifically in Tests 1, 5 and 6, the user node from CERTH participated with one TVM representation. By observing the respective outcomes (Table 7, Table 18, Table 21), one can extract useful findings with respect to the capabilities and operability of the TVM technology/pipeline. TVM, i.e. mesh-based volumetric video, is the most demanding user representation medium of VRTogether with respect to the amount of data, composed of approx. 15.000 vertices per frame, faces and UV coordinates, along with *4* x *640x360* coloured textures, one per device, per user instance. On top of that, beyond raw data acquisition, further processing steps are required to reconstruct the 3D surface in real-time, map and blend the textures.

To this end, many efforts has been devoted to overcome these challenges, developing optimized and highly effective, state-of-the-art techniques and implementing most of the operations of the pipeline with parallel GPU programming. Nevertheless, the main challenge is the transmission of such amounts of data in an efficient way. Indeed, despite that the TVM capturing, compression and

encoding reaches rates higher than 15 FPS (Average of message rate in), being stable during the duration of the experiments despite its varying nature, the receiving, decoding, decompression and rendering rate cannot be higher than 5 FPS (Frames per second in Unity) in remote sessions. In addition, TVM decompression/deserialization και rendering time is low (see Unity pipeline metrics tables in tests 1, 5, 6). Despite the fact that receiving rate (MB/second) is stable during the experiments, delays are present due to network bandwidth limitations, given the massive amount of data, especially for UV texturing, being the main bottleneck in the end-to-end delay issue. In the experiments, we report 0 average skipped frames and TVMs processing time in Unity side is lower than the frame receiving time that depends on the network.

Resources Consumption Metrics:
With respect to the resources, consumption metrics, it can be concluded that all PCs were able to process the considered scenarios without major inconveniences. Of course, the different used PCs have different capabilities, and they were in charge of generating different user representation formats, so the results from them cannot be correlated. However, it can be observed that the usage of resources increased for the more complex setups.

4.2 Web Pipeline: VO-3.1

The main objective is to test the web client performance across different Connected Nodes. Therefore, we conducted a set of user sessions in a realistic setting connecting 4 and 6 users (from Netherlands, France and Germany). The details of the user endpoints used can be found in Table 22. We conducted 6 sessions with a duration of at least 25 minutes each. An example of the user test is shown in Figure 23 and Figure 24. The 4 user sessions were conducted between the nodes NL1, NL2, FR1 and FR2 in 4 conditions: P2P with 2D and 3D presentation, MCU with 2D and 3D representation. As the performance is not stable enough for a 6 user P2P condition, we only tested 2 conditions: MCU with 2D and 3D representation. The 6 user sessions were conducted between the nodes NL2, NL3, NL4, FR1, FR2 and DE1. The results of these tests are presented in the following chapters. CPU, GPU and network performance were measured with the same Resources Consumption Metrics (RCM) measurement tool[3] as in the simulation evaluation, the frame rate was measured via the Aframe stats and the WebRTC delay was measured via the Chrome WebRTC stats.

3

Figure 23. Example of 6 user web client test with MCU transmission in 3D conference room[4].

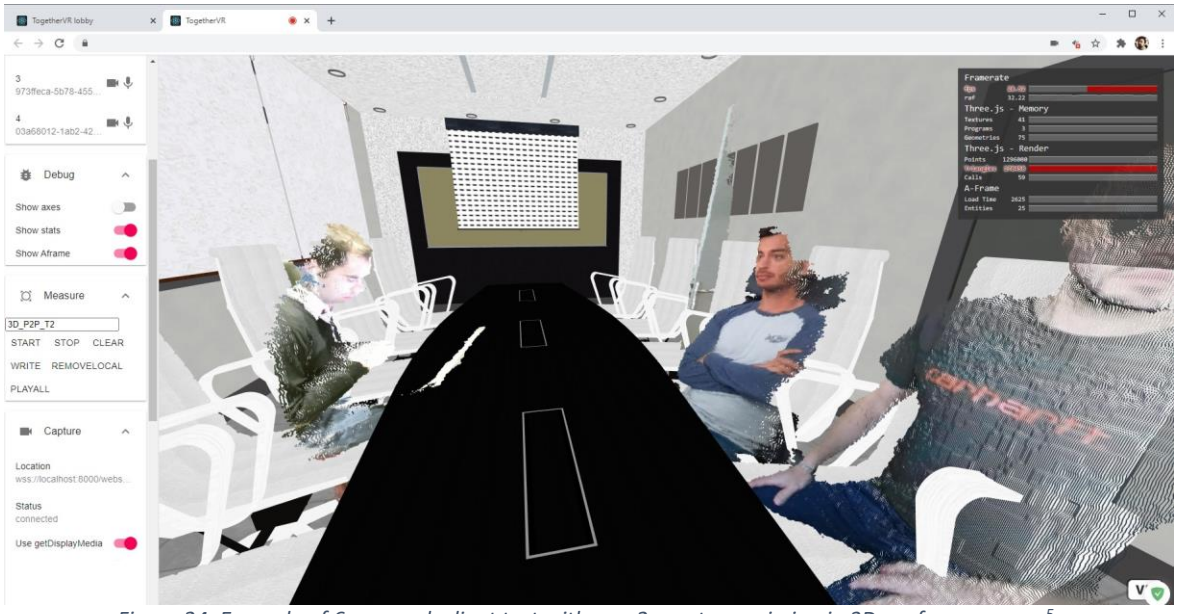


Figure 24. Example of 6 user web client test with peer2peer transmission in 3D conference room[5].

4.2.1 Methodology

Prior to this study we conducted a simulation experiment testing various conditions of the web pipeline and its maximum supported users. In this measurement study we selected the most relevant setup (Criterion Sampling) to test the web pipeline under realistic conditions (True Experiment) with distributed users of the different user labs. Due to covid-19 restrictions and to showcase a wide applicability of the Web Client we use endnotes within the company connected network and in-home network connections. For the measurement we selected 6 conditions (see Experiment sessions) of at least 25 minutes, testing both p2p and MCU transmission as well as 2D and 3D user representations. To simplify the testing and "cognitive load" on the users we

[4] 3D Background Model, Olam Conference Room by Gideon Abochie licensed under CC Attribution
[5] https://github.com/ETSE-UV/RCM-UV

performed our test without a VR HMD. Further we used Google Chrome as browser and both Kinect v2 and Azure Kinect RGBD sensors.

For each condition we followed the following testing procedure:

- Supported browser: chrome (i.e. version 86.X)
- Depth sensor: Kinect V2 or Azure Kinect
- Modern VR capable PC/Laptop
- Procedure for test
    - Start all the capture
    - Start browser and join session
    - (optional) enter VR (requires setup HMD and steam VR)
    - Check if everyone is in the session and visible / audible
    - start all measurement tools
    - min 25 minutes session
    - stop all measurement tools
    - upload measurement files to gdrive (central collection)
        - browser measurements from download folder
        - Process measurements from capture module folder

Experiment sessions
We conducted 6 sessions in the following conditions:

Session Name	Number of user nodes	Type of transmission	Type of user representation
4ppl 2D P2P	4: (NL1, NL2, FR1, FR2)	peer-to-peer (P2P)	2D (Chroma)
4ppl 3D P2P	4: (NL1, NL2, FR1, FR2)	peer-to-peer (P2P)	3D (RGBD)
4ppl 2D MCU	4: (NL1, NL2, FR1, FR2)	Central MCU	2D (Chroma)
4ppl 3D MCU	4: (NL1, NL2, FR1, FR2)	Central MCU	3D (RGBD)
6ppl 2D MCU	6: (NL2-4, FR1/2, DE1)	Central MCU	2D (Chroma)
6ppl 3D MCU	6: (NL2-4, FR1/2, DE1)	Central MCU	3D (RGBD)

Table 22. Measurement conditions.

With the following set of user end point configurations:

Name	CPU	GPU	Memory	Sensor	Location
NL1	Intel Core i7-8750H CPU @ 2.20GHz (6 cores / 12LPUs)	NVIDIA GeForce GTX 1070 Max-Q	32 GB	Azure Kinect	NL, Amsterdam
NL2	Intel Core i7-8700 CPU @ 3.20GHz	NVIDIA GeForce RTX 2080	32 GB	Azure Kinect	NL, Katwijk
NL3	Intel Core i7-6700K CPU @ 4.00Ghz (4 cores / 8 LPUs)	NVIDIA GeForce GTX 980 Ti	16 GB	Kinect V2	NL, The Hague

NL4	Intel Core i7-7820HK CPU @ 2.9GHz (4 cores / 8LPUs)	NVIDIA GeForce GTX 1070	24 GB	Azure Kinect	NL, Enschede
FR1	Intel Core i7-8750H CPU @ 2.20GHz	NVIDIA GeForce GTX 1070	16 GM	Kinect V2	FR, Rennes
FR2	Intel Core i7-47770K CPU @ 3.50GHz	NVIDIA GeForce GTX 1050 Ti	16 GB	Kinect V2	FR, Paris
DE1	Intel Core i7-8750H CPU @ 2.20GHz (6 cores / 12LPUs)	NVIDIA GeForce GTX 1070 Max-Q	32 GB	Azure Kinect	DE, Berlin

Table 23. Overview of measurement endpoints.

4.2.2 Metrics

In each condition collected the following metrics:

- WebGL/Aframe Rendering stats (https://github.com/aframevr/aframe/blob/master/docs/components/stats.md)
  - fps: frames per second, framerate.
  - requestAnimationFrame (raf): Latency.
  - Textures: number of three.js textures in the scene. A lower count means the scene is using less memory and sending less data to the GPU.
  - Programs: number of GLSL shaders in the scene.
  - Geometries: number of three.js geometries in the scene. A lower count means the scene is using less memory.
  - Vertices: number of vertices in the scene.
  - Faces: number of faces in the scene.
  - Calls: number of draw calls on each frame.
  - Load Time: how long it took for the scene to start rendering, in ms.
  - Entities: number of A-Frame entities.
- WebRTC stats
  - Video upload delay
  - Video upload jitter
  - Audio upload delay
  - Audio upload jitter
- Chrome Process (CPU / GPU / memory)
- Depth Sensor Process (CPU / GPU / memory)
- Capture Module Process (CPU / GPU / memory)
- Network
  - total upload bandwidth
  - total download bandwidth

4.2.3 Results

In the following we present the most relevant results from this measurement study.

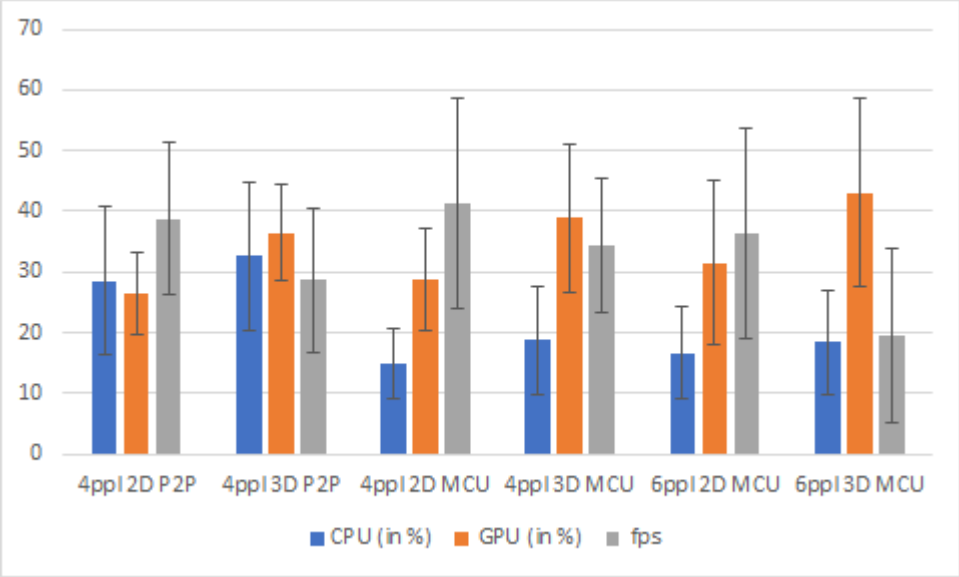


Figure 25. performance statistics.

Figure 25 shows the overall performance average per condition (of all user endpoints) of the Chrome instance running the Web client in terms of CPU, GPU and rendering frame rate. The CPU and GPU are overall in an acceptable range with a clear trend in between the MCU and P2P: The MCU condition allows to reduce the CPU load on the cost of GPU usage. As CPU resources are sparser, and necessary for many more processes (like capture and the VR HMD) this is particularly beneficial to support more simultaneous users and constant high quality rendering. Important to note is that the frame rates are only indicative and not realistic for the rendering performance in an VR HMD. This is, we conducted the evaluation without a VR HMD to simplify the measurements and interactions between users. Further the browser executes various optimizations strategies to balance performance load with visual rendering quality. None of the users perceive stuttering or visual impact due to performance and the CPU/GPU load was low enough to allow higher frame rates in VR mode.

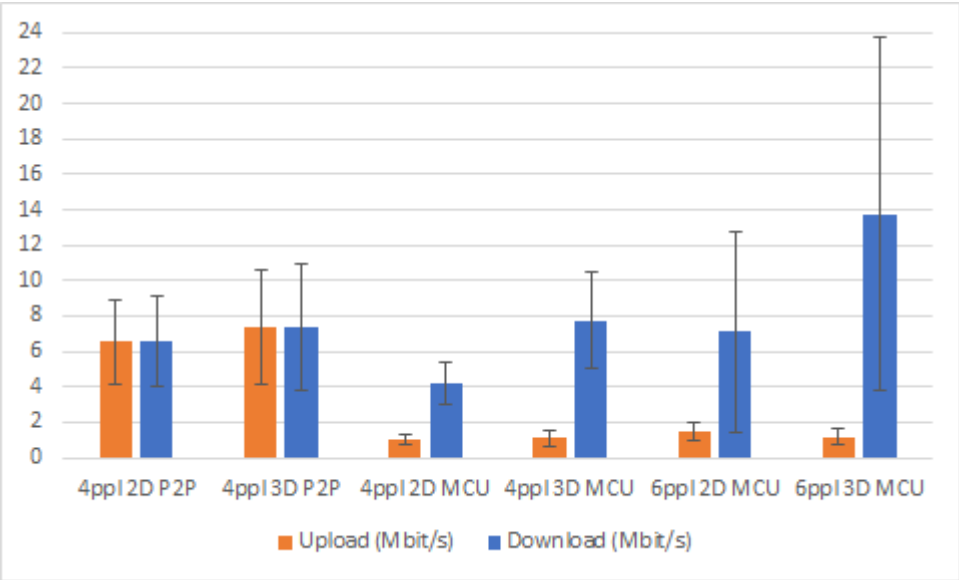


Figure 26. Network upload / download utilization.

Figure 26 shows the overall average per condition (of all user endpoints) in network traffic. Our results reflect the main benefits of a central WebRTC approach (MCU) as the upload traffic is significantly decreased. This is, in the MCU condition the representation of a user is only uploaded

once, while in the P2P condition the user representation has to be uploaded to each other endpoint. Overall, the MCU is capable to use network resources much more efficiently (on the cost of central computation in the server).

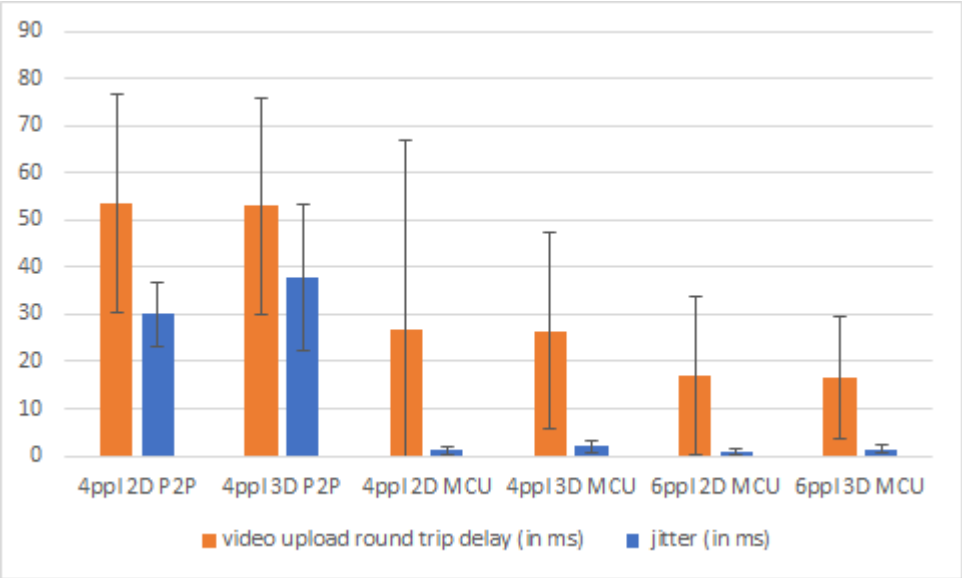


Figure 27. Video upload latency and jitter.

Figure 27 shows the overall average per condition (of all user end points) of the video upload round trip delay and jitter. These delays are additional to the overall glass-to-glass delay (peer-to-peer delay 2D: 396ms & 3D: 384ms; MCU delay 2D: 564ms & 3D: 622ms). We can observe an overall higher delay and jitter for P2P transmission compared to central MCU transmission. However, one condition "4 users MCU with 2D representation" shows a high standard derivation as one client (FR1) observed higher delay values. This is to be expected in such a test (with a realistic and varying internet connection) and is in the normal boundaries of delay to expect for WebRTC transmission (and in the delay range acceptable for remote communication).

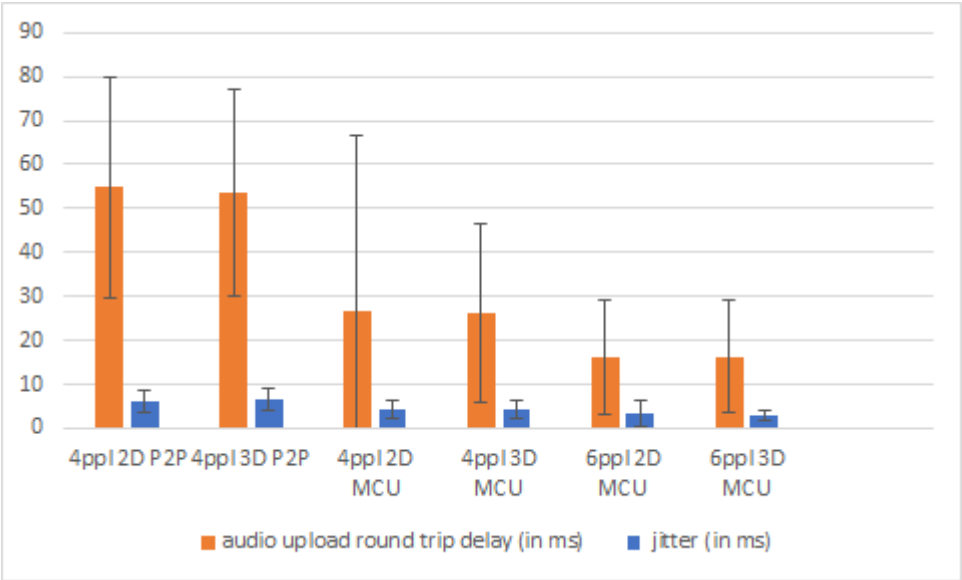


Figure 28. Audio upload latency and jitter.

Figure 28 shows the overall average per condition (of all user endpoints) of the audio upload round trip delay and jitter. This is, the delays are practically the same (or at least very similar) to the video

delay measurements, with lower jitter in the peer-to-peer condition (wish to be expected as the data rate of audio streams is significantly lower than for video).

4.2.4 Conclusion

Overall, the Web client connected user lab measurement study confirms the technical readiness of the system. It shows reasonable performance load under realistic user conditions both in terms of resource usage and network utilisation. The results show that we can achieve VR communication over the Web with similar network transfer rates and slightly more CPU/GPU resource usage compared to video conferencing solutions while being able to render users both in 2D and volumetric 3D (captured with one RGBD sensor).

5 USER LABS: PILOT ACTIONS

During the third year of the project we have run a number of pilot actions that help us paving the way to the deployment and execution of Pilot 3. Overall, nine experiments have been coordinated for exploring different aspects of the project:

- Benchmarking of Social VR Systems (VRT-3.1)
- Evaluation of the Users and Environment
    - CERTH-3.2 = comparison of TVMs
    - I2CAT-3.1 = Subtitling 3D VR Content
    - I2CAT-3.2 = Evaluation of use cases
- Evaluation of the underlying technology
    - CWI-3.2 = Point Clouds Tiling
    - TNO-3.1 = Evaluating tethered and non-tethered HMD's
    - I2CAT-3.3 = Evaluation of the PC-MCU
    - CERTH-3.3 = Transmission rate TVM comparison
    - CERTH-3.4 = Pre-Pilot Technology Test

These activities required of major managerial effort, due to the global pandemic, forcing the research team to re-evaluate priorities and objectives around March 2020. The work package leaders closely followed the situation and put in place the mechanisms to monitor the situation, predict risks, and palliate them. In the end, a number of initially proposed experiments could not be run, new experiments were defined, and others were slightly modify to follow existing health-related regulations (see D2.3).

5.1 VRT-3.1: Benchmarking of Social VR Systems

5.1.1 Motivation / Objective

The VR-Together project not only targets the development of outstanding Social VR solutions, but also considers the analysis and comparison of existing reference platforms and tools in order to: validate the benefits and innovative aspects of our solutions; to extract valuable ideas; to assess potential integration possibilities, thus feeding the technology transfer and exploitation activities; and to analyse the target use cases in which our platform could be used.

In this context, this pilot action reports on the analysis and benchmarking of seven relevant state-of-the-art Social VR platforms, taking into account relevant technical and user experience related aspects.

5.1.1. Previous Related Activities in the VR-Together project

The project has devoted previous efforts to analysing and testing existing Social VR platforms, even comparing them to traditional shared media watching platforms. Next, pilot actions executed in this context are summarized:

1. Pilot Action 1: Experiment on Photo Sharing (CWI-2, see D4.2)
    - Scenario Description: evaluate a photo sharing experience between two users using different available tools/platforms.
    - Platforms Used:
        1) Baseline (Face-to-Face scenario);

2) Traditional social video platform (Skype);
3) State-of-the-art Social VR platform (Facebook Spaces)

- Sample: N=52 participants
- Results / Insights:
  - The designed evaluation methodology for Social VR is appropriate
  - Social VR provides a better experience than Skype
- Related Publication: *J. Li, Y. Kong, T. Roggla, F. De Simone, S. Ananthanarayan, H. de Ridder, A. El Ali, and P. Cesar, "Measuring and Understanding Photo Sharing Experiences in Social Virtual Reality", ACM CHI, Glasgow, UK, May 4-9, 2019*

2. Pilot Action 2: Experiment on Shared Video Watching (CWI-3)
   - Scenario Description: evaluate a shared video watching experience between two users, by using movie trailers.
   - Platforms Used:
     1) Baseline (Face-to-Face);
     2) State-of-the-art Social VR platform (Facebook Spaces);
     3) VR-Together Social VR Platform (Web Platform)
   - Sample: N=32 participants
   - Results / Insights:
     - The designed evaluation methodology for Social VR is appropriate
     - The VR-Together Social VR Platform (Web) provides a better experience than Facebook Spaces, and comparable Quality of Interaction (QoI) to Face-to-Face scenarios

3. Pilot Action 3: Pilot 1
   - Scenario Description: evaluate a shared video watching experience between two users, by using an immersive VR story.
   - Platforms Used:
     1) VR-Together Social VR Platform (Native Platform, TVMs)
   - Sample: N=30 participants
   - Results / Insights:
     - The designed evaluation methodology for Social VR is appropriate
     - Although using a different VR scenario and content, the VR-Together Social VR Platform (Native Platform, TVMs) provides comparable levels of immersion, social presence and Quality of Interaction (QoI) to VR-Together Social VR Platform (Web Platform), and thus to Face-to-Face scenarios (according to Pilot Action 2).
   - Related Publication (Under Review): *M. Montagud, J. Li, G. Cernigliaro, A. El Ali, S. Fernández, P. Cesar. 2020. Towards SocialVR: Evaluating a Novel Technology for Watching Videos Together, Submitted in September 2020, 26 pages*.

4. Pilot Action 4: Online Events through Social VR
   - VR-Together members have co-organized international Workshops on Social VR (https://www.socialvr-ws.com/) and tracks of the ACM IMX 2020 conference (http://imx.acm.org/2020) using state-of-the-art Social VR platforms, like Mozilla Hubs.
   - Sample:
     - Social VR workshop: N=20 participants

    - ACM IMX Social VR sessions: around N=190 attendees, but distributed over different connected virtual rooms in a ring to prevent more than 20 simultaneous users in each one.
- Results / Insights:
    - These actions proved that Social VR is an appropriate and promising medium to host virtual events, especially in pandemic times.

In addition, within the umbrella of Work Package 5 (WP5), Deliverable D5.2 has provided a comparison chart between 16 state-of-the-art Social VR platforms as of November 2019, elaborated by Ryan Schultz: https://ryanschultz.com/2019/11/12/an-updated-comparison-chart-of-sixteen-social-vr-platforms-first-draft-november-2019/

The comparison chart is also available as a Google Sheet, and is sketched in Table 24. In particular, the aspects considered in that comparison are:

- Name (and Company developing / offering the platform)
- Purpose of the Platform
- Desktop (non VR) OS Support.
- VR Headset Support
- Mobile Support
- Avatars Can Freely Move Around (i.e. 6DoF, 6 Degrees of Freedom)
- Default Avatars
- Default Dressable Human Avatars & Fashion Market
- Can Create Custom Rigged Avatars
- Shopping
- Currency
- In-World Building tools (i.e. ability to create complex objects entirely within the platform itself, and not using external tools such as Blender or Unity and then importing the externally-created objects into the platform).
- Architecture/game engine
- Scripting
- Open/closed source

This is a very valuable comparison for VR-Together, but it is highly focused on high-level features and commercial aspects. Within the umbrella of WP4, and more particularly as part of the "Benchmarking" sub-task considered in *Task 4.3. Evaluation*, it was decided to further elaborate on this comparison to gain deeper insights about state-of-the-art solutions in the Social VR field.

Another very useful list of Social VR platforms, with demo videos, is provided here: https://medium.com/immersively/vr-for-virtual-meetings-the-ultimate-guide-d0c9ebe634d4

Finally, it should be remarked that after conducting this analysis, it was published another relevant analysis (published in May 2020, and found in September 2020) that also compared existing Social VR platform. The full report can be found here, while a summary categorization table can be found here.

Our analysis has also been enriched by leveraging the insights from these additional ones.

Comparison Chart of 16 Social VR Platforms (November 2019) : Sheet1

Comparison Chart of 16 Social VR Platforms (First Draft © Ryan Schultz, Published to RyanSchultz.com, November 12th 2019)

Name	Company	Purpose of Platform	Desktop (Non-VR) OS Support	VR Headset Support	Mobile Support	Avatars Can Freely Move Around	Default Avatars	Default Dressable Human Avatars & Fashion Market*	Can Create Custom Rigged Avatars	Shopping	Currency	In-World Building Tools
AltspaceVR	Microsoft	General Purpose	Windows	Oculus Rift, Oculus Quest, Oculus Go, HTC Vive, Windows MR, Gear VR, Google Daydream	Android	Yes	Cartoon-Like Avatars	No	No	No	None	No
Anyland	Anyland	General Purpose	Oculus Rift, HTC Vive, Valve Index, Windows MR	Oculus Rift, HTC Valve, Valve Index, Windows MR	None	Yes	A Pair of Hands (You Build Your Own Avatar)	No	No	No	None	Yes
Bigscreen	Bigscreen, Inc.	Media Consumption (Video, TV, Movies)	Windows	Oculus Rift, Oculus Quest, Oculus Go, HTC Vive, Windows MR, Gear VR	None		Cartoon-Like Avatars	No	No	No	None	No
Cryptovoxels	Nolan Consulting Ltd.	General Purpose	Windows, MacOS, Linux (Web Based)	Any that support WebVR: Oculus Rift, Oculus Go, HTC Vive, Windows MR, Gear VR, Google Daydream	iOS and Android	Yes	Artist Mannequin Avatar (Not Changeable)	No; Avatar Attachments Only	No	Blockchain-Based Land Parcels and Avatar Attachments	Ethereum (Cryptocurrency)	Yes (Voxel Building)
Engage	VR Education Holdings PLC	Education	Oculus Rift, Oculus Quest (via SideQuest), HTC Vive, Valve Index, Windows MR	Oculus Rift	None	Yes	Human Avatars	No	No	No	None	Yes (IFX System)
High Fidelity	High Fidelity	Business/Remote Workteams (Formerly General Purpose)	Windows, MacOS	Oculus Rift, HTC Vive, Valve Index, Windows MR	Android	Yes	Artist Mannequin Avatar (You Can Create an Avatar Based on a Selfie Using an iOS or Android App)	No; Avatar Attachments Only	Yes	In-Client Shopping and High Fidelity Marketplace	High Fidelity Coin (Cryptocurrency)	Yes (Limited)
JanusVR	JanusVR	General Purpose	Windows, MacOS, Linux	Any that support WebVR: Oculus Rift, Oculus Go, HTC Vive, Windows MR, Gear VR, Google Daydream	iOS and Android	Yes	Artist Mannequin Avatars	No	No	No	None	No
Mozilla Hubs	Mozilla	General Purpose	Windows, MacOS, Linux (Web Based)	Any that support WebVR: Oculus Rift, Oculus Go, HTC Vive, Windows MR, Gear VR, Google Daydream	iOS and Android	Yes	Cartoon-Like Avatars	No	No	No	None	Yes
NeosVR	Neos VR Metaverse	General Purpose	Oculus Rift, HTC Vive, Valve Index, Windows MR	Oculus Rift, HTC Vive, Valve Index, Windows MR	None	Yes	Cartoon-Like Avatars	No	Yes	No	Neos Credits (Plans to Move to Cryptocurrency)	Yes
Rec Room	Against Gravity	Games	Windows	Oculus Rift, Oculus Quest, HTC Vive, Windows MR, PSVR	iOS	Yes	Cartoon-Like Avatars	No	No	In-World Store	Rec Room Tokens	No
Sansar	Linden Lab	General Purpose (New Focus on Live Events)	Windows	Oculus Rift, HTC Vive, Windows MR (unofficial)	None	Yes	Human Avatars	Yes	Yes	In-Client Sopping and Sansar Store	Sansar Dollars	No
Sinespace	Sine Wave Entertainment	General Purpose	Windows, MacOS, Linux	Oculus Rift, HTC Vive, Windows MR (unofficial)	Android and iOS	Yes	Human Avatars	Yes	Yes	In-World Stores, In-Client Shopping, and Sinespace Shop	Gold Credits and Silver Credits	
Somnium Space	Somnium Space	General Purpose	Windows	Oculus Rift, HTC Vive, Windows MR, Gear VR, Google Daydream	None	Yes	Head-and-Shoulders Avatars (Full-Body Avatars are Planned)	No	No	Blockchain-Based Land Parcels and Objects	Cubes (Cryptocurrency)	Yes
VRChat	VRChat Inc.	General Purpose	Windows	Oculus Rift, Oculus Quest, HTC Vive, Windows MR	None	Yes	Cartoon or Human Avatars (You Can Create a Custom Avatar Using the Tafi Beta App)	No	Yes	No (Active External Market for Custom Avatars)	None	No
vTime XR	vTime Limited	Chat	Windows	Oculus Rift, Oculus Go,	iOS	No, Locked to	Human Avatars	No	No	No	None	No

Published by Google Sheets – Report Abuse – Updated automatically every 5 minutes

[NOTE: Facebook Spaces is not included in this table, because it is not available anymore since October 2019, and few details about Facebook Horizon were available at the time the table was created]

Table 24. Comparison of Social VR platforms, by Ryan Schultz (November 2019).

5.1.2 Methodology

Using the above comparison chart as a starting point, this pilot action consisted of analysing and testing one Social VR platform per partner, adding further technological and user experience aspects in the benchmark and comparison analysis.

On the one hand, beyond the aspects reviewed in the above comparison, our analysis includes the next additional aspects (if these aspects are publicly available and/or can be known with an external analysis):

- Architecture/game engine
- Availability of SDK
- 3D Environments
- 6DoF Support
- End-User Representation Type (Cartoon Avatars, Human Avatars, 2D Human Reconstruction; 3D Volumetric Human Reconstruction)
    - If Avatars: Is Customization Possible? Can avatars be imported?
    - If Realistic Representations: Types of media, types and number of capturing cameras.
    - If realistic representations: Also valid for AR / XR?
- Other interaction modalities supported, for instance: audio chat, text chat, shared board, shared video watching, interaction with the environment

- Supported Media Types (More than 1?)
- Is it possible to select between VR environments?
- Is it possible to import VR environments, or just a default set is available?
- Is it possible to have a private room or environment
- Targeted Use Cases
- Exploitation Strategy (Link for online service?)
- Known Adopters / Clients
- Technological Aspect aspects
    - Delivery Technology (DASH, WebRTC, proprietary...)
    - Scalability (number of supported users per session)
    - Bandwidth consumption for a pre-defined scenario (if possible)
    - CPU/GPU/RAM consumption for a pre-defined scenario (if possible)
    - Latency for a pre-defined scenario, if not possible to measure – rank it subjectively.
- Is Integration with VR-Together solutions possible?
- Remarkable limitations (ideally addressed in VR-Together)

On the other hand, the list of selected and analysed Social VR platforms / solutions by the consortium are detailed in Table 25. Beyond adding other relevant and non-considered aspects for the Social VR platforms compared to the ones in existing comparisons, our analysis incorporates additional recent Social VR platforms, like: Virbela, Vive Sync, and Spatial.io.

Partner	Social VR platform / solution
Motion Spell	AltSpaceVR
Viaccess Orca	BigScreen
CWI	Mozilla Hubs
CERTH	NeosVR
I2CAT	Spatial.io
TNO	Virbela
ARTANIM	Vive Sync

Table 25. Analysed Social VR platforms.

5.1.3 Results

Next, the results from each of the Social VR platforms listed in Table 25 are presented.

5.1.3.1. AltspaceVR platform

- Name (and Company developing / offering the platform): AltspaceVR is a company founded in 2013, and bought by Microsoft in 2017.
- Link: https://altvr.com/
- Desktop (non-VR) OS Support: Yes, 2D mode is available from web browsers or from a Windows app (Windows Store).
- Mobile Support: Yes, using Samsung Gear VR.
- VR Headset Support: HTC Vive, Oculus Rift CV1, Samsung Gear VR, Oculus Go, Oculus Quest, Windows Mixed Reality.
- Architecture/game engine: Several references to Unity are added in the documentation.
- Scripting: No. A SDK is available to create VR spaces.

- Open/closed source: Closed source.
- Availability of SDK: Yes. See: https://github.com/AltspaceVR/AltspaceSDK Relevant qoute: "*the AltspaceVR SDK can be used together with Three.js or A-Frame to create holographic, multi-user web apps for virtual reality*."
- 3D Environments: Yes, and it is possible to import glTf objects.
- 6DoF Support: Yes, and compatible with 6DoF headsets.
    - Teleport available
- End-User Representation Type (Cartoon Avatars, Human Avatars, 2D Human Reconstruction; 3D Volumetric Human Reconstruction):
    - Human-looking avatars, with customisable faces, but no clothes
    - Avatars cannot be imported.
    - No Realistic Users' Representations

Figure 29. Avatars in AltSpaceVR.

- Other interaction modalities supported:
    - Video Watching: yes.
    - Audio chat: yes, but everyone seems to be in the same room.
    - Text chat: no.
    - Shared board: somehow with slides.com "Remote Control Link".
    - Emojis
    - Raise Hands
    - Interaction with environment: very limited: only limits of the physical space.

Figure 30. Slides sharing in AltSpaceVR.

- Is it possible to select between VR environments?: Yes, several rooms are available.
- Is it possible to import VR environments, or just a default set is available?: Yes, it is possible to import VR environments by using the SDK.
- Is it possible to have a private room or environment?: Yes.
- Supported Media Types:
    - Full 3D VR.

    - Avatars have a 3D look.
    - 2D media
- Targeted Use Cases: AltspaceVr define themselves as a generalist platform. Events and rooms include mediation, conferences, open-talk…
- Exploitation Strategy (Link for Online Service?): Bought by Microsoft. Looks like an analysis about future Social VR adoption rather than an actual commercial service.
- Known Adopters / Clients: None, but bought by Microsoft. From the Reddit channel, the activity seems to have decreased until September 2019. Since then more activity.
- Integration with VR-Together solutions possible?: Maybe. AltspaceVR used to have some Kinect integration (not tested) to mimic users body language.
- Remarkable limitations (ideally addressed in VR-Together)
    - Non-realistic avatars.
    - Content is poor (rooms, interactions…)
    - Most of the VR world seem to be filled just by teenagers.
- Technological aspects
    - Delivery Technology (DASH, WebRTC, proprietary...): WebRTC
    - Scalability (number of supported users per session): only experience with sessions with up to 30 users, which were still stable.
        - According to documentation: "*Spaces can hold up to 70 avatars and when that is reached, avatars can be mirrored over to replicated spaces and the presenters still appear present via the Front Row feature.*"
    - Bandwidth consumption (for a pre-defined scenario): between 200Kbps and 20Mbps, in a range of scenarios
    - CPU/GPU/RAM Usage / Requirements: CPU 30.2%, GPU 74.7%, RAM 1.4GB (tested on a PC with Windows 10, Core i7 4710HQ, nVidia GTX860M, 16GB RAM)
- User Experience aspects
    - Latency: it was low, and not annoying. Around 300ms (tested by clapping hands)
    - Quality of the VR environment: Poor. It looks like an old video game.
    - Quality of end-users' representations: Avatars are OK, and work good enough for people to identify themselves in the world.
    - Quality of Auditory Interaction: Good. Spatial audio, but the environment easily becomes noisy when other discussions are happening around.
    - Quality of Visual Interaction: Ok. The cartoon aspect of the scene makes it high-contrast. Users and content are easy to locate.
    - Quality of 6DoF (Quality of Navigation): Satisfactory.
    - Smooth playout: Good. Only a few minor glitches in the first minutes
    - Ease of setup: Very easy.
    - Usability: Good.
        - Aside from the meditation room and the technical presentations, other rooms were rather disorganized.

5.1.3.2. Bigscreen platform

- Name (and Company developing / offering the platform): Developed by Bigscreen inc.
- Link: https://www.bigscreenvr.com/
- Desktop (non-VR) OS Support: No, it is not possible to play Bigscreen without HMD
- VR Headset Support: Valve Index, HTC Vive, Oculus Rift, Oculus Quest, Oculus Go, Windows Mixed Reality.

- Architecture/game engine: Unity.
- Scripting: No
- Open/closed source: Closed source.
- Availability of SDK: No
- 3D Environments: It is possible watch movies and interact with peoples in fourteen 3D environments and ten 360 environments (static Equirectangular texture):
    - Available 3D environments
        - Lobby, the cinema entrance
        - The home (personal room)
        - Retro cinema which have a 60s design aspect
        - Modern cinema (more futuristic)
        - Grand cinema (looks like a small opera room)
        - Wood
        - Home theater
        - Campfire
        - Luxury theater,
        - Apartment
        - Balcony
        - Bedroom with a screen on the ceiling
        - Kitchen
        - Side room
    - Available 360°environments: pegasus / cassiopeia / andromeda / sagittarius / redshift / sunset / the moon / canyon / mars / the planet
- 6DoF Support: Full 6Dof support and teleport in specific areas and on seats.
- End-User Representation Type (Cartoon Avatars, Human Avatars, 2D Human Reconstruction; 3D Volumetric Human Reconstruction):
    - Human cartoon avatars in 3D (Oculus Avatar SDK)
    - Avatars cannot be imported, but customization is possible: Glasses / mouth / hat / body / skin color / eyes / eyebrows / eye color / hair / hair color / mustache / sidebruns / beard / shirt / clothes color
    - No Realistic Users' Representations
- Other interaction modalities supported:
    - Video Watching: yes.
    - Audio chat: yes, but everyone seems to be in the same room.
    - Furthermore the App integrates "interactive tools" to play with:
        - Draw in 3D with a tilt bush
        - Stream your computer desktop screen in VR
        - Play with some props (popcorn, tomatoes, cameras…)
- Is it possible to select between VR environments?: Yes, several rooms are available.
- Is it possible to import VR environments, or just a default set is available?: Only the ones in the list are available.
- Is it possible to have a private room or environment?: Yes, it is possible to create and edit your personal room:
    - Choose the name
    - Add a description
    - Choose a category in a list
    - Choose if it is a private or public room

- Choose the maximum number of users in the room

Figure 31. Creation of rooms in Bigscreen.

- Supported Media Types:
  - Full 3D VR.
  - Avatars have a 3D look.
  - 2D & 3D videos can be streamed from your Desktop
- Targeted Use Cases:
  - Watch movies together in a multiplayer virtual cinema
  - Create meetings in Social VR chatrooms
  - Multiplayer VR communication

Figure 32. Watching Movies in Bigscreen.

- Exploitation Strategy (Link for Online Service?): Possibility to purchase tickets to attend events or watch 3D movies in VR. Examples:
    - https://www.bigscreenvr.com/movies
    - https://www.bigscreenvr.com/events
- Known Adopters / Clients:
    - Not yet a thorough identification, but it seems that mainly teenagers use the service
    - However, Bigscreen recently made a partnership with Paramount Pictures, MGM, and other major movie studios, so this can change.
- Integration with VR-Together solutions possible?: Unknown a priori. The integration with VR together seems not possible without directly contacting its development and support teams, because Bigscreen is not an open source solution
- Remarkable limitations (ideally addressed in VR-Together)
    - Non-realistic avatars.
    - Setting customization is very poor (e.g. audio controls, playback settings, social settings)
    - Limited target use cases at the moment.
- Technological aspects
    - Delivery Technology (DASH, WebRTC, proprietary...): Not specified
    - Scalability (number of supported users per session): It is possible to have up to 12 users in a Social VR room. The maximum number of parallel sessions is not specified, but potentially it should be around hundreds.
    - Bandwidth consumption (for a pre-defined scenario): around 2-3Mbps (up & down) for a two user scenario with a shared 1080p resolution video
    - CPU/GPU/RAM Usage / Requirements: Bigscreen recommends a minimum of 8 GB of RAM for a very good experience
- User Experience aspects
    - Latency: Around 300ms for audio communication
    - Quality of the VR environment: VR environment of a good quality, especially in terms of lighting.
    - Quality of end-users' representations: Cartoon avatars, but with a wide range of customisation options.
    - Quality of Auditory Interaction:
        - Good. Spatial audio.
        - However, while watching a movie the communication is impossible because of the noise of the movie (Audio controls are missing).
    - Quality of Visual Interaction:
        - Avatar mouth is animated following user speech, which gives some feedback and quality to the interaction.
        - However, avatars are not that much expressive (cartoon style limitation)
    - Quality of 6DoF (Quality of Navigation): Ok. But it’s not easy to understand which parts of the rooms are accessible or not by teleportation
    - Smooth playout: Good
    - Ease of setup: Hardware and configuration is quite complex for a beginner in VR experiences
    - Usability: Not very difficult, but some functionalities and inputs are not very intuitive.

5.1.3.3. Mozilla Hubs platform

- Name (and Company developing / offering the platform): Mozilla (stylized as *moz://a*), which is a free software community founded in 1998 by members of Netscape. Mozilla VR is the Mozilla team focused on bringing VR tools, specifications and standards to the open Web, and maintains e.g. A-Frame. In 2018, the Mozilla VR released Hubs, a VR chat room designed for every headset and browser, but also an open source project that explores how communication in virtual/mixed reality can be accomplished by using web browsers (WebVR, or WebXR)
- Link: https://hubs.mozilla.com/
- Desktop (non-VR) OS Support: Yes. Windows, MacOS, Linux (Web Based)
- Mobile Support: iOS and Android
- VR Headset Support: Any VR headset that supports WebVR/WebXR, including Oculus Rift, Oculus Go, HTC Vive, Valve Index, Windows MR, Gear VR, Google Daydream.
- Architecture/game engine: Three.js plus A-frame (Ammo.js for physics).
- Scripting: No, at the moment, but planned for future releases
- Open/closed source: Open source.
- Availability of SDK: o SDK available. Hubs is a web-based platform
- 3D Environments: Yes, combining 3D and 2D content. It is possible to import glTF scenes.
- 6DoF Support: Yes, but limited.
    - Teleport is supported using both HMD and desktop-based screens
- End-User Representation Type (Cartoon Avatars, Human Avatars, 2D Human Reconstruction; 3D Volumetric Human Reconstruction):
    - 3D Cartoon-like avatars, with customization capabilities. Avatars can be easily modified and customized by users on the Hubs website or with 3D modeling tools. Two examples of such customization tools are listed as follows:
        - Wolf3D: https://readyplayer.me
        - IEEE VR 2020 conference: https://rhiannanberry.github.io/Avatar-Customizer/
    - Live windowed 2D videos from the webcam

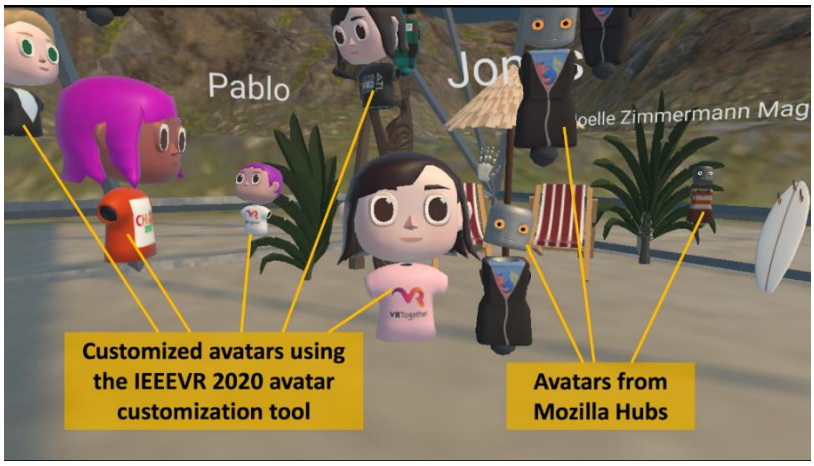


Figure 33. Avatars in Mozilla Hubs.

- Other interaction modalities supported:
    - Video Watching: yes (even live videos from Youtube and Twitch)
    - Audio chat: yes.
    - Text chat: yes (Integration of Discord is possible)

    - Draw in 3D
    - Share content from websites
    - Selfie tool
    - Interaction with the VR environment: users can click on the icons / pics to turn pages, can fly, can swim in the water, and can add, turn, and pin 3D objects to the virtual environment. Users can change the scene of the VR environment. Users can leave a room and enter another room in the VR environment. Users can enter separate rooms (with separate audio from the main room).
- Is it possible to select between VR environments?: Yes, several rooms are available.
- Is it possible to import VR environments, or just a default set is available?: Yes, it is possible to import new VR environments (glTF). Even, since 2018, Mozilla also provides the Spoke editor for the creation of new 3D and 2D spaces. Spoke lets users quickly and easily take 3D content from websites like Sketchfab and Google Poly, and compose it into a custom scene. Users can also use their own 3D models, exported as glTF. In addition, in October 2019, Mozilla released *Spoke: the Architecture Kit*, which enables users to create more realistic 3D scenes without using external tools.
- Is it possible to have a private room or environment?: Yes, it is possible to create and edit your personal room.
- Supported Media Types:
    - Full 3D VR.
    - Avatars have a 3D look.
    - Integration of live 2D from the webcam
    - 2D pictures
    - Stored and live 2D videos
- Targeted Use Cases:
    - Watch movies together
    - Social VR chatrooms
    - Online Events (e.g. IEEE VR 2020, ACM IMX 2020…)
- Exploitation Strategy (Link for Online Service?): Open-Source Solution and Community; Donation; Consultancy, customization and deployment (Hubs Cloud) can be provided on demand.
- Known Adopters / Clients:
    - Open-Source Development Community;
    - Many scientific and academic conferences (e.g. IEEE VR, ACM IMX…)
- Integration with VR-Together solutions possible?: Hubs is an open-source project, but web-based. Thus, potential integration could happen with the VR-Together web platform.
- Remarkable limitations (ideally addressed in VR-Together)
    - Non-realistic users' representations.
    - Scalability is limited to 15-20 people, and require a good server.
    - Transition between spaces is not optimal (it requires to open a new URL).
- Technological aspects
    - Delivery Technology (DASH, WebRTC, proprietary...): WebRTC, A-Frame
    - Scalability (number of supported users per session): Around 20 users per room.
    - Bandwidth consumption (for a pre-defined scenario): -
    - CPU/GPU/RAM Usage / Requirements: -
- User Experience aspects
    - Latency: Delay levels within the boundaries for a good communication

- Quality of the VR environment: Some of the available are poor, but they can be edited and new ones can be created.
- Quality of end-users' representations: Cartoon-like avatars, but with customisation options.
- Quality of Auditory Interaction:
  - Good. Spatial audio (that can be enabled and disabled dynamically).
- Quality of Visual Interaction:
  - Avatars are not that much expressive (cartoon style limitation).
  - Communication via 2D windowed videos from the webcam is possible.
- Quality of 6DoF (Quality of Navigation): Good.
- Smooth playout: Satisfactory. No major problems encountered in stable sessions.
- Ease of setup: Easy
- Usability: Good.

5.1.3.4. NeosVR platform

- Name (and Company developing / offering the platform): Solirax Ltd. NeosVR has a long story, but it was released in 2018 as a closed beta launch, looking for supporters on Steam and Patreon.
- Link: https://neos.com/
- Desktop (non-VR) OS Support: Yes. Windows, Linux
- Mobile Support: Android
- VR Headset Support: It supports all of the major headsets, including SteamVR/OpenVR, Oculus, and Windows Mixed Reality (via SteamVR).
- Architecture/game engine: Most of Neos is running on the FrooxEngine, a custom-made engine crafted over the span of numerous years since around 2015 and around five hundred thousand lines of code, not including third party libraries. Unity is used primarily for its renderer, the runtime environment (Mono/.NET), and interfacing with the audio system.
- Scripting: Yes, but in progress.
  - Logix is the node-based visual scripting language of Neos.
  - Future planned features for NeosVR LogiX Scripting Tool include support for C#, JavaScript and LUA.
- Open/closed source: Closed source, but open source envisioned.
- Availability of SDK: Yes. It is available to Patrons of the Gunter tier, and is presently in early access, and mostly useful for batch importing pre-existing Unity assets. A proper and more fully-featured SDK is planned to be created at a later date.
- 3D Environments: Yes. New 3D environments can be created and important, and customization is possible. LogiX Tooltip, a scripting tool to assign specific attributes and form interactions between game objects, is available.

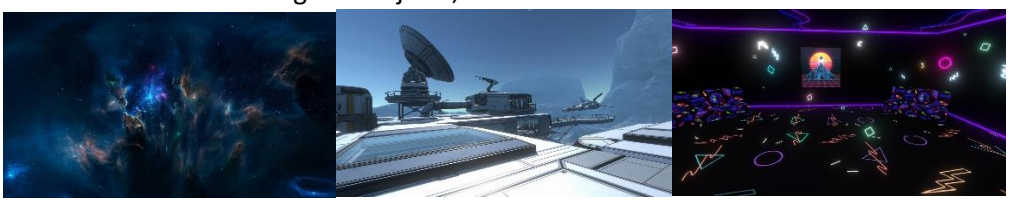


Figure 34. NeosVR in-game 3D world screenshots using the camera tool.

- 6DoF Support: Yes.

- NeosVR also provides interactive dynamic bones with collisions and grabbing, as well as the fully configurable full body of 11 tracking points (head, hands, hips, feet, chest, elbows and knees).

- End-User Representation Type (Cartoon Avatars, Human Avatars, 2D Human Reconstruction; 3D Volumetric Human Reconstruction): Customisable cartoon-like avatars (Figure 35) that can be imported as game assets with different formats (fbx, obj, blend).

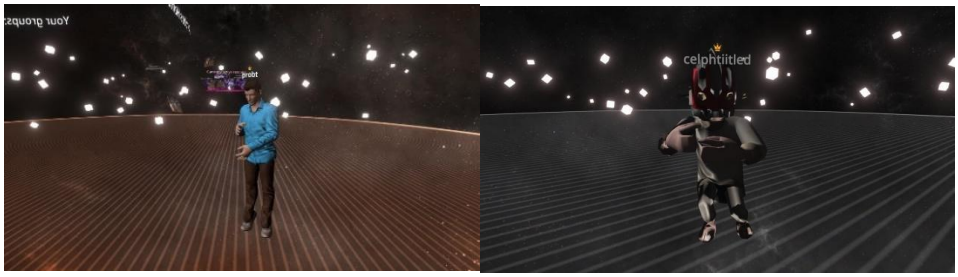


Figure 35. NeosVR in-game avatars.

- Other interaction modalities supported:
  - Video Watching: yes (even live videos from Twitch)
  - Audio chat: yes.
  - Text chat: yes (Twitch chat integration).
  - Screen Share is planned as a future feature
- Is it possible to select between VR environments?: Yes, several rooms are available and can be imported.
- Is it possible to import VR environments, or just a default set is available?: NeosVR allows the manipulation (importing/exporting) of a variety of assets, such as 3D models, audio streams, point clouds, images, videos and texts.
- Is it possible to have a private room or environment?: Yes, it is possible to create and edit your personal room.
- Supported Media Types:
  - Full 3D VR.
  - Avatars have a 3D look.
  - Integration of live and stored 2D video
- Targeted Use Cases:
  - Educational institutes
  - Entertainment industry (movies, games…)
  - Streaming services and social VR interaction (e.g. integration with Twitch)
- Exploitation Strategy (Link for Online Service?): Patrons, Research Projects / Funds (EU Regional Development Fund), Projects with Industry, Funds by Ventures and Accelerator strategies.

  NeosVR positions itself as a versatile and feature rich metaverse for VR, designed so everybody can find his way in a social setting.
  - Entered the gaming market by launching DelVR, an immersive system to play games like Dungeons and Dragons within VR, in January 2020
  - Creation of the ambitious MetaMovie “Alien Rescue” using the NeosVR platform in November 2019.
- Known Adopters / Clients: See above.
- Integration with VR-Together solutions possible?: At the moment, NeosVR is closed source. In addition, since the current state of scripting support is limited to LogiX Tooltip, the

rendering of real-time volumetric data is not possible. Planned features for NeosVR LogiX Scripting Tool include support for C#, JavaScript and LUA. That may open the door for future possible integrations.

- Remarkable limitations (ideally addressed in VR-Together)
  - Non-realistic users' representations.
  - Not a so smooth and stable performance yet, with noticeable latency.
  - Slow progress.
  - Missing relevant features for Social VR, but mostly envisioned as VR world.
- Technological aspects
  - Delivery Technology (DASH, WebRTC, proprietary...): Unknown (FrooxEngine)
  - Scalability (number of supported users per session): Unknown (but platform envisioned for virtual worlds, so it is expected that >20 users).
  - Bandwidth consumption (for a pre-defined scenario): -
  - CPU/GPU/RAM Usage / Requirements: See obtained results in Table 26, by using a PC with the resources outlined in Table 27.

Resources consumption	
CPU %	36.3
GPU %	13.1
RAM (MBs)	1325.6

Table 26. Resources consumption when using NeosVR.

PC's specifications	
Processor	Intel® Core™ i5-8600K CPU @ 3.6GHz
GPU	NVIDIA GeForce GTX 1070
RAM	16 GB

Table 27. Specifications of the computer used for the evaluation of NeosVR.

- User Experience aspects: For NeosVR, a subjective study was conducted with the participation of 16 users (8 with previous experience in VR, 8 without previous experience in VR), testing the platform during 15min. Before the start of a session, the users observed an exhibition on how to install and use NeosVR with Oculus Rift. An experiment facilitator guided the participants during the test, with the VR equipment setup (Oculus Rift), launching of the platform, and also for an appropriate usage of its functionalities. At the end of the experiment, the participants filled in a questionnaire to assess a range of perceptual metrics in a scale from 1 to 5 (5 being the best grade). The results are in Table 28, averaged for all, but also clustered for the VR experts and beginners. Overall, the users were quite satisfied with regard all considered metrics and aspects, but the perceived latency, the quality of end-users' representations of and visual interaction were less positive.

Users	Latency	Quality of virtual environment	Quality of end-user 's representation	Quality of auditory interaction	Quality of visual interaction	Quality of 6DOF (navigation)	Smooth playout	Ease of setup	Usability
All Users	3.6	4.375	3.5	4	3.5	4.2	4	5	4.75
Prior VR Experience	3.5	4.25	3	4	3.2	4	4	5	4.5
No prior VR experience	3.7	4.5	4	4	3.8	4.4	4	5	5

Table 28. Average Scores for the considered metrics and aspects.

5.1.3.5. Spatial.io platform

- Name (and Company developing / offering the platform): Spatial Systems, Inc.
    - Spatial was founded in 2016 from investors including iNovia Capital, White Star Capital, Expa (founded by Garrett Camp), Kakao Ventures, Lerer Hippeau, Leaders Fund, Samsung NEXT as well as angels including Mark Pincus (Founder of Zynga), Andy Hertzfeld (Co-Inventor of the Macintosh) and Mike Krieger (Co-Founder of Instagram).
- Link: https://spatial.io/
- Desktop (non-VR) OS Support: Yes. Web Version available
- Mobile Support: Yes. iOS and Android.
- VR Headset Support: Oculus Quest.
    - Also AR/MR headsets: Hololens, Magip Leap, nreal.
- Architecture/game engine: Unity.
- Scripting: Unknown
- Open/closed source: Closed source.
- Availability of SDK: Unknown
- 3D Environments: It is possible to upload 3D models (*.fbx, .gltf, .glb*)
- 6DoF Support: Yes.
- End-User Representation Type (Cartoon Avatars, Human Avatars, 2D Human Reconstruction; 3D Volumetric Human Reconstruction):
    - Human-like avatars in 3D, from a selfie
    - Without headsets, users can also join through the webcam

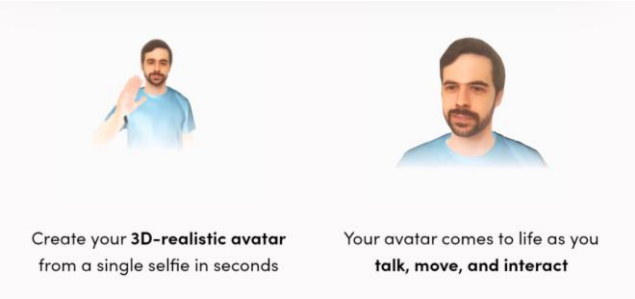


Figure 36. Avatars in Spatial.

- Other interaction modalities supported:
    - Video Watching: yes.
    - Audio chat: yes.
    - Whiteboard
    - Board for sharing notes and organising ideas
    - Integrations with external tools, e.g. Drive, Slack, etc.
    - Screen share any app and window from your computer
- Is it possible to select between VR environments?: Yes, 3D models can be uploaded.
- Is it possible to import VR environments, or just a default set is available?: Yes, 3D models can be uploaded.
- Is it possible to have a private room or environment?: Yes, it is possible to create and edit your personal room: Yes
- Supported Media Types:
    - Full 3D VR/AR/XR.
    - Avatars have a 3D look.

    - 2D & 3D videos
    - Pictures
    - Boards
    - Screen sharing
- Targeted Use Cases:
    - Collaborative meetings
    - Co-work
- Exploitation Strategy (Link for Online Service?): Free service, with Pro and Enterprise version with premium features
    - Pricing: https://spatial.io/pricing/
- Known Adopters / Clients:
    - Mattel
    - BNP Paribas
    - Ford
    - Enel
    - Etc.
- Integration with VR-Together solutions possible?: Unknown a priori. The integration with VR-Together seems not possible without directly contacting its development and support teams.
- Remarkable limitations (ideally addressed in VR-Together)
    - Non-realistic avatars.
- Technological aspects
    - Delivery Technology (DASH, WebRTC, proprietary...): Not specified
    - Scalability (number of supported users per session):
        - Host up to 30 people in a room, using the free version
        - Host up to an additional 20 people from the webapp, using the free version.
        - It is possible to increase the scalability in Enterprise versions
    - Bandwidth consumption: -
    - CPU/GPU/RAM Usage / Requirements: -
- User Experience aspects
    - Latency: Acceptable for a good interaction.
    - Quality of the VR environment: Good for the envisioned scenarios.
    - Quality of end-users' representations:
        - Although being avatars, the quality is acceptable
    - Quality of Auditory Interaction:
        - Good.
    - Quality of Visual Interaction:
        - Although being avatars, the quality of visual interaction is acceptable.
        - However, avatars are not that much expressive
    - Quality of 6DoF (Quality of Navigation): Good for the tested scenario.
    - Smooth playout: Good for the tested scenario.
    - Ease of setup: Easy, no complications were found.
    - Usability: Usability is good.

5.1.3.6. Virbela platform

- Name (and Company developing / offering the platform): eXp World Technologies, LLC d/b/a Virbela.
- Link: https://www.virbela.com/
- Desktop (non-VR) OS Support: Windows and Mac
- VR Headset Support: Support since 2020. Still very recent, but HTC VIVE and Oculus are supported (https://www.virbela.com/blog/virtual-reality-comes-to-virbela).
- Mobile Support: mobile application (iOS/iPadOS only) available, called VirBELA Intercom, but provides only voice input to the virtual world.
- Architecture/game engine: Unity.
- Scripting: Yes
- Open/closed source: Closed source.
- Availability of SDK: No
- 3D Environments:
    - The VirBELA Open Campus is a graphical 3D world, with an open space, buildings, rooms, a beach, a soccer field and some fun elements like a lighthouse and speed boats.
    - For large enterprise use or conferences a dedicated world can be created by VirBELA
- 6DoF Support: Yes.
    - Users (avatars) can move quite naturally: walk, run, stand, pose, clap, raise hand, shake hands and even dance.
    - Shortcut keys (F1-F10) or commands control the avatar.
- End-User Representation Type (Cartoon Avatars, Human Avatars, 2D Human Reconstruction; 3D Volumetric Human Reconstruction):
    - Live-like (human-like) 3D characters, with head and limbs
    - Avatars are customizable, in the “VirBELA Avatar Creation System”. An avatar needs to be created on first use, but can be changed at any moment
    - No custom avatar can be uploaded. Therefore, the avatars often tend to resemble, making it difficult to recognize each-other.
    - The name displayed above the avatar, helping to recognise users.

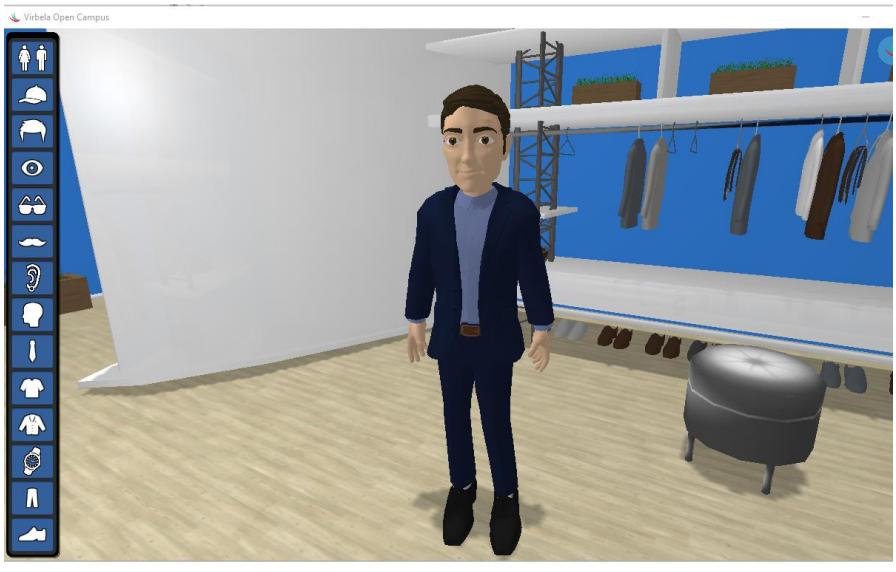


Figure 37. Avatar creation and adaptation in VirBELA.

Figure 38. Natural movements of avatars, and displayed name, in Virbela.

- Other interaction modalities supported:
  - Video Watching: yes (e.g. Youtube videos).
  - Audio chat: yes. Voice Over IP (based on Teamspeak).
  - Directional voice, spatial audio
  - Text Chat
  - Collaborative Web Browsing
  - Google Slides Presentations
  - Google Docs
  - Whiteboards
  - Virtual Laser Pointer
  - Record/Replay System
- Is it possible to select between VR environments?: Virbela provides a virtual world (Campus), and users can freely navigate therein.
- Is it possible to import VR environments, or just a default set is available?: Only the default campus is available. Other VR environments can be provided on demand.
- Is it possible to have a private room or environment?: You need to request it on demand.
- Supported Media Types:
  - Full 3D VR.
  - Avatars have a 3D look.
  - 2D and presentations
  - Web Browsing
  - Laser Pointing
  - Text Chat
- Targeted Use Cases:
  - Universities: this was the initial target market addressed by the founders with MBA students as targeted users. The students can roam around on the campus, meet each-other and follow courses in the auditoria and smaller rooms. Stanford, The Rady School of Management and Davenport University are using VirBELA.
  - Enterprise: more focus is now on corporate users since eXp Realty bought VirBELA. eXp World is an ideal showcase for multinational companies. The way a billion dollar company can run its entire business only in a virtual world is probably unique and shows the capabilities of the VirBELA platform. In 2018 eXp Realty was named Best Place to Work by Glassdoor. Other examples of enterprise customers are: The Honor Foundation and the US Army.

- Event organisers: VirBELA has been hosting virtual events, conferences and exhibitions, like Laval Virtual 2020 and VR Days 2020.

- Exploitation Strategy (Link for Online Service?): VirBELA positions itself as a 3D virtual world, not as VR conferencing. The difference being that the virtual world is a persistent environment, comparable to Second Life: actions in the 3D world remain even when a user leaves. VR is only recently an added capability.

  The benefits of virtual worlds compared to video conferencing, as seen by Virbela, are discussed here.
  VirBELA positions itself as the full service provider of the solution including technical support, user analytics and customisation.
- Known Adopters / Clients:
  - See: https://www.virbela.com/customers Examples are:
    - PwC
    - Many online events: Laval Virtual, VR Days, etc.
- Integration with VR-Together solutions possible?: Closed source solution, but using Unity as the engine, so it may be possible.
- Remarkable limitations (ideally addressed in VR-Together)
  - Avatars cannot be imported.
  - You cannot choose between different VR environments and import new ones, without contacting the company.
  - Technical issues occurred in previous events (e.g. Laval Virtual)
- Technological aspects
  - Delivery Technology (DASH, WebRTC, proprietary...): Unity and TeamSpeak for VoIP. Other details unspecified.
  - Scalability (number of supported users per session): Around hundreds, and even thousand. For instance, 6600 attendees to Laval Virtual 2020 event.
  - Bandwidth consumption and CPU/GPU/RAM Usage / Requirements:
    - Virbela provides a System Metrics Panel (https://virbela.zendesk.com/hc/en-us/articles/360036324331-System-Metrics-Panel) that can inform about the magnitudes of delays, packet loss, bandwidth, and CPU and memory usage.
    - VirBELA features a lot of settings to find the right balance between quality and resource consumption. Performance measures are therefore just an indication of what the system requires.

    Test using a PC with the next resources: MSI, NVIDIA GTX1070, 16GB RAM, Intel i7.
    - Scenario 1: Acceptable lowest quality, desktop based. Settings (chosen from VirBELA):
      - Resolution: 1280x720
      - 3D avatar
      - Fastest quality

      Results: CPU: 9.8%; GPU: 16.6%; RAM: 585 MB
    - Scenario 2: Highest quality desktop:
      - Resolution: 1680x1050
      - 3D avatar
      - Best quality

      Results: CPU: 15.7%; GPU: 40.2%; RAM: 569 MB

        - Scenario 3: Highest quality HMD:
            - native VR +1920x1080
            - 3D avatar
            - Best quality

        Results: CPU: 24.8%; GPU: 29%; RAM: 636 MB
- User Experience aspects
    - Latency: Acceptable latency for satisfactory interaction
    - Quality of the VR environment: Good. They are specialized in replicating conference-like or events-like environments.
    - Quality of end-users' representations: Customisable human-like avatars, but avatars cannot be imported.
    - Quality of Auditory Interaction:
        - Good. Spatial audio.
    - Quality of Visual Interaction:
        - Quite natural gestures of avatars.
    - Quality of 6DoF (Quality of Navigation): Ok
    - Smooth playout: Good, but some technical issues happened in previous events.
    - Ease of setup:
        - It requires to install a (heavy) application, which is available for Windows and Mac. Often, when starting the application updates and patches are installed, which can take a lot of time (this cannot be aborted).
    - Usability:
        - Good, but some functionalities and inputs are not so intuitive.

5.1.3.7. ViveSync platform

- Name (and Company developing / offering the platform): Vive Sync, developed by HTC.
- Link: https://sync.vive.com/
- Desktop (non-VR) OS Support: Windows, including a viewer mode
- VR Headset Support: VIVE, VIVE pro, VIVE pro Eye, VIVE Cosmos ,VIVE Cosmos Elite.
- Mobile Support:
    - VIVE Sync supports VIVE untethered mobile headsets: VIVE Focus, VIVE Focus Plus
    - And Android and iOS viewer mode apps are listed as “Coming Soon”, likely to provide similar functionality as the “viewer mode” of the default Windows application.
- Architecture/game engine: While not explicitly stated, the installed application clearly has the structure and content of a Unity3D build.

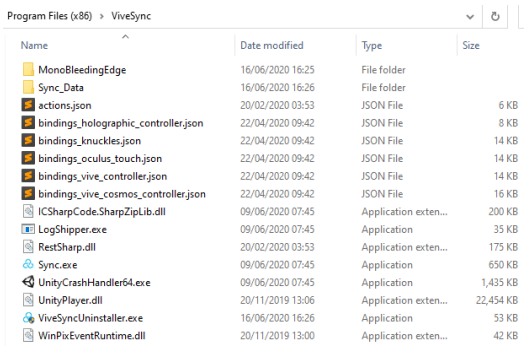


Figure 39. File Structure of the Vive Sync installation.

- Scripting: No
- Open/closed source: Closed source.
- Availability of SDK: No
- 3D Environments: When scheduling a meeting there is a choice of 3 predefined environments. A SciFi room, an outdoor platform with a view overlooking water and a room with a platform in the clouds

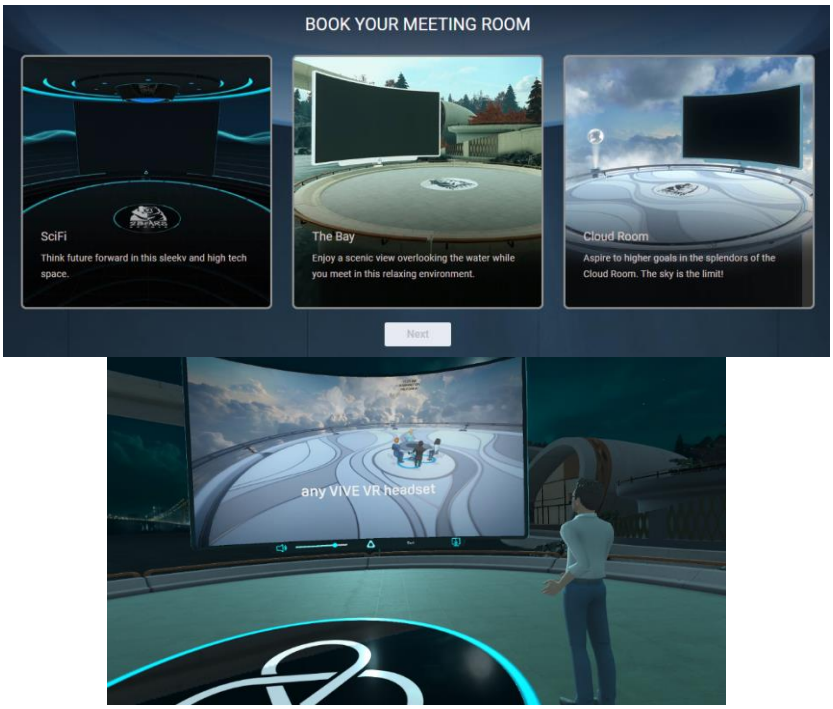


Figure 40. 3D Environments in Vive Sync.

- 6DoF Support: Yes.
    - There is full 6DoF support. Users can walk around freely within the bounds of their local room, but there is also the functionality to teleport freely, or to teleport to a specific seat when tables or an auditorium are added to a meeting (this can be done at any time during the meeting).
- End-User Representation Type (Cartoon Avatars, Human Avatars, 2D Human Reconstruction; 3D Volumetric Human Reconstruction):
    - Live-like (human-like) 3D characters
    - A set of predefined default avatars is available
    - Avatars are customizable via the Sync XR Avatar Creator. This is an app available for Android or iOS. Based on a selfie photo an avatar will be created which you can then further customize. You go through the following steps:
        - Select a body (male/female)
        - Take a selfie (Without glasses, even lighting, no hair in face, mouth closed)
        - Select hair, hair color, eye color and whether or not you wear glasses
        - This results in an avatar ID which you can then use inside the application
    - Although based on a selfie, the avatars are not necessarily realistic. In some cases the facial recognition seems inaccurate, cause for example poor placement of mouth textures.

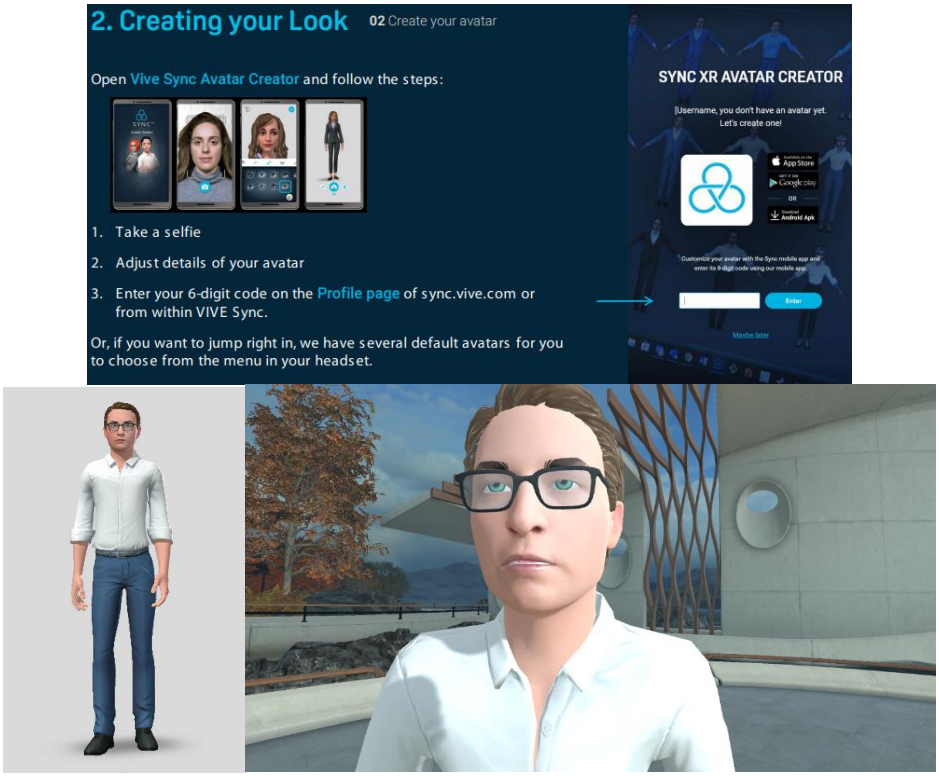


Figure 41. Avatar creation and customization in Vive Sync.

- Other interaction modalities supported:
    - VIVE Sync supports a number of media formats for sharing in the environment on a large screen. See the section on Supported Media Types.
    - Audio chat: yes. Even for non-VR viewers who can join in a viewer mode.
    - Take both public and private notes.
    - Virtual laser point to point at anything in the environment.
    - Virtual pencil to draw freely in the environment's space, in full 3D.
    - Sharing files through a cloud service such as OneDrive, or by sharing a local folder, up to a maximum of 2GB per meeting room
    - Any data captured during these meetings (photos, speech-to-text recordings, etc.) are available for download as well
- Is it possible to select between VR environments?: Yes, VIVE Sync comes with 3 predefined environments which can be selected when planning a meeting.
- Is it possible to import VR environments, or just a default set is available?: Just a default set is available.
- Is it possible to have a private room or environment?: Rooms have a specific URL available after a meeting has been planned. These seem to be private as long as the details are not made public.
- Supported Media Types: VIVE Sync supports a number of media formats that allow media sharing through a big screen:
    - PDF
    - PowerPoint
    - Video (MP4, AVI, MOV)
    - Images (PNG, JPEG, BMP, TGA)
    - 3D Models (FBX, OBJ, Unity Asset Bundles)

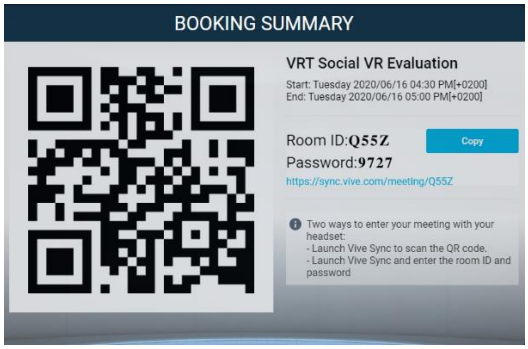


Figure 42. Room info in Vive Sync.

Figure 43. Media Sharing in Vive Sy.

- Targeted Use Cases: Per the official VIVE Sync blurb:

"VIVE Sync is the all-in-one meeting and collaboration solution for VR. With VIVE Sync, it's easy to customize your avatar, create a private meeting room, and begin working face-to-face with colleagues around the world. And with our suite of 3D interactive meeting tools, you can review 3D interactive content in ways that have never been possible."

- Exploitation Strategy (Link for Online Service?): For 2020 the service has been made available in beta for free. The application is part of Vive's newly announced XR Suite of applications, an upcoming subscription package in the style of Microsoft Office with a free version available for those users only requiring basic features. The launch of the service is expected in Q3 2020 for China with other regions following later.
- Known Adopters / Clients: Unknown
- Integration with VR-Together solutions possible?: The solution is closed source and does not provide an SDK. Immediate integration without direct support from HTC seems impossible.
- Remarkable limitations (ideally addressed in VR-Together)
    - Unrealistic users' representations
- Technological aspects
    - Delivery Technology (DASH, WebRTC, proprietary...): Unclear. Likely proprietary. Servers are currently hosted in 3 regions: US, Asia and China, available for selection when creating a meeting
    - Scalability (number of supported users per session): Up to 30 users per session are currently supported
    - Latency, bandwidth consumption and CPU/GPU/RAM Usage / Requirements:

        - Speech and movement latency on a US-server seemed to range between 0.5 and 1 second.
            - This was measured by users communicating in adjacent rooms in the office, counting down to the start of a gesture
        - For a 2-user scenario the bandwidth averaged around 0.3Mbps, with peaks at about 0.5Mbps. This is likely to increase with the number of participants increasing.
        - The Windows app does allow for Low/Medium/High quality settings. On the High quality setting:
            - The amount of RAM used in the 2-user scenario was around 1.2GB
            - The CPU seemed to occupy a single core fully, as is usual for Unity-based applications
            - GPU Memory usage peaked at around 5GB
- User Experience aspects
    - Latency:
        - Acceptable. As said between 0.5 and 1.0 seconds on a US server for 2 users. It never hindered communication or interaction
    - Quality of the VR environment:
        - Good, the environments looked polished and pleasant to be in
    - Quality of end-users' representations:
        - Good. The avatar generation based on a selfie has some rough edges, but overall the avatar quality looked clean. Animation left a bit to be desired. Driven by controller positions as well as head height meant that animation wasn't always the best. The head height is used to determine if a user is standing or seated, and on occasion the system gets this wrong.
    - Quality of Auditory Interaction:
        - Good for two users. Spatialization worked but wasn't excellent. It remains to be seen how a system like this works with an auditorium full of 30 users.
    - Quality of Visual Interaction:
        - Very good. All actions of other users (presenting content, taking notes, drawing in mid-air, etc.) were all perfectly visible and usable.
    - Quality of 6DoF (Quality of Navigation):
        - Very good. Users can freely move around or instantly teleport to locations in the environment. When tables and seats are used, the seats are visualized as potential teleport locations.
    - Smooth playout:
        - No real disturbing hiccups were experienced. Audio was smooth, framerate was constant. The animation of other users left to be desired on occasion, but nothing which hindered the meeting.
    - Ease of setup:
        - Very simple to set up. No particularly complex interactions are required. The avatar generation app is also very easy to use.
    - Usability:
        - Very usable. The menu is cleanly laid out. Instructions are available, but the interactions are very intuitive. In general, they all work as you expect they would.

5.1.4 Conclusions

This section has shown that there are many available Social VR platforms in the market. This is a clear proof of the high interest Social VR is awakening. It was been also shown that these platforms typically perform well, and their usability is satisfactory. This is also a proof of market readiness.

Almost every platform supports 3D environments, provides support for desktop and VR modes, allows for media sharing, and allows to live stream VR sessions to other 2D platforms, like Youtube and Twitch. The VR-Together platform also provides these widely supported features. Interestingly, it has been shown that all of platforms rely on the use of (either cartoon-like or human-like, even customisable) 3D avatars for the end-users' representations, and few of them (e.g. Mozilla Hubs and Spatial.io) also support the integration of live 2D windowed videos from the webcam. The VR-Together platform not only also supports these features, but it is an outstanding platform with regard to the support of realistic volumetric end-users' representations, both as Time Varying Meshes (TVM) and Point Clouds. It also supports the integration of users with just audio communication (no visual representation) and as pure spectators (no audio, no video, as ghosts).

In addition, the VR-Together platform supports the integration of live broadcasted streams, with background removal (Chroma keying), without having to integrate external platforms like Youtube and Twitch, which provides higher control and much lower delays. Moreover, the VR-Together platform enables the integration of live stereoscopic 180º/360º feeds, which is not supported in existing Social VR platforms. The table below summarizes the conducted analysis and comparison, by paying attention to key aspects related to supported platforms and content formats.

In essence, the analysis not only served us to better known the available solutions, but also to confirm and prove the outstanding features provided by the VR-Together platform.

Platform / Features	**HMD / Desktop Support**	**Users' Representation**	**3D Environ -ment**	**Integration of Live Broadcasted Video**	**Chroma Keying for Live Videos**	**Live 180º / 360º video**	**Is Possible to Broadcast Live Sessions?**
AltspaceVR	Y / Y	Human-like Avatars (customizable clothes, but no faces)	Y	Partially (Youtube integration)	N	N	Y (e.g. Twitch)
BigScreen	Y / N	Cartoon-like avatars (customizable)	Y	Yes (but integrating third-party platforms and TV channels)	N	N (Only static 360º scenes for the environment)	Y (e.g. Twitch)
Mozilla Hubs	Y / Y	Cartoon-like Avatars (customizable) and live 2D video from webcam	Y	Partially (Youtube and Twitch integration)	N	N (Only static 360º scenes for the environment)	Y (e.g. Twitch)
Spatial.io	Y / Y	Human-like Avatars and 2D videos from webcam	Y	Partially (integration of video and screenshare)	N	N	-
NeosVR	Y / Y	Cartoon-like Avatars (customizable)	Y	Partially (Twitch integration)	N	N	Y (e.g. Twitch)
Virbela	Y / Y	Human-like Avatars (customizable)	Y	Partially (Youtube integration)	N	N	Y (e.g. Twitch, Youtube)
Vive Sync	Y / Y	Human-like Avatars (customizable)	Y	-	N	-	Y (e.g. Youtube)
VR-Together	Y/ Y	Realistic Volumetric Representation, 3D Avatars (no customization at the moment), live 2D video from webcam, just audio, or no audio and video but just presence (ghost)	Y	Y+ (Own live broadcasting pipeline)	Y	Y	Y (Youtube)

Table 29. Summary of comparison between state-of-the-art Social VR platforms.

5.2 Evaluating the Users and the Environment

5.2.1 CERTH-3.2: Visual Comparison of TVMs

5.2.1.1 Objective

The aim of these experiments is to objectively compare the visual quality between TVM instances produced by two different sets of RGB-D sensors, Kinect 4 Azure (TVM v3) and Intel RealSense D415 (TVM v2) in terms of fidelity with the reconstructed subject. This experiment constitutes a benchmark between the different types of off-the-shelf RGBD sensors supported by our VRTogether 3D Capture, allowing for the selection of the best performing combination of sensors for multi-view 3D-capturing and reconstruction and its evolution during the project period (new version per year).

5.2.1.2 Methodology

Using Volumetric Capture (VolCap - https://github.com/VCL3D/VolumetricCapture, CERTH's in-lab developed multi-RGB-D sensor 3D capturing, streaming and recording application), the collected RGB and depth data from a set-up of sensors are gathered in a centralized processing where the different views are merged to a 3D reconstruction and rendered to an RGB image of the desired view in real-time.

Figure 44 Eight (8) RGB-D sensors spatio-temporally aligned to capture various performances.

5.2.1.3 Setup

For this experiment, two types of cutting-edge consumer grade RGB-D sensors (*Kinect4Azure* and *Intel RealSense D415*) are placed in the setup depicted in Figure 44. The 4-sensor cross-like setup of each sensor type is the typical multi-view setup used in CERTH's multi-view reconstruction system. Both sensor type setups simultaneously captured the subjects from different views as the viewpoints of each sensor type setup are tilted 45◦degrees compared with the viewing position of the other sensor type.

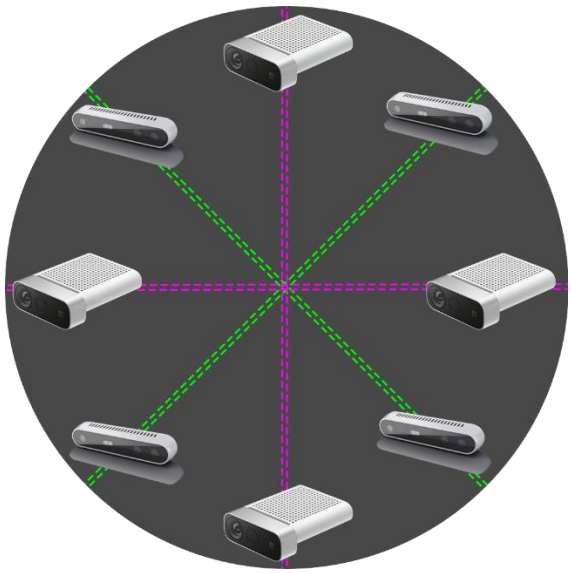


Figure 45 Cross-like setup with two 4-sensor sets.

This results to the star like setup of the whole system (that includes both sensor types) depicted in Figure 45. We placed the two distinctive setups with a 45^{o} degrees rotational offset in order to overcome the lack of ground-truth data for the comparison process. The reconstruction result of each sensor type setup was rendered to the views of the other's. Then, the rendered reconstruction image is compared to the actual captured colour image from the other type of sensors. This way the captured frames of one sensor type serve as the "ground-truth" frame for the evaluation of the reconstructed frame of the other sensor type.

We captured short-duration sequences of 3 subjects that performed according to predefined scenarios. The volunteers were 2 male and 1 female actors with different body types (height, mass, hairstyle, clothes, etc.).

After the data collection process, the 3D reconstruction rendered frames of each sensor type were evaluated in comparison with the actual videos gathered by the other sensor type. The fidelity of the rendered frames with the captured frames is evaluated using objective image metrics. The results are aggregated across frames and sequences and performance results are produced.

For the experiments purposes we defined 9 performance categories depending on the subjects posture and movement speed:

1) *Static movement - standing position*
2) *Static movement - sitting position*
3) *Slow movement - standing position*
4) *Slow movement - sitting position*
5) *Fast movement - standing position*
6) *Fast movement - sitting position*

The duration of each performance depends on the category it belongs to. Categories (1) and (2) lasted 3", (3) and (4) 13" duration, while (5) and (6) 8". Additionally, we defined simple performance scenarios to trigger the subjects to move: each subject had to form 3 capital English letters using their limbs and body. The assigned letters were: **A, B, C, E, F, M, P, R, T, Y, X, W**. As an experiment session, we define a subject executing one of the performance categories. In total, the experiment session where 19 (# of subjects x # of performance categories + one static object

capture). In total, we processed 350 group of 8 RGBD views to extract the total PSNR and SSIM between the TVM v3 and v2 TVM rendered images.

5.2.1.4 Metrics

For the evaluation process we used two different metrics, *PSNR* and *SSIM*. Both metrics are widely used to evaluate image quality in an objective manner.

5.2.1.5 Results

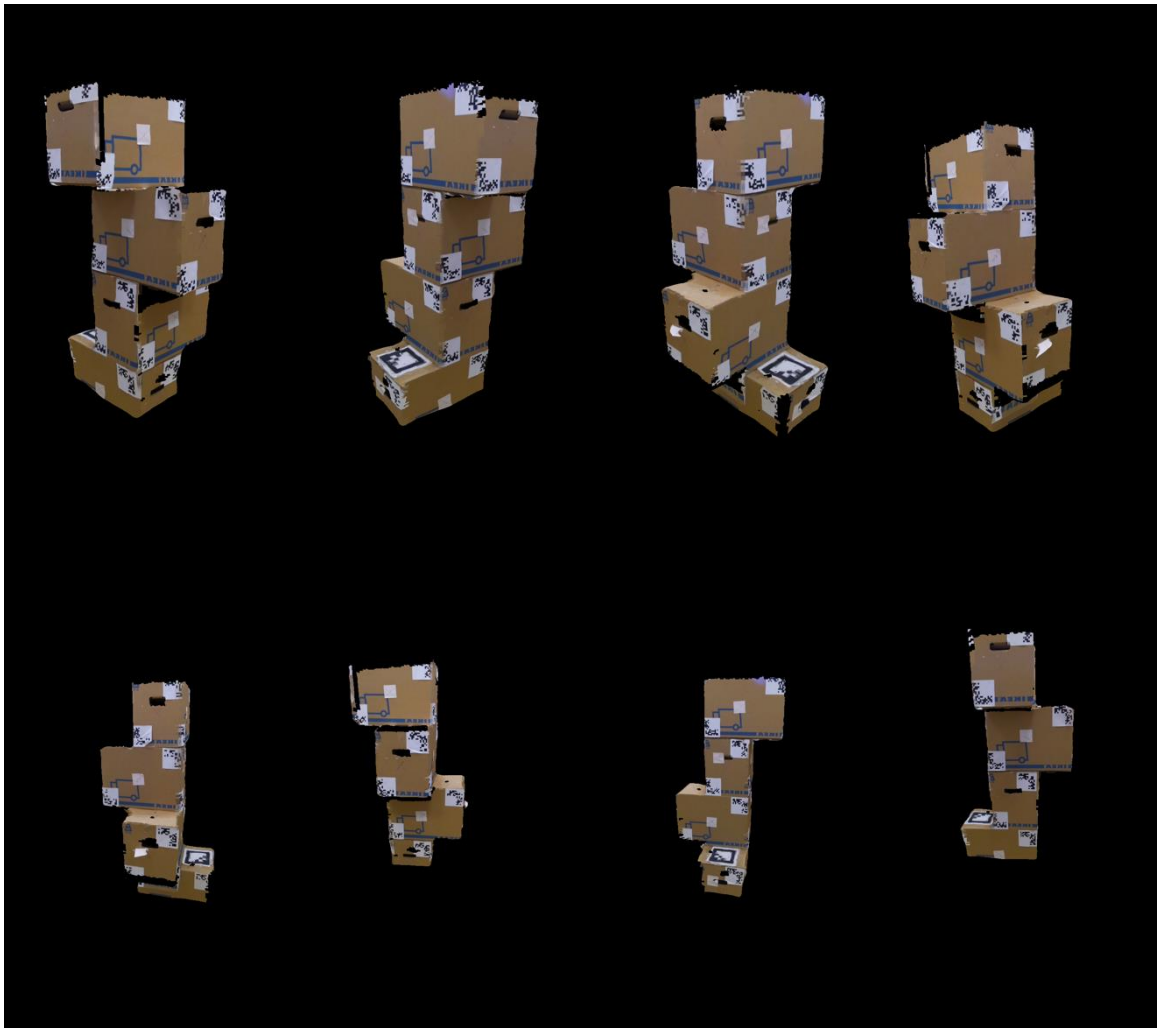


Figure 46 TVM v3 reconstruction from 8 camera viewpoints. Top and bottom rows show the views from 4 Intel RS D415 and 4 MS Kinect4Azure devices, respectively.

TVM Versions	PSNR	SSIM
TVM v2 (4 x RS D415)	23.059	0.918
TVM v3 (4 x MS Kinect4Azure)	**27.333**	**0.963**

Table 30 Performance results of the proposed metric in the cross-validation.

Figure 47 Left: Rendered TVM v3. Right: Real colour view (Ground truth) and 4 MS Kinect4Azure devices, respectively.

5.2.1.6 Conclusion

From this assessment, we objectively prove that the new version of the TVM using Kinect4Azure achieves higher fidelity scores (Table 30). Given the higher quality of the depth streams, the geometry is shaped more appropriately in comparison v2, resulting in better aesthetic quality. All and all, we prove that our TVM capturing system is a software ready to capture point cloud and mesh-based volumetric video, while with the evolution of the low-cost and consumer-grade depth sensors in the market will get better and better quality.

5.2.2 I2CAT-3.1: Accessibility in VR: Subtitling 3D VR Content

5.2.2.1 Objective

Accessibility is a key requirement for every (multimedia) service, but it has been scarcely explored for VR environments. This applies to all existing access services: subtitling, audio description, and sign language interpreting.

This experiment is pioneering in exploring two relevant Research Questions (RQ) with regard to the applicability of subtitling, as the most widespread access service, in 3D VR environments:

- RQ1) what are the most appropriate subtitling presentation modes in 3D VR environments?
- RQ2) do visual indicators, like arrows, provide benefits for guiding the users toward the target speakers in 3D VR environments?

These research questions arise from previous research initiatives and insights on exploring the same aspects for 360º video subtitling [Mon20b] [Mon20c] [Mon21].

5.2.2.2 Methodology

Presentation Modes.

Three subtitling presentation modes have been implemented and evaluated (see figure below):

- *Mode A. Fixed-positioned subtitles*: subtitles are presented at fixed positions and planes in the scene. In 360º videos, this is typically implemented by replicating the subtitles presentation, e.g. equally spaced every 120º. However, this strategy can lead to content

blocking in 3D VR environment, and the depth dimension also plays a key role. Therefore, subtitles have been presented in a position fixed position closed to the speakers.

- *Mode B. Comic-style*: subtitles are presented as bubbles attached to the associated speaker, as in comics.
- *Mode C. Always-visible subtitles*: subtitles follow the user's viewpoint at anytime, regardless of were the user is looking at.

Figure 48 Considered 3D VR subtitling presentation modes: (top-left) Mode A. Fixed-positioned (in front of the mirror); (top-right) Mode B. Comic-style; (bottom-left) Mode C. Always-visible; (bottom-right) indicator.

Indicators.
A 3D arrow can be enabled to guide the user toward the target speaker.

Stimuli.
As stimuli, the full 3D version of the pilot 1 content (see D4.2) was used, including the two interrogation scenes to each one of the suspects. The video was presented to the participants without audio, so the subtitles become key to understand the story, although not having hearing impairments.

This video shows how the considered presentation modes and indicators look like in the pilot 1 content.

Test Conditions 1. Presentation Modes.
The first part of the experiment explored the appropriateness of the considered 3D VR subtitling presentation modes (Modes A, B, C). Given that three presentation modes are considered, the two interrogation scenes were divided into two parts of the same length:

- Clip 1A: First Part of Interrogation to Zeller – suspect 1 (~240s)
- Clip 1B: Final Part of Interrogation to Zeller – suspect 1 (~240s)
- Clip 2A: First Part of Interrogation to Christine – suspect 2 (~240s)
- Clip 2B: Final Part of Interrogation to Christine – suspect 2 (~240s)

In order to avoid order effects, the presentation of test conditions was counterbalanced, always starting by the first part of the interrogation scenes to be able to adequately follow the story:

Participant ID	Clip 1A	Clip 1B	Clip 2A
1-7-13-19	Mode A	Mode B	Mode C
2-8-14-20	Mode C	Mode A	Mode B
3-9-15-21	Mode B	Mode C	Mode A
4-10-16-22	Mode A	Mode C	Mode B
5-11-17-23	Mode B	Mode A	Mode C
6-12-18-24	Mode C	Mode B	Mode A

Table 31 Counterbalancing of test conditions 1 (subtitling presentation modes) to avoid order effects.

Test Conditions 2. Use of Indicators.

The second part of the experiment consisted of adding a visual indicator (arrow) that dynamically points at the target speaker, regardless of the user's position and viewpoint. The indicators were added to the Modes B and C, in a counterbalanced manner. That test condition was not considered for Mode A, as in such a mode it may be possible to have the subtitles and the target speakers in different Field of Views (FoV), and therefore the use of the indicator could cause confusion in such situations.

Participant ID	Clip 2A	Clip 2B
Odd ID (1, 3, 5…)	Mode B	Mode C
Even ID (2, 4, 6…)	Mode C	Mode B

Table 32 Counterbalancing of test conditions 2 (use of indicators) to avoid order effects.

Equipment and Setup:

- Gaming laptop (MSI, i7-10750H, 16GB DDR4-2666MHz, GeForce® GTX 1660 Ti, GDDR6 6GB)
- Oculus Quest connected to the laptop via the Oculus link cable
- Noise cancelling headphones

The participants sit in a comfortable swivel chair in a spacious room, with an appropriate lighting and temperature, and with just the presence of the experiment facilitator.

Procedure:

Step 1 (5min). Participants are welcome, and introduced to the project and test to be conducted.
Step 2 (2min). Participants fill in the consent form
Step 3 (3min). Participants fill in the demographic and background information questionnaire
Step 4 (3min). Participants fill in the simulation sickness questionnaire (SSQ)
Step 5 (4min). Part 1 - First Test Condition
Step 6 (5min). Participants fill in the IPQ and SSQ questionnaires
Step 7 (4min). Part 1 - Second Test Condition
Step 8 (5min). Participants fill in the IPQ and SSQ questionnaires
Step 9 (4min). Part 1 – Third Test Condition
Step 10 (5min). Participants fill in the IPQ and SSQ questionnaires

Step 12 (8min). Participants fill in the ad-hoc questionnaire on subtitling presentation modes
Step 13 (4min). Part 2 – First Test Condition
Step 14 (4min). Part 2 – Second Test Condition
Step 15 (8min). Participants fill in the ad-hoc questionnaire on indicators
Step 15 (2min). Participants are thanked and said goodbye.

Before each test condition, the experiment facilitator helped the participants with the appropriate setting and collocation of the equipment (HMD, headphones), and launched the experienced.

Overall, the experiment session for each user had a duration between 60-80 minutes.

Forced Camera Movements.
In order to test the effect of distance and different (especially sided) viewing perspectives, for each one of the subtitling presentation modes, smooth camera movements and transitions were programmed and triggered in each test condition, for the two clips. These transitions between positions and implicit viewpoint are sketched and listed in the figure below.

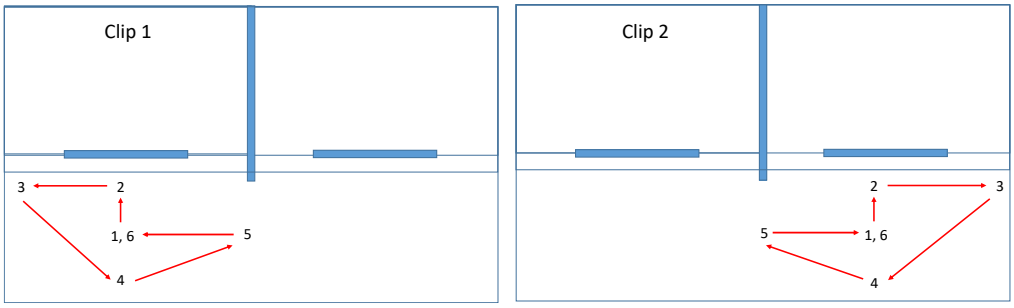


Figure 49 Forced Camera Movements in Clips 1 and 2, with smooth transitions: P1(0-30s): centered position; P2(30-60s): close to window; P3(60-90s): corner, low visibility; P4(90-120s): centered position, but far; P5(120-150s): sided, far, crosswise (sided viewing perspective); P6(150-180s): centered position.

5.2.2.3 Sample of Participants

In total, 24 users participated in the study, 12 of them were female, and they were aged between 18 and 65 years old (average of 35.12, standard deviation of 14.72), being 13 of them young adults (18-35 years), 9 of them middle-age adults (36-55 years) and 2 old adults (>55 years).

Regarding their study level, 20.8% of them had a secondary school level, 25% were undergraduate university students, 29.1% held a university degree, 20.8% held a PhD degree, and 4.1% prefer not to indicate the study level. All participants were hearing users, but the audio was muted, so the subtitles become a key element to understand the story. 58.3% of the participants had previous experience with VR content consumption using HMDs (half of them less than once a year, 28.5% of them between 1-5 times per year, 14.2% of them around a monthly basis, and 7.1% in around a weekly basis).

It must be remarked that no particular filter was applied for the participants' recruitment, beyond having good English level, as the subtitles were presented in that language. It is due to the fact that the considered subtitling modes and use of indicators can potentially provide benefits to the general audience, being potentially applied in many different scenarios, as highlighted in [Mon20b]. Future experiments will be targeted at assessing the appropriateness and potential differences between different users' profiles and VR scenarios.

5.2.2.4 Results

Results – Part 1.
The impact on presence of each of the considered subtitling presentation modes was assessed by using the Igroup Presence Questionnaire (IPQ), http://www.igroup.org/ It is composed of 14 statements/items to be rated on a seven–point scale (1 to 7). In turn, the 14 items are distributed into four sub-scales:

- *General Presence*: One single item that assesses the general 'sense of being there'.
- *Spatial Presence*: five items that measure the sense of being physically and bodily present in the virtual environment.
- *Involvement Scale*: four items that measure the attention that the subject pays to the virtual environment and the involvement experienced.
- *Experienced Realism*: four items that measure the subjective experienced sense of realism attributed to the virtual environment.

The tables below provide a summary of the mean and standard deviation values of the answers by the participants for each of the scales, together with the statistical analysis to determine whether significant differences exist between the results obtained for each test condition, by using a Wilcoxon Signed Rank test (with 95% Confidence Interval). Similarly, Figure 50 shows the boxplots of the obtained results for each test condition and IPQ scale.

The results for all three tested conditions were quite positive, especially for the Mode C (Always-Visible) and Mode B (Comic Style). Indeed, the statistical analysis show that, in the considered scenario and for the considered implementation, Mode C provides higher presence than Mode A for each one of the IPQ scales. It is also the case for Mode B when compared to Mode A in terms of the General Presence and Involvement scales. Mode C also provides higher presence than Mode B with regard to the Experienced Realism scale. It may be due to the fact that the bubbles where subtitles are presented resemble comics and animation content, thus having a negative impact on realism.

	Mean	**Standard Deviation**	**Compared to Always-Visible**	**Compared to Fixed-Positioned**	**Compared to Comic-Style**
Always-Visible	6.291	0.806	-	**p-value = 0.0032**	p-value = 0.0726
Fixed-Positioned	5.583	0.928	**p-value = 0.0032**	-	**p-value = 0.0125**
Comic-Style	5.958	0.859	p-value = 0.0726	**p-value = 0.0125**	-

Table 33 IPQ – General Presence Scale.

	Mean	**Standard Deviation**	**Compared to Always-Visible**	**Compared to Fixed-Positioned**	**Compared to Comic-Style**
Always-Visible	5.05	1.873	-	**p-value = 0.004**	p-value = 0.4054
Fixed-Positioned	4.91	1.785	**p-value = 0.004**	-	p-value = 0.0609
Comic-Style	5.025	1.907	p-value = 0.4054	p-value = 0.0609	-

Table 34 IPQ – Spatial Presence Scale.

	Mean	**Standard Deviation**	**Compared to Always-Visible**	**Compared to Fixed-Positioned**	**Compared to Comic-Style**
Always-Visible	5.146	1.025	-	**p-value = 0.015**	p-value = 0.3458
Fixed-Positioned	5.01	1.041	**p-value = 0.015**	-	**p-value = 0.004**
Comic-Style	5.188	1.019	p-value = 0.3458	**p-value = 0.004**	-

Table 35 IPQ – Involvement Scale.

	Mean	Standard Deviation	Compared to Always-Visible	Compared to Fixed-Positioned	Compared to Comic-Style
Always-Visible	3.635	1.37	-	**p-value = 0.04249**	**p-value = 0.03**
Fixed-Positioned	3.541	1.321	**p-value = 0.04249**	-	p-value = 0.1187
Comic-Style	3.448	1.329	**p-value = 0.03**	p-value = 0.1187	-

Table 36 IPQ – Experienced Realism Scale.

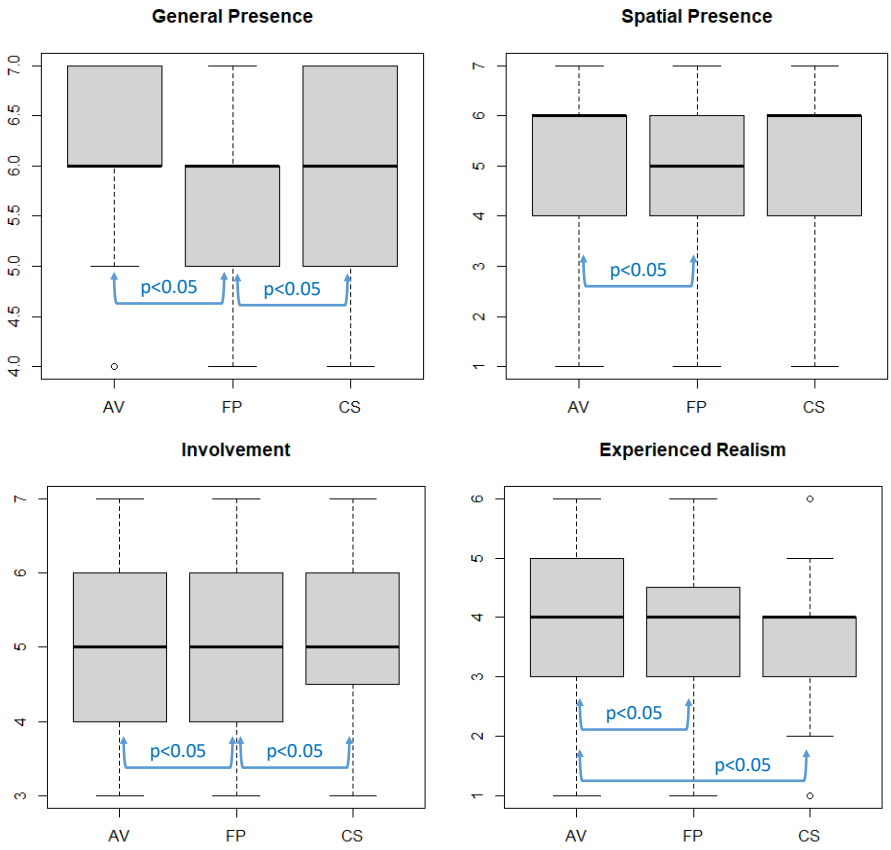


Figure 50 Boxplots of IPQ Scales for each VR Subtitling Presentation Mode (AV=Always-Visible; FP=Fixed-Positioned; CS=Comic-Style).

With regard to the SSQ, no significant effects / symptoms were noticed to be caused by the experience, in any of the test conditions.

With regard to ad-hoc questionnaire on preferences, the answers can be checked in the bar charts in Figure 51. In terms of avoiding content blocking / obstruction, the most preferred mode is Always-visible closely followed by the other two ones. In terms of ease of reading, Always-visible was the most preferred mode, followed by Fixed-Positioned subtitles. The reason why Comic-styles were the third preference may be due to the fact that they were a bit small, and thus not easy to read once the distance between the user and the speaker was large. In terms of reading comfort, the Fixed-Positioned subtitles were the third preference. It may be due to the fact that users had to deviate the viewing patters to read the subtitles and see the speakers when using that mode. In terms of speaker identification, Comic-style was the most preferred option as subtitles were associated to the target speaker in such a mode, and Always-visible subtitles are the second preferred option. As expected, Fixed-Positioned subtitles scored worst with regard to that aspect. In terms of integration with the VR content, Comic-style was again the preferred mode, as subtitles were indeed integrated with the associated speaker(s) on purpose. No significant differences were obtained between the other two modes with regard to that aspect. In terms of the impact of distance and viewing perspective, Always-visible subtitles were clearly the preferred mode, as

these two factors heavily have an impact on the other two modes. These other two modes could be adjusted such that the subtitles rendering planes are always kept normal/orthogonal to the user's viewing perspective, and their size is dynamically adapted according to the distance, but this needs further investigation and fine tuning. Finally, in overall terms, Always-visible was the most preferred mode, closely followed by Comic-style. This is in line with the results from the IPQ questionnaire, and confirms that these two innovative proposed 3D VR subtitling presentation modes do provide benefits to the user experience.

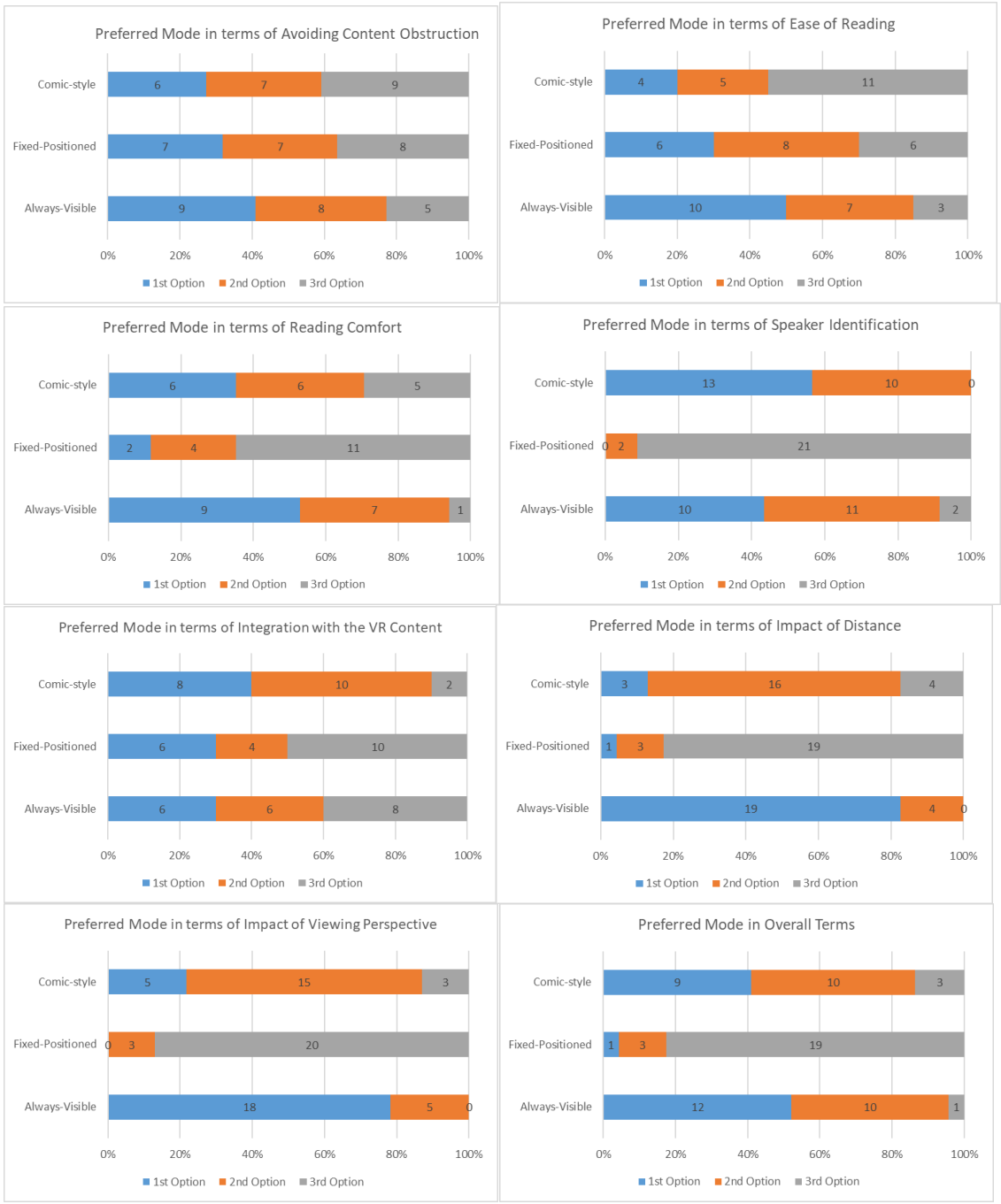


Figure 51 Preferences on VR Subtitling Presentation Modes.

Finally, the user's position and camera's position (i.e. the viewpoints) were recorded every 0.5s during all test conditions. By doing this, the goal was to compare the users' viewing patters for each one of the considered subtitling modes in order to assess whether these modes have any impact

on the omnidirectional 3D scene exploration. The obtained results for a sub-sample of eight users (those having watched each subtitling mode for the two parts of Clip 1, Table 31) are in Figure 52. These results indeed confirm that when using the always-visible mode the users did explore further the omnidirectional 3D environment, as the subtitles were never missed out. In contrast, a free exploration of the 3D environment when using the fixed-positioned modes, including the comic-style one, can cause the subtitles being out of the FoV and thus a loss of information (audio is not available, and users may have hearing impairments or not understand the spoken language). Due to this, a full exploration of the 3D environment was less common for these two modes. The first three graphs provide the heat maps from 3D front views, while the last three ones are 2D aerial projections of the previous ones.

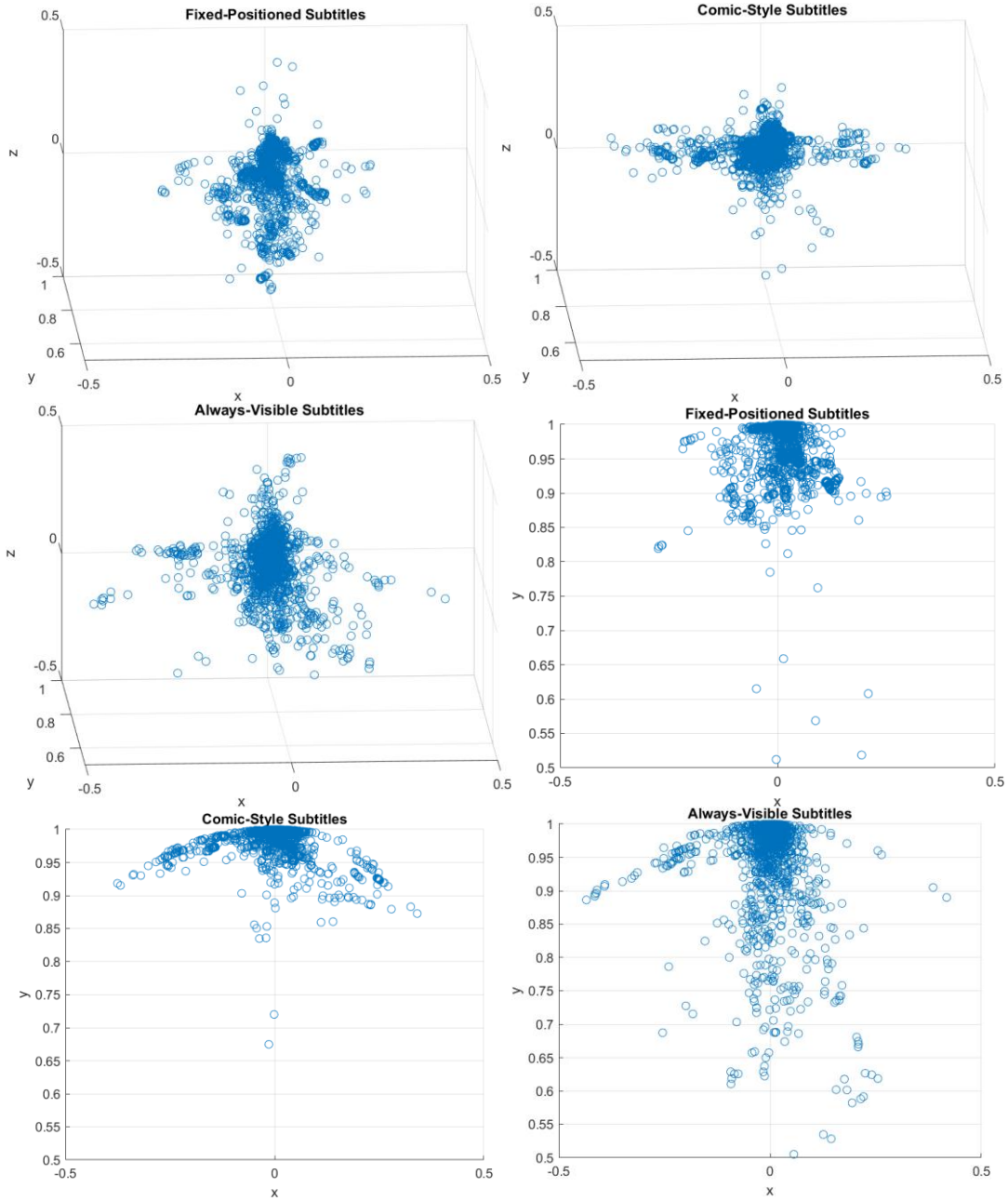


Figure 52 Heat maps of viewing patterns when using each of the considered 3D VR subtitling presentation modes in Clip. First three graphs are 3D front views for each mode, while the last three ones are 2D aerial projections of the former.

Results – Part 2. Benefits of using Indicators

In Part 2 of the experiment, the participants watched again the two parts of Clip 2 (2A, 2B), in that order (to appropriately understand the story), with the arrow indicator enabled for the Modes B and C, in a counterbalanced manner (Table 37). The arrow was positioned at the bottom right of the FoV to not block the subtitles in Mode C. In these test conditions, the forced camera movements were especially useful to demonstrate how the arrow can point anytime at the target speaker, regardless of the user's position and viewpoint.

Participant	Clip 2A	Clip 2B
Odd ID (1, 3, 5…)	Mode B	Mode C
Even ID (2, 4, 6…)	Mode C	Mode B

Table 37 Conditions for Testing the Benefits of Indicators.

Table 38 provides the answers to the ad-hoc questionnaire focused on assessing the benefits of using indicators. The obtained results shows that many participants (above 60%) believe that the arrow is beneficial for a better positioning within the VR environment and also for a better identification of the active speaker, both when using always-visible and comic-style presentation modes. Participants generally agreed that the use of indicators can contribute to a better content comprehension (almost 70%), and that their inclusion does not have a negative impact on immersive (above 80%), but even might have a positive impact (especially if relevant for content comprehension). Participants were also quite satisfied with the graphic design of the arrow, although it seems to have room for improvement. Participants were additionally asked about improvement suggestions with regard to the (use of) indicators, and the next ones can be highlighted: use of colors associated with subtitle colors (12.5%); assess the use of 3D gaming radars (12.5%); and improve the design (8.3%), but no concrete suggestions were given to achieve this, beyond enlarging a little bit the size (8.3%) and adding intermittence effects to the arrow (8.3%).

Finally, 91.6% of participants considered that personalization should be enabled for the activation / deactivation of indicators, so each consumer can choose to show or hide them based on the preferences or the content being watched.

Question Items / Answers	Totally Agree	Partially Agree	Neutral	Partially Disagree	Totally Disagree
The availability of indicators (arrow) is beneficial for a better positioning within the VR environment	7 (29.2%)	8 (33.3%)	5 (20.8%)	4 (16.7%)	-
The availability of indicators (arrow) is beneficial for a better identification of the active speaker when using always-visible subtitles	9 (37.5%)	9 (37.5%)	4 (16.7%)	2 (8.3%)	-
The availability of indicators (arrow) is beneficial for a better identification of the active speaker when using comic-style subtitles	6 (25%)	7 (29.2%)	7 (29.2%)	3 (12.5%)	1 (4.2%)
The availability of indicators (arrow) can contribute to a better content comprehension	6 (25%)	8 (33.3%)	6 (25%)	4 (16.7%)	-
The inclusion of indicators (arrow) can have a negative impact on immersion while watching VR content	-	1 (4.2%)	3 (12.5%)	8 (33.3%)	12 (50%)
The inclusion of indicators (arrow) can have a positive impact on immersion while watching VR content	6 (25%)	7 (29.2%)	5 (20.8%)	4 (16.7%)	2 (8.3%)
The provided indicator (arrow) is visually attractive	6 (25%)	8 (33.3%)	5 (20.8%)	5 (20.8%)	-

Table 38 IPQ – Benefits and Appropriateness of Indicators.

Final Questions

Finally, participants were asked general questions about the relevance of subtitling for hearing users, and of VR subtitling in particular (Table 39). Over 90% believe both that subtitles are beneficial for hearing users and that an appropriate subtitling of VR content is a relevant feature to be explored and provided. Participants were asked about the particular benefits of subtitles according to their opinion, and the main answers are listed next: when consumers do not speak the content language (33.3%); for language learning and improvement (25%); in noisy environments (20.8%) or if/when the audio volume is low (16.7%); to train the reading skills (16.7%); to understand / get the spelling of specific uncommon words and names (12.5%); and when the audio quality is not good (8.3%). Participants were also encouraged to make final free comments, and key ones were: subtitles helped them to understand the story (25%), in fast action scenes it was difficult to read subtitles (16.7%), and that hybrid modes (e.g. combining Modes B and C) could help to overcome the impact of distance and viewing perspectives but also to increase the immersion once having the speakers within the current FoV (20.8%).

	Totally Agree	Partially Agree	Neutral	Partially Disagree	Totally Disagree
Subtitles can contribute to a better content comprehension for hearing users	14 (58.3%)	8 (33.3%)	2 (8.3%)	-	-
In general, I believe that VR subtitling is a relevant feature	15 (62.5%)	7 (29.2%)	2 (8.3%)	-	-

Table 39 Relevance of (VR) Subtitling.

5.2.2.5 Discussion, Conclusions and Future Work

As with recent studies on 360º video subtitling [Mon21], this experiment has shown that an always-visible presentation mode is by far more appropriate than a fixed-positioned presentation mode for subtitles. The differences are even larger in 3D VR environments than in 360º video environments, especially due to two key factors: depth, and 6DoF (i.e. freedom to explore). Unlike in 360º video environments where subtitles can be e.g. replicated every 120º in the sphere (strategy proposed and adopted by the BBC channel) [Mon21], this is clearly not an appropriate and sufficient solution in 3D VR environments, because of potential content blocking and of the depth dimension, respectively. This experiment has shed some light on these relevant issues. In addition, a novel comic-style presentation mode has been proposed and been very well received by the users. Indeed, this mode has been shown to provide similar levels of presence than the always-visible mode, for all scales of the IPQ questionnaire, but for the Experienced Realism one (maybe due to the fact that the addition of bubbles impact the experience realism, as the story resembles more a comic or an animation film).

In addition, the comic-style mode has been the second most preferred mode by users, but has been the most preferred option in terms of relevant aspects, like speaker identification and integration with the VR content. This reflects the potential of this mode, and the convenience of keep researching on assessing its benefits, by ideally also overcoming some of its identified drawbacks (e.g. ease of reading, impact of distance and viewing modes, content blocking…).

It should be also remarked that the considered scenario brings its own particular conditions that may potentially impact the obtained results. First of all, all actions happen within a delimited area (inside the interrogation room), where the speakers hardly move. The area is also separated by a mirror from the user's area, so the user e.g. cannot move around the speakers. In addition, despite forcing camera movements, the scenario does not incite the users to navigate around the 3D VR environment, and therefore the scenario has limited 6DoF.

In general, all these findings about 3D VR subtitling presentation modes provide valuable answers to RQ1, and open to door to further research on this topic.

On the other hand, the obtained results and feedback with regard to the use of indicators are quite positive, but the hypothesis is that they would had been even more positive in scenarios with 6DoF and/or in scenarios in which the speakers and the user can freely and significantly move around the VR environment. This also provides valuable answers and insights with regard to RQ2.

Given the learned lessons and still open questions, future work will be targeted at overcoming identified drawbacks (e.g. impact of distance and viewpoints, ease of reading…) and at exploring these aspects in scenarios with unlimited – or less limited - 6DoF. Finally, two further aspects will be explored: 1) the benefits of Easy-to-Read subtitles [Onc20] in (fast action) 3D VR environments, due to identified difficulties in following the subtitles for specific scenes; and 2) exploring hybrid and advanced presentation modes and strategies, e.g. based on the user's and speaker's positions and viewpoints.

References:

[Mon20b] M. Montagud, F. Boronat, D. Marfil, J. Pastor, "Web-based Platform for a Customizable and Synchronized Presentation of Subtitles in Single- and Multi-Screen Scenarios", Multimedia Tools and Applications, Springer, 2020

[Mon20c] C. Hughes, M. Montagud, "Accessibility in 360° video players", Multimedia Tools and Applications, Springer, October 2020.

[Mon21] M. Montagud, O. Soler, I. Fraile, S. Fernández, VR360 Subtitling: Requirements, Technology and User Experience, IEEE ACCESS, 2021.

[Onc20] E. Oncins, M. Montagud, R. Bernabeu, "Accessible scenic arts and Virtual Reality: A pilot study in user preferences when reading subtitles in immersive environments", MonTI, 2020.

5.2.3 I2CAT-3.2: Evaluation of use cases: HoloConferencing / HoloMeetings

5.2.3.1 Objectives

Social VR is a novel communication and interaction medium. As confirmed in the conducted focus groups (see e.g. D4.4), the applicability of Social VR spans from shared media consumption to other highly interactive use cases, like holoconferencing and holomeetings.

In the project, many experiments to assess both the quality of interaction and of the end-users' representation have been conducted (e.g. see D4.2, D4.4, and other experiments reported in this deliverable). On the one hand, the quality of interaction / communication effectiveness was evaluated under the context of a specific content consumption and/or VR storytelling (e.g. watching movie trailers, and pilots 1-3). On the other hand, the same happened for evaluating the end-users' representation, or the quality was assessed for specific self or other representations.

The main objective of this pilot action is to evaluate the appropriateness and benefits of the VR-Together technology in an additional use case: holoconferencing or holomeetings. While the key focus is on evaluating the quality of communication and interaction, the experiment also evaluates presence and togetherness, as well as the task completion effectiveness, naturalness and comfortability. In addition, the impact of the number of users is also assessed in the experiment, by considering sessions with 2 and of 4 users. Finally, an additional goal was to prove that the VR-Together technology is robust enough to hold many consecutive sessions of both four simultaneous users, and of two simultaneous sessions with two users, run in 2 days (4 sessions of 4 users, and 8 sessions of 2 users per day).

The holomeetings are conducted in a newly created 3D meeting room, in which the users are sit around a round table, as shown in the figure below. No dynamic VR content is added. By doing this, we exclusively move the focus to the end-users' representations and to the audiovisual communication, and not to the presented content. In addition, users are asked to complete tasks by performing verbal and non-verbal gestures, which increases the audiovisual attention towards them.

. Figure 53 Overview of the multi-party holoconferencing / holomeeting scenario in Social VR.

5.2.3.2 Methodology

Setup and Equipment:

- Capture System: single-camera point cloud with K4A sensor.
- Setup: four users located in different spacious rooms at the i2CAT lab premises (Barcelona), with an appropriate lighting and temperature. The users were sit in a chair. The room only had the presence of the experiment facilitator before and after the test conditions.
- Equipment:
    - PCs: PCs and gaming laptop with enough resources to run the VR-Together technology and scenarios smoothly.
    - 2 Oculus Rift, and 2 Oculus Quest 2 connected to a PC via the Oculus link cable
    - Noise cancelling headphones
- Delivery protocol: socket.io

Stimuli:

No dynamic VR content was presented to the participants, but they were just teleported to the newly created virtual 3D meeting room, with chairs around a round table.

Test Description: Test Conditions and Tasks

A key goal of this experiment was to evaluate the impact of the number of users on both the system's performance and on the user experience. Therefore, two test conditions were prepared, and presented in a counterbalanced manner:

- Test Condition A: Sessions of 2 users
- Test Condition B: Sessions of 4 users

As the groups were of 4 people, two simultaneous sessions of 2 users were held when it was the time for Test Condition A.

Another key goal was to have highly interactive sessions to richly evaluate the quality of interaction / communication in the shared virtual experiences. In order to boost interaction, and to ensure that

visual, verbal and non-verbal (e.g. gestures) interaction happened during the sessions, “Guess What?” gamification tasks were planned, for the following categories:

1. Films
2. Animals
3. Sports (optional)
4. Jobs
5. Celebrities
6. Cities (optional)

In iterative rounds, each participant had to choose 1 option of each category to be guessed by the others. The participants were requested to: 1) not choose too easy solutions; 2) to also provide and answer with verbal hints, and request for them; 3) once guessed, to boost a conversation of why that option was important for them, linking to related news, stories or anecdotes.

The “Guess What?” categories were always played in that order, and thus were not linked always to the same test condition. However, the categories 3 and 6 were optional, and just considered in the case of availability of remaining time in the session.

The duration of each test conditions was approximately 8 minutes.

Figure 54 Setup, equipment and recreated multi-user holomeeting scenario.

Procedure / Evaluation Steps:

Step 1 (5min). Participants are welcome, introduced to the project and to the test to be conducted.
Step 2 (2min). Participants fill in the consent form
Step 3 (3min). Participants fill in the demographic and background information questionnaire
Step 4 (3min). Participants fill in the simulation sickness questionnaire (SSQ)
Step 5 (8min). First Test Condition (A for odd sessions, B for even sessions)
Step 6 (15min). Participants fill in the Social VR questionnaire, SSQ and IPQ questionnaires
Step 7 (20min). Participants fill in the SSQ, SUS, NASA TLX and an ad-hoc questionnaire, and participate in a semi-structured interview.
Step 8 (2min). Participants are thanked, asked for their willingness to participate in future experiments, given an Amazon voucher of 30 euros, and said goodbye.

Before each test condition, the experiment facilitator helped the participants with the appropriate set up of the equipment (HMD, headphones), and launched the experienced.

Overall, the experiment session for each user had a duration between 60-80 minutes

5.2.3.3 Sample of Participants

In total, 32 users participated in the study, 17 of them were female, and they were aged between 18 and 40 years old (average of 24.31, standard deviation of 6.65). 25 of them were right handed, 6 left handed, and 1 was ambidextrous.

With regard to their efficiency using computers, 4 of them declared to be novice, 20 intermediate, and 8 advanced. Half of them stated not having had any previous VR experience.

5.2.3.4 Results

This section reports on the results for the sessions with 4 participants.

Experience Questionnaire

After each test condition, each participant was asked to complete the Experience Questionnaire. As detailed in D4.2, D4.4, and in the pilot 3 section, the questionnaire includes question items about emotions, feelings, perception and opinion regarding crucial aspects of VR-Together, and is categorized in for main sections:

- quality of interaction (including emotional experience, quality of the communication, and naturalness of the communication)
- social connectedness (including feeling of togetherness, feel of emotional closeness, and enjoyment of the relationship)
- presence / immersion (including plausibility and place illusion…)
- additional issues (realism, how much the contents like to the users…)

Most of the question items for each part of the questionnaire had to be answered by using a 5-level likert scale, with the potential answers detailed in Table 40. The acronyms in that table are then used in the next tables providing the results from each part of the Experience Questionnaire.

Acronym	Meaning	Assigned Score
TD	Totally Disagree	1
PD	Partially Disagree	2
NN	Neither Agree nor Disagree	3
PA	Partially Agree	4
TA	Totally Agree	5

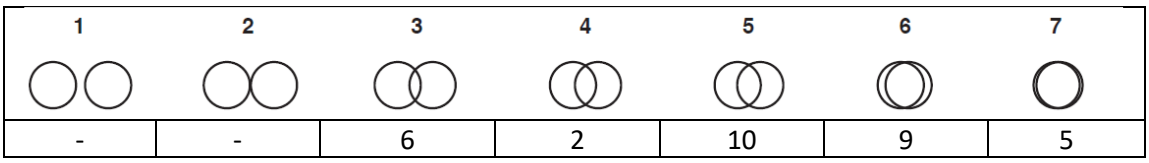
Table 40 Acronyms used for the ratings in the Experience Questionnaire (Table 41).

Questions	TD	PD	NN	PA	TA
Part 1. Quality of Interaction					
Q2. "I was able to feel the other users' emotions in the shared VR scenario."	-	-	2	**16**	14
Q3. "I was sure that the other users often felt my emotion."	-	-	4	**18**	10
Q4. "The virtual experience with the other users seemed natural."	-	-	-	**23**	9
Q5. "The actions used to interact with the other users were similar to the ones in the real world."	-	-	2	**19**	11
Q6. "It was easy for me to contribute to the conversation."	-	-	1	8	**23**
Q7. "The conversation with the other users seemed highly interactive."	-	-	1	**16**	15
Q8. "I could readily tell when the other users were listening to me."	-	-	1	12	**19**
Q9. "I found it difficult to keep track of the conversation."	**19**	11	2	-	-
Q10. "I felt completely absorbed in the conversation."	-	-	-	12	**20**
Q11. "I could fully understand what the other users were talking about."	-	-	-	11	**21**
Q12. "I was very sure that the other users understood what I was talking about."	-	-	1	13	**19**
Q13. "I often felt as if I was all alone in the virtual experience."	**26**	6	-	-	-
Q14. "I think the other users often felt alone in the virtual experience."	**22**	10	-	-	-
Part 2. Social Connectedness					
Q15. "I often felt that the other users and I were sitting together in the same space."	-	-	-	12	**20**
Q16. "I paid close attention to the other users."	-	-	-	**18**	14
Q17. "The other users were easily distracted when other things were going on around us."	9	7	**14**	1	-
Q18. "I felt that the having the VR experience together enhanced our closeness."	-	-	3	**21**	7
Q19. "Having the VR experience together created a good shared memory between us."	-	-	-	**16**	**16**
Q20. "I derived little satisfaction from the virtual shared experience."	**16**	12	4	-	-
Q21. "The VR shared experience with my partner felt superficial."	10	**15**	6	-	1
Q22. "I really enjoyed the time spent with the other users."	-	-	-	6	**22**
Q24. "In the virtual world I had a sense of 'being there'."	-	-	-	12	**20**
Q25. "Somehow I felt that the virtual world was surrounding me and my partner."	-	1	3	12	**16**
Q26. "I had a sense of acting in the virtual space, rather than operating something from outside."	-	-	**3**	11	**16**
Q27. "My virtual shared experience seemed consistent with a real world experience."	-	-	6	**18**	8
Q28. "I did not notice what was happening around me in the real world."	-	3	2	**15**	11
Part 3. Presence / Immersion					
Q29. "I felt detached from the outside world while having the VR experience."	-	-	9	9	**14**
Q30. "At the time, the shared VR experience with the other users was my only concern."	-	-	5	13	**14**
Q31. "Everyday thoughts and concerns were still very much on my mind."	**11**	9	**11**	-	1
Q32. "It felt like the VR shared experience took shorter time than it really was."	-	-	3	12	**17**
Q33. "When having the VR experience together, time appeared to go by very slowly."	**20**	8	4	-	-
Extra Questions					
Q34. "I liked the created VR scenario for holding virtual meetings."	-	-	1	14	**17**
Q35. "The created VR scenario is realistic (i.e. resemble a real scenario),"	-	-	4	**15**	13
Q36. "The spatiality in the VR scenario (i.e. perceived distances and sizes of elements, including the participants' bodies) is consistent with a real-life scenario."			5	11	**16**

Table 41 Results from the Experience Questionnaire (Holoconferencing Experiment).

The obtained results are very satisfactory, for each part of the Experience questionnaire. This reveals that the VR-Together technology is ready to provide high: 1) quality of interaction; 2) social connectedness / togetherness; and 3) presence / immersion. The participants were also very satisfied with the created VR scenario, in terms of its quality, realness and spatiality.

In addition to the questions listed in Table 41, Q23 asked about *how emotionally close to the other users did each user felt*, using a 7-point scale with two circles separated by different distances (from separated to totally overlapped), as seen in the figure below. From the results, as the majority of answers were given to the options 5 and 6 with significant overlapping between the circles, it can be concluded that participants did feel quite emotionally close, which is also a sign of the feeling of togetherness and intimacy.



-	-	6	2	10	9	5

Figure 55 Emotional closeness between participants.

Simulation Sickness

With regard to the SSQ, no significant effects / symptoms were reported to be caused by the experience, in any of the test conditions. This was confirmed during the semi-structured interviews when asking explicitly about this. In addition, none of the participants stated to have felt dizziness and got tired as result of the two experienced test conditions.

Usability

With regarding to **usability**, the SUS average score is 91 (above average, as the average is 68), as shown in Figure 56. The letter grade is A+, and the obtained score corresponds to the percentile range: 96-100.

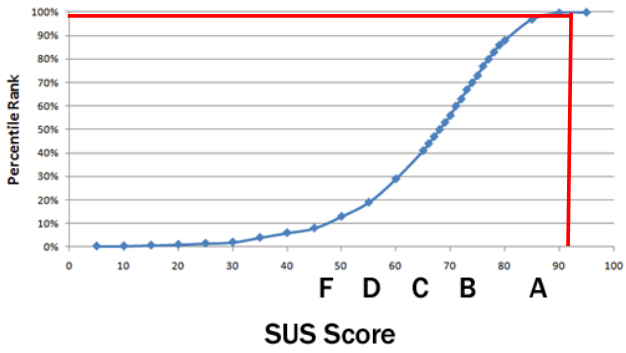


Figure 56 SUS score and percentile rank for the Social VR platform.

Ad-Hoc Questionnaire

An ad-hoc questionnaire (Table 42) was designed to capture insights about additional features and aspects of the evaluated technology and scenario.

The first part of the questionnaire focused on the perceived audio-visual quality. In general, participants were quite satisfied with the visual quality of the scenario and with the audio quality from all users, regardless of their positions. When referred to the visual quality of the end-users' representation, participants were quite happy with their representation, and with the others' representations, especially noticing a better quality for the users located in front of them. This is reasonable, as a single-camera capture system was used in the experiment, placed in front of the users. Therefore, the full volume was not captured, especially with regard to the lateral and posterior parts of the body. Still, the quality was acceptable by the participants in such cases.

The second part of the questionnaire focused on comparing the perception of the virtual scenario / experience to a real one. The visual quality of the VR scenario, as well as the audio quality and naturalness of the gestures in the virtual scenario, were generally rated as equivalent to the ones in a real scenario. This was almost the same regarding the fluidity of gestures, although many users

rated it as slightly worse than in real scenarios. Likewise, most of the participants rated the overall communication quality and virtual experience as equivalent to the ones in real scenarios.

Finally, and interestingly, most of the participants rated the overall experience with two users as equivalent (46.9%) or slightly better (40.6%) to the one with your users. It reveals that the performance of the system with 4 users was still satisfactory, and that despite perceiving noticeable limitations in the visual quality of the users placed next to oneself, the experience with 4 users was still very satisfactory, mainly due to the higher interaction possibilities.

The third part of the questionnaire assesses the potential and impact of the developed technology and considered scenario. Most of the participants believe the system is ready to hold online virtual meetings. Besides, the participants generally agreed that the distance between them in the virtual scenario was appropriate. Despite of the noticeable visual quality limitations of the users' representation, the participants generally believe that the current quality is enough to enable effective and comprehensive interactions and collaborations. The fact of wearing an HMD was still perceived as a barrier for some participants, but not a limiting one (most of the participants affirmed to have felt comfortable during the experience in the interviews conducted at the end of the experiment). In general, participants seem to be very interested in using systems like the evaluated one for holding online meetings, including scenarios with more than 4 users. Finally, most of the participants believe that these kinds of systems can contribute to a more sustainable environment, and to save time and money.

Part 1. Ad-hoc questionnaire – Audio-visual Quality					
Questions (Rating Scales: 1 = bad, 2= poor, 3= fair, 4= good, 5= excellent)	1	2	3	4	5
The visual quality of the virtual scenario	-	-	1	**22**	9
The visual quality of my representation	-	2	13	**15**	2
The visual quality of the representation of the user(s) next to me	-	4	**16**	12	0
The visual quality of the representation of the user(s) in front of to me	-	-	7	**21**	4
The audio quality from the user(s) next to me	-	-	-	**19**	13
The audio quality from the user(s) in front of to me	-	-	-	**20**	12

Part 2. Ad-hoc questionnaire – Comparison to a Real Scenario					
Questions (Rating Scales: 1 = much worse, 2= slightly worse, 3= equivalent, 4= slightly better, 5= much better)	1	2	3	4	5
The visual quality of the virtual scenario, compared to a real one	-	12	**16**	4	-
The overall experience with 2 users compared to the one with 4 users.	-	2	**15**	13	2
The overall virtual experience with one in real life	-	2	**19**	3	1
The visual representation of the users in the virtual experience compared to in a real scenario	2	**19**	11	**-**	-
The audio quality in the virtual experience compared to the one in a real scenario	-	5	**20**	7	-
The naturalness of the gestures in the virtual scenario, compared to a real scenario	-	8	**22**	2	**-**
The fluidity of the gestures in the virtual scenario, compared to a real scenario	-	11	**19**	2	**-**
The overall communication quality in the virtual scenario, compared to a real scenario	-	5	**19**	7	1

Part 3. Ad-hoc questionnaire – Potential and Impact					
Questions (Rating Scales: 5 = Strongly Agree, 4= Agree, 3= Neutral, 2= Disagree, 1= Strongly disagree)	1	2	3	4	5
This system is effective to hold virtual meetings.	-	-	-	6	**26**
The distances between the participants was appropriate	-	-	5	8	**19**
The quality of the users' representation is enough to enable effective and comprehensive interactions and collaborations	-	1	1	**21**	9

Wearing a HMD prevented from an effective and satisfactory interaction experience (due e.g. to the unfeasibility of seeing each others' eyes / sights)	6	6	**16**	4	-
I would use a system like this one for meetings and collaborative tasks in virtual scenarios with up to four users	-	-	-	10	**22**
I would use a system like this one for meetings and collaborative tasks in virtual scenarios with more than four users	-	-	4	13	**15**
These kind of systems can contribute to a more sustainable environment	-	-	5	12	**15**
These kind of systems can contribute to save time and money.	-	-	1	8	**23**

Table 42 Results from the Ad-Hoc Questionnaire (Holoconferencing Experiment).

Semi Structured Interviews:

After having finished experiencing with the two test conditions, and having filled in the associated questionnaires, the participants took part in a semi-structured interview with projects researchers. The audio recordings of the interviews were transcribed and coded, following an open coding approach [Tho06].

Since the interviews were conducted with all participants of each group together, their answers have been transcribed and coded both as individual participants (labelled P1-P32) and as groups (labelled G1-G8). From the coded transcripts, several themes emerged, for which we elaborate further next.

Key Impressions

Participants were firstly asked their impressions and keywords about the virtual meeting experience. The next ones can be highlighted: Impressive (46.9%), futuristic (43.8%), innovative (43.8%), amazing (31.3%), funny (31.3%), WoW! (25%), next-generation videoconferencing (25%), interesting (25%), surprising (21.9%), technology enabler for new possibilities (15.6%), incredible (12.5%).

Benefits and Potential of Social VR

All participants thought that the Social VR platform enabled them to experience "social presence", having felt identified with the end-users' representations, including their own and the other's representations. "It was not an avatar, but me!", stated by P4, P9, P19 and P28. The participants generally felt "being together" with the others, which enriched the overall experience. G1, G4 and G7 stated "We felt together, sharing an activity and experience, and this is really an added value to VR". "That is super! You can meet with whoever you want, anywhere, anytime!", G2 expressed. "This really gives the feeling of being in the same place, together", mentioned by G6. Both the scenario and experience was found immersive by all participants.

The participants in general felt comfortable in the virtual environment. A few participants (12.5%) mentioned to had felt a bit tense at the first contact with the Social VR platform, because of the uncertainty, but then they rapidly felt more relaxed. None of the participants felt uncomfortable by the fact of hearing the HMD, even many of them state to have forgotten about wearing it (37.5%).

The quality of communication and interaction was also found satisfactory in general. Even though visual artefacts were noticed, it was impressive to see themselves and the other users represented inside the VR scenario. "I could even see my tattoos and details of my clothing", participants from G3 and G5 stated. Participants also pointed out that the perceived delays were a minor issue (25%) and that although facial expressions were partially blocked by the visual quality and the HMD occlusion (31.25%), all of them stated that having realistic representations of users is impressive and enabled natural and rich interaction. None of the participants reported on lack of fluidity or audio artefacts during the experiences. "It was clear that the audio was directional, as you could identify from where it came", stated users from G2 and G8. In general, the interactions between

the participants were perceived as natural. “The interaction was natural. It is not the same as in a real scenario, as eye contact is missing, but we were able to effectively communicate and perform the requested tasks clearly and effortlessly”, stated by G1 and G5. Indeed, none of the participants reported on communication and interaction problems during the test sessions.

Participants also found out the limitations of having used a single-camera capture system in the experiment. First, the quality of the representations of the users placed next to them was perceived as lower to the one of the users in front of them. Second, artefacts in the self-representations were also mentioned (by G1, 63, G7), especially with regard to the arms, hands and feet. P7, P11 and P26 mentioned: “The quality of my partner’s representation seemed better than mine”. Three participants (P1, P13, P31) suggested to decrease the sizes of the points in the end-users’ representations (Point Cloud format).

All participants believed that the photo-realistic representations for the end users can help maintain, strength, and even create new, relationships in life. G1, G3 and G6 stated, "It is a very innovative and useful solution. We have friends and family members living apart. This would enable us to meet and share experiences, overcoming distance barriers, and saving time". In general, participants believe that these systems can be applied to interact with both known people and new contacts. Suggested applicability use cases for this Social VR technology are enumerated later.

Many participants (50%) affirmed it was an amazing experience for them, and that Social VR can be a powerful tool to evade from the real world in certain situations (37.5%), but especially as a very valuable communication medium to overcome the social distancing measures brought by the pandemic, as remarked by seven out of the eight groups.

Missing aspects / Weaknesses in Social VR

Many participants (62.5%) explicitly mentioned that higher visual quality would be desired, especially for the self-representation and for the representations of the lateral users. “The current level of quality is enough to effectively communicate and interact, but this is not yet as in real scenarios”, stated by four of the groups. “The quality of the end-users’ representation should improve in the future”, declared G7. None of the participants claimed on lack of fluidity and levels of delays affecting the overall experience.

Integration of multi-sensory stimuli, like scents (12.5%) and especially haptic feedback (75%), was identified as a missing aspect. G1, G3 and G8 stated: “It would be great if you could touch things, and if the haptic interactions indeed have an effect on the VR environment or story”. “The fact of having the table gave the impression that I could pick up objects from it, or leave objects on it”, stated by two participants. Participants in four out of the eight sessions tried to shake hands in the virtual scenario.

75% of participants would also like to move freely in VR (e.g., 6DoF). “It would be great if you could move around, get closer to elements and participants in the shared environment”, stated by participants in G2 and G4.

With the combinations of haptic feedback and 6DoF features, participants mainly pursue enjoying more interactive and active experiences. “If you can actively explore things and complete collaborative tasks together, as well as influence the VR environment, then you would be able to perform very useful activities through Social VR”, P2 and P25 remarked.

Potential Use Cases

In general, the participants foresee a big impact of Social VR. They identified the following use cases as the most interesting for Social VR: meeting and conferencing (87.5%), gaming (50%), training and education (37.5%), virtual consultation (37.5%), and virtual events (37.5%), like conferences (60%), dating (25%), and shared video watching (25%).

Participants believed that Social VR is a powerful medium to meet with known users, but also to meet new contacts. All participants shown interest in using Social VR in the future. “It would be great to have such a system at home!” stated members of many of the groups. “This is the next generation conferencing tool”, stated by P5 and P16. Participants of two of the groups were also concerned about the price of the technology: “Of course, it is great, but adoption will depend on its cost”. A few participants (12.5%) also declared that Social VR might be more adequate in corporate environments in the short term, and not yet for domestic environments. Other ones (12.5%) shown concerns about Social VR contributing to sedentariness.

All the participants stated that haven more than 2 participants in the same session is very useful and provides added-value, although the visual representation of the users placed at the sides would need to improve, with the whole volume.

Next generation of Social VR
The next generation of Social VR is envisioned by participants as:

- Evolved version with higher quality volumetric representations, ideally not blocking the facial expressions, i.e. with HMD removal (75%).
- Scenarios with 6DoF, and with no cables (62.5%).
- eXtended Reality (XR) environments where the boundaries between the real and the virtual worlds are blurred (50%).
- Multi-sensory environments, stimulating all senses, especially touch and ideal smell (37.5%).

Two groups also suggested adding a feature to personalize your meeting environment and or event import the desired ones. “It would be great if you could choose the place where you want to meet, and personalize it”, stated P8 and P29.

Finally, all participants declared their willingness in participating in future experiments on this Social VR / holoportation technology.

[Tho06] D. R. Thomas. 2006. A general inductive approach for analyzing qualitative evaluation data. American journal of evaluation 27, 2 (2006), 237–246.

5.2.3.5 Discussion and Conclusions

This experiment has shown the readiness and potential of the VR-Together technology in an additional relevant use case: multi-party holoconferencing or holomeetings.

Despite of some known limitations, like the current level of visual quality, the limitations of single-camera capture systems for non-frontal views, and HMD blocking issues, the participants were in general very satisfied with the experience, being very impressed and showing high interest. The system also perform smoothly and in a robust manner for all sessions, which is also a proof of evidence of the maturity of the technology.

The fact that the single-camera capturing system already provides these satisfactory results is very promising, as it significantly lowers the deployment costs, in terms of required space, setup time and monetary inversion to enjoy the provided Social VR experiences.

5.3 Underlying Technology

5.3.1 CWI-3.2: Point Cloud Tiling

To optimize the delivery of the large amount of data required to provide volumetric photorealistic point cloud reconstructions for social VR applications, we evaluated the effectiveness of viewport adaptive streaming techniques. To perform the evaluation, we used pre-recorded point cloud sequences from the MPEG 8i dataset that contains four sequences; longdress, loot, red&black and soldier that are played back at 30 frames per second. We designed a low complexity tiling approach to create spatial segments of point cloud objects suitable for real-time applications (see Figure 57). We used the MPEG anchor codec as it allows for low-delay encode and decode. The tiles were then encoded at multiple quality levels to prepare an adaptation set. The user can then select an appropriate quality level for each tile and decode the tiles independently before rendering.

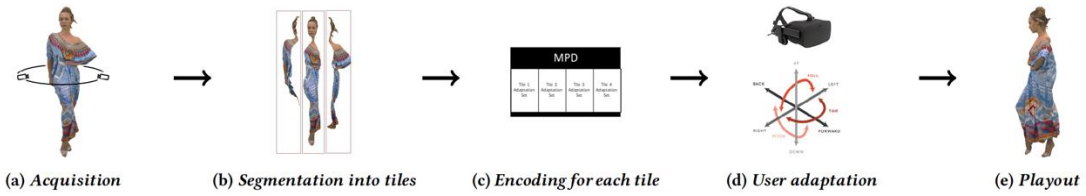
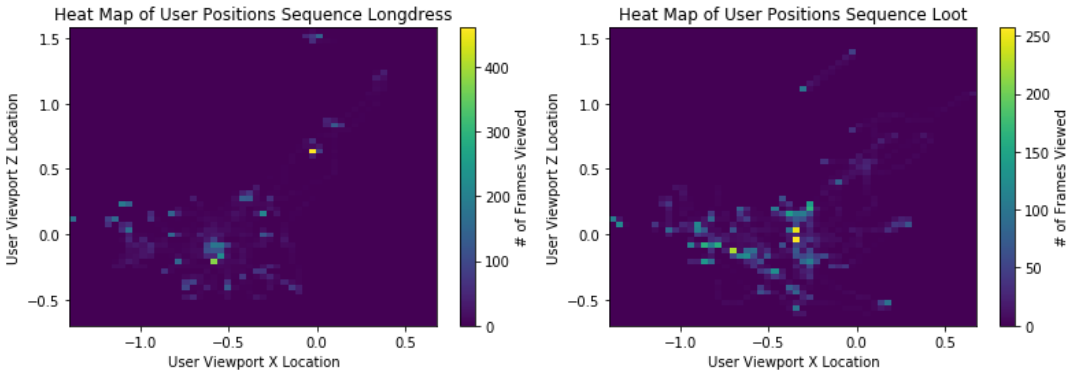


Figure 57 Overview of the proposed tiling approach.

To perform the evaluation, we first collected navigation patterns of 26 users viewing the original pre-recorded point cloud sequences while they were free to navigate the scene with 6 degrees of freedom. The variation in movements and interaction behaviour shown below in the form of a heat map of user positions on the floor plane indicate that a user-centred adaptive delivery mechanism could lead to significant gains in terms of perceived quality.



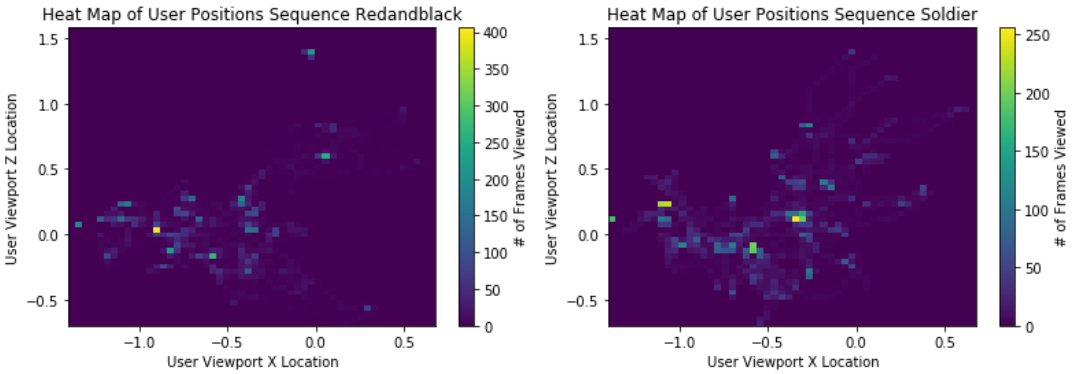


Figure 58 Heatmaps of user positions on the XZ (floor) plane during playout of each of the sequences.

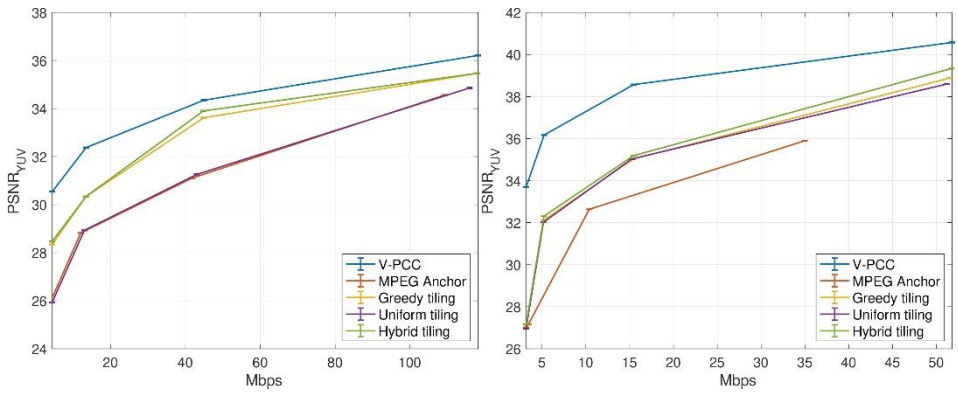
We then perform an objective evaluation using three tile selection algorithms. We use the image based PSNR metric to evaluate the point clouds visualized from each of the recorded navigation paths. All of the frames of each of the contents were rendered on the native player at a resolution of 1920x1080. The scene was set to have a uniform of RGB 0, 177, 85 to provide maximal contrast with the point clouds being rendered. The evaluation was performed only on parts of the image containing the point cloud. We compare the original content with the decoded content in the YCbCr colour space averaged across the channels using the weights proposed in (Ohm et al, 2012) as this has been shown to be closely correlated with human perception.

The evaluation was performed at static bitrate targets used in the MPEG PCC standardization activity. We defined the utility of a tile by comparing the orientation of the tile with the orientation of the user's viewport with the highest utility tile directly facing the user. We evaluated the following tile selection algorithms.

- Greedy bitrate allocation: The highest quality representation is provided for the highest utility tile and then we move on to the next highest tile until the bitrate budget is spent
- Uniform bitrate allocation: The representation of tiles is increased one step at a time starting with the highest utility tile

Hybrid bit rate allocation: The representations of tiles oriented towards the user are first uniformly increased in order of utility. The representations of the remaining tiles are then uniformly increased until the bitrate budget is spent.

The results of the evaluation are shown below:



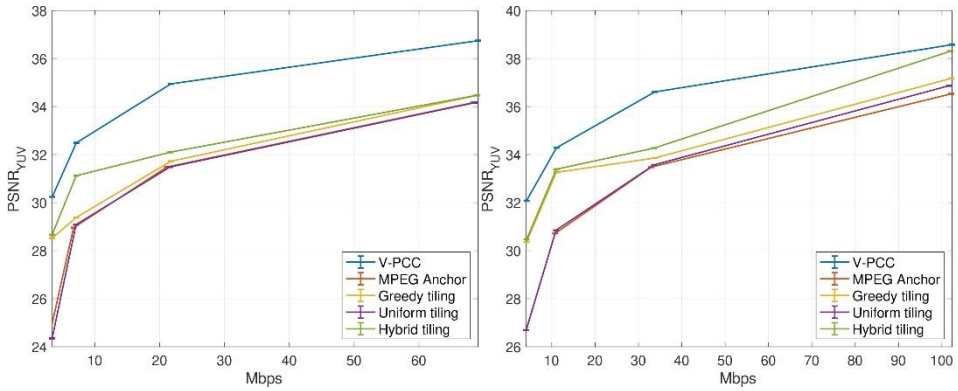


Figure 59 PSNR computed on the YUV channels against achieved bit-rate, expressed in Mbps, averaged across frames and navigation paths.

The analysis confirms that considerable gains can be achieved by exploiting user adaptive streaming of point clouds. We observed the most gains with the hybrid bit rate allocation. We observe Bjontegaard rate savings of up to 57% over the non-adaptive approach for the test sequences (Longdress: 57.15%, Loot: 50.43%, Red&black: 42.85% and Soldier: 46.04%).

The complete study and results have been published in ACM Multimedia 2020:

- S. Subramanyam, I. Viola, A. Hanjalic, and P. Cesar, "User Centered Adaptive Streaming of Dynamic Point Clouds with Low Complexity Tiling," in Proceedings of the ACM Multimedia Conference (ACM MM 2020), Seattle, USA, October 12-16, 2020.

5.3.2 TNO-3.1: Evaluating tethered and non-tethered HMD's

We see the VR device market is changing: mobile based VR (Google Daydream, Samsung Gear) has lost his attractiveness, but non-tethered devices (like Oculus Quest/Quest 2) became more popular. The issue is that non-tethered devices have more limited compute resources, so this might have an impact on the capabilities and quality when using for the VRTogether system. In this experiment, we compared tethered to non-tethered performance using the lightweight web-based pipeline in a conference room scenario.

The evaluation is based on the VRTogether Web Client is, which is fully web-based and thus can be used on many mobile devices (i.e. an Android based Oculus Quest device), with only little modifications. One thing that will not work is to support the capture and upload of the user capture directly on the device. For this we developed an upload only client that is not part of this evaluation. The upload only client can run on any Windows-based laptop or PC (considering it will support an RGBD capture sensor, i.e. it needs to support at least USB 3.0 connections).

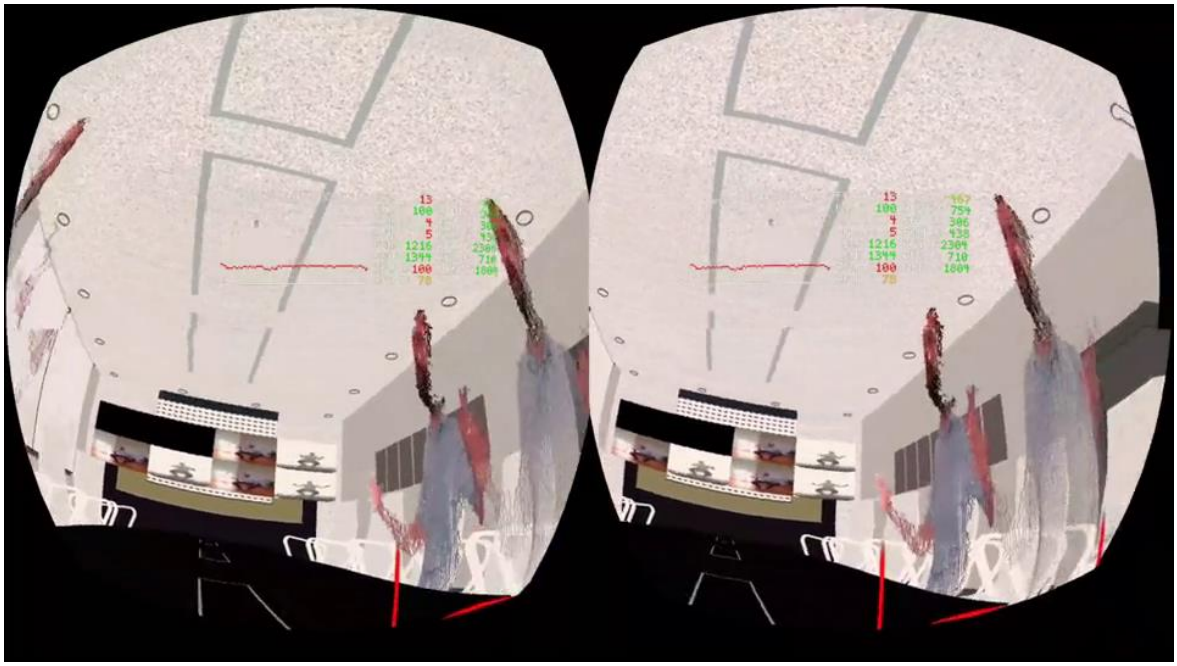


Figure 60. Example image of VRTogether Web-Client running on the Oculus Quest.

5.3.2.1 Pre-Test

Prior to setting up the evaluation we conducted a set of tests in order to identify the suitability of running the web client on the Oculus Quest HMD. In these tests we simply run our Web Client on the device under different browser configurations with two main results:

- Oculus browser vs. Firefox Reality: Not many VR capable browser currently exist (browsers that support the WebXR API and thus display Three.js-based web content in VR mode). While on regular android phones Chrome would be the browser of choice for (also for WebXR content), this option is not available on a Oculus Quest. Thus, we were only to test 2 options: Oculus Browser (comes with the device and is some derivate from Chromium) and Firefox Reality (needs to be installed manually). Both browsers fully support the WebXR APIs. However, in our testing Firefox Reality showed a better initial performance and thus was used for our further testing.
- Performance, background and user representation: simply running the web client over the Firefox Reality Browser did not result into acceptable rendering performance (high CPU / GPU and low frame rate). Therefor we reconfigured the Web Client to offer a simple 360-degree office room image as background and configured the rendering of users to 2D sprites accordingly for our evaluation.

Figure 60 shows how the Web-Client is shown on the Oculus Quest. The image was taken in pre-test showing users in 3D in a 3D room with performance overlay active. The test shows with a framerate of 13fps that this configuration is not suitable, resulting into a 2D 360-degree rendering approach for the evaluation.

5.3.2.2 Methodology

In this evaluation we compare two scenarios, regular PC with web client (without VR HMD) and Oculus Quest (without upload). In both cases a central MCU is used for stream connection, while all user streams are simulated pre-recorded video streams.

The regular web client evaluation had the following configuration:

- Device: Laptop MSI GS65 8RF Stealth Thin (CPU: Intel Core i7-8750H CPU @ 2.20GHz (6 cores / 12LPUs; GPU: NVIDIA GeForce GTX 1070 Max-Q; Memory 32 GB)
- Network connection: wired Internet access with (30Mbit/s up and 50 Mbit/s down)
- Room: 3D GLTF model of a conference room
- User representation: 2D sprites (RGB+Chroma background in 540x800px resolution blend into the environment)
- Browser: Google Chrome
- Measurements: Resources Consumption Metrics (RCM) measurement tool (https://github.com/ETSE-UV/RCM-UV)

The Oculus Quest evaluation had the following configuration:

- Device: Oculus Quest
- Network connection: Internet over USB connection (gnirehtet; 30Mbit/s up and 50 Mbit/s down)
- Room: 360-degree 2D image of a conference room
- User representation: 2D sprites (RGB+Chroma background in 540x800px resolution blend into the environment)
- Browser: Firefox reality browser
- Measurements: OVRMetricsTool_v1.4 (https://developer.oculus.com/downloads/package/ovr-metrics-tool/)

5.3.2.3 Results

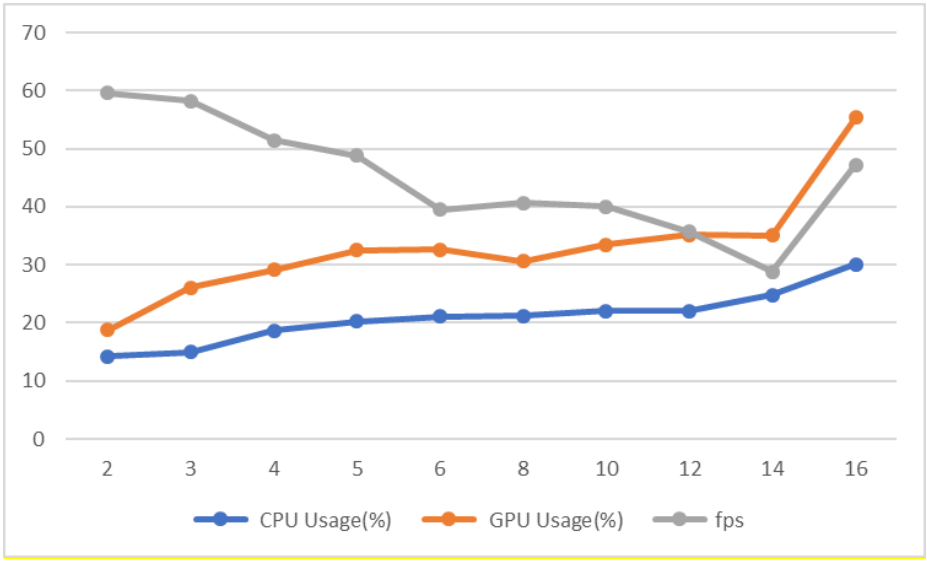


Figure 61. Results of Regular Client on Laptop.

On a "normal PC" the results show an overall suitable utilization of resources (see Figure 61). This is the client and MCU can support up to 16 simultaneous users. It is important to mention here that therefore that we did not use a VR HMD not all resources are utilized and we can expect that higher resource usage and rendering framerates will occur while in VR mode (this is due to internal optimisation strategies of the Google chrome browser).

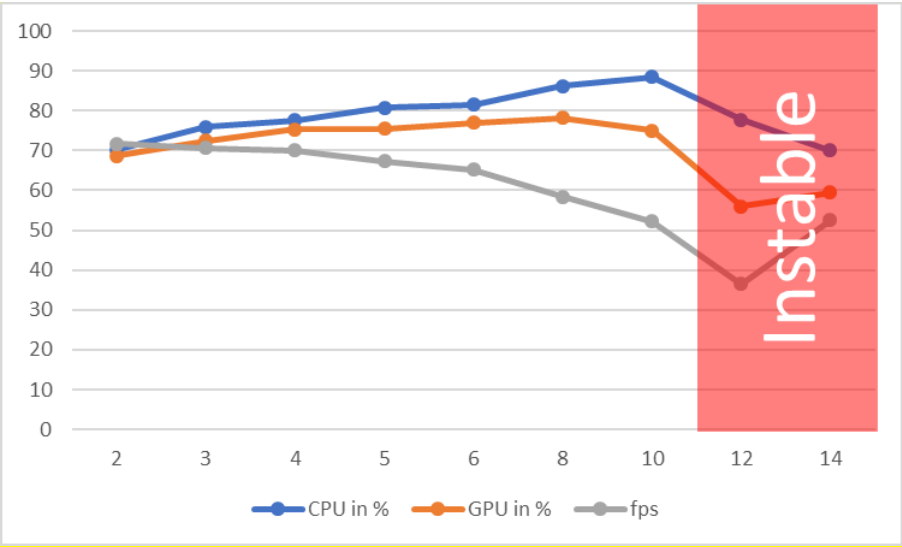


Figure 62. Results of Non-Tethered client on Oculus Quest.

The evaluation shows a much higher resource utilization (see Figure 62) for the Oculus Quest condition. In particular, we can see instabilities from above 10 simultaneous users. This is the device cannot keep up the rending of frames and we can consider this unusable. Thus, in total our current client and MCU combination can support up to 10 simultaneous users.

5.3.2.4 Conclusion

With our evaluation we show that while on a "normal VR-capable PC" we can support up to 16 users the non-tethered mobile HMD (i.e. Oculus Quest) can only support a maximum of 10 simultaneous users (2D+Chroma rendered sprites). With advances in technology and the next generation of devices (e.g. Oculus Quest 2) and with a better integration of WebXR functionality in the Oculus Browser software we can expect higher performance values and full support of 3D volumetric room and user representations.

5.3.3 I2CAT-3.3: PC-MCU

The presented PC-MCU is a virtualized cloud-based component applicable to any distributed multi-user Virtual/Augmented/Mixed/eXtended Reality (VR/AR/MR/XR) service, which aims at alleviating the demands at the client side, by minimizing the computational resources and bandwidth consumption. A set of innovative and coordinated features have been proposed and implemented to achieve this:

- Multiple Volumetric Video de/coding, compatible with most common Point Cloud compression strategies.
- Reception and delivery of multiple MPEG Dynamic Adaptive Streaming over HTTP (DASH) streams.
- Level of Detail (LOD) adjustment: the incoming Point Cloud representations can be down-sampled, providing the most appropriate resolution based on the users' relative position, activity and underlying context.
- Removal of non-visible volumetric video: the non-visible parts of the volumetric environment can be removed from the specific stream delivered to each user.
- Fusion of volumetric videos: the incoming volumetric videos are decoded, appropriated processed and fused as a single volumetric video for the scene, which can be delivered as a single (personalized) stream to the client devices.

This set of features allows providing an optimized stream for each involved user, depending on their position, viewpoint and available resources, thus alleviating the requirements at the client side, and ensuring a smooth experience.

The PC-MCU and its features have been evaluated in a realistic scenario with two remote virtual users in order to get initial evidence on its potential benefits, when compared to the same scenario using a peer-to-peer communication, as baseline. The obtained results show a significant reduction in terms of computational resources (RAM, CPU, GPU) and bandwidth consumption, thus proving its benefits and encouraging further research on this area.

The contributions of this experiment can provide relevant societal and economic benefits to our society, by enabling hyper-realistic virtual meetings using inexpensive hardware, while overcoming spatial barriers and travel requirements, and minimizing the environmental burden.

The experimental assessment presented in this deliverable is based on the use of a first implementation of the proposed PC-MCU, comparing it to a system in which the volumetric videos are delivered in a peer-to-peer fashion. The goal is to assess the benefits of using the PC-MCU in a scenario with two volumetric videos compared to a baseline scenario without the use of the PC-MCU.

5.3.3.1 Experimental Setup

The setup used in this experiment considers a set of 10 simulated holoconferencing sessions where one of three end users receives the Point Clouds of two other users remotely connected (although the networking aspects are out-of-scope of the paper). The sequences considered in the tests are two among the ones available at the 8i Voxelized Full Bodies database[6]: *Red and Black* and *Longdress* (see also 5.3.1). In order to facilitate a real-time processing of the whole pipeline, the resolution of the two sequences have been downsampled to 65k points and 78k points, respectively. In order to recreate a typical holoconferencing system, a set of actions are forced to activate and assess the benefits of the PC-MCU features. The duration of each single session was 34 seconds. The forced actions in the holoconferencing scenario are the following:

- Step 1: initial position purposely defined to receive the two Point Clouds at their maximum resolution (65k and 78k points) within the viewport.
- Step 2: viewpoint panning, excluding *Red and Black* from the viewport to evaluate the computational load reduction when the *Non Visible Area Removal* feature is active.
- Step 3: viewpoint panning, excluding *Longdress* from the Viewport to evaluate the computational load reduction when the *Non Visible Area Removal* function is active for this second simulated user.
- Step 4: user's viewpoint back to the initial position.
- Step 5: simulation of user moving away from both Point Clouds to evaluate the benefits when the *LoD Selection* feature is active.

The PC-MCU Fusion function, in charge of merging the sequences of several users into one, is always active. For the comparison, the same sequence of actions is simulated also when the volumetric videos are delivered in a peer-to-peer fashion, without the PC-MCU. All the features introduced above have been implemented according to a CPU oriented sequential programming model. At this stage, the implementation follows a sequential calls composition without any parallelization technique, nor GPU implementation. The GPU involved in this study is the one used at the end-user client machine, for the volumetric video rendering. The specifications of the

[6] Eugene d'Eon, Bob Harrison, Taos Myers, and Phil A. Chou. 2017. 8i voxelized full bodies-a voxelized point cloud dataset. ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document WG11M40059/WG1M74006 (2017).

machine used as the client are the following:

- CPU: Intel Core i7-6700 @ 3.40GHz
- RAM: 16 GB @ 2133MHz
- GPU: NVIDIA Quadro K4200 4GB GDDR5
- NET: Realtek PCIe GbE Family Controller

5.3.3.2 Evaluation Methodology and Metrics

The tests consisted of 10 iterations with the PC-MCU plus 10 iterations without the PC-MCU. In each iteration, samples of the following set of metrics were collected by using a tool from Montagud *et al.*[7]:

- CPU usage at the client machine (in %).
- GPU usage at the client machine (in %).
- Memory usage at the client machine (in MB).
- Bandwidth consumption (in Mbps).

5.3.3.3 Results

In order to provide a thorough comparison, we have performed a simulation of 10 sessions without the PC-MCU and 10 sessions using the PC-MCU. During each session, the mentioned metrics have been sampled along the whole experience. The final results show the average of each sample over the 10 iterations, showing the different benefits of the PC-MCU. Table 43 shows the reduction of exploiting the PC-MCU on CPU usage when the PC-MCU is included in the holoconferencing session, when compared to the baseline condition (without the PC-MCU). The intervals in which the PC-MCU features are active are specified on the top part of the figure. When the PC-MCU is not used, the percentage of CPU usage does not suffer strong fluctuations along the duration of the 10 sessions, due to the constant incoming stream from the two Point Clouds. When the PC-MCU is used, it is possible to notice that, initially, when only the Fusion feature is active, there is already a considerable gain in terms of CPU usage. The gain is considerable when, afterwards, the PC-MCU performs the Non Visible Areas Removal actions, first excluding the Longdress sequence, and then excluding Rad and Black. The CPU usage increases again when the 2 Point Clouds are newly available (see sample 15 in **Figure 63.a**) and then starts being furtherly reduced when the LoD Selection function is active. **Figure 63.b** shows the evolution for the percentage of GPU usage. In this case, the benefits of the Non Visible Areas Removal function are less noteworthy than for the CPU usage, because the GPU is in charge of the rendering of the visible part of the 3D scenario; however, when the LoD Selection function is active, it is possible to notice a gain. The GPU load is indeed reduced thanks to the lower amount of voxels needed to represent the Point Clouds. The overall average results for the 10 iterations are summarized in Table 43, by also including additional ones regarding the RAM and bandwidth consumption. It is possible to observe how the introduction of the PC-MCU resulted in a reduction of 69% of the CPU usage, 7% of the GPU usage, 18% of memory usage and of 84% of bandwidth consumption. For completeness, the additional latency introduced by the PC-MCU has been also evaluated. In average, the PC-

-	No PC-MCU	PC-MCU	Δ Reduction
CPU (%)	25	7.8	68.7%
GPU (%)	27.2	25.4	6.6%
RAM (MB)	445.5	366.3	17.85%
BW (Mbps)	92.3	14.9	83.9%

Table 43: Average results in terms of resources consumption

[7] M. Montagud, J.A. De Rus, R. Fayos-Jordan, M. Garcia-Pineda, and J. Segura- Garcia. 2020. Open-Source Software Tools for Measuring Resources Consumption and DASH Metrics. In Proceedings of the 11th ACM Multimedia System Conference.

MCU adds a latency of 89 ms when the Point Clouds are both at their lowest resolution and an average of 160 ms when they are both visible at the maximum resolution. For the intermediate cases the latency is kept in between those values. A demo showing the virtual scenarios, the set of simulated actions, and the registered metrics can be watched here: https://youtu.be/qEENaFVeLrk.

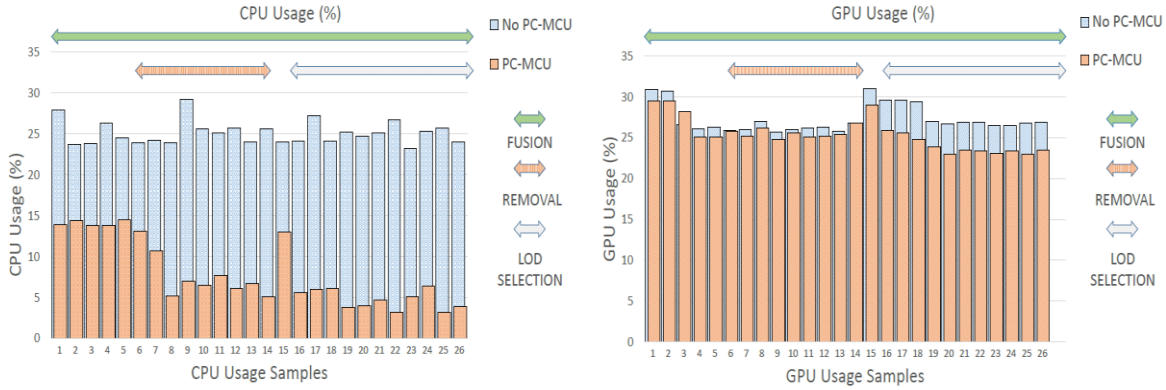


Figure 63: Percentage of CPU and GPU usage: comparison between sessions with and without the PC-MCU.

(a) CPU usage *(b) GPU usage*

5.3.4 CERTH-3.3: Transmission rate TVM comparison

5.3.4.1 Objective

The purpose of these experiments is to compare the transmission's frame rate achieved given a configuration regarding the resolution and the compression of the multi-view RGB colour data (textures), as well as the voxel grid dimensions (geometry) for the reconstruction of a TVM. The expected results are going to show the trade-off between performance and visual quality. Furthermore, the comparison between the different parametrization sessions is going to guide us to a conclusion regarding the best combination, with respect to the application's target frame rate (real time), while achieving the highest possible quality.

5.3.4.2 Methodology

Two capturing nodes were set up inside CERTH's laboratory, using a set of four sensors respectively, one with Kinect 4 Azure and the other with Intel RealSense 2.0 D415. In order to make the comparison as fair as possible while focusing on the different parameters affecting the transmission and, as a result, the frame rate, a set of calibration boxes, working as a static scene, was used as a subject for the reconstruction. As far as the parametrization of the TVMs goes, the following were included:

- RGB data resolution with HD and HD/2 corresponding to 1280x720x3 and 640x360x3 color image resolution respectively
- JPEG compression rate with the values of 20 corresponding to a high compression rate and 80 to a low one
- Voxel-Grid size with values 32x64x32, 64x128x64 and 128x256x128 resulting to a respective increase to geometry 's resolution as well as to the average number of vertices per TVM

In this final version of the ***VolReco*** software, as mentioned earlier, only the Draco compression library is used, while, for VolCap, Corto compression library is deployed. Every experiment was conducted using the VRTogether application, with the Pilot0 VR scenario, while metrics were captured for almost 5 minutes per session. At last, the machine used to run the VRTogether platform had the following specifics:

CPU: Intel(R) Core(TM) i7-8700K @ 3.70GHz
GPU: Nvidia GeForce GTX 1070
RAM: 32 GB

5.3.4.3 Metrics

At this point, we present the list of transmission related metrics computed:

- Frames per second as an average of the frames received over a time window of one second
- Missed/Skipped frames per second as a result of other frames being rendered
- Average frametime per TVM (ms) along with its standard deviation
- Frames per second as a fraction of the number of TVMs received and their total frame time
- Average end-to-end delay between the timestamp of a reconstructed TVM by the Volumetric Reconstruction tool and the one captured when the TVM is received by the VRTogether's application along with its standard deviation
- Average number of vertices per TVM
- Average received compressed buffer's size and its standard deviation
- Average decompressed buffer's size and its standard deviation
- Average decompression and deserialization function's execution time and its standard deviation
- Average message rate in and out of the RabbitMQ exchange assign to the TVM's data along with their standard deviation
- Average receiving rate between the RabbitMQ server and the machine running VRTogether's application along with its standard deviation

5.3.4.4 Results

Below, we present the results of these sessions in the form of tables. All those metrics were extracting via the VRTogether platform.

Cam type	Colour resolution	Voxel grid size	Jpeg rate (compression)	Frames per second	Skipped frames / second	Average frame time (ms)	StD of frame time (ms)	Frames / second (frame time)
K4A	HD	32x64x32	80	3.201	1.3	178.938	11.271	5.589
K4A	HD	64x128x64	80	2.705	1.5	204.516	10.561	4.896
K4A	HD	128x256x128	80	2.202	1.1	232.182	17.874	4.307
K4A	HD/2	32x64x32	80	14.905	10.7	46.96	5.359	21.295
K4A	HD/2	64x128x64	80	13.209	10.1	52.523	5.401	19.039
K4A	HD/2	128x256x128	80	7.807	5.5	82.256	12.137	12.157
K4A	HD/2	32x64x32	20	18.346	9.9	34.41	6.185	29.061
K4A	HD/2	64x128x64	20	15.286	9.3	43.474	5.253	23.002
K4A	HD/2	128x256x128	20	9.32	3.9	67.151	11.308	14.891

Cam type	Colour resolution	Voxel grid size	Jpeg rate (compression)	Frames per second	Skipped frames	Average frametime (ms)	StD of frametime (ms)	Frames / second (frametime)
D415	HD	32x64x32	80	3.496	2.1	179.68571	7.641	5.565
D415	HD	64x128x64	80	3.127	2.2	195.6	14.870	5.112
D415	HD	128x256x128	80	2.806	2.6	267.71429	15.479	3.735
D415	HD/2	32x64x32	80	9.6	0	63.530303	4.631	15.74
D415	HD/2	64x128x64	80	6.804	0	79.18	7.574	13.682
D415	HD/2	128x256x128	80	4.907	2.3	119.77551	10.320	8.348
D415	HD/2	32x64x32	20	16.0136	9.1	41.717	5.772	23.828
D415	HD/2	64x128x64	20	13.711	9	50.905	5.393	19.644

D415	HD/2	128x256x128	20	7.689	3	101.714	6.155	9.831

Cam type	Colour resolution	Voxel grid size	Jpeg rate (compression)	Average E2E delay (ms)	StD of E2E delay (ms)	Average number of vertices	Average decompression function execution (ms)	StD of decompression function execution (ms)
K4A	HD	32x64x32	80	245.31	25.162	6723	150.151	7.808
K4A	HD	64x128x64	80	253.25	30.451	26385	158.8	8.220
K4A	HD	128x256x128	80	265.77	25.478	109109	166.451	10.244
K4A	HD/2	32x64x32	80	111.087	28.683	6113	34.464	2.704
K4A	HD/2	64x128x64	80	131.545	21.895	25600.6	40.901	2.206
K4A	HD/2	128x256x128	80	181.244	21.156	111036	47.205	3.557
K4A	HD/2	32x64x32	20	105.983	15.703	6159.3	23.489	3.248
K4A	HD/2	64x128x64	20	127.217	15.503	25455.9	25.366	1.054
K4A	HD/2	128x256x128	20	173.592	24.341	105431	35.0002	3.609

Cam type	Colour resolution	Voxel grid size	Jpeg rate (compression)	E2E delay (ms)	StD of E2E delay (ms)	Average number of vertices	Average decompression function execution (ms)	StD of decompression function execution (ms)
D415	HD	32x64x32	80	245.71	12.863	10190.5	143.74	4.897
D415	HD	64x128x64	80	251.7	25.388	42734.8	147.432	6.683
D415	HD	128x256x128	80	255.6	28.135	191276	176.002	9.333
D415	HD/2	32x64x32	80	175.68	20.462	12273.5	40.720	1.596
D415	HD/2	64x128x64	80	201.51	17.251	43918	50.993	3.072
D415	HD/2	128x256x128	80	234.32	10.341	190363	62.076	4.32
D415	HD/2	32x64x32	20	106.28	19.138	10084	25.819	3.177
D415	HD/2	64x128x64	20	157.28	18.99	43096	35.571	1.751
D415	HD/2	128x256x128	20	192.52	16.032	189317	48.582	3.076

Cam type	Colour resolution	Voxel grid size	Jpeg rate (compression)	Average compressed buffer size (MB)	StD of compressed buffer size (MB)	Average decompressed buffer size (MB)	StD of decompressed buffer size (MB)
K4A	HD	32x64x32	80	1.011	0.0053	10.8	0.076
K4A	HD	64x128x64	80	1.094	0.0146	12.554	0.264
K4A	HD	128x256x128	80	1.21	0.0624	14.697	1.297
K4A	HD/2	32x64x32	80	0.148	0.0041	2.871	0.0634
K4A	HD/2	64x128x64	80	0.222	0.0135	4.612	0.241
K4A	HD/2	128x256x128	80	0.355	0.061	6.837	1.177
K4A	HD/2	32x64x32	20	0.063	0.0042	2.874	0.064
K4A	HD/2	64x128x64	20	0.146	0.014	3.903	0.253
K4A	HD/2	128x256x128	20	0.257	0.05	6.674	1.007

Cam type	Colour resolution	Voxel grid size	Jpeg rate (compression)	Average compressed buffer size (MB)	StD of compressed buffer size (MB)	Average decompressed buffer size (MB)	StD of decompressed buffer size (MB)
D415	HD	32x64x32	80	0.517	0.0024	10.947	0.0157
D415	HD	64x128x64	80	0.813	0.01	13.598	0.082
D415	HD	128x256x128	80	1.018	0.04	17.852	0.282
D415	HD/2	32x64x32	80	0.287	0.01	4.273	0.0819
D415	HD/2	64x128x64	80	0.432	0.042	7.89	0.333
D415	HD/2	128x256x128	80	0.695	0.036	9.8	0.783
D415	HD/2	32x64x32	20	0.085	0.0024	3.027	0.0173
D415	HD/2	64x128x64	20	0.282	0.01	6.093	0.075
D415	HD/2	128x256x128	20	0.584	0.04	9.877	0.314

Cam type	Colour resolution	Voxel grid size	Jpeg rate (compression)	Average message rate in of RMQ exchange (MB)	StD of message rate in of RMQ exchange (MB)	Average message rate out of RMQ exchange (MB)	StD of message rate out of RMQ exchange (MB)
K4A	HD	32x64x32	80	10.025	0.446	9.85	0.277
K4A	HD	64x128x64	80	6.577	1.282	6.533	1.228
K4A	HD	128x256x128	80	4.733	0.316	4.733	0.8
K4A	HD/2	32x64x32	80	26.711	0.843	26.755	1.112
K4A	HD/2	64x128x64	80	21.533	0.583	21.244	0.638
K4A	HD/2	128x256x128	80	13.711	0.388	13.777	0.595
K4A	HD/2	32x64x32	20	27.955	1.24	27.866	1.166
K4A	HD/2	64x128x64	20	23.466	0.632	23.266	2.124
K4A	HD/2	128x256x128	20	13.533	0.509	13.533	0.632

Cam type	Colour resolution	Voxel grid size	Jpeg rate (compression)	Average message rate in of RMQ exchange (MB)	StD of message rate in of RMQ exchange (MB)	Average message rate out of RMQ exchange (MB)	StD of message rate out of RMQ exchange (MB)
D415	HD	32x64x32	80	23.688	1.634	23.155	1.981
D415	HD	64x128x64	80	17.3	1.175	6.725	7.349
D415	HD	128x256x128	80	8.875	0.103	8.5	0.595
D415	HD/2	32x64x32	80	6.46	0.298	6.44	0.206
D415	HD/2	64x128x64	80	6.266	0.1	6.266	0.1
D415	HD/2	128x256x128	80	8.2	0.2	8.288	0.105
D415	HD/2	32x64x32	20	27.155	0.909	27	0.714
D415	HD/2	64x128x64	20	24.2	0	24.133	0.509
D415	HD/2	128x256x128	20	8.111	0.105	8.088	0.105

Cam type	Colour resolution	Voxel grid size	Jpeg rate (compression)	Average receiving rate from the RMQ (MB)	StD of receiving rate from the RMQ (MB)
K4A	HD	32x64x32	80	15.276	0.28
K4A	HD	64x128x64	80	9.223	3.724
K4A	HD	128x256x128	80	12.559	1.229
K4A	HD/2	32x64x32	80	12.875	0.092
K4A	HD/2	64x128x64	80	12.819	0.173
K4A	HD/2	128x256x128	80	12.9211	0.045
K4A	HD/2	32x64x32	20	10.149	0.52
K4A	HD/2	64x128x64	20	10.418	0.0244
K4A	HD/2	128x256x128	20	9.879	0.2654

Cam type	Colour resolution	Voxel grid size	Jpeg rate (compression)	Average receiving rate from the RMQ exchange (MB)	StD of receiving rate from the RMQ exchange (MB)
D415	HD	32x64x32	80	21.798	0.947
D415	HD	64x128x64	80	20.672	1.796
D415	HD	128x256x128	80	13.687	6.957
D415	HD/2	32x64x32	80	12.844	0.0029
D415	HD/2	64x128x64	80	12.904	0.056
D415	HD/2	128x256x128	80	12.799	0.0289
D415	HD/2	32x64x32	20	9.658	0.00122
D415	HD/2	64x128x64	20	9.674	0.0058

D415	HD/2	128x256x128	20	9.722	0.0402

5.3.4.5 Conclusion

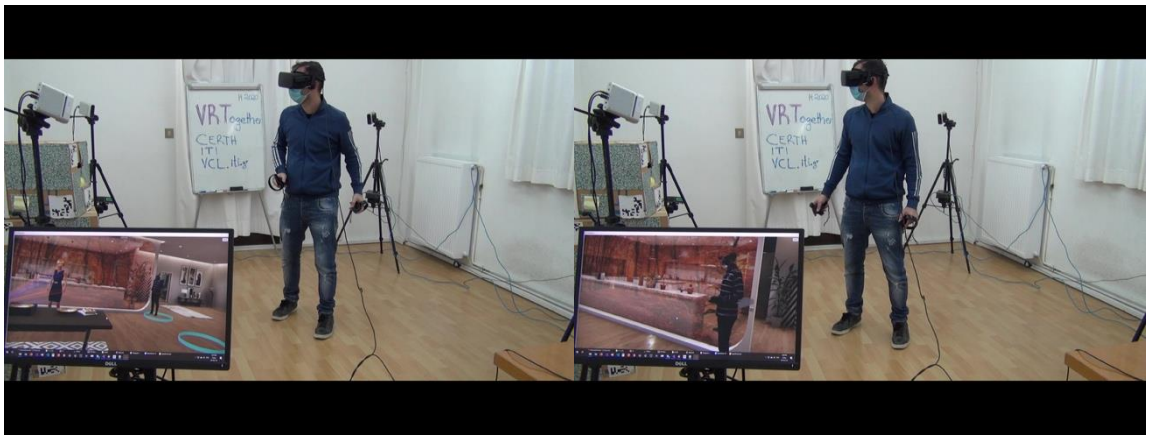
After examining the results, someone can easily observe the high effect that the colour image resolution has to the transmission and the frame rate. For both sensors, the HD quality seems forbidden for our application and its real-time requirements. Furthermore, the impact of the jpeg compression on the VolCap's side is also quite important both for Kinect 4 Azure and Intel RealSense sensors, as it reduces the amount of data transmitted, although resulting in an inferior texture quality for the TVMs in general. From the effect the change in voxel grid size has on the above results, it is understandable that the geometry's impact on the transmission is significantly lower, although the reduction on the voxel grid size boosts the performance, in exchange for TVMs with much lower number of vertices. Summarizing the above observations, we selected the values below for the parameterization of the TVMs, with respect to the overall trade-off between performance and quality:

- **Colour image resolution:** HD/2 (640x360x3)
- **Jpeg compression rate:** 70
- **Voxel grid size:** 64x128x64

Regarding the jpeg compression and the voxel grid size, the choice was made based mostly on the target quality, while, for the colour image resolution, performance was the decisive factor.

As far as the sensors go, it is obvious that the overall performance of the Kinect 4 Azure is superior that the one Intel RealSense has. This is mostly based on the fact that Intel RealSense RGBD sensors tend to track a larger number of points in the 3D space, as their precision is lower than Kinect 4 Azure sensors. This affects the quality, at first, but also the performance, as more data have to be transmitted for an inferior overall result. Therefore, we choose Kinect 4 Azure over Intel RealSense as far as the sensors go.

5.3.5 CERTH-3.4: Pre-Pilot Technology Test



5.3.5.1 Objective

The pilots are checkpoints to evaluate the creative and technical challenges of the project, aiming at:

- assessing the performance of the technological developments;
- validating and refining the defined evaluation methodology;

- assessing the appropriateness of the technology and created scenarios / content to provide truly realistic and interactive social VR experiences.

Pilot 1 was focused on a 2-participant scenario with stored content, and aimed at evaluating the potential of the contributions of the project to offer not only a feeling of being there (immersion), but also a feeling of being there together, intimacy and closeness (togetherness) between the participants. Pilot 2 continued the same storyline, but it was mostly focused on integrating more than two users and a live factor (the news presenter). The live factor also added an extra level of interaction, with the ability to talk with the live presenter and to control the presentation of media assets based on the dynamics of the conversation. In Pilot 3, the experience attempts to increase the sense of immersion and togetherness even more. While having a maximum of five users with a live representation and a variety of different types (3D avatar, 2D video, single and multi-camera point clouds, TVMs), users have also the ability to join a session as spectators or even with no representation at all. The Pilot 3 VR scenario resumes and concludes the same storyline, while taking part in a virtual apartment of great detail. It also includes three 3D avatars which interact with the users through voice instructions for specific actions using the Oculus Rift controllers and a simple form of dialog resulting in binary answer questions.

5.3.5.2 Methodology

We set up two capturing nodes inside CERTH's laboratory, which used two sets of four Kinect 4 Azure sensors. While following the protocol constructed by CWI and all the hygiene preparation required, we were able to conduct two two-user sessions using the Pilot3 VR scenario of VRTogether's application. Both users were represented as TVMs and were able to complete the scene's interactions through the Oculus Rift's controllers and its microphone. The parametrization of the TVMs was the following:

- RGB data resolution set to HD/2 (640x360x3 colour image dimensions)
- Jpeg compression parameter set to 70
- Voxel grid size set to 64x128x64

At last, the machines used to run the VRTogether platform had the following specifics:

- CPU: Intel(R) Core(TM) i7-8700K @ 3.70GHz
- GPU: Nvidia GeForce GTX 1070
- RAM: 32 GB

We were not able to proceed with a bigger number of participants, as a result of the local COVID19 restrictions.

5.3.5.3 Metrics

Besides the questionnaires which the participants filled before and after the sessions, as instructed by the used protocol, all the metrics integrated in VRTogether's platform were captured.

5.3.5.4 Results

Below, we present the results of these sessions in the form of a table. All those metrics were extracting via the VRTogether platform.

Frames per second (number of frames / time)	14.715	Average of connection_readers memory (MB) of RMQ server 's node	0.906
Missed / Skipped frames per second	9.8	Standard deviation of connection_readers memory (MB) of RMQ server 's node	0.0265
Average frametime per TVM (ms)	49.177	Average of connection_writers memory (MB) of RMQ server 's node	6.234
Standard deviation of average frametime per TVM (ms)	10.039	Standard deviation of connection_writers memory (MB) of RMQ server 's node	0.137
Frames per second (1 / Average frametime in seconds)	20.335	Average of connection_channels memory (MB) of RMQ server 's node	1.214
Average end to end delay (ms)	134.095	Standard deviation of connection_channels memory (MB) of RMQ server 's node	0.032
Standard deviation of end to end delay (ms)	25.002	Average of connection_other memory (MB) of RMQ server 's node	0
Average number of vertices per TVM	18576.966	Standard deviation of connection_other memory (MB) of RMQ server 's node	0
Standard deviation of number of vertices per TVM	3065.593	Average of queue_procs memory (MB) of RMQ server 's node	9.213
Average size of received compressed TVM (MB)	0.101	Standard deviation of queue_procs memory (MB) of RMQ server 's node	0.095
Standard deviation of received compressed TVM 's size (MB)	0.0085	Average of queue_slave_procs memory (MB) of RMQ server 's node	26.017
Average size of decompressed TVM (MB)	3.348	Standard deviation of queue_slave_procs memory (MB) of RMQ server 's node	0.235
Standard deviation of decompressed TVM 's size (MBs)	0.106	Average of plugins memory (MBs) of RMQ server 's node	0.748
Average total deserialization-decompression time per TVM (milliseconds)	38.358	Standard deviation of plugins memory (MB) of RMQ server 's node	0

Standard deviation of deserialization-decompression time per TVM (milliseconds)	8.889	Average of other_proc memory (MB) of RMQ server 's node	8.07
Average deserialization-decompression function execution time per TVM (milliseconds)	31.381	Standard deviation of other_proc memory (MB) of RMQ server 's node	0.0626
Standard deviation of deserialization-decompression function execution time per TVM (milliseconds)	8.13	Average of metrics memory (MB) of RMQ server 's node	0.192
Average marshalling time for the texture data per TVM (milliseconds)	2.429	Standard deviation of metrics memory (MB) of RMQ server 's node	0
Standard deviation of marshalling time for the texture data per TVM (milliseconds)	2.526	Average of mgmt_db memory (MB) of RMQ server 's node	2.338
Average marshalling time for the geometry data per TVM (milliseconds)	3.234	Standard deviation of mgmt_db memory (MB) of RMQ server 's node	0.000217
Standard deviation of marshalling time for the geometry data per TVM (milliseconds)	2.686	Average of mnesia memory (MB) of RMQ server 's node	13.065
Average marshalling time for the extra parameters per TVM (milliseconds)	1.293	Standard deviation of mnesia memory (MB) of RMQ server 's node	0.503
Standard deviation of marshalling time for the extra parameters per TVM (milliseconds)	1.293	Average of other_ets memory (MB) of RMQ server 's node	0.0285
Average rendering time per TVM (milliseconds)	4.238	Standard deviation of other_ets memory (MB) of RMQ server 's node	0
Standard deviation of rendering time per TVM (milliseconds)	2.005	Average of binary memory (MB) of RMQ server 's node	27.138
Average CPU % consumption	31.311	Standard deviation of binary memory (MB) of RMQ server 's node	0
Standard deviation of CPU % consumption	4.256	Average of msg_index memory (MB) of RMQ server 's node	1.071

Average GPU % consumption	36.2	Standard deviation of msg_index memory (MB) of RMQ server 's node	0
Standard deviation of GPU % consumption	9.167	Average of code memory (MB) of RMQ server 's node	12.192
Average RAM consumption (MB)	1299.730	Standard deviation of code memory (MB) of RMQ server 's node	0.0063
Standard deviation of RAM consumption (MB)	3.289	Average of atom memory (MB) of RMQ server 's node	52.314
Average of Message rate in / second of RMQ exchange	25.622	Standard deviation of atom memory (MB) of RMQ server 's node	1.496
Standard deviation of Message rate in / second of RMQ exchange	0.972	Average of other_system memory (MB) of RMQ server 's node	0.778
Average of Message rate out / second of RMQ exchange	50.8	Standard deviation of other_system memory (MB) of RMQ server 's node	1.301
Standard deviation of Message rate out / second of RMQexchange	1.897		
Average of Receiving rate (MB / second) of RMQ connection	10.849		
Standard deviation of Receiving rate (MB / second) of RMQ connection	0.037		

5.3.5.5 Conclusion

Regarding the gathered metrics, as expected, they are similar to the ones acquired during the experiments. We would mostly like to underline the impact that texture (colour) data can have on the transmission performance. As one can observe, textures overhead even the process of marshalling, requiring more time compared to the volumetric data. The deserialization-decompression function affects the frame rate consuming the most of the processing time needed for a TVM to be rendered, starting from the time data were received. Another point is the comparison between the averages of message rate in and out, which indicate the presence of two users during the session.

As far as the user's experience goes, all 4 participants expressed great interest in the platform. The immersion along with the interaction in the virtual environment resulted in a great experience. All and all, given the measures we gathered above, VRTogether constitutes one of the few real-time VR platforms that allow real-time, mesh-based, photorealistic volumetric video transmission.

6 PILOT 3

This section reports the work on Pilot 3, including the content creation, the deployment and evaluation, and the analysis of the results. It concludes with a report on how the technology has been showcased in several professional events, maximizing the impact.

6.1 Scenario and Content Creation

This is the last episode of the three pilots, where a thriller is told as a means to offer a unique and innovative virtual reality experience. The story behind the three episodes covers the investigation of the murder of Ms. Armova and all the details around it. This last episode has many innovative aspects in terms of content, from the real representation of the users to the interaction in them, all integrated in a scenario that seems real and that makes us want to participate with other users and share experiences. The storytelling was a risk and a gamble, and it shows the importance of adapting the new script techniques to this type of immersive experiences. Further information about the content creation can be found in D4.5 and the final version of all the content in D4.7.

The 4 characters:

- Sarge Hoffsteler: A police inspector, aged, highly professional and strict rigor. His training and experience in murder cases has made him an excellent profiler. He is able to know in a glance the weaknesses of the suspects and how to respond to the questions to make mistakes and reach the truth.
- Elena Armova: The victim. She was extremely rich and living in a constant opulent lifestyle. Her friends were close to her because of her social position. Elena appears in the form of an interactive hologram as part of an Artificial Intelligence (AI) software program, that is why she has an active role in the pilot.
- Rachel Tyrell: Policewoman working in the same department as Sarge. She helps him during the investigation of the murder.
- Evans: Forensic technician that helps Sarge in the investigation of the murder. He wears a white suit and gloves to protect found evidence.

For the generation of the characters we followed three steps.

- The first one to record the acting of the real actors.
- The second one to capture movements of the face using an iPhone and Reallusion's LIVE FACE app. This all combines nicely into the Character Creator ecosystem we use for our characters.
- The third step was a body mocap during post production.

The participants follow Sarge and his team to find clues regarding the murder of the victim. This configuration allows users to be part of the experience, being able to make decisions regarding the story. They need their cooperation in terms of trying to understand what could happen between Elena Armova and the suspects. Just like in Pilot 1 and 2, the experience presented by the VRTogether project allows us to experience the sensation of togetherness in a virtual reality environment. One step beyond, users will be allowed to interact with objects, unlike the previous pilots. In total, Pilot 3 has 18 animation controllers, 62 animation clips, 22 timelines and 31 synchronized triggers to drive a multiplayer experience with 4 characters in 3 rooms of the apartment over 4 scenes for a duration of about 10 minutes.

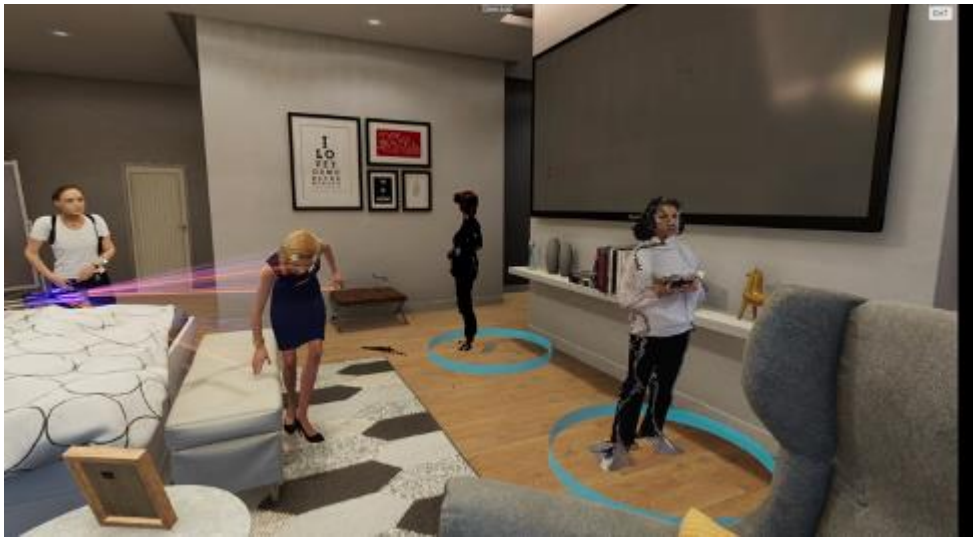


Figure 64. The Bedroom of Pilot 3.

The main goal is to get the best 3D generated content possible, achieving the best quality possible and, therefore, taking care of the rendering process. We have 3 different spaces: living room, kitchen and bedroom, all three of them in full CGI and optimized for Unity. The work included the generation of the 3D assets. We worked on the lighting, an important element for hyper-realistic results, so we added different emission points of a natural lighting direction coming from the windows, and other emission points of light like the lamps. It's very important for hyper-realism to work hard and to focus on highlighting each element, and the lighting structure and distribution was altered to give the room a more natural aspect and quality look. Years of experience in the field of VFX for feature films allowed us to test with textures and colours that give us this treatment in a more effective way. The lighting and texturing process is long and intense, as it requires many hours of work. The scenario layout design included representation of the users, locating as well the presence of the characters in every room. For these reasons, it is essential to break down the script, the list of actions, interactions, and therefore the need for rehearsals with the actors. These actions allow adaptations of the script, always respecting the original. This is a delicate and essential process to obtain high quality and consistent content, which also allows the development of all the elements to provide a VR experience with the best quality possible.

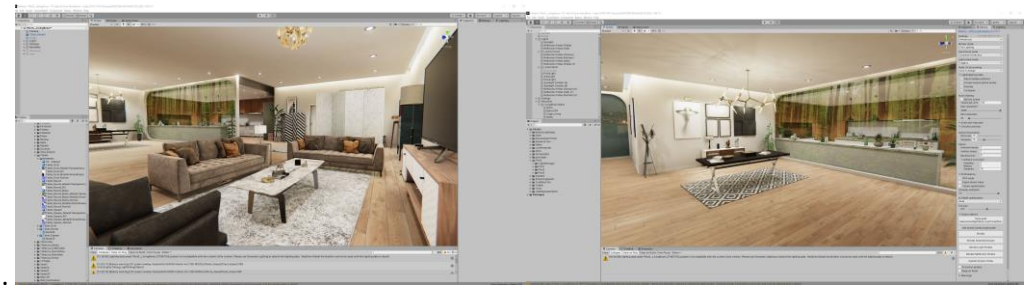


Figure 65. Living-room Scenario Generation for Pilot 3.

The visual effects should enhance the hyperrealism of the CGI and improve the appearance of the characters and their animations. The VFX needed in this case can be grouped in the following categories: lighting, shading and other VFX with the goal of providing a better hyper-realistic definition of characters and the environment. Also, the postproduction sound: sound design, sound effects, sound editing. This section summarizes the actions taken regarding the elaboration of VFX and animations, including pictures to show the process and/or the final look of some effects. Creating these effects is a very important process to give the scene and the characters a more

realistic and defined look. It also helps the viewer to get inside the environment of high-quality photorealistic content.

Figure 66. Creating Volumes for Pilot 3.

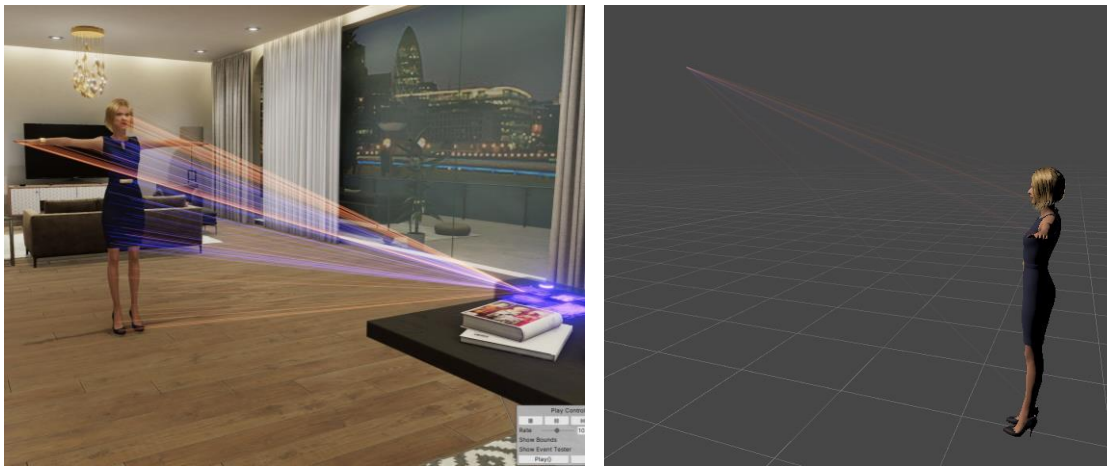


Figure 67. Hologram of Elena Armova and apparition effect.

6.2 Technology and Setup

The Pilot 3 experiments were conducted in 4 labs at the CWI premises in Amsterdam. All the labs were equipped with workstation PCs (GPU: Nvidia GeForce 1080Ti, CPU: Intel i7-9700k, Memory: 32GB DD4 or better). In addition, all 4 labs had a point cloud capture setup with three Azure Kinect sensors. The captured point clouds were used to provide a self-view and were also encoded and sent to all other users sharing the experience. The participants in two labs had an Oculus Rift HMD, these users were able to walk around in a limited space and could make unrestricted head movements to move their viewport. These users could also teleport to unoccupied player spawn locations and interact with some objects in the scene using the Oculus controllers. The participants in the remaining two labs were watching the experience on 55" TV screens and were able to move around without restrictions using a Logitech F710 wireless controller although they could not interact with objects in the scene. All participants had microphones and headphones that they could use to talk to each other during the experience.

Figure 68. HMD (left) and non-HMD (right) user labs used for Pilot 3.

All sessions were run in the Pilot 3 scenario with Socket.IO communication for audio and point clouds. In addition, a fifth spectator user was used to create the sessions and record the experience with screen capture.

6.3 Evaluation with Users

In this section, we present the methods and results of the Pilot 3 experiments with 48 users. The goal of the Pilot 3 experiment is to evaluate the VRTogether platform and the Pilot 3 content. In the experiment, we measured and compared the experiences of users using the Head Mounted Display (HMD) with the experiences of the users using the desktop version of the VRTogether platform. We aim to answer the following research questions through the Pilot 3 experiments:

(1) Research questions about subjective data:

- What are the differences in user experience (e.g., quality of interaction, presence/immersion, social connectedness) of the users (HMD vs. Desktop)?
- How do the users rate their own representation compared to those of other users?
- What are users' ratings and opinions towards their photorealistic representations and the Pilot 3 content?
- How do users perceive the quality of the movie characters?

(2) Research questions about objective data:

- How does user behaviour evolve over time?
- How is user behaviour affected by the device (HMD versus Desktop, different controllers)
- (If we log latency etc) how does user behaviour correlate with quality of service?

6.3.1 Methodology

The Pilot 3 experiments with users were conducted in four labs: Lab A, B, C, D at CWI, located in the Science Park, Amsterdam. Two Labs (A, B) were equipped with the Oculus Rifts HMDs, and the other two (C, D) with desktop computers and game controllers (see Figure 69). Each lab has three depth cameras to capture, render and deliver users' volumetric representations to the virtual environment. We invited a total of 48 users, and conducted 12 sessions of the experiment. The users came to the experiment in groups of four people. Two of them were assigned to use the HMD and the other two to use the desktop.

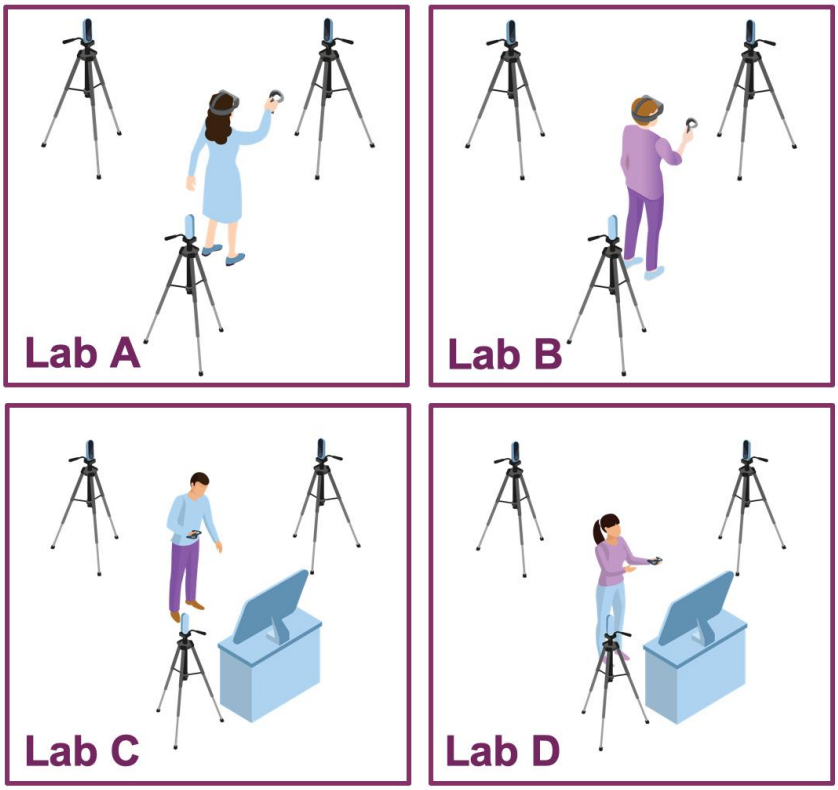


Figure 69. Visualization of the four lab setups for Pilot 3.

The virtual environment of the Pilot 3 experience was a virtual crime scene (the home of the victim - Elena Armova) in which the point cloud representations of participants were placed in four distinct positions (Figure 72). The four participants were integrated into the experience as members of Citizen Oversight Committee as they enter the apartment of the victim to witness a murder investigation lead by pre-recorded virtual characters: the detective Sarge, two assistants Rachel and Evans, and the hologram of the victim Elena Armova.

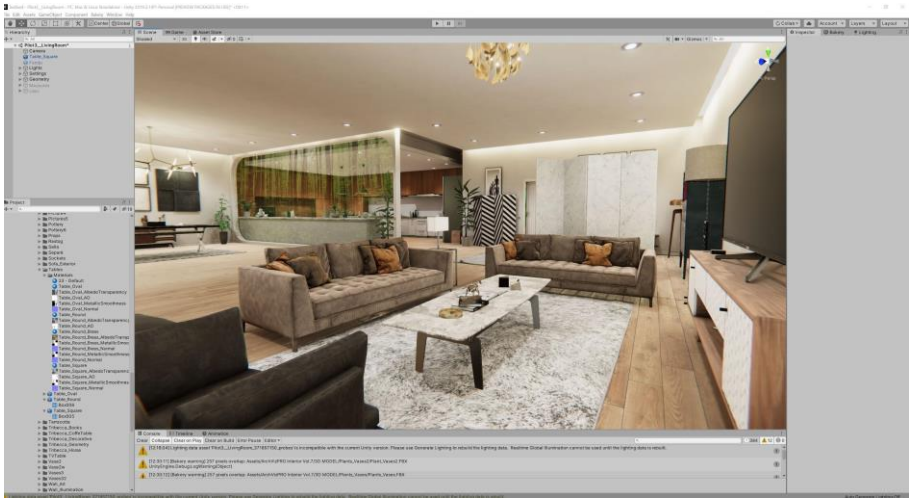


Figure 70. The crime scene (the virtual living room of Elena Armova) in Pilot 3.

Figure 71. The four pre-recorded movie characters, from left to right: Sarge, Evans, Elena and Rachel in Pilot 3.

Figure 72. User positions in the beginning of the experience of Pilot 3 and the placement of the interactive element of the environment: User 1(HMD), Position 1 next to the light switch; User 2 (HMD), Position 2, next to the phone finder); User 3 (Desktop), Position 3; User 4 (Desktop), Position 4.

For each session, we assign position numbers to each user to distinguish their roles and the devices they used. Since the two HMD users had to perform some tasks (i.e., switch on the light, click on the phone finder), the two HMD users were always User 1 and 2 in the experience. User 1 is assigned to Position 1 next to the light switch. User 2 is assigned to the Position 2 next to the phone finder. User 3 and 4 joined the virtual apartment between User 1 and User 2.

6.3.2 Participants

We recruited 48 users/participants (23 males, 25 females), who aged between 21-56 (Mean=34.9, SD=10.3). 12 female and 12 males users used the HMD, 13 female and 11 male users used the Desktop. All the users can read, speak, write fluently in English, have no visual/hearing/motor impairment. The users who were long-/short-sighted were instructed to come to the experiment with their lenses or glasses. 13 of the participants had never used VR, 33 had used 1-3 times and 2 were experienced VR users.

6.3.3 Evaluation Procedure

Step 1 (~10 min). The experimenter welcomed participants and briefly explained the goal and procedure of the experiment including an introduction about the scenario and the characters. The

participants read and signed an experiment consent form, agreed that they understand the risks of VR technology and the HMD, and they have the right to quit the experiment at any time. Afterwards, they were asked to fill in two questionnaires:

- A Background Info Questionnaire including a colour blindness screening test and a vision acuity test
- A pre-study Simulation Sickness Questionnaire

Step 2 (~10 min). Experimenters trained the participants to use the HMD and desktop controllers within a training scenario.

Step 3 (~10 min). Each participant was taken to a separate lab room and the experience was launched.

The Pilot 3 experience can be summarized in the next steps:

- The participants are placed into the virtual environment and have a moment to familiarize themselves with the space.
- The scenario starts.
- Sarge asks User 1 to turn on the light.
- Rachel enters with the hologram of the victim.
- Sarge asks User 2 to click the phone finder.
- User 4 has a chance to answer Sarges question
- The users are split into two groups. Two users that are familiar with each other are put into the kitchen while two others into the bedroom.
- User 1 in the kitchen and User 2 in the bedroom both get a chance to answer a question.
- The users are reunited in the living room and have extra time to interact before the scenario starts again.
- The murderer is revealed the scenario ends.

Step 4 (~20 min). After the experiment, all four participants returned to the meeting room and fill in the following questionnaires:

- Post-study Simulator Sickness Questionnaire
- Social VR questionnaire for Part 1 (4 users), and Social VR questionnaire for Part 2 (2 users)
- Presence Questionnaire
- Visual Quality Questionnaire
- NASA-TLX (Task Load Index)

Step 5 (~10 min). The experimenter conducted a semi-structured interview with 4 users to understand their experiences. The interview is audio recorded. The experimenter thanked the participants for coming.

Overall, each experimental session takes between 60 and 90 minutes.

The objective data was analysed and gathered from:

- Record of the overall completion time of crime solving (all 4 users)

- Record of the completion time of each task for the users (e.g., Who switched on the button? How long did it take for the user to find the button and switch it on?)
- Joystick/controller data (keystrokes etc)
- Objective metrics of the downloaded point clouds for each user: timestamp of frames downloaded/decoded (which granularity can we achieve? Can we log it at 30Hz)
- Phantom user recording the scene in 360

6.3.4 Pilot 3 Results: Semi-Structured Interviews

A group interview was conducted after participants filled in the questionnaires. The goal of the interview was to understand the overall user experience, the comparison of the real-world experiences, and the visual quality of their photorealistic representations. It further gathered participants' recommendations for improving the social VR experiences. We label the group by the date and their starting time, for instance, 1123-14 stands for the group who participated in the Pilot 3 experiment on November 23, started at 14:00.

6.3.4.1 Quality of Experience (QoE)

We evaluated the general user experience of this system, as well as comparing the differences in user experience between HMD users and desktop users. For the general experience, most participants gave positive feedback for the novel ideas of photorealistic representation (*1124-14: "(With photorealistic representation) I feel more like I was talking with real people instead of virtual characters."*) and fine details of movie contents (*1123-14: "The details of movie characters were pretty good especially when they were moving their fingers."*). Communication and interaction between users was less active for the lack of triggers (i.e., collaborative assignments) and opportunities (i.e., buffer time with the movie for free talk) for social interaction. For the differences between HMD users and desktop users, we figured out two main insights. Firstly, we found that desktop users had better mobility than HMD users. Desktop users moved around freely to explore the VE and such free movement improve the user experience (*1125-11: It was nice to move everywhere and check the room.*), while HMD users reported the limitation in movement for the sake of teleportation function and limited walking space in VR labs (*1123-14: I was too stuck on my position.*). The second point was that the presence and immersion of HMD users were better than those of the desktop users. (*1125-11 desktop user: "I was jealous of HMD users, because they were actually there (the virtual scene)."*)

6.3.4.2 Representation quality

The concept of photorealistic representation enhanced the co-presence and emotional connection for remote social VR users. (*1127-11: "I can see her (another HMD USER) clothes and VR headsets, then I feel that it is definitely her!"*). Another interesting finding was that HMD users perceived representations of other uses better than the ones of themselves. Users paid more attention to the individual pixels of their own avatar (*1123-14: "I saw myself as a crowd of pixels."*), but perceived others' representation as a complete 3d image made up with point clouds (*1126-11: "I saw the colour he was wearing as well as his outfit"*).

6.3.4.3 New possibilities of use cases

Most users were positive on use of social VR for a wider range of use cases, such as entertainment, professional collaboration, medical training and virtual tour. One interesting insight was that people had diverse opinions on how social VR was applied in social interaction between strangers. Some people were optimistic about using VR as an opportunity to extend social contact and interact

with strangers (*1123-14: "I won't specifically meet strangers in real life, so it is meaningful to interact with them in VR."*) However, some users were still concerned with the security issues and invaded personal boundaries when meeting strangers in VR (*1126-15: "I would still feel that they (strangers) are in my personal spaces in VR (even though they are actually not in real life"*).

6.3.4.4 Recommendation

We concluded promising recommendations for improving this prototype based on the feedback of users. Firstly, multisensory interaction (i.e., haptic feedback) could be included for enriching user experience. Secondly, game mechanism and storytelling of the virtual movie could be more engaging, with more challenging collaborative tasks between users (i.e., solving a puzzle together), more interaction between users and virtual characters (i.e., more human-like dialogue with movie character), more self-exploration in the virtual environment (i.e., freely move and grab virtual objects in VE), as well as more impacts of users on the game stories (i.e., make choices to influence how the story is going on).

6.3.5 Pilot 3 Results: Questionnaires

6.3.5.1 The simulator sickness questionnaire (SSQ)

The simulator sickness questionnaire was analysed, where four representative scores can be calculated. Nausea-related subscore (N), Oculomotor-related subscore (O), Disorientation-related subscore (D) are the scores for the symptoms for the specific aspects. Total Score (TS) is the score representing the overall severity of cybersickness experienced by the users of virtual reality systems. We checked the normality of all the scores obtained before and after the Pilot 3 using Shapiro-Wilk normality test. None of them were normally distributed. We compared the scores of each category (i.e., N, O, D and TS) before and after the Pilot 3 experiences within each device group (i.e., HMD and Desktop), using Wilcoxon signed rank test with continuity correction. None of the categories shows significant differences. We also compared the changes of the TS before and after between HMD and Desktop users, again, no significant differences were found. Since Pilot 3 was a short (about 9 minutes) social VR experience, users reported no differences in terms of nausea, disorientation before and after the Pilot 3.

6.3.5.2 The social VR questionnaire

The social VR questionnaire has three factors, namely Quality of Interaction (QoI, Q1-Q11), Social Connectedness (SC, Q12-Q22), and Presence/Immersion (PI, Q23-Q32). We first checked the internal consistency of the questionnaire items by applying Cronbach's alpha on the 96 copies of filled questionnaires (two versions together). The questionnaire items were highly internal consistent. The cronbach's alpha scored 0.81, 0.74 and 0.79 for QoI, SC and PI, respectively.

For the first half of the Pilot 3, where 4 users were together in the virtual living room. No significant differences in QoI and SC were identified between the HMD and the Desktop users. For PI, both HMD scores and Desktop scores are normally distributed based on the Shapiro-Wilk normality test. We applied the two-sample T-Test, and identified a significant difference between the HMD and the Desktop users (M_{HMD} =38.33, SD_{HMD}=5.30; M_{Desk}=33.50, SD_{Desk}=6.28, p<.01, Cohen's *d*= 0.51). Similarly, for the second half of the Pilot 3, where 4 users were separated into two groups, we used again the two-sample T-Test with normality assumed data sets, and only identified a significant difference between the HMD and the Desktop users in terms of PI (M_{HMD} =40.67, SD_{HMD}=5.11; M_{Desk}=33.75, SD_{Desk}=6.11, p<.001, Cohen's *d* = 0.74). So, HMD users tended to feel more immersed in the virtual environment.

In both parts of the Pilot 3 (4 users versus 2 users), HMD users reported higher scores in Presence/Immersion experiences than the Desktop users (see Figure 73).

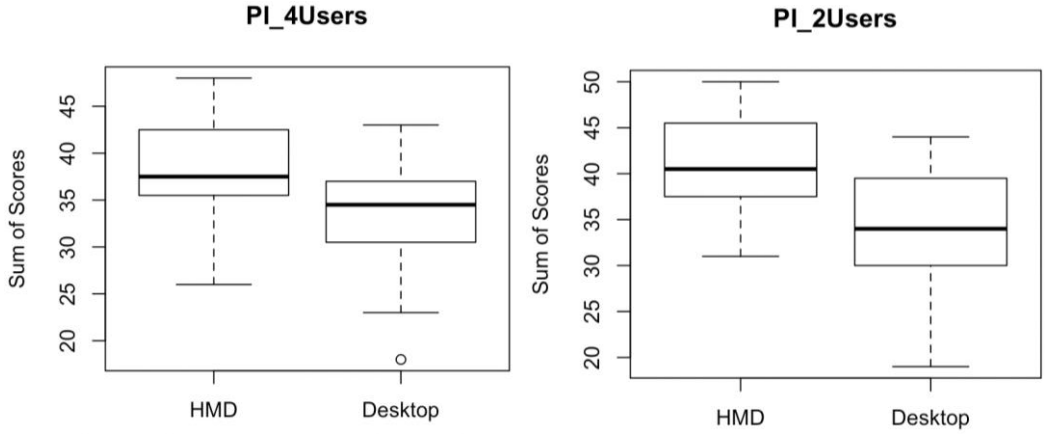


Figure 73. Pilot 3: Presence Immersion (P/I) Results.

6.3.5.3 The Presence Questionnaire

We removed the last two items of the presence questionnaire, as we did not include any haptics in the Pilot 3 social VR experiences. The 22 -item questionnaire was constructed as follows:

- Realism: Items 3 + 4 + 5 + 6 + 7 + 10 + 13
- Possibility to act: Items 1 + 2 + 8 + 9
- Quality of interface: Items (all reversed) 14 + 17 + 18
- Possibility to examine: Items 11 + 12 + 19
- Self-evaluation of performance: Items 15 + 16
- Sounds: Items 20 + 21 + 22

We compared the scores of the above-mentioned factors, only "Possibility to examine" exhibited a significant difference between the HMD users and the Desktop users (Figure 74). We first ran a Shapiro-Wilk normality test to confirm normality of the "possibility to examine" sub data sets, then we performed a two-sample t-test (M_{HMD} =11.92, SD_{HMD}=3.02; M_{Desk}=14.63, SD_{Desk}=3.16, p<.01, Cohen's *d*= 0.55). Desktop users had more possibilities and freedom to explore the virtual environment, because they could move freely in the scene. In contrast, the HMD users can only teleport between fixed positions (i.e., the blue circles in Figure 72).

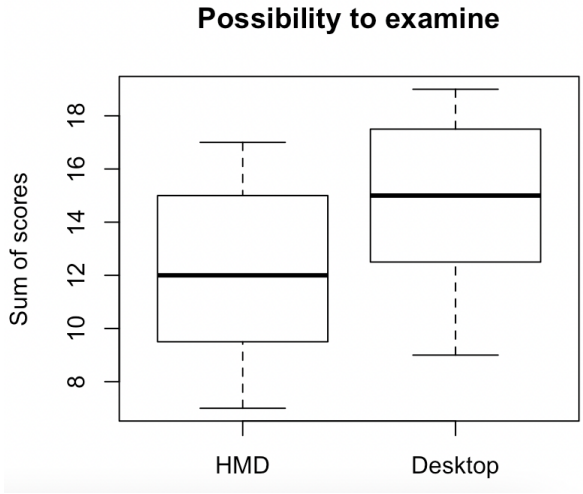


Figure 74. On the Presence Questionnaire, Desktop users reported to have significantly more possibilities to examine the virtual environment.

6.3.5.4 NASA Task Load Index (TLI)

To compare the TLI between HMD users and Desktop users, we first examined the normality of the data by using Shapiro-Wilk normality test. Since both data sets are normally distributed, we applied two-sample t-test and identified a significant difference between the HMD users and the Desktop users in terms of task load (M_{HMD} =20.02, SD_{HMD}=5.11; M_{Desk}=16.63, SD_{Desk}=3.12, p<.01, Cohen's *d*= 0.52). HMD users reported heavier task load compared to the desktop users (Figure 75).

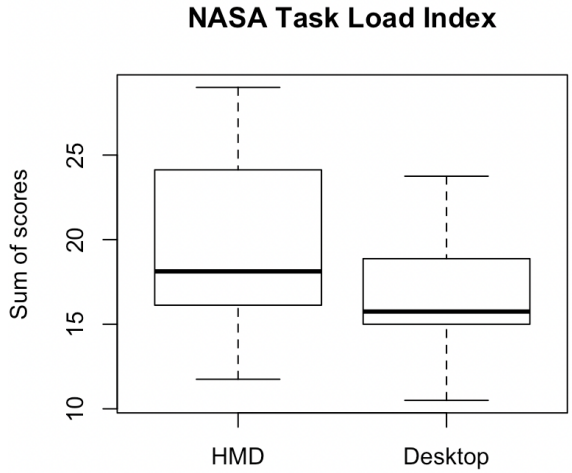


Figure 75. On the NASA TLI, the HMD users reported to have a significantly heavier task load than the Desktop users.

6.3.5.5 The Visual Quality Questionnaire

We first examined the self-visual quality ratings and the ratings of other users' representations of the HMD users. Since the data does not follow the normal distribution, we conducted Wilcoxon signed rank test, and found a significant difference between the ratings of the self-representation and those of the others' representations (M_{Self}=2.96, SD_{Self}=1.08; M_{Others}=3.58, SD_{Others}=0.51, p<.01, Cohen's d=0.57). The other comparisons, including the self-ratings and others ratings of the Desktop users, self-ratings between HMD and Desktop users, and others ratings between HMD and Desktop users, do not show significant differences.

For the quality of the virtual characters, both the HMD and the desktop users gave high ratings. There were also no significant differences found between the HMD users and the desktop users (M_{HMD} =4.50, SD_{HMD}=0.72; M_{Desk}=4.33, SD_{Desk}=0.96, p>.05).

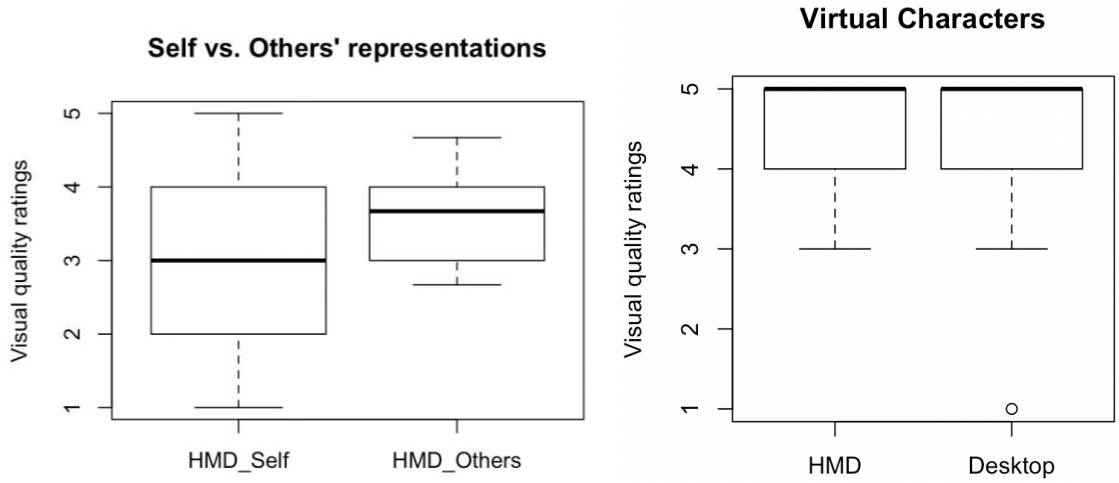


Figure 76. Pilot 3: Visual Quality Results.

Figure 76 shows the results on visual quality (scale: 1=poor, 2=bad, 3=fair, 4=good, 5=excellent). In the left, HMD users rated their own volumetric representations significantly worse than those of other users. In the right we can see that both the HMD and the Desktop users rated the visual quality of virtual characters as between good and excellent. No significant differences were found between the ratings of the HMD and Desktop users.

6.3.5.6 The Pilot 3 content

We have added three extra questions regarding the content of the Pilot 3 at the end of the Social VR Questionnaire, which asked participants to rate, on a 5-point likert scale (1) whether they like the virtual movie, (2) whether the virtual movie was realistic, and (3) whether the spatiality was consistent with the real world. No significant differences were found between the HMD and the Desktop users towards the ratings of the Pilot 3 content. The Pilot 3 content received decent scores on the three asked questions (Figure 77 and Table 44. The mean scores and standard deviations of the users' ratings of the Pilot 3 content (1=fully disagree, 5=fully agree).

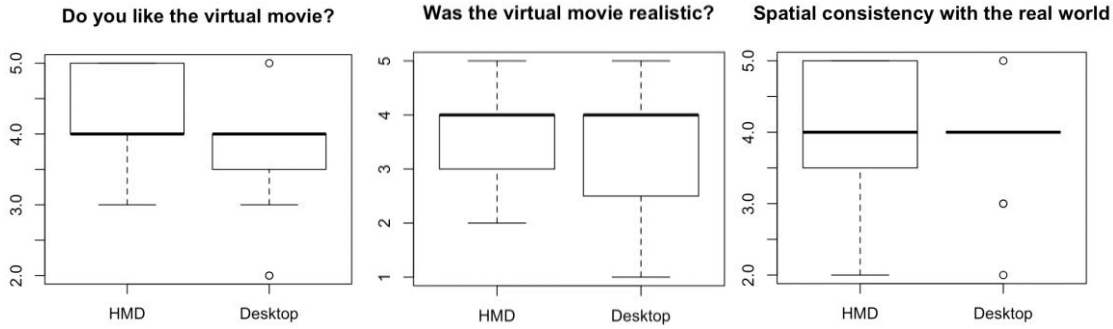


Figure 77. Users' ratings of the Pilot 3 content (1=fully disagree, 5=fully agree).

	Do you like the virtual movie?	Was the movie resembling the real world scenario?	Was the spatiality consistent with the real world?
HMD users	M=4.08, SD=0.71	M=3.63, SD=0.92	M=4.00, SD=0.93
Desktop users	M=3.71, SD=0.69	M=3.38, SD=1.21	M=3.88, SD=0.54

Table 44. The mean scores and standard deviations of the users' ratings of the Pilot 3 content (1=fully disagree, 5=fully agree).

6.3.6 Pilot 3 Results: Objective data

In the context of Pilot 3, the objective data described in sections 2.2 (objective metrics) and 2.3 (performance metrics) were gathered.

6.3.6.1 User movements in 3D space

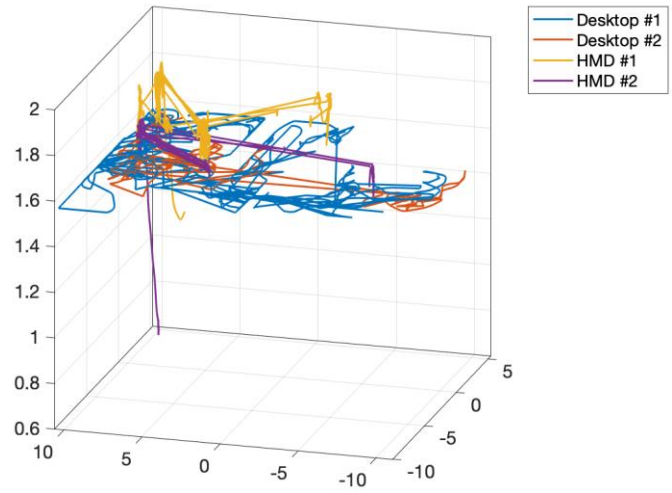


Figure 78. 3D view of head trajectories for a group of users.

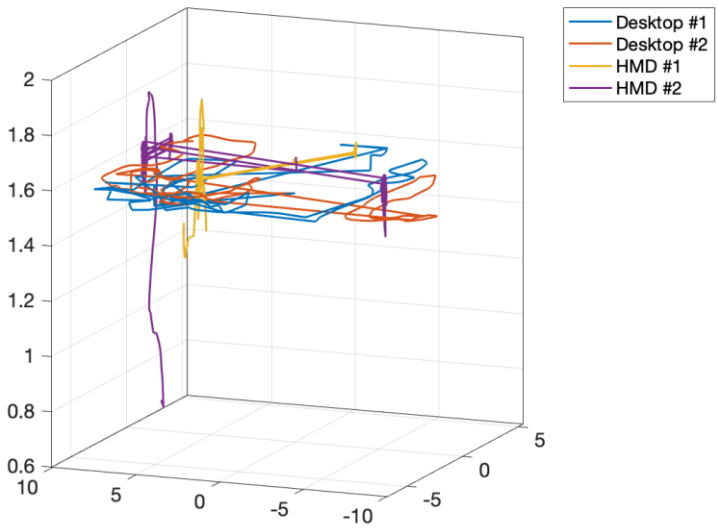


Figure 79 3D view of head trajectories for a group of users.

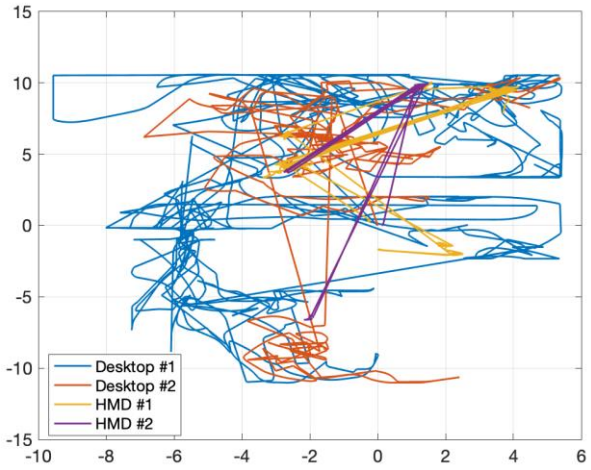


Figure 80 Floor view of head trajectories for one group of users.

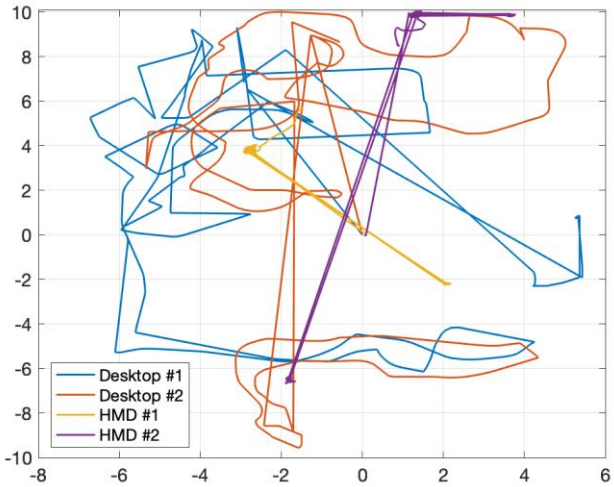


Figure 81 Floor view of head trajectories for one group of users.

Figure 78, Figure 79, Figure 80, and Figure 81 show the head trajectories of two representative groups, for the entire duration of the pilot. The first two represent 3D view, whereas the latter two show the floor movement. A main difference can be observed in exploration behaviour between desktop users and HMD users. This can be explained by the fact that HMD users were restricted to teleport in fixed locations in the scene, provided that they were not occupied by another user. As such, they were restricted regarding the locations that could be visited. Desktop users, on the other hand, could roam freely within the scene. Users were not obliged to explore the scene, which resulted in varying behaviour between different groups.

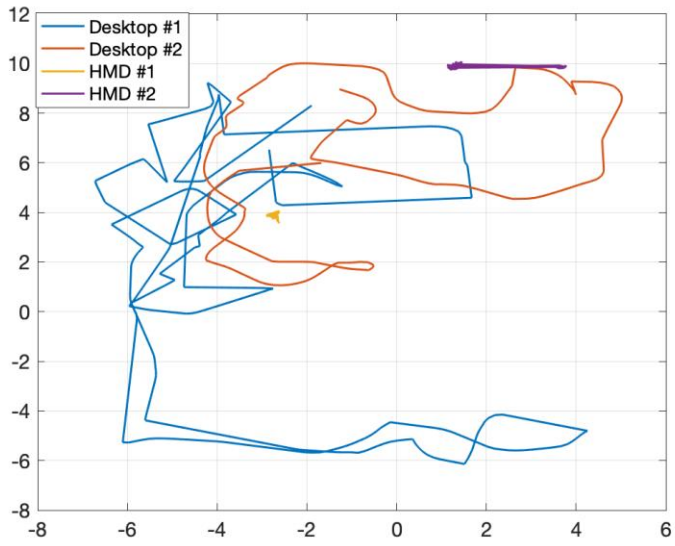


Figure 82 Floor view of user behaviour in scene #1.

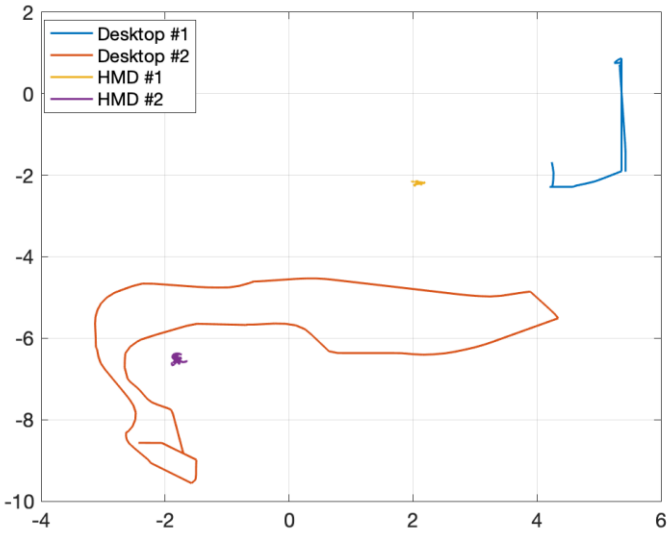


Figure 83 Floor view of user behaviour in scene #2.

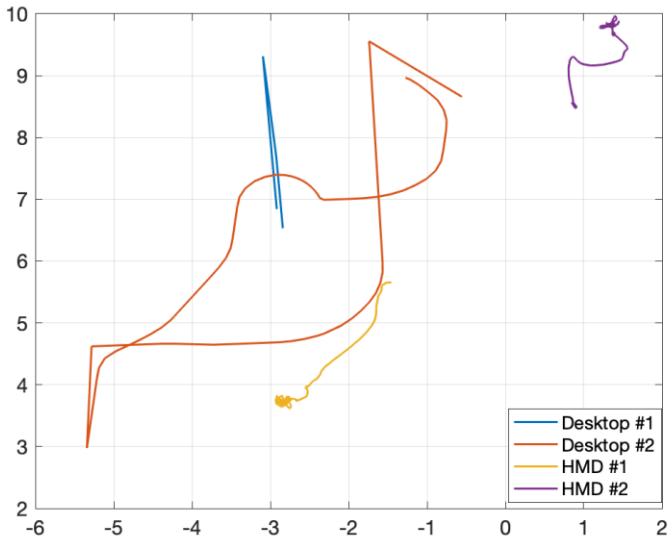


Figure 84 Floor view of user behaviour in scene #3.

Figure 82, Figure 83, and Figure 84 show the floor movement for the users in the three main scenes. It can be observed that the exploratory behaviour was much more varied in the first scene, whereas it became progressively less dynamic as the game progressed.

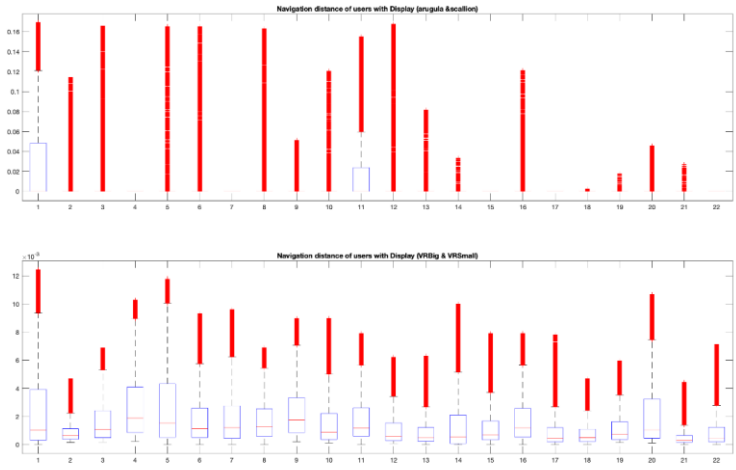


Figure 85 Distance with respect to the origin, for each user.

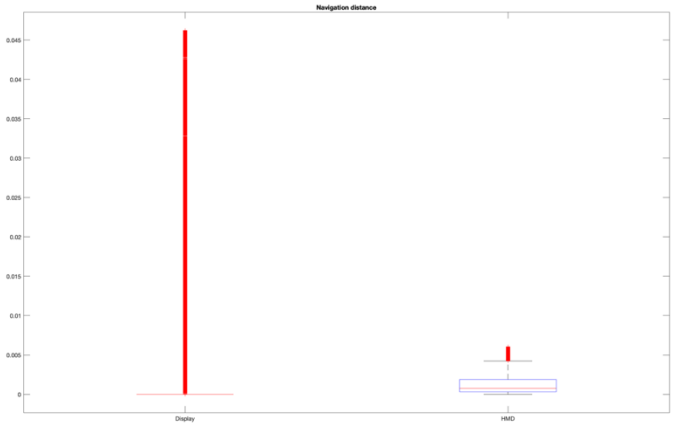


Figure 86 Distance with respect to the origin, for each group

Figure 85 and Figure 86 show the boxplot of the Euclidean distance of each user with respect to the origin, divided by the type of display (desktop vs HMD), to give an idea of how much the position of the users varied in each session. It can be observed that a more varied behaviour can be seen for Desktop users with respect to HMDs, which are more consistent. As observed before, the Desktop users were able to move freely around the scene, which explains the large number of outliers in the relative boxplot. On the other hand, HMD users were confined to predefined zones, thus making their position more uniform across the sessions.

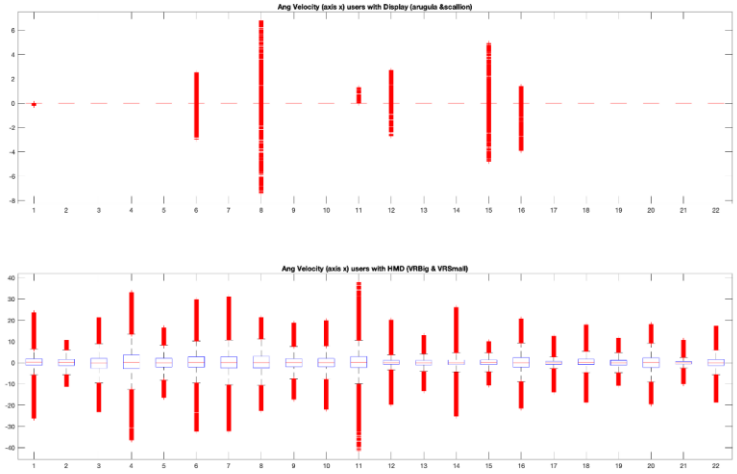


Figure 87 Angular velocity with respect to the x axis, for each user.

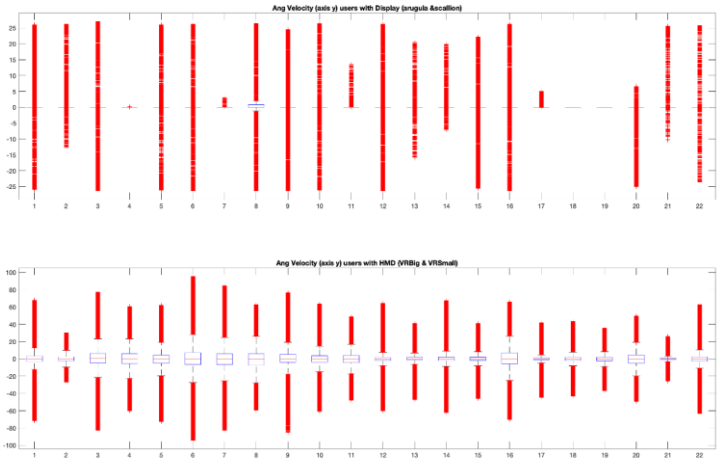


Figure 88 Angular velocity with respect to the y axis, for each user.

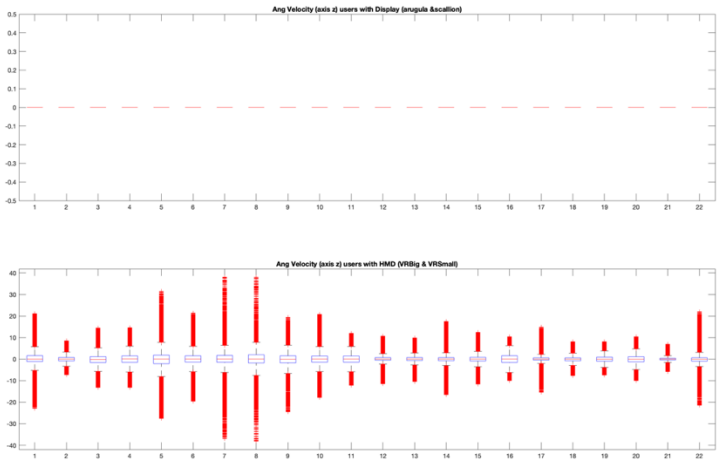


Figure 89 Angular velocity with respect to the z axis, for each user.

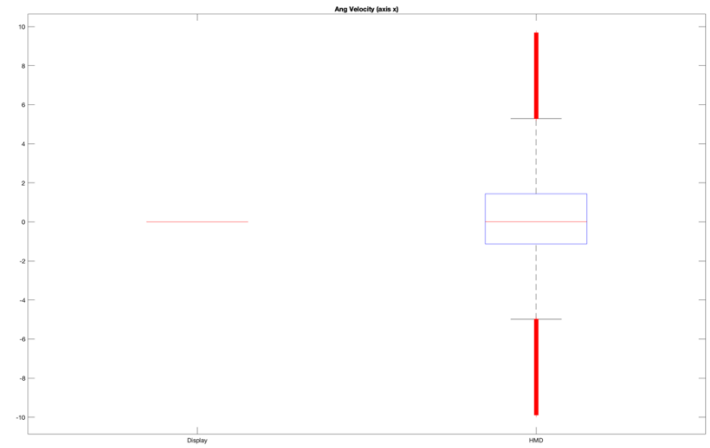


Figure 90 Angular velocity with respect to the x axis, for each group.

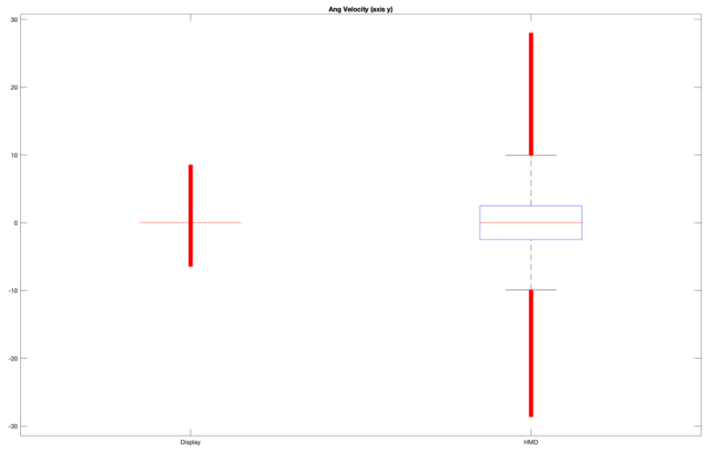


Figure 91 Angular velocity with respect to the y axis, for each group.

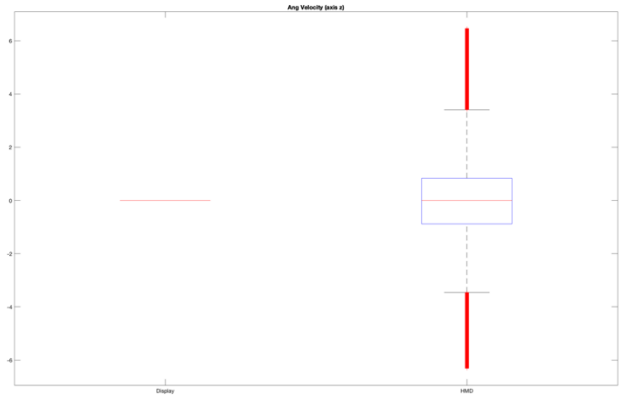


Figure 92 Angular velocity with respect to the z axis, for each group.

Figure 87, Figure 88, Figure 89, Figure 90, Figure 91, and Figure 92 show the boxplots of the angular velocity along the three axes. In this case, HMD users present a more varied behaviour; this is due to the fact that the type of display offers a natural way of looking around and changing the angle from which the scene is being viewed. Desktop users, on the other hand, had to rely on their controller to rotate their position, which leads to smaller angular velocity.

6.3.6.2 End-to-end latency

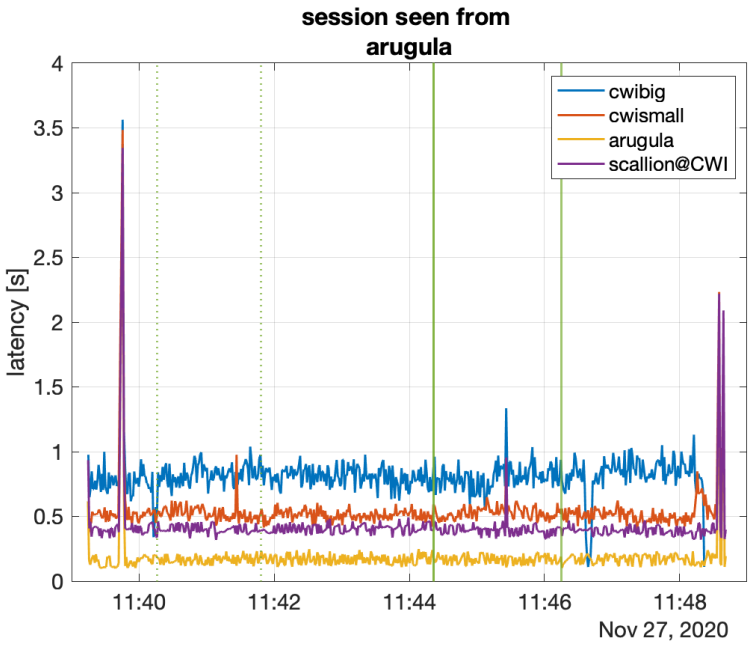


Figure 93 Latency of all device, as observed by machine arugula.

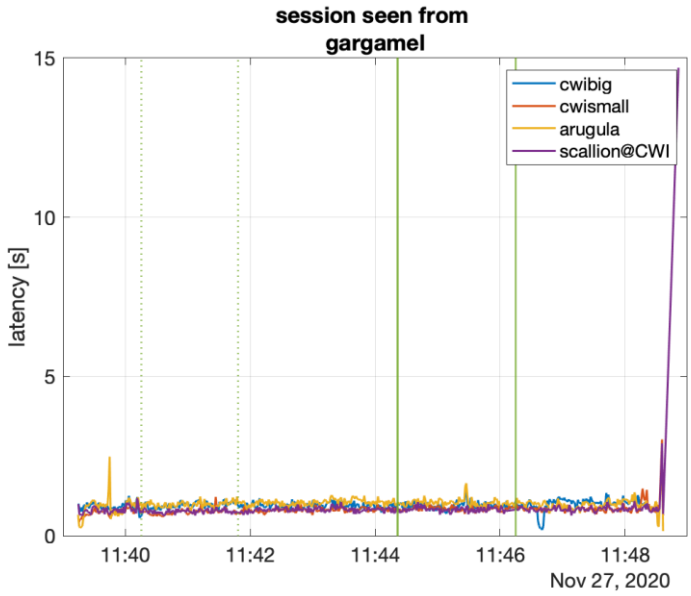


Figure 94 Latency of all device, as observed by machine gargamel.

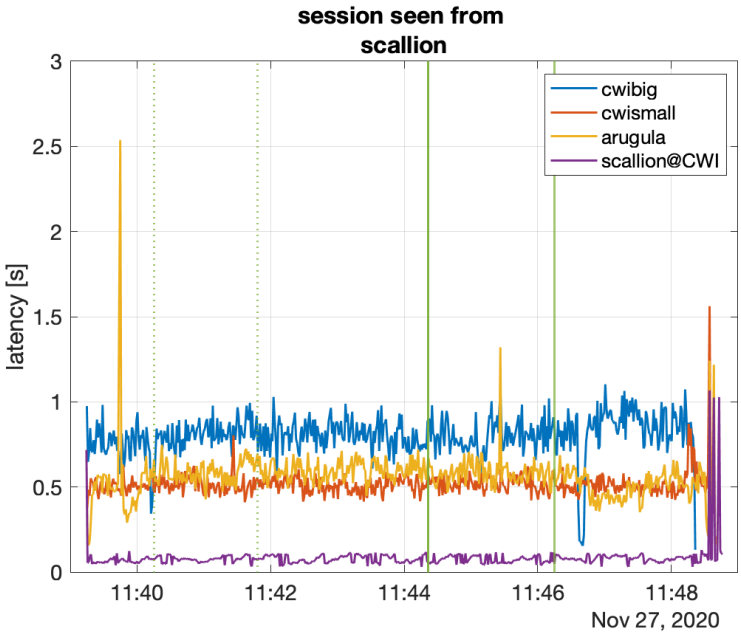


Figure 95 Latency of all devices, as observed by machine scallion.

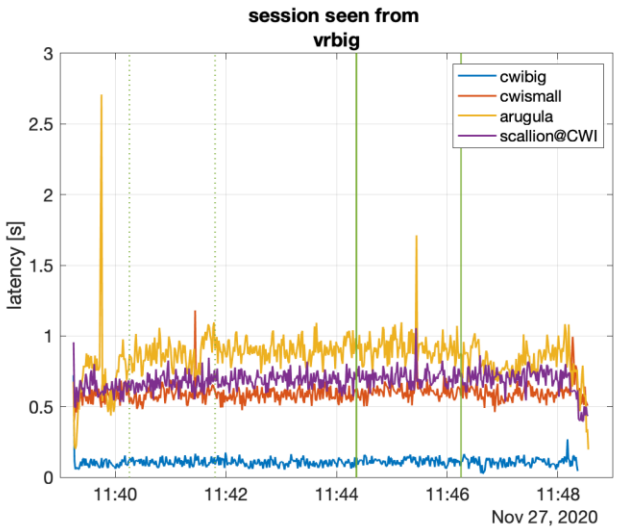


Figure 96 Latency of all devices, as observed by machine vrbig.

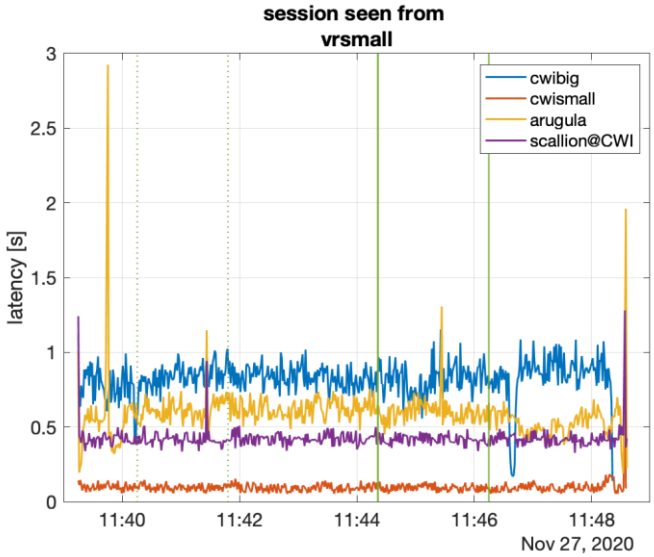


Figure 97 Latency of all devices, as observed by machine vrsmall.

Figure 93, Figure 94, Figure 95, Figure 96, and Figure 97 depict the latency observed in one session, as recorded from the machine in the session. Vertical lines indicate user triggers (dotted) and change of scene (solid). The values were synchronized to the orchestrator time to allow to compare latencies across devices. It can be observed that small values of latency were observed throughout, often lower than 1 second.

6.3.6.3 Correlation between objective and subjective measurements

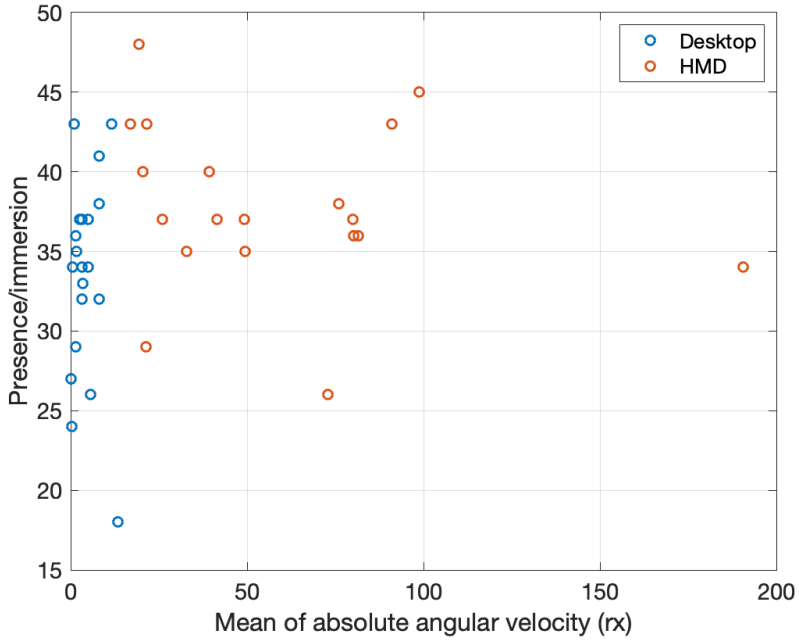


Figure 98 Mean of absolute angular velocity on the x axis vs Presence/Immersion values for each user

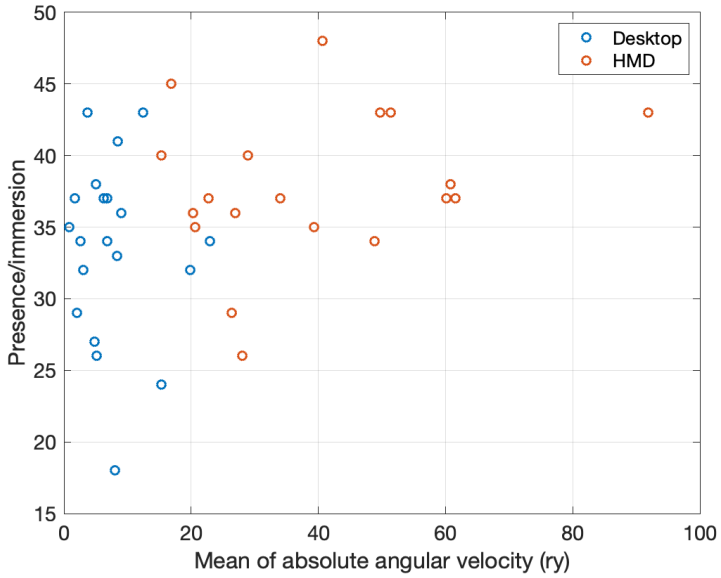


Figure 99 Mean of absolute angular velocity on the y axis vs Presence/Immersion values for each user

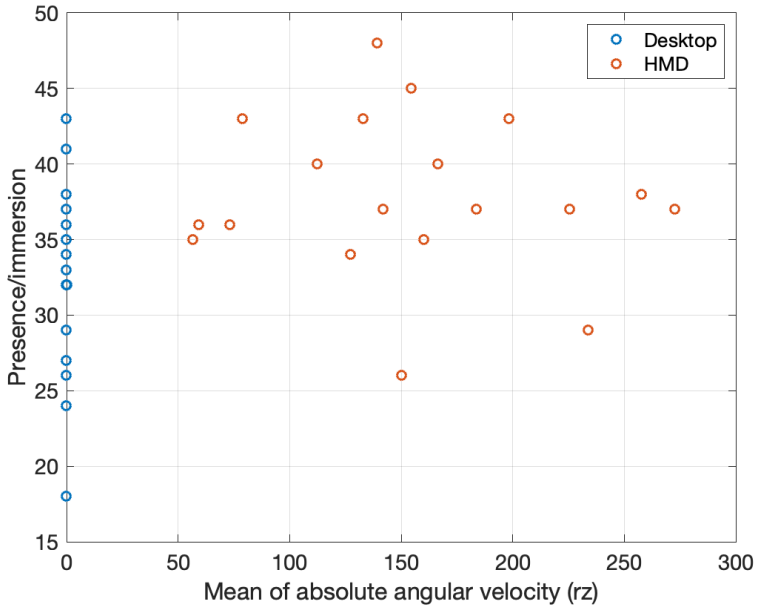


Figure 100 Mean of absolute angular velocity on the z axis vs Presence/Immersion values for each user

Figure 98, Figure 99, and Figure 100 show a scatterplot of the mean of absolute angular velocity with respect to the x, y, and z axis, respectively, shown against the values of Presence/Immersion as collected with the Social VR questionnaire. Greater mean angular velocity values can be observed for HMD users with respect to Desktop users. A slight trend towards a higher sense of presence for larger mean angular velocity can be observed; however, in general, there doesn't seem to be a strong correlation between mean absolute angular velocity and sense of presence.

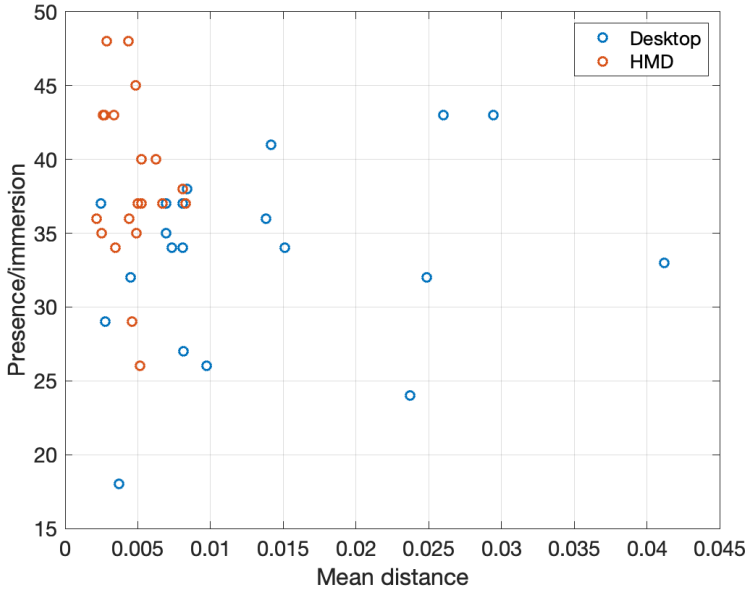


Figure 101 Relative distance between consecutive frames vs Presence/Immersion values for each user

Figure 101 depicts the relative distance between consecutive frames, plotted against the Presence/Immersion values. Whereas for HMD users there seems to be no correlation among the two values, for Desktop users a general trend can be observed that their sense of presence is positively correlated with the average distance among frames. This could indicate that a more exploratory behaviour could possibly lead to a heightened sense of presence; however, further studies and analyses are needed to confirm the hypothesis.

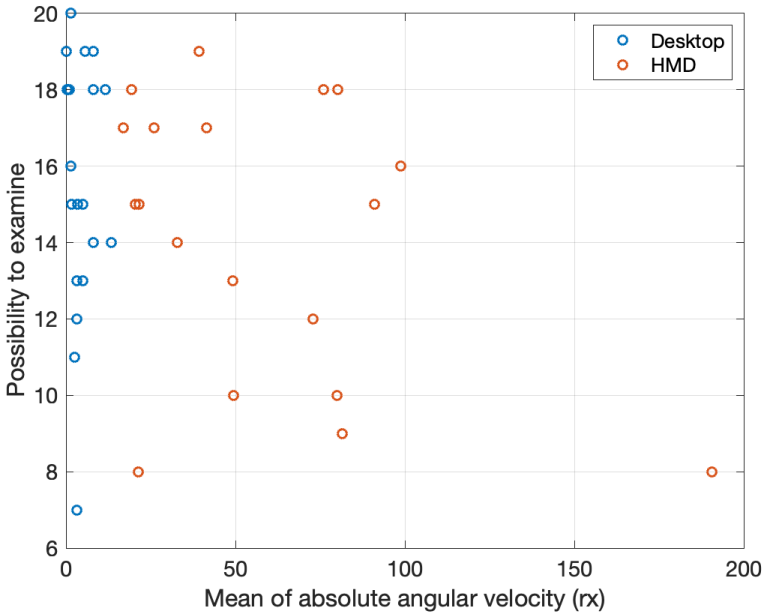


Figure 102 Mean of absolute angular velocity on the x axis vs Possibility to examine values for each user

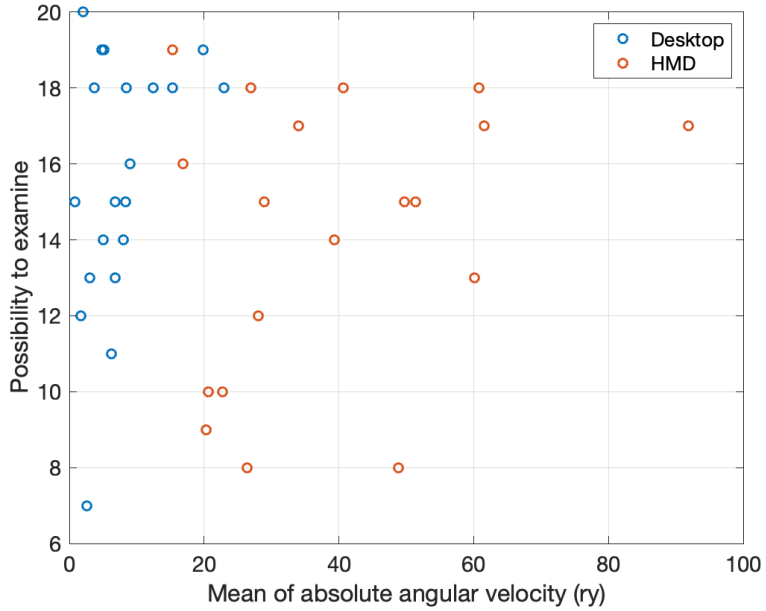


Figure 103 Mean of absolute angular velocity on the y axis vs Possibility to examine values for each user

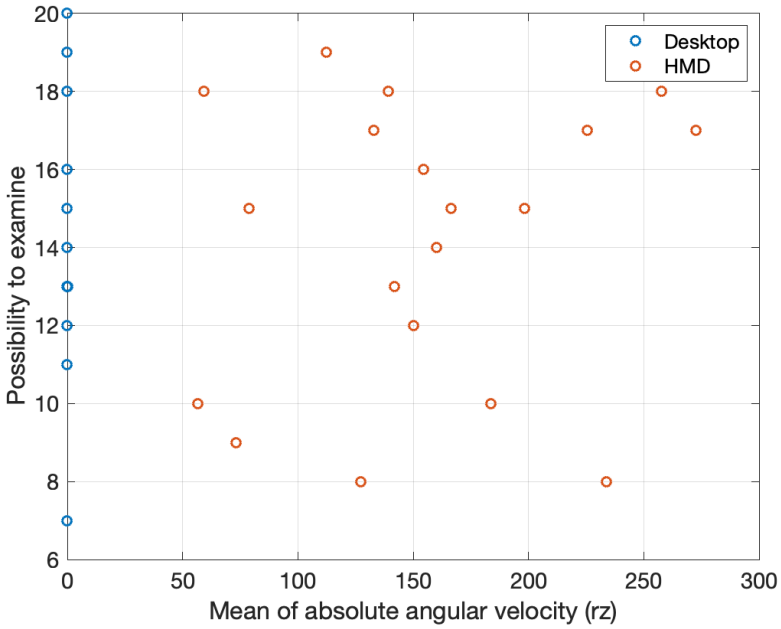


Figure 104 Mean of absolute angular velocity on the z axis vs Possibility to examine values for each user

Figure 102, Figure 103, and Figure 104 depict the mean of absolute angular velocity with respect to the x, y and z axis, respectively, against the Possibility to examine values, as collected from the Presence Questionnaire. There does not seem to be a correlation between the values, which could indicate that the average head rotation is not an indicator of the possibility to examine. That might signify that moving across space, as opposed to exploring through head rotations, would impact more this value.

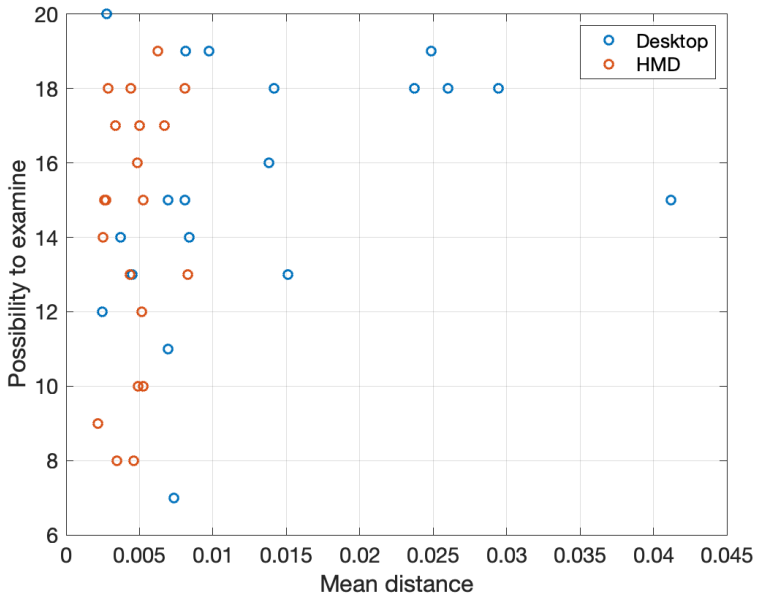


Figure 105 Relative distance between consecutive frames vs Possibility to examine values for each user

Figure 105 shows the relative distance between consecutive frames against the possibility to examine value. Whereas for HMD users distance values are quite low, due to the fact that exploration in 3D space was only possible through teleportation, distance values for Desktop users are generally higher, and seem to show a slight positive correlation with respect to the possibility to examine values.

6.4 Evaluation with Professionals

As in previous years, the project has been present in a number of events, demonstrating the results and new exploitations opportunities. In particular, this year, VRTogether has attended IMX2020, MMSys2020, and VRdays2020.

6.4.1 IMX2020

As part of the exploitation activity, CWI colleagues approached medical professionals and developed a social virtual reality (VR) clinic for patients to remotely access healthcare services. The results were demonstrated at ACM IMX 2020 to over 150 participants, and obtained the Best Demo Award.

- T. Xue, J. Li, G. Chen, and P. Cesar, A Social VR Clinic for Knee Arthritis Patients with Haptics. In Adjunct Proceedings of the ACM International Conference on Interactive Media Experiences (ACM IMX 2020), Barcelona, Spain, June 17-19, 2020.

The motivation of developing a social VR clinic is to support patients with limited physical mobility to travel fewer times to the hospital but still communicate well with doctors and nurses. Patients with knee arthritis are the target user group of this work. The final goal is to build a Social VR clinic that simulates the real consultation room and facilities in the hospital, in which patients can interact with the doctors or nurses with visualized information, such as surgery preparation procedures, 3D anatomical models, and a tour in the surgery room.

For this demonstration, we started with a series of ethnographic studies at the Reinier de Graaf hospital in Delft, which led to a better understanding of the complete patient journey (Figure 106), and to the identification of the requirements for the design and prototyping of a Social VR solution. It supports the four main identified activities within the patent journey (Figure 107): (1) visualization of the intervention process, (2) "walking into" a 3D virtual surgery room to "meet" the medical staff and to get familiar with the equipment, (3) interacting with an animated virtual 3D knee anatomical model and a virtual knee prosthesis model to see what the differences are before and after the surgery, and (4) learning to use an injection tool. The first prototype of the social VR clinic included the first three activities (Figure 108).

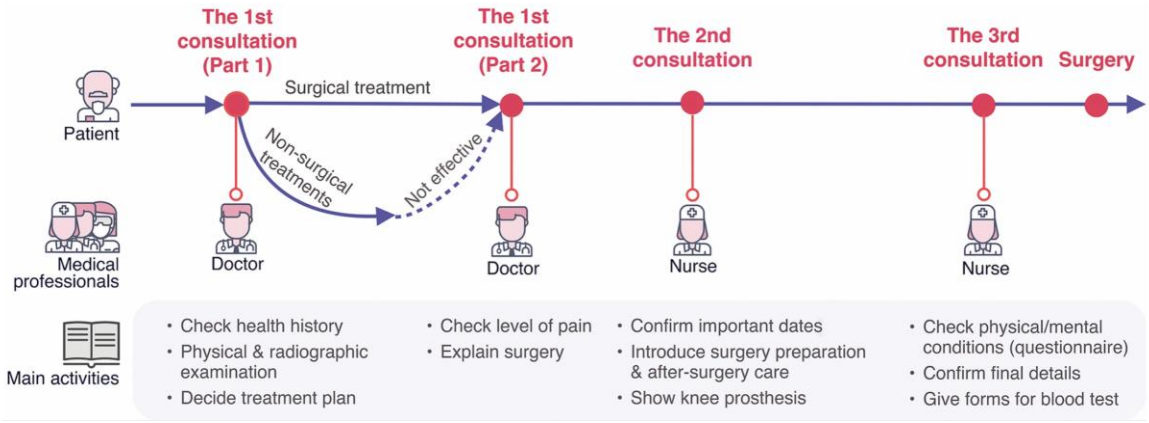


Figure 106. A typical treatment journey for knee arthritis patients.

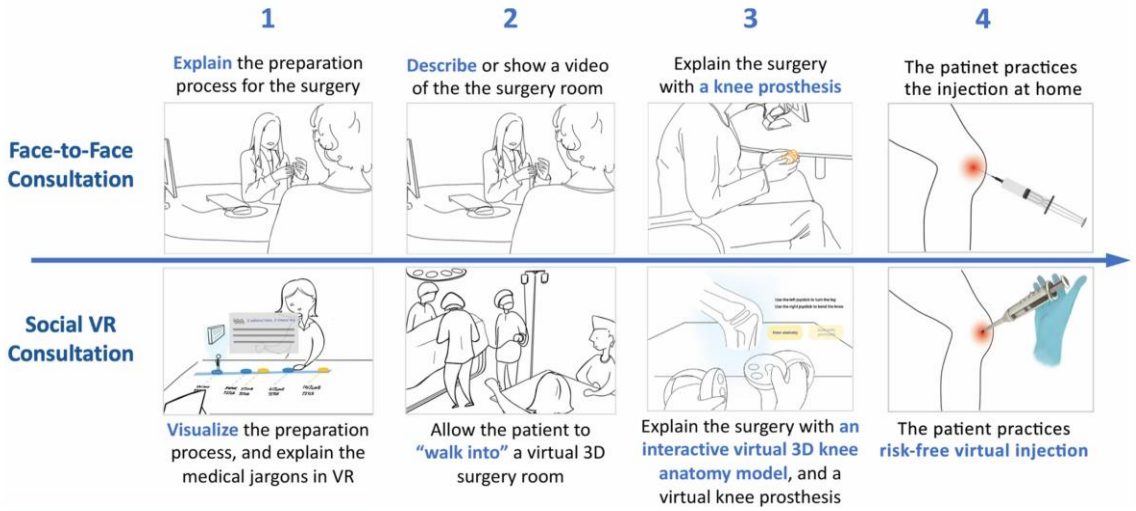


Figure 107. The four main activities related to a medical consultation: comparing the differences in the face-to-face (F2F) consultation with the social VR consultation.

Figure 108. The first social VR clinic prototype: (a) visualized surgery preparation timeline; (b) 3D "walk-in" surgery room; (c) 3D interactive knee anatomical and prosthesis models.

Based on the first prototype, we extended the experience to include a virtual injection tool to train patients to practice injecting medicine to their knees. By wearing a pair of mechanical VR gloves (SenseGlove), patients are able to use the virtual injection tool with realistic haptic feedback. The video shows the second prototype with the virtual injection tool (Figure 109). This project and the two prototypes show the potential of social VR as a new tool to help patients receive remote personalized medical care, as a potential exploitation opportunity for the project.

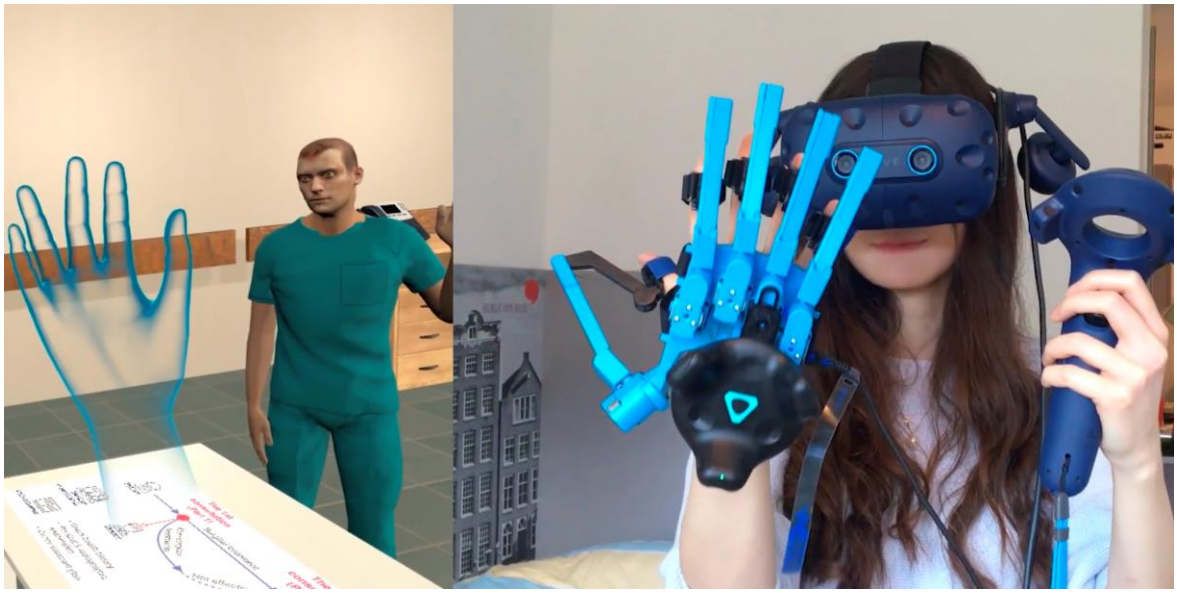


Figure 109. The second social VR clinic prototype.

6.4.2 MMSys2020

The native point cloud pipeline developed in the VRTogether project was demonstrated at ACM MMSys 2020 to around 200 participants. The associated paper is:

- J. Jansen, S. Subramanyam , R. , G. Cernigliaro, M. Martos, F. Pérez, and P. Cesar, "A Pipeline for Multiparty Volumetric Video Conferencing: Transmission of Point Clouds over Low Latency DASH," in Proceedings of the ACM Multimedia Systems Conference (ACM MMSys 2020), Istanbul, Turkey, June 8-11, 2020

The work was awarded the Best Demo Award
The video can be seen here: https://youtu.be/noL9pc4OzFY

The paper describes the architecture used for our pipeline (Figure 110) and how it achieves low delay through use of our point cloud codec and Low Latency DASH implementations.

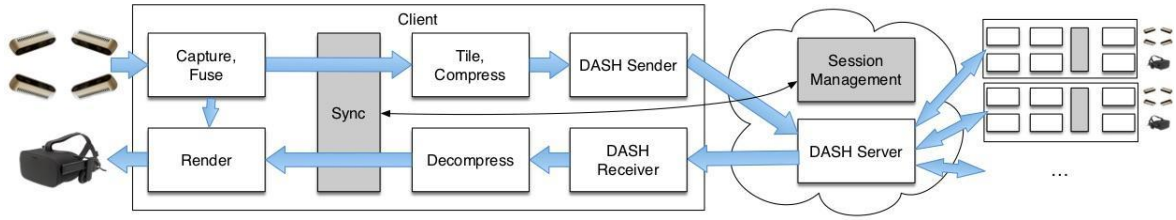


Figure 110. Architecture of Pointcloud Transmission Pipeline

How the system works from an end-user point of view, and how participants share a virtual world with their point cloud representation (Figure 111) is also demonstrated. Initial performance measurements of the pipeline in operation are included.

Figure 111. Point cloud Pipeline View in Virtual and Real World.

6.4.3 VRDays 2020

In November 2020, CWI and Sound did a premier by showing the world's first live volumetric video conference, using point clouds, over a commercial 5G network. The live demo was presented at VRDays Europe 2020 conference, in partnership with KPN, to over 1500 remote participants. A video from the event can be watched here: https://www.youtube.com/watch?v=EuoRfy18_Fc.

The live demo proved the extensibility and customization capabilities of our native point cloud pipeline, as we were able to extend it within only two weeks with a mobile phone as a new

capturing device. In this case, it was a commercial Android smartphone (Samsung Galaxy S20 Ultra 5G). The demo showcased volumetric video conferencing (based on point clouds) between a doctor, in a medical examination room, and an (acting) injured patient outside in the street.

The doctor was captured at the VRDays stage using Azure Kinect cameras. In the same stage the experience was captured by a virtual camera, projected on a big screen and broadcasted in real time. The patient joined the session outside near Science Park in Amsterdam, was captured with a Samsung Galaxy S20 Ultra 5G smartphone, and streamed over the standard 5G network of KPN in real-time. Figure 112 shows a diagram of the system used in the demo.

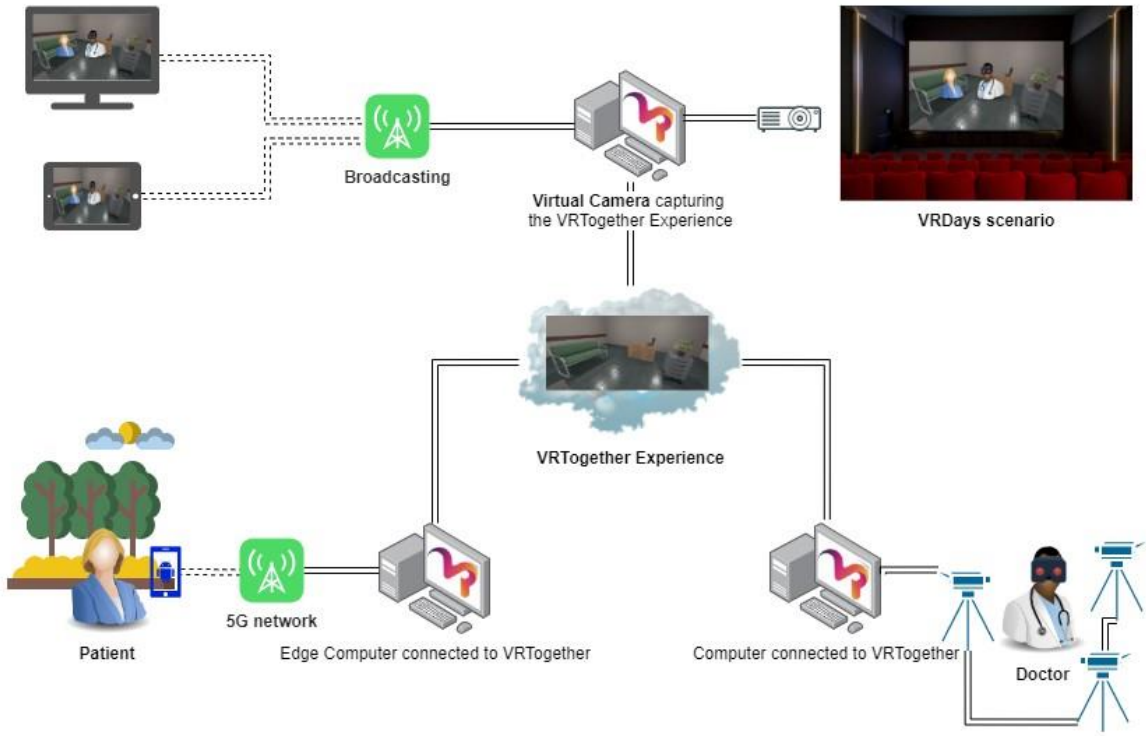


Figure 112. Diagram of the Demo showcased at VRDays Europe 2020

The implementation of the new mobile capturing source consisted in two main modules: An Android App, and a Proxy server implemented in Python. First, we implemented a custom Android App that uses both the colour and depth cameras available in the smartphone to generate 3D point clouds on the fly. The application performs a live streaming of these point clouds to a Proxy server using sockets. The streaming works over a commercial 5G network, allowing the phone user to be anywhere. Second, the proxy server was implemented in Python. It listens to incoming socket requests and stablishes a stable connection that allows receiving the stream of point clouds from the phone. These point clouds are then redirected and rendered in one of the rooms available in our VRTogether application.

Figure 113 shows the demo setup, the two locations and the coordination of all the infrastructure.

Figure 113. Demo at VRDays Europe 2020 in action

7 CONCLUSION

The third year of the project has been successful in terms of action pilots, even though the world has stopped due to a global pandemic. Pilot 3 has been conducted with 48 users in Amsterdam, gathering relevant and useful results about social VR experiences. Moreover, the project results have been demonstrated in IMX 2020 (exploring a new use case around healthcare) and MMsys (presenting the native pipeline for volumetric video) both getting the best demo award. Finally, the project premiered at VRDays 2020 the first volumetric video conferencing over a public 5G network, showcasing the possibilities of the solution developed during the project for remote consultation. Overall, Pilot 3 has been a success, completing the work in the previous years. The project has been able to explore three pilots, including:

- How social VR can be used for providing novel experiences for co-watching media remotely (Pilot 1)
- How social VR can change the broadcast model, but placing users in a studio set (Pilot 2)
- How social VR can be used for game-style applications like virtual scape rooms (Pilot 3)

In addition, during this period the project has run a number of experiments (nine), benchmarking other social VR platforms to better understand the added-value of the VRTogether platform, for better understanding the user and environment representation (including an experiment to explore other use cases than the ones defined in the project proposal), and testing the underlying technical infrastructure (including a technical pre-pilot test). Moreover, the project has maintained active two functional connected user labs (Barcelona-Thessaloniki-Amsterdam, and Rennes/Paris-The Hague) running a number of experiments to better understand the performance and has further evaluated the business potential with the local advisory board in a focus group.

Finally, the project has delivered a number of new metrics and protocols for evaluating social VR. Social VR is a new medium for communication and collaboration, and thus new manners of assessing the experience of users are needed. In particular, the project has proposed a complete protocol that includes novel questionnaires and objective data gathering regarding the user behaviour. In the same direction the first reduce-reference metric for evaluating point clouds has been proposed, allowing for objective and automatic monitoring of the perceptual quality of this new medium.

Overall, the results of this work package are positive and provide a solid baseline for others to further study social VR.

8 ANNEXES