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*

D4.2-Technical Report on First Pilot.

Revision: 3.0

Authors: Mario Montagud (i2CAT) and Pablo Cesar (CWI)

Delivery date: M13 (14-11-18)

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 deliverable reports the work during the first year of the proyect toward the execution of the pilot 1. In particular, it details the evaluation metrics and methods that have been defined and adopted for pilot 1, the 12 experiments that have led to the pilot, and the specifics of the pilot evaluations.

REVISION HISTORY

Revision	Date	Author	Organisation	Description
0.1	23-07-2018	Mario Montagud, Pablo Cesar	I2CAT, CWI	Structure and Table of Contents
0.2	11-09-2018	Tom De Koninck, Hans Stokking	TNO	Input TNO experiments
0.3	14-09-2018	Henrique Galvan Debarba, Caecilia Charbonnier	Artanim	Input Artanim experiment and pilot 1 motion capture
0.4	14-09-2018	Mario Montagud	i2CAT	Inputs about the Objective Performance Metrics
0.5	17-09-2018	Henrique Galvan Debarba	Artanim	Objective metrics
1.0	21-09-2018	Many	CERTH, CWI	First full draft of the deliverable
1.1	03-10-2018	Henrique Debarba	Artanim	Added tables to 2.2
1.2	08-10-2018	Mario Montagud	i2CAT	Selection of Objective Performance Metrics to be Used in Pilot 1
1.3	11-10-2018	Rick Hindriks	TNO	Elaboration and improvement of TNO experiments descriptions
1.4	15-10-2018	Manos Christakis	CERTH	Elaboration and improvement of CERTH experiments descriptions
1.5	21-10-2018	Pablo Cesar	CWI	Inclusion of CWI-1, CWI-3 and Artanim1-2
2.0	30-10-2018	Pablo Cesar	CWI	Second full draft of the deliverable (complete)
2.5	07-11-2018	Pablo Cesar and Mario Montagud	CWI, i2CAT	Definitive Version
2.6	11-11-2018	Henrique Debarba	Artanim	Internal review

3.0	14-11-2018	Pablo Cesar and Mario Montagud	CWI, I2CAT	Reviewed version

Disclaimer

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

Statement of originality:

This document contains original unpublished work except where clearly indicated otherwise.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

EXECUTIVE SUMMARY

During the first year, the project has successfully run the first pilot. The pilot took place in Amsterdam (during IBC2018) in September with professionals and in Barcelona with end-users in October. The coordination between WP2 and WP4 allowed for a successful outcome. In particular, the activities that have taken place in the first year include:

- full content creation, production and post-production, of high-quality assets for pilot 1. Such content has been created in different forms (see D4.1)
- pilot deployment in Barcelona (end-users) and in Amsterdam (professionals) and pilot evaluation based on a set of new protocols and metrics for social VR (sVR)

This deliverable reports the work that has made possible to successfully run the pilots. In particular, it details the metrics and methods developed for the pilot, the 12 experiments that have led to the pilot, and the specifics of the pilot evaluations.

Section 2 focuses on the metrics and evaluation methods. During the first year, the project has created a new protocol and set of metrics, and the associated data analysis toolset, for evaluating social VR with end-users. The impact of this result may go beyond the project, since it can be become the de-facto standardised manner for evaluating a new genre of experiences: social VR. The protocol and metrics include both quantitative and qualitative aspects, such as a new questionnaire combining presence, immersion, and togetherness; a set of objective metrics based on the behaviour of the user, focusing on speech analysis, neck rotation, body movement, etc.; and performance metrics for profiling the system aspects. This novel evaluation method has been iteratively developed and validated through a human-centred process, including a number of experiments (with around 100 users in total).

Section 3 describes the twelve experiments that have taken place in the first year in order to help the project constructing the trial, evaluating the innovation value of the system and providing technical requirements. In particular, three types of experiments have taken place: technology requirements, experience design and evaluation, and innovation and entertainment value. Examples of the experiments include:

- A technical evaluation assessing per-module the distribution evaluation of the time-varying mesh (TVM) pipeline;
- A data capturing system to acquire RGB-D faces with and without HMDs, for developing, training and evaluating an algorithm for real-time HMD removal;
- Two experiments aimed at developing and testing the subjective and objective methodologies to evaluate and compare social VR systems; and
- Presence in relevant events such as VRDays2017 and MMSys2018

Section 4 reports the pilot evaluations with both end-users and professionals, and provide some initial evaluation of them. The evaluations include the user experience, the perceived added value, and the performance of the system between remote locations. Results are extremely positive and encouraging, showing how social VR is a promising area for the media industry. Finally, the annexes provide all the questionnaires and consent forms used for the reported evaluations.

During the second year, further pilot actions, using the pilot 1 material, are expected at relevant events such as VRDays 2018 in Amsterdam or ICT2018 in Vienna, and festivals (an initial dialogue with Sundance has already started). Moreover, preparations for pilot 2 have already started and the second pilot will run by the end of the second year.

CONTRIBUTORS

First Name	Last Name	Company	e-Mail
Mario	Montagud	I2CAT	mario.montagud@i2cat.net
Guillermo	Calahorra	Entropy Studio	guillermo@entropystudio.net
Tom	De Koninck	TNO	Tom.dekoninck@tno.nl
Hans	Stokking	TNO	Hans.stokking@tno.nl
Henrique	Galvan Debarba	Artanim	henrique@artanim.ch
Caecilia	Charbonnier	Artanim	caecilia@artanim.ch
Jie	Li	CWI	Jie.Li@cwi.nl
Francesca	De Simone	CWI	Francesca.De.Simone@cwi.nl
Pablo	Cesar	CWI	P.S.Cesar@cwi.nl
Vladimiros	Sterzentsenko	CERTH	vladster@iti.gr
Rick	Hindriks	TNO	rick.hindriks@tno.nl
Manos	Christakis	CERTH	manchr@iti.gr

CONTENTS

TABLES OF FIGURES AND TABLES

LIST OF ACRONYMS

Acronym	Description
CGI	Computer Generated Imagery
DCR	Degradation Category Rating
DOF	Degrees of Freedom
EFA	Exploratory Factor Analysis
F2F	Face To Face
FB	Facebook Spaces
GSR	Galvanic Skin Response
HMD	Head Mounted Display
HMFA	hierarchical multiple factor analysis
LoD	Level of Detail
MOS	Mean Opinion Score
MU	Media Unit
P/I	Presence/Immersion (PI)
PCA	Principle Component Analysis
QoE	Quality of Experience
QoI	Quality of Interaction
RGB	Red, Green and Blue
RGB-D	Red, Green, Blue and Depth
SOS	standard deviation of opinion score
SM	Social Meaning
sVR	social VR
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 first pilot of the project, which took place in Amsterdam during IBC2018 (September 2018) with professionals and in Barcelona with end-users (October 2018). 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 and the initial results. This is a first version of the deliverable that will be subsequently updated in the second year for pilot 2 and in the third year for pilot 3. Overall, the objectives of this WP have been met, with high-quality production of content (see D4.1) and two successful pilot evaluations both with end-users and with professionals. Pilot 1 is still active, and the project is trying to bring it to a festival (Sundance or Venice) and to a number of relevant events (VRDays 2018 and ICT2018). In parallel, work towards pilot 2 has already started.

1.2. Scope of this document

This document reports all the activities leading towards the first 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 evaluations. It includes as well an initial plan for pilot 2.

1.3. Status of this document

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

1.4. Relation with other VR-Together activities

This document gathers the outputs of all the activities of WP4 during the first year (T4.1 to T4.3). D4.1, from the same WP, further details 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 first year of the project, applied to the different pilot actions and user evaluations. In particular:

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

2.1. Subjective Evaluation

Introduction

Successful design of social Virtual Reality (VR) products requires insights into the user experience that take place while using the products. Although the interest in evaluating user experience of social VR is high, currently there are no common methodologies and metrics. Therefore, there is a strong need to analyse what User Experience (UX) evaluation methodologies and metrics are currently available, what is missing, and what needs to be developed. In this section, we specially focus on subjective evaluation metrics that can be used for social VR. We start from understanding the user experience in social VR, based on UX definitions and frameworks from other related fields. After that, we provide an overview of the relevant subjective evaluation metrics from related fields, such as mediated social communication. Based on these metrics, a new questionnaire to evaluate user experience in social VR is proposed.

User experience in social VR

According to [ISO 9241-110:2010] (clause 2.15), user experience is defined as: ‘a person’s perceptions and responses that result from the use and/or anticipated use of a product, system or service’. However, this definition is not sufficient when we take interpersonal communication into consideration. The development of interpersonal communication technologies has put emphasis on the social aspects of user experience. This user experience cannot only be seen as an individual’s reaction, but also as something constructed by social interaction. [Battarbee 2003] therefore created a definition for ‘co-experience’ as the experience that users create together in social interaction. The definition of user experience in mediated social interaction is ‘the various types of experiences people have when using the system, product or service for social communication [Steen et al., 2012]. For virtual environments, the emphasis of user experience is on the ability to produce a sense of presence, or ‘being there’ [Bowman et al., 2007]. Although a commonly accepted definition of user experience for social VR cannot be found, these related definitions help to create an initial common ground of understanding.

Research has also been done to understand what dimensions of experience are important for mediated social interaction. According to [Steen et al., 2012] people’s experiences of mediated social interaction can be divided into three categories: 1) Aesthetics: people’s experiences of the sensorial qualities of the system that enables social communication; 2) Interacts: people’s experiences of interacting with the system and with others via the system; 3) Meaning: people’s experiences of social communication in the broader context of daily life. These three types of experience also correspond to the three groups of UX evaluation methods: Sensory characteristics, Emotional reactions and Meaning [Vermeeren et al., 2010]. On the other hand, the important dimensions of experiences in virtual reality have also been discussed in the literature. According to [Heim, 1998], who defined VR with 'three I's', the three characteristic of VR are immersion, interactivity and information intensity. [Steuer, 1992] define virtual reality based on the concept of 'Telepresence', and two dimensions of experiences were vividness and interactivity. Even though there are no commonly accepted user experience frameworks for social VR, the frameworks from the field of mediated social communication and virtual reality, as mentioned above, can help us to propose a list of important dimensions of user experience in social VR. Figure 1 shows the relationships among these fields of knowledge.

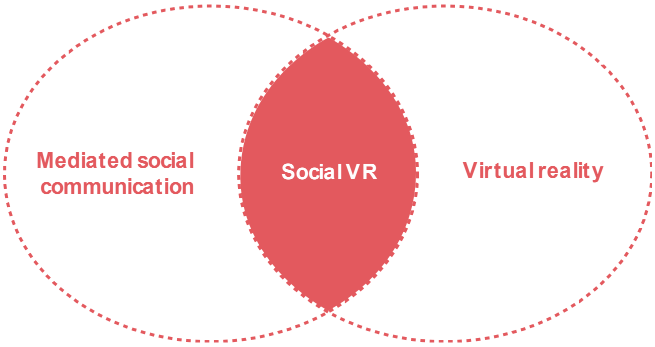


Figure 1- Relationships between different fields of knowledge

Related Questionnaires

The focus of understanding UX is not only on utilitarian aspects (e.g., user cognition and user performance) of human-technology interactions, but also on user affect, sensation, and the meaning as well as value of such interactions [Law et al., 2009] [Vermeeren et al., 2010]. A common methodology to evaluate UX is to collect self-reported responses from users through a combination of metrics usually in the form of a questionnaire [Albert & Tullis 2013]. With reference to a recent survey, 53% of the user experience studies have employed questionnaires to yield quantifiable results [Bargas et al., 2011] [Vermeeren et al., 2010]. An overview of the existing questionnaires that can be used to evaluate user experience in social VR is very helpful for understanding the current state of art.

Currently, there are no commonly acknowledged metrics for evaluating UX in social VR. However, based on the frameworks of general UX, social UX, UX in mediated social communication systems as discussed above [Vermeeren et al., 2010] [Steen et al., 2012] [Steuer, 1992] [Heim, 1998], and a series of user studies [Kong, 2018], we propose to focus on three dimensions that can properly address UX in social VR, namely quality of interaction, social meaning and presence/immersion. In the following part, definitions for each category are provided and the related questionnaires are discussed.

Dimension 1 - Quality of interaction

Quality of interaction in social VR is described as the ability of the user to interact with the virtual world and to interact with other users in that virtual world [Steuer, 1992] [Steen et al., 2012]. It assesses how well the interactions in VR resemble the face-to-face interactions in terms of quality of communication, mutual sensing of emotions and naturalness.

Quality of communication

[Garau et al., 2001] investigated the impact of visual and behavioral realism in avatars on perceived quality of communication with post-experiment questionnaires, which aim at evaluating 1) naturalness of the conversation as compared to face-to-face; 2) degree of involvement in the conversation; 3) sense of co-presence; 4) evaluation of the conversation partner. This questionnaire was partly based on the previous questionnaire designed to elicit subjective responses to mediated communication [Stellen, 1995]. While [Garau et al., 2001] emphasizes the communication process, the questionnaire from [Steen et al., 2012] focuses more on the results and influences of the communication, including 1) understanding and being understood; 2) being able to communicate one's intentions and having the feeling the others can do the same; 3) knowing how the other is feeling during the social interaction and having the feeling the other knowing your feelings as well.

Experienced emotion

Emotional experience is believed to contribute to the overall user experience. The valence, intensity of the emotions and whether the emotions are in control influence the overall experience [Vermeeren et al., 2010]. [Desmet, 2018] provided an overview of emotion measurement approaches, being sorted into two categories: non-verbal instruments that measure either the expressive or the physiological component of emotion (e.g., facial or vocal expressions, blood pressure, skin or heart responses); and verbal self-report instruments that respondents to report their emotions with the use of a set of rating scales or verbal protocols. A rich and easy-to-use pictorial instrument developed by [Vastenburg et al., 2011] was used to allow participants to self-report their emotions and the intensity of their emotions.

Naturalness of interaction

Naturalness describes the degree to which users perceive the interaction as predictable, logical or in line with expectations [Skalski et al., 2011]. Perceived naturalness is often used to assess how well virtual tools can simulate real world interaction, which is influenced by individual differences as well as the technology itself. In the study by [Nilsson et al., 2013], perceived naturalness of leg movements in VR was assessed with a self-report questionnaire with items covering the following four topics: 1) naturalness (e.g., how natural the experience was), 2) physical strain (e.g., whether the virtual actions required muscle activities), 3) self-motion compellingness (e.g., whether felt as if they were actually moving) and 4) acclimatization (e.g., how quick forgot they were not really walking). Naturalness is also recognized as a contributing factor to presence. [Witmer & Singer, 1998] included items that assess naturalness of the human-IVEs interactions and virtual controlling mechanism in their presence questionnaire.

Dimension 2 - Social meanings

According to the UX definition by [Law et al., 2009] and [Vermeeren et al., 2010], the meaning or meaningfulness of interactions is an essential factor contributing to UX. In this project, social meaning is defined to enclose the experience of "being together" both mentally and physically.

Social connectedness

Mentally "being together" refers to social connectedness as defined by [van Bel et al., 2009], which is a short-term experience of belongingness and relatedness based on social appraisals and relationship salience. To assess the social connectedness at the individual level (toward a particular individual), a questionnaire with 29 items was developed by [van Bel et al., 2009], which was identified with two main factors: the sense of sharing and involvement and dissatisfaction with contact quality. The first main factor contained four sub-factors: relationship salience, shared understandings, knowing experiences of each other and feelings of emotional closeness. The second main dimension only consisted of dissatisfaction with contact quality.

Togetherness

Physically “being together” or co-presence refers to the degree to which people believe she/he is not alone, but being together with others in a shared space [Durlach & Slater, 2000] [Biocca, et al., 2001]. Co-presence also includes the level of peripherally or focally awareness of each other [Biocca et al., 2001]. Three categories of questions were created for assessing co-presence/togetherness: isolation/aloneness, mutual awareness and attention allocation.

Dimension 3 – Presence and immersion

Initiating a sense of presence is believed to be a core differentiator that sets VR apart and takes traditional computing interfaces to the next level [Berg and Vance, 2017]. Immersion is considered as an objective measurement of the technology systems. The more a system preserves fidelity in relation to their equivalent real-world sensory modalities, the more “immersive” experience it leads to [Slater et al., 1999]. [Slater et al., 1999] stressed that presence and immersion are not the same. Presence is a human reaction to immersion, which is subjective experience towards a system of a certain level of immersion. However, [Witmer & Singer, 1998] argued that rather than an objective measurement, immersion is a psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment. When subjects are requested to report their subjective sense of presence and immersion, the questionnaires from [Witmer & Singer, 1998] and [Schubert et al., 2001] can be used, assessing the experience based on following factors: 1) control factors, 2) sensory factors, 3) distraction factors, 4) realism factors, 5) spatial presence and 6) involvement. The assessment of immersion can refer to the immersive tendency questionnaire [ITQ] [Witmer & Singer, 1998] and the immersion for gaming [Jennett et al., 2008]. The ITQ focuses on four factors: 1) isolation from the physical environment, 2) perception of self-inclusion in the virtual environment, 3) natural modes of interaction, and 4) control and perception of self-movement. According to [Jennett et al., 2008], the immersive experience was indicated by lack of the awareness of time, lack of the awareness of the real world, involvement and a sense of being in the task environment.

Design of a new questionnaire

Through comparing the related questionnaires, we found that questionnaires used for similar dimensions of experience have different question items and different factors. In order to make these questionnaires more unified, clear definitions for each dimensions of experience need to be created and adopted. On the other hand, many overlapped question items were found among some categories of experiences. The fact that different experiences are able to influence each other may explain for the overlaps. However, the overlaps might be caused by a lack of consistent framework for user experience in social VR.

The requirement for developing a new questionnaire that can be commonly used to evaluate the user experience in social VR is identified. Therefore, we propose the design of the following questionnaire, based on the related work discussed above, and our own findings from user studies [Kong, 2018].

Dimension 1 - Quality of interaction

Questionnaire items 2-12 were for measuring Quality of interaction, where items 2,3,9,10 were designed by us, items 4-8 adapted from [Garau et al., 2001], and items 11-12 adapted from [Nilsson et al., 2013]. Questionnaire item 1 was a graphical self-report emotion rating questionnaire adapted from [Vastenburg et al., 2011]. Presence and immersion in VR environment has been shown to play an important role in emotional reactions [Diemer et al., 2015], which led us to measure this subjectively. This questionnaire was chosen (despite many existing emotion report questionnaires) as it is both rich (captures multiple categories) and an easy-to-use pictorial mood-reporting instrument.

Q2. “I was able to feel my partner’s emotion while watching the contents.”
Q3. “I was sure that my partner often felt my emotion.”
Q4. “The experience of watching the contents with my partner seemed natural.”
Q5. “The actions used to interact with my partner were similar to the ones in the real world.”
Q6. “It was easy for me to contribute to the conversation with my partner.”
Q7. “The conversation with my partner seemed highly interactive.”
Q8. “I could readily tell when my partner was listening to me.”
Q9. “I found it difficult to keep track of the conversation.”
Q10. “I felt completely absorbed in the conversation.”
Q11. “I could fully understand what my partner was talking about.”
Q12. “I was very sure that my partner understood what I was talking about.”

Dimension 2 – Social meaning

Questionnaire items 13-23 were for measuring Social meaning, where items 13-17 were adapted from [Biocca et al., 2001], items 20-22 adapted from van [Bel et al., 2009], item 23 adapted from [Steen et al., 2012], and items 18-19 designed by us.

Q13. “I often felt as if I was all alone while watching the contents.”
Q14. “I think my partner often felt alone while watching the contents.”
Q15. “I often felt that my partner and I were sitting together in the same space.”
Q16. “I paid close attention to my partner.”
Q17. “My partner was easily distracted when other things were going on around us.”
Q18. “I felt that watching the contents together in VR enhanced our closeness.”
Q19. “Watching the contents together created a good shared memory between me and my partner.”
Q20. “I derived little satisfaction from the content watching experience with my partner.”
Q21. “The content watching experience with my partner felt superficial.”
Q22. “I really enjoyed the time spent with my partner.”
Q23. “How emotionally close to your partner do you feel now?”

Dimension 3 – Presence and immersion

Questionnaire items 24-33 were for measuring Presence and immersion, where item 24 was adapted from [Slater et al., 1997], items 25-26 from [Schubert et al., 2001], item 27 from [Witmer & Singer, 1998], and items 28-33 from [Jennett et al., 2008].

Q24. “In the virtual world I had a sense of ‘being there’.”
Q25. “Somehow I felt that the virtual world was surrounding me and my partner.”
Q26. “I had a sense of acting in the virtual space, rather than operating something from

outside."

Q27. "My content watching experience in the virtual environment seemed consistent with a real-world experience."
Q28. "I did not notice what was happening around me in the real world."
Q29. "I felt detached from the outside world while watching the contents."
Q30. "At the time, watching the contents with my partner was my only concern."
Q31. "Everyday thoughts and concerns were still very much on my mind."
Q32. "It felt like the content watching experience took shorter time than it really was." [Duration of contents is ~7min]
Q33. "When watching the contents with my partner, time appeared to go by very slowly."

The 5-point likert-scale questionnaire has been created for this project, as a novelty for evaluating social VR, and has been validated in a number of experiments, including the pilot. A 5-point (and not 7-point) scale was chosen since with coarser measurements on lower sample sizes (typical of lab studies), there is lower variance. Furthermore, previous work has shown that 5- and 7-point scales are comparable if rescaling is performed [Dawes, 2008].

References:

[Albert & Tullis 2013] Albert, W., & Tullis, T. (2013). Measuring the user experience: collecting, analyzing, and presenting usability metrics. Newnes.
[Bargas et al., 2011] Bargas-Avila, J.A., & Hornbæk, K. (2011). Old wine in new bottles or novel challenges? A critical analysis of empirical studies of user experience. In Proc. CHI'11. Vancouver, Canada.
[Battarbee 2003] Battarbee, K. (2003, April). Co-experience: the social user experience. In CHI'03 extended abstracts on Human factors in computing systems (pp. 730-731). ACM.
[Berg and Vance, 2017] Berg, L.P. and Vance, J.M. (2017). Industry use of virtual reality in product design and manufacturing: a survey. Virtual Reality, 21(1), pp.1-17.
[Biocca et al., 2001] Biocca, F., Harms, C., & Gregg, J. (2001, May). The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In 4th annual international workshop on presence, Philadelphia, PA (pp. 1-9).
[Bowman et al., 2007] Bowman, D. A., & McMahan, R. P. (2007). Virtual reality: how much immersion is enough?. Computer, 40(7).
[Dawes, 2008] Dawes, J. (2008). Do Data Characteristics Change According to the Number of Scale Points Used? An Experiment Using 5-Point, 7-Point and 10-Point Scales. International Journal of Market Research 50, 1 (2008), 61–104. DOI: http://dx.doi.org/10.1177/147078530805000106
[Desmet, 2018] Desmet, P. (2018). Measuring emotion: Development and application of an instrument to measure emotional responses to products. In Funology 2 (pp. 391-404). Springer, Cham.
[Diemer et al., 2015] Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., & Mühlberger, A. (2015). The impact of perception and presence on emotional reactions: a review of research in virtual reality. Frontiers in psychology, 6, 26.
[Durlach & Slater, 2000] Durlach, N., & Slater, M. (2000). Presence in shared virtual environments and virtual togetherness. Presence: Teleoperators & Virtual Environments, 9(2), 214-217.
[Garau et al., 2001] Garau, M., Slater, M., Bee, S., & Sasse, M. A. (2001, March). The impact of eye gaze on communication using humanoid avatars. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 309-316). ACM.
[Heim, 1998] Heim, M., "Virtual Realism", Oxford University Press, 1998, p.7.

[ISO 9241-110:2010] ISO DIS 9241-210:2010. Ergonomics of human system interaction - Part 210: Human-centred design for interactive systems (formerly known as 13407). International Standardization Organization (ISO). Switzerland.

[Jennett et al., 2008] Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. International journal of human-computer studies, 66(9), 641-661.

[Law et al., 2009] Law, E., Roto, V., Hassenzahl, M., Vermeeren, A., and Kort, J. (2009). Understanding, Scoping and Defining User eXperience: A Survey Approach. Proc. CHI'09, ACM SIGCHI conference on Human Factors in Computing Systems.

[Nilsson et al., 2013] Nilsson, N. C., Serafin, S., & Nordahl, R. (2013, November). The perceived naturalness of virtual locomotion methods devoid of explicit leg movements. In Proceedings of Motion on Games (pp. 155-164). ACM.

[Schubert et al., 2001] Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. Presence: Teleoperators & Virtual Environments, 10(3), 266-281.

[Slater et al., 1997] Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. Presence: Teleoperators & Virtual Environments, 6(6), 603-616.

[Slater et al., 1999] Slater, M., Lotto, B., Arnold, M. M., & Sánchez-Vives, M. V. (2009). How we experience immersive virtual environments: the concept of presence and its measurement. Anuario de Psicología, 2009, vol. 40, p. 193-210.

[Sellen, 1995] Sellen, A. J. (1995). Remote conversations: The effects of mediating talk with technology. Human-computer interaction, 10(4), 401-444.

[Skalski et al., 2011] Skalski, P., Tamborini, R., Shelton, A., Buncher, M., & Lindmark, P. (2011). Mapping the road to fun: Natural video game controllers, presence, and game enjoyment. New Media & Society 13, 2, 224–242.

[Steen et al., 2012] Steen, M. (2012). D8. 8 User Evaluations of TA2 Concepts. Public Deliverable from the EU project, TA2: Together Anywhere Together Anytime (ICT-214793).

[Steuer, 1992] Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. Journal of communication, 42(4), 73-93.

[van Bel et al., 2009] Van Bel, D. T., Smolders, K. C. H. J., IJsselsteijn, W. A., & de Kort, Y. (2009). Social connectedness: concept and measurement. Intelligent Environments, 2, 67-74.

[Vastenburg et al., 2011] Vastenburg, M., Romero Herrera, N., Van Bel, D., & Desmet, P. (2011, May). PMRI: development of a pictorial mood reporting instrument. In CHI'11 Extended Abstracts on Human Factors in Computing Systems (pp. 2155-2160). ACM.

[Vermeeren et al., 2010] Vermeeren, A. P., Law, E. L. C., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010, October). User experience evaluation methods: current state and development needs. In Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (pp. 521-530). ACM.

[Witmer & Singer, 1998] Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. Presence, 7(3), 225-240.

[Kong, 2018] Kong Y. (2018). User experience in social virtual reality: Exploring methodologies for evaluating user experience in social virtual reality. Master Thesis, Delft University of Technology. http://resolver.tudelft.nl/uuid: 0501aba7- 57e0- 4b01- 9e4a- 8ebfd1c96185

References for the Interviews

[Elizabeth, 2012] Elizabeth B-N Sanders and Pieter Jan Stappers. 2012. Convivial toolbox: Generative research for the front end of design. BIS Amsterdam

2.2. Objective Evaluation

Introduction

Objective data, describing different aspects of user’s behavior, such as head and body motion, content navigation, and audio-visual interaction with other users, can be collected in an automatic way while the user is using a social VR application. We refer to this data collection as “objective evaluation”, since an evaluation of high-level user experience and behavior can be inferred by the collected data.

For example, it is known that some objective metrics, such as the frequency of the speech and visual interaction between two or multiple users are directly correlated with the level of social interaction that is happening in between them. Therefore, extracting the visual field of view of each user and detecting how often users look at each other, or analyzing the audio channels of all users to detect the frequency of alternate talking, can give objective indication of the level of social interaction.

Related Objective Metrics

This section details the objective metrics that can be used to assess aspects of user experience with social VR systems. In general, objective metrics of presence and co-presence rely on the assumption that, if someone feels present in a virtual environment and co-present with other people in that environment, he/she will respond to the environment, people and events within that environment in the same way as he/she would in reality. Therefore, these metrics require the implementation of additional logging tools and the presentation of specific events or tasks in the virtual environment, often constraining the design of the experience. Events are meant to trigger a measurable response, for instance, the increase in arousal associated to a physical threat. Tasks are meant to assess user performance in a controlled manner and help to answer the question of how well a given system solves the problem it was designed for. As a consequence, aspects of the experience, such as content, may need to be tailored so that a given measurement can be collected.

The metrics are organized according to the UX dimensions introduced in the subjective evaluation section: quality of interaction, social meaning (co-presence in particular), and presence. We further categorized the metrics into physiological (relative to how the body functions), behavioral (relative to how a person behaves), and user performance measures.

Quality of the Interaction

We focus on quality of the interaction as a measure of how well a social VR system accomplishes communication tasks between end users.

a. Performance metrics:

System efficacy:

- Completion time and volume of information: The application support a controlled task, such as solving as many simple collaborative puzzles - or any short task requiring the interaction between the users of a system - as possible in an allocated time interval. An alternative design is to measure how long users take to finish a fixed set of tasks (completion time). While the most significant results might come from comparing the proposed interface with a state of the art or other commonly used interface (e.g. skype or face to face), secondary results may consider whether the puzzle requires good spatial representation (e.g. video-based VR lacks motion parallax, could this affect performance?) or benefit from 1:1 visualization scale. To some extent, performance depends on the human to human bandwidth that the system affords. Example: [Pejsa et al., 2016] proposed a conferencing system, an experiment comparing it with Skype and face to face is presented.

- Error rate, precision and accuracy: How well the user performs an assigned task. It may serve as an indicator of miscommunications caused by the system. For instance, skype does not present 1:1 scale, this can cause misinterpretation of the physical properties of content presented through this specific system.

Presence

Objective metrics of presence rely on the assumption that if someone feels present in a virtual environment (s)he will respond to the environment, people and events within that environment in the same way as (s)he would in reality. Therefore, these metrics generally rely on whether the user's physiological, behavioral, and performance responses were realistic considering the situation (s)he was submitted to. Note that to provoke strong/measurable responses we often need to insert content and/or events that are specific to the particular characteristics of that measure to the virtual experience.

a. Physiological Measures

Skin conductance:

- Galvanic Skin Response (GSR): variation in arousal in response to a threat or other highly stimulating situations.

Heart rate:

- Increase in heart rate: increase in mean heart rate in response to a threat or other uncomfortable situations.

The analysis of physiological measures often consists of computing the change from a baseline reading, before a specific event, to a reading following the time of the event. It generally presents a slow response, requiring the acquisition for a time window of a few seconds before and after the event. We can compute a summary statistic from each interval and use the difference between them as the response variable. If a single system is being evaluated, statistical analysis can be performed on this value alone to know if the physiological response is the one we expect (based on literature). Note, however, that it is not trivial to define the expected response. A more appropriate experiment design would include a reference setting (e.g. the real experience or a second system that is known to be effective). In addition, the inter-subject variability for these measures tends to be high, favoring a within-subject experiment design. Example: [Meehan et al., 2002] used the aforementioned measures in a VR experience including a pit room (a room that provides visual and/or tactile feedback to the user indicating that (s)he is standing at the boarder of a pit).

b. Behavioral measures

Movement patterns:

- Postural response: For instance, dodging an apparent physical threat.

Analysis: may consist of evaluating the magnitude and correctness of body sway in response to a stimulus and comparing it with the response for a reference system. Example: head motion behavior in response to a descending ceiling fan in [Pomes and Slater, 2013].

Co-presence (togetherness)

There are no clear standard or recurrent methods to objectively assess co-presence in literature. Generally, we can start from a similar assumption to the one we make for presence: if we obtain similar physiological and behavioral responses for a real situation and a mediated social VR system, then the system affords a high level of co-presence. One of the goals of the VR together project is to propose and test new objective metrics to assess co-presence.

a. Physiological measures

Skin conductance:

- Galvanic skin response: Socially uncomfortable situations, such as breaking social norms (subverting expectations), cause alterations in arousal. For instance, having your intimate space invaded by a strange person, or being stared at by a stranger. Example: [Llobera et al., 2010] demonstrated the increase in arousal as computer-controlled characters (agents) walk towards the user. The increase was proportional to the distance that the agents would stop from the users, i.e. the closer to the user, the stronger the GSR. However, the effect was similar for non-anthropomorphic agents.

b. Behavioral measures

Movement patterns:

- Motion synchronization: Literature suggests that co-actors synchronize movements when interacting together. Example: [Llobera et al., 2016] performed an experiment exploring the synchronization of co-actors.
- Mimicking a co-actor: Literature demonstrates nonconscious mimicry of posture and behaviors of social partners [Chartrand and Bargh, 1999], which can be used to assess the social engagement of co-actors. In [Forbes et al., 2016], subjects had to replicate a task presented by a virtual agent. They conclude that an engaged agent, which recognizes the presence of the user by reacting to him/her, influenced the trajectory of a replication task that the user had to perform. That is, users replicated a task performed by the agent more closely when the agent demonstrated a social behavior. We note, however, that the effect size was very small.

Gaze patterns:

- Joint attention: when users simultaneously react to a shared object/content in the environment as the focus of interest (e.g. TV, whiteboard).
- Eye contact: when users stare at one another.
- Averted attention: when users are not acting jointly, i.e. measure the lack of engagement.
- Averted gaze: when a user deviates the gaze to preserve a social distance with another social entity, it is normally related to unknown people, e.g. when crossing with a stranger in the street or confined in a crowded public space people tend to avert gaze to compensate physical proximity and preserve social distance. In the context of an It indicates a social pressure imposed by the second person/agent [Pelphrey et al., 2004].

- Analysis: Joint attention, Eye contact and Deviated gaze can be computed as a proportion of the total experiment/experience time in which these actions were performed. The measure can indicate engagement and recognition of a social agent. Example: some of the gaze metrics are explored in [Pelphrey et al., 2004].

Speech Activity:

- Conversational turnover: when users switch turns in a conversation. It may indicate engagement and fluidness of the conversation and high information throughput.
- Speech interruptions: when a user abruptly interrupts another user and forces a conversational turnover. If intentional, it may indicate a good conversation dynamic. Unintentional interruptions may be a consequence of high latency affecting the inter-subject perception-action coordination.
- Speech semantic: as a measure of whether the system enforces a natural conversation exchange and is perceived as transparent to the user, e.g. the system is not the subject of the conversation.
- Voice tone: as a measure of emotional response to content and/or to the interaction with other users.

Analysis: see [Smith and Neff, 2018] for an example.

3. Performance measures

Reaction time:

- Reduction in reaction time in a competitive task: a competitive task between co-actors. Example: the whac-a-mole game described in [Fribourg et al., 2018], motivate users to react quickly in recognition of the other social entity and may indicate engagement with a co-actor.

Memory encoding:

- Effect of co-actor in memory encoding: Acting side-by-side with a co-actor in a task changes the encoding of memories. Example: [Eskenazi et al., 2013] [Elekes et al., 2016] and [Wagner et al., 2017] have shown an enhancement of memory encoding of information irrelevant to one's own task, but relevant to a co-actor task, when performing the task simultaneously and in the same room.

Adoption and analysis in the pilots

Overall, we selected the following signal data for the objective evaluation of a social VR experience in pilot 1 actions:

- The speech of the user captured from the user's microphone (D1)
- The viewport visualized via the HMD by the user over time (D2)
- The positions and rotations associated to the user's head poses over time when wearing the HMD (D3)
- The visual (RGB) recording of the user body in the real world (D4)
- The visual (RGB) recording of the user body used as input for user representation in the system (D5)
- The depth recording of the user body used as input for user representation in the system (D6)

An overview of these signals is provided in Table 1, the associated metrics of quality of experience, presence and co-presence that we expect to assess with these signals are provided in Table 2. We will evaluate the validity of these metrics as indicators of quality of experience, presence and co-presence based on current literature as well as on the consistency with user's reported experience (i.e. subjective evaluation).

Moreover, we expect to add the following objective data for the objective evaluation of a social VR experience in pilot 2 actions:

- The eye gaze of the user if supported by available HMD hardware (D7)
- The positions and rotations associated to the torso and limbs poses of the user over time (D8)
- The skin conductance of the user during the experience (D9).
- The heart beats of the user during the experience (D10).
- The time performance of the user associated in a social task supported by the system (D11).
- The error performance of the user associated in a social task supported by the system (D12).

Table 1- List of signals that might be collected in the project during pilot 1 and 2 actions

ID	Signal	Source	Current Software	Tested on:				Pilot 1	Pilot 2
				TNO system	TVM system	PC system	other systems		
D1	User's speech	HMD microphone	OBS client (Exp-CWI3)	Yes	Yes	Yes	Facebook spaces	Yes	Yes
D2	User's viewport	Screen capture	OBS client (Exp-CWI3)	Yes	Yes	Yes	Facebook spaces	Yes	Yes
D3	HMD poses	HMD built-in tracking (inertial sensors and camera)	Custom software external to application (Exp-CWI3)	Yes	No	No	Facebook spaces	Yes	Yes
D4	User's body RGB video feed	External camera (Logitech camera)	Logitech camera SW (Exp-CWI3)	Yes	-	-	Facebook spaces	Yes	Yes
D5	User's body RGB video feed 2	Kinect or realsense camera	Custom software external to application	No	No	No	No	Yes	Yes
D6	User's body depth video feed	Kinect or realsense camera	Custom software external to application	No	No	No	No	Yes	Yes
D7	User's eye gaze	Not defined (eye tracking add-on)	Custom software integrated	No	No	No	No	No	Maybe

			to application						
D8	User’s torso and limbs poses	External tracking (e.g. controllers)	Custom software external to application	No	No	No	No	No	Yes
D9	User’s skin conductance	Not defined (Arduino based if low cost)	Openvibe	No	No	No	No	No	Yes
D10	User’s heart beats	Not defined (Arduino based if low cost)	Openvibe	No	No	No	No	No	Yes
D11	Task completion time	does not apply	Task and logging integrated to application	No	No	No	No	No	Yes
D12	Task error rate, precision and accuracy	does not apply	Task and logging integrated to application	No	No	No	No	No	Yes

Table 2- Associations of signals to metrics and related quality of experience constructs.

ID	Associated metrics	Measured constructs (presumed)	Notes
D1	Speech activity	Co-presence, Quality of interaction	how often, how much, how efficiently users communicate
D2, D3, D7	Gaze patterns	Co-presence	patterns related to what the users is looking at
D2, D3, D7	Gaze movement features	Presence, Co-presence	gaze behaviors and response to events in the environment and actions of other users
D6, D8	Movement features	Presence, Co-presence	postural behavior and response to events in the environment and actions of other users
D4. D5	Movement semantics	Co-presence	visual inspection of actions, such as mimicking or pointing
D9	Event related galvanic skin response	Presence, Co-presence	Increase in skin conductance following a meaningful event (risk or social interaction related)
D10	Event related changes in heart rate	Presence	Variations to heart rate following a stressful event.
D11, D12	System efficacy	Quality of interaction	How well the system allows users to achieve an objective interaction goal

References

[Meehan et al., 2002] Meehan, M., Insko, B., Whitton, M., & Brooks Jr, F. P. (2002). Physiological measures of presence in stressful virtual environments. ACM Transactions on Graphics (TOG), 21(3), 645-652.

[Pomes and Slater, 2013] Pomés, A., & Slater, M. (2013). Drift and ownership toward a distant virtual body. Frontiers in human neuroscience, 7, 908.

[Llobera et al., 2010] Llobera, J., Spanlang, B., Ruffini, G., & Slater, M. (2010). Proxemics with multiple dynamic characters in an immersive virtual environment. ACM Transactions on Applied Perception (TAP), 8(1), 3.

[Llobera et al., 2016] Llobera, J., Charbonnier, C., Chagué, S., Preissmann, D., Antonietti, J. P., Ansermet, F., & Magistretti, P. J. (2016). The Subjective sensation of synchrony: an experimental study. PloS one, 11(2), e0147008.

[Chartrand and Bargh, 1999] Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: the perception–behavior link and social interaction. Journal of personality and social psychology, 76(6), 893.

[Forbes et al., 2016] Forbes, P. A., Pan, X., & Hamilton, A. F. D. C. (2016). Reduced mimicry to virtual reality avatars in Autism Spectrum Disorder. Journal of autism and developmental disorders, 46(12), 3788-3797.

[Pelphrey et al., 2004] Pelphrey, K. A., Viola, R. J., & McCarthy, G. (2004). When strangers pass: processing of mutual and averted social gaze in the superior temporal sulcus. Psychological science, 15(9), 598-603.

[Smith and Neff, 2018] Smith, H. J., & Neff, M. (2018). Communication Behavior in Embodied Virtual Reality. CHI 2018.

[Fribourg et al., 2018] Fribourg, R., Argelaguet, F., Hoyet, L., & Lécuyer, A. (2018, March). Studying the Sense of Embodiment in VR Shared Experiences. In IEEE Virtual Reality and 3D User Interfaces (pp. 1-8).

[Eskenazi et al., 2013] Eskenazi, T., Doerrfeld, A., Logan, G. D., Knoblich, G., and Sebanz, N. (2013). Your words are my words: effects of acting together on encoding. Q. J. Exp. Psychol. 66, 1026–1034. doi: 10.1080/17470218.2012.725058

[Elekes et al., 2016] Elekes, F., Bródy, G., Halász, E., & Király, I. (2016). Enhanced encoding of the co-actor's target stimuli during a shared non-motor task. The Quarterly Journal of Experimental Psychology, 69(12), 2376-2389.

[Wagner et al., 2017] Wagner, U., Giesen, A., Knausenberger, J., & Echterhoff, G. (2017). The Joint Action Effect on Memory as a Social Phenomenon: The Role of Cued Attention and Psychological Distance. Frontiers in psychology, 8, 1697.

[Pejsa et al., 2016] Tomislav Pejsa, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andy Wilson. (2016). Room2Room: Enabling Life-Size Telepresence in a Projected Augmented Reality Environment. In Proc. of ACM Conference on Computer Supported Cooperative Work (CSCW 2016).

2.3. Objective Performance Metrics

For the evaluation of the experiments / pilot actions of VR-Together it is also important to take into account objective performance metrics. This section lists potential metrics to be used in the project to give an idea of the performance of the whole VR-Together platform and of its key components (described in D2.1).

Apart from *what* to measure (i.e., the metrics themselves), this section describes *why* to measure these metrics (i.e. their relevance), *where* to measure them (i.e. for which components they need to be measured), *when* to measure them (i.e. measurement granularity and number of samples) and *how* to measure and report on them (i.e., measurement methodology and units).

After having listed and described all the metrics and their measurement, the selected metrics used in pilot 1 will be indicated. The selection is based on key performance issues. Based on the obtained results and the involved complexity of the measurements, the selection of the other metrics will be discussed after pilot 1.

List of Metrics

OP1: Delay

Why to measure it?

It is a key metric in VR-Together to guarantee a feeling of immediacy and interactive communications.

Where to measure it?

It can be measured: i) for each time-sensitive component; ii) for each part of the end-to-end platform (server-side, network-side, cloud-side (processing in networked elements: orchestration, MCU…), client-side); and iii) most importantly, what really matters is the **end-to-end delay** (from capture to rendering), which partially determines the **startup delay** (from starting app or pressing the play button to actually watching/hearing the contents).

When to measure it?

It can be measured per Media Unit (e.g. for video) or period interval (e.g. for audio) basis, and throughout the duration of the session.

How to measure and report on it?

Delay is quite easy to measure. For **end-to-end delays**, the whole end-to-end chain should be considered (i.e. from capture to rendering) in order to report on measurements that accurately resemble the perceived delays.

OP2: Jitter

Why to measure it?

It is an important metric, because it can have an impact on the smoothness, continuity and natural evolution of the media playout process, and even result in Media Units being discarded because of late arrival. To a given extent, jitter (i.e. delay variation) can typically be compensated for by using proper playout buffering strategies. Large playout buffers will be able to compensate for high jitter values, but will also add extra latency to the service, which is clearly undesired. Therefore, having knowledge about the typical magnitudes of jitter in the envisioned scenarios in VR-Together will help in setting the proper playout buffering strategies. This will allow guaranteeing a smooth and natural playout process, while minimizing the end-to-end delays.

It can be assumed that jitter will not have a big impact in the VR-Together scenarios, due to the employed delivery technologies, such as DASH (segments of a longer duration than the expected jitter magnitudes) and WebRTC (multiplexed streams). Anyway, its measurement can provide interesting insights about the magnitudes of jitter when delivering VR contents in the envisioned scenarios and about how properly setup the playout buffers, as mentioned.

Where to measure it?

It can be measured: i) for each time-sensitive component; ii) for each part of the end-to-end platform (server-side, network-side, cloud-side (processing in networked elements: orchestration, MCU…), client-side); and iii) mostly important, what really matters is the end-to-end jitter, which is the accumulated jitter at the client-side and the magnitude that needs to be compensated for before playout. The compensation of jitter at the server- and network-sides can also contribute to this process.

When to measure it?

It can be measured per Media Unit (e.g. for video) or period interval (e.g. for audio) basis, and throughout the duration of the session.

How to measure and report on it?

As for delay, jitter is quite easy to measure. For **end-to-end jitter**, the whole end-to-end chain should be considered in order to report on measurements that accurately resemble the perceived jitter.

OP3: Media Unit (MU) Loss Rate

Why to measure it?

It is an important metric, because it will have an impact of the continuity of the media playout, and can result in specific pieces of contents not being presented to the end-users. As mentioned, large jitter values can also contribute to the MU loss rate, so as to prevent from too large end-to-end delays.

Having knowledge on the occurrence of MU losses and the specific loss rate is very relevant, because it is a sign of network and/or end-systems congestion, for which either preventive or reactive actions should be taken (e.g., decreasing the number of streams, selecting lower qualities…).

Packet or MU loss can also be compensated by enabling retransmission techniques, which is an intrinsic feature in TCP connections, used both in DASH and WebRTC delivery. The number and rate of retransmissions can also be measured and reported. In VR-Together scenarios, it will be relevant to adopt a trade-off between latency minimization and retransmission strategies. If latency does matter, re-transmission should be fast or even avoided, even at the cost of potential MUs not being presented because of losses.

Even with the existence of retransmissions, MU loss rate can happen due to end-system's congestion.

This metric should be measured for traditional 2D video and audio. For 3D video formats, a comparison between the generation and reconstruction rates (explained later) can give insights about the MU rate.

Where to measure it?

It can be measured: i) for each bandwidth-sensitive component; ii) for each part of the end-to-end platform (server-side, network-side, cloud-side (processing in networked elements: orchestration, MCU…), client-side); and iii) mostly important, what it really matters is the MU loss rate measured at the client-side, which is the accumulated value for all components of the end-to-end chain. TCP retransmission and playout buffering strategies to compensate for the effect of MU losses occur at this step of the end-to-end chain.

When to measure it?

It can be measured per Media Unit (e.g. for video) or period interval (e.g. for audio) basis, and throughout the duration of the session.

How to measure and report on it?

MU loss rate is easy to measure, by using ids (e.g. sequence numbers, timestamps, flags…) contained in the Media Units. This is possible both for DASH and WebRTC (although media streams could be encrypted in this case).

The total number of MU losses and the loss rate can be measured.

OP4: Bandwidth Consumption Metrics

Why to measure them?

Bandwidth consumption related metrics need to be measured to gain insights about the bandwidth-related requirements in the VR-Together scenarios, as well as about the achieved performance and quality.

VR-Together includes interactive bi-directional scenarios, so both upload and download rates become relevant. In addition, bandwidth is typically measured in bps, but it can also be useful to measure it in MU/s to acquire a deeper knowledge about the streams being transmitted and received. In relation to this, the following metrics will be used:

OP4.1. Upload Rate [bps]

OP4.2. Download Rate [bps]

OP4.3. Upload MU Rate [MU/s]

OP4.4. Download MU Rate [MU/s]

OP4.5. Reconstruction Rate [MU/s]

OP4.1 and OP4.3 refer to the upload rate in bps and MU/s, respectively. OP4.2 and OP4.4 refer to the download rate in bps and MU/s, respectively. These metrics can be used for audio and video streams. In addition, *OP4.5. Reconstruction Rate* can be used for 3D video.

A comparison between the upload and download rate can also give information about the MU loss rate.

Apart from the bandwidth consumption for each media stream, the total amount of bandwidth being used is also important. Accordingly, an additional metric will be used for it:

OP4.6. Total Bandwidth [bps]

Similarly, the amount of traffic being transmitted/received apart from the bandwidth used for media delivery is also important. This traffic can be used for control information, signalling, synchronization, interactions with the environment, etc. A metric for it has been also considered:

OP4.7. Traffic Overhead [bps]

Where to measure it?

These bandwidth-related metrics can be measured: i) for each bandwidth-sensitive component; ii) for each part of the end-to-end platform (server-side, network-side, cloud-side (processing in networked elements: orchestration, MCU…), client-side); and iii) mostly important, what it really matters is the bandwidth consumption at the client side.

When to measure it?

It can be measured per Media Unit (e.g. for video) or period interval (e.g. for audio) basis, and throughout the duration of the session.

How to measure and report on it?

As mentioned, the bandwidth-related metrics can be measured per Media Unit basis, or per period interval, based on the amount of uploaded / downloaded data.

OP5: DASH Quality

Why to measure it?

When using DASH for media delivery, the media contents are commonly available in multiple qualities. Thus, the client can adaptively select the best quality for each independent segment, based on the current network and end-system conditions. The selected quality is an important metric, as it is a KEY factor that determines the perceived media (video and audio) quality.

Moreover, the transitions between selected qualities for consecutive segments should be smooth, as aggressive transitions can have a negative impact on the perceived Quality of Experience (QoE), especially when selecting lower qualities. In this process, the duration of segments and the designed/adopted quality switching algorithms will play a key role.

By registering the selected quality for each segment, two related metrics can be measured a posteriori [Bentaleb et al., 2016]:

- Average Video Quality: it represents the total average quality of the downloaded segments.

- Video Quality Switches: it represents the magnitude of the quality differences between successively downloaded segments. The frequency of quality switches is also an interesting parameter.

Where to measure it?

It must be measured at the client side, where DASH segments are requested and downloaded. It can be measured either at the network client or player components, depending on the implementation.

When to measure it?

It must be measured for each selected/downloaded DASH segment.

How to measure and report on it?

It can be measured at the client side, where the DASH logic is implemented. If this module cannot be accessed, the DASH quality can be measured by inspecting the incoming DASH segments at the network client component.

It is enough with the measurement of the DASH quality in a quality index or in bps, together with a DASH segment id (e.g. timestamp, sequence number) and/or the current time. Then, the proper conversions and calculations can be done.

OP6: Clock Synchronization Accuracy [ms; time]

Why to measure it?

Accurate Clock Synchronization is very relevant for accurate delay measurements, media synchronization and scheduling of tasks.

Relevant aspects related to Clock Synchronization are: clock offsets, skews (i.e. deviation rates), drifts (i.e. non-linear fluctuations), the added traffic overhead, and the magnitude and frequency of adjustments to achieve synchronization.

Depending on the employed clock synchronization technology or protocol (e.g. NTP, PTP, ad-hoc solutions….), and of their settings, different accuracy levels can be obtained.

Where to measure it?

It can be measured for each component of the platform handling synchronization-related tasks (e.g. timestamps insertion or interpretation, playout process…), and especially between the clocks used by end-systems or entities involved in a VR-Together scenario. In particular, these components are: Synchronization components at the Capturing and Playout blocks, MCU component at the Delivery block (when present) and Session Manager component at the Orchestrator block.

When to measure it?

It must be periodically measured, with a high granularity (e.g. every 0.2s).

How to measure and report on it?

It can be easily measured by periodically registering clock instances at the involved components.

It can be reported as an array of time values, obtained from the employed global clock, per time interval. Then, the proper calculations can be done.

OP7: Intra-Media Sync [ms]

Why to measure it?

It is strongly related to delays and jitter, as it refers to the time differences between the real presentation times of Media Units when compared to their nominal or ideal value. It also has a big impact on the smoothness, continuity and naturalness of the media playout, and can result in either specific pieces of contents not being presented to the end-users or in playout interruptions/stalls (freezing effects).

Having knowledge on the magnitude on the intra-media sync accuracy (i.e. of end-to-end jitter) will allow adopting the proper synchronization and playout buffering strategies, thus preserving the original temporal dependences at the playout side, while minimizing end-to-end delays.

Where to measure it?

It can be measured: i) for each component of the platform actually processing media streams; ii) for each part of the end-to-end platform (server-side, network-side, cloud-side (processing in networked elements: orchestration, MCU…), client-side); and iii) mostly important, what it really matters is the intra-media synchronization accuracy or asynchrony at the client-side, which is the accumulated value for all components of the end-to-end chain.

When to measure it?

It can be measured per Media Unit (e.g. for video) or period interval (e.g. for audio) basis, and throughout the duration of the session.

How to measure and report on it?

Intra-media asynchrony is easy to measure, by using ids (e.g. sequence numbers, timestamps, flags…) contained in the media streams. This is possible both for DASH and WebRTC (although media streams could be encrypted in this case).

It will be measured in ms per measurement interval (e.g. ms/MU).

OP8: Inter-Media / Inter-Source Sync [ms]

Why to measure it?

It refers to the end-to-end delay differences between correlated media streams. It can be measured by using the inserted timestamps for each Media Unit at the capturing side and then inspecting them at the client side. This metric is very relevant to enable coherent media sessions and to avoid confusion and users' annoyance (e.g. audio not aligned to the scenes, offset between streams in stereoscopic video…).

Where to measure it?

It can be measured: i) for each component of the platform actually processing media streams; ii) for each part of the end-to-end platform (server-side, network-side, cloud-side (processing in networked elements: orchestration, MCU…), client-side); and iii) mostly important, what it really matters is the inter-media synchronization accuracy or asynchrony at the client-side, which is the accumulated value for all components of the end-to-end chain.

When to measure it?

It can be measured per Media Unit (e.g. for video) or period interval (e.g. for audio) basis, and throughout the duration of the session.

How to measure and report on it?

Inter-media asynchrony is easy to measure, by using ids (e.g. sequence numbers, timestamps, flags…) contained in the media streams. This is possible both for DASH and WebRTC (although media streams could be encrypted in this case).

It will be measured in ms per measurement interval (e.g. ms/MU).

OP9: Inter-Device / Inter-Destination Sync [ms]

Why to measure it?

It refers to the end-to-end delay differences between the same or correlated media streams being played out on different devices. It can be measured by using the inserted timestamps for each Media Unit at the capturing side and then inspecting them at the client side. This metric is very relevant to provide coherent and interactive media sessions, and to avoid users' annoyance (e.g. the same media scenes or elements being presented at different times at each one of the involved clients).

Where to measure it?

It will be measured at the client side, as this is the point where end-to-end delays can be measured, and ideally compensated for. Similarly, if a MCU component is involved, it can be helpful to measure this metric there.

When to measure it?

It can be measured per Media Unit (e.g. for video) or period interval (e.g. for audio) basis, and throughout the duration of the session.

How to measure and report on it?

Inter-destination asynchrony is easy to measure, by using ids (e.g. sequence numbers, timestamps, flags…) contained in the media streams. If no inter-device / inter-destination synchronization solution is adopted, it can still be helpful to measure the end-to-end delays, and then compare them a posteriori to acquire knowledge about the sync levels that have been achieved. This can help in understanding the requirements of the VR-Together scenarios.

Its measurement is possible when using both for DASH and WebRTC (although media streams could be encrypted in this case).

It will be measured in ms per measurement interval (e.g. ms/MU).

OP10: Playout Process Metrics

Metrics about the playout process are relevant, as they have an impact on the latency, perceived quality and user's perceived QoE. The following metrics will be taken into account:

OP10.1. Playout Buffer Occupancy

Why to measure it?

It is an important metric that has an impact on the probability of playout interruptions and on the latency of the service. A low playout buffer occupancy can result in buffer underflow situations, and thus in playout interruptions. A high playout buffer occupancy can result in buffer overflow situations, and thus in playout interruptions, some pieces of contents not being presented, and in a high end-to-end latency.

Where to measure it?

It must be measured at the client side, where DASH segments are downloaded and buffered for their ultimate presentation.

When to measure it?

It must be measured for each downloaded/buffered DASH segment.

How to measure and report on it?

It can be measured at the client side, where the playout buffer logic is implemented.

The playout buffer occupancy should be reported in time units (e.g. ms), although it can also be reported in terms of size or number of DASH segments.

OP10.2. Playout Interruptions

Why to measure it?

It is an important metric, because it will have an impact on the continuity of the media playout process, and can result in either playout pauses (i.e. freezing effects) or playout skips (i.e. specific pieces of contents not being presented to the end-users).

The playout interruptions can occur due to high/low playout buffer occupancy levels (e.g., because of network and/or end-system congestion), due to end-system congestion situations (e.g. high CPU load), or due to playout adjustments to achieve synchronization.

Where to measure it?

It must be measured at the client side, by periodically monitoring the playout time evolution. Its measurement is strongly related to the ones for the jitter and intra-media synchronization accuracy metrics.

When to measure it?

It must be periodically measured, with a granularity equal or higher that the length of DASH segments (e.g. every 0.2s).

How to measure and report on it?

It can be measured at the client side, by monitoring the evolution of playout times.

It can be reported as an array of playout times per time interval. Then, the proper calculations can be done.

OP10.3. Playout Adjustments

Why to measure it?

The frequency and magnitude of playout adjustments are key factors that determine the perceived QoE.

This metric is strongly related to the playout buffer occupancy and playout interruptions metrics.

However, the playout adjustment reported here are the ones performed due to the clock and media synchronization processes.

Where to measure it?

It must be measured at the client side, at the Synchronization component, although it can also be measured by periodically monitoring the playout time evolution.

When to measure it?

It must be periodically measured, either for each Media Unit or with a granularity, or every time playout adjustments are calculated and enforced at the Player component.

How to measure and report on it?

It can be measured at the client side, by monitoring the calculations made by the Synchronization component and the evolution of playout times at the Player component.

It can be reported as an array of playout adjustments (in ms) per time interval. Then, the proper calculations (mean, max and min values, distribution of values, frequency…) can be done.

OP11: Computational Resources Metrics

It is important to measure how much computational resources are required / used when running the VR-Together scenarios. The usage of resources can be measured in time (e.g. duration of CPU intensive periods, processing delays…), in percentage or in terms of memory usage (bytes).

OP11.1. Processing Time

Why to measure it?

It is related to the OP1 (delays), but specifically focused on the time it takes to conduct a specific task for a specific component. It is important to measure it in order to know about the magnitudes of delay sources along the end-to-end chain and about the performance / limitations of specific components.

Where to measure it?

It must be measured at each component of the VR-Together platform where data-intensive processes are being executed (e.g. encoding / decoding, scene compositions), as detailed in the associated table.

When to measure it?

It can be measured per period interval basis, and throughout the duration of the session.

How to measure and report on it?

It can be measured by monitoring the input and output times for each involved component.

OP11.2. Processing type

Why to measure it?

Apart from the processing time, it is also useful to indicate the processing type being used when reporting on *OP11.1. Processing Time*. The indications about where, when and how to measure it do not apply to this metric.

OP11.3 and OP11.4. CPU Usage [% and bytes]

Why to measure it?

The measurement of the CPU usage when performing data-intensive tasks along the end-to-end chain is important to know about the requirements of specific components and the resources that are needed to meet them. The CPU usage can be 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) will also be measured.

Where to measure it?

It must be measured at each component of the VR-Together platform where data-intensive processes are being executed (e.g. encoding / decoding, scene compositions, rendering…). This is detailed in the associated table in this sub-section.

When to measure it?

It can be measured per period interval basis, and throughout the duration of the session.

How to measure and report on it?

It can be measured by monitoring the CPU load when performing specific data-intensive tasks at each involved component.

OP12: RAM Usage

Same rationale and methodology than for OP11, but in terms of RAM Usage in bytes.

OP13: GPU Usage

Same rationale and methodology than for OP11, but focused on the GPU Usage when performing video- or graphic-related processing tasks. It will be measured and reported in terms of percentage (OP13.1) and bytes (OP13.2).

OP14: Video Quality Metrics

It is important to objectively measure the video quality at the client side in order to determine the achieved performance and infer the user's perceived QoE and satisfaction. In VR-Together, different video formats, including traditional and immersive ones, are being considered for both the shared virtual environment and the end-users' representation. Each video format may have its own or most adequate video quality metrics, and determining the most adequate ones for 3D and immersive video contents is currently a hot research topic, addressed in VR-Together. Examples of typical metrics for traditional formats are Peak signal-to-noise ratio (PSNR) and Structural similarity (SSIM), which have their variants for stereoscopic / 3D and 360º video, and examples of metrics for Point Clouds are the number and density of points, Level of Details, compression level, etc. The appropriateness of each metric, and the need of re-defining / extending the existing ones for the considered video formats, will be considered in the experiments of VR-Together.

Where to measure them?

It must be measured at the Player and Self-Stream Render components, P1 and P2 of the generic component diagram presented in the deliverable D2.4, and also reflected in Table 4.

When to measure them?

It can be measured per period interval basis, and throughout the duration of the session.

How to measure and report on them?

Each selected metric can have its own measurement method (e.g., for each video frame, or recording the video and then computing it…) and unit (e.g., in dBs, a single number within a given range…).

OP15: Audio Quality Metrics

It is important to objectively measure the audio quality at the client side in order to determine the achieved performance and infer the user's perceived QoE and satisfaction. In VR-Together, different audio formats, including traditional and immersive/spatial ones, are being considered. Each audio format may have its own or most adequate audio quality metrics (e.g. PESQ, variants for immersive / spatial audio formats…). The appropriateness of each metric, and the need of re-defining / extending the existing ones for the considered audio formats, will be considered in the experiments of VR-Together.

Where to measure them?

It must be measured at the Player and Self-Stream Render components, P1 and P2 of the generic component diagram presented in the deliverable D2.4, and also reflected in Table 4.

When to measure them?

It can be measured per period interval basis, and throughout the duration of the session.

How to measure and report on them?

Each selected metric can have its own measurement method (e.g., for each audio sample, or recording the audio and then computing it…) and unit (e.g., in dBs, a single number within a given range…).

To conclude this sub-section, Table 3 lists the previously introduced metrics, by assigning a code to them, and indicating which ones will be taken into account in Pilot 1 (also highlighted in bold). In addition, Table 4 relates the metrics to the targeted components of the VR-Together metrics in which they can be measured. An overview of the components is presented in the deliverable D2.4.

Table 3- List of and Codes for the Objective Performance Metrics

Metric Code	Metric Name	Used in Pilot 1
OP1	**Delay**	**Yes**
OP2	**Jitter**	**Yes**
OP3	Media Unit (MU) Loss Rate	-
OP4.1	**Upload Rate [bps]**	**Yes**
OP4.2	**Download Rate [bps]**	**Yes**
OP4.3	**Upload MU Rate [MU/s]**	**Yes**
OP4.4	**Download MU Rate [MU/s]**	**Yes**
OP4.5	**Reconstruction Rate [MU/s]**	**Yes**
OP4.6	**Total Bandwidth [bps]**	**Yes**
OP4.7	Traffic Overhead [bps]	-
OP5	Dash Quality [bps or Quality Index]	-
OP6	Clock Synchronization	-
OP7	Intra-Media Sync	-
OP8	Inter-Media / Inter-Source Sync	-
OP9	Inter-Device / Inter-Destination Sync	-
OP10.1	Playout Buffer Occupancy	-
OP10.2	Playout Interruptions	-
OP10.3	Playout Adjustments	-

OP11.1	**Processing Time**	**Yes**
OP11.2	**Processing Type**	**Yes**
OP11.3	**CPU Usage [%]**	**Yes**
OP11.4	**CPU Usage [bytes]**	**Yes**
OP12	**RAM Usage [bytes]**	**Yes**
OP13.1	**GPU Usage [%]**	**Yes**
OP13.2	**GPU Usage [bytes]**	**Yes**
OP14	Video Quality Metrics	-
OP15	Audio Quality Metrics	-

Table 4- Objective Performance Metrics to be Measured for the Components of the VR-Together Platform

Component (- Platform Part)	**Performance Metrics**
(C1) Visual Sensor data input - Capturing	**(OP4.1) Upload Rate [bps], (OP4.3), Upload MU Rate [MU/s], (OP11.1) Processing Time, (OP11.2) Processing Type, (OP11.3) CPU Usage [%], (OP11.4) CPU Usage [bytes], (OP12) RAM Usage [bytes], (OP13.1) GPU Usage [%], (OP13.2) GPU Usage [bytes]**
(C2) Audio sensor data input - Capturing	**(OP4.1) Upload Rate [bps], (OP4.3), Upload MU Rate [MU/s]**
(C3) Content reconstruction - Capturing	-
(C4) Synchronization - Capturing	**(OP1) Delay, (OP2) Jitter**, (OP6) Clock Synchronization, (OP7) Intra-Media Sync, (OP8) Inter-Media / Inter-Source Sync
(E1) Encoder – Enc/Encap	**(OP4.1) Upload Rate [bps], (OP4.3), Upload MU Rate [MU/s], (OP11.1) Processing Time, (OP11.2) Processing Type, (OP11.3) CPU Usage [%], (OP11.4) CPU Usage [bytes], (OP12) RAM Usage [bytes], (OP13.1) GPU Usage [%], (OP13.2) GPU Usage [bytes]**
(E2) Encapsulator – Enc/Encap	-
(E3) Packager – Enc/Encap	-
(D1) Ingest brick - Delivery	-
(D2) Web Server - Delivery	-
(D3) Multi Control Unit (MCU) - Delivery	**(OP1) Delay, (OP2) Jitter**, (OP3) Media Unit (MU) Loss Rate, **(OP4.1) Upload Rate [bps], (OP4.2) Download Rate [bps], (OP4.3) Upload MU Rate [MU/s], (OP4.4) Download MU Rate [MU/s], (OP4.5) Reconstruction Rate [MU/s], (OP4.6) Total Bandwidth [bps]**, (OP4.7) Traffic Overhead [bps], (OP6) Clock Synchronization, (OP7) Intra-Media Sync, (OP8) Inter-Media / Inter-Source Sync, **(OP11.1) Processing Time, (OP11.2) Processing Type, (OP11.3) CPU Usage [%], (OP11.4) CPU Usage [bytes], (OP12) RAM Usage [bytes], (OP13.1) GPU Usage [%], (OP13.2) GPU Usage [bytes]**

(O1) Connection Manager - Orchestration	-
(O2) Session Manager - Orchestration	**(OP1) Delay, (OP2) Jitter**, (OP3) Media Unit (MU) Loss Rate, (OP6) Clock Synchronization
(O3) Stream Manager - Orchestration	**(OP11.1) Processing Time, (OP11.2) Processing Type, (OP11.3) CPU Usage [%], (OP11.4) CPU Usage [bytes], (OP12) RAM Usage [bytes]**
(O4) Metadata Constructor - Orchestration	-
(O5) Session Logic Manager - Orchestration	-
(O6) Non-live Content Manager - Orchestration	-
(P1) Player (Renderer) – Play-out	**(OP1) Delay, (OP2) Jitter**, (OP3) Media Unit (MU) Loss Rate, **(OP4.1) Upload Rate [bps], (OP4.2) Download Rate [bps], (OP4.3) Upload MU Rate [MU/s], (OP4.4) Download MU Rate [MU/s], (OP4.5) Reconstruction Rate [MU/s], (OP4.6) Total Bandwidth [bps]**, (OP4.7) Traffic Overhead [bps], **(OP5) Dash Quality [bps or Quality Index]**, (OP6) Clock Synchronization, (OP7) Intra-Media Sync, (OP8) Inter-Media / Inter-Source Sync, **(OP11.1) Processing Time, (OP11.2) Processing Type, (OP11.3) CPU Usage [%], (OP11.4) CPU Usage [bytes], (OP12) RAM Usage [bytes], (OP13.1) GPU Usage [%], (OP13.2) GPU Usage [bytes]**
(P2) Self-Stream Renderer – Play-out	**(OP4.2) Download Rate [bps], (OP4.4) Download MU Rate [MU/s], (OP4.5) Reconstruction Rate [MU/s], (OP13.1) GPU Usage [%], (OP13.2) GPU Usage [bytes]**
(P3) Network Client – Play-out	**(OP4.6) Total Bandwidth [bps]**, (OP4.7) Traffic Overhead [bps], **(OP5) Dash Quality [bps or Quality Index]**, (OP10.1) Playout Buffer Occupancy, (OP10.2) Playout Interruptions, (OP10.3) Playout Adjustments
(P4) Demuxer – Play-out	-
(P5) Content Decoder – Play-out	**(OP4.2) Download Rate [bps], (OP4.4) Download MU Rate [MU/s], (OP4.5) Reconstruction Rate [MU/s], (OP11.1) Processing Time, (OP11.2) Processing Type, (OP11.3) CPU Usage [%], (OP11.4) CPU Usage [bytes], (OP12) RAM Usage [bytes]**
(P6) Synchronization – Play-out	**(OP1) Delay, (OP2) Jitter**, (OP3) Media Unit (MU) Loss Rate, (OP6) Clock Synchronization, (OP7) Intra-Media Sync, (OP8) Inter-Media / Inter-Source Sync, (OP9) Inter-Device / Inter-Destination Sync
(P7) Metadata Extractor – Play-out	-

References:

[Bentaleb et al., 2016] A. Bentaleb, A. C. Begen, R. Zimmermann, "SDNDASH: Improving QoE of HTTP Adaptive Streaming Using Software Defined Networking", ACM on Multimedia Conference (MM '16), October 2016, Amsterdam (The Netherlands)

2.4. Added-Value Evaluation

For the evaluation of added value, the main focus is on providing questionnaires to professionals. The project will use this questionnaire, or a version derived from this, at trade shows and events like IBC and VRDays. The focus with professionals is not so much on the experience, but on the consumer market.

Another way of measuring added value is by slowly moving towards a market introduction. Even though it is still too for a real market introduction, discussing about the opportunities after a demonstration of the system is a good way of better understanding the added value of the system. This strategy has led the project to, for example, a try out at an IT company.

The project has developed an Added Value questionnaire (for the Advisory Board), based on a number of iterations and through presence in a number of relevant events. It is included in Annex I.

3. EXPERIMENTS

During the project we run a number of pilot actions (12) that help us construct the trial, evaluate the innovation value of the system and provide us technical requirements. We have divided them into three main categories:

- Technology requirements: typically with our without users, with a focus on the technology
- Experience design and evaluation: with users, with a focus on their experience
- Innovation and Entertainment value: with professionals, with a focus on the added value

3.1. Technology Evaluation

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

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

3.1.1. CWI-1

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

Methodology

We perform a subjective experiment using mixed methodology to evaluate point cloud compression quality. Our quantitative study aims to collect the ground truth subjective scores to evaluate our metrics with. Our qualitative study aims to explore in more detail user's perception of point cloud quality, and uncover some quality dimensions that we may consider in our objective metrics. The employed Consent Form in this experiment can be found in Annex II.

We first explain the dataset that we use to conduct our experiments, and continue with the setup of our quantitative and qualitative study.

Dataset

Figure 2 shows the point cloud sequences that we use in our subjective study. We use 6 frames from 6 different sequences of full-body humans from [Eugene, 2017]. We limit our experiments to these sequences, and do not include other types of sequences (for example, non-human objects, or close-up human objects) available through other datasets for the following reasons. Firstly, we wish to avoid content categories becoming a confounding factor in our analysis. Secondly, sequences taken from different datasets often vary in perceptual quality due to different acquisition techniques. Each frame of point clouds from the dataset is compressed into 4 different levels of compression using the compression algorithm in [Mekuria, 2017]. The compression parameters used are Level of Detail (LoD) 10, LoD 9, LoD 8, and LoD 7. LoD 10 means that the compression uses a 10-b octree setting, i.e., 10 quantization bits per direction. In the rest of this report, we refer to compression with LoD 10 as compression level 1, and so on, compression with the lowest LoD (LoD 7) as compression level 4. We obtain a total of 24 point cloud sequences to be rated by our participants. During the rendering of the point clouds, we assign different point sizes for each compression level, such that each object can be seen in full (i.e., no hollow parts due to missing points can be seen).

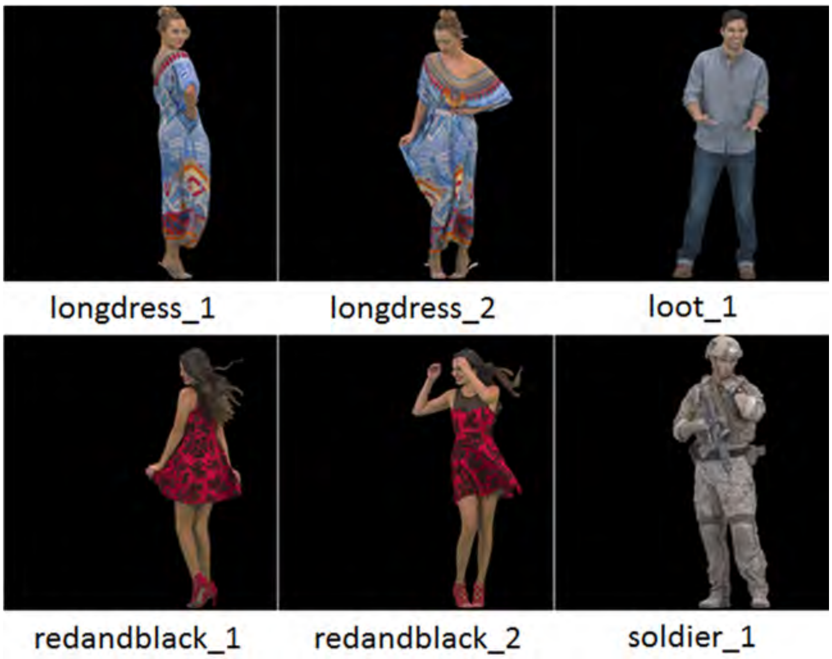


Figure 2- Point cloud sequences used in the experiment, from the publicly available 8i Voxelized Full Bodies (8iVFB v2) dataset [Eugene, 2017]

We then created video sequences for each point cloud (reference and compressed), which show the point cloud rotated along the vertical axis. This allows users to have a 360 degree view of the point clouds. The resolution of the video sequences is 1280x720, and were 40 seconds long each, with 30 frames per second. Figure 3 illustrates some of the viewpoints taken from one of the resulting test videos.

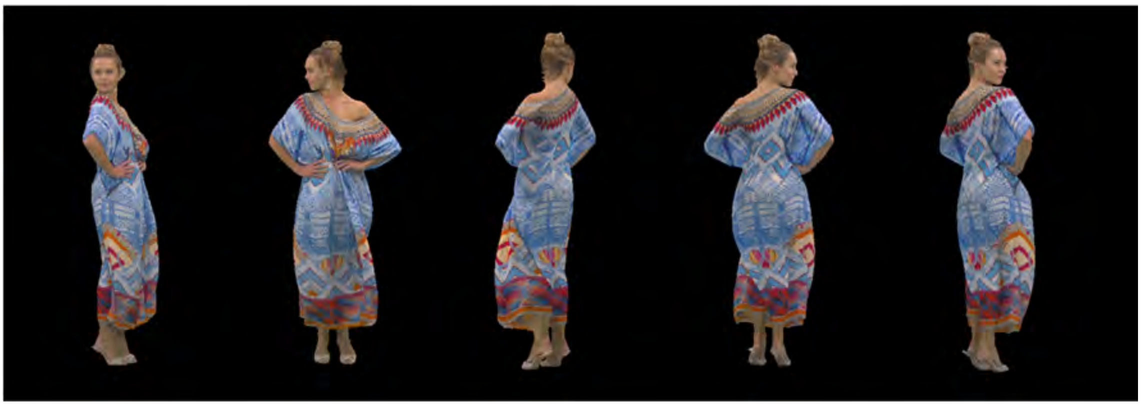


Figure 3- In the test video sequences, each point cloud was rotated along the vertical axis such that users had a 360 degree viewpoint of it

Quantitative Subjective Study

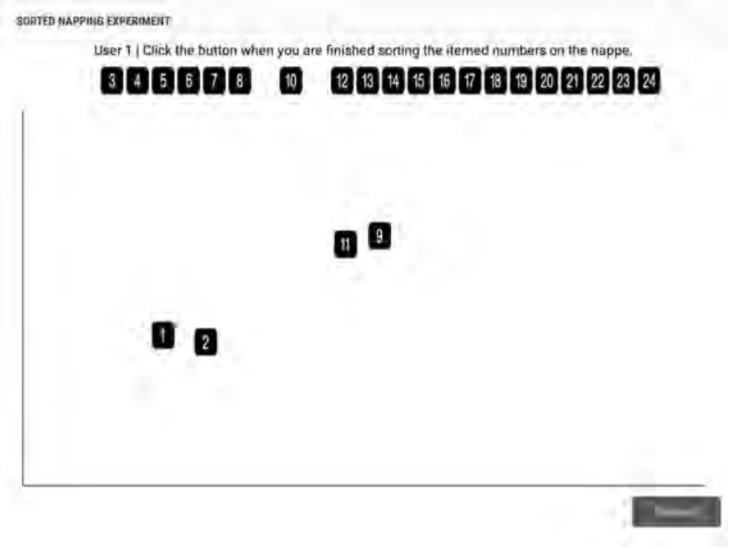
We perform a quantitative subjective experiment to obtain ground truth subjective scores of our point cloud sequences. 23 people participated in the experiment, 6 of which are female and 17 are male. The participants' age ranges from 22 to 33 years old, and most of them naive to point clouds. A training session was given before the test, so that participants could familiarize themselves with the task and rating interface. We used different point clouds from our test set for the training session. Users had full control of the time taken to evaluate a test point cloud, as they had to press a button on the test screen to display the next pair of test sequences. A 32 inch DELL UHD 4K monitor was used in the experiment, and users were seated at a distance of twice the height of the monitor, compliant with the recommendation in [ITU-T P. 910]. The experiment room had controlled lighting as recommended in [ITU-R BT. 500-13].

Participants were asked to evaluate the quality of point clouds using a Simultaneous Double Stimulus presentation, in which the reference point cloud is shown alongside the test point cloud. Participants were then asked to rate the level of degradation for the test point cloud using a 5-point Degradation Category Rating (DCR) scale [ITU-T P. 910]. Following the current standard in subjective assessments [ITU-T P. 910], the point cloud sequences were shown in a non-interactive manner, meaning that participants could not interact with the stimulus and change its viewpoint or its scale. We created two different viewpoint sequences for our point clouds, to prevent the collected scores biasing one particular viewpoint. The viewpoints rotate the point clouds along the vertical axis, and slowly zooms in or out on the point clouds at certain moments. A reference point cloud would always be shown with the same viewpoint as the test point cloud.

Qualitative Subjective Study

A qualitative subjective study is performed to allow users to express more freely what they perceive when asked to judge point cloud quality, and give us a better understanding of how they perceive point cloud quality. In this experiment, we use the Descriptive Sorted Napping method ([Cadoret, 2010], [Le, 2015]). This method presents users or participants with a number of test items, and asks them to sort and group the items on a Nappe, or a blank space, based on how similar or dissimilar users find them. The closer two items are placed on the Nappe, the more similar they are perceived by a user, and vice versa. After all items are placed on the Nappe, users are asked to draw a circle around the different groups of items they have decided on, and write down or explain the similar characteristics attributed to items in the same group.

This method was originally developed in the food sciences field [Cadoret, 2010], and recently has been used in the quality assessment field to explore the way users construct their judgment of Quality of Experience (QoE) ([Strohmeier, 2010], [Strohmeier, 2013]). We use this method to learn what attributes users would associate with degraded point clouds when asked to group point clouds based on their similarity in quality. Participants were given an interface through a tablet device in which they are presented with a set of numbered blocks and a blank space that represent the Nappe. On a big computer screen, participants could watch the point cloud sequences, each corresponding to a numbered block on the tablet. Participants could then place the numbered blocks on the Nappe based on how similar they perceive the quality of point clouds corresponding to each block. Users were allowed to watch the point cloud sequences multiple times, and were not restricted in time to complete the task. Once users were done placing the numbered blocks on the Nappe, they could draw circles around numbered blocks that they consider to belong in the same group, and type in the attributes or characteristics related to the perceived quality of that group. This is done until all numbered blocks have been included in at least one group. This whole process was explained to participants before they started the task, and they were instructed specifically to sort and group the point clouds based on the perceived quality. Figure 4 shows the interface used for the study.



SORTED NAPPING EXPERIMENT

User 1 | Click the button when you are finished sorting the itemed numbers on the nappe.

Figure 4- The interface used to perform the sorted napping experiment. Users could watch the videos corresponding with each number on a computer screen in front of them

24 people participated in the experiment, however, 3 people were excluded from the data analysis since they did not understand the task and categorized the point clouds based on content instead of quality. Our first 5 participants were asked to sort and group all 24 point cloud videos in our dataset, i.e. the whole quality range in our dataset. Preliminary analysis on their data showed that participants separated the bad quality and good quality sequences using descriptions that are biased towards the bad quality sequences (for example, describing only how blurry/blocky the images appear). According to literature ([Strohmeier, 2013]), this happens when a wide test range is introduced to users. We then asked the rest of our participants to sort and group only 12 point clouds which belong to the higher quality range. This distinction was made in order to better understand detailed attributes that users associate with the point cloud quality.

Before analyzing the data, we coded users' responses, such that long full sentences or sentences with repetitions were shortened to emphasize only the adjectives or attribute words in the sentence. For example, a user described a group of sequences as: ”The videos are not colorful and not detailed.” We would shorten the sentence into: ”Video not colorful and detailed”. Another user described his/her groups with how blurry the face or clothes looked. And so, we coded his/her description: ”Both face and clothes are not blurred but in high quality”, into ”Both face and clothes in high quality". In this way, we do not alter the choice of words used by users, but still shorten the sentences to easily compare keywords used across users.

Results

We conducted an outlier analysis on the scores obtained from our quantitative study, according to the recommendation in [ITU-R BT. 500-13]. No outlier was found in the analysis. We then calculated the Mean Opinion Score (MOS) and standard deviation of opinion score (SOS) of each image. To compare the level of user agreement in this task with user agreement in other quality assessment tasks, we calculated the SOS hypothesis alpha [Hossfeld, 2011] of our collected data. The SOS hypothesis alpha represents how the standard deviation of opinion scores (SOS) changes with the mean opinion scores (MOS) values using a parameter α. A higher value of α would indicate higher disagreement among user scores. The level of user agreement obtained in our study lies within the range of alpha values for visual quality assessment tasks. This indicates that the use of Simultaneous Double Stimulus presentation with Degradation Category Rating (DCR) scale yields reliable scores for assessing point cloud compression quality in our study.

Table 5 - Results from CWI-1

Application and Subjective Study	SOS α
Point cloud compression quality assessment (our study)	0.146
Image quality assessment, JPEG images of LIVE dataset [Sheikh, 2006]	0.0400
Image quality assessment, IRCCyN/IVC Scores on Toyama [Tourancheau, 2008]	0.1715
Video streaming quality assessment, H.264 codec [Brandao, 2009]	0.1078
Video streaming quality assessment, MPEG2 codec [Brandao, 2009]	0.1137

Figure 5 plots the MOS distribution across impairment levels and content. At the higher levels of quality (low compression rate), we observe that the content ”Soldier” is significantly rated higher than other contents for the same compression level. We confirm this through an ANOVA test on the sequences with compression levels 1 and 2 separately; opinion scores being the dependent variable and content being the independent variable. For both compression levels, the MOS for content ”Soldier” has a statistically significant difference with the MOS for other contents ($p<0.05$). This may indicate at the higher range of quality, certain characteristics of the content ”Solider” could mask artifacts that users otherwise perceive in other content.

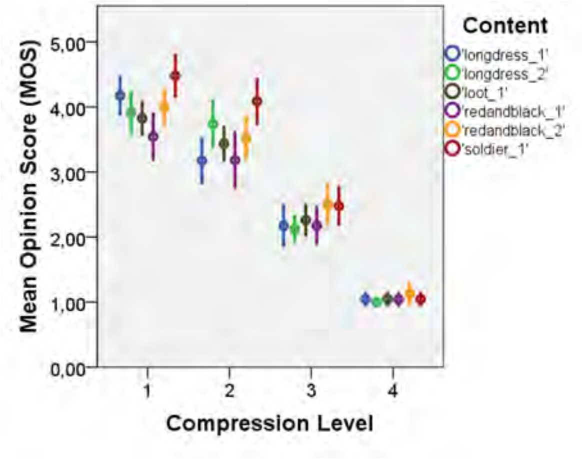


Figure 5- Mean Opinion Scores (y-axis) over the different compression levels (x-axis), and content (colored). Compression level 1 indicates the lowest compression level, and 4 indicates the highest compression level. Error bar indicates 95% confidence interval

Next, we perform a hierarchical multiple factor analysis (HMFA) on our collected data from the qualitative study, to find shared structures among the individual sortings. Figures 6a and 6b shows Confidence Ellipses of the sorted napping configurations obtained through the study on 24 and 12 point cloud sequences, respectively. The study on 24 point cloud sequences includes the whole quality range in our dataset, while the study on 12 sequences only includes the upper-half of the quality range in our dataset. Each colored ellipse in the figures represent the spread of each item along the two dimensions that explain most of the variance of the data. The two dimensions of the plot for the 24 items explains 57.51% of variance in the data together, while the two dimensions of the plot for the 12 items explains 35.59% of variance of the data together. When possible, we place the name of a point cloud sequence at the center of its corresponding ellipse. The suffix ”CLn” at the end of a sequence name indicates the compression level of the point cloud, with n=1 being the lowest compression level (and thus, highest quality).

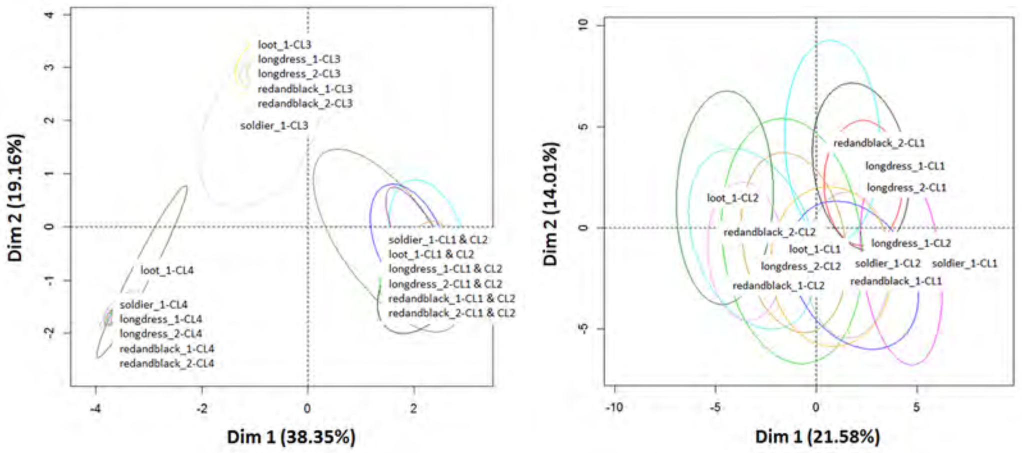


Figure 6- Confidence ellipses of the sorted napping configuration for the study on (a) the whole quality range (24 items) and (b) the upper-half of the quality range (12 items) in our dataset. The suffix ”CLn” indicates the compression level with n=1 indicating the lowest compression level

From the plot for the 24 items, participants seem to agree on three clear separate clusters of items. The two clusters on the negative side of the Dim 1 axis show clusters of point clouds with highest level of compression (most left-side cluster), and point clouds of the second highest level of compression. We look into the attributes that participants associated with these three different clusters, and find that for the high levels of compression (the two lowest quality range), participants noted that the point clouds appear blurry or blocky (due to the large point sizes assigned to the point clouds during rendering).

Meanwhile, the higher quality point clouds were described as being clear. Some participants specifically mentioned facial features and patterns in clothing as cues to judge the clarity of the point cloud videos. As shown in the Confidence Ellipses plot for the 24 items (Figure 6a), the sequences belonging to the higher levels of quality fall into the same cluster at the positive side of the Dim 1 axis. To look more closely at how users perceive the quality of these sequences, we conducted the Sorted Napping study only on these 12 sequences with different participants. The Confidence Ellipse plot for the 12 items shows that there is still no clear separate clusters formed from the 12 items.

Looking at the attributes that participants associated with the sequences that fall into the negative side of the Dim 1 axis for the 12 point clouds, we find that half of the participants mentioned color distortion to describe these sequences. When asked to explain what they meant by color distortion, participants commented that some of the sequences had colors from their clothes projected onto their skin. For example, they could see some red dots in the skin of the woman in the “Red and Black” sequence, or some blueish tones on the skin of the man in the “Loot” sequence.

Analysis

Considering the use of color point clouds in real-world applications, we can conclude that it is important to also evaluate perceived point cloud quality based on colors. In addition, different point cloud compression algorithms may use different color coding schemes, and thus create different color artifacts. Point cloud quality evaluation based on color has not been explored in previous studies as most of the stimuli used were not colored. For this reason, we look into objectively quantifying the quality of compressed point clouds based on their color properties, and later on show how it improves overall quality prediction together with geometry-based metric.

To evaluate the quality of a point cloud, we propose to compare its color distribution information with that of its reference. Luminance, or the perceived brightness of color, has often been used in image quality metrics to estimate users perception of color [Sheikh, 2006]. However, some studies have also suggested the use of hue or chrominance as relevant to quality perception [deSimone, 2009]. We perform our experiments using both luminance and chrominance values of point clouds, and compare the performance of both in predicting point cloud quality. The luminance value is represented by the Y-channel values of the YUV space, while chrominance is represented by the H-channel values of the HSV space.

Further information about the new metric can be found here: https://repository.tudelft.nl/islandora/object/uuid:d0a8f1b0-d829-4a34-be5a-1ff7aa8679ca

References

[Mekuria, 2017] R. Mekuria, K. Blom, and P. Cesar, “Design, implementation, and evaluation of a point cloud codec for tele-immersive video,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 27, no. 4, pp. 828–842, 2017.

[Eugene, 2017] T. M. Eugene d ´ Eon, Bob Harrison and P. A. Chou, “8i voxelized full bodies - a voxelized point cloud dataset,” ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG), Tech. Rep. WG11M40059/WG1M74006, January 2017.

[ITU-T P. 910] I. T. U. (T), “Recommendation ITU-T P. 910,” Subjective video quality assessment methods for multimedia applications, 2008.

[ITU-R BT. 500-13] I. T. U. (R), "Recommendation ITU-R BT. 500-13," Methodology for the subjective assessment of the quality of television pictures, 2009.

[Cadoret, 2010] M. Cadoret, S. Le et al., "The sorted napping: A new holistic approach in sensory evaluation," Journal of Sensory Studies, vol. 25, no. 5, pp. 637–658, 2010.

[Le, 2015] S. Le, T. Le, and M. Cadoret, "Napping and sorted napping as a sensory profiling technique," in Rapid sensory profiling techniques. Applications in New Product Development and Consumer Research. Woodhead Publishing Cambridge, 2015, pp. 197–213.

[Strohmeier, 2010] D. Strohmeier, S. Jumisko-Pyykko, and K. Kunze, "Open profiling of quality: a mixed method approach to understanding multimodal quality perception," Advances in multimedia, vol. 2010, 2010.

[Strohmeier, 2013] D. Strohmeier, K. Kunzem, K. Gobel, and J. Liebetrau, "Evaluation of differences in quality of experience features for test stimuli of good-only and bad-only overall audio visual quality," in Image Quality and System Performance X, vol. 8653, 2013.

[Hossfeld, 2011] T. Hossfeld, R. Schatz, and S. Egger, "Sos: The mos is not enough!" in 3rd International Workshop on Quality of Multimedia Experience (QoMEX 2011). IEEE, 2011, pp. 131–136.

[Sheikh, 2006] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," IEEE Transactions on Image Processing, vol. 15, no. 11, pp. 3440–3451, 2006.

[Tourancheau, 2008] S. Tourancheau, F. Autrusseau, Z. P. Sazzad, and Y. Horita, "Impact of subjective dataset on the performance of image quality metrics," in 15th IEEE International Conference on Image Processing. IEEE, 2008, pp. 365–368.

[Brandao, 2009] T. Brandao, L. Roque, and M. P. Queluz, "Quality assessment of h.264/avc encoded video," in Proc. of conference on telecommunications-ConfTele, Sta. Maria da Feira, Portugal, 2009.

[deSimone, 2009] F. De Simone, F. Dufaux, T. Ebrahimi, C. Delogu, and V. Baroncini, "A subjective study of the influence of color information on visual quality assessment of high resolution pictures," in Proc. Fourth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM-09), no. MMSPL-CONF-2009-001, 2009.

3.1.2. CERTH-1

A very early technical experiment was conducted in order to assess the total distribution performance of the TVM pipeline, allowing for a better understanding of the required improvements for providing connected social VR.

Methodology

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

Experimental Sessions in different configurations

The sessions used in this experiment are detailed in Table 6. The duration of each session was 2 minutes.

Table 6- Sessions used in the CERTH-1 Experiment

Sequence	Voxel Grid Resolution	Texture Downscale
Session2T3	High (64x128x64)	D1 (no downscale)
Session3T3D2	High (64x128x64)	D2 (width/2, height/2)
Session9D4	Low (32x64x32)	D4 (width/4, height/4)
Session10D8	Low (32x64x32)	D8 (width/8, height/8)
Session11	Low (32x64x32)	D1 (no downscale)

Metrics

- Frame rate
- Frame rate in (frame uploading)

Results

Table 7 - Results from CERTH-1 Experiment

Voxel Grid Resolution	Texture Downscale	**Average Bandwidth (uploading) (mb/s)**	**Average Frames per second (fps)**	**Average Bandwidth per frame (mb/s)**
High (64x128x64)	D1 (no downscale)	3.3	10	0.33
High (64x128x64)	D2 (width/2, height/2)	2.09	10	0.209
Low (32x64x32)	D4 (width/4, height/4)	1.51	19	0.079
Low (32x64x32)	D8 (width/8, height/8)	1.12	19	0.059
Low (32x64x32)	D1 (no downscale)	3.1	10	0.31

Analysis

From the statistical analysis of the initial RabbitMQ experiments, using the TVM simulation, we reach the conclusion that real time TVM distribution between remote countries over the web was feasible. The findings of the analysis were:

- TVM frame rate depends on a) the TVM reconstruction rate and b) the available bandwidth. To this experiment, when TVM voxel grid resolution is high (64x128x64), TVM is distributed at 10 FPS, since the reconstruction delay is higher, while for lower resolution (32x64x32), the TVM frequency is at 19 FPS.
- TVM Texture is the main consumer of the bandwidth, since mesh data configuration slightly changes the bandwidth needs.

Finally, it should be mentioned that these experiments have been conducted using the initial versions of the modules for TVM production and transmission, thus, it is significant to achieve better results using the new versions of the components developed for the first pilot.

3.1.1. CERTH-2

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

Methodology

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

Experimental Sessions in different configurations (1 minute)

Sequence	Voxel Grid Resolution	Texture Downscale
CERTH_D1_High	High (64x128x64)	D1 (no downscale)
CERTH_D2_High	High (64x128x64)	D2 (width/2, height/2)
CERTH_D1_Low	Low (32x64x32)	D1 (no downscale)
CERTH_D2_Low	Low (32x64x32)	D2 (width/2, height/2)
VO_D1_High	High (64x128x64)	D1 (no downscale)
VO_D2_High	High (64x128x64)	D2 (width/2, height/2)
VO_D1_Low	Low (32x64x32)	D1 (no downscale)
VO_D2_Low	Low (32x64x32)	D2 (width/2, height/2)

Metrics (see Section 2.3)

- 3D Capture Delay
- 3D Capture Jitter
- TVM Encoding Delay

- TVM Encoding Jitter
- TVM Serialization Delay
- TVM Serialization Jitter
- RMQ Server Delay
- RMQ Server Jitter
- TVM Networking Delay
- TVM Networking Jitter
- TVM Player Delay
- TVM Player Jitter
- Bandwidth TVM Upload Rate [Bytes/s]
- Bandwidth TVM Download Rate [Bytes/s]
- Bandwidth TVM Upload MU Rate [MU/s]
- Bandwidth TVM Download MU Rate [MU/s]
- TVM Reconstruction Rate [MU/s]
- Total TVM bandwidth

Results

CERTH_D1_High				
Metrics at various stages of the TVM			**Bandwidth Metrics**	
3D Capture	**Delay (ms)**	95.74	**TVM Upload Rate [Bytes/s]**	3121751.78
	Jitter (ms)	8.11	**TVM Download Rate [Bytes/s]**	3117251.15
TVM Encoding	**Delay (ms)**	209.77	**TVM Upload MU Rate [MU/s]**	6.225
	Jitter (ms)	32.37	**TVM Download MU Rate [MU/s]**	6.225
TVM Serialization	**Delay (ms)**	1.18	**TVM Reconstruction Rate [MU/s]**	6.83
	Jitter (ms)	5.70	**Total TVM bandwidth**	6239002.938
RMQ Server	**Delay (ms)**	25.2		
	Jitter (ms)	16.02		
TVM Networking	**Delay (ms)**	154.26		

	Jitter (ms)	11.87		
TVM Player (renderer)	**Delay (ms)**	17.18		
	Jitter (ms)	32.52		
Total Time	503.36			

CERTH_D2_High				
Metrics at various stages of the TVM			**Bandwidth Metrics**	
3D Capture	**Delay (ms)**	93.68	**TVM Upload Rate [Bytes/s]**	2900307.5
	Jitter (ms)	5.89	**TVM Download Rate [Bytes/s]**	2910354.286
TVM Encoding	**Delay (ms)**	64.49	**TVM Upload MU Rate [MU/s]**	10.25
	Jitter (ms)	5.60	**TVM Download MU Rate [MU/s]**	10.28
TVM Serialization	**Delay (ms)**	0.74	**TVM Reconstruction Rate [MU/s]**	10.3
	Jitter (ms)	3.40	**Total TVM bandwidth**	5810661.78
RMQ Server	**Delay (ms)**	8.39	**Traffic Overhead [Bytes/s]**	-
	Jitter (ms)	13.00		
TVM Networking	**Delay (ms)**	84.76		
	Jitter (ms)	22.97		
TVM Player (renderer)	**Delay (ms)**	10.99		
	Jitter (ms)	8.12		
Total Time	263.09			

CERTH_D1_Low				
Metrics at various stages of the TVM			**Bandwidth Metrics**	
3D Capture	**Delay (ms)**	66.76	**TVM Upload Rate [Bytes/s]**	2436066.18

	Jitter (ms)	4.96	**TVM Download Rate [Bytes/s]**	2435258.44
TVM Encoding	**Delay (ms)**	186.89	**TVM Upload MU Rate [MU/s]**	6.71
	Jitter (ms)	23.64	**TVM Download MU Rate [MU/s]**	6.68
TVM Serialization	**Delay (ms)**	0.82	**TVM Reconstruction Rate [MU/s]**	6.85
	Jitter (ms)	3.50	**Total TVM bandwidth**	4871324.62
RMQ Server	**Delay (ms)**	14.58		
	Jitter (ms)	21.11		
TVM Networking	**Delay (ms)**	129.69		
	Jitter (ms)	19.71		
TVM Player (renderer)	**Delay (ms)**	16.75		
	Jitter (ms)	20.89		
Total Time	415.53			

CERTH_D2_Low				
Metrics at various stages of the TVM			**Bandwidth Metrics**	
3D Capture	**Delay (ms)**	64.57	**TVM Upload Rate [Bytes/s]**	1967990.78
	Jitter (ms)	4.45	**TVM Download Rate [Bytes/s]**	1966503.33
TVM Encoding	**Delay (ms)**	52.00	**TVM Upload MU Rate [MU/s]**	14.08
	Jitter (ms)	4.12	**TVM Download MU Rate [MU/s]**	14.1

TVM Serialization	Delay (ms)	0.49	TVM Reconstruction Rate [MU/s]	14.71
	Jitter (ms)	1.10	Total TVM bandwidth	3934494.11
RMQ Server	Delay (ms)	3.10		
	Jitter (ms)	80.18		
TVM Networking	Delay (ms)	61.25		
	Jitter (ms)	171.09		
TVM Player (renderer)	Delay (ms)	9.95		
	Jitter (ms)	9.03		
Total Time	191.39			

VO_D1_High				
Metrics at various stages of the TVM			Bandwidth Metrics	
3D Capture	Delay (ms)	94.17	TVM Upload Rate [Bytes/s]	3003071.85
	Jitter (ms)	6.81	TVM Download Rate [Bytes/s]	2639755.3
TVM Encoding	Delay (ms)	207.537	TVM Upload MU Rate [MU/s]	5.94
	Jitter (ms)	30.64	TVM Download MU Rate [MU/s]	6.01
TVM Serialization	Delay (ms)	9.96	TVM Reconstruction Rate [MU/s]	6.71
	Jitter (ms)	30.65	Total TVM bandwidth	5642827.157

RMQ Server	Delay (ms)	30.43		
	Jitter (ms)	91.424		
TVM Networking	Delay (ms)	7839.933		
	Jitter (ms)	3367.541		
TVM Player (renderer)	Delay (ms)	20.3234		
	Jitter (ms)	7.58		
Total Time	8202.37			

VO_D2_High				
Metrics at various stages of the TVM			Bandwidth Metrics	
3D Capture	Delay (ms)	92.584	TVM Upload Rate [Bytes/s]	2905817.28
	Jitter (ms)	6.366	TVM Download Rate [Bytes/s]	2808023.65
TVM Encoding	Delay (ms)	63.5212	TVM Upload MU Rate [MU/s]	10.34
	Jitter (ms)	5.3835	TVM Download MU Rate [MU/s]	10.7
TVM Serialization	Delay (ms)	2.482	TVM Reconstruction Rate [MU/s]	10.37
	Jitter (ms)	11.134	Total TVM bandwidth	5713840.93
RMQ Server	Delay (ms)	5.38		
	Jitter (ms)	21.909		

TVM Networking	**Delay (ms)**	1337.868		
	Jitter (ms)	856.377		
TVM Player (renderer)	**Delay (ms)**	10.349		
	Jitter (ms)	18.608		
Total Time	1512.186			

VO_D1_Low				
Metrics at various stages of the TVM			**Bandwidth Metrics**	
3D Capture	**Delay (ms)**	67.034	**TVM Upload Rate [Bytes/s]**	2382693.745
	Jitter (ms)	5.842	**TVM Download Rate [Bytes/s]**	2335756.491
TVM Encoding	**Delay (ms)**	189.701	**TVM Upload MU Rate [MU/s]**	6.54
	Jitter (ms)	24.156	**TVM Download MU Rate [MU/s]**	6.58
TVM Serialization	**Delay (ms)**	0.7508	**TVM Reconstruction Rate [MU/s]**	6.66
	Jitter (ms)	1.245	**Total TVM bandwidth**	4718450.236
RMQ Server	**Delay (ms)**	1.284		
	Jitter (ms)	1.048		
TVM Networking	**Delay (ms)**	1241.34		
	Jitter (ms)	800.561		
	Delay (ms)	12.397		

TVM Player (renderer)	**Jitter (ms)**	38.474		
Total Time	1512.508			

VO_D2_Low				
Metrics at various stages of the TVM			**Bandwidth Metrics**	
3D Capture	**Delay (ms)**	64.531	**TVM Upload Rate [Bytes/s]**	2049145.893
	Jitter (ms)	5.519	**TVM Download Rate [Bytes/s]**	2032779.173
TVM Encoding	**Delay (ms)**	51.613	**TVM Upload MU Rate [MU/s]**	14.7
	Jitter (ms)	4.813	**TVM Download MU Rate [MU/s]**	14.8
TVM Serialization	**Delay (ms)**	0.543	**TVM Reconstruction Rate [MU/s]**	14.7
	Jitter (ms)	0.673	**Total TVM bandwidth**	4081925.066
RMQ Server	**Delay (ms)**	0.681		
	Jitter (ms)	0.506		
TVM Networking	**Delay (ms)**	117.814		
	Jitter (ms)	34.475		
TVM Player (renderer)	**Delay (ms)**	9.88		
	Jitter (ms)	7.242		
Total Time	245.065			

Analysis

Comparing the above experiments, useful deductions can be derived, helping us figure the best tradeoff between quality and speed. In all the experiments with the RabbitMQ server located in Thessaloniki, the one-third of the total time is consumed in the P3 stage (Network Client) of the pipeline. In contrast in all the experiments conducted with the RabbitMQ server located in France, the P3 stage's time ranges from 48% to 95%. Changing the voxel grid resolution and the texture's downscale factor affects as expected the pipeline's time. More specifically, the most time-consuming stage of the pipeline without texture downscaling is the E1-Encoder stage almost regardless the voxels grid resolution, although the geometry is compressed as well. Oppositely, when the texture is being downscaled by 2 the most time-consuming stage of the pipeline is the C3 (Content Reconstruction). In addition, the derived results clearly indicate that the texture size is the one which mostly affects the total time. The total time is reduced to 50% with the usage of downscaling.

Finally, the best result in both cases is derived from the experiment which the voxels' resolution is set to 32x64x32 and we downscale the texture by two. This is also the only time which is comparable between the two cases. It should be mentioned that the experiments have been conducted using the pilot 1 versions of the modules (v1) for TVM production and transmission, also implementing the official VRT objective performance metrics in the pipeline.

3.1.2. CERTH-3

The goal of this experiment is to compare different hardware configurations, in order to select the most relevant for the pilot.

Devices used in the experiments

Kinect for Xbox One

Figure 7- Kinect for Xbox One Sensor

The time-of-flight system modulates a camera light source with a square wave. It uses phase detection to measure the time it takes light to travel from the light source to the object and back to the sensor and calculates distance from the results. The timing generator creates a modulation square wave. The system uses this signal to modulate both the local light source (transmitter) and the pixel (receiver).

The Xbox One Kinect is the second-generation Microsoft 3D image and audio sensor. It is integral to the Xbox One system. The 3D image and audio sensors and the SoC computation capabilities operating in parallel with games and other applications provide an unprecedented level of voice, gesture, and physical interaction with the system.

Intel RealSense D415

Figure 8- Intel RealSense D415 Sensor

The Intel® RealSense™ Depth Camera D400-Series uses stereo vision to calculate depth. The D415 is a USB-powered depth camera and consists of a pair of depth sensors, RGB sensor, and infrared projector. It is ideal for makers and developers to add depth perception capability to their prototype development.

With its rolling image shutter and narrow field of view, Intel® RealSense™ Depth Camera D415 offers high depth resolution when object size is small and precise measurements are required.

HTC Vive

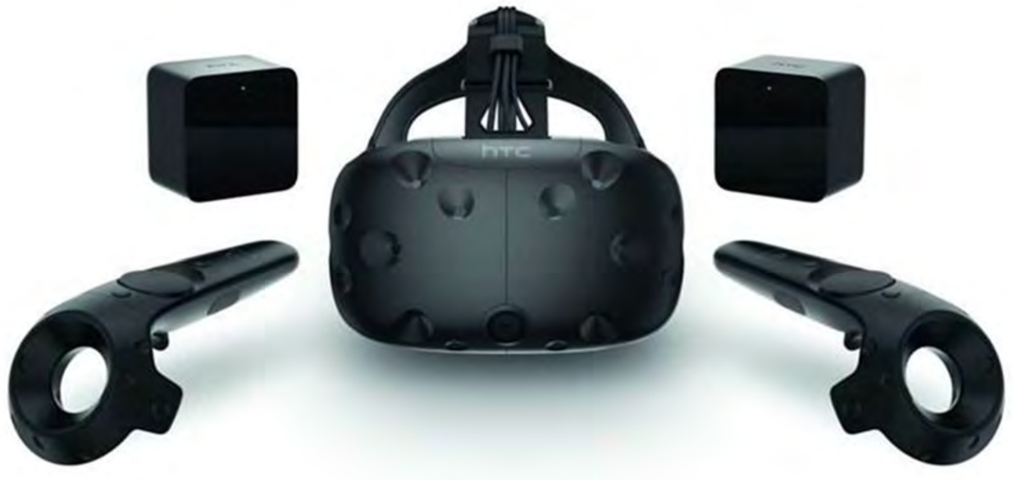


Figure 9- HTC Vive components

Vive Headset: The Vive headset has a refresh rate of 90 Hz and a 110-degree field of view. As well the head set has multiple sensors, the headsets outer-shell has divots, inside these divots are dozens of infrared sensors that detect the base stations to determine the headset's current location in a space. Other sensors include a G-Sensor, gyroscope and proximity sensor.

Vive Controllers: The wireless controllers are the hands of virtual reality, making a more immersive experience for the user. The controller has multiple input methods included a track pad, grip buttons, and a dual-stage trigger. Across the ring of the controller are 24 infrared sensors that detect the base stations to determine the location of the controller.

Vive Base Stations: Also known as the Lighthouse tracking system are two black boxes that create a 360-degree virtual space up to 5m radius. The base stations emit timed infrared pulses at 60 pulses per second that are then picked up by the headset and controllers with sub-millimeter precision.

Oculus Rift

Figure 10- Oculus Rift HMD

Oculus Rift is a HMD developed and manufactured by Oculus VR and Facebook Inc.. Two Pentile OLED displays, 1080×1200 resolution each at 90 Hz and a 110° field of view constitute the main specifications of the device. Rift also support head rotational and positional tracking, while integrated headphones provide a 3D audio effects.

Results

Scenario 1: Kinects + HTC Vive

1. Standard Placement

Placement: 4 Kinect cameras placed around the user. The angle between the cameras and the floor is approximately equal to 0 degrees.

Result: Does not work. When the user enters the capture station, the HTC Vive cannot function properly, the display switches off. When out of the station, the display works again.

2. Other Placements

Placements:

- 4 Kinect cameras placed around the user. High-angle shot (approximately equal to 45 degrees).
- 4 Kinect cameras placed around the user. Low-angle shot (approximately equal to 45 degrees).

  [The difference between the angle-shots is the reflections that the floor can cause to the infrared lights.]

Result: It does not work. When the user enters the capture station, the HTC Vive cannot function properly, the display switches off.

Scenario 2: Kinects + Oculus

Placement: 4 Kinect cameras placed around the user. The angle between the cameras and the floor is approximately equal to 0 degrees.

Result: It works. There were some changes to the 3D coordinate system, but the experience was good. It should be highlighted that this version of Oculus is not the latest commercial one.

Scenario 3 – D415 + HTC Vive

Placement: 4 D415 cameras placed around the user. The angle between the cameras and the floor is approximately equal to 0 degrees.

Result: It works properly.

Scenario 4 – D415 + Oculus

Placement: 4 D415 cameras placed around the user. The angle between the cameras and the floor is approximately equal to 0 degrees.

Result: It works properly.

Analysis

Oculus Rift HMD worked properly with both RGB-D devices, thus, it has been considered the appropriate device for Pilot 1.

3.1.3. CERTH-4

When a user is immersed in sVR, he/she wears a VR HMD, thus the face of the full body 3D user representation is occluded. This results in a major loss for discriminating facial information. The presence of the HMD during multi-user communication in the virtual environment weakens the feeling of co-presence and prevents the user from being fully immersed. The main goal of this experiment was to create a dataset in order to develop, train and evaluate an algorithm that will perform efficient and real-time HMD removal, exploiting the full information medium (i.e., color (RGB) and depth data). A special data capturing system was designed to acquire RGB-D faces with and without HMDs. The dataset will be publicly available and will be utilized for the HMD removal task, in the context of VRTogether.

Methodology

Data acquisition setup: The acquisition setup consisted of three RGB-D sensors placed at an approximate 1.5 meters "head-to-device" distance (as depicted in **¡Error! No se encuentra el origen de la referencia.**1), operating at ~30Hz, and capturing RGB and depthmap frames at *1920x1080* and *512x424* pixel resolution, respectively. The dataset was recorded under controlled environmental conditions, i.e., with negligible illumination variations (no external light source was present during the experiments) and a homogeneous static background. The three sensors captured facial data along with skeleton and eye data from a group of 62 individuals, without gender or age limitations (one subject per session). Each session lasted 2 minutes and was divided in two sub-sessions. At first, the subject started the session by facing the central RGB-D sensor without wearing any HMD and was asked to freely rotate/translate the head and change facial expressions (i.e., open mouth, lift eyebrows, move jaw, etc.). For the second sub-session, the same subject wore a HMD (i.e., FOVE) and was asked to perform the same (not identical, but close enough) head movements and facial expressions. From each sub-session, 200 frames were sampled.

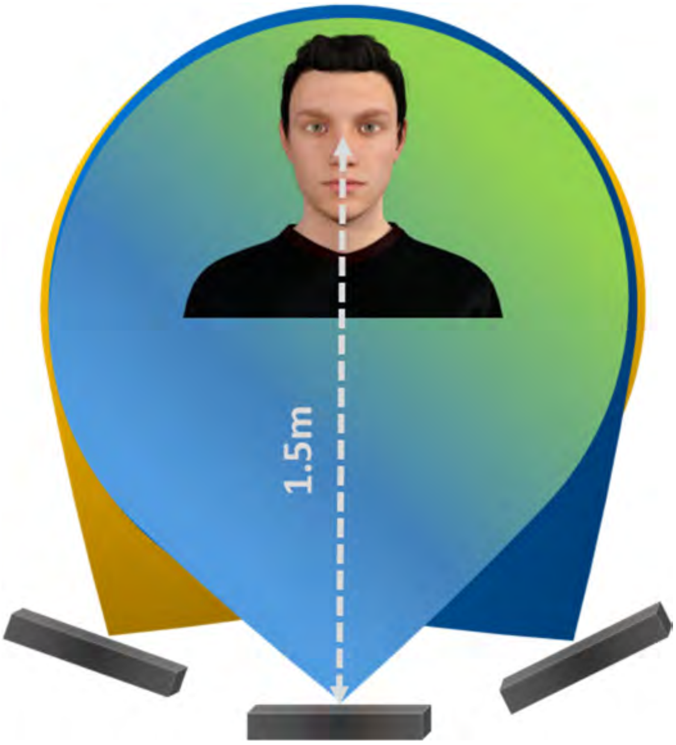


Figure 11- Visualization of the capturing setup

The employed forms in CERTH-4 experiment are included in Annex III.

Resulted Dataset

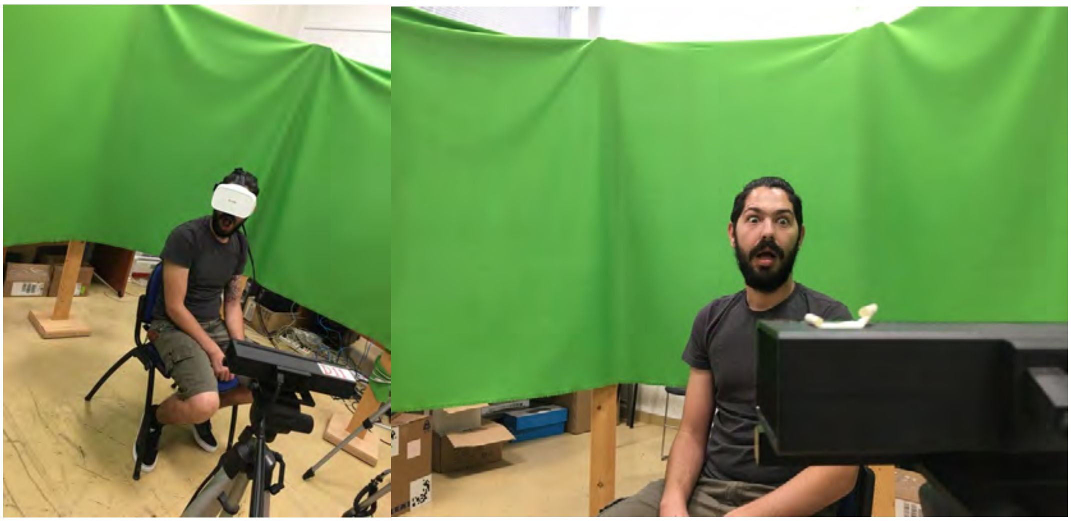


Figure 12 Pictures from the captures.

The following details the acquitted dataset:

- Per RGB-D sensor (on average):

  - 62 subjects
  - 1500 frames per subject without HMD
  - 1500 frames per subject with HMD
  - 1500 skeleton frames per subject without HMD
  - 1500 skeleton frames per subject with HMD
- Per subject (on average):
  - 3000 eye frames captured by the HMD (FOVE)

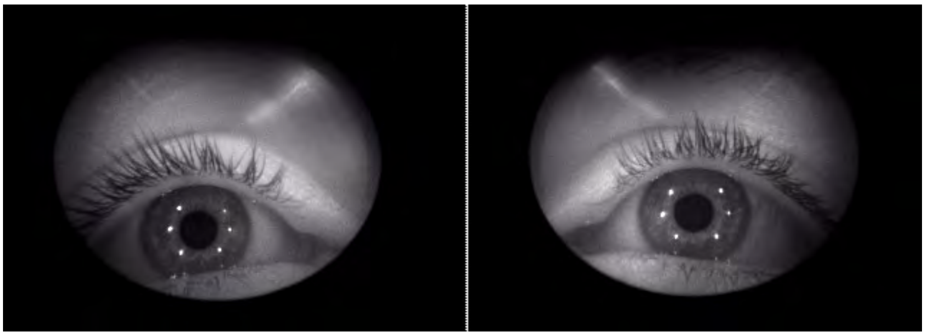


Figure 13 Eye frame example captured by the HMD (FOVE).

3.2. User Experience Evaluation

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

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

3.2.1. ARTANIM-1 and ARTANIM-2

We present two experiments to assess the relative impact of different levels of body animation fidelity of a user controlled virtual avatar (ARTANIM-1) and of a virtual character that is not controlled by the user (ARTANIM-2) to plausibility illusion (Psi). Psi concerns the feeling that events in a virtual environment may be really happening and is part of Slater's proposition of two orthogonal components of presence in virtual reality (VR) [Slater, 2009][Slater et al., 2010]. We emphasize that these experiments only address self and others representation based on 3D rigged meshes, which will be used as a baseline for experiments evaluating self-representation (e.g. compare self-representation using 3D rigged mesh avatars with TVM and point clouds), as well as for content production in the next pilots (e.g. pre-recorded and live actors interacting with the users).

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

In the second experiment (ARTANIM-2) we address the question: to what extend the animation fidelity of a character that is not controlled by the user affects Psi? By improving our understanding of users' perception of character animation we can refine the minimum requirements for offline and live performance capture in the context of the project.

The employed Consent Forms in these experiments can be found in Annex IV.

Methodology

Overview: In the experiments, the face, hands and upper and lower bodies of the avatar (ARTANIM-1) or animated character (ARTANIM-2) were manipulated with different degrees of animation fidelity, such as no animation, procedural animation and motion capture.

Participants started the experiment by experiencing the most complete animation setting, then, the animation features were set to a basic configuration by limiting the amount of captured – tracking and speech – information available to the system. Participants could then improve animation features step by step towards a more complete animation configuration until either the avatar animation realism felt equivalent to the initial – and most complete – condition or all settings were maxed.

Experimental conditions: In the experiment we manipulated four different animation fidelity factors: upper body (UB), lower body (LB), facial (FA) and hands (HA) animation. We chose to divide the whole body into these four factors based on usual separation of tracking equipment. We denote each possible configuration in our experiment by a vector c = [UB, LB, FA, HA]. The manipulations related to the four animation fidelity factors are described below.

Upper Body: upper body animation comprises animation fidelity of pelvis, torso, neck, head, shoulders, arms and wrists.

- (UB=0): only wrists and HMD tracking information is available. The library FinalIK (http://root-motion.com) uses this information to generate valid poses for the upper body joints that are not being tracked.
- (UB=1): pelvis and elbows tracking information is provided in addition to wrists and HMD tracking. The library FinalIK uses this information to generate valid poses for the upper body joints that are not being tracked.
- (UB=2): all tracking information provided by VICON Shogun is used to animate the upper body, that includes wrists, elbows, shoulders, clavicles, chest, neck, head, spine and pelvis tracking information.

Lower Body: lower body animation comprises animation fidelity of feet, knee, hip, and if set at the highest level, the pelvis.

- (LB=0): no tracking information related to the lower body limbs is available. The locomotion functionality provided by FinalIK is used to procedurally animate the legs based on pelvis pose.

- (LB=1): feet tracking information is available. The library FinalIK uses this information to generate valid knee and hip joints rotations. The knee bending axis is perpendicular to the plane defined by the principal component of each foot and the hip joint.
- (LB=2): all tracking information provided by VICON Shogun is used to animate the lower body, that includes feet, knees and hip joints, as well as pelvis pose if not yet available as part of upper body levels 1 and 2.

Facial: facial animation comprises animation fidelity of the eyes and mouth.

- (FA=0): face is not animated, and a static facial expression is used.
- (FA=1): for experiment ARTANIM-1, mouth is animated based on speech captured by a microphone built in the Oculus and gaze is animated based on the relative position of the eyes with regard to a mirror present in the virtual environment. For experiment ARTANIM-2, an iPhone X with the ARKit library was used to track facial expressions of the actor.

Hands: hand animation comprises the movement of fingers and thumb.

- (HA=0): thumb and fingers are not animated and a predefined pose is used instead.
- (HA=1): procedural animation is used to move thumb and fingers when close enough to the object used on subtask 2.
- (HA=2): the bending sensors integrated to the Manus VR are used to animate thumb and fingers (only used for ARTANIM-2).

In total, experiment ARTANIM-1 had 54 possible combinations of animation features (3 for Upper body, 3 for Lower Body, 2 for Face, and 3 for Hands - 3 x 3 x 2 x 3 = 54), while experiment ARTANIM-2 had 36 possible combinations of animation features (3 for Upper body, 3 for Lower Body, 2 for Face, and 2 for Hands - 3 x 3 x 2 x 2 = 36).

Trial Structure: In a trial, participants would start in a basic configuration and were asked to perform configuration transitions until their feeling of animation realism matched that of the most complete configuration. To prevent participants from carelessly moving to the most complete configuration we imposed the following constraints: transitions could only be made in one direction, by increasing the level of an animation feature; only one step could be taken at any given transition, that is, participants could not go from (UB=0) to (UB=2) with a single transition, instead, they had to first reach (UB=1) to then transition to (UB=2); participants had to complete a task (ARTANIM-1) or watch the character animation clip once (ARTANIM-2) before any single transition in a trial; when choosing a feature to improve, participants were urged to reflect on the feature that they were missing the most at that moment, and to improve that first. They were explicitly told that the order with which they improved the animation features was important for the experiment.

ARTANIM-1

Setup: participants were equipped with a motion capture suit and reflective markers to track their movements in real time with a Vicon optical motion capture system. They also wore a pair of Manus VR gloves for fingers tracking and an Oculus consumer version HMD (Figure 14a). During the experiment, participants had to execute a number of tasks (walk, grab an object, speak in front of a mirror), these are detailed below.

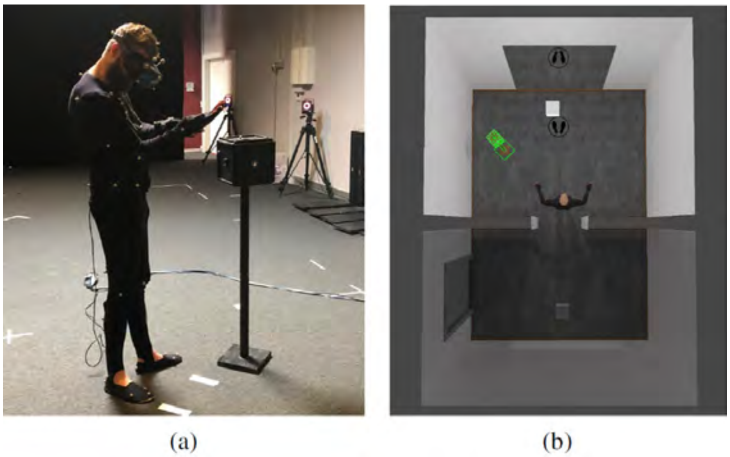


Figure 14: Participant wearing the HMD, ManusVR gloves and the motion capture suit with retro-reflective markers (a) and an overview of the virtual environment (b).

Task: Participants had to complete a task and answer two questions repeatedly, until he/she felt that the animation quality was equivalent to the start condition or no more animation improvements were possible, i.e. the absorbing configuration [UB=2,LB=2,FA=1,HA=2] was reached. The task was divided into two sub-tasks and two questions that the participants had to answer.

The first sub-task consisted of stepping over a pair of footprints on the floor facing the mirror and repeating the phrase "my name is (a), and I am feeling (b)", where they were asked to replace (a) with their name, and (b) with an expression of how they were feeling about the experience and virtual avatar (Figure 15.a.). The experimenter presented words such as "good", "bad", "weird", "OK", "worst" and "better" as suitable alternatives but did not constrained participant's choices to these words. Longer sentence formulations were also allowed. This piece of information was noted by the experimenter and used to understand if something was not working as expected.

The second sub-task consisted of grabbing and moving an object from its current location to a new location, indicated by a spotlight (

Figure 15b.). Participants were asked to look at their hands while carrying the object.

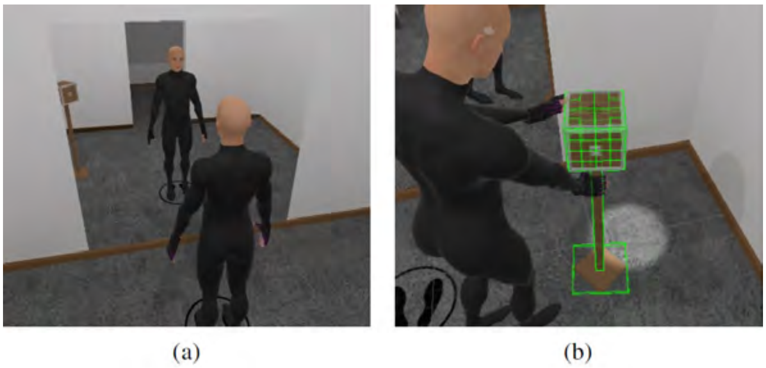
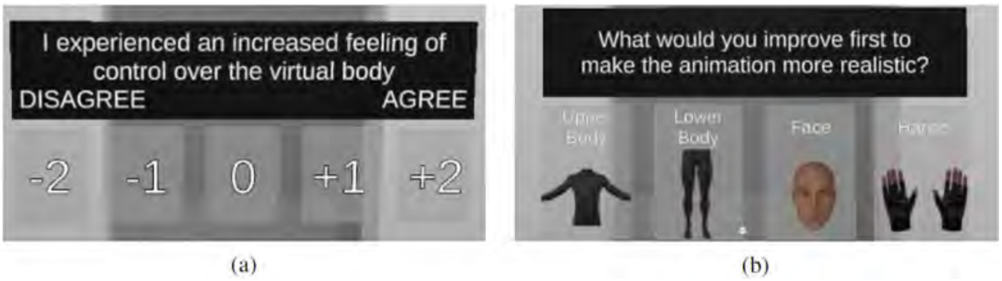


Figure 15: Overview of the two subtasks, in (a) the participant walked to the mirror and repeated a phrase, in (b) the participant had to grab an object and carry it to a spotlight.

The first question concerned how their feeling of control over the virtual body has changed when comparing the current experience to the immediately previous one (Figure 16.a.). Participants had to disagree or agree in a 5-point likert scale to the affirmation "I experienced an increased feeling of control over the virtual body”. In the first trial this meant to compare the most complete setting with a low fidelity setting. As participants are expected to disagree with the affirmation, the answer to this question is used as a control. After that, the comparisons concerned the felt difference in control due to the most recent animation feature improvement.

Finally, the participant was prompted to select an animation improvement option by the question “What would you improve first to make the animation more realistic?” (

Figure 16b.). Participants could either improve an animation feature, or state that the animation realism already felt equivalent to the initial condition.



I experienced an increased feeling of control over the virtual body

DISAGREE AGREE

What would you improve first to make the animation more realistic?

Figure 16: After completing the task, the participant had to report whether the latest change has increased the feeling of control over the virtual body (a), and decide whether and which feature will be improved next (b).

Results

Participants: 24 male individuals participated in the experiment.

Match Configurations: With the match configurations response variable we calculate the probability P(UB=u, LB=l, FA=f, HA=h | match), where P(A|B) represents the probability of A given B. Then, the probability P(match | u, l, f, h) that participants will declare that a given configuration is felt to be equivalent to the most complete condition (i.e. that the configuration feels as realistic as the most complete condition). We also use P(u, l, f, h | match) to assess the marginal probability that a given feature will be active in the matching configuration, for instance, P(u=2|match) = 0.6 describes that, given the feeling of equivalent plausibility, there is a probability of 60% that the upper body feature will be at the animation level (UB=2).

Figure 17 presents the probability of a Psi match P(match | u, l, f, h) in red, and P(u, l, f, h | match) in blue. Note that only configurations with 10 or more occurrences are presented. Considering the configurations preceding the absorbing condition [2,2,1,2], four configurations achieved a probability close or above to 50% of being accepted as a match configuration, P(match | u=0, l=2, f=1, h=2) = .5, P(match | u=1, l=1, f=1, h=2) = .49 and P(match | u=2, l=1, f=0, h=2) = .46, while P(match | u=2, l=1, f=1, h=2) = .66. Moreover, the marginal probabilities that a given animation feature level would be active at the match condition are presented in Table 8.

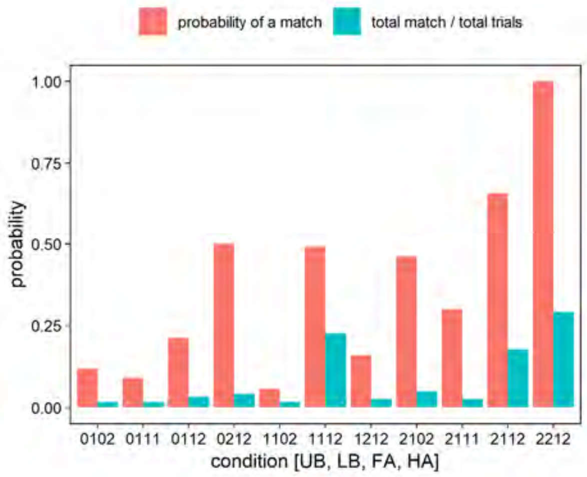


Figure 17: The probability of a Psi match for a given configuration [u, l, f, h], or P(match | u, l, f, h) is presented in red. The probability of a given configuration [u, l, f, h] when the participant felt a Psi match, or P(u, l, f, h | match), is presented in blue. Only configurations with 10 or more occurrences are presented.

Table 8: Probability that any given animation setting will be active in a match configuration.

Feature	Level		
	0	1	2
Upper Body	.117	.283	.6
Lower Body	0	.575	.425
Face	.133	.867	-
Hands	.025	.05	.92

Transitions: Participants were told that the transition order was important, and that they should decide for the feature that they miss the most. With the record of transitions between configurations we can reconstruct the path preferred by participants to go from an initial configuration to a match configuration. We model it with a sparse 54 x 54 matrix containing a total of 135 transition probabilities. With this matrix we represent the transitions as a Markov chain in Figure 18. The following sequence of transitions was the most probable to happen in our experiment: [0,0,0,0] → [0,1,0,0] → [1,1,0,0] → [1,1,0,1] → [1,1,0,2] → [1,1,1,2] → [2,1,1,2] → [2,2,1,2]. This path describes a single improvement for Lower and Upper Body, followed by a maxing the Hand feature to achieve fingers movement control, then maxing the Face animation feature, and finally maxing Upper and Lower Body.

In Figure 18, the representation of the transition probabilities as a Markov Chain is shown. Participants could only move from the nodes in the left to nodes in the right. [2, 2, 1, 2] is an absorbing condition and no transition was possible once it was reached. The thicker the line is, the most probable it is that a transition from a node in the left to a node in the right will be performed. Gray shaded nodes are configurations that have not been visited in any trial.

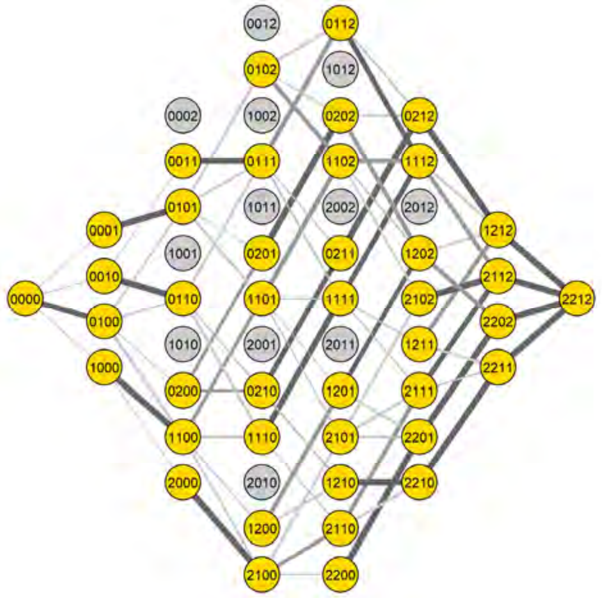


Figure 18: Representation of the transition probabilities as a Markov Chain.

As participants could perform up to 7 transitions, the probabilities of achieving any given configuration after 1, 2, 3, 4, 5 or 6 transitions are presented in Figure 19.

Sense of Agency: To assess the change in sense of control for a given animation factor and level the participant answered to the statement “I experienced an increased feeling of control over the virtual body” in a scale from -2 to 2, where -2 means “Disagree” and 2 means “Agree”. The summary of results, with the average score per participant, is shown as a box and whiskers plot in Figure 20. We run the Wilcoxon signed-rank test to identify whether participants generally agreed that the features added to the sense of control of the virtual body. The responses to all settings expressed statistically significant agreement with the statement (i.e. p < .05), except for (HA=1) (p > 0.7). As shown in Figure 20, (LB=1) received the highest overall scores, but the difference is not guaranteed to be statistically higher than (HA=2) and (LB=2) (both p > .11).

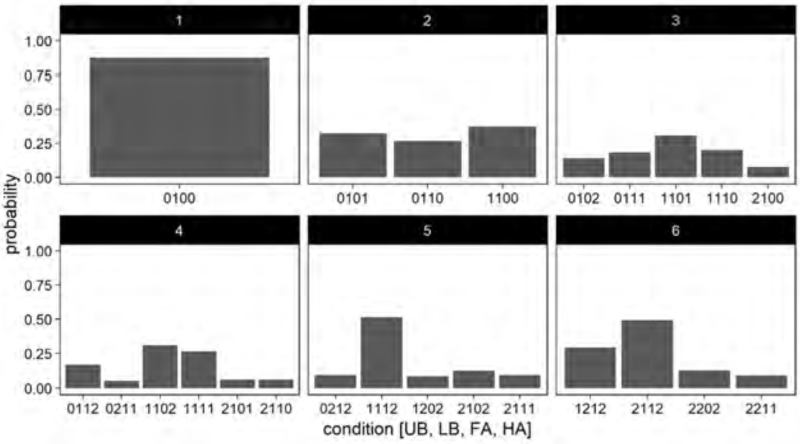


Figure 19: Probability distribution of a given configuration after each transition. Only probabilities greater than 0.05 are presented.

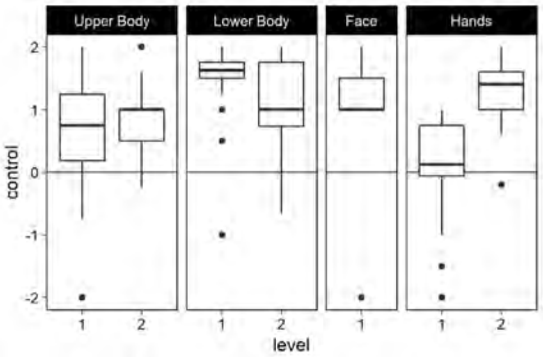


Figure 20: Box and whiskers plot for participants' agreement to the statement "I experienced an increased feeling of control over the virtual body", where 2 means "Agree" and -2 means "Disagree".

Analysis

We found that a virtual body with upper and lower body animated using 5 to 8 tracked rigid bodies and inverse kinematics (IK) was often perceived as equivalent to a professional capture pipeline relying on 53 markers. Compared to what usual VR kits in the market are offering (headset and controllers tracking), feet tracking, followed by the use of the built-in microphone for mouth animation and gloves for finger tracking, were the features that most added to the sense of control of a virtual body, and were often among the first to be improved.

ARTANIM-2

Setup: participants were equipped with an Oculus Consumer Version HMD and Oculus Touch motion controllers. During the experiment, participants had to repeatedly complete a task consisting of watching a clip of pre-recorded motion capture data produced with the same upper and lower body tracking equipment used in ARTANIM-1, with the addition of an iPhone X for face tracking.

Task: the task was divided into two parts, in the first part the participant had to watch a 34 seconds animation clip (part of the content produced for Pilot 1), in the second part participants had to answer a question.

The animation clip shows a police inspector presenting a murder case to the user. In the animation, the inspector walks in the interrogatory room, and shows a photo of the victim while describing the case (Figure 21). This animation clip is an extract of the content produced for the first pilot of the VR Together project.

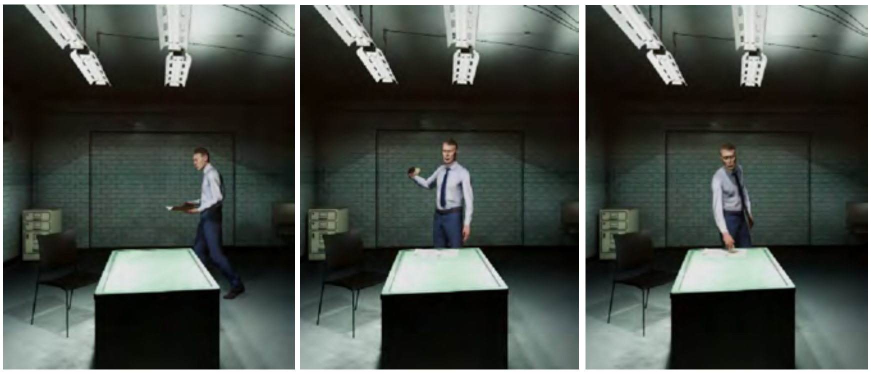


Figure 21: Frames from the animation clip, the detective (virtual character) walks in and presents a photo of the victim.

At the end of the animation clip, the participant was prompted to select an option by the question “What would you improve first to make the animation more realistic?” (Figure 22). Participants could either improve an animation feature, or state that the feeling of control was already equivalent to the initial condition. The task was repeated until the participant stated that the character animation felt equivalent to the most complete configuration, or the absorbing animation condition [2,2,1,1] was reached.

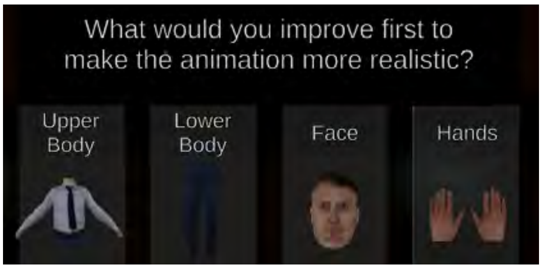

What would you improve first to make the animation more realistic?

Figure 22: After the animation clip, the participant had to decide whether and which feature to improved next.

Results

Participants: 13 individuals (five female) participated in the experiment.

Match Configurations: Figure 23 presents the probability of a Psi match P(match | u, l, f, h) in red, and P(u, l, f, h | match) in blue. Note that only configurations with 6 or more occurrences are presented. Considering the configurations preceding the absorbing condition [2,2,1,1], four configurations achieved a probability above 50% of being accepted as a match configuration, P(match | u=1, l=1, f=1, h=1) = .73, P(match | u=2, l=1, f=1, h=0) = .61 and P(match | u=2, l=1, f=1, h=1) = .8. Moreover, the marginal probabilities that a given animation feature level would be active at the match condition are presented in Table 9.

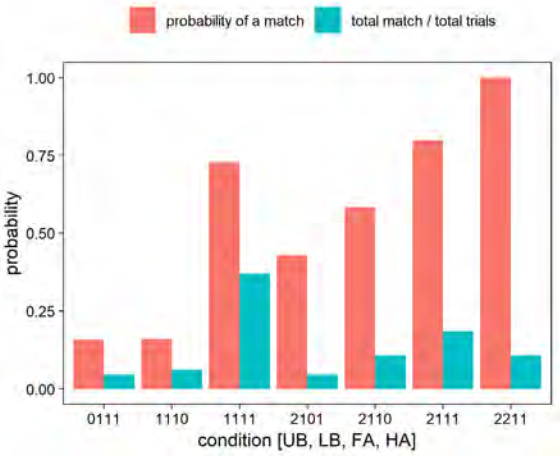


Figure 23: The probability of a Psi match for a given configuration [u, l, f, h], or P(match | u, l, f, h) is presented in red. The probability of a given configuration [u, l, f, h] when the participant felt a Psi match, or P(u, l, f, h | match), is presented in blue. Only configurations with 10 or more occurrences are presented.

Table 9: Probability that any given animation setting will be active in a match configuration.

Feature	Level		
	0	1	2
Upper Body	.077	.477	.446
Lower Body	.015	.815	.169
Face	.046	.954	-
Hands	.185	.815	-

Transitions: Participants were told that the transition order was important, and that they should decide for the feature that they miss the most. With the record of transitions between configurations we can reconstruct the path preferred by participants to go from an initial configuration to a match configuration. We model it with a sparse 36 x 36 matrix containing a total of 84 transition probabilities. With this matrix we represent the transitions as a Markov chain in Figure 24. As illustrated, the following sequence of transitions was the most probable to happen in our experiment: [0,0,0,0] → [0,1,0,0] → [0,1,1,0] → [1,1,1,0] → [1,1,1,1] → [2,1,1,1] → [2,2,1,1]. This path results in a single improvement for Lower body, Face, Upper body and Hands respectively, followed by a maxing the Upper body and then Lower body.

In Figure 24, the representation of the transition probabilities as a Markov Chain is shown. Participants could only move from the nodes in the left to nodes in the right. [2, 2, 1, 1] is an absorbing condition and no transition was possible once it was reached. The thicker the line is, the most probable it is that a transition from a node in the left to a node in the right will be performed. Gray shaded nodes are configurations that have not been visited in any trial.

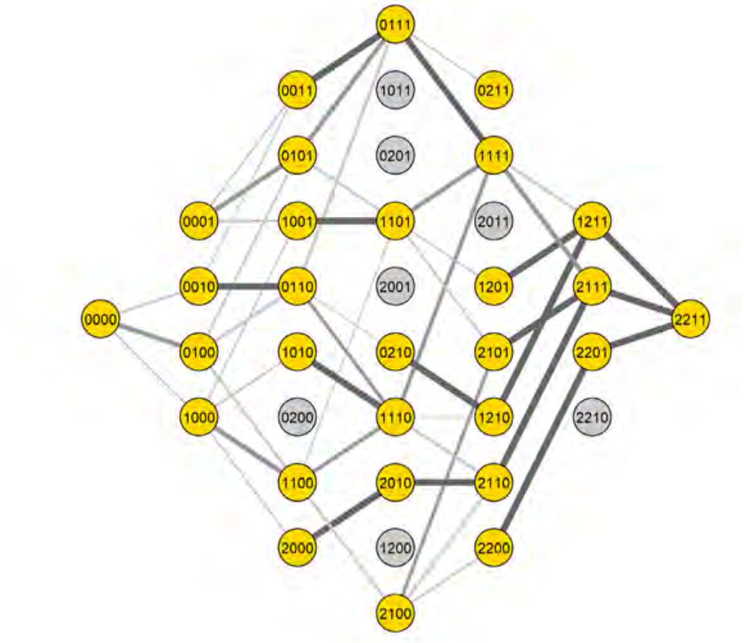


Figure 24: Representation of the transition probabilities as a Markov Chain.

As participants could perform up to 6 transitions, the probabilities of achieving any given configuration after 1, 2, 3, 4 and 5 transitions are presented in Figure 25.

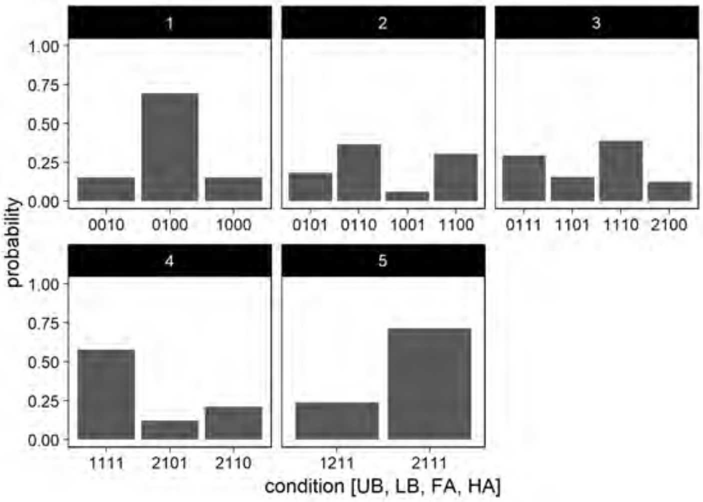


Figure 25: Probability distribution of any given configuration after each transition. Only probabilities greater than 0.05 are presented.

Analysis

Similarly to the experiment ARTANIM-1, we found that an animated character with upper and lower body animated using 8 tracked rigid bodies and inverse kinematics (IK) was often perceived as equivalent to a professional capture pipeline relying on 53 markers. Compared to what usual VR kits in the market are offering (headset and controllers tracking), feet tracking, followed by mouth animation and procedural finger animation, were the features that most added to the sense of control of a virtual body, and were often among the first to be improved.

References

[Slater, 2009] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 364(1535):3549–3557, 2009. doi: 10.1098/rstb.2009.0138

[Slater et al., 2010] M. Slater, B. Spanlang, and D. Corominas. Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. ACM Trans. Graph., 29(4):92:1–92:9, July 2010. doi: 10.1145/1778765.1778829

3.2.2. CWI-2

The goal of the experiment is to understand the user experience of photo sharing in social VR, comparing with face-to-face photo sharing and Skype photo sharing, benchmarking different existing mediated communication systems. This experiment also helped in developing the new methods for evaluating social VR.

Research questions

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

Methodology

A within-subject controlled user study was conducted to compare photo sharing experiences in three conditions: face-to-face (F2F), Facebook Spaces (FBS), Skype (SKP). The resulting data is then: (a) used in an exploratory factor analysis (EFA) [Costello & Osborne, 2005] to better understand the important factors in our questionnaire (b) provide empirical findings comparing photo sharing across study conditions.

While many platforms have incorporated social capabilities in their VR experiences, FBS allows users to be immersed in a 360-degree photo-based virtual and shared space, using their self-customized avatar. The body and lip movements of the avatar are coordinated with user movements, and users can express a selection of facial emotions on the avatar using a controller. Users can also access their own photos, videos, and any media shared, which made FBS suitable for our photo sharing activity. While systems such as vTime (https://vtime.net) and High Fidelity (https://highfidelity.com) both have an in-built avatar-based social VR community, where people can socialize with friends/family in 3D virtual destinations or in self-selected 360 photos, they have limitations in user control of facial emotions of the avatar. High Fidelity has no photo sharing function yet, and other social VR platforms like SineSpace (http://www.sine.space), Sansar (https://www.sansar.com) and AltSpaceVR (https://altvr.com) focus on inviting users to be creators and explorers of 3D virtual worlds, rather than supporting F2F photo sharing. While user avatars in these platforms can mimic users' body language, they are limited in what facial emotions can be expressed. Finally, we used FBS as it was one of the high-fidelity commercially available systems that required minimal setup. SKP was chosen as it is currently a standard in 2D video conferencing for friends/family [Forghani et al., 2014].

Participants were asked to select three different photos on their smartphone for sharing purposes. Each participant shared one photo in each condition. The sequence of conditions was counterbalanced according to Latin Square design, and all sessions were video recorded (with consent). After each condition, participants filled in our developed 5-point Likert-scale questionnaire about the experience in that condition. A semi-structured interview was conducted when participants completed the three conditions and were sitting together. Interviews were audio recorded. Four main questions were asked during the interview: 1) Compared with F2F photo sharing, what do you think is missing in SKP and FBS? 2) How did you experience photo sharing activities in FBS and what else would you want to do in social VR? 3) To what extent are you satisfied with the virtual environment? 4) What for you is the future of immersive social communication media?

The employed questionnaires and forms in Exp-CWI-2 can be found in Annex V.

Setup

Our experiment room was divided into two separate rooms by a movable wall. Both rooms have the same layout consisting of a pair of identical chairs, placed side by side. A computer with a 55" TV screen was placed in front of the chairs (Figure 26). For SKP, participants wore noise-cancelling headphones, and for FBS heard audio from the Oculus headset, so any potential proximity-related noise did not influence study participants. In the F2F condition, two participants were sitting together and showing each other photos on their smartphones and telling stories behind each photo. In the SKP condition, two participants were sitting in different rooms, where each was equipped with one laptop, and requested to share photos on their smartphones through SKP. In the FBS condition, photos were selected by participants and uploaded to the FBS system. Two participants were sitting in two different rooms, but they entered the same virtual space to share their photos (represented as physical photographs which are similar to the photo contents of a smartphone display). Although we could have uploaded photos to SKP, we opted for maintaining the same embodiment and self-representation (as in sharing photos in person) in all conditions. This also aligns with prior work that showed that users would spend more time in a video call if media sharing was easier [Forghani et al., 2014]. Admittedly, there is a trade-off here with image quality and size but this approach preserves the tangible and physical nature of photograph sharing.

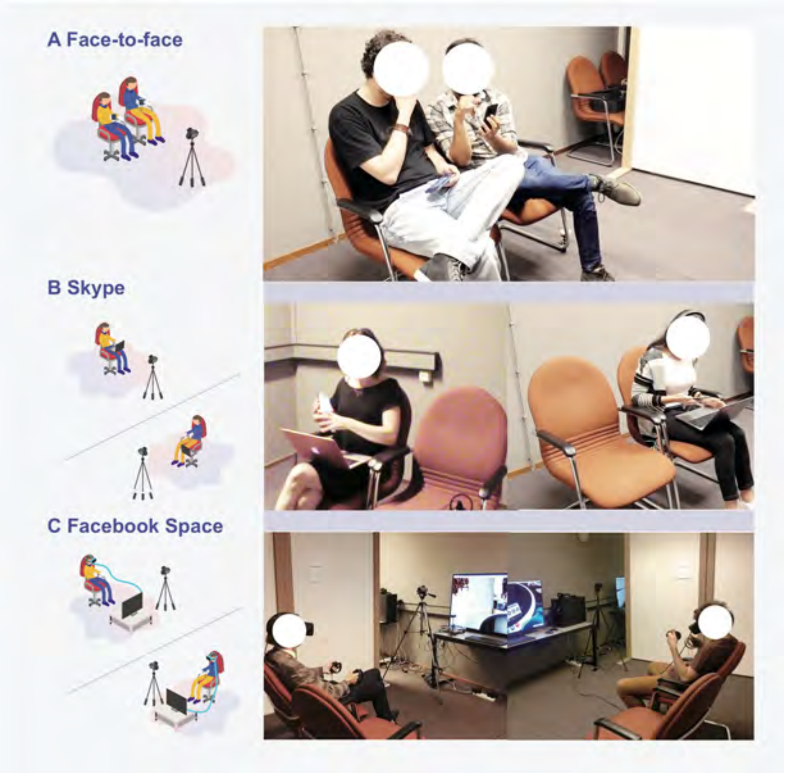


Figure 26- The setup of three experiment conditions

Description of F2F setup A:
Two participants sit side-by-side and used their mobile phones to share photos with each other. Each participant shared one photo.

Description of SKP setup B:

Two participants sit in different rooms, and each person saw the other person through Skype in a laptop, displaying the upper part of the body. They still shared photos with each other using mobile phones. Each participant shared one photo, but a different one.

Description of FBS setup C:
Two participants sit in different rooms. Each person uploaded one photo into Facebook Space, and they entered a virtual room together to share one photo with each other. The photo was different from the previous two.

In the virtual space, the avatars of participants were created by the researcher, according to their appearance. The participants only saw their avatars' hands. Only the upper part of their partners' body was visible to them. The virtual bodies of the two participants were positioned side-by-side around a table. The background image was a 360 image of the real environment.

Participants

Twenty six (N=52) participant pairs (29 m, 23 f; M=27.6, SD=7.9) were recruited. All had diverse backgrounds and education levels. Recruiting criteria was: (a) Participant pairs for each experiment session should know each other and have a smartphone with at least three photos on it (b) Participants should be willing to share those photos with their partner during sessions (c) Participants should not have visual or auditory problems to avoid motion sickness in VR

Procedure

Step 1 Introduction: The participants were introduced about the background of the experiment and the process they need to go through.

Step 2 Sign the consent form: Participants signed the consent form, explaining the data usage and possible risks.

Step 3 Background information questionnaire: Participants filled in the background questionnaire, collecting basic information of them.

Step 4 Face-to-face condition: Chose a photo from the phone and share it with partner. After that, filled in a questionnaire about the experience.

Step 5 Skype condition: Chose a photo from the phone and share it with partner through Skype. After that, filled in a questionnaire about the experience.

Step 6 Facebook social VR condition: Chose a photo from the phone and uploaded it onto Facebook space. Get training about how to use the social VR. And finally enter the social VR to share the photos with each other. After that, fill in a questionnaire about the experience.

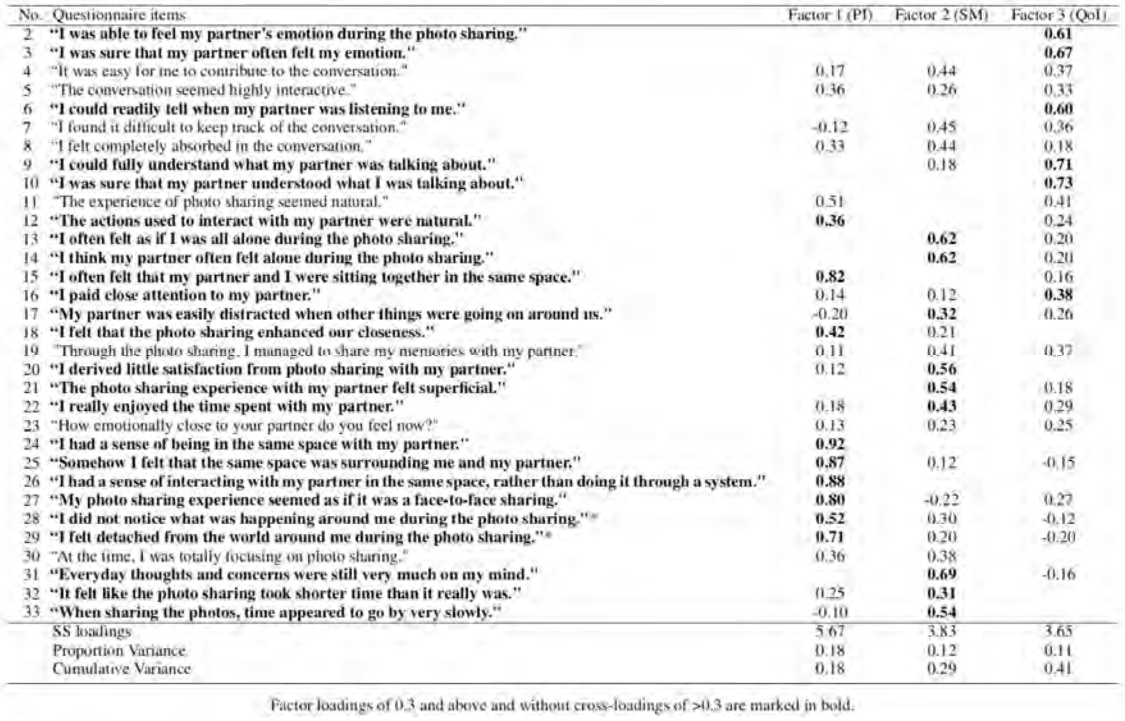
Step 7 Face-to-face interviews: The experimenter conducted a semi-structured interview with participants which took about 10 minutes.

Results

Exploratory Factor Analysis

We ran an exploratory factor analysis (EFA) [Costello & Osborne, 2005] to better understand the important factors in our questionnaire. EFA is a statistical technique within factor analysis commonly used for scale development involving categorical and ordinal data, and serves to identify a set of latent constructs underlying a battery of measured variables [Fabrigar et al., 1999; Norris & Lecavalier, 2010]. Given that our focus was on evaluating SKP and FBS, and that they contained the complete list of questions, we ran our analysis only on data from these two system evaluations. Since Bartlett's Sphericity test was significant (χ2(2,496) = 2207.187, p<0.001) and Kaiser-Meyer-Olkin was greater than 0.5 (KMO=0.85), our data allowed for EFA. Given our earlier literature reviews showed groupings of questionnaire items based on three factors, we tested our model fit based on three factors corresponding to each set of questionnaire items. Furthermore, since we assumed that factors would be related, we used oblique rotation ('oblimin') along with standard principal axes factoring. Standardized loadings are shown in Table 10.

Table 10 - Exploratory Factor Analysis (EFA) applied to our questionnaire items, where questions in bold indicate that these items are kept for the final analysis



No.	Questionnaire items	Factor 1 (PI)	Factor 2 (SM)	Factor 3 (QoI)
2	**"I was able to feel my partner's emotion during the photo sharing."**			**0.61**
3	**"I was sure that my partner often felt my emotion."**			**0.67**
4	"It was easy for me to contribute to the conversation."	0.17	0.44	0.37
5	"The conversation seemed highly interactive."	0.36	0.26	0.33
6	**"I could readily tell when my partner was listening to me."**			**0.60**
7	"I found it difficult to keep track of the conversation."	-0.12	0.45	0.36
8	"I felt completely absorbed in the conversation."	0.33	0.44	0.18
9	**"I could fully understand what my partner was talking about."**		0.18	**0.71**
10	**"I was sure that my partner understood what I was talking about."**			**0.73**
11	"The experience of photo sharing seemed natural."	0.51		0.41
12	**"The actions used to interact with my partner were natural."**	**0.36**		0.24
13	**"I often felt as if I was all alone during the photo sharing."**		**0.62**	0.20
14	**"I think my partner often felt alone during the photo sharing."**		**0.62**	0.20
15	**"I often felt that my partner and I were sitting together in the same space."**	**0.82**		0.16
16	**"I paid close attention to my partner."**	0.14	0.12	**0.38**
17	**"My partner was easily distracted when other things were going on around us."**	-0.20	**0.32**	0.26
18	**"I felt that the photo sharing enhanced our closeness."**	**0.42**	0.21	
19	"Through the photo sharing, I managed to share my memories with my partner."	0.11	0.41	0.37
20	**"I derived little satisfaction from photo sharing with my partner."**	0.12	**0.56**	
21	**"The photo sharing experience with my partner felt superficial."**		**0.54**	0.18
22	**"I really enjoyed the time spent with my partner."**	0.18	**0.43**	0.29
23	"How emotionally close to your partner do you feel now?"	0.13	0.23	0.25
24	**"I had a sense of being in the same space with my partner."**	**0.92**		
25	**"Somehow I felt that the same space was surrounding me and my partner."**	**0.87**	0.12	-0.15
26	**"I had a sense of interacting with my partner in the same space, rather than doing it through a system."**	**0.88**		
27	**"My photo sharing experience seemed as if it was a face-to-face sharing."**	**0.80**	-0.22	0.27
28	**"I did not notice what was happening around me during the photo sharing."***	**0.52**	0.30	-0.12
29	**"I felt detached from the world around me during the photo sharing."***	**0.71**	0.20	-0.20
30	"At the time, I was totally focusing on photo sharing."	0.36	0.38	
31	**"Everyday thoughts and concerns were still very much on my mind."**		**0.69**	-0.16
32	**"It felt like the photo sharing took shorter time than it really was."**	0.25	**0.31**	
33	**"When sharing the photos, time appeared to go by very slowly."**	-0.10	**0.54**	
	SS loadings	5.67	3.83	3.65
	Proportion Variance	0.18	0.12	0.11
	Cumulative Variance	0.18	0.29	0.41

Factor loadings of 0.3 and above and without cross-loadings of >0.3 are marked in bold.

To ensure the factors are meaningful and redundancies eliminated (removing collinearity effects), we only took items with factor loadings of 0.3 and above, and with cross-loadings not less than 0.2 across factors. The cumulative explained varance of the three factors is 41%. The 24 questionnaire items in bold were used for our evaluation of the three conditions (F2F, FBS, and SKP) along the identified concepts: Quality of Interaction (QoI), Social Meaning (SM), and Presence/Immersion (PI). We furthermore tested each set of items for internal reliability by measuring Cronbach's alpha, and our final item sets show high reliability coefficients: F2F QoI (α=0.8), F2F SM (α=0.89), F2F PI (α=0.74), FBS QoI (α=0.79), FBS SM (α =0.83), FBS PI (α =0.76), SKP QoI (α =0.78), SKP SM (α=0.79), SKP PI (α=0.75)

Questionnaire Response Analysis

We consider the effects of the three factors (F2F, FBS, SKP) on each 5-point likert-scale measure: Quality of Interaction (QoI), Social Meaning (SM), and Presence/Immersion (PI).

Quality of Interaction (QoI)

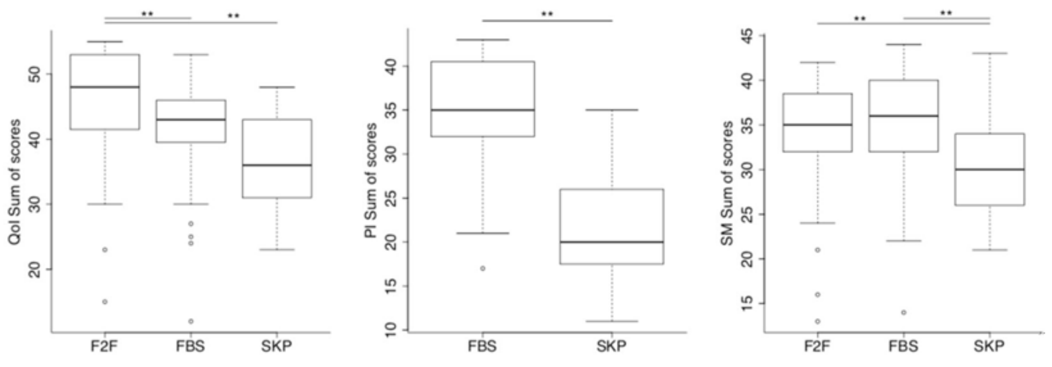
The sum of scores means and standard deviations for the QoI questions (6 items) for each tested photo sharing condition are: F2F=25.9(4.3), FBS=22.6(4.3), SKP=21.4(4). The sum of scores are compared in Figure 27.a. The horizontal lines within each box represent the median, the box bounds the Inter-quartile (IQR) range, and the whiskers show the max and min non-outliers. A Shapiro-Wilk-Test showed that our data is not normally distributed (p<0.001). As we compare three matched groups within subjects, we directly performed a Friedman rank sum test. Here we found a significant effect of photo sharing conditions on QoI (χ2(2)=39.1), p<0.001). A post-hoc test using Mann-Whitney tests with Bonferroni correction showed significant differences between F2F and FB Spaces (p<0.001, r = 0.41), between F2F and SKP (p<0.001, r = 0.54), but not between FBS and SKP (p=0.07). These results indicate that with respect to QoI, F2F photo sharing was perceived to be better than both FBS and SKP.

Presence/Immersion (PI)

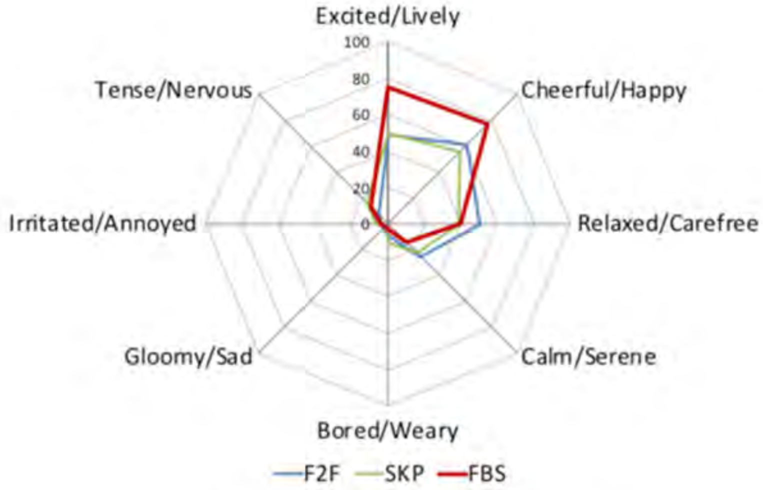
Since the F2F condition did not involve interacting with a photo sharing system and required four PI item omissions from our questionnaire, a comparative analysis is not applicable here. The sum of scores means and standard deviations for the PI questions (9 items) for each tested photo sharing system interaction are: FBS=34.9(6), SKP=21.7(5.9). The sum of scores are compared in Figure 27.b. A Shapiro-Wilk-Test showed that our data is not normally distributed (p<0.001). As we compare two matched groups within subjects, we directly performed a Wilcoxon rank sum test. Here we found a significant effect of photo sharing conditions on PI (W=2511, p<0.001, ηp2=0.74). These results indicate that with respect to PI, FBS was perceived to be more immersive and result in higher feelings of presence than SKP.

Social Meaning (SM)

The sum of scores means and standard deviations for the SM questions (9 items) for each tested photo sharing condition are: F2F=34.2(6.2), FBS=35.2(5.7), SKP=30.4(5.4). The sum of scores are compared in Figure 27.c. A Shapiro-Wilk-Test showed that our data is not normally distributed (p < 0.001). As we compare three matched groups within subjects, we directly performed a Friedman rank sum test. Here we found a significant effect of photo sharing conditions on SM (χ2(2)=27.2), p<0.001). A post-hoc test using Mann-Whitney tests with Bonferroni correction did not show significant differences between F2F and FBS (p=0.55), however did between F2F and SKP (p<0.001, r = 0.37) and between FBS and SKP (p<0.001, r = 0.43). These results indicate that with respect to SM, FBS photo sharing was comparable with F2F interactions, and significantly different than with SKP.



(a) Quality of Interaction (QoI). **(b) Presence/Immersion (PI).** **(c) Social Meaning (SM).**



(d) Reported emotions.

*Figure 27 - (a)-(c) Sum of scores boxplots for across photo sharing conditions for face-to-face (F2F), FB Spaces (FBS), Skype (SKP). (d) Self-reported emotion ratings for each condition. ** = p<.001*

Emotion Ratings

We plotted the emotion ratings on a radar chart (Figure 27.d), which shows FBS was perceived to be more exciting and cheerful than F2F and SKP. The sum of scores means and standard deviations (after subtracting the negative ratings) for emotions (N=8) under each photo sharing condition are: F2F=17.6(8.9), FBS=16.8(8.7), SKP=15.6(9.3). These are visually compared in Figure 27.d. A Shapiro-Wilk-Test showed that our data is not normally distributed ($p < 0.001$). As we compare three matched groups within subjects, we performed a Friedman rank sum test. We found a significant effect of photo sharing condition on emotion ratings ($\chi2(2)=8.74$), $p<0.05$). However, running post-hoc Mann-Whitney tests with Bonferroni correction did not show any significant interaction effects.

Other Influential Factors

Apart from different technology conditions, the influence of gender factor on different dimensions of experience was also explored (see Figure 28). Participants reported their genders in the Background questionnaire. The pairs were divided into three groups: male-male, female-female and male-female. Two-way ANOVA was performed. Only the technology factor has significant influence on the quality of interaction level ($F=10.229$, $p<0.001$). Gender factor had no significant influence on the quality of interaction, neither did the interaction factor of gender and technology. For the other two dimensions of experience, 'social meaning' and 'presence and immersion', the findings were the same. Therefore, the gender factors do not influence different dimensions of experience.

The factor 'Length of relationship' was also explored with the three dimensions of experience (see Figure 29). There are three levels of relationships: 1) knowing each other less than one year; 2) knowing each other between 1 and 3 years; 3) knowing each other more than 5 years. (Another level 4-5 years was not selected by any participants) With two-way ANOVA, significant influences of length of relationship factor ($F=5.496$, $p=0.005$) on 'social meaning' were found.

Participants knowing each other more than 5 years scored 'social meaning' significantly lower than participants who know each other less than 1 year or 1~3 years. This indicates that people knowing each other for a long time have higher requirements for 'social meaning'.

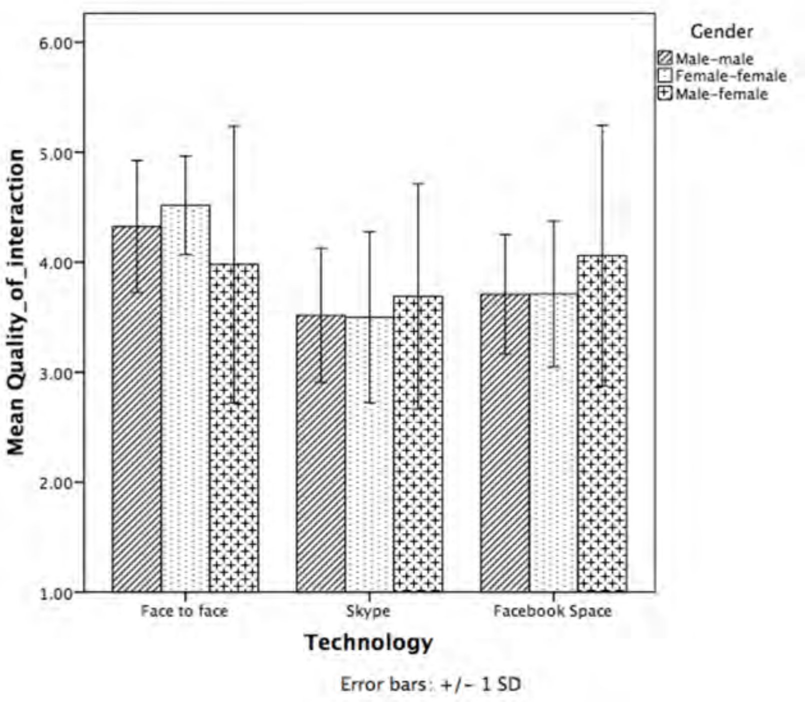


Figure 28 - The influence of gender factor on quality of interaction, score values range from 1 to 5 (SD indicated in graph)

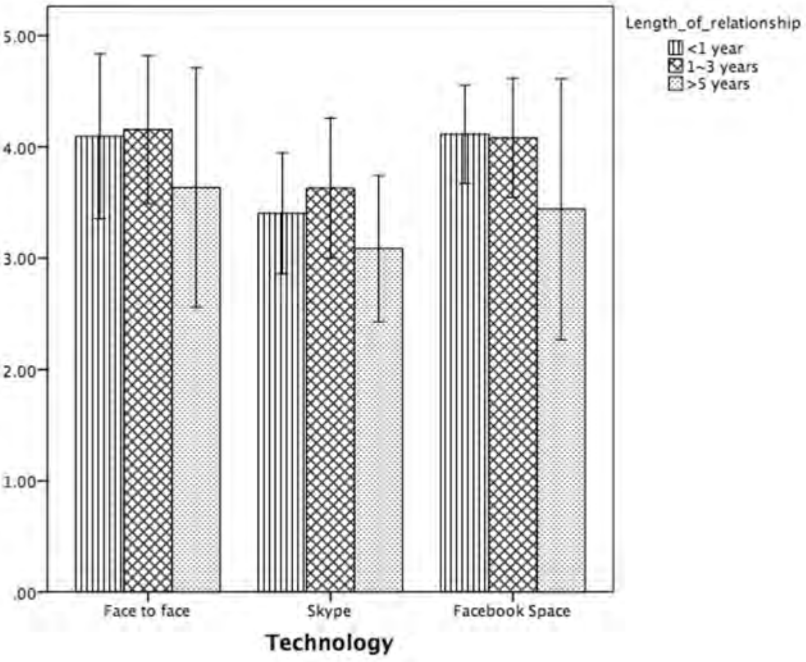


Figure 29 - The influence of length of relationship on social meaning, score values range from 1 to 5 (SD indicated in graph)

Interview Analysis

Audio recordings of the semi-structured interviews were transcribed and coded by two researchers, following an open coding approach [Sanders & Stappers, 2012]. From the coded transcripts, four main themes emerged, which we discuss below. Twenty-six pairs of participants are labeled P1A(B)-P26A(B).

Limitations of screen-based communication. Half the participants (52%) expressed concerns over using SKP for photo sharing, where they felt separated communicating via a monitor screen. They also felt distracted by the environment and became less focused on the conversation (P23B: "Skype is kind of curtain between you two."). Furthermore, the mis-match between the camera and the screen of the laptop made it impossible for participants to have eye contact (P9B: "The position of the camera makes the eye contact impossible in Skype."). Moreover, while is a broader issue with communication over an internet connection that can be affected by bandwidth issues, some participants (21%) complained the occasional delay in the SKP audio stream influencing the fluency of the conversations. (P5A: "In Skype, I can read what they are saying from the lips, but the voice comes later.").

Avatar-based embodiment in social VR. For FBS, a few participants (13%) complained that the avatar appearance is too detached from reality (P1B: "It is difficult to link avatars with human beings. I will put more attention on the voice, rather than the avatar face."). Participants (29%) did not feel that they actually saw each other (P1A: "Avatar reminds you that you are still in VR, which is different from real life."). Another missing aspect mentioned by some participants (35%) is the limited ability to show facial expressions on their avatars via the Oculus Swift controllers, where these sometimes felt unnatural and restrictive in expressiveness (P25A: "In VR, you cannot express your emotions because you only have a few options, and you need to control them with buttons."). Furthermore, timely choosing as emotion was found to be difficult (P2A: "It is difficult to think about the emotions I need to show on my avatar, when I'm pointing to the photo or telling stories."). Instead, participants relied on extracting emotional cues from voice (P22B: "It's a bit difficult to show emotions, but you can still hear his [partner's] voice and interpret it."). Some participants (19%) found pointing and touching gestures in FBS to be natural (P2A: "The hands and pointing feel so natural in VR...when you touch something, it provides a little buzz which feels natural."). However, other hand gestures were less intuitive, such as holding, picking up, dropping, which requires participants to operate a few buttons to perform (P4A: "I needed to think about the gestures and operate many buttons.").

Social VR immersiveness and novelty effects. Compared with the SKP, participants felt physically and emotionally closer using FBS (P18B: "In VR, you feel much closer than in Skype, like staying in the same room."). An interesting aspect raised here is that the more detached from the real world made participants felt, the more they could focus on each other (P13A: "In VR, I felt we were in the same space, and focusing on the same activity. No distraction from the environment."). Throughout the interviews, the "wow" effect of FBS was raised (38%). Participants tended to feel more excited and happier because of the new technology (P9A: "VR experience was very exciting. It's new and you're not familiar with it."), However, this came at a technological habituation cost (P24A: "I guess, with time, I will get used to it.").

Beyond photo sharing in social VR. Participants believed that social VR can bring new forms of social interactions (25%). Using such technology, participants would like to do activities that were not possible in the real world. For instance, the most frequently mentioned activities included gaming (31%), collaborating in 3D spaces (25%), family or friend gatherings (21%), and exploring the world (19%). Participants also suggested that social VR should explore novel social activities, aside from everyday interactions like photo sharing (P16B: "We need some novel interactive approaches to live another person's life, for instance."). Furthermore, some participants wanted to have intimate activities with close relationships, while with strangers, gaming was considered as a good option (P12A: "For a long-distance relationship, I'd like to hug my girlfriend in VR. For friends/strangers, doing something together like board games.").

Future of social VR. Most participants (87%) were satisfied with the socialVR environment. Even though the resolution of the background 360◦ image was not ideal, they felt immersed (P3B: "The borderless view made me it feel very real."). They mentioned that immersion and sense of presence could be improved by using high resolution images and by reducing the weight of the HMD (P4B: "My HMD is not comfortable. It is heavy and tight, which reminds me that I am in VR."). Importantly, participants (33%) suggested that future socialVR platforms for meeting strangers and close friends should be separated due to safety and privacy concerns. Indeed, this echoes previous work that showed users' privacy concern when sharing photos in collocated [Lucero et al., 2011] and public settings [Holopainen et al., 2011]. For friends or families, they wanted intimate interactions (P2A: "I think social media should be a platform for friends and family only, and have another platform for meeting new people."). Some participants (21%) believed that F2F interactions cannot be replaced, especially for special occasions like dating or family gatherings (P4B: "Families can gather and meet, because I'm from a culture that embraces this.").

References

[Bordens, 2002] Bordens, K. S., & Abbott, B. B. (2002). Research design and methods: A process approach. McGraw-Hill.

[Costello & Osborne, 2015] Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical assessment, research & evaluation, 10(7), 1-9.

[Fabrigar et al., 1999] Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. Psychological methods, 4(3), 272.

[Forghani et al., 2014] Forghani, A., Venolia, G., & Inkpen, K. (2014, November). Media2gether: Sharing media during a call. In Proceedings of the 18th International Conference on Supporting Group Work (pp. 142-151). ACM.

[Field, 2009] Field, A. (2009). Discovering statistics using SPSS. Sage publications.

[Holopainen et al., 2011] Holopainen, J., Lucero, A., Saarenpää, H., Nummenmaa, T., El Ali, A., & Jokela, T. (2011, November). Social and privacy aspects of a system for collaborative public expression. In Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology (p. 23). ACM.

[Lucero et al., 2011] Lucero, A., Holopainen, J., & Jokela, T. (2011, May). Pass-them-around: collaborative use of mobile phones for photo sharing. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 1787-1796). ACM.

[Norris & Lecavalier, 2010] Norris, M., & Lecavalier, L. (2010). Evaluating the use of exploratory factor analysis in developmental disability psychological research. Journal of autism and developmental disorders, 40(1), 8-20.

[Sanders & Stappers, 2012] Sanders, E. B. N., & Stappers, P. J. (2012). Convivial toolbox: Generative research for the front end of design. Amsterdam: BIS.

3.2.3. CWI-3

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

Scenario

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

Social VR Systems

We included in the experiment two social VR (sVR) systems using different virtual user representations:

- The Facebook Spaces system, where each user is depicted as half-body cartoon-like customizable avatar [Facebook, 2007]. The avatar was personalized by the facilitator before the experiment starts, to look like the user. Figure 30 shows the view by one user in Facebook Spaces system.
- The Web Player-based sVR system developed by TNO [VRTogether, D2.4] where each user's 2D real texture is captured and segmented, by means of a Kinect sensor, while the user is wearing the HMD and participating to the experience. Figure 31 shows the view by one user in TNO system.

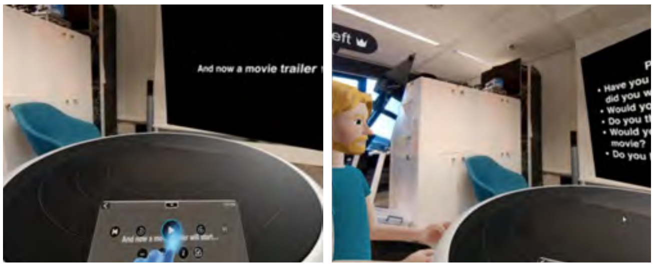


Figure 30- View by one user in the Facebook Spaces system. Depending on his viewing direction, the user sees the other user avatar, a virtual table where the playout interface is appearing, and a virtual screen where the movie trailer is played.

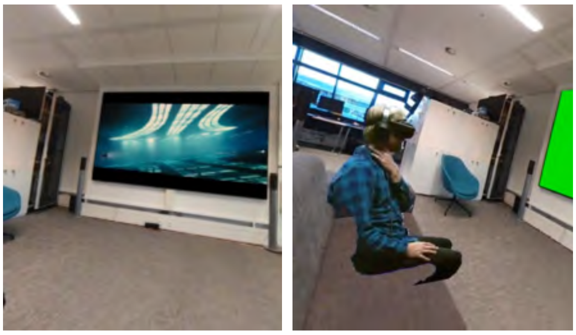


Figure 31- View by one user in the TNO system. Depending on his viewing direction, the user sees the other user segmented texture, sitting on a sofa, and a virtual screen where the movie trailer is played.

The VE was the same in the two systems, meaning that the same 360-degree background image was used in both systems, while some system design differences could not be modified. The differences included:

- in Facebook Spaces: the users wear hand controllers that allow to see their own virtual hands; the users appear as sitting on virtual chairs that are positioned around a virtual round table; the virtual screen is appearing at a fixed position on the other side of the table; the video playout has to be started by one user, using a virtual touch player interface; the video is loaded in the application as a media file appearing in the timeline of the Facebook account of the user, thus it can be re-compressed and thus show lower visual quality.

- in TNO system: the users appear as sitting on a coach and the virtual screen is covering the wall in front of them; the users did not need to hold/wear any controller; there is no self-representation, i.e., the user does not see any part of her/his own body; in the user virtual representation the HMD was visible and occluding large part of the user's face; the video playout has to be started by an operator external to the VE.

Test room and recording setup

In both cases, each user was sitting on a chair fixed to the floor in two separate rooms, wearing an HMD and noise-cancelling headphones. The two rooms were isolated controlled environment, with no background noise. One facilitator operated the VR systems using a central computer and helped the user in each of the room. The two facilitators could communicate with each other over Skype. The setup is depicted in Figure 32.

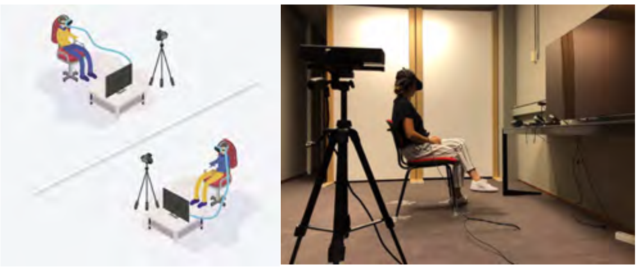


Figure 32- Test room setup to test the sVR systems

In order to collect the data for the objective evaluation described at section 2.2., the following external applications and recording devices were used:

- an ad-hoc Python application developed by Artanim [available at https://bitbucket.org/hgdebarba_artanim/steam_vr_recorder] was used to log the quaternions representing the user's head rotation when wearing the HMD at fixed time intervals

- the external application OBS (https://obsproject.com) was used to capture the viewport attended by the user when wearing the HMD as well as the audio channel of each user.

- a Logitech webcam and related recording software was used to record the user's body.

As a benchmark for the analysis of users' behaviour, we included the actual face to face scenario in the experiment, where the users where sitting on two chairs in the same conference room with a screen in front of them. Both users were recorded using a webcam. The room layout was similar to that of the 360-degree image used as background in the other systems. Figure 33 depicts the face to face room setup.

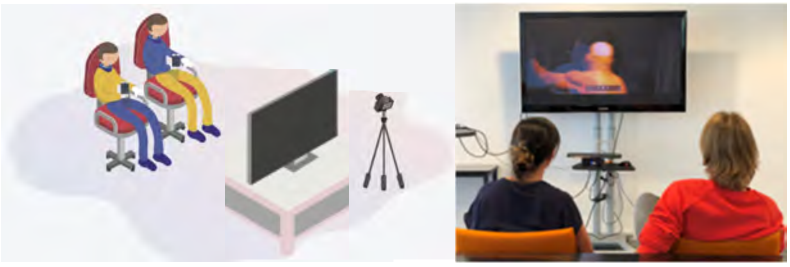


Figure 33- Test room setup for the face to face condition

Test design and users

We will refer to the three cases, i.e., Facebook Spaces (FB), TNO and face to face (f2f), as the *conditions*.

We recruited 16 pairs of users, so that users in each pair knew each other. The users received monetary compensation to participate in the test. The users average age was 31.06 years old (std = 7.39), and we had 17 women and 14 men participants.

Three video trailers where selected, being trailers of action/science fiction movies with approximately the same number of views on YouTube. Table 11 summarizes the trailer information.

Table 11- Trailer Youtube Link and info (T1 = trailer 1, T2 = trailer 2, T3 = trailer 3).

Trailer	Movie	YouTube Link	Duration (minutes: seconds)
T1	Blade Runner 2049	https://www.youtube.com/watch?v=gCcx85zbxz4	02:21
T2	Mad Max Fury Road	https://www.youtube.com/watch?v=hEJnMQG9ev8	02:32
T3	Ready Player One	https://www.youtube.com/watch?v=cSp1dM2Vj48&frags=pl%2Cwn	02:25

We performed a within-subjects design: each pair of users experienced all three conditions, watching a different video trailer in each condition. A fully counter-balanced test design was applied so that each condition was experienced first the same amount of times across the user set and the association trailer-condition was also balanced (i.e. each trailer was associated to each system the same number of times). Table 12 summarizes the test conditions and content allocation per user pair. Note that in practice 16 user pairs out of 18 were recruited, so only the 16 first combinations of stimuli and conditions were used.

Table 12 - Randomized order of experimental conditions (FB = Facebook Spaces, TNO = Web-player based TNO sVR system, f2f = face to face) and content (T1 = trailer 1, T2 = trailer 2, T3 = trailer 3).

Number of pair	Order of the three conditions/trailers
1	FB-T1 TNO-T2 f2f-T3
2	FB-T2 TNO-T1 f2f-T3
3	FB-T3 TNO-T1 f2f-T2
4	FB-T1 f2f-T2 TNO-T3
5	FB-T2 f2f-T1 TNO-T3
6	FB-T3 f2f-T1 TNO-T2
7	TNO-T1 FB-T2 f2f-T3
8	TNO-T2 FB-T1 f2f-T3
9	TNO-T3 FB-T1 f2f-T2
10	TNO-T1 f2f-T2 FB-T3
11	TNO-T2 f2f-T1 FB-T3
12	TNO-T3 f2f-T1 FB-T2
13	f2f-T1 FB-T2 TNO-T3
14	f2f-T2 FB-T1 TNO-T3
15	f2f-T3 FB-T1 TNO-T2

16	f2f-T1 TNO-T2 FB-T3
17	f2f-T2 TNO-T1 FB-T3
18	f2f-T3 TNO-T1 FB-T2

Process (before the experiment)

Before the experiment started the facilitator explained to the users in a pair what the experiment was about, i.e. watching three movie trailers together in three different conditions, two VR and one face to face, and what the process would be, i.e., after each condition the users will have to fill in some forms and at the end of the entire experiment the users will be gathered together to perform a quick interview to get their feedback on the overall experiment. The users were instructed to feel free to interact and talk to each other before, during and after the video trailer playout. They were also informed that in order to facilitate and trigger the discussion between the users, some example questions appeared at the end of each trailer playout on the screen where the trailer was rendered. The same questions appeared after each trailer.

Then, the users were asked to read and fill three forms:

- a consent form explaining the scope of the experiment and asking for agreement upon the ownership of the data collected during the experiment and their use for research purposes
- a general information form to collect background information upon the user, any previous experiences with VR applications that the user might have had and general information upon the level of familiarity with the other user
- a Social Anxiety (SAD) form, to profile the social skills of the user.

Process (after each test condition)

After experiencing each test condition each user was asked to fill in the questionnaire designed for CWI-2, in order to assess the quality of interaction (10 questions), the social connectedness (9 questions), and the sense of presence/immersion (5 questions - used only for VR conditions), as well as the familiarity and appreciation of the video trailer.

When the experienced condition was a VR system, a simulator sickness questionnaire [Kennedy, 1993] was also filled in, to gather feedback upon eventual problems and discomfort when wearing the HMD.

Process (after the entire experiment)

After the user pair experienced all three conditions, both users were gathered together with the facilitator to perform a semi-structured interview to collect overall feedback on the experience. The interview's audio was recorded.

The employed questionnaires and forms in CWI-3 experiment can be in Annex VI.

Results from subjective data

Figure 34 shows the boxplot of the subjective scores for the factors quality of interaction (10 questions) and social connectedness (9 questions), for each of the three conditions, i.e., face2face (F2F), Facebook Spaces (FB Spaces) and TNO system (TNO). A Wilcoxon-Mann-Whitney Test for non-parametric statistical significance showed that statistically significant difference can be found between the face2face and both VR conditions (p-value = 1.107e-07; p-value = 0.0005626), as well as in between Facebook and TNO conditions (p-value = 0.002691). This allows to conclude that in terms of subjective reported quality of interaction, none of the sVR systems could compete with the actual face2face experience, while at the same time a more realistic avatar representation, such as that used by TNO system, improves the quality of communication with respect to a puppet-based avatar representation. This is inline with the findings of existing studies in literature that correlate avatar's realism to the quality of the interaction in mediated communications [Heidicker, 2017] [Smith, 2018] [Latoschik, 2017] [Roth, 2016] [Garau, 2003]. No statistically significant difference could be found in terms of social connectedness, and in terms of presence/immersion between the two sVR systems (Figure 35).

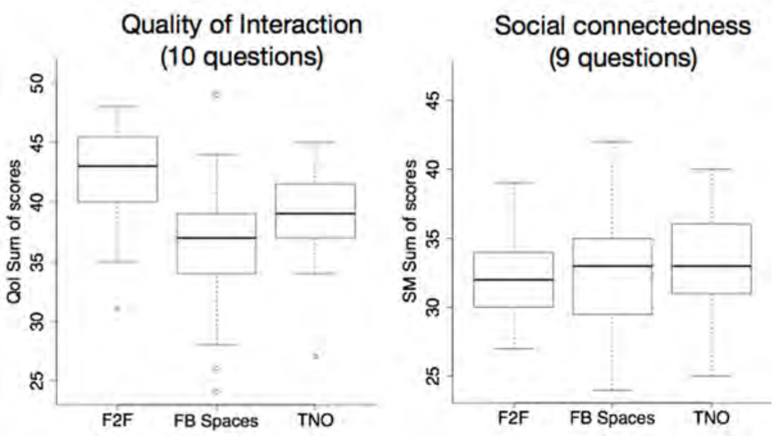


Figure 34- Box plot of the subjective scores for the factors quality of interaction and social connectedness for the three conditions: face2face (F2F), Facebook Spaces (FB Spaces) and TNO system (TNO).

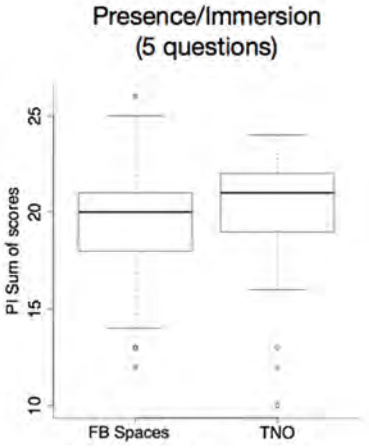


Figure 35- Box plot of the subjective scores for the factor presence/immersion for the three conditions: face2face (F2F), Facebook Spaces (FB Spaces) and TNO system (TNO).

Results from objective data

Figure 36 shows the boxplot of the temporal complexity [ITU-T P910] of each video recording of the users bodies during each condition (i.e., this index quantifies "how much did they move their body?"). By performing a repeated measures ANOVA and multiple comparison test with Bonferroni correction, we can observe a statistically significant difference between the face2face and the two sVR conditions (p = 0.4799e-07; 0.0043). This outcome indicates that the users moved significantly more in the sVR conditions rather than in the real face2face experience and it could be explained by the fact that having a limited field of view when wearing the HMD, they did move more their heads, as well as to the novelty effect of being in a VE, which brings them to visually explore the surrounding scene. Finally, the discomfort linked to wearing the HMD could also be a cause for more body movements.

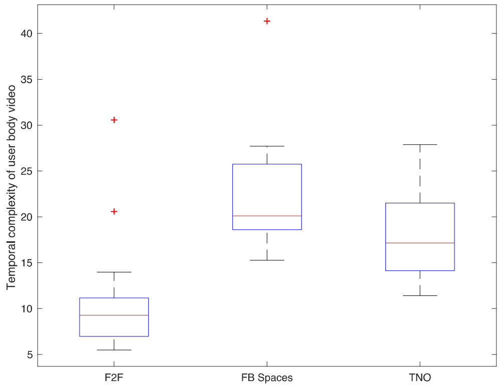


Figure 36- Box plot of the temporal complexity of each video recording users bodies during each condition as indicator of amount of body movements.

Figures 37 and 38 show the box plot for the percentage of time spent talking to each other and looking at each other, respectively, over the entire duration of the experience, for each condition. These are based on the processing of only 30% of the collected recordings, so the final outcomes, which will be submitted to IEEE VR 2019, might change. The current figures show no significant difference between the three conditions and correlated behaviour between the speech and gaze behaviours.

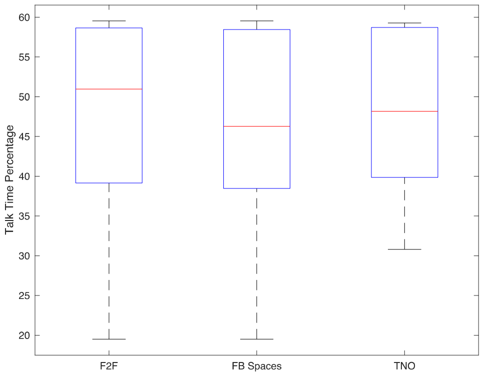


Figure 37- Box plot of the percentage of time spent talking to each other over the entire duration of the experience.

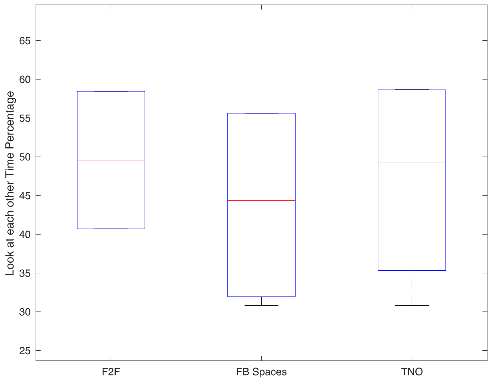


Figure 38- Box plot of the percentage of time spent looking at each other over the entire duration of the experience.

Additional objective data that has been collected but not processed yet include the logs of the head rotations which can be used to extract more detailed information upon "how much did the users move their head?" as well as statistics upon the angular velocity of head rotations.

Interview Results

Audio recordings of the semi-structured interviews were transcribed and coded by two researchers, following an open coding approach [Elizabeth, 2012]. From the coded transcripts, several main themes emerged, which we discuss below. The sixteen pairs of participants are labeled P1A(B) - P16A(B).

Almost half of the participants (47%) expressed concerns that the avatars in the Facebook system were not realistic. Some of the participants (22%) explained that the facial expressions were limited or missing, which influence the communication. Others (19%) mentioned that the body language of avatar was also missing. Therefore, they felt that the avatar was not helpful in communication (P4B: " We didn't look at each other while watching the trailer.") The user display in the TNO system was believed by some participants (28%) to be more personal and natural, compared with the Facebook system. With the photo-realistic display, the participants (25%) were able to interpret the emotions of each other (P9A: " If you looked into each other in the TNO system, you can somewhat interpret the emotions."). Some participants (41%) felt the blocked eyes were ok, while others (38%) were still bothered by the blocked eyes. Another significant difference between the two systems reported by the participants was the controller. They felt the controllers were difficult to use (16%), and do not want to hold the controllers all the time (22%). However, they (25%) did mention that the controllers increased the realistic feelings. Based on these differences, half of the participants (50%) prefer the TNO system for activities such as watching a movie. Others (34%) prefer the Facebook system because they believe it will be good for gaming.

Around 25% of the participants said the quality of VR environments was ok. The only problem with the Facebook system environment was found to be the 'table'. Participants (16%) mentioned that the virtual table looks fake and weird, since it did not match the realistic background. On the other hand, the problem reported most often with TNO environment was that the participants (34%) felt anxious with it (P9A: "If I move like this, I felt I could fall down"). When inside the VR environment, 38% of the participants mentioned that they did have the feeling of presence. However, the feeling of presence in the TNO system was strongly influenced by the blurry and pixels of the images, as was reported by 28% of the participants.

Due to the drawbacks mentioned above, most of the participants believed that the systems should to be improved. One of the important improvements was providing better body representations and enable automatic gesture recognition (22%). The head mounted display was also suggested to be more comfortable and light-weighted (28%). Other improvements were also suggested, such as including multisensory experience (9%), expanding the field of view (13%) and making the system more interactive (9%).

Analysis

The CWI-3 experiment allowed to develop and test a comprehensive test method to evaluate the quality of experience of users using sVR systems. The processed data so far show interesting outcomes that are inline with existing findings in literature, concerning the importance of realistic avatars [Heidicker, 2017] [Smith, 2018] [Latoschik, 2017] [Roth, 2016]; [Garau, 2003].

References

[Facebook, 2007] https://developers.facebook.com/videos/f8-2017/the-making-of-facebook-spaces/

[VRTogether, D2.4] VRTogether Deliverable D2.4-Integrated Software platform, v1

[Kennedy, 1993] R.S. Kennedy, N.E. Lane, K.S. Berbaum, and M.G. Lilienthal, Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, 1993.

[ITU-T P910] ITU-T, "Subjective video quality assessment methods for multimedia applications", Recommendation ITU-P 910, September 1999.

[Heidicker 2017] P. Heidicker, E. Langbehn, and F. Steinicke. 2017. Influence of avatar appearance on presence in social VR. In 2017 IEEE Symposium on 3D User Interfaces (3DUI).233–234.

[Smith, 2018] H. J. Smith and M. Neff. 2018. Communication Behavior in Embodied Virtual Reality. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). ACM, New York, NY, USA, Article 289, 12 pages.

[Latoschik 2017] M.E. Latoschik, D. Roth, D. Gall, J. Achenbach, T. Waltemate, and M. Botsch. 2017. The Effect of Avatar Realism in Immersive Social Virtual Realities. In Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (VRST '17). ACM, New York, NY, USA, Article 39, 10 pages.

[Roth 2016] D. Roth, J. Lugrin, D. Galakhov, A. Hofmann, G. Bente, M. E. Latoschik, and A. Fuhrmann. 2016. Avatar realism and social interaction quality in virtual reality. In 2016 IEEE Virtual Reality (VR). 277–278

[Garau 2003] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M. A. Sasse. 2003. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03). ACM, New York, NY, USA, 529-536.

3.3. Feedback from Professionals

Experiments intended for gathering feedback from professionals and experts, at fairs and exhibitions. In particular, we report:

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

3.3.1. TNO-1: Initial use-case study

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

Research questions:

- RQ1: Is social VR relevant for people?
- RQ2: What are the most important social VR use cases?

- RQ3: How do you measure the user experience in social VR?

Hypothesis:

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

Scenario

With the increased interest for Virtual Reality in the market (both in terms of hardware and software), interest in social VR has also emerged. This is showcased through VRChat, which attracted "10000 concurrent users" in January 2018 (*https://twitter.com/vrchatnet/status/949453320200052737*). The demand for more social VR is not surprising as humans are highly social beings. However, current VR systems that allow communication in VR (Facebook Spaces, VRChat and AltspaceVR, to name a few) have severe limitations when it comes to communication interactions (*https://extendedmind.io/social-vr*).

One limitation is that users are represented as artificial (sometimes comic-like) avatars. Even though this might be beneficial for some use cases, this might not be beneficial for many communication settings such as business meetings, or sharing experiences with family or friends. Based on current scientific literature and industrial approach, it is still unclear which use cases are relevant to the different methods by which users can be represented. Thus, more research is necessary to better understand social VR requirements. As a first step to close the above gap, we conducted a social VR study where participants tried a photo-realistic social VR experience in sessions of 3-10 min followed by a questionnaire and informal discussion. The experience was created to give people a better idea of social VR. The main contribution of this experiment is the study of use cases in social VR.

User panel: size and characteristics

We conducted our requirements gathering at the European VR exhibition, VR Days 2017 in Amsterdam. In this way, we ensured that our participants at this stage are people who at least have an interest in VR, and/or have experience using VR applications. The participants had the following characteristics:

- Total participants: 91 Experienced VR before: 80
- Gender distribution: 20 F, 69 M, 2 N/A
- Age range:
    - 20 between 18 and 30,
    - 49 between 30 and 45
    - 21 between 45 and 60

Stimuli

In the virtual environment, the two participants would appear to be sitting side by side on an office couch. They could see each other and communicate verbally with each other. Moreover, they could see a screen in front of the couch, and could either watch a video clip together or play a game together on the screen. After trying out the demo and taking off all the equipment, participants were then asked to fill in our questionnaires through a tablet device. Figure 39 shows the setup of the demo space and the screen in the picture shows the view of the participant within the HMD.

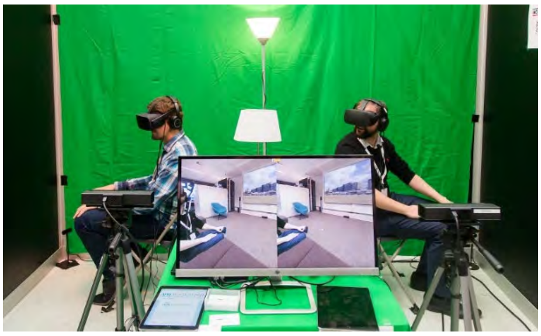


Figure 39- Two participants trying a social VR Experience. In a virtual environment, they appear to sit next to each other on an office couch, and can interact with each other

Environment

In our setup (Figure 39), each user has a specific and similar setup. Each user has a laptop (MSI GT62VR), Oculus Rift HMD (CV1), Kinect camera, headset (Sennheiser HD 201), unidirectional microphone (Power Dynamics PDT3), and gamepad (Xbox 360). It is important to note that the physical environment of the user is aligned with the virtual environment, i.e. if the user looks into the camera, he will look at the other person in the virtual environment.

In the virtual environment, the users sit on either the left or right side of a sofa. Thus, the view in the virtual room is different for each of the users. Furthermore, the other user is placed to the right or left of the user according to their view. The placement of users is done by alpha-blending people into the environment using WebGL shaders. We use this system to record users with a Kinect 2 RGB-plus-depth camera, replace the background with an alpha channel before transmission, and apply alpha-blending after reception to remove the background in the receiving browser (leaving us with a transparent image showing just the user without his/her physical background). Currently, for capture and transmission we use a resolution of 960x540 pixels.

Test protocol

We conducted the experiment in a demo space in an informal setting at the European VR exhibition with the following procedure:

- Short introduction of setup and safety
- Setup people with the HMD and microphone
- People experience and communicate in VR
- People get help to step out of VR
- People answer questionnaire: https://goo.gl/forms/DPKQXx4Vrimm6YzF3
- Short informal interview

The employed questionnaire in TNO-1 experiment is included in Annex VII.

Results

Further details of the results can be found in the following publication:

Gunkel, S., Stokking, H., Prins, M., Niamut, O., Siahaan, E., & Cesar, P. (2018, June). Experiencing Virtual Reality Together: Social VR Use Case Study. In Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video (pp. 233-238). ACM.

47.25% of the participants expressed that they are extremely interested in social VR experiences. Only 6 people were neutral or slightly interested, while no people expressed low to no interest through our survey.

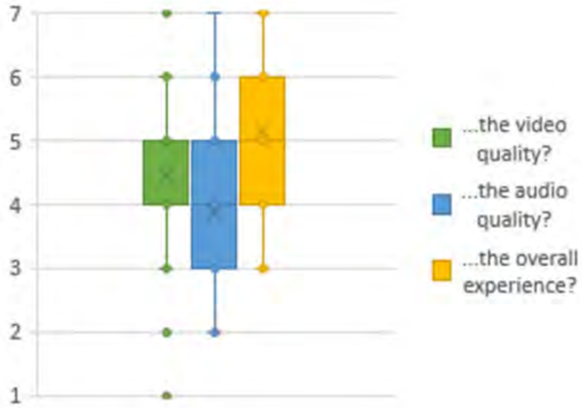


Figure 40- Users response when asked about audio, video, and overall quality of the experience on a 7-point Likert scale.

Figure 41 shows the histogram of responses for the questionnaire items asking users what they would consider to be the most important factors in (social) VR experiences. Based on the charts, "interaction within the experience" and "enjoyment of overall experience" seem to be considered extremely important by more than half of our participants.

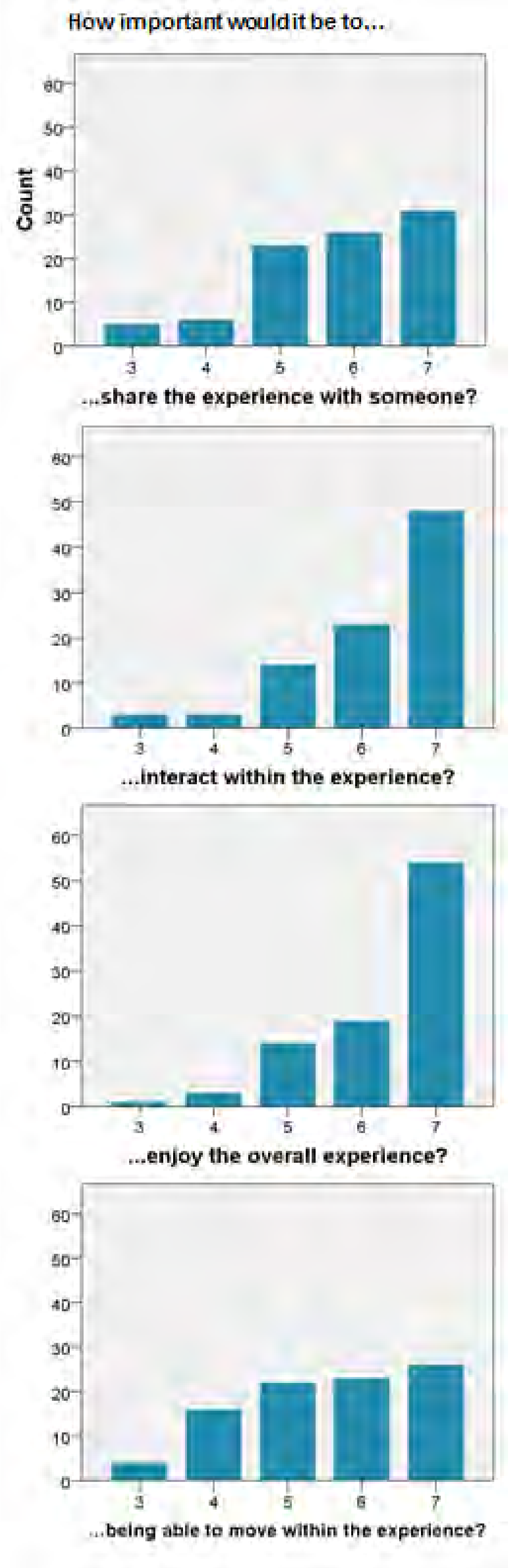


Figure 41- Important factors in (social) VR experiences

To compare the responses regarding interest in different application contexts, we coded the responses into the score range [-3,3], with neutral (the middle of the scale) as 0 value. We then took the average score across participants and plotted a chart comparing the average scores. Figure 42 shows an overview of the results for the different application contexts proposed in our questionnaire. From the charts, we see that the highest interest is shown for video conferencing and education applications, followed by video games, music experiences and movies.

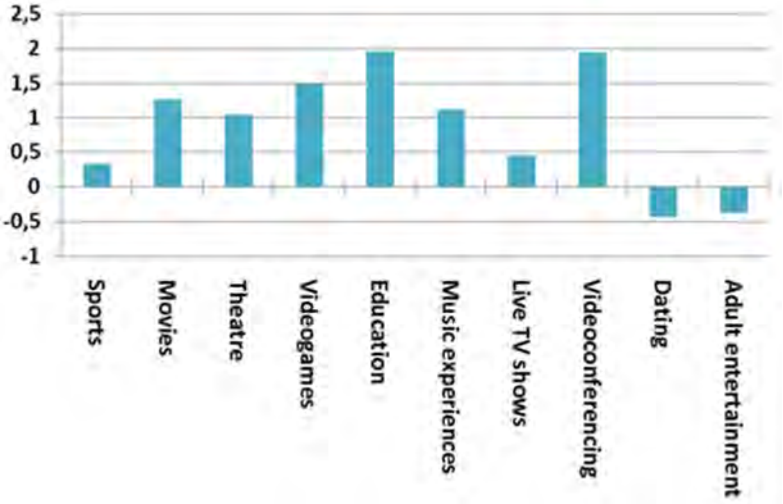


Figure 42- Potential application contexts for social VR experiences

The boxplot in Figure 43 lays out the statistical properties of the survey questions related to the quality of the social VR experience. We mapped the answers on the Likert scale to a value from 1 to 7.

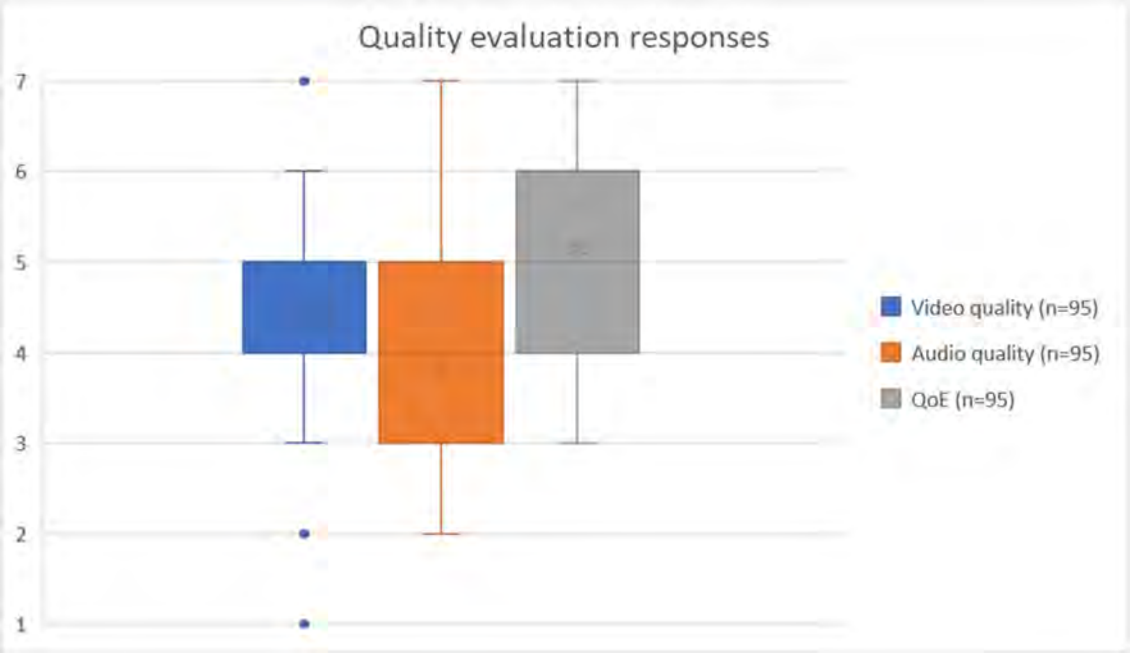


Figure 43- Statistical properties of the survey questions related to the quality of the social VR experience

The interest questions are distributed according to the following boxplot, where we mapped the answers on the Likert scale to a value from 1 to 7 (Figure 44):

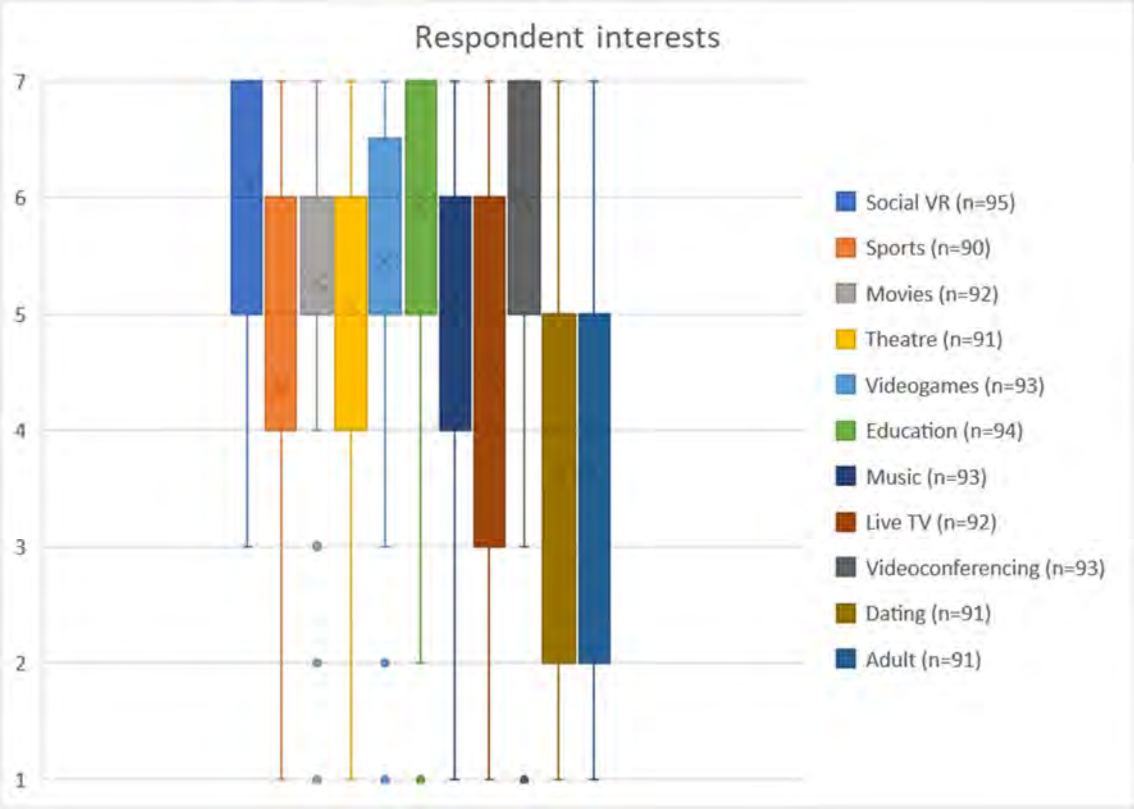


Figure 44- Scenarios / Use cases of Interest

Finally, the answers regarding which factors in social VR the participants deemed important show the following distribution, where we mapped the answers on the Likert scale to a value from 1 to 7 (Figure 45):

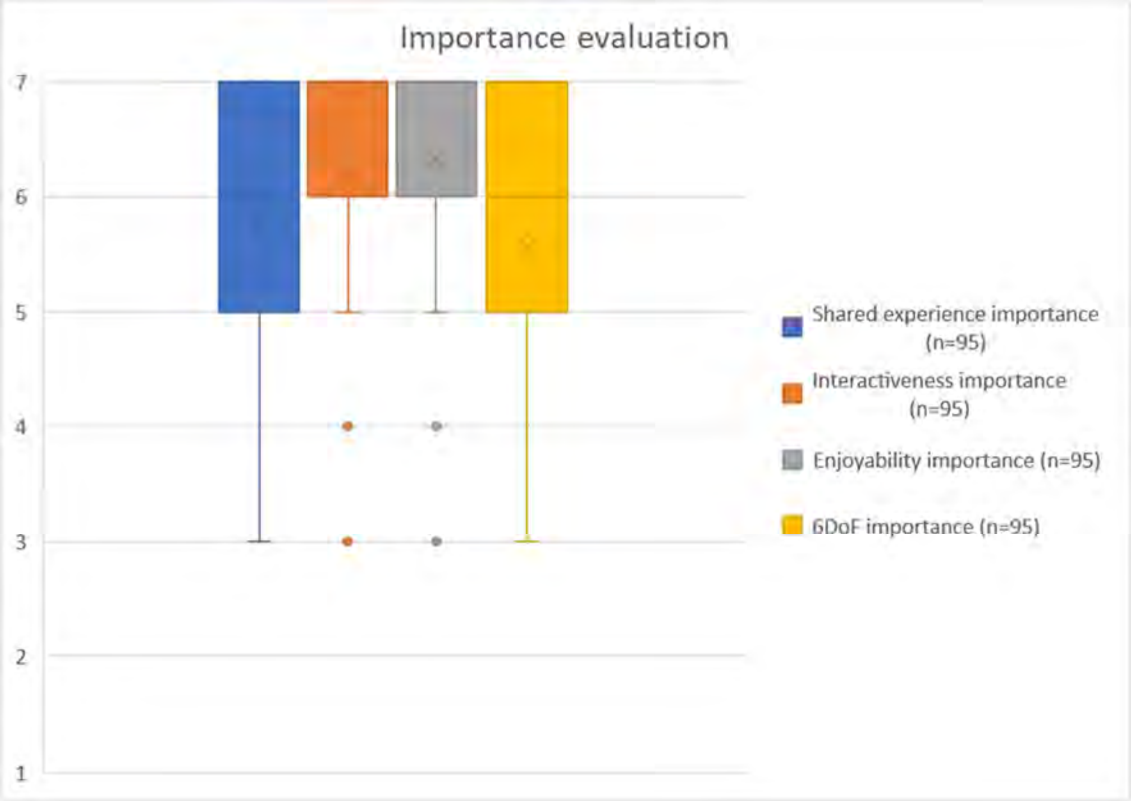


Figure 45- Important factors in social VR

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

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

The main goal of the experiment is to determine if a field trial would be suitable, and if so, under what conditions. A secondary goal of the experiment is to gather feedback on our system from a market party, in this case a potential buyer of such a social VR system. The experiment is thus also about requirements gathering.

Expectations / Hypotheses

The expectation was that the audio and video quality of the system will be sufficient for interaction/communication purposes. Also, it was expected that the system as offered, would not be sufficient for a field trial.

During an intake with the company, we discussed various issues with the current setup:

- Maximum number of participants of 4, while most teams within the company are between 6 and 8 persons.
- The HMD is visible during communication, which prevents eye-contact. The expectation was that HMD removal would be needed.

- Many teams use some kind of Kanban board, the company also often uses whiteboards and markers during these sessions. For a field trial, it is expected that some additional functionality (i.e. shared interactivity) is needed.

Scenario

The scenario is to have a stand-up in VR with 4 people standing around a table. To facilitate the stand-up, a shared PowerPoint is included, projected on the table in the virtual environment. One participant will receive a remote control, i.e. PowerPoint clicker, to be able to navigate through the slides. The envisioned duration of each stand-up would be between 20 and 30 minutes, in which the group can discuss on the presentation.

Conditions included in the comparison

All participants are given the same conditions.

Each participant is physically in his/her own room, acoustically separated from the others. All four rooms were adjacent to one another. In each room, a VR capable PC is placed, with attached to this an Oculus Rift CV1, a Microsoft Kinect V2, and using Bluetooth, a Sony MDR-1000X noise-cancelling headphone. Participants were wearing the Oculus and the Sony headphones during the experiment.

The users are asked to stand in a certain place in front of the Kinect camera. For this purpose, a thin circular tube with approx. 60 cm diameter, is placed on the floor, so people can feel their position with their feet, without a chance of tripping over the tube.

User panel: size and characteristics

Ten persons participated in the experiment, in two groups of four, and one group of three. One person participated in two sessions.

Stimuli (e.g, content of the photos)

The company provided its own PowerPoint presentation for discussion during the experiment. This was a technical presentation about a new IT architecture they are working on. The presentation was not specifically adapted for use in VR.

Environment

The physical rooms were small meeting rooms of approximately 2 by 3.5 meters. The rooms were typical office rooms.

The virtual environment is a 360-degrees image of a meeting room from the company, normally used for these types of meetings. This makes the virtual environment a familiar environment for the participants, and suitable for the virtual stand-ups. We captured this room with an Insta360 Pro camera. Each participant will stand in the same position in the room, seeing the other participants across the table shown in the 360-degrees image.

Test protocol

Introduction and training

All participants were explained beforehand about the experiment, and volunteered to participate.

Before each stand-up, we gathered the participants for that stand-up in the hallway between the rooms. We explained the general process of the experiment:

- First, some explanation about the equipment to be used.
- Then, up to half an hour of virtual stand-up.
- Afterwards a short questionnaire, followed by a group feedback session.

Then, one person was chosen to receive the remote control for controlling the slideshow.

The explanation of the equipment was given individually. This included:

- How to adjust the Oculus HMD properly to have it placed comfortably.
- How to adjust the headphones and the audio volume.
- Where to stand.

We also instructed our participants that they could stop the experiment at any time if they would like to, e.g. when they would feel dizzy, nauseated or the like. After this, we helped the participants with their HMDs and headphones, and left them to their session.

Test Methodology

The test consists of an explanation before, a test session, a short questionnaire and a group discussion afterwards. The group discussion was kicked off with the question: 'So, what do you think, how was it, what was good, what should be better?', the groups needing no more invitation to speak freely.

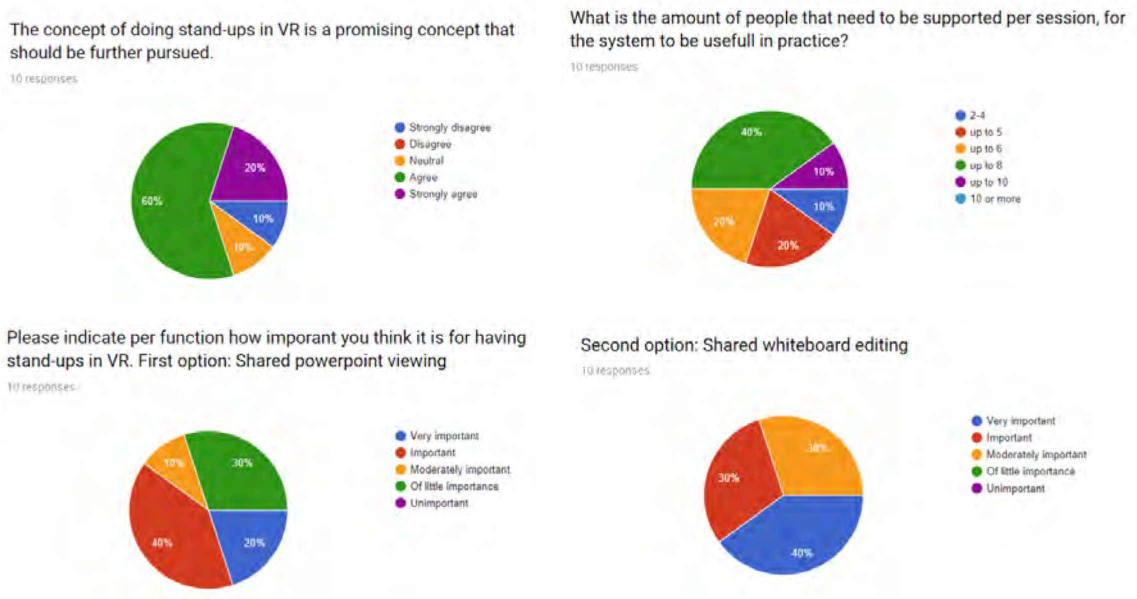
The employed questionnaire in TNO-2 experiment is included in Annex VIII.

Length of the test session

Each stand-up session lasted approximately 25 minutes. We ended each test by standing next to the participants, waving to the capture camera (so they would talk about this in the virtual environment), then tapping them on the shoulder and helping them take off the HMD and headphone.

Results

Based on the questionnaire and the group discussion, the following results were recorded (Figure 46):



The concept of doing stand-ups in VR is a promising concept that should be further pursued.

10 responses

What is the amount of people that need to be supported per session, for the system to be usefull in practice?

10 responses

Please indicate per function how imporant you think it is for having stand-ups in VR. First option: Shared powerpoint viewing

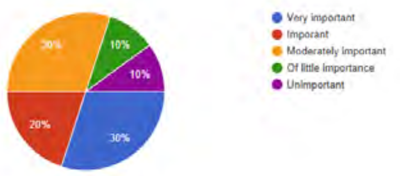
10 responses

Second option: Shared whiteboard editing

10 responses

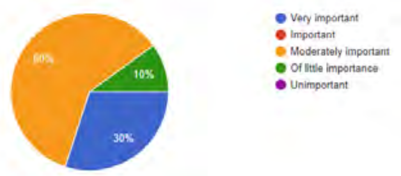
Third option: Personal note-taking

10 responses



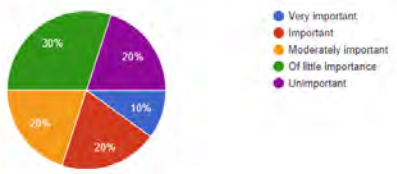
4th option: Shared document editing

10 responses



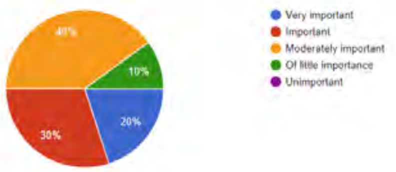
5th option: personal device usage (i.e. personal phone screen visible VR)

10 responses



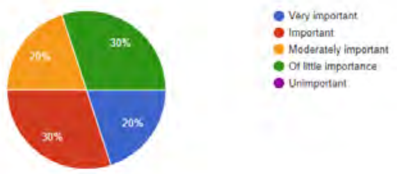
6th option: HMD removal (i.e. use computer vision technology to remove the HMD visually from the face)

10 responses



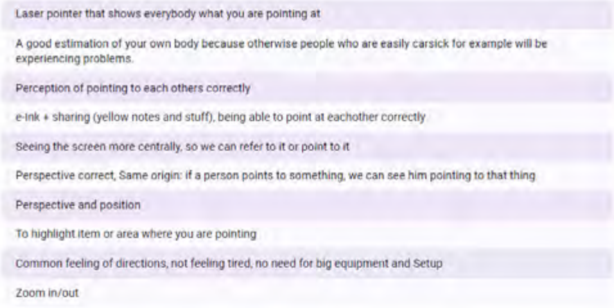
Final option: Self view (i.e. see your arms/hands/body in VR)

10 responses



What other functionality is important or very important to having these stand-ups in VR?

10 responses



Laser pointer that shows everybody what you are pointing at

A good estimation of your own body because otherwise people who are easily carsick for example will be experiencing problems.

Perception of pointing to each others correctly

e-Ink + sharing (yellow notes and stuff), being able to point at eachother correctly

Seeing the screen more centrally, so we can refer to it or point to it

Perspective correct, Same origin: if a person points to something, we can see him pointing to that thing

Perspective and position

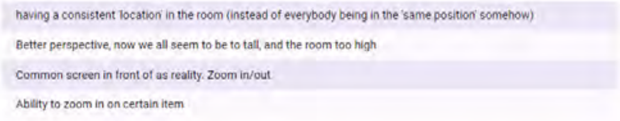
To highlight item or area where you are pointing

Common feeling of directions, not feeling tired, no need for big equipment and Setup

Zoom in/out

What other functionality would be nice-to-have for having these stand-ups in VR?

4 responses



having a consistent 'location' in the room (instead of everybody being in the 'same position' somehow)

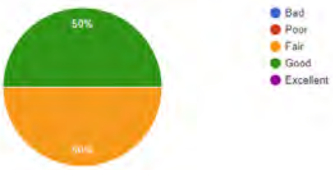
Better perspective, now we all seem to be to tall, and the room too high

Common screen in front of as reality. Zoom in/out

Ability to zoom in on certain item

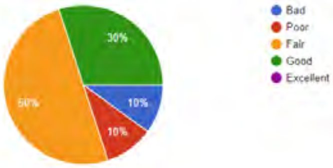
Now follow some questions on quality aspects. How would you classify the overall experience?

10 responses



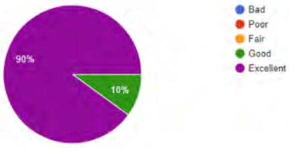
How would you classify the overall video quality?

10 responses



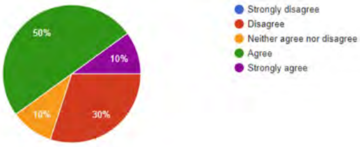
How would you classify the overall audio quality?

10 responses



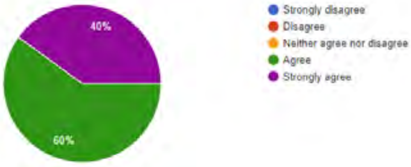
I really felt immersed in the experience, it felt as if I was actually in the room shown in VR.

10 responses



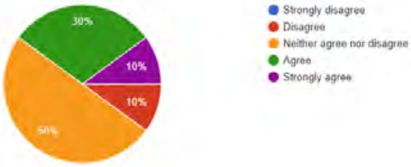
I felt connected to the other(s) in the virtual environment.

10 responses



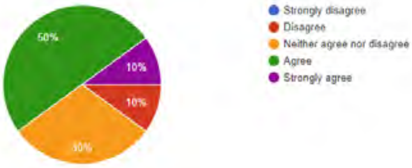
The other(s) paid close attention to me.

10 responses



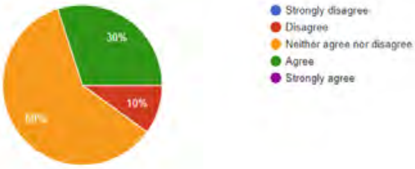
I paid close attention to the other(s).

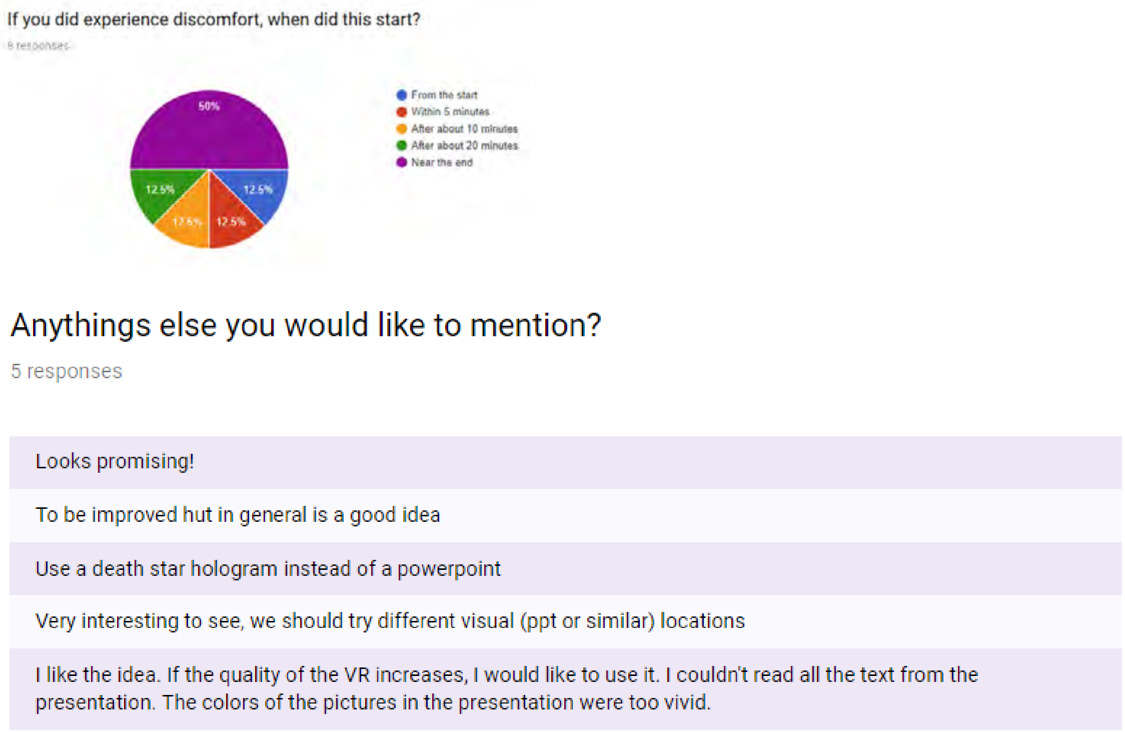
10 responses



Physically, the stand-up in VR was a pleasant and relaxed experience.

10 responses




If you did experience discomfort, when did this start?

8 responses

Anythings else you would like to mention?

5 responses

Looks promising!

To be improved hut in general is a good idea

Use a death star hologram instead of a powerpoint

Very interesting to see, we should try different visual (ppt or similar) locations

I like the idea. If the quality of the VR increases, I would like to use it. I couldn't read all the text from the presentation. The colors of the pictures in the presentation were too vivid.

Figure 46- Overall Results from TNO-2: Try-out of VR stand-up

Analysis

From the user feedback, we draw the impression / conclusion that social VR is suitable for doing Stand-ups in VR. Stand-up meetings are ideal for social VR as the duration of the interaction is limited, thereby avoiding user discomfort from prolonged sessions in VR. Users indicate that they would like to have access to more interactive features, mostly related to enhance the actual work they would like to do while in social VR.

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

This experiment has three main goals:

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

Research questions

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

Hypothesis

H1: 3D representations lead to a better user experience.

Scenario

Based on our previous experiments [Gunkel 1, 2017] [Gunkel 2, 2017], representing users in 2D (with chroma-key cut-out) and blending them into 360-degree static VR environments, we wanted to investigate the possibility to allow people a photo-realistic communication experience in full 3D (6DoF). In this experience, we compare the difference between the two approaches (360-degree static background with 2D user representations and a 3D scene with point cloud user representations, as shown in Figure 47). Thus we build a demo experience that allows users to switch between the 3D environment and the 360-degree environment, thereby facilitating the comparison of both approaches.

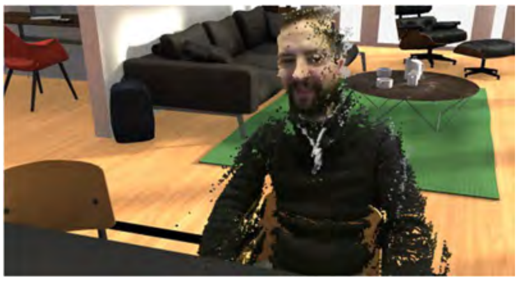


Figure 47- 3D point cloud in a 3D living room

From our previous version we mainly changed two components: (A) the Kinect v2 camera capture to record both colour and depth, and (B) the WebGL shader to display the user video. Regarding the camera capture (A), we combined the colour image with the depth image. However, as the browser and current WebRTC implementation does not support depth encoding we need to use an intermediate step. We map the 16 bit depth value from the Kinect into the RGB colour space using the red colour only. An example of such a mapping can be seen in Figure 48 (bottom right). To display this image as a point cloud (B) we changed the WebGL shader, in order to not display a flat image, but each pixel with coordinates in the 3D space, based on the depth image. Our shader is based on the work by George MacKerron [MacKerron].

Conditions included in the comparison

In order to test our 2D vs 3D approach we created a demonstrator and presented this in informal testing sessions at ACM Multimedia Systems (ACM MMsys) 2018. Arbitrary visiting participants tried our photo-realistic social VR experience in sessions of 3-10 minutes, followed by a questionnaire and informal discussion. In the VR environment, users stand in front of each other in two conditions:

A. A 2D 360-degree environment with users represented in 2D and
B. A 3D environment with users represented as a 3D Point cloud.

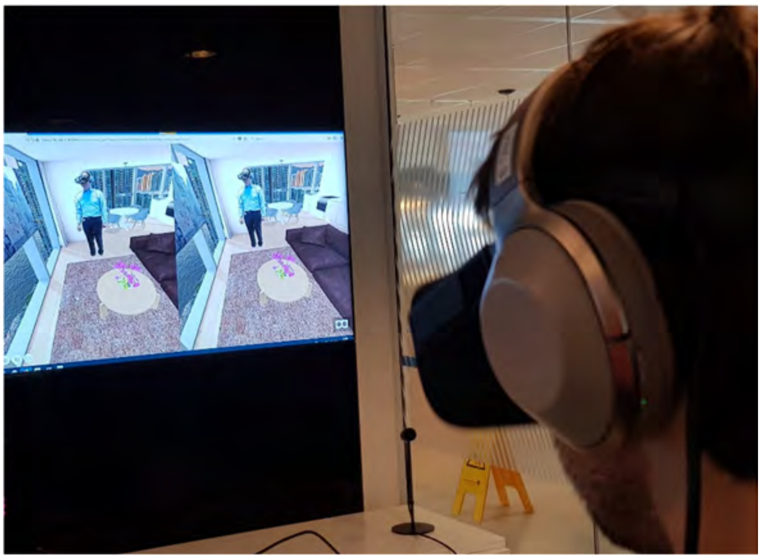


Figure 48- User experiencing our social VR WEB version in 2D and 3D (view of user is shown on the screen)

2D 360-degree environment with users presented in 2D: This condition is very similar to our previous experiences [Gunkel 1, 2017] [Gunkel 2, 2017]. This is, two users can communicate within a 360-degree VR environment. In the environment, a user is standing and sees the other user standing in front of him. Users are represented in 2D and blended into the 360-degree background. The 360-degree background was created as a simple image snapshot of the 3D room environment (which was taken from [Archilogic]). Two such 360-degree background images were created to reflect the two different user positions. These positions match the exact position of users experiencing the 3D view.

3D environment with users presented as 3D Point cloud: Similar to the 2D case, two users can communicate with each other in a VR environment. However in this configuration, the VR environment is represented using a full 3D room (taken from [Archilogic]]) and users are visualized as a 3D point cloud. A user is standing within the environment and sees the other user standing in front of him/her.

Testing Environment

Each user has a laptop (MSI GT62VR), Oculus Rift HMD (CV1 including a microphone), Kinect camera, noise-cancelling headset (Sony WH-1000X). It is important to note that the physical environment of the user is aligned with the virtual environment, i.e. if the user looks to the front into the camera, he or she will look at the other person in the virtual environment. Each user is recorded from the front, while standing up and thus both users will face each other standing in the virtual environment.

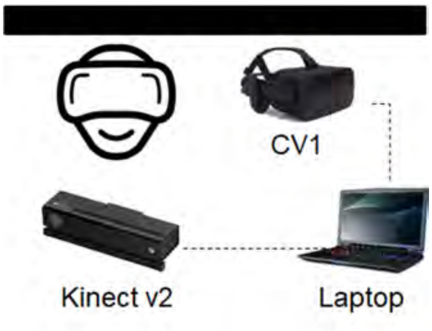

[redacted]

Figure 49- User's hardware setup

Test protocol

A. Users will get a brief introduction to our research
B. Users will be equipped with hardware
C. Users can experience our VR solution while being able to interact with each other through our system
D. The view of the users will be switched every 30 seconds between the 2 conditions (2D / 3D)
E. Users get help stepping out of VR (each session lasts roughly between 3 and 15 min)
F. Users answer a questionnaire
G. (optional) open discussion with users about the experience and conditions

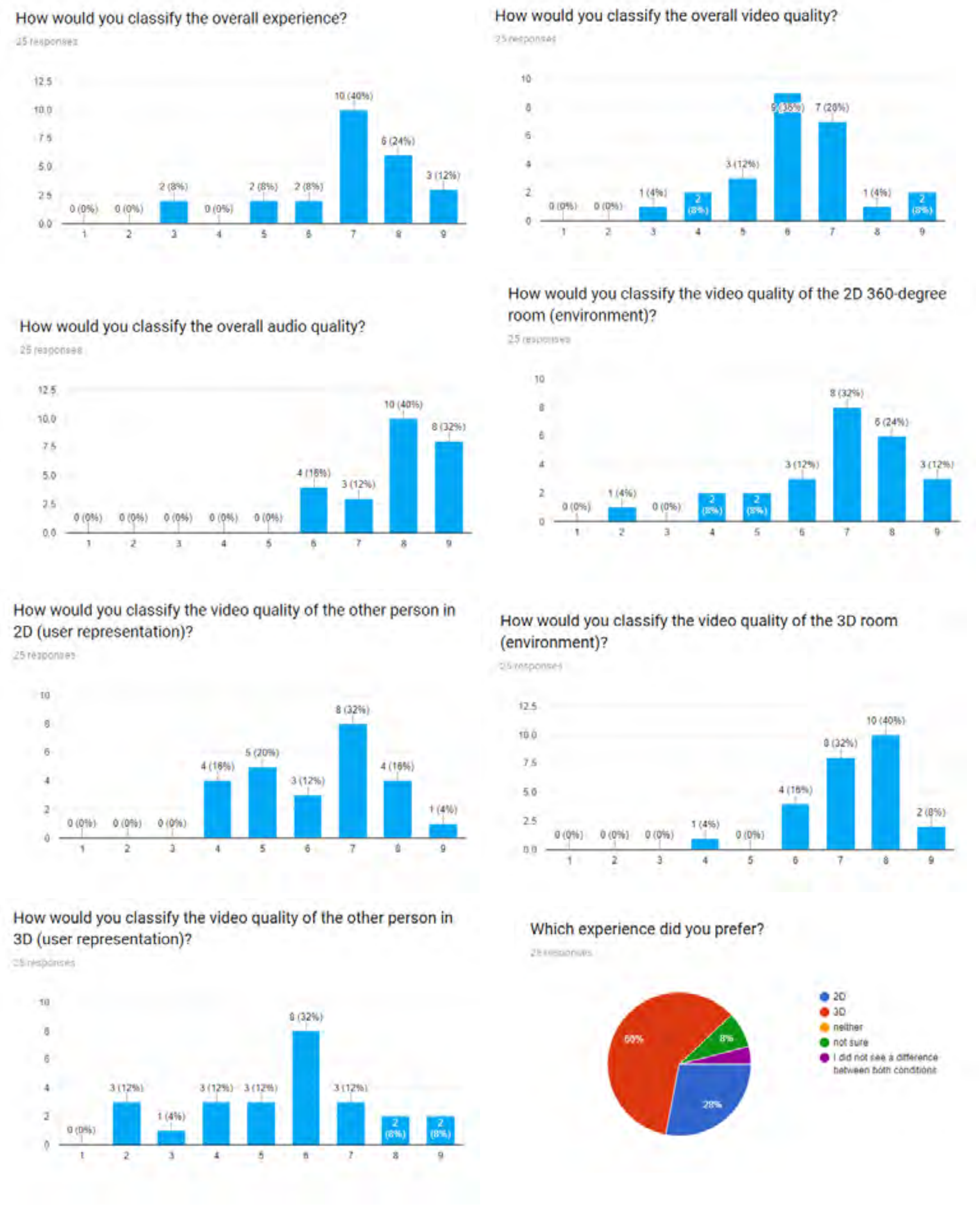
The employed questionnaire in TNO-3 experiment is included in Annex IX.

Initial Results

With our current test we identified the following initial results:

- Technically it is possible with little overhead to show the environment and users in 3D
- Users rate 3D room in higher quality (parallax, moving around)
- Users rate 2D user in higher quality (people look sharper and better, however you lose the ability to point accurately)
- We can make the 3D point cloud representation look better, but not in the current constraints of the laptop and browser performance
- Results are very dependent on the use case
- Overall we were able to showcase that more complex 3D VR applications can easily be enhanced with photo-realistic user communication in web-based applications.

Results


How would you classify the overall experience?

How would you classify the overall video quality?

How would you classify the overall audio quality?

How would you classify the video quality of the 2D 360-degree room (environment)?

How would you classify the video quality of the other person in 2D (user representation)?

How would you classify the video quality of the 3D room (environment)?

How would you classify the video quality of the other person in 3D (user representation)?

Which experience did you prefer?

Figure 50- Overall results from TNO-3: Representing the environment and users in either 2D or 3D

Analysis

A summary of the results obtained with the questionnaire is presented in Figure 50. Based on the limited amount of responses, it is unlikely that the results from the measurements can be generalized to end-users. From the answer data, we make the following observations:

- On average, the 3D representation is valued higher.
- The environment valuation for both the 2D and 3D environments have the same median (7), but the average valuation for the 3D environment is higher on average (7.28 vs 6.76) with a lower standard deviation (1.10 vs 1.71).
- The overall experience valuation is more strongly correlated to the 3D user representation valuation (0.62) than to the 2D user representation valuation (0.44).

- The overall experience valuation is more strongly correlated to the 2D environment representation valuation (0.32) than to the 3D environment representation valuation (0.23).
- The video quality valuation is more strongly correlated to the 3D user representation valuation (0.59) than to the 2D user representation valuation (0.25).
- The video quality valuation is much more strongly correlated to the 3D environment representation valuation (0.36) than to the 2D environment representation valuation (-0.05).
- The population is skewed towards males without vision problems.

The seeming discrepancy between observations 1st+2nd and 4th suggests that there may be more factors at play than just people on average preferring the 3D representations, this should be investigated more thoroughly. Along this line of thought, the relation between perceived video quality and the usage of 3D content is also an interesting link to investigate.

Extra Material

Technical Performance reporting is done in the MMsys 2018 demo paper:

https://drive.google.com/open?id=1I8pF_N2adNKJLL8mBIqZgXEsiySbac9Q

Questionnaire feedback can be found here:

Raw answers: https://drive.google.com/file/d/1WVnrVAwriE8qoOiSDUJt-iheGYiKwuar/view?usp=sharing

Google form diagrams:
https://drive.google.com/open?id=1mxZcbNRooYur1gYZDVVY0f8QrZFqreaX

References:

[Gunkel 1, 2017] Simon Gunkel, Martin Prins, Hans Stokking, and Omar Niamut. 2017. WebVR meets WebRTC: Towards 360-degree social VR experiences. In Virtual Reality (VR), 2017 IEEE. IEEE, 457–458.

[Gunkel 2, 2017] Simon NB Gunkel, Martin Prins, Hans Stokking, and Omar Niamut. 2017. Social VR Platform: Building 360-degree Shared VR Spaces. In Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video. ACM, 83–84.

[MacKerron] http://blog.mackerron.com/2012/02/03/depthcam-webkinect/

[Archilogic] https://archilogic.com/

4. Pilot 1

The VR-Together project is illustrated through three pilots that address specific objectives in terms of technical challenges and evaluation purposes. Pilots are project checkpoints to evaluate the creative and technical challenges identified towards the creation of truly realistic social VR experiences. The structure and plot complexity of the pilots is linked to a gradually increasing technical difficulty, being the first pilot the most simple to produce and in terms of technological aspects, and the third one the most complex.

These three pilots were initially planned as follows: A first offline pilot, intending to offer, not only the feeling of being together, but also intimacy and closeness. The second pilot, simulating a live production of immersive content from multiple sources that aimed to virtually transfer the user to the location of the news and share the experience with other users. Lastly, the third pilot will present a test to users through an interactive and totally immersive experience, where users can participate in the scene, interact between them, make conclusions, etc.

All three pilots will tell a great story about a murder investigation. In the first pilot of the VR-Together project, two users are located in a 70's look-a-like police interrogation room, behind a two-way mirror, where they can observe the place where interrogations are going to take place in. Even though the participants are located in the same room, they are separated from each other, because each one will enjoy a different part of a police interrogation. Both users are able to see each other in the room created specifically for this.

The users, represented as inspectors of the case, are witnesses of the interrogations made to the main suspects of a murder. Each user will be able to listen to a witness and, subsequently, interact with each other to deduce who is guilty of the murder. In this way the sense of togetherness is fostered, since users can see each other not only during the interrogation, but also after the interrogation is finished. The interaction between them is necessary to reach a conclusion.

Both users will have to pay attention to every move the witnesses are doing, what they say and how they act, in order to make a decision about who they think is guilty of the murder of Ms. Armova. The experience presented by the project VR-Together allows to experience the sensation of togetherness in a virtual reality environment.

Our aim and goal is getting the users to experience a togetherness feeling, seeing each other and knowing that the other user is experiencing a slightly different adventure, since each one will have to pay attention to a different witness. Both of them won't hear the same story of the murder, and after hearing what the witnesses have to say, they will have to agree on what they heard and saw, translating the experience not only in the artificial environment created, but also afterwards, making the most of the experience. As the name of the project itself, we want to transmit a feeling of togetherness before and after of the experience lived by the users.

Figure 51 illustrates the release dates for the pilots, including pilot 1 and the upcoming ones.

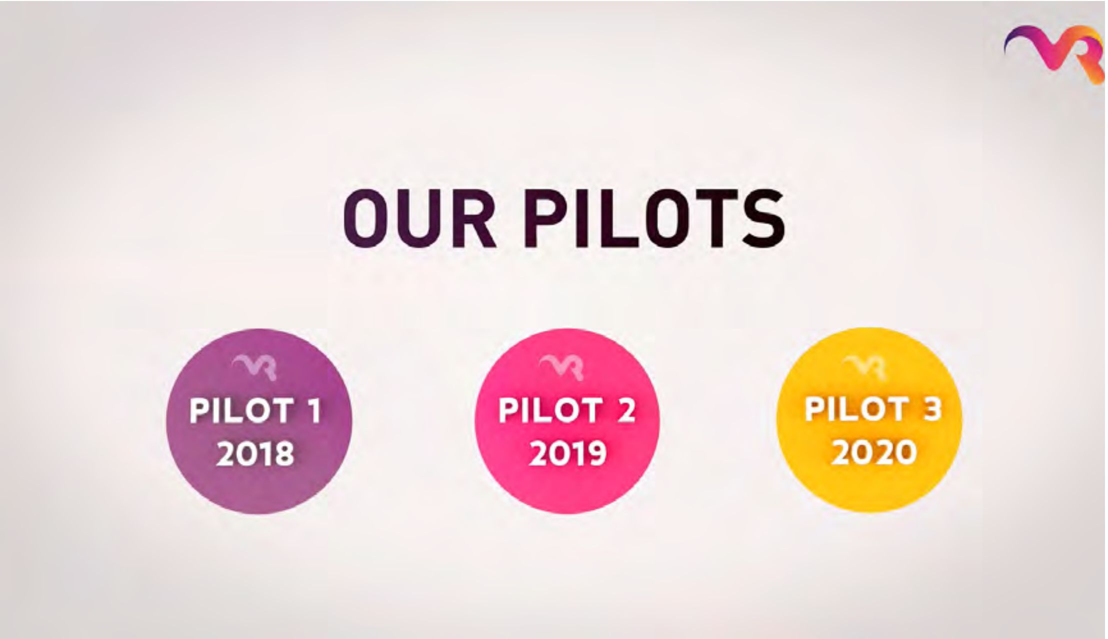


Figure 51- Release dates of the VR-Together pilots

4.1. Scenario and Content Creation

Please refer to D4.1 for all the details regarding the content production and post-production process. Here we provide the list of all the contents created, which have been uploaded to the Zenodo server.

- Media Unity 3D Scene GLTF Format (https://zenodo.org/record/1452519)
- Media Unity 3D Scene Unity format (https://zenodo.org/record/1456486)
- Media User 1 _ 360 Stereo Panorama (https://zenodo.org/record/1456521)
- Media User 2 _ 360 Stereo Panorama (https://zenodo.org/record/1456521)
- Media User 1 _ 360 Mono Panorama (https://zenodo.org/record/1456521)
- User 1 _ Video Billboard (https://zenodo.org/record/1456687)
- User 2 _ Video Billboard (https://zenodo.org/record/1456785)
- Ambisonic track _ User 1 (https://zenodo.org/record/1456542)
- Ambisonic track _ User 2 (https://zenodo.org/record/1456542)
- User 1_360 Mono V (https://zenodo.org/record/1456866)
- User 2_360 mono video (https://zenodo.org/record/1457163)
- User 1_360 stereo video
- User 2_360 stereo video

During the post-production months (from April to June 2018), Entropy Studio prepared and edited a promotional video to help to explain what is the VR-Together project and the story behind the first pilot. In less than 4 minutes, the whole process is explained, trying to detail every single part of the preparation, shooting and post-production of the first pilot of the project.

This video was made thinking about being shown in different congresses, events and to teach the members that make up the Advisory Board of the project how it was evolving and the decisions that were being taken at the artistic and technical level.

The video also intends to act as a "behind-the-scene" video, where it can be seen how the shooting was prepared, paying special attention to the motion capture process. The video can be watched by clicking here (https://www.youtube.com/watch?v=aHO5M1qNmjY)

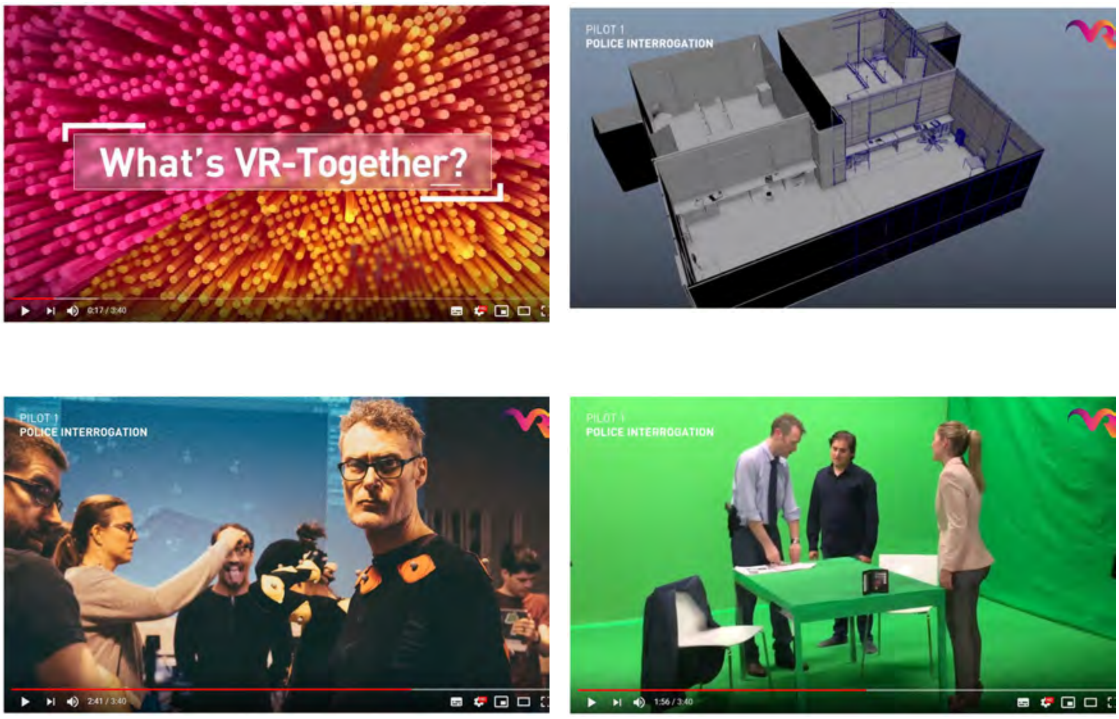


Figure 52 – Screenshots from the Video

To give more details and a more explained version of the previous video, a longer version of it is being developed. This video will showcase interviews from every partner of the project. The release date is to be announced soon. Table 13 shows the name of the organization and the people who got interviewed.

Table 13- Organizations and people who were interviewed

Organization	Person
i2CAT	Sergi Fernández, Mario Montagud
TNO	Tom Koninck
Viaccess-Orca	Patrice Angot, Vincent Lepec

Motion Spell	Romain Bouqueau
Artanim	Caecilia Charbonnier
CWI	Francesca De Simone
CERTH	Anargyros Chatzitofis
Entropy Studio	Ignacio Lacosta

4.2. Pilot with End-Users in Barcelona

This section reports on the pilot with end-users conducted in October 2018, using the end-to-end VR-Together platform and the contents created for pilot 1. The pilot was conducted in a Living Lab in Barcelona (see Figure 53) in October 2018.

Figure 53- Living Lab where the VR-Together experiments for pilot 1 where conducted

In particular, the section provides details about the evaluation goals and setup (system configuration and content variants being used) in these tests, the protocol that was defined for the evaluation, based on the methodology described in Section 2 and on the insights from previous experiments and the obtained results (Section 3), paying special attention to the subjective evaluation. Brief conclusions are also provided.

4.2.1. Goals

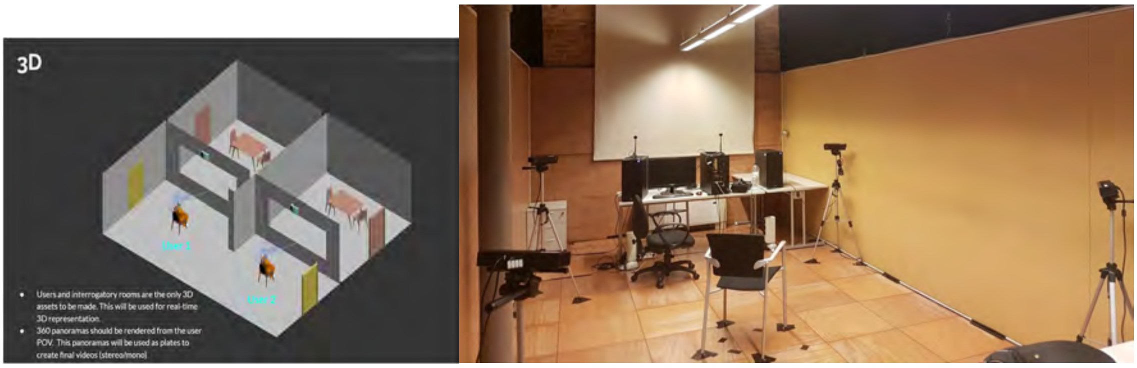
As mentioned, pilots are project checkpoints to evaluate the creative and technical challenges identified towards the creation of truly realistic social VR experiences. The first pilot intends to offer not only a *feeling of being there*, but also a *feeling of being there together*, intimacy and closeness between the participants.

The pilots with end-users, making use of the end-to-end VR-Together platform (see D2.4) and the contents created for pilot 1 (see D4.1), aim at: 1) validating the technological developments; 2) validating and refining the evaluation methodology defined in the project, and being (partially) used in previous experiments; and 3) assessing the appropriateness of the technology and created scenarios / contents to provide these mentioned benefits and feelings to the users. Likewise, the pilots with end-users are also a great opportunity to gather relevant feedback from the users, like about their behaviours, perception, their (potential use cases of) interest, as well as for getting insights about the features that need to be improved throughout the duration of the project and other additional suggested ones.

4.2.2. Setup

Different systems configurations for the VR-Together platform (more details in D2.4) and different variants of the contents (more details in D4.1) have developed and created, respectively, in the project. The tests in Barcelona were prepared with the use of the native and TVM pipeline as the platform configuration, and using the Unity3D scene with stereo billboards for the interrogation rooms as the VR contents.

According to the scenario ideated for pilot 1, two different rooms for each one of the two users involved in the experiment were required (see picture on the left in Figure 54). For each room, 4 Kinects and 5 (1 per Kinect + 1 controller) PCs were required for the TVM-based end-user's reconstruction (see picture on the right in Figure 54). A laptop was also used to record the audio and video from each participant via its integrated webcam (see picture on the left in Figure 55). Both rooms were located in different dependences (different levels) of the Citilab Living Lab, interconnected via its networking infrastructure. The rooms had no background / surrounding noise. Each user was equipped with an Oculus Rift, with an integrated microphone for the audio interaction and noise-cancelling headphones to experience with the positional audio (from the other user). The users had to sit in a chair placed in the centre of the effective capturing region (see picture on the right in Figure 54). The two rooms, with the participants immersed in the created shared experiences, are shown in Figure 55.



- Users and interrogatory rooms are the only 3D assets to be made. This will be used for real-time 3D representation.
- 360 panoramas should be rendered from the user POV. This panoramas will be used as plates to create final videos (stereo/mono)

Figure 54- Virtual (left) and real (right) setup of the pilot 1 tests.

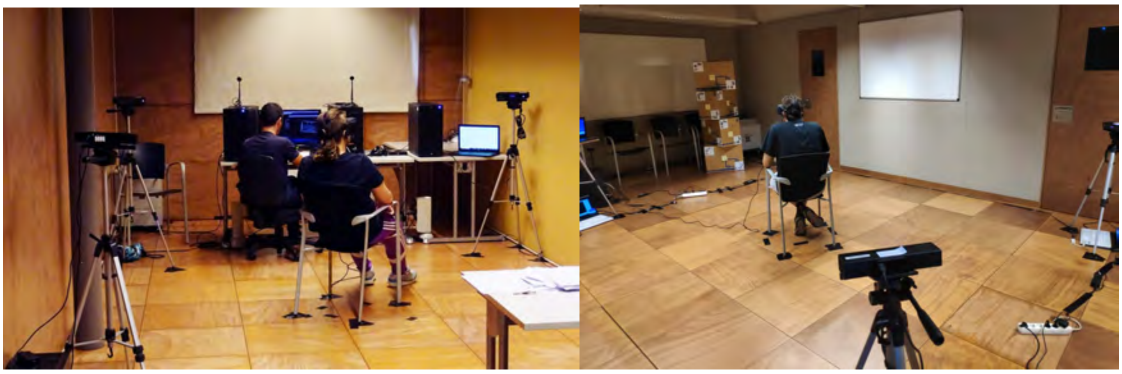


Figure 55- The two rooms with two participants immersed in the shared VR experience

4.2.3. Evaluation Protocol and Procedure

As mentioned, the tests were conducted by pairs. The users were recruited based on the following criteria:

- The participants had to be older than 18 years old
- The participants needed to have a good English level
- The participants for each pair had to know each other.

The tests consisted of the following steps:

Step 1 (~10min). The facilitators welcome the participants, and briefly describe them the project, the story and evaluation process.

After the explanations, the participants were asked to fill in:

- A Background Info Questionnaire
- A Consent Form
- A Societal Anxiety Questionnaire

The participants were informed that their participation was voluntary, and that they could stop the experiment at any time, if they would like to do so, for whatever reason.

Step 2 (~5min). The participants were brought to each one of the rooms. With the help of the facilitator(s), they were equipped with HMDs and audio headsets, and indicated where to sit down (see Figure 55).

Step 3 (~15min). The facilitator started the recording of the audio+video from the laptop and then launched the shared VR experience, after having confirmed everything was ready in the other room (via Slack communication with the other facilitator).

The participants were instructed to feel free to interact and talk to each other before, during and after the interrogations that they were going to watch. In particular, three main parts in the shared watching experience can be differentiated:

- *Initial Phase*: the participants are initially immersed in the virtual scenario to get familiar with it, and after a short period, the interrogations start. This initial period is expected to trigger high interactions in order to exchange the first impressions.
- *Interrogation Phase*: during the interrogations, participants can also freely interact and share insights/impressions. Triggers, like short light off scenes, images from a security camera from the day of the murder, and cross-references between the interrogation scenes were added to stimulate this.

- *End Phase*: After the interrogations, the participants have extra time to interact within the shared VR environment and to explore it. Richer interactions can happen at this phase, because participants have already experienced the whole interrogation scenes, so they know the whole story. In addition, they feel more relaxed at this stage, as no cross-audio or extra contents are provided then.

Step 4 (~2min): With the help of the facilitator(s), participants stepped out of VR.

Step 5 (~10min): Participants filled in a questionnaire about their experience (see Metrics).

Step 6 (~15min): The facilitators coordinated a semi-structured interview and discussed about the experience with the participants. The interview's audio was recorded.

Step 7 (1min): The facilitators thanked the participants, gave a reward and say goodbye to them.

Overall, the duration of the tests was around 1 hour.

The employed questionnaires and forms in Pilot 1 are included in Annex X.

4.2.4. Results

In this sub-section, the participants sample and results are reported.

Sample and Participants' Data

30 participants took part in the tests. Next, background information about the participants, and pairs, is provided:

- Aged between 18 and 35 (average = 21.66, standard deviation = 3.88) years
- 27 males and 3 females
- 5 were left handed and 25 right handed
- None of them expressed to have audio-visual impairments

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

- 1 participant stated to be novice regarding the use of computers, 12 intermediate and 17 experts.
- Only 3 participants stated not having previous experience in VR, 22 affirmed to have some experience, and 5 expressed to be very experienced. The VR products being used in the past by participants are indicated in Figure 56 (multiple answer option). When referring to *Other products*, 2 participants indicated Oculus Go and 3 indicated cardboards with smartphones.

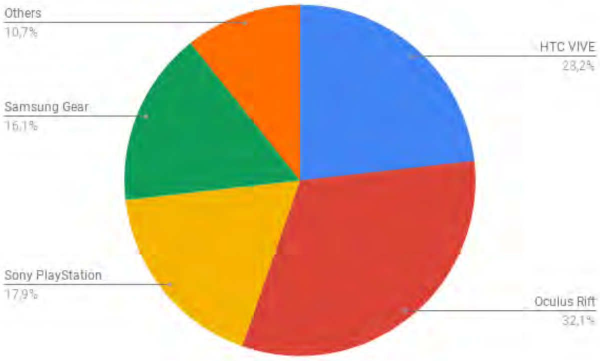


Figure 56- VR products with which participants had previous experience

Finally, the participants / pairs were asked about their relationships to better understand, and potentially correlate, their behaviour in the VR environment and their willingness in enjoying this kind of shared experiences:

- 2 pairs indicated to be friends, while the other 13 stated to be colleagues (including classmates in this category).
- 1 pair indicated to know each other since 4-5 years ago, 3 since 1-3 years ago, and 11 since less than 1 year ago.
- Regarding their main contact method (multiple option answer), all the participants indicated face-to-face, although Social Media (mainly Facebook and WhatsApp) was commonly indicated too. Phone call was indicated by 1 pair.
- Similarly, talk to each other was indicated to be the main shared activity to maintain the relationship for all participants, but other activities were additionally indicated (multiple option answer), such as: go to cinema (1 pair), use cooperation tools, like Slack and WhatsApp (5 pairs), play computer games (3 pairs), and have a drink (3 pairs).

Note that the participants' sample does not include a huge variety in terms of ages, and few females took part in the tests. It was due to the fact that the Living Lab recruited the participants from two courses related to VR development and 3D animation. Although this became an issue in terms of narrower diversity in the participants sample (ages, gender balance, different relationships between them…), this also turned into an opportunity to test the system and contents with experienced and very interested users, and to get useful feedback from them (confirmed especially in the interviews).

Results from Societal Anxiety Questionnaire

As mentioned, a Societal Anxiety Questionnaire was passed to the participants in order to get insights about their willingness, feelings and attitudes in social settings. This can be very useful to better understand their interest and behaviour in social VR scenarios.

The results from such a questionnaire are summarized in Table 14.

Table 14 - Results from Societal Anxiety Questionnaire

Question	Yes	No

Q1. I feel relaxed even in unfamiliar social situations	20 (66.66%)	10 (33.33%)
Q2. I try to avoid situations which force me to be very sociable	10 (33.33%)	20 (66.66%)
Q3. It's easy for me to relax when I am with strangers	19 (63.33%)	11 (36.66%)
Q4. I have no particular desire to avoid people	13 (43.33%)	17 (56.66%)
Q5. I often find social settings upsetting	6 (20%)	24 (80%)
Q6. I usually feel calm and comfortable in social situations	21 (70%)	9 (30%)
Q7. I am usually at ease when talking to someone of the opposite sex	26 (86.66%)	4 (13.33%)
Q8. I try to avoid talking to people unless I know them well	7 (23.33%)	23 (76.66%)
Q9. If the chance comes to meet new people, I often take it	26 (86.66%)	4 (13.33%)
Q10. I often feel nervous or tense in casual get-togethers in which both sexes are present	2 (6.66%)	28 (93.33%)
Q11. I am usually nervous with people unless I know them well	11 (36.66%)	19 (63.33%)
Q12. I usually feel relaxed when I am with a group of people	22 (73.33%)	8 (26.66%)
Q13. I often want to get away from people	7 (23.33%)	23 (76.66%)
Q14. I usually feel uncomfortable when I am in a group of people I don't know	14 (46.66%)	16 (53.33%)
Q15. I usually feel relaxed when I meet someone for the first time	18 (60%)	12 (40%)
Q16. Being introduced to people makes me tense and nervous	9 (30%)	21 (70%)
Q17. Even though a room is full of strangers I may enter it anyway	21 (70%)	9 (30%)
Q18. I would avoid walking up to and joining a large group of people	14 (46.66%)	16 (53.33%)
Q19. When my superiors want to talk to me, I talk willingly	23 (76.66%)	7 (23.33%)
Q20. I often feel on the edge when I talk to a group of people	7 (23.33%)	23 (76.66%)
Q21. I tend to withdraw from people	2 (6.66%)	28 (93.33%)
Q22. I don't mind talking to people at parties or social gatherings	19 (63.33%)	11 (36.66%)
Q23. I am seldom at ease in a large group of people	12 (40%)	18 (60%)
Q24. I often think up excuses in order to avoid social engagements	4 (13.33%)	26 (86.66%)
Q25. I try to avoid formal social occasions	5 (16.66%)	25 (83.33%)
Q26. I usually go to whatever social engagements I have	22 (73.33%)	8 (26.66%)
Q27. I find it easy to relax with other people	27 (90%)	3 (10%)

Results from Experience Questionnaire

After the shared VR experience, each pair was asked to complete the *Experience Questionnaire* (see Annex X), which included questions about their emotions, feelings, perception and opinion regarding crucial aspects of VR-Together (see Figure 57), categorized as (more details in Section 2):

- 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…).

The results are categorized and summarized in Table 15, and the meaning of the acronyms used in (the first row of) that table are defined in Table 16.

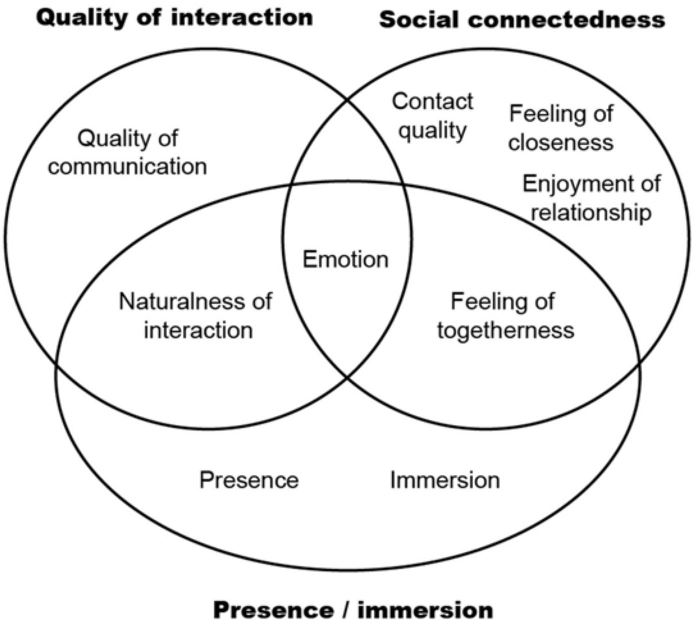


Figure 57- Evaluated aspects in the Experience Questionnaire.

Table 15- Results from the Experience Questionnaire.

Questions	TD	PD	NN	PA	TA
Part 1. Quality of Interaction					
Q2. “I was able to feel my partner’s emotion while watching the contents.”	0	4	8	**17**	1
Q3. “I was sure that my partner often felt my emotion.”	2	4	**11**	**10**	3
Q4. “The experience of watching the contents with my partner seemed natural.”	0	1	9	**14**	6
Q5. “The actions used to interact with my partner were similar to the ones in the real world.”	0	4	**11**	**10**	5
Q6. “It was easy for me to contribute to the conversation with my partner.”	0	1	2	**15**	**12**
Q7. “The conversation with my partner seemed highly interactive.”	2	2	7	**14**	5
Q8. “I could readily tell when my partner was listening to me.”	0	7	1	**17**	5

Q9. “I found it difficult to keep track of the conversation.”	5	**7**	**13**	5	0
Q10. “I felt completely absorbed in the conversation.”	0	1	9	**14**	6
Q11. “I could fully understand what my partner was talking about.”	0	0	2	**14**	**14**
Q12. “I was very sure that my partner understood what I was talking about.”	0	1	4	**16**	**9**
Q13. “I often felt as if I was all alone while watching the contents.”	6	**12**	7	5	0
Q14. “I think my partner often felt alone while watching the contents.”	5	**9**	**11**	5	0
Part 2. Social Connectedness					
Q15. “I often felt that my partner and I were sitting together in the same space.”	1	3	6	**11**	**9**
Q16. “I paid close attention to my partner.”	0	2	**13**	**13**	3
Q17. “My partner was easily distracted when other things were going on around us.”	1	**9**	**11**	6	3
Q18. “I felt that watching the contents together in VR enhanced our closeness.”	1	0	8	**15**	6
Q19. “Watching the contents together created a good shared memory between me and my partner.”	1	2	5	**18**	4
Q20. “I derived little satisfaction from the content watching experience with my partner.”	3	5	**12**	7	3
Q21. “The content watching experience with my partner felt superficial.”	1	**17**	8	4	0
Q22. “I really enjoyed the time spent with my partner.”	0	1	3	**15**	**11**
Q24. “In the virtual world I had a sense of ‘being there’.”	0	0	6	**16**	**8**
Q25. “Somehow I felt that the virtual world was surrounding me and my partner.”	0	1	6	**15**	**8**
Q26. “I had a sense of acting in the virtual space, rather than operating something from outside.”	0	1	**12**	**12**	5
Q27. “My content watching experience in the virtual environment seemed consistent with a real world experience.”	0	1	11	**17**	1
Q28. “I did not notice what was happening around me in the real world.”	0	3	5	**13**	**9**
Part 3. Presence / Immersion					
Q29. “I felt detached from the outside world while watching the contents.”	0	2	7	**12**	**9**
Q30. “At the time, watching the contents with my partner was my only concern.”	0	4	9	7	**10**
Q31. “Everyday thoughts and concerns were still very much on my mind.”	5	6	**11**	6	2
Q32. “It felt like the content watching experience took shorter time than it really was.” *[Duration of contents is ~7min]*	0	2	3	**11**	**14**
Q33. “When watching the contents with my partner, time appeared to go by very slowly.”	**11**	6	9	2	2
Extra Questions					
Q34. “I liked the created VR contents.”	0	0	2	**11**	**17**
Q35. “The created VR contents are realistic (i.e. resemble a real scenario).”	1	0	2	**23**	4
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.”	0	1	10	**11**	8

Table 16 - Acronyms used in (the first row) of Table 15.

Acronym	Meaning
TD	Totally Disagree
PD	Partially Disagree
NN	Neither Agree nor Disagree
PA	Partially Agree
TA	Totally Agree

From the results of Table 15, it seems that the VR-Together platform and contents were able to provide the targeted emotions and feelings to the participants, and that they were quite satisfied and impressed in general. These results will be discussed later, and a statistical analysis will be conducted when evaluating the other system configurations to be able to compare them.

Apart from the questions in Table 15, two additional questions were included in the *Experience Questionnaire*, Q1 and Q23. On the one hand, Q1 consisted of asking the participants about their emotions and their partners' emotions, by using a diagram with 8 types of emotions (4 positive and 4 negative ones), with a scale 0-100. Table 17 shows the results from such a question.

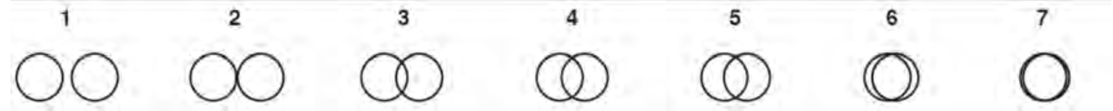
In general, it seems evident that participants experienced more positive than negative emotions, due to the higher scores on the one hand, and that the positive ones were also more commonly reported by participants on the other hand. Interestingly, it seems that the participants generally felt that their partners were happier, more relaxed and calm than themselves. Similarly, slight differences between scores reflect that participant might felt that their partners were less tense and nervous compared to themselves. This could be due to the fact that the participant felt tense or pressure about what to say or how to behave / react during the shared VR experience. However, this feeling was not explicitly expressed by participants, neither in Q1 nor in the final interviews.

Overall, the results corroborate the previous findings from CWI-2 experiment. Statistical analysis to determine significant differences between partners, partners' profiles and between test conditions, will be conducted after evaluating the other system configurations.

Table 17- Reported Levels of Emotions.

Emotion	Own Emotions			Partner's Emotions		
	Avg. Value	Std. Value	Nº Answers	Avg. Value	Std. Value	Nº Answers
Exited/Lively	74.14	17.4	25	73.12	11.49	25
Cheerful/Happy	75.57	11.89	27	78.89	12.55	25
Relaxed/Carefree	78.19	16.91	24	79.21	18.02	26
Calm/Serene	79.64	17.99	24	83.28	10.95	22
Bored/Weary	15.17	7.07	14	21.43	11.95	12
Gloomy/Sad	8.92	5.05	14	14.28	14.28	12
Irritaded/Annoyed	9.28	6.77	15	11.9	9.52	12
Tense/Nervous	28.23	14.08	23	27.8	26.19	19

On the other hand, Q23 asked about *how emotionally close to the partner did each partner feel*, 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, it seems that participants did feel neither too far nor completely overlapped. Instead, they reported on different levels of closeness with different amount of overlapping, which is a sign of the feeling of togetherness and intimacy (although with different intensity).



0	4	11	7	2	5	1

Figure 58- Emotional closeness between participants.

Results from Participant's Activity / Behaviour

TBC: The analysis of results of the participants' activity / behaviour is ongoing.

Objective Performance Results

TBC: The analysis of the results of the objective evaluation is ongoing.

Regarding performance metrics, the TVM pipeline (capturing, encoding, distribution, decoding and rendering) is the most challenging part of the evaluated system configuration of the VR-Together platform. The results for different resolution and texture downscale configurations, in terms of delay, jitter and bandwidth consumption, for different components along the end-to-end VR-Together platform, are reported in CERTH-2 (section 3.1.1).

Results from Interviews

TBC: The analysis of the results of the semi-structured interviews is ongoing.

4.3. Pilot with Remote End-Users (Spain – Greece)

Apart from local tests with end-users, remote communication tests with end-users were conducted, with the same Native + TVM pipeline and pilot 1 contents. 3 researchers from i2CAT and 3 researchers from CERTH were the participants of such tests, in three different sessions. In each of the sessions, 1 participant of the shared VR experience was located at i2CAT premises (Barcelona, Spain) and the other one at CERTH premises (Thessaloniki, Greece). The researchers knew each other, were involved in the project and had previously experienced the local setup for the sake of comparison.

Goals of the Experiments

The goals of these remote communication tests were:

- To validate the communication between the two remote sites.
- To gather objective performance metrics. This will allow gaining insights about the requirements and limitations of the VR-Together platform to enable shared VR experiences between remote participants.
- To allow the researchers experiencing in first person the perceived performance and media quality in such remote scenarios. This will allow better understanding the technological aspects that already perform satisfactorily and the ones that need to be optimized to enable social VR scenarios for remote users.

Methodology

The methodology briefly consisted of:

- The remote participants watched the contents.
- The objective performance metrics were registered, as for the local tests.
- The participants had to fill in a newly created Experience Questionnaire for the remote communication tests.
- The participant had a final short discussion about the experience, perception aspects related to it, and things to be improved in next iterations of the platform, considering their priority levels and/or importance for an enjoyable experience.

The employed questionnaire in Remote Communication Test (I2CAT <--> CERTH) is included in Annex XI.

Objective Performance Results

TBC: The analysis of the results of the objective evaluation is ongoing.

Regarding performance metrics, the TVM pipeline (capturing, encoding, distribution, decoding and rendering) is the most challenging part of the evaluated system configuration of the VR-Together platform. The results for different resolution and texture downscale configurations, in terms of delay, jitter and bandwidth consumption, for different components along the end-to-end VR-Together platform, are reported in CERTH-2 (section 3.1.1).

Results from Questionnaires

Technical Issues

Q1. Did you encounter any technical issues in the remote communication test?

If yes, please briefly describe the issues and if/how you solved them

In the three sessions, there were two technical issues that not prevented the realization of the tests. In one of the sessions, the audio communication was interrupted after few seconds of starting the experience. The session was re-started and the communication was stable and good for the second session. [The cause of this issue was inspected by the researchers in order to solve it for future releases of the platform].

In the second session, there was a sporadic freezing effect in the remote participant's representation (the one at CERTH premises), but the fluidity of playout was soon recovered, and the session continued smoothly until the end.

Questions about Audio Quality

Q2. Compared to the Local Scenario Test, the perceived audio quality for the communication between participants is:

No differences have been perceived	A bit lower, but still acceptable	Much lower, and needs to be improved
5	1	-

Q3. In general, the perceived audio quality for the communication between participants is satisfactory:

Totally disagree	Partially disagree	Neither agree nor disagree	Partially agree	Totally agree
-	-	-	3	3

Questions about Video Quality

Q4. Compared to the Local Scenario Test, the perceived video quality for the remote end-user's reconstruction is:

No differences have been perceived	A bit lower, but still acceptable	Much lower, and needs to be improved

2	4	-

Q5. Compared to the Local Scenario Test, the fluency and naturalness of the movements and gestures of the reconstructed remote end-user are:

No differences have been perceived	A bit worse, but still acceptable	Much worse, and needs to be improved
2	4	-

Q6. The movements and gestures of the reconstructed remote end-user are perceived as fluent and natural:

Totally disagree	Partially disagree	Neither agree nor disagree	Partially agree	Totally agree
-	1	-	4	1

Q7. In general, the perceived video quality for the remote end-user's reconstruction is satisfactory:

Totally disagree	Partially disagree	Neither agree nor disagree	Partially agree	Totally agree
-	1	-	4	1

Questions about Delays and Synchronization

Q8. Compared to the Local Scenario Test, the perceived delays for the end-to-end audio communication are:

Same order	A bit higher, but still acceptable	Much higher, and need to be reduced
4	2	-

Q9. Compared to the Local Scenario Test, the perceived delays for the end-to-end video communication is:

Same order	A bit higher, but still acceptable	Much higher, and need to be reduced
1	5	-

Q10. The synchronization levels between the end-to-end audio and video interaction channels between users are satisfactory

Totally disagree	Partially disagree	Neither agree nor disagree	Partially agree	Totally agree
-	-	1	2	3

Q11. The synchronization levels between the audiovisual contents from the end-users' capturing/reconstruction and the ones for the shared environment are satisfactory
[For instance, think whether the users' comments and movements were time-aligned with the contents being presented, like users reacting immediately to event, triggers and questions]

Totally	Partially disagree	Neither agree nor	Partially agree	Totally agree

disagree		disagree		
-	-	-	3	3

Q12. Based on the experience, I believe that synchronization between participants will be a key requirement for pilot 2, in which more participants will be involved in the shared social VR scenario. *[Note that DVB-CSS support has been added just before the pilots]*

Totally disagree	Partially disagree	Neither agree nor disagree	Partially agree	Totally agree
-	-	-	2	4

Questions about the Experience

Q13. “The audiovisual communication channels between users enable high quality and interactive conversations”.

Totally disagree	Partially disagree	Neither agree nor disagree	Partially agree	Totally agree
-	-	-	1	5

Q14. “I often felt that my partner and I were together in the same space”.

Totally disagree	Partially disagree	Neither agree nor disagree	Partially agree	Totally agree
-	-	-	1	5

Q15. In general terms, the VR-Together platform already enables satisfactory shared and interactive experiences between remote participants.

Totally disagree	Partially disagree	Neither agree nor disagree	Partially agree	Totally agree
-	-	-	1	5

Final Discussion between Researchers

At the end of the Questionnaire, participants were asked to discuss and indicate their thoughts about the technological aspects that already perform satisfactorily and the ones that need to be optimized to successfully enable social VR scenarios for remote participants. They were indicated to pay special attention to performance issues influenced by the existence of bandwidth and delay limitations in the remote scenarios. In addition, they indicated specific directions/actions to optimize the system performance and improve the overall experience for next pilots.

Summary of the Discussion

The researchers that participated in the remote shared VR experience were quite satisfied, in general, and believed that the already developed technological components enable to communicate thoughts and have a feeling of togetherness in a shared scenario.

Regarding the issues to be improved in next releases of the platform, the researchers highlighted:

- *The mentioned issues with the audio communication.*
- *The delays in the TVMs, especially the ones for the self-representation are noticeable.*
- *The visual quality for the TVMs, especially for the self-representation. Loss of details (fingers) and artifacts (holes in the body) were perceived.*
- *Provide more interaction features with the environment, and enable 6 DoF to explore the environment.*

Analysis

By taking into account the results from the Experience Questionnaire, it can be concluded that the researchers that participated in the remote communication tests were quite satisfied in general with the experience and perceived quality and feelings. This was especially true for the audio communication, in terms of delays and perceived quality. For the video-based interaction channel, using the TVM pipeline, the researchers noticed a slightly lower performance in terms of delays, quality and fluidity of the movements / gestures. However, the perceived quality and performance was still acceptable to them. The researchers were also quite satisfied with the perceived inter-media synchronization accuracy, and believe that synchronization across participants will be a key requirement for next pilots. Finally, the researchers believe that the developed technological components enable high quality and interactive conversations, as well as to communicating thoughts and providing a feeling of togetherness in shared VR scenarios.

4.4. Pilot with Professionals

In addition to the pilot action in Barcelona, the pilot has been:

- Showcased at IBC2018 and VRDays2018
- Screened for participation in Sundance Festival 2018

Pilot action at IBC2018

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

Research questions

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

Hypothesis

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

Scenario

At IBC 2018, we set up the web-based Pilot 1 version. This consists of 2 users, experiencing the pilot 1 content of the police interrogation, while being virtually together in the police station. Users were seated for the experience, we helped them with the setup of their equipment, consisting of an Oculus Consumer Version HMD and a Sony noise-cancelling headphone. After they were comfortable, we started the pilot 1. The pilot one consisted of an introduction by a virtual video avatar, in this case a TNO colleague explaining the demo inside the demo. After that, the virtual avatar disappeared, and the 2 users were moved to the virtual seating behind the virtual mirror. Next, the 2 parallel interrogations were shown, after which the users were again moved closer together to discuss what they saw. After a short discussion, we ended the experiment and helped people with taking of the equipment again.

Afterwards, people were asked to fill in an online questionnaire. For this purpose, we had 2 iPads to easily fill out the questionnaire. The questions for this questionnaire were derived from the questionnaire for professionals. A more limited set was chosen, on the one hand to be more aligned with the expected visitors, on the other to limit the time asked of participants, as participants at a trade show normally do not have too much time.

The employed questionnaire in Pilot with Professionals (IBC 2018) is included in Annex XII. The link to the Google Form is: https://goo.gl/forms/Dp5T9ktKKnysDZGF3

User panel: size and characteristics

We conducted our requirements gathering at the IBC 2018 in Amsterdam. In this way, we ensured that our participants at this stage are people who have an interest in media and entertainment, and likely have experience using VR applications. The participants had the following characteristics (totals not adding up, as not all participants filled out all questions, we did not make any questions mandatory):

- Total participants: 109
- Gender distribution: 14 F, 87 M, 2 N/A
- Age range:
    - Below 20: 21
    - 26-35: 32
    - 36-45: 25
    - 46-55: 16
    - 56-65: 5
    - Above 65: 2

Environment

The demonstration was held at the booth from the project I3 European Media Innovation. The demonstration area for the VRTogether project was approximately 4 by 3. At the back, a table was set up with two PCs, two Kinects on tripods and a large screen. About 2 meters away, two chairs were set up in front of the kinect cameras. The large screen showed the viewpoint of one of the participants. On a separate table at the front of the booth, a large screen continously showed the VRTogether promo movie. The setup is shown in Figure 59.

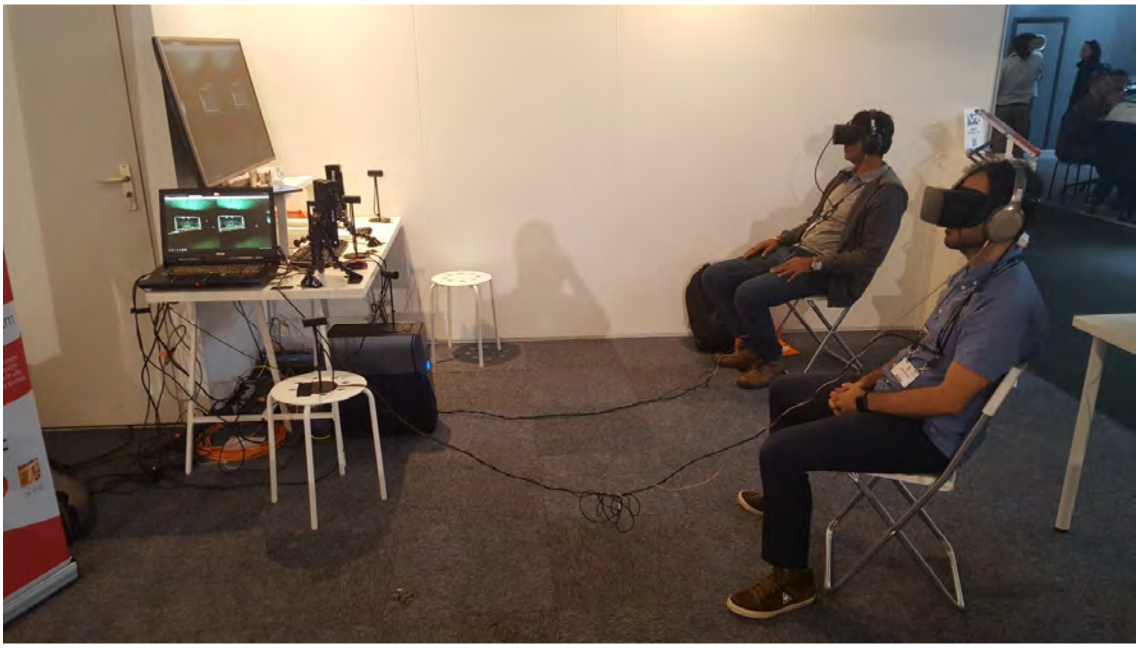
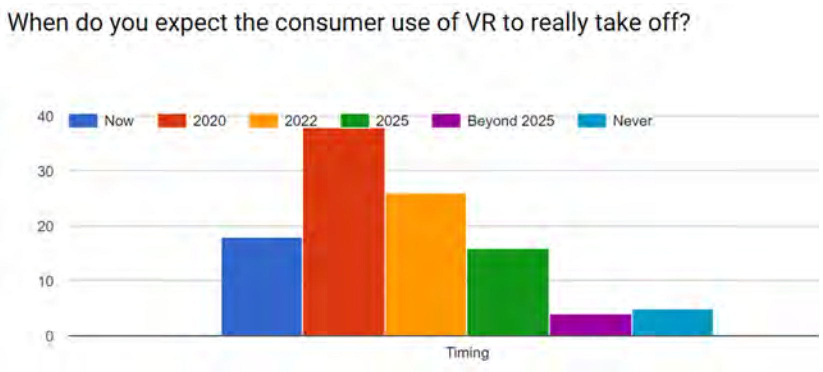


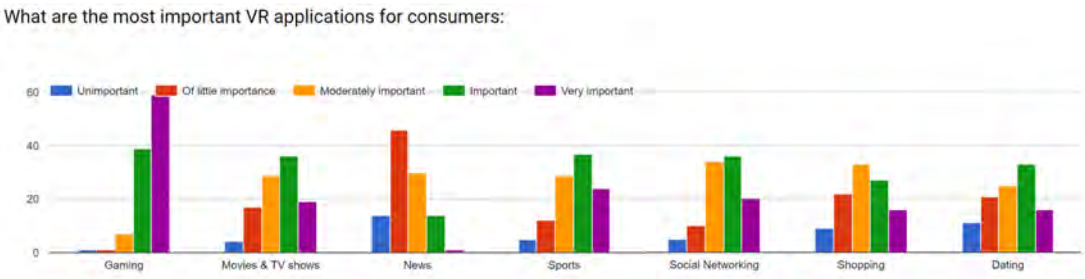
Figure 59 The VRTogether demonstration area at the I3 booth

Results

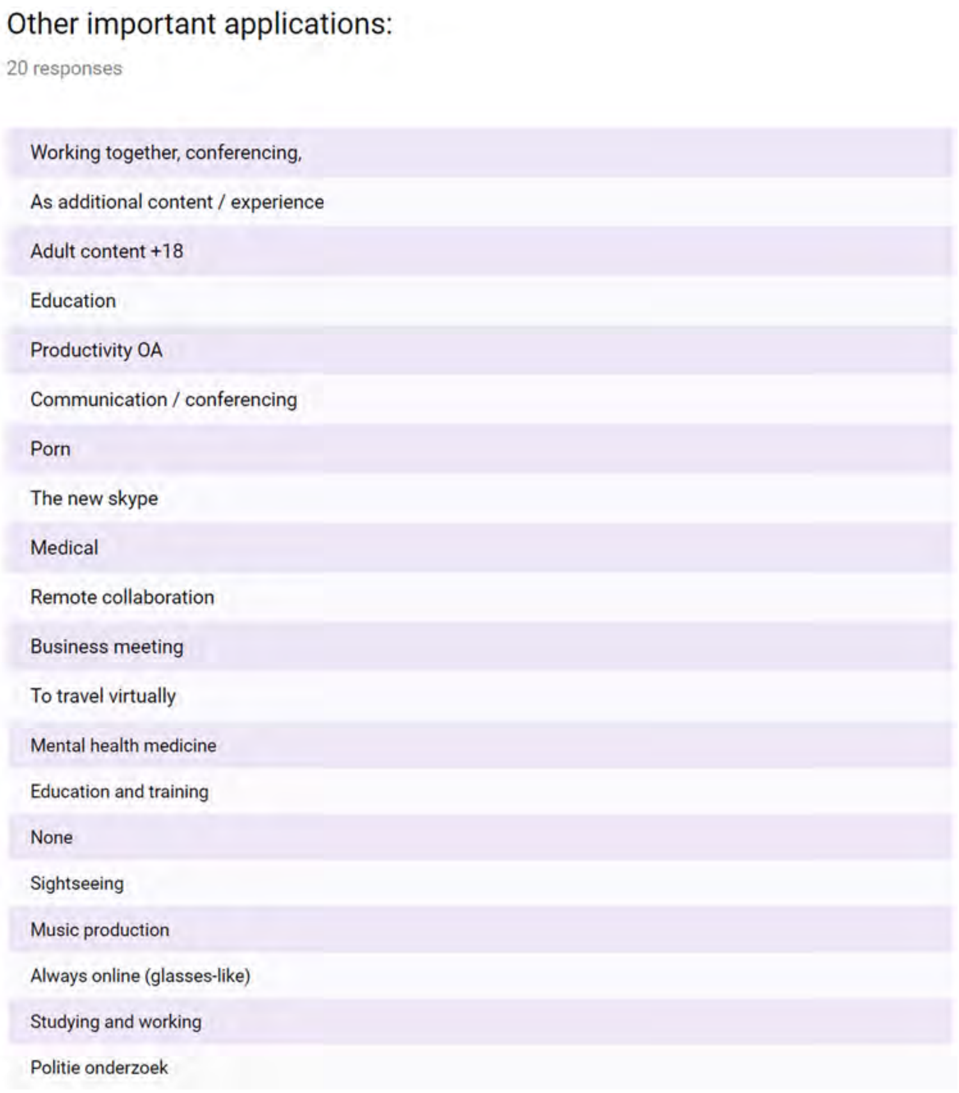
A summary of the results is presented in Figure 60.



When do you expect the consumer use of VR to really take off?



What are the most important VR applications for consumers:



Other important applications:

20 responses

Working together, conferencing,

As additional content / experience

Adult content +18

Education

Productivity OA

Communication / conferencing

Porn

The new skype

Medical

Remote collaboration

Business meeting

To travel virtually

Mental health medicine

Education and training

None

Sightseeing

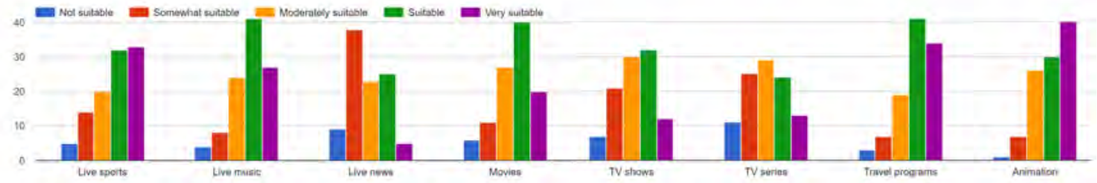
Music production

Always online (glasses-like)

Studying and working

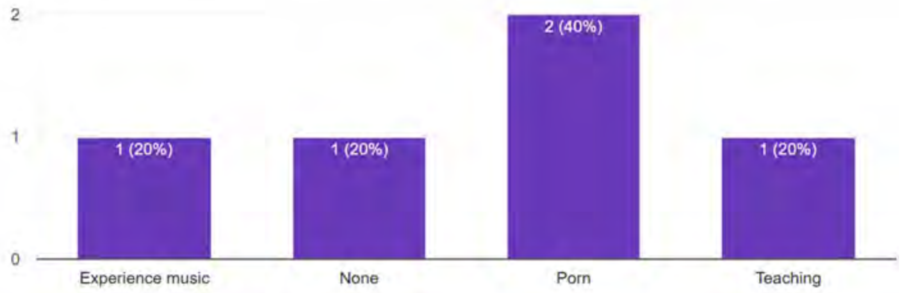
Politie onderzoek

Which types of content do you think are most suitable for VR:



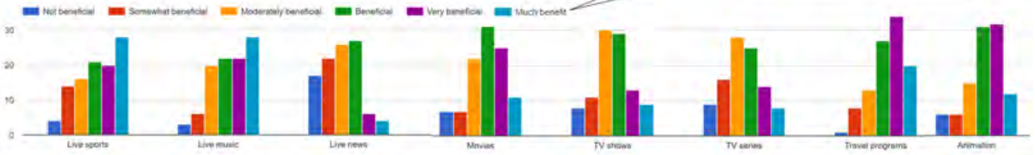
Other suitable content:

5 responses



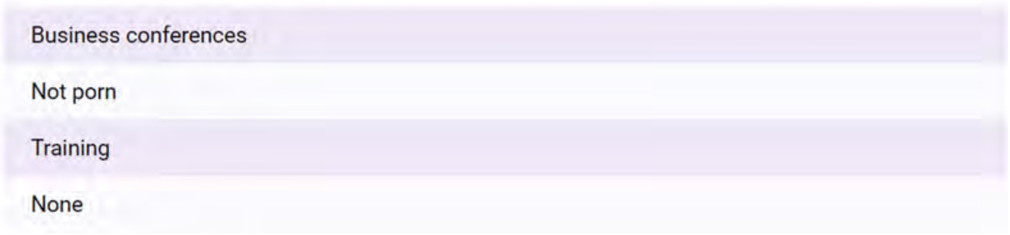
Which types of content will have the most benefit from experiencing it together, e.g. with a partner, with family, with friends?

This scale somehow ended up as six-point likert with two similar ones on top.



Other content that benefits from experiencing it together:

4 responses



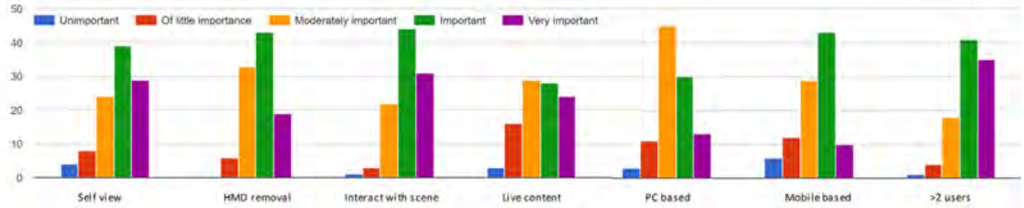
Business conferences

Not porn

Training

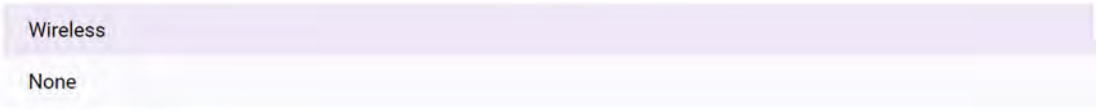
None

There are various aspects that are currently in scope of the VRTogether project. Please indicate the importance of these aspects on a shared VR experience:



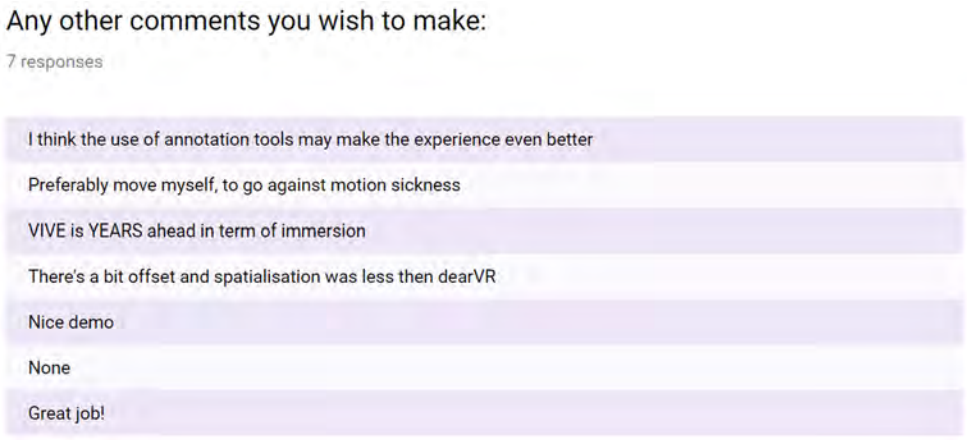
Other important aspects for the VRTogether project:

2 responses



Wireless

None



Any other comments you wish to make:

7 responses

I think the use of annotation tools may make the experience even better
Preferably move myself, to go against motion sickness
VIVE is YEARS ahead in term of immersion
There's a bit offset and spatialisation was less then dearVR
Nice demo
None
Great job!

Figure 60 - Overall results from IBC2018 pilot action

Analysis

The data collected does not have many potentially usefull correlations, the answers to the questions speak for themselves. Still, some interesting observations can be made:

- Most people expect VR to still take some time to take off. This is a good indication on the timing of the VRTogether project, as the project finishes when most people expect VR to take off.
- Even though gaming is by far the most important VR application, people foresee other applications such as movies, sports and social networking as important applications as well.
- Remarkable is, that people do not see 'news' as an important application for VR. In the original VRTogether project proposal, the second pilot use case was 'news'. This strengthens our choice to go with the police TV series (interrogation, crime scene, prison scene) as content.
- Multiple people also mentioned business meetings as potential application for our technology. This also matches our own beliefs, as also investigated in experiment TNO-2. This may be a direction to follow up on.
- As expected, live sports and live music performances seem interesting for VR, also for experiencing it together. As a nice third category, travel programs also seem of interest.
- Of the various aspects, multi-user, interaction capabilities and self view scored highest. On the other hand, no aspects were deemed unimportant or of little importance. The VRTogether project incorporates all the aspects mentioned, and it seems it rightly does so.
- We did check for correlations between the answers and the various age groups. The only significant results were between shopping (positively correlated with age, r=0.26) and movies and TV series (negatively correlated with age, r=-0.25 and r=-0.27). Even though this does not tell us much, it is good to keep in mind which age groups to target with certain kinds of services, and also good to follow VR adoption and to see if this focusses on specific age groups.

5. PILOT 2

This section will be filled in the third year of the project.

6. PILOT 3

This section will be filled in the third year of the project.

7. CONCLUSION

The first year of the project has been incredibly successful in terms of action pilots. A full first pilot is ready and has been tested with users and shown to professionals. The pilot action is still active, since the project will continue to showcase it at relevant events such as VRDays2018 and ICT2018, and has the intention of bringing it to festivals (Sundance or Venice). Moreover, the project has run a large number of experiments (12) for better developing the infrastructure supporting the pilot, understanding the performance of the system, and evaluating the experience of users (both end-users and professionals). Finally, the project has developed a number of methodologies and protocols for evaluating sVR, a novelty with huge potential impact.

For the second and third year, a similar approach will be followed. Early on the consortium will begin the discussions about pilot 2, aiming at a concept and script in early 2019. This activity has already kicked-off, with a number of phone conference and a face-to-face meeting. A parallel, supportive, action is the design and schedule of relevant experiments by the partners of the project to pave the path towards the pilot. These will include experiments with professionals, user experience evaluations, and performance of the system assessments. This activity started in the last GA meeting in Madrid in October 2018. Finally, the metrics and methods will be updated for the second pilot, paying particular attention to making automatic the data analysis and to incorporating sensor technology for more accurately gathering data from the users.

The second pilot (2019) will continue with the plot of the previous one, expanding and illustrating the crime investigation introduced in the first pilot. During the kick-off meeting in Barcelona in mid-October 2017, it was decided to create a coherent storyline that runs across the three pilots, where each one of them is representing a scene of an overarching story plot. The hypothesis is that by changing the original concept and plot line of the pilots, in the end we will offer a more attractive and engaging experience to the end-user. Moreover, this will allow the project to provide a concrete and coherent novel “product” that can be showcased in film festivals and other artistic venues. It is expected to draw the attention of the consumers, enabling them as participants in the experience show, served by the elaborated plot and interaction between players.

Starting in the second year, apart from showcasing the pilots at scientific (e.g., VRDays) and commercial (e.g., IBC) events, the project aims to bring them to film festivals. The project has created a list of the most relevant ones and is purposefully better understanding the timelines and processes. The first pilot was privately presented to a selection committee of the Sundance Film Festival (January 24 – February 3 at Park City in Utah, USA) on October 19th in order to understand if we could enter, with an initial positive response. The decision if the project is able to attend the festival now depends on the jury. Other considered alternatives include South by Southwest SxSW (March 8-17 at Austin in Texas, USA), the Tribeca Film Festival (April 24 – May 5 in NYC, USA), Cannes Film Festival (Mid May – June in Cannes, France) and the Biennale di Venezia (end of August – September in Venice, Italy).

Overall, the results in this work package 4, about the pilots of the project, are positive and encouraging, beyond the initial expectations for the first year.

8. Annexes

The following Annexes are included in this deliverable:

Annex I. Added Value Questionnaire

Annex II. CWI-1 Questionnaires and Forms

Annex III. CERTH-4 Form

Annex IV. Consent Forms in ARTANIM Experiments

Annex V. CWI-2 Questionnaires and Forms

1. (Pre-Test) Consent Form
2. (Pre-Test) Basic Information Questionnaire
3. (Pre-Test) Questionnaire about Face-to-face setup
4. (Post-Condition) Questionnaire about Skype setup
5. (Post-Condition) Questionnaire about VR setup
6. (After-Test) Interview Questionnaire

Annex VI. CWI-3 Questionnaires and Forms

1. (Pre-Test) Consent Form
2. (Pre-Test) Basic Information Questionnaire
3. (Pre-Test) Societal Anxiety Questionnaire
4. (Post-Condition) Questionnaire Real World setup
5. (Post-Condition) - Questionnaire VR setup
6. (After-Test) Semi-Structured Interview

Annex VII. TNO-1 Questionnaire

Annex VIII. TNO-2 Questionnaire

Annex IX. TNO-3 Questionnaire

Annex X. End-Users Pilot 1 Questionnaires and Forms

1. (Pre-Test) Consent Form
2. (Pre-Test) Basic Information Questionnaire
3. (Pre-Test) Societal Anxiety Questionnaire
4. Experience Questionnaire
5. Semi-Structured Interview

Annex XI. Remote Communication Pilot 1 Test Questionnaire

Annex XII. Pilot 1 with Professionals Questionnaire

<END OF DOCUMENT>