Utrecht University

TNO innovation for life

MASTER'S THESIS HUMAN COMPUTER INTERACTION

UTRECHT UNIVERSITY

DEPARTMENT OF INFORMATION AND COMPUTING SCIENCES

Towards guidelines for facilitating engaging Social VR business meetings

Author:
Simardeep Singh

Supervisors:
Dr. Christof van Nimwegen
Dr. Sylvie Dijkstra-Soudarissanane

Second Marker:
Prof. dr. ir. Judith F.M. Masthoff

September 15, 2022

Abstract

Videoconferencing technologies currently fall short in keeping employees of organizations engaged throughout business meetings. A team of researchers at TNO has therefore built a Social VR communication system including a photorealistic representation of users which can be used by employees of organizations to hold remote meetings with each other in virtual reality. The current exploratory study is the first to investigate the proposed communication system in a real-world setting for a longer period of time. Eight of the systems are deployed at four TNO offices spread throughout The Netherlands. The systems were used by six pairs who held their private meetings with each other. After each meeting, the participants completed a questionnaire assessing multiple elements. Some textual data was also gathered using open-ended questions to gain further insights into the participant's answers. A mixed-methods approach was taken to investigate whether the elements measured are related and which have a high linear relationship with engagement. Moreover, multiple meetings were held by the participants to investigate whether a first timer effect regarding any of the elements could be found when employees use the system progressively. The study provides a novel way to subjectively investigate the engagement and quality of experience (QoE) in social VR in real-world settings and long-term use. Correlation analyses showed that there are strong linear relationships between the quality of communication, embodiment, immersion, social presence, and meeting engagement. Moreover, a first timer effect was not found in the current study. However, this may be different in other studies when many employees use the system for an even longer timeframe. Low usability, quality of service or quality of interaction could then result in a first timer effect. The study concludes with a ranking of the previously mentioned elements, which lays the groundwork for developing guidelines for meeting engagement in social VR business meetings.

Keywords: Social Virtual Reality, meeting engagement, Quality of Experience, business meetings, photorealistic user representation

Contents

List of Figures

List of Tables

Chapter 1

Introduction

With a newfound sense of remote and hybrid working due to the COVID-19 pandemic, the need to battle Zoom fatigue has arisen. Zoom fatigue has been defined as "a feeling of exhaustion from participating in video conference calls" [21]. Bailenson (2021) has proposed four factors that could be causing Zoom fatigue, namely: excessive close-up eye gaze, cognitive load, increased self-evaluation and constraints on physical mobility [3]. All of these factors can be attributed to the use of videoconferencing tools. Because of the absence of body language cues, employees have a harder time in keeping themselves engaged in video meetings [38]. These videoconferencing technologies have also allowed employees to multitask more during remote meetings, which contributes to the fatigue experienced by the employees [10]. This fatigue translates to low employee engagement. A recent study suggested that camera-use and self-presentation have fatiguing effects on employees, which in turn leads to lower engagement during meetings [56]. Therefore, videoconferencing tools currently fall short in keeping employees engaged throughout meetings.

There are numerous studies on how meetings can be done remotely in other ways than videoconferencing. The use of extended reality (XR) is a big alternative which provides many possibilities that are currently being researched. XR is an umbrella term used for describing virtual, augmented and mixed reality [40]. Virtual reality (VR) is achieved when users are immersed in a virtual 3D artificial environment which is completely separate and digitally created [42]. The environment can be seen using a head mounted display (HMD). Augmented reality (AR) can be approached in 2 ways. The first approach is video-seethrough where the world is captured with a camera and virtual objects are placed onto the captured images [53]. Optical-seethrough is the second approach and is achieved by letting the users perceive the outside world through a transparent projection surface which displays AR content [53]. Mixed reality is achieved when there is a mix between AR and VR hardware or when real-world objects are used for interaction with AR or VR hardware [53]. An example of mixed reality is the AR pass-through mode on VR headsets which displays the real environment using integrated camera sensors in an HMD [42].

In the context of meeting with others digitally, social virtual reality has shown promising results for social presence [40]. In the most basic terms, social virtual reality allows people who are geographically separated from each other, to meet in a virtual space and communicate with each other as if they were communicating face-to-face (f2f) [47]. This VR space can be considered a parallel virtual reality universe created from computer graphics, which anyone from around the world can access through headsets. A term for this space, the Metaverse, was also coined years ago [42]. This term is conceived as the 3D Internet or Web 3.0. Initial small steps on implementing social virtual reality in different contexts are being taken, such as for having meetings. A case study conducted by Meetingroom and Bank of Ireland showed that virtual reality meetings improved feelings of presence, closeness and arousal for VR environments [9]. Another important insight was related to speech and discussions. It was noticed that many participants from the case study mentioned that the flow of the conversation was much more natural than in video meetings. Normally, during video meetings the participants noticed that it was required to wait a moment before saying something in response to another person. This moment of silence was not required in the VR meetings. The researchers of the case study hypothesize that employees in VR

meetings will be better engaged and focused than during video meetings [9].

Besides social VR meetings, real-time remote meetings in AR and MR are being extensively researched as well. These technologies are more often used when design tasks need to be carried out, such as in product design and architectural design [53]. Hybrid work challenges are solved using AR and MR as well. In these scenarios a task is being done by a user who is wearing AR glasses while an expert on the task joins in on the experience remotely using a VR HMD to help with the task. Some AR and MR collaboration systems have been created to support such use cases. These AR and MR based remote collaboration systems can support more natural and intuitive interaction, improve the performance time, provide more engaging user experiences and share important nonverbal cues [60]. However, in the context of business meetings, these technologies are not appropriate since most use cases for AR and MR collaboration systems involve working together on physical tasks. So, a trend has been noticed by researchers in which VR technology is more focused on designing and having meetings and that AR and MR systems are mainly focused on supporting remote expert scenarios [53].

All in all, it is becoming extremely important to research different case studies using VR, AR and MR to improve engagement during meetings. Being engaged during meetings can be defined as 'an attentive state of listening, observing, and giving feedback, leading into protagonistic action in group interaction' [22]. This is different to the definition of employee engagement, which refers to the degree to which employees are immersed in serving their organization [64]. It has long been acknowledged that the most effective and engaging way to do business meetings is f2f [17]. While it is now known that in the context of holding business meetings, videoconferencing technologies do not hold up to the same level of engagement, the next step is to research engagement in social VR meetings.

The Media Networks team at TNO, the Netherlands Organisation for applied scientific research, has long been researching social VR experiences. Even long before the COVID-19 pandemic, a group of researchers already saw potential in using social VR communication systems. Therefore, in previous research a framework was built that enables communication in VR with real-time photorealistic user representations [27]. Graphical avatar-based approaches in social VR have made a lot of progress, however some social cues are still missing [46]. In the context of engagement, the photorealistic avatars could also lead to higher levels of engagement during meetings since these social cues, or more specifically, body language cues, can more easily be recognized. However, at TNO, no research has been done yet on how Social VR business meetings can be conducted and whether they will be more engaging than video meetings or not. Therefore, the aim of the current study is to research engagement in business meetings using social VR communication systems with photorealistic avatars that have been deployed in offices throughout The Netherlands. Aspects that are highly important for engagement in social VR meetings will be identified and investigated in a longitudinal study. These aspects include the user experience and usability of the system.

1.1 Preliminary research

Knowledge on the previous social VR studies was gathered by conducting short interviews using Microsoft Teams with TNO experts who were involved in the research. The researchers were asked on their experiences researching and working on the novel social VR communication framework. They were also asked on the potential further developments of the framework and any research areas that are especially of interest. The findings are categorized under technical findings and under user experience and usability findings.

1.1.1 Technical findings

Pilot studies and other small studies on social VR with photorealistic avatars have been conducted at TNO using various head-mounted displays (HMDs). These HMDs occlude people's eyes when they put them on and enter a virtual environment. So, when one participant of a study looks at another person in VR, he cannot see the other person's eyes. However, the eye-tracking capabilities of some of the HMDs was used in previous small studies to try and mimic people's eyes onto

their HMDs. So, in the VR environment, googly eyes were put on top of the HMDs, which can be seen in Fig. 1.

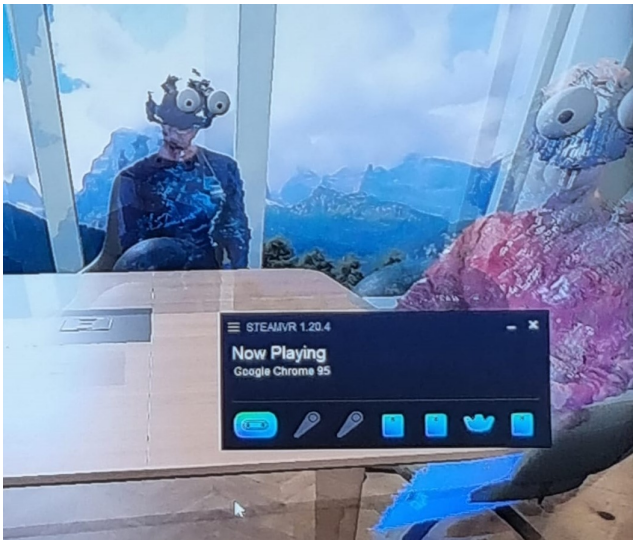


Figure 1: Solving HMD occlusion with googly eyes

Figure 2: Demo system with tactile feedback at VRDays

The pupils of the googly eyes follow the same line of direction as the person's own eye gaze. One of the researchers commented that while this did look somewhat terrifying and silly, it also contributed to feelings of co-presence of the participants. Moreover, some other small studies have been conducted on the use of tactile feedback in social VR as well. This was done using a glove with actuators that vibrate. The vibration was initiated when 2 people reached out to each other in the virtual environment to, for example, high-five or shake hands. Studies done with the

elderly highlighted that tactile feedback would contribute a lot to their feelings of co-presence.

A system which included both of these solutions was presented at the VRDays conference [1] in November 2021. Visitors of the conference could quickly try out the system with each other, after which some of their feedback on the system was collected through a survey. The survey that was used can be found in Appendix A. A set-up of one person using the system can be seen in Fig. 2. The spot marked with the number 1 indicates the glove with the actuators that has already been put on by a visitor on their own hand. The participant is looking at their own hand in VR, which can be seen in the screen. The number 5 indicates the hand of the participant in the VR environment. It is also marked with a yellow circle because that indicates that the glove has been put on and is working. When another circle, indicating another user's hand, enters the area of the circle around the participant's hand, the glove will vibrate in real life, indicating that someone else's hand is near in VR. Moreover, another participant in VR has been marked with the number 6 on the screen. The googly eyes on that participant's HMD can also be seen. The camera that captures the participant has been marked with the number 2. There are 2 cameras capturing the participant and there are also lights around the camera, indicated with the number 4. The HMD that the participant has put on has been marked with the number 3.

There are many gimmicks that could be included in a social VR system with photorealistic avatars to create a more lifelike experience but not all of them are easy to use and deploy on a mass scale. Therefore, TNO chose to focus their system on "capturing the user alone with a single RGB+depth camera and a web-based system to transmit and display users in a virtual environment" [26]. This solution allows the system to support 8 3D users as 3D point clouds, and 16 2D users in real-time on the web [27, 28]. To be able to conduct experiments on an even larger scale, TNO partnered with Connec2 [2] so that the processing of the photorealistic avatars moves from client to the cloud. Connec2 has a virtual collaboration platform in place with a stable network. The platform and stable network take away scalability issues with the social VR system. Connec2 was tasked with integrating TNO's current framework into their platform so that much more research can be done on social VR systems that are deployed at different locations in The Netherlands. Up until now, the Connec2 platform has only worked with graphical avatars. Their clients could already hold some of their meetings or presentations in VR using these graphical avatars. In the end, it is not TNO's goal to create a commercial system that is ready for the consumer market. However, this partnership was essential to be able to deploy social VR systems at TNO and continue to do research on a larger scale using these systems.

1.1.2 User experience and usability findings

From the previous studies it has become clear to the team at TNO that there was a lot of focus on the technical details of the system, while evaluating the usability of the system for end-users was pushed back. Quality of service (QoS) and quality of experience (QoE) were investigated. However, QoS assesses system parameters while QoE describes all of the different aspects that are related to the experience that a user has in VR [51]. Therefore, not much is known specifically on usability and user experience of the physical system and of the VR environment. Some small studies have been done on what the effect of the photorealistic avatars is on feelings of presence and immersion [28, 26]. However, longitudinal studies on the usability and the user experience of the system are yet to be done. One of the researchers noticed that technological implementations sometimes conflict with the needs and values of the end users. However, this is a mere assumption. Therefore, it is necessary to finally test the usability and user experience aspects of the system, especially since the issue of scalability has now been resolved.

There were some ways indicated in previous research by which the user experience of the system may be improved. These findings were collected from a small demo in which 3 users could participate in a social VR experience as point clouds [28]. One of the steps to a higher sense of immersion and presence may be to include self-view of user's hands or even full-body self-view in the VR environment. Moreover, HMD removal techniques could be investigated as well so that people are able to look each other in the eyes, which would make the conversations more similar

[1] https://vrdays.co/
[2] https://connec2.nl/

to f2f conversations. Lastly, the first timer effect and real-world use cases have to be researched as well. A longitudinal study would show whether the higher sense of immersion and presence that users feel when first entering a VR environment persists or whether it would diminish over time. In a follow-up study participants used the system for one of their own business meetings [26]. The participants highlighted that they would like to interact with the environment just like in a normal f2f meeting, by, for example, using a whiteboard.

1.2 Research goals

Based on the findings from the preliminary research, the main research question was formed. The focus of this study will be to conduct a longitudinal study in an informal, real-world setting. The longitudinal study will be conducted by allowing users to use the system developed by TNO and Connec2 for their own small-project business meetings, which can be regarded as real-world use cases. The participants will have to conduct several meetings with the system and, afterwards, deliver their feedback.

Thus, the main question is as follows:

RQ1 *What elements lead to meeting engagement in a social VR environment in which business meetings take place over a longer timeframe?*

To answer this main question, several sub-questions have been formulated based on the preliminary findings as well. These questions will be taken into account when writing the literature review.

The sub questions are as follows:

RQ2 How can meeting engagement in social VR business meetings be captured and evaluated?

RQ3 What were the conditions for meeting engagement in face-to-face, hybrid and online meetings?

RQ4 How can the quality of experience be defined for social VR business meetings?

RQ5 How can the quality of service be defined for social VR business meetings?

RQ6 How can the first timer effect be defined?

1.3 Outline

The next section of this manuscript contains a literature review on f2f and remote meetings, the lack of engagement in remote meetings, enhancing meeting engagement using social VR communication systems and lastly, the evaluation of meeting engagement in social VR business meetings. Next, the methods section details how the current study evaluates the social VR communication system that TNO has built together with Connec2 by using a questionnaire that assesses multiple elements that are of importance during VR experiences. The research design, the system design and the procedure of the study are outlined in this section as well. It is also explained how the collected data will be analysed. Afterwards, the results are visualized based on the collected data from the questionnaire. The results are then discussed in the discussion section. The limitations of the current study and directions for future work are also presented. Lastly, the conclusion ends with a ranked overview of the elements that are necessary for creating high engagement in social VR business meetings. These elements lay the groundwork for developing guidelines on engaging social VR business meetings.

The overview will also be used after this study by TNO to continue their research on the different building blocks necessary to conduct meetings in social VR. The systems will also be used after this study to collect more feedback centered around other research questions. The participants of this study and, possibly, other TNO colleagues, will also be able to continue their meetings in VR after the data collection period has ended.

Chapter 2

Literature Review

Different concepts related to engagement are first explored in the context of face-to-face (f2f), video and hybrid meetings. Several concepts revolved around social VR are then also explored. The first part of the literature review highlights what the shortcomings are to videoconferencing, why f2f meetings are more effective and why social VR can therefore be a more effective way than videoconferencing to meet remotely. Next, several social VR elements are investigated. It becomes clear how engagement should be measured in social VR meetings using the knowledge on these different elements that are of importance in social VR meetings. Quality of service and quality of experience aspects are also explored. The next section presents how other social VR systems create engaging meeting experiences. Lastly, the first timer effect is also explored by comparing first timer experiences and repeaters experiences with social VR.

2.1 Evaluating face-to-face, video and hybrid meetings

2.1.1 Shortcomings of videoconferencing and hybrid meetings

It has long been established that face-to-face (f2f) meetings have been the most effective way to communicate with others. Through f2f communication a more personal bond can be created and immediate feedback can be produced [17]. This is because colleagues in a workplace can more easily communicate their attitudes and feelings to each other in f2f meetings [52]. Denstadli, Julsrud and Hjorthol had already established back in 2012 that videoconferencing tools can be used for various purposes but they could never replace f2f meetings [17]. There is a lack of engagement and effectiveness in remote meetings because of several issues. Some of these issues are background noise, not knowing who is speaking or a lack of social presence and can most often be a problem in meetings with many meeting participants [45]. Employees of an organization mention that the lack of engagement is caused by technological constraints [33].

Technological constraints and a lack of engagement can also be found in hybrid meetings where some colleagues are physically present in a meeting room and other colleagues join the meeting remotely using videoconferencing software. Employees in hybrid meetings often feel that it is more difficult for others who are joining the meeting remotely to participate in the conversation as it is very hard for them to indicate that they want to say something [33]. On the other hand, it is difficult for employees who are joining remotely to understand what is going on during meetings. There are multiple reasons for this such as multiple speakers speaking at the same time in the physical meeting room and people in the meeting room speaking too far from microphones [45].

However, there are not only technological reasons for why engagement in video and hybrid meetings is much lower than in f2f meetings. When digging somewhat deeper, other reasons can be found as well. Being engaged or not in a meeting can be an active choice, which is sometimes signaled by a simple act such as turning the camera on or off [33]. This deliberate act of choosing to be less engaged can also have several reasons. Some people want to multi-task when they expect that a meeting is not important for them and others simply do not want to actively participate in the meeting [33]. The study by Cao et al. also shows that multitasking is a common

behaviour in video meetings and that having your camera on or off impacts the multitasking behaviour [10]. It was also reported that multitasking during meetings has become more frequent once more meetings had to be done remotely because of the COVID-19 pandemic. Yet, multitasking does not always have a negative impact on the meeting, especially if the multitasking is related to the meeting. Taking notes during a meeting or navigating the internet for relevant information related to the meeting can sometimes boost productivity levels [10]. However, people have also reported that their multitasking behaviour has led to a loss of attention, and therefore also engagement, during meetings [10].

While remote meetings have led to both positive and negative experiences, it is essential to know how remote meeting experiences can foster more productive and engaging attitudes [10]. People have been trying to repurpose video to engage in shared activities, instead of simply communicating with each other. However, videoconferencing still lacks a sense of co-presence since there is no shared visual context [8] and there are offsets between screens and cameras which makes it difficult to establish mutual eye gaze [57]. Therefore, the shortcomings of videoconferencing and hybrid meetings can also be led back to another important factor that is present in f2f meetings, the body and all it entails [39].

2.1.2 Cognitive load caused by lack of nonverbal cues

Because of our social nature as human beings we constantly rely on social cues to enhance our day-to-day interactions with others. Nonverbal cues such as body language, slight shifts in posture, the use of gestures and initiating or withdrawing from conversations are all an essential part in such interactions [39]. These cues are therefore naturally present in f2f meetings and can create psychological closeness between people in initial interactions because they are believed to communicate interest and warmth [15]. Video meetings have successfully captured some of these nonverbal cues more than in audio-only interactions. However, in video meetings some of these important nonverbal cues are still lost [15, 39, 60]. Especially postural cues, gesture cues and gaze direction is hindered in video meetings [57]. These cues and other aspects of the body can best be displayed and reciprocated in f2f communication where the entire body is visible [15].

In f2f interactions nonverbal cues make sure that communication flows naturally. People understand a lot from each other's head and eye movements because these signal turn-taking, agreement, and other cues [3]. Mutual eye gaze lets people know who the conversation is directed to and it lets the speaker know that they are being listened to. So, gaze is an important factor for conversational management [57]. Moreover, gaze lets a user control what scenery they are looking to and therefore gaze is an effective tool in a shared visual context, especially for collaborative tasks [53]. Additionally, it has been found that when deictic utterances (words such as: 'that', 'there', 'those') and gestures are well supported in a working/meeting environment, resulting communication is much more efficient, task performance increases and users rate the quality of experience higher [57].

However, the same cues can often have a different meaning in video meetings than in f2f meetings [3]. This can be caused by the technical constraints that videoconferencing software imposes [15]. So, the nonverbal communication that was produced may be perceived differently than was intended. Another reason is that some cues are lost entirely because of the limits of videoconferencing software, and therefore the meaning of the remaining cues becomes distorted [15]. Participants from the case study by Meetingroom and Bank of Ireland also mentioned that communication flows less naturally and becomes more difficult in video meetings [9]. This could be because of the loss of nonverbal cues or the distortion of the meaning of the cues.

All of these issues in the transmission of nonverbal cues in video meetings can lead to cognitive load. Cognitive load is a highly contributing factor to zoom fatigue because people in video meetings have to make much more of an effort to send and receive nonverbal signals [3]. While in f2f meetings nonverbal signals are transmitted and picked up automatically, these signals have to be produced and interpreted with quite some effort in video meetings [20]. It has been found that people, for example, tend to smile more animatedly and speak more loudly when they are doing video meetings [15]. This suggests that people tend to compensate for the lack of physical pres-

ence and the technological constraints by being more expressive in their verbal and nonverbal communication.

2.1.3 Social presence and engagement

Besides being together physically, it has been found in previous studies that shared visual context also creates a sense of co-presence [8]. However, with videoconferencing tools people are not in each other's direct physical environment, so their visual contexts are different to each other. This creates a lack of physical co-presence [15], meaning that people are less aware of each other's presence [5]. In a context in which people are geographically separated from each other but can still communicate with each other, mediated interactions take place. Mediated interactions are interactions with spaces and people that are not immediately present in the physical environment, so typically a virtual space [5]. The 'moment-by-moment awareness of the co-presence of another sentient being accompanied by a sense of engagement with the other' can be referred to as social presence [6]. In the most succinct manner, social presence can be defined as 'a sense of being with another in a mediated environment' [5]. So, co-presence is replaced with the term social presence in, for example, video meetings.

Mechanisms of social presence rely mostly on visual cues [32]. People having a meeting with each other might receive different visual cues because of their differing visual contexts. These cues can refer to the body language cues mentioned in the previous section. The visual contexts might differ a lot because of multitasking. Multitasking is a common behaviour in video meetings [10]. By constantly switching screens and not paying attention to the speaker, the meaning of the cues may become distorted. So, the distortion of cues cannot only be attributed to technological constraints. In turn, social presence is affected. With a shared visual context, visual information can be presented to everyone in the same way. Because of a shared visual context and the presence of visual information, immersive virtual environments may evoke more social presence than an email [5]. Video meetings have shown that social presence depends on both the medium that it uses (text, audio or video) and also on the interactions that are enabled and supported by this medium [32]. However, videoconferencing software is still limited in the interactions that it can enable, which is why feelings of social presence are lower for video meetings than f2f meetings.

Visual information is important for attention as well [32] and attentional engagement can be linked to social presence [5]. Attention is an important aspect to meeting engagement as it is directly related to engagement [22]. The ability to perform specific attention actions and the accessibility of corresponding functions of attention contribute to the overall sense of engagement that a person has in a meeting [32]. Therefore, social presence and attention are both important aspects to meeting engagement.

Precisely for these two concepts it has become clear that virtual reality-based systems could offer a solution. Feelings of immersion, engagement and presence are low in video meetings. VR systems, on the other hand, provide a heightened sense of immersion and can also provide meeting spaces for users to interact [38]. Immersion is often explained in terms of presence and interactivity [39]. So, immersion is also linked to social presence. Immersive virtual environments that allow people to interact with each other will allow users to moderate their social interactions in more subtle ways and enhance social fluidity by making interactions more clear and direct [39]. These social VR applications will also allow users in meeting environments to provide situational cues for attention [32]. Especially with the use of head-mounted displays (HMDs) people feel a higher sense of presence in an immersive VR meeting [58]. Feelings of social presence are enhanced by verbal and nonverbal behaviour in immersive meetings [32]. Users are able to use their hands and body in immersive environments and prefer using their own body over a keyboard and a mouse [63]. This enables nonverbal communication as well. These nonverbal cues are key features for engagement and social presence in social VR [63]. All of the lessons learned from comparing f2f and video meetings lead to creating more engaging meeting experiences using social VR.

2.2 Creating more engaging remote meeting experiences

2.2.1 Base elements for social VR experiences

It is widely acknowledged that social VR has the potential to support communication at almost the same level as f2f communication [31, 47]. It has even been suggested that social VR meeting alternatives could pose as an antidote to Zoom fatigue [40]. Research into social VR has therefore grown considerably in the past few years and there are multiple scenarios for which social VR is being used and tested. In such scenarios social presence is seen as an essential ingredient for creating realistic and engaging social VR experiences [63]. But first, a high level of social presence also needs to be achieved in the VR experiences.

Both social presence and immersion are necessary elements to perfect in social VR experiences. Where social presence can be seen as a state of mind since it regards the feeling of being connected to other social users, immersion is considered to be an experience over time [55]. Immersion is characterized by a user being enveloped by, included in, and interacting with a continuous stream of stimuli and experiences in a mediated environment [62]. In designing VR experiences it is necessary to understand what opportunities the VR systems can provide. Immersion and presence are both such opportunities that can be perceived by users [55]. In turn, together, these concepts create engaging experiences in which the line between reality and imagination is blurred [54]. However, the usability of VR systems can greatly affect engagement. Usability concerns the affective (frustration) and cognitive (effortful) aspects of using the system and therefore it contributes to an overall sense of user engagement [61]. Moreover, usability regarding VR is also defined by how comfortable users feel about immersion and presence [55]. Therefore, both of those concepts influence usability as well. So, it is safe to say that for VR experiences it is necessary to look at both social presence and immersion and determine how these affect user experiences in virtual reality and the usability of such systems.

On the other hand, embodiment plays a key part in realising high social presence in social VR experiences [36]. Embodied VR where a full-body avatar was used and movements of the users were rendered onto the avatar led to conversation patterns and interactions with other users that were almost similar to f2f communication [57]. The reason why these patterns might be similar is because embodied experiences in VR create the feeling of personally having the experience with your own body [54]. To the users it seems that their avatar's actions are their own. Users can feel that VR components are part of their own bodies and they then see themselves as part of the VR environment [4]. Immersion also plays a big role in embodiment. VR experiences that are fully immersive offer a better sense of embodiment [54]. This could be the reason why people who use HMDs feel a higher sense of presence in an immersive VR meeting [58]. They are fully immersed in the VR environment using the HMDs, whereas without the HMDs they would not be as immersed. All in all, embodiment seems to contribute positively to engagement [54], as do social presence, immersion and usability [38, 39, 55].

2.2.2 Quality of experience aspects

All of the different aspects that are related to the experience that a user has in VR can be bundled in a multidisciplinary indicator that is widely recognized as QoE, quality of experience [51]. This is not exactly the same as the concept of user experience, even though the two are strongly related. Quality of experience is similar to user experience as the latter is part of a user's quality of experience. However, quality of experience is also the perceptual quality of service (QoS) [11]. This means that all of the technical parameters can either objectively or subjectively be evaluated in a QoE metric.

The interactive and immersive nature of VR disrupts this QoE metric as VR makes a greatly different impact than, for example, normal 2D videos [51]. Immersion, embodiment, and usability, are part of the QoE and can be measured either objectively or subjectively. While immersion is objectively quantifiable, presence is a subjective experience and can only be quantified by the user experiencing it [43]. It is important to note here that measures for attention, social presence and co-presence can be used to measure engagement. Attentional engagement can therefore be

measured through subjective self-reports, but also by correlated behaviours such as eye tracking, body postures etc., [5]. Both body posture and gestures have been identified as important elements to nonverbal behaviour that can indicate whether someone is engaged [22]. These non-intrusive objective measures can then be evaluated and used to validate subjective self-reports.

Another important component for experience, specifically for social VR, is quality of interaction (QoI). Because of VR systems in the consumer market, it is nowadays becoming easier for remote users to inhabit the same virtual environment and engage in a common activity [36]. Engagement can then be felt at an individual level but also in the context of group-social interaction [61]. Therefore, the quality of interaction is important for such feelings of engagement. The QoI is defined by the ability of the user to interact with the virtual world and with other users in that virtual world [36]. On the other hand, quality of communication (QoC) is also important for engaging in activities with other users in social VR. Quality of communication can be defined by the naturalness of a conversation, the degree of involvement in the conversation, a sense of co-presence with the other and an evaluation of the conversation partner [23]. Again, both of these elements can be measured objectively and subjectively as well. Besides body position and gestures, facial expressions and voice can also be objectively evaluated to measure engagement [22]. Direct measures of how much participants speak or do not speak plays a major role in objective meeting outcomes [34]. This signals a turn-taking structure, which can be measured to evaluate whether shorter turns in speech and more overlapping speech has a higher engagement rating than meetings in which there are longer turns and not a lot of overlapping speech.

2.2.3 The influence of the quality of service on the quality of experience

While QoE can be measured both objectively and subjectively, there are also objective quality of service (QoS) metrics that can be used to indicate system performance [11]. These QoS metrics are system factors and are considered to hugely impact QoE [51]. In the past it has, for example, been argued that improved image quality leads to viewers experiencing increased presence [54]. Embodiment was also enhanced in another study by ensuring that the latency between the movement of a user's body and their virtual avatar was minimal [38]. Other studies have also investigated QoS parameters, such as buffering and bitrate, and show that they correlate greatly with QoE [51].

There are a number of QoS parameters that have an influence on the QoE of social VR systems. There are both network-related parameters and media configurations [51]. The main purpose is to identify the QoS parameters that contribute to user perceptual quality and then map them to user QoE [11]. Bitrate and framerate are, for example, QoS parameters that are often taken into consideration for QoE evaluation. In the case of VR, video-streaming is a very important technical aspect as well. Packet loss, delay, and jitter are also major factors to evaluate because these can cause video quality degradation [11].

However, it may be possible that the technical quality of VR systems may not directly influence performance, value or experiences that users have in VR [54]. It has been shown that higher bitrate does not always lead to high QoE. Moreover, framerate affects QoE depending on the temporal and spatial characteristics of the video content [11]. Jitter also has nearly the same effect on QoE as packet loss. All in all, especially for VR systems, aspects such as immersion have a huge effect on QoE as well besides QoS parameters [54]. However, no literature can be found as of yet that shows what the relationship of concepts such as immersion, presence, embodiment and usability of VR systems is to QoE and QoS.

2.3 Evaluating the use of Social VR systems

2.3.1 VR systems for business meetings

Social VR applications have become increasingly more popular over the recent years. Besides research prototypes, some social VR applications have been released commercially since 2016 as well [31]. Big players such as Facebook and Microsoft have either developed their own social VR platforms or acquired smaller VR companies with platforms that are relatively mature.

AltspaceVR was acquired by Microsoft and offers stand-alone hosting for small-scaled VR events such as meetings [39]. Spaces also rapidly developed a platform for hybrid virtual experiences to host meetings when the COVID-19 pandemic started. Another system that is very popular for remote meetings is Mozilla Hubs [39]. The openness of the platform provides wide opportunities for creating and setting up your own virtual space and the platform can be accessed using a web browser [19]. Users can move through the environment, which provides a spatial interaction between the user and the world, mimicking the experience of real-world environments as well.

What all of these platforms offer can be categorized into three different categories in which they slightly differ from each other. These categories are: the representation of the self, interaction possibilities with others, and the VR environment itself [31]. All of these aspects have an effect on social presence. There are several levels to social presence which designers designing VR systems need to focus on. These include one's own presence, others' presence, interactive presence and shared presence [63]. There is an extensive body of literature that focuses on self-embodiment and on nonverbal cues, which are both known aspects of enhancing social presence [63]. Proxemics, meaning the location and distance between users, and association, which refers to creating collaborative activities, are both also important aspects for communicating in social VR.

The appearance of avatars can also have a significant influence on communication and interaction with others and the environment itself in social VR [29]. Avatar representations can differ a lot where some applications depict a full body graphical avatar and others depict a partial body [31]. Moreover, nonverbal communication is also popular through facial expressions that user's graphical avatars can make in some applications. Gestures are popular as well. By using both of these nonverbal cues, conversations patterns that are similar to those in f2f interaction can be found in social VR as well [57]. Moreover, some applications enable users to manipulate the environment themselves. The layout of the VR environment can sometimes also be very open so that users are free to traverse to different spaces or rooms. All of these applications make it easy for the masses to engage in social VR activities [31].

While some platforms can be used by the masses now, there are still several technical and usability or user experience issues that lowers users intent on using the system regularly. An efficient communication platform is essential to connect users in an immersive XR environment [38]. The use of Mozilla Hubs has shown quite some technical problems related to audio, user capacity, lag and performance issues [19]. On the other hand low resolution with documents, confusing chat messages and identifying others were some of the usability and user experience issues that were found.

2.3.2 Avatar appearance and embodiment

Typically in VR environments, avatars are quite limited, non-lifelike versions of users because these are graphical and sometimes abstract representations of them [39]. However, people have reported that they would feel more involved or present, in other words engaged, if avatars looked more realistic [63]. Avatar realism can be divided into both appearance and behavioural realism [49]. Embodiment is already contributing to the latter element of behavioural realism, which is why avatars are now perceived differently. Users feel a stronger sense of ownership over their own avatar's body in embodied VR [35]. Because of a stronger sense of ownership, immersion is also increased [63].

On the other hand, there is also research being done on avatar appearance. There is research dedicated on how to develop photorealistic representations of users for social VR meetings [39]. Facebook is for example using high-resolution image gathering and machine learning to develop photorealistic computer-generated avatars. The VRTogether project also developed photorealistic representations of users where each user's 2D real image was captured and segmented with a Kinect sensor [16]. Scanned avatars also seem to be more human-like and therefore create an even stronger sense of virtual body ownership [35]. The appearance of others' virtual bodies also seems to have an effect on the users self-perception towards their own body. Moreover, avatars that represent user's full body contribute to a significantly higher social presence [63]. These

insights show how important avatar appearance can be in creating more engaging social VR systems.

Avatar appearance is also important for the transmission of nonverbal communication, which can be marked by gestures, head nods, eye gaze and direct body orientation [15]. Some of these cues could be transmitted by simply using photorealistic avatars. However, an essential aspect to effective communication is still missing with photorealistic avatars, which is eye gaze. Gaze can signal a desire to communicate while gaze aversion can reveal a disinterest [15]. Yet, since the eyes are still occluded by the HMDs when using photorealistic avatars this communication is made more difficult. Thus, it could also be hard for users to transition from feeling others' presence to interactive, or social presence [63]. It is unclear whether this has a big impact on the use of photorealistic avatars for engaging in social VR experiences. Even though such important cues like facial expressions and eye gaze are still missing [49], more and more research is being done in the field of VR with a focus on eye contact, the conveyance of facial expressions and the implementation of mutual hand gestures to name a few [53].

2.3.3 Interactive capabilities of VR environments

Besides interaction with others, interaction with the VR environment and objects in the VR environment is important as well to make the environment feel more real [63]. Interaction with objects in the VR world partly determines the level of immersion and presence that users feel [19]. The use of different technologies also has a great effect on the interaction possibilities in VR. It is, for example, preferred by people to use body and hand gestures instead of a keyboard and mouse as this feels more natural [63]. Users rate the quality of experiences higher when communication is more efficient with the use of gestures [57]. Even interacting with a system using eye gaze is sometimes preferred compared to traditional methods since it is much more natural and intuitive [60].

There are quite some interaction possibilities that are already available to people who are working with VR technologies in the construction sector [44]. Examples include the addition, resizing, erasure and manipulation of objects using VR controllers so that designs for new buildings can be made. It is also being explored how VR systems can make an impact on educational environments. In this context it is believed that interaction would make people feel more engaged when trying to learn something new [13]. In learning environments this interactivity can for example be achieved by creating a VR game in which embodied interactions are also present in the form of gestures. On the other hand, a note-taking feature in VR systems is seen to be an important element in learning as well [12].

Note-taking is an important element in business meetings as well. It is a way for people to engage in the meeting as it makes the meeting experience more positive [10]. Chen et al., (2019) were able to implement this feature in VR using a digital tablet and a stylus [12]. The stylus' position and orientation were tracked and rendered into a virtual stylus in VR. Other features that are focused on for more engaging VR meeting experiences are more relevant for design meetings. An interesting insight from research on VR systems for design meetings is that VR alone is not necessarily better than traditional media such as drawing [37]. VR needs to be supported by tools that help to create visualizations. An example of such a feature is a multitouch table that facilitates interactive design and immediate feedback [50]. For normal business meetings, it is not clear yet whether these features for design meetings would be relevant as well or whether other types of features need to be created to support a better VR experience for business meetings.

2.3.4 First timers and repeaters in social VR

While there are numerous ways to enhancing engagement in VR meeting experiences, various insights have indicated that a first timer effect or habituation effect might occur in VR experiences. A first timer effect can be defined as an experience in VR for the first time while only spending a short amount of time in the VR environment [25]. On the other hand, a habituation effect is the counterpart of the first timer effect. A habituation effect occurs when someone has immersed himself in the same VR environment for at least 2 times [14]. Both effects have similar outcomes. The first timer effect can create a 'wow' effect and lead to a lot of excitement when users first use

a VR system [2]. However, during a second interaction the excitement already somewhat drops. The same can also be said for other VR-induced symptoms and effects such as motion sickness. These symptom levels are highest on first immersion and reduced to negligible by the third immersion [14]. This then implies that there is a habituation effect.

Exploration also plays an interesting part in first interactions with VR systems and environments [2]. The novelty of a system or environment encourages full exploration of the environment, which leads to users making many head movements to take everything in [14]. This is also reduced in subsequent VR experiences in the same environment. Usability research has already established a while back that user requirements for first-time, casual, and expert users may all be different for various systems and that the requirements of these users may change as they move from one level to another [24].

Therefore, it is important to be cautious of the first timer effect for longitudinal studies with social VR systems. This is why Cobb et al. (1999) recommend that first experiences of VR systems and VR environments should be used as an experience for the users to familiarize themselves with the system and environment for a short amount of time [14]. Ideally, this familiarization process should take less than 10 minutes. If possible, most of the interaction possibilities of the environment should then be experienced by the user. This would also take away any barriers for users who do not know what to expect and can become intimidated by all of the different functionalities that they do not know about when using the system for the first time [41]. As a result, all of this previous literature indicates that the effects on engagement of the first timer effect still need to be investigated.

2.4 Aim of the current study

The literature review has highlighted a gap in the current literature on the relationships between different elements regarding QoE and engagement. There is no universal method yet for evaluating engagement in social VR meetings either. Moreover, the first timer effect also needs to be researched in relation to engagement. The preliminary findings in the previous section also supported these gaps in knowledge. Thus, the aim of the current study is clear.

The following sub-questions have been formulated to guide the study further; questions highlighted in blue are answered by collecting and analysing data from participants using a questionnaire. The questions highlighted in red are answered by analysing both the literature study and the collected data.

RQ7 How can the relationship between quality of service (QoS) and quality of experience (QoE) be defined for social VR business meetings?

RQ8 How can the relationship between quality of experience and meeting engagement be defined for social VR business meetings?

RQ9 How can the relationship between embodiment (using the photorealistic user representation) and meeting engagement be defined for social VR business meetings?

RQ10 How can the relationship between the quality of interaction (QoI) and the quality of communication (QoC), and meeting engagement be defined for social VR business meetings?

RQ11 How can the relationship for social presence and immersion be defined for social VR business meetings?

RQ12 Do feelings of engagement diminish over a longer period of using social VR for business meetings when compared to using it for the first time?

RQ13 How usable is the social VR system to use for a longer period of time compared to other VR systems?

RQ14 What do users experience when working with a social VR system for a longer period of time compared to other VR systems?

Chapter 3

Methods

This chapter outlines the details of how the research was conducted and what decisions have been made regarding the system design and the procedure of the study. The research has been focused on a subset of meeting types and the experiment was conducted under the supervision of TNO at several TNO locations in The Netherlands. The measures for analysing the outcomes are also explained. Lastly, it is also explained how the data was prepared for the data analysis and how the data was analysed.

3.1 Study design

This exploratory longitudinal study takes a mixed-methods approach with both a between-subjects and within-subjects design. The aim is to understand whether people reach an engaged state in VR using photorealistic avatars and how they reach that engaged state based on different elements regarding QoE and engagement. Moreover, the aim is also to evaluate whether people remain in an engaged state for multiple meetings or whether that engaged feeling diminishes after their first meeting in VR.

Participants in the study had to use the proposed social VR system by TNO for their own small-project business meetings or informal business meetings. The number of meetings an employee of TNO has on a daily basis, varies immensely because of their role within the company. An employee's role also dictates whether a participant of the study can be present at the locations where the VR systems have been placed. Additionally, as a result of COVID-19, a hybrid way of working at all TNO locations in The Netherlands was adopted. Because of all these factors, it was decided that the participants would be given a timeframe of eight weeks to conduct their meetings with the system. The first meeting was regarded as a trial meeting in which participants were guided on how to enter the *TNO Metaverse* using the social VR systems.

The mixed-methods approach is used for this study to systematically understand the factors contributing to feelings of engagement and to understand the consequences of having remote meetings with the social VR system. Both the quantitative and qualitative data was gathered from the meetings done with the VR system and from surveys that participants had to fill in after each meeting they had done with the system. The within-subjects design was used to investigate whether there were any changes in the participants' attitudes when they had more meetings in VR, compared to when they had their first meeting with the social VR system. Moreover, their experiences were also compared to each other so that some insights could be generalized. The data that was compared for this purpose stems from the survey that was filled in after every meeting in VR.

3.1.1 Quantitative research design

The conceptual model in Fig. 3 has been made to portray the relationship between the different variables of this research used for the quantitative analysis. These relationships are based on the findings from the literature study. Perceived meeting engagement, quality of experience and quality of service are composite variables that are made up of indicators. These indicators were

measured using survey questions or using system data. The survey questions that were used for an indicator are noted in the figure as well next to the indicator. Perceived meeting engagement is made up of the indicators immersion and social presence. These are important aspects to enhancing engagement in social VR environments. Quality of service is made up of the indicators subjective system performance and objective system performance. Moreover, the quality of experience variable is made up of the indicators usability, embodiment, quality of interaction and quality of communication. All of these elements have been defined in the literature study.

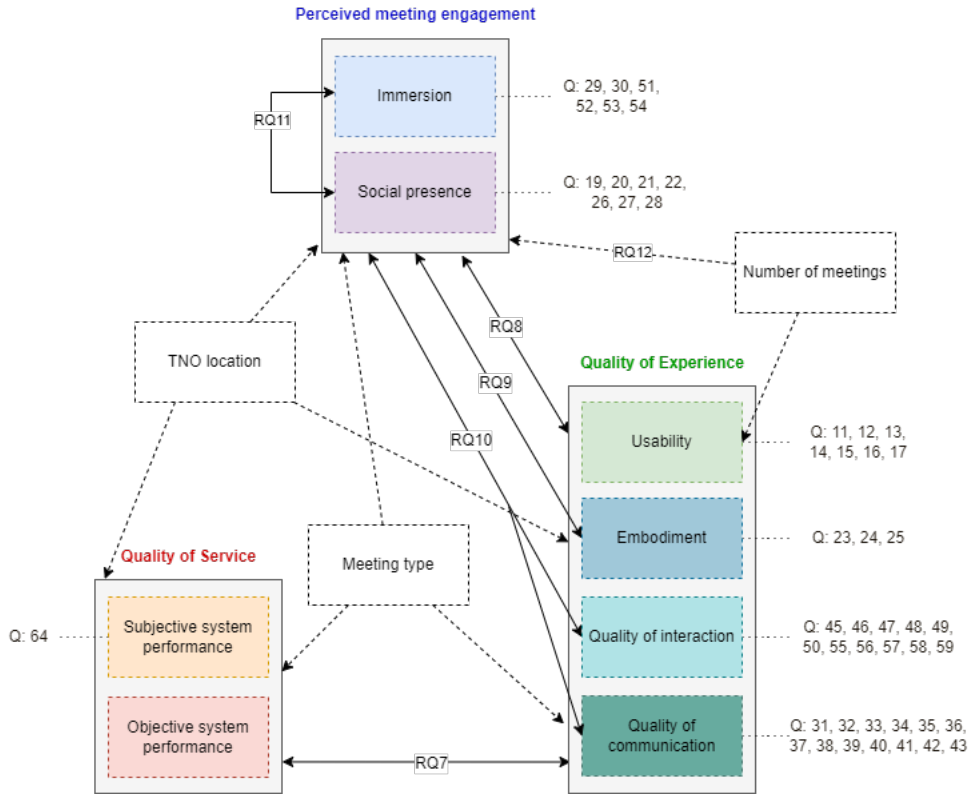


Figure 3: Relationship between the variables of the quantitative research design

Next to the composite variables, there are also three situational variables: *meeting type*, *number of meetings* and *TNO location*. These situational variables may have an effect on the study since for each meeting, the meeting type was different, the physical office locations from which colleagues met each other in VR varied, and the number of meetings held per participant also varied. The *meeting type* variable is necessary to include because there are different types of meetings that employees usually conduct f2f or using videoconferencing software. The type of meeting can be defined by the purpose of the meeting [1]. The purpose of having brainstorm sessions, for example, is to brainstorm for ideas or solutions. The purpose of project meetings is to discuss ongoing projects. The content of the communication is also an important factor in determining the meeting type [17]. This has already been an important factor for employees of organizations to decide whether a meeting would be held f2f or with videoconferencing software.

With new technologies being added to the list, namely social VR, it is important to consider what types of meetings these technologies could be beneficial for and what types of meetings they are not beneficial for. The communication content of different types of meetings is therefore considered as well in the current study. Previous studies have shown that meetings with a purpose of information exchange, management and training are held using videoconferencing software. F2f meetings are usually held for meetings where the purpose is to negotiate, demonstrate, or have discussions [17]. As the current social VR system is more focused on simulating f2f meetings, it is investigated whether there are any differences or similarities with meetings that are usually done f2f. Therefore, it has been chosen to evaluate meetings in which people have to negotiate, demonstrate or have discussions. Thus, the meeting types that have been investigated in this research

are:

- Brainstorm meetings
- Ongoing project meetings
- Presentations

The conditions of the physical rooms at the office locations, in which the systems have been deployed, are slightly different from each other, especially regarding the network connectivity for each system. So, this may have an effect on various elements. Therefore, the *TNO location* and *meeting type* variables are investigated in relation to the composite variables; *perceived meeting engagement, quality of experience* and *quality of service*. The results can then show whether these variables had an impact on the different variables of the research. It was not chosen to investigate the effect of *TNO location* and *meeting type* on the indicators of the composite variables because it needs to be established first whether there was any effect at all from these situational variables.

Moreover, the *number of meetings* variable was used to compare the results of people's first experiences with the social VR system to their later experiences with the system. For this variable it was chosen to investigate what the outcome of *perceived meeting engagement* is after each meeting in VR per participant. Moreover, it was also chosen to investigate whether the participants' attitude towards the usability of the system changed after each meeting.

3.1.2 Qualitative research design

In addition to the quantitative research design, it was planned to also conduct a qualitative analysis in the form of content analysis. This enables the evaluation of engagement in nonverbal behaviour in an unobtrusive way. Behavioural data was gathered from body tracking videos made with a stereo camera, which was also used for capturing the users and streaming their point cloud to the VR environment. Body language is an important element in f2f meetings, and should therefore be included in the assessment of VR meetings [48]. The body tracking module of the camera only focuses on detecting people's bones, tracking it and representing a bone using two endpoints, which are connected using a line. An example of a body tracking image can be seen in Fig. 4.

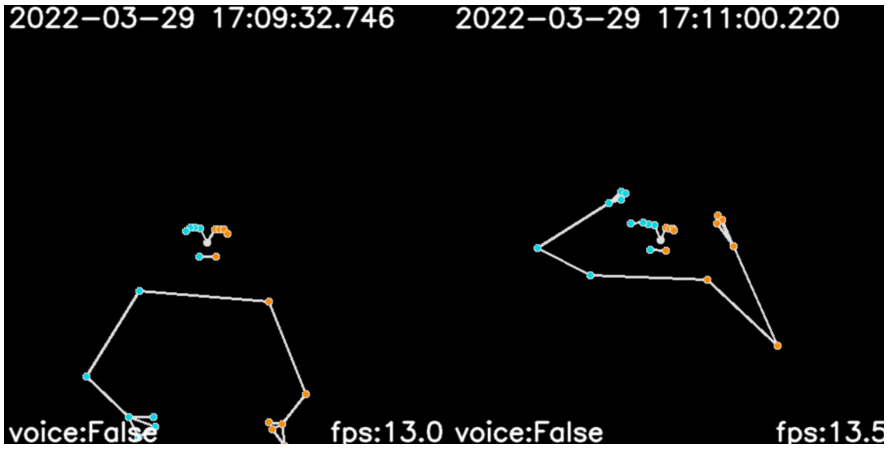


Figure 4: Images of a body tracking video in which a person presumably puts on an HMD

Moreover, textual data was also gathered from the surveys. The textual data was coded using open coding, which means that the codes for the coding analysis have been identified from the data itself. The behavioural data was deductively coded using a codebook based on the Engagement Framework [22]. With this behavioural data the aim is to detect patterns or repeated behaviour. Certain characteristics in these videos have been used to code the participant's behaviour in the videos in different engagement states. The following diagram, Fig. 5, shows how the behavioural data is coded. This diagram has been made using the information on the Engagement Framework.

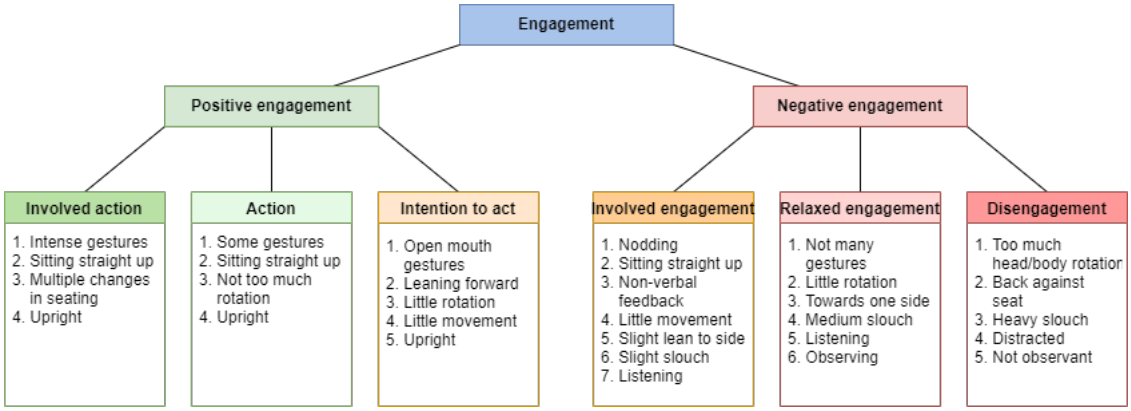


Figure 5: Codebook for coding body tracking data

3.2 Study population

Since it is TNO's goal to work with the social VR system internally at first, the participants for the study have been recruited from TNO and therefore consist of only TNO employees. This group of participants is representative of the end-user group since TNO employees have a lot of meetings with colleagues who are situated at other office locations in The Netherlands. So, there are many meetings held by people who are geographically distanced from each other.

Purposive sampling was used to select the participants since they needed to fulfill the criteria listed below. Pairs or teams of three were preferred over individuals because it was assumed that the pairs or teams would already have meetings with each other. Thus, it would be less likely that they would drop out of the study. Moreover, the system was not yet able to support more than three users at the same time.

The following locations from which employees took part in the research have been chosen beforehand because there is already much cooperation by employees from these locations. So, there are already some employees from these locations who are working together on projects. Employees from the following TNO locations have taken part in the research:

- TNO New Babylon, situated in The Hague
- TNO Oude Waalsdorperweg, situated in The Hague
- TNO Soesterberg
- TNO Groningen

3.2.1 Inclusion criteria

An employee was able to take part in the research if he/she/they:

- Was willing to travel to one of the locations where a system has been placed
- Could name one or two colleagues to do their meetings with
- Was prepared to do the meetings at least once per week
- Was prepared to do at least four meetings in total with the social VR system
- Knew that they could only hold brainstorm meetings, ongoing project meetings or presentations with the social VR system for a maximum of one hour

3.2.2 Exclusion criteria

An employee could not take part in the research if he/she/they:

- Needed to show external objects in the meetings
- Needed to take many notes during the meetings
- Needed to have a meeting about classified work, for example, for the Ministry of Defense
- Needed to have a meeting that is longer than an hour
- Could not fill in the surveys

3.2.3 Sample size

There were eight social VR systems deployed over the four previously mentioned office locations. So, there were two VR systems per location. The sample size of the study was limited to a maximum of 40 participants to ensure that there were enough systems free for the employees to plan their meetings flexibly with each other. Moreover, since this was a pilot study, the minimum required sample size was set at 15 participants. With a maximum of 40 participants, ten employees per location could make use of the two systems. Moreover, the minimum number of participants was set at 15 so that at least five teams of two to three employees could use the system.

3.2.4 Dropout and replacement procedures

If a participant wished to drop out of the research, it was sadly not possible with this research to only replace that employee. A team of three, or pair, then needed to be replaced. If there were enough other participants, so more than the minimum number of required participants, the team would not have to be replaced. However, if a team dropped out and the number of participants was lower than the required number of participants, it was possible to still contact another team wishing to use the system if the previous team dropped out within one month of the study. However, a minimum number of meetings done in VR also dictated whether a team/pair needed to be replaced or not. If the team that dropped out had done at least three meetings in VR, filled in their surveys, and agreed that the results up until the dropout could be used for the research, the data was still used.

So, another team/pair was contacted when the minimum number of 15 participants was not reached. However, another team/pair was only able to use the system and hold the meetings if it was sure that they were able to do at least two or three meetings with the system before the end of the study. At the end of the study, the testing period was extended from eight to twelve weeks so that more participants could be recruited for the study, as many participants had dropped out.

3.2.5 Recruitment period

To spark the interest of TNO colleagues to participate in the study, a flyer on the *TNO Metaverse* was created in which a description of the system and the purpose of the study were outlined. The inclusion and exclusion criteria were also mentioned in this flyer. This flyer was then sent out in an email to everyone at the Unit ICT of TNO and to every colleague from each of the locations where a system was placed. The created flyer and accompanying email can be found in Appendix B and Appendix C.

The initial emails to recruit participants for the study were sent out in the first week of April. Over a period of four weeks there were responses from 46 people in total from different teams and individuals. However, even though the requirements were listed in the email and in the flyer, most people only replied to do one meeting in the *TNO Metaverse*. Moreover, sometimes teams of four or eight employees wanted to try the system out. However, it was not possible with the system to host more than three people simultaneously. So, at first all of these people had to be rejected from taking part in the study. Yet, over the course of the next four weeks many people dropped out of the study, even after filling in the consent form, because they did not find the time to meet in VR or because there were sometimes difficulties with the VR system which made them

cancel their meetings. So, in total, twelve participants were left who took about twelve weeks to have either one, two or four meetings in VR. These twelve participants were six pairs.

After a period of eight weeks, it was decided that the search for new participants would be stopped because it was clear that TNO colleagues could not hold a VR meeting every week due to their personal high work load and holidays. In addition, continuing the research for more than twelve weeks was not feasible due to the time limit of this research. It was also not feasible anymore to let participants have one meeting with the system. Most TNO employees who were contacted were extremely busy in the months of May and June and could therefore not plan a meeting quickly. There were some employees from the higher management at TNO who tested out the system before. However, these meetings were not comparable to the meetings that the participants of the study had since the systems were already started up for the higher management and they only spent about 5 minutes each in a VR meeting.

3.2.6 Participants

In Fig. 6, it has been visualized from which TNO locations the colleagues took part in the study. One connection represents the connection between two participants from those locations. So, for example, there was one couple who met with each other in VR while one of the participants was situated in Groningen and the other was situated in Soesterberg. This connection can be seen in the figure below. The number of connections between each of the locations has also been added in the figure.

Figure 6: Map of The Netherlands showing the different participant connections

As previously mentioned, there were twelve participants who took part in the study. There were four participants from New Babylon, five from Soesterberg, two from Groningen and one from Oude Waalsdorperweg. The individuals from the pairs already knew each other to some extent. They were either currently working on projects together or had previously worked on projects together and wanted to catch up using the social VR system. Moreover, eleven participants were male while only one participant was female. The age range varied much as there were participants from all age categories. Four participants were between 20 and 29 years old, two participants were between 30 and 39 years old and three participants were between 40 and 49 years old. Only one participant was in between 50 and 59 years old. Lastly, there were two participants who were 60 or older.

3.3 System design

During the deployment phase, it was necessary to make sure that all of the setups at the different TNO locations were correctly placed and that the conditions at all locations were kept consistent. It was ensured that all of the employees accessed the system in the same way, even though they worked at different locations. To make sure that this happened, a deployment plan was created, which is detailed below. However, due to some restrictions and regulations, not all of the requirements outlined in the deployment plan were met. The details of this deployment process have also been outlined below.

3.3.1 Materials & Setup for deployment

The physical social VR system placed in a room is pictured in Fig. 7. One system consists of the following equipment: an HMD with its respective controllers shown by the numbers 1 and 2, a laptop or PC (with a monitor, keyboard and mouse) represented by the numbers 3, 4 and 5 and, lastly, a camera that captures the user represented in the figure with the number 6. Number 7 is a cleaning kit meant to clean the VR headset and the controllers for hygienic purposes.

Figure 7: TNO's Social VR system at New Babylon

A platform, built by Connec2, was used to interact with the VR environment, which was also built by Connec2, and to transmit video and audio streams. To show how the TNO network and the Connec2 platform communicate, TNO colleagues built an overview which was similar to Fig.

8. The figure shows an updated version of the overview.

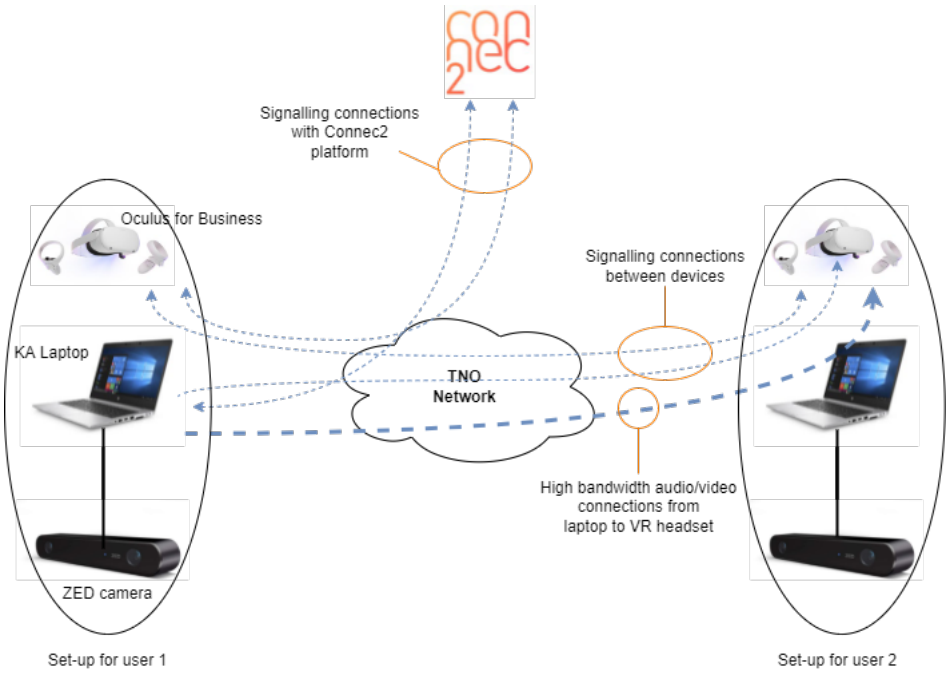


Figure 8: TNO's social VR setup

The users of the system were captured with the 3D sensor from the ZED camera. The capture module that is used for the system was developed by TNO and is described in [18]. It allows RGBD image processing for 3D user representations, as well as processing and storing for body tracking purposes. The 3D point clouds of users could then be seen in the VR environment. This environment was rendered in the HMDs of the users in which they could then also see other users' 3D point clouds. The system also allowed for movement in a 3D space with reasonable visual and auditorial spatial awareness. As there was only one RGBD camera in the setup, users were encouraged to keep in a confined space to avoid distortions in the 3D point cloud, which would be a normal situation for a business meeting.

The list below shows what the system consisted of and what the requirements were for the setup of the system so that it would work well in the way it was intended. However, not all requirements could be met. For each requirement, it is explained whether the requirement was met or not in the following list:

1. *Every VR system consisted of a working PC/laptop on which Connec2's platform, Uplink, was downloaded.* Uplink was needed to connect the camera to the VR environment and transmit the images for the point cloud to another user's HMD.

    - *If PCs were used, monitors, keyboards and mouses needed to be gathered.* In total there were 3 PCs used for which monitors, keyboards and mouses needed to be arranged.
    - *The PCs/laptops needed to have an NVIDIA graphics card.* The NVIDIA card was needed because CUDA was required for the cameras. CUDA needed to be installed on all of the PCs/laptops.
    - *The OBS virtual camera also needed to be installed on all of the PCs/laptops.* This was also necessary for the camera.
    - *The ZED SDK needed to be installed on all of the PCs/laptops as well.* The ZED SDK was necessary since ZED cameras were used.
    - *All of the PCs/laptops also needed to be able to connect to WiFi.* It was not possible to connect the PCs/laptops to the network that was being used through an Ethernet cable.

- *The TNOCaptureStarter, which is the capture module for the point clouds, also needed to be installed on all of the PCs/laptops.* The capture module made sure the point cloud was captured correctly by the camera. This module was built internally by TNO colleagues. Connec2 made sure that their platform, Uplink, used the module.

2. *A stereoscopic ZED 2i camera was included in the setup to capture the user.* This camera was placed in such a way that the user could only be captured from the front. The camera was clamped onto a desk in the room. Each setup had its own desk on which the equipment was placed.

    - *The camera stayed put in one place throughout the whole study.* The camera was always facing the user's chair. The height was also set after first placing the camera. So, the height and place of the camera would not have to be adjusted again.
    - *The camera needed to stay plugged into the PC/laptop.* By letting it stay plugged in, it would be ensured that the camera was plugged into the correct USB port on the PC/laptop.

3. *The HMD used for the system was Oculus Quest 2, which came accompanied with VR controllers.* These HMDs and controllers were from Oculus for Business. Several features were included in the HMDs such as hand-tracking and a pass-through mode.

    - *Software from Connec2 needed to be installed on the HMDs before deploying the systems.* Connec2 installed the software once at their office. After deploying the systems they were updated automatically over the cloud.
    - *There needed to be enough power outlets for all of the equipment, especially for plugging in the HMD as well.* The HMD could stay unplugged for the duration of a meeting so it would be easier to wear. However, it needed to be plugged back into a charger after each meeting. This ensured that the HMD was always charged fully for anyone else who was participating in the pilot.

4. *For each VR system there was at a location, there should have been a dedicated physical XR room in which the VR setup was positioned.* The locations New Babylon and Groningen were able to facilitate 2 rooms for each of the VR systems placed at those locations. So, there were dedicated rooms for each system. However, Soesterberg and Oude Waalsdorperweg could not facilitate this requirement. Therefore, there was only 1 room in which both VR systems were placed at these locations. This means that both systems could not simultaneously be used by participants who each had a different meeting.

    - *The VR systems needed to be easily findable.* To make sure that participants of the pilot knew where the systems were placed, logos were made and printed out to put them on the doors of the rooms. So, besides the room numbers, the logos made the rooms more visible and findable as well.
    - *The purpose of the VR systems and the study needed to be clear to every TNO colleague.* Besides logos, the flyer for recruiting participants was also printed out and taped to the doors of the rooms. Moreover, warnings were also placed on the desk that alerted other colleagues, who were not participating in the study, that they should not move the equipment or use it. This made the purpose of the systems and the study clear to every colleague.
    - *Per VR system there needed to be a desk in the rooms and a non-rotatable chair.* The equipment was then placed on the desk and the user sat behind the desk on the chair. The chair was non-rotatable to prevent people from rotating in the VR environment using their body because that would disrupt the image of their point cloud. The camera only captured the users from the front.
    - *The rooms in which the VR systems were placed also needed to have a stable connection to the Internet.* Unfortunately, an Ethernet cable could not be placed since no cables could be routed to the Eduroam network, which is a public network used for the systems. The internal TNO network could not be used since WiFi accounts would only be opened up for employees and not for systems. This was a TNO-wide policy. In the rooms where Eduroam was not even detectable by the system as a WiFi connection, wireless access points were placed to strengthen the connection. The deployment showed that this was only necessary for one room at New Babylon.

- *The available bandwidth in the rooms needed to be stable for both uplink and downlink video streams.* The bitrate could be as high as 15 Mbits/s up from the PC/laptop and 15 Mbits/s down to the HMDs.
- *The rooms should have been lockable.* It was intended that only the employees who were participating in the pilot had access to the room. However, this requirement could not be met by all of the locations since some of the rooms could not be locked at all. Because of this, no room was locked anymore so that the conditions of using the system were kept the same at each location.
- *Besides each VR setup, a guide with instructions needed to be placed on the desk so that users could figure out on their own how to use the system.* An employee could always look at it when something did not work or so that they could read through it to understand how they should use the system.
- *Moreover, a cleaning kit also needed to be present in the room to clean the VR headset and the controllers.* This cleaning kit was essential to ensure good hygiene for all of the participants of the pilot.

5. *Participants needed to log into Connec2 Uplink using system-unique login credentials.* The login credentials were formed with emails from TNO that were created specifically for each VR system. In the Uplink platform a general password for all of the accounts was formed as well to create the Connec2 accounts.

    - *The login credentials needed to be created for each individual VR system at each location.* This means that there were 8 Connec2 accounts which were used by multiple participants. So, the login credentials were not linked to the participants and the participants did not have to provide any personal information for these accounts.
    - *TNO also needed to have their own workspace ID on Connec2 Uplink to ensure that privacy regulations were met.* This means that the workspace in Uplink was only used by TNO colleagues since only they knew what the workspace ID was and which accounts were linked to it. So, no third party had access to the VR environment.

For many requirements mentioned above, there was much contact with the location managers from the different TNO locations. By discussing different needs from both parties, many of the requirements became clearer. These discussions also led to realizations about which requirements could not be met so those had to be adjusted. Furthermore, the location managers made sure that the meeting rooms were reserved for the pilot and they also arranged some of the furniture in the rooms such as the desks and chairs. They also made sure that the cleaning kits were available in the rooms.

3.3.2 Requirements for a meeting in a VR environment

Fig. 9 shows a summary of some use-cases created to ensure that aspects for a successful experience of having a business meeting in the VR environment were considered. These use cases only concern the experience that the users have in VR. This is because the usability and user experience of the system were not fully researched in the current context as the systems were already fully set up for the users before they interacted with them. All the necessary software had already been downloaded. So, these aspects were not taken into account when researching the usability and user experience of the current system. How the systems have been set up is already fully explained in the previous section. The only details that are missing in the figure are regarding the turning on and off of the equipment. This has been left out in the figure because the figure highlights the interaction with the platform itself and between two users in the VR environment.

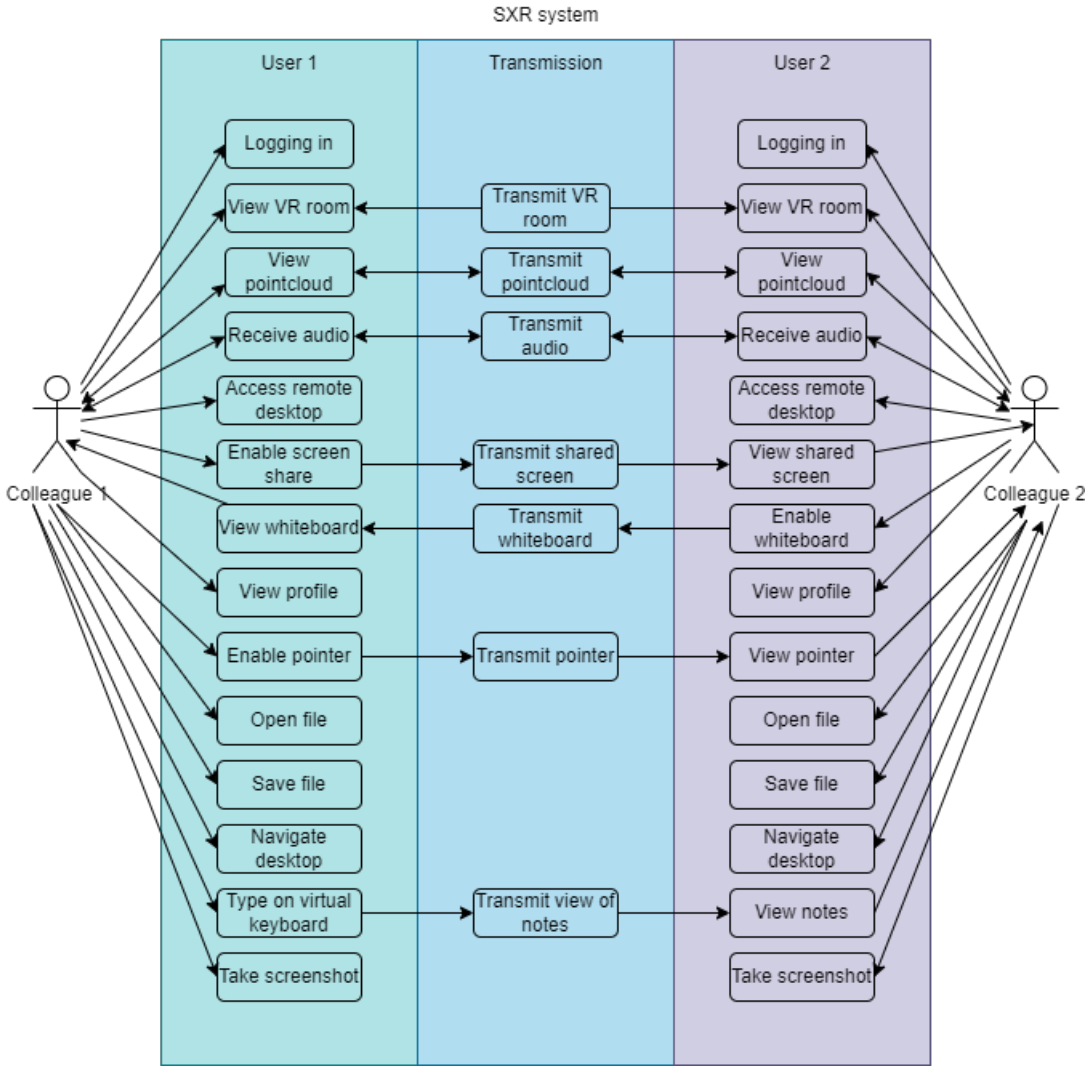


Figure 9: The use case diagram

Moreover, Fig. 10 shows how the participants should have used the system and what functionalities the system should have had for them to interact with each other and with the VR environment. Based on these models and the features that were implemented in the system, a guide was created to assist participants with using the system. Based on the figures, a list of requirements for the system had also been created and the requirements were categorized using the MoSCoW-method. Table 1 presents these requirements. The table also shows which features could not be added to the system due to a lack of time. These features are coloured red.

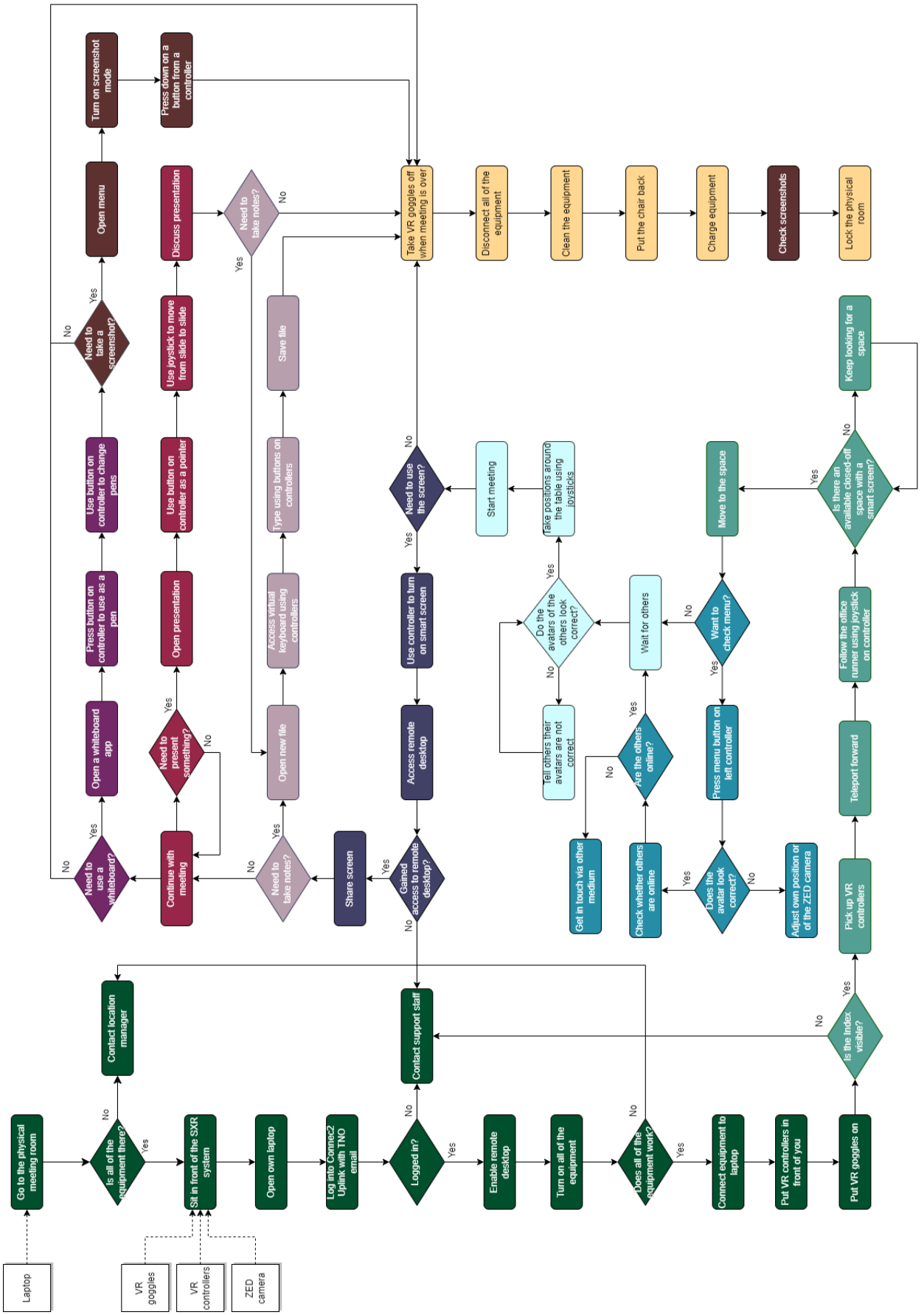


Figure 10: The user flowchart

Table 1: System requirements for a VR meeting

Must have	Seeing colleagues in the VR environment Hearing colleagues in the VR environment Viewing avatars as point clouds Accessing own personal device in VR (remote desktop)
Should have	Screen-sharing option on a smart screen in the VR environment Closed-off meeting rooms in the VR environment Profile menu with self-view of point cloud Visibility of colleagues online/offline status Menu with buttons for all functionalities (buttons) Note-taking on remote desktop (on own personal device) Being able to save files (through remote desktop) Taking screenshots on the Quests
Could have	Opening up a whiteboard in the VR environment Using the controller as a pointer Viewing the use of a pointer by other colleagues Using the controller to go to the next slide in a presentation Note-taking using a physical keyboard
Would like to have	Multiple pens, highlighters and an eraser to use for a whiteboard Everyone being able to write on the whiteboard simultaneously

Most of the requirements were taken into account for the research setup that was used. The main goal was that users were able to communicate with each other and see each others point clouds. Otherwise, a meeting could not take place at all. An example of a point cloud in the VR environment can be seen in Fig. 11. Users were able to hear each other from speakers in the HMDs.

Figure 11: A point cloud in the VR environment

It was also important that the Connec2 platform was stable and that people could accurately use the VR controllers to teleport forward in the VR environment. Moreover, during meetings it is typical that employees use their laptop to access different files. So, it was also necessary that users had access to their own laptops. Users could then access important files that were on their laptop in the VR environment. They could do so by enabling the functionality in Uplink on their own private laptop. So, they needed to install Uplink on their own laptop first, which can be done from the Connec2 website. After Uplink had been installed, the users could log into the platform with the account that they were using for streaming their point cloud on the VR system. By logging in with the same account, they could then enable the 'remote desktop' functionality on their own laptop and access their files by using the 'remote desktop' feature on special 'smart screens' in the VR environment, such as the one that can be seen at the top of Fig. 12. Uplink is also shown in Fig. 12. This functionality is similar to the screen-sharing functionality that can already be found in videoconferencing software. Using the same functionality users could also use the Internet on the smart screens in the VR environment.

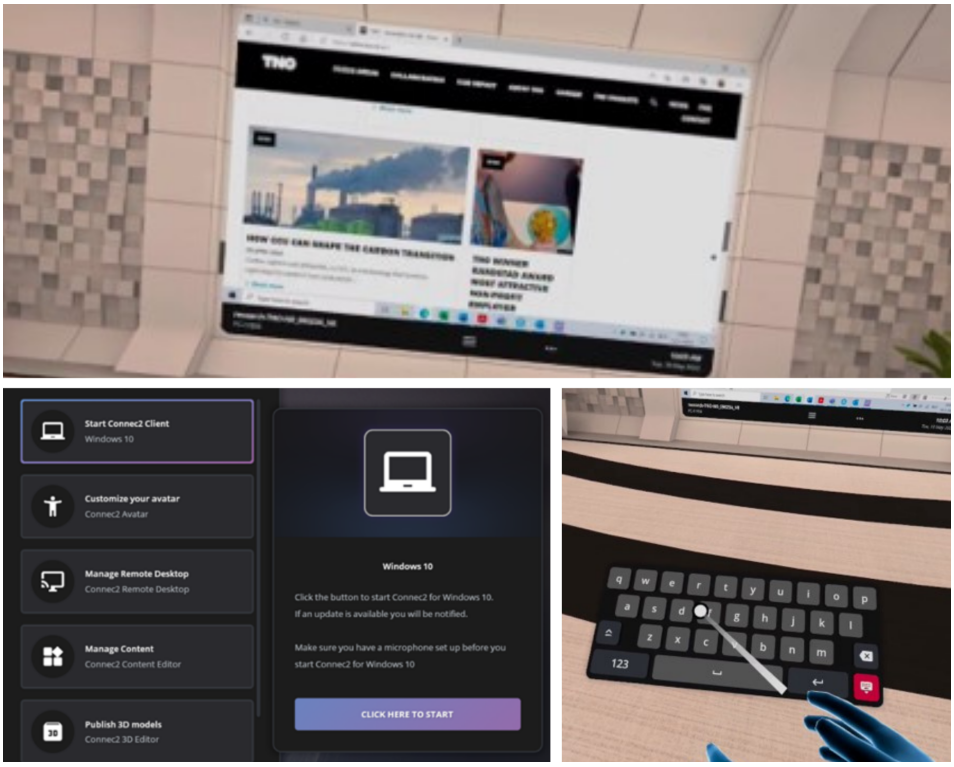


Figure 12: The Uplink menu on a laptop, a smart screen in VR and a virtual keyboard

Furthermore, note-taking was done by, for example, opening a Word-document with the 'remote desktop' functionality. Whenever it was possible to type something on the screen, a virtual keyboard popped up in front of the user. The VR controllers could then be used to type in a Word document using the virtual keyboard. Fig. 12 shows the virtual keyboard that users could use in the VR environment. Even though it was possible to type, it could not be done in the same way as typing when using a physical keyboard to take notes. So, only small note-taking was possible at the time. Taking notes on a physical keyboard was not yet possible with the system.

Moreover, Fig. 13 shows the location where the users were first spawned in the VR environment. The users could then decide between two spaces to hold their meetings. These two spaces were opened up so that two teams of up to three users could use the platform simultaneously. As previously mentioned, the system could not host more than three users at the same time in one space. So, this solution was brought into existence. However, later on, it was found that Connec2 had placed a constraint of only allowing three users to log in at the same time. Thus, teams could not hold their meetings simultaneously. The users could still move between the two spaces and decide where to hold their meetings. The first space was the Index, seen at the bottom in Fig. 13,

Figure 13: An Oculus menu on the left, the Connec2 menu on the right and the TNO Index

which shows the office building and the multiple meeting rooms. This space was mostly used by the participants of the study. Spatial audio was used in the Index to make sure that point clouds that were close to each other could hear each other in a room and that point clouds who were further away could not hear the other room. Besides the Index, there was also the Venue space which could be accessed with an elevator from the Index. Users were then streamed to a completely separate space where they could not reach the point clouds of the users who were in the Index, except for when they used the elevator again to go to the Index.

There was also a menu in the VR application which could be accessed using a button on the left VR controller. This menu showed whether there were other users online in the VR environment. There were also options to mute yourself, etc. The Connec2 menu can be seen in Fig. 13. The regular Oculus menu, which could be accessed using a button on the right VR controller, also allowed users to take screenshots of their view in the VR environment. The screenshots were then saved on the HMD. This menu can also be seen in Fig. 13.

All in all, almost all of the must have and should have requirements were taken into account. The only requirement that was not implemented anymore from those was the self-view. This feature would have let users check on themselves to see whether their point cloud appeared correct or not. If it looked too vague or if there were too many or too little points, the user would then have been able to see that and try to adjust their point cloud by moving the chair closer to the ZED camera so that the body could fully be captured.

On the other hand, most of the could have and would like to have requirements were not included in the system as there was already much time spent on making sure that the point clouds were correclty captured and streamed to the HMDs. It took much time to create a smooth experience of viewing the point clouds. The alignment of audio and video streams also took much more time than expected to make it seamless. So, much time was spent on perfecting the technical aspects of the system and integrating the TNO capture module and the Uplink platform well. Especially everything related to the whiteboard functionality was left out due to the time constraints and because that would have probably taken a lot of processing power as well.

3.4 Procedure of research

This subsection contains the steps that were taken to test the system internally and externally, meaning first with the team and then other TNO colleagues. It also includes a detailed instruction on what the participants of the research had to do to try the system out. This not only includes the steps they had to take before using the system but also the steps they had to take while using the system before they could have their meeting. Moreover, it also explains how the data from the meetings was collected, processed and how it was analysed.

3.4.1 Internal & external testing periods

The internal testing period in February and March was meant to put the system to the test with the team from TNO and Connec2 that was involved in integrating TNO's capture module with the Uplink platform. The aim of this internal testing period was to make sure that the system worked accurately and well-enough to hold real-time collaborative meetings of 30 minutes to one hour. It also needed to be assured that the users would be able to see the photorealistic avatars during the meetings. There were multiple meetings to investigate why, at first, the photorealistic avatars did not appear, why the connection of the capture module was not working well with the Uplink platform or why the 'remote desktop' functionality would not work. Moreover, there were also some hardware issues that needed to be resolved before the deployment could begin. It took quite some time to gather all of the supplies due to a limited budget and the limited contact that TNO employees had because of COVID-19 regulations. The supplies sometimes had to be gathered from different teams and departments at TNO. The main cause for these problems was that multiple TNO and Connec2 employees fell victim to COVID-19, which hindered the process a lot. Many meetings had to be rescheduled. Because of this, a two-month delay on finishing the system on time was accumulated.

After continuously readjusting the system, going over multiple tests and acquiring all of the hardware, it was finally decided that users of the system could start using the systems from April onwards. So, the systems were quickly deployed at the four chosen locations at the end of March. Sometimes, there were still hick-ups in the deployment because some of the PCs and laptops that were acquired still had some issues. However, in the end, it was still possible to start recruiting participants for the study from the first of April.

Thus, the external testing period could finally begin. This period was regarded as the first roll-out of the system as it is TNO's intention to roll out the systems at more locations later in 2022 after the current study. This first roll-out period was used to collect feedback on the system and test its capabilities so that the system can be improved for the second phase of the project. With a working end-to-end pipeline, communication via the VR system was possible with minimal audio and video delays which were unnoticeable for users if there was a stable WiFi connection.

3.4.2 Actions taken before and while using the system

An overview of the communication towards (potential) participants and the steps they had to go through to take part in the study can be seen in Fig. 14. First, the recruitment email and accompanying flyer were sent out. These can be found in Appendix B and Appendix C. After some colleagues expressed their interest in trying out the system, an informed consent was sent to them, which they had to fill in using Adobe Fill & Sign and send back. The informed consent contained information on how the system worked and how the data of the participants would be gathered. The informed consent can be found in Appendix D. After filling in the consent form, the participants received the log-in information that was needed for them to use the systems at their locations. It was also explained to the participants how they needed to book the systems for their meetings. This email can be found in Appendix E.

It was necessary that the systems were bookable so that different meetings were not held at the same time while the systems were already in use by other colleagues. So, Sharepoint team sites were opened up with the help of the IT servicedesk from TNO. These team sites were used

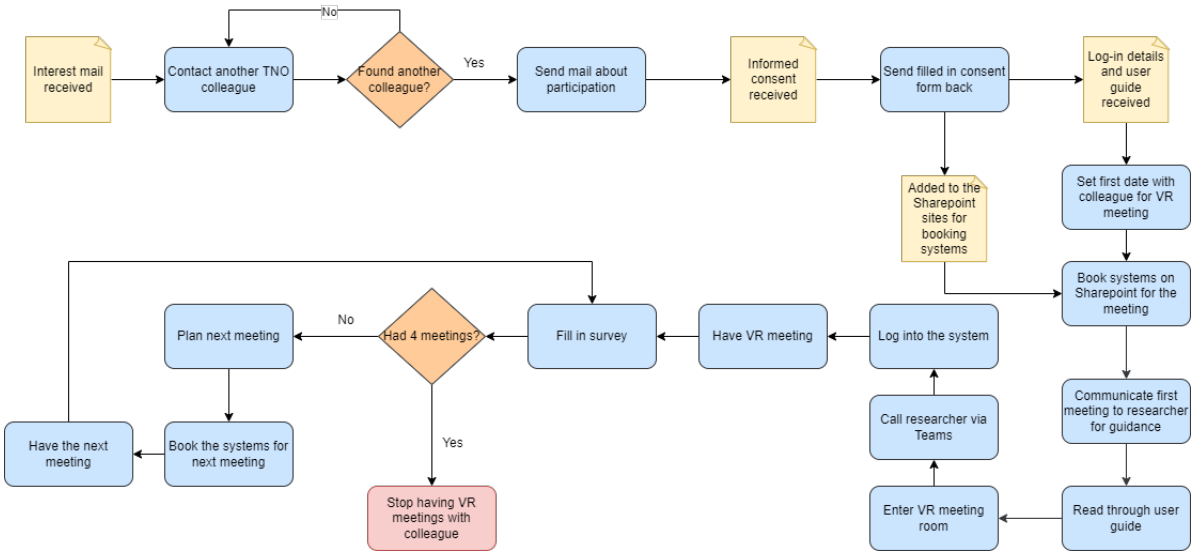


Figure 14: Overview of steps participants have to take to complete the study

to form the emails that were used for logging into the systems and they were also used for booking the systems. An example of one of the Sharepoint sites can be seen in Fig. 15. A calendar was added to the site which could be used to set appointments with the system. In this way the systems were made bookable. Anyone from the participants who wanted to use a system at a certain location, needed to set an appointment in the calendar of the Sharepoint site of that specific system first. The Sharepoint sites were private, so the participants only had access to these sites when they had filled in the consent forms and were added to the sites by the team.

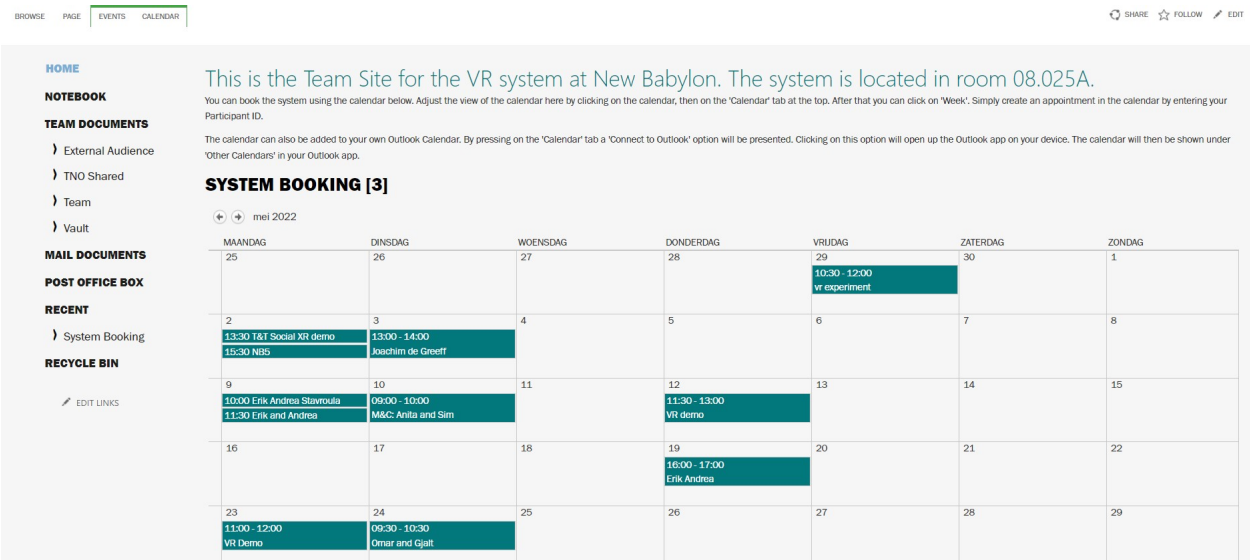


Figure 15: The Sharepoint site and calendar for the system in room 08.025A at New Babylon

For booking the systems, the participants also needed to set a date with each other. The date of the first meeting also needed to be communicated to the researcher because they would guide the participants in using the system and tackling problems if any occurred. For the rest of the meetings, it was assumed that the participants would not need any help anymore. So, the first meeting was regarded as a trial meeting. However, even after this trial meeting the participants still had to fill in the survey. The survey was sent to the participants before the first meeting because they had to fill it in after every meeting they had in VR. As a reminder to fill in the survey, a QR code with a link to the survey was also placed on the desks of each of the VR systems. These surveys were used to compare the participants' experiences with the system. By comparing the partici-

pants' answers to the questions in the surveys, a negative, neutral or positive score for different elements could be analysed and a conclusion could be drawn for whether a first timer effect was found.

Additionally, a unique participant ID was also sent out to each of the participants. They could fill in this participant ID in the survey. This ID was used to track how many meetings the participants had and when their meetings took place. Especially the indication of when the meetings took place was important for linking the survey data to the body tracking videos that were made of the participants in each of their meetings. This was essential for analysing the body tracking data. The survey was also used to check what the conditions for all of the meetings were.

Lastly, with all of the previous information at their disposal, the participants could then finally use the system on the day and time they had booked the system to hold their meeting. The participants first had to turn on the laptop/PC in the physical room and find the Uplink platform on it. Once they were logged into the platform, they could navigate to the 'enable depth camera' button in the menu and turn on the depth camera. After that was done, they could put on the HMD and find the Connec2 application in the Oculus menu using the VR controllers. The participants were sometimes first also prompted by the HMD to recalibrate the headset. However, this was not necessary, so the 'stationary calibration' setting on the HMDs had to be selected. This was communicated to the participants in their first meetings. When they had found the Connec2 application in the menu, the users could select it and they would automatically be logged into the account for that specific system. Once they saw the Index, they could use the VR controllers to move in the VR environment, find their partner and find the perfect meeting room for them to hold their meeting in.

3.4.3 Data collection & measurement schemes

As previously mentioned, body tracking videos and a survey were used to evaluate the different quantitative and qualitative variables mentioned in sections 3.1.1 and 3.1.2. The survey was made using the Gorilla Experiment Builder [1], which is a cloud-based research platform. So, all of the data was stored on that platform and was downloaded from the Gorilla website. The software that captured and stored the body tracking videos was made by a TNO colleague, who also helped in building the software for the capture module. Using the capture module the body tracking data could be stored into videos locally on each of the VR systems. Next to that, some system performance data could also be stored locally on the systems. Both the system performance data and body tracking data from these systems have been transferred to a USB stick after all the VR meetings took place.

Unfortunately, due to unanticipated privacy regulations from Connec2 prohibiting them from storing data in log files, and a limited amount of time for TNO, it was not possible to collect the ideal data on some QoS variables, such as network latency and packet loss. Therefore, solely, the system performance data that was stored locally on the systems was collected for the analysis. Data on the upload framerate, the date and time, average CPU usage (%), average GPU usage (%), average RAM memory (GBs), average received bytes/s and average sent bytes/s was stored per meeting. This data was intended to be used to check if there were any outliers which might have caused technical issues and reduced the QoS.

Moreover, the specific data that would be analysed from the body tracking videos were body postures, body rotation, hand gestures and head rotations. However, there were a lot of videos of at least 30 minutes to one hour long. So, it was decided that body tracking videos would only be further analysed if the survey results deemed it necessary to do so. The following criteria would be used to decide whether to look at the body tracking data or not:

- The survey answers were vague and confusing
- The survey answers were more on the positive side compared to the rest
- The survey answers were more on the negative side compared to the rest

[1] https://app.gorilla.sc/

- The participants reported that there were technical issues during the meeting
- The participants reported that they were very engaged
- The participants reported that they were not engaged enough

Besides objective measurements, self-report measures continue to be popular as well for characterising the psychological state of user engagement. It is important to use multiple measures of engagement to capture both behavioural and psychological aspects of engagement [61]. There is no measurement scheme yet for evaluating engagement in social VR business meetings. Therefore, for each subjective QoE metric and element regarding engagement, corresponding existing subjective measurement scales were found. These were used to create the survey that the participants of this study filled in. The following measurement schemes were used:

- The holistic mediated social communication questionnaire [59] (14 items used)
- The networked minds measure of social presence [6] (6 items used)
- The experience of presence [43] (2 items used)
- Measuring presence in virtual environments [62] (7 items used)
- The experience of immersion in games [30] (6 items used)
- The impact of eye gaze [23] (6 items used)
- The system usability scale [7] (6 items used)
- The engagement in video game-based environments scale [61] (4 items used)

The full survey can be found in Appendix F. Some of the questions were slightly rephrased to make the survey more coherent in its tone towards the participants. The first section of the survey concerned questions on the demographics of the participants and the meeting characteristics of their meeting in VR. These results were used to link the survey results to the body tracking data. Moreover, the participants were also asked about the perceived time that they spent interacting with the system. This question was included to investigate whether people experienced the interaction with the system to be shorter or longer than the actual time spent with the system. Shorter perceived times may have indicated that the participant was engaged in the meeting, and did not notice how much time was going by. Longer perceived times may have indicated that the participant was not engaged in the meeting, and was possibly somewhat bored.

The next section included questions that evaluate social presence. Four items were used from [6], one item was used from [23] and four items were used from [59]. Following those questions, a section with questions that assessed immersion were also included. This section included six items from [30]. All of these questions were rephrased from a gaming context to a virtual reality meeting context. Moreover, the next section contained questions that assessed embodiment. However, since the users could not see themselves in the VR environment, only three items were included. One item from [43] was included and two items from [59] were included. On the other hand, 11 items for assessing the QoI variable were included in the next section of the survey. One item was included from [61], five items were included from [62], one item was included from [43] and four items were included from [59]. The next section assessed QoC and it contained 11 items. Two items were included from [6], five items were included from [23] and four items were included from [59]. Furthermore, usability was also assessed in a following section. Seven items were used, four items were from [7] and three items were from [61]. Lastly, to phrase the open questions on user experience, some questions were adapted. Two items each from [62] and [7] were adapted.

Because the holistic mediated social communication questionnaire is not yet fully validated, it is necessary for TNO to include the questionnaire in this research. For each item that was used from the questionnaire, there was a similar validated question in the survey used in this study. The reason for including the questionnaire was so that other TNO colleagues were able to validate it with participants of a study who took part in VR meetings over a longer period of time. So, the questionnaire needed to be validated with people who had multiple meetings in VR. Even

though the data collected from this study will be used to try and validate the holistic mediated social communication questionnaire, this was not done in the current study. This work will be taken up by other TNO colleagues since the validation of the questionnaire falls out of the scope of the current research.

Lastly, all of the categories of the survey and the questions in each category were randomized to combat survey bias. This ensured that the participants were alert when they had to fill in the survey after each meeting. There were also two questions added to the survey at random intervals in the middle of the survey to make sure that the participants were paying attention while filling it in. In these questions they had to fill in a specific answer that was mentioned in the question. If the participants had filled in an answer that was different to the one asked to fill in, the survey would not be taken into account during the analysis. Next to that, a question at the end of the survey was also added to evaluate the subjective QoS of the system. The participants' experience with the system was important to consider. It was necessary to include their perception of the system and whether they thought that the system functioned well enough or not. They could also explain why the system did or did not function well to them.

3.5 Data analysis method

Different types of data were collected during this study. The sections below show how the different types of data were cleaned and processed and what decisions were made while processing the data. It is also explained how the cleaned data of this study was analysed.

3.5.1 Data preparation

The data on the system performance measures was stored in a log file per meeting. However, if the camera was turned off and on more than once during the meeting, new log files were made for that meeting. So, for many VR meetings there were multiple log files instead of only one per meeting. If this was the case, it was investigated how much data was stored in the files. If the file contained less than 100kB of data, the file was deleted as there were other files which contained more than 300kB of data. Even after deleting some files, there were still multiple files for some meetings which contained much data. In that case, those log files were kept for that meeting and bundled per meeting. All of the log files were also renamed to indicate for which participants and which meetings the files were made. This was possible because the files stored the date of the meeting and it was also noted which file was created at which location. By linking these files to the respective survey results, the data could be analysed more completely.

Moreover, there were also multiple body tracking videos for one meeting if the camera had been turned off and on. These videos were bundled per meeting and per participant as well. All of these videos were played using the 'poseplayer' software that was made by a TNO colleague. If the poseplayer was not able to play a certain video, due to the file being corrupted, then that file was deleted. The body tracking videos were then also linked to the survey results of each participant. Afterwards, the body tracking videos would be coded. Since some videos were up to an hour long, it was likely that many different codes would have to be used for one video. However, the video would eventually be coded into the codes that occurred most often. The coding strategy was explained in section 3.1.2.

Besides the data for the study, there was also much data gathered on system performance measures and body tracking data from one-time demos with higher management employees at TNO and with people who are not employees from TNO. These demos were organized to showcase TNO's Social XR research to others. This data was not used in the current study since the one-time demos were done with people who did not participate in this study and have therefore not filled in the consent form. However, the data was still gathered because it was not possible to stop the data collection on the systems for a limited time. The gathered data was deleted immediately after gathering all of the data from the different systems. The date and time of the one-time demos was noted so that the correct data was deleted.

All of the questions in the survey were used to assess social presence, immersion, embodiment, QoI, QoC and usability were closed questions that were answerable using a 7-point Likert scale. To prepare the data for the analysis, the Likert scale scores were changed from 1 to 7 to -3 to 3. This was done so that the data better represented a negative, neutral or positive score of the participants to the elements. Moreover, there were five open questions and one closed question for assessing the user experience, of which two questions were related to the subjective QoS. The survey data was collected per participant and per meeting. The data was cleaned and prepared using R in RStudio [2]. Almost all the survey data was exported from Gorilla into one CSV file. There was one extra CSV file exported from Gorilla because that file contained data from a previous version of the questionnaire, to which the QoS questions were not yet added. This survey was still filled in by one participant, resulting in a 'NA' to the answers for the QoS questions for this participant. Gorilla had also collected some data which was not necessary to use for the study, such as which browser was used to fill in the survey etc. This data was omitted, resulting in only the data on the survey questions and answers.

After cleaning the data, it also needed to be aggregated for the different elements of the study. As seen in Fig. 3, there were multiple questions for certain elements. The results of these questions needed to be aggregated to find a value for each element per participant and per meeting. Furthermore, there were some negative items for some of the elements in the questionnaire. The scores to these items had to be inverted first. So, the results to questions 14 to 17 regarding usability were inverted. This was also done for the results of questions 51 and 53 regarding immersion. Moreover, for the results of questions 46, 48, 55 and 58 regarding QoI, this was also done. By inverting the results; the questions were rephrased. To check whether a question needed to be inverted, the question was compared to the other questions of the element and it was rephrased to see what the effect would be on that element if the question was inverted. For example, this was done for question 53 of the questionnaire in the following way:

- *Question 53: I felt the urge to stop the meeting.* A participant filled in that he disagrees. However, this contributes to a decrease in the score of immersion, even though the questions show that the participant was immersed.
- *Inversion of question 53: I did not feel the urge to stop the meeting.* Someone then agrees, which contributes to a positive score for immersion.

3.5.2 Analysis strategy

Multiple sub-questions were drafted for the current study, which have been stated in sections 1.2 and 2.4. These questions are either answered using the literature review, the data analysis or a combination of both. Table 2 shows what the answers from the literature study are to some of the sub-questions. The table also shows how the rest of the sub-questions are answered through the data analysis. The answers to those questions can then be found in the Conclusion section. The questions in black were answered using the findings from the literature review. The questions highlighted in blue are answered using the data analysis and in the table it is explained how the analysis is done. The questions highlighted in red are answered through both the literature review and the data analysis. For these questions, it is explained in the table what the findings from the literature review are and how the data is analysed.

[2] https://www.rstudio.com/

Table 2: Sub-questions including literature review findings and data analysis strategy

Sub-question	Literature findings	Data analysis strategy
RQ2: How can meeting engagement in social VR business meetings be captured and evaluated?	This can be done by analysing body language cues such as body postures and gestures. Because of the VR aspect, social presence and immersion are also important elements related to engagement.	No further analysis was needed for this sub-question as this was part of the method to evaluate engagement.
RQ3: What were the conditions for meeting engagement in f2f, hybrid and online meetings?	In f2f meetings engagement was highest because of factors, such as body language cues and eye gaze. These were either missing or very hard to interpret in hybrid and online meetings.	No further analysis was needed for this sub-question as it was only necessary to know why f2f meetings are most successful.
RQ4: How can the quality of experience be defined for social VR business meetings?	It can be defined as as a multidisciplinary indicator that includes all the different aspects related to the experience that a user has in VR. It also includes the user experience.	No further analysis was needed for this sub-question as the definition was found from the literature review.
RQ5: How can the quality of service be defined for social VR business meetings?	The quality of service comprises system performance measures that can impact the quality of experience.	No further analysis was needed for this sub-question as the definition was found from the literature review.
RQ6: How can the first timer effect be defined?	As a wow effect because some experience is novel to its users. It happens when people experience VR for the first time and only spend a short amount of time in it. This effect is diminished when people again use VR technologies.	No further analysis was needed for this sub-question as the definition was found from the literature review.
RQ7: How can the relationship between quality of service (QoS) and quality of experience (QoE) be defined for social VR business meetings?	Literature has shown that QoS can have an impact on QoE. However, this is not always the case.	A Pearson correlation is done with the data of all meetings for these two elements to investigate whether in this case there is a relationship or not.
RQ8: How can the relationship between quality of experience and meeting engagement be defined for social VR business meetings?	Since QoE comprises all of the experiences that a user has in VR, it can be assumed that engagement is related to it as well. However, this has not yet been shown in other studies.	A Pearson correlation is done on the data of all meetings for this question as well to investigate whether there is a relationship or not.

RQ9: How can the relationship between embodiment (using the photorealistic user representation) and meeting engagement be defined for social VR business meetings?	An assumption is made in the literature that photorealistic avatars enhance engagement. However, it needs to be investigated whether that is the case with the photorealistic avatars used in this study.	Again a Pearson correlation is done using the embodiment and meeting engagement data from all meetings. Moreover, textual data regarding the user representation is also analysed.
RQ10: How can the relationship between the quality of interaction (QoI) and the quality of communication (QoC), and meeting engagement be defined for social VR business meetings?	Literature shows that these are important elements for social VR meetings. So, it is interesting to investigate their relationship to engagement.	A Pearson correlation is done with the data of all meetings. Moreover, textual data regarding interaction with the VR environment is also analysed.
RQ11: How can the relationship for social presence and immersion be defined for social VR business meetings?	These two elements are both needed for creating successful VR experiences. So, it is interesting to investigate the relationship in a meeting context.	A Pearson correlation is used for this as well with all of the meeting data.
RQ12: Do feelings of engagement diminish over a longer period of using social VR for business meetings when compared to using it for the first time?	Based on the literature findings, this is possible if a first timer effect occurs.	It was intended to do a MANOVA analysis to answer this question. However, since only one pair had done more than 2 meetings, only their results are compared to each other with line graphs.
RQ13: How usable is the social VR system to use for a longer period of time compared to other VR systems?	The literature review shows that there are no commercial social VR systems yet that support photorealistic avatars. These are supported in systems in a research setting but the systems are still in initial stages where people cannot efficiently communicate with each other yet over big distances.	Again, it was the intention to do a MANOVA analysis. However, only the results of one pair are analysed. This was done with line graphs. Moreover, textual analysis provides more insights on usability as well.
RQ14: What do users experience when working with a social VR system for a longer period of time compared to other VR systems?	The literature review shows that users can sometimes experience a first timer effect, and can also sometimes have some issues with using the systems for a longer period of time, especially for business meetings.	Only textual data is used to answer this research question since the textual data provides detailed insights.

Chapter 4

Results

This chapter presents the results from the data analysis. First some descriptive statistics are shown to illuminate how many meetings and which kinds of meetings the data was collected from. Next the average scores for the elements over all the meetings are shown. Pearson correlation results are also presented. The results for one pair in particular have also been visualized. Lastly, the textual data has also been visualized in word clouds centered around specific topics, which were found by deductively coding the data.

4.1 Descriptive statistics

Table 3: Total number of meetings per meeting number, type of meeting and location

Meeting 1	Type of meeting	GR	NB	SB	OWW	Total 1st meetings
	Update	0	2	3	0	5
	Presentation	0	0	0	0	0
	Brainstorm	0	0	1	0	1
	Informal	2	2	1	1	6
Total 1st meetings		2	4	5	1	12
Meeting 2	Type of meeting	GR	NB	SB	OWW	Total 2nd meetings
	Update	0	0	2	0	2
	Presentation	0	0	1	0	1
	Brainstorm	0	1	1	0	2
	Informal	1	0	0	0	1
Total 2nd meetings		1	1	4	0	6
Meeting 3	Type of meeting	GR	NB	SB	OWW	Total 3rd meetings
	Update	0	2	0	0	2
Meeting 4	Type of meeting	GR	NB	SB	OWW	Total 4th meetings
	Update	0	1	1	0	2
All meetings	Type of meeting	GR	NB	SB	OWW	All meetings
	Update	0	5	6	0	11
	Presentation	0	0	1	0	1
	Brainstorm	0	1	2	0	3
	Informal	3	2	1	1	7
All meetings		3	8	10	1	22

Table 3 shows all of the meetings that have been done using the social VR system. The table also shows the number of meetings that have been done from the different locations where the systems were placed and it also shows the types of meetings that were done.

Only one pair, two participants, was able to hold the third and fourth meeting using the VR system. Their results have been added to the table as well. All of the other rows would contain 0 results for the third and fourth meeting. Therefore, these rows have been omitted from the table. The categorization of whether a meeting was either an update meeting, brainstorm meeting, presentation or informal meeting, was done by the participants themselves. Because of this, subjects of a pair sometimes categorized the meeting differently than their partner.

In the end, update meetings were held the most with the system. Moreover, many informal meetings were held as well, which were categorized as 'other' in the questionnaire. To make it obvious that these meetings categorized as 'other' did not have anything to do with projects, the meetings have been categorized as 'informal' for further analysis.

In total, 22 meetings were held with the system. One pair held four meetings with the system and two other pairs held two meetings with the system. Moreover, there were three other pairs who were only able to hold one meeting with the system. Lastly, in total, there were ten meetings held from the location Soesterberg, eight meetings held from New Babylon (The Hague), three meetings held from Groningen and one meeting held from Oude Waalsdorperweg (The Hague).

4.2 Average questionnaire results

As stated in section 3.5.1, the answers to the questions that corresponded to either embodiment, QoS, usability, QoI, immersion, social presence, and QoC, have been aggregated for their respective elements to analyse the scores of these elements. Fig. 16 shows the average score over all 22 meetings for the elements that the questionnaire collected data on. Embodiment received the lowest score on average with a score of **-1.44**, ranging on a scale from -3 to 3, with 0 being a neutral value. Negative values indicate an insufficient or poor reaction of the participants to the element, while positive values indicate a satisfactory or excellent reaction to the element. Moreover, QoC scored the highest of all the elements with a score of **0.9**.

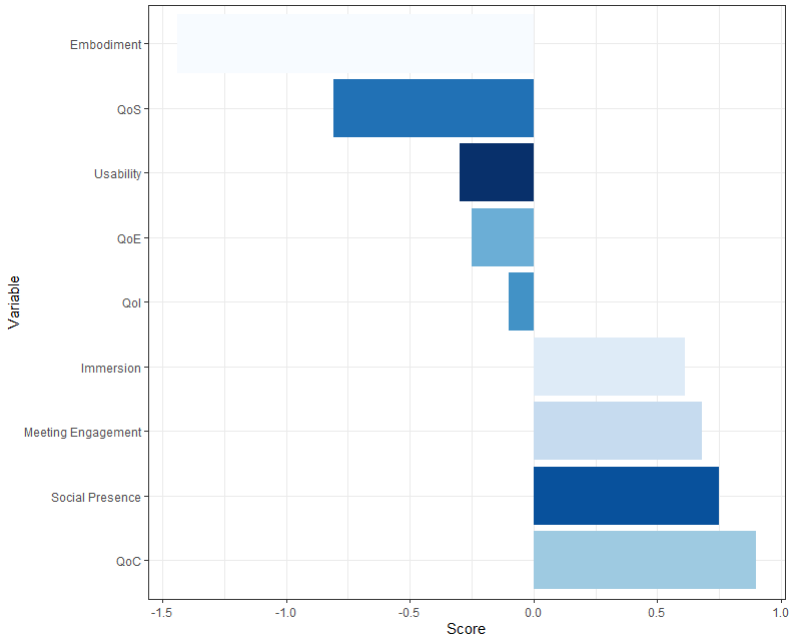


Figure 16: The average scores for the variables over all meetings

Furthermore, the scores for meeting engagement and QoE have been aggregated based on the model in Fig. 3. Meeting engagement scored positively with a score of **0.68** on average. On the

other hand, QoE scored slightly negatively with a score of **-0.25**. The score for the QoE is based on contradicting results from, mostly, the elements embodiment and QoC.

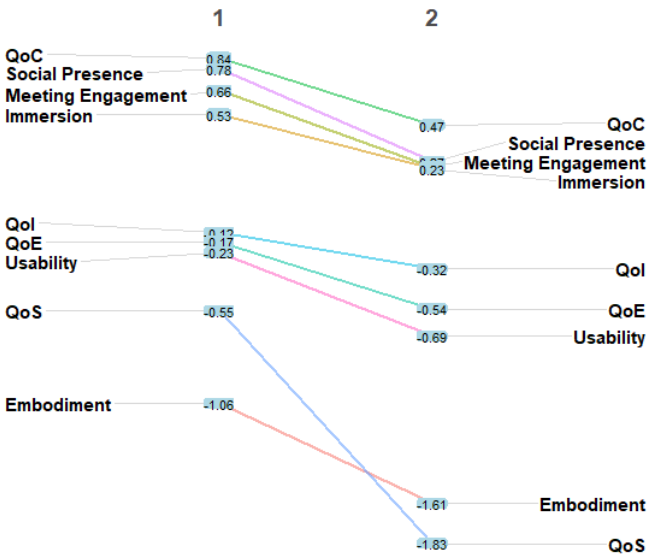


Figure 17: The change in scores from meeting 1 to meeting 2

While the previous figure shows the average scores over all the meetings, Fig. 17 presents the change in the scores of the participants from the first meeting to the second meeting. However, the scores of the second meeting are from half the participants who took part in the first meeting. There were twelve participants who held the first meeting and only six participants who also held the second meeting. Based on that data, the figure shows that the scores of all the elements fell for the second meeting. Especially the score for the QoS decreased to a greater extent compared to the other elements.

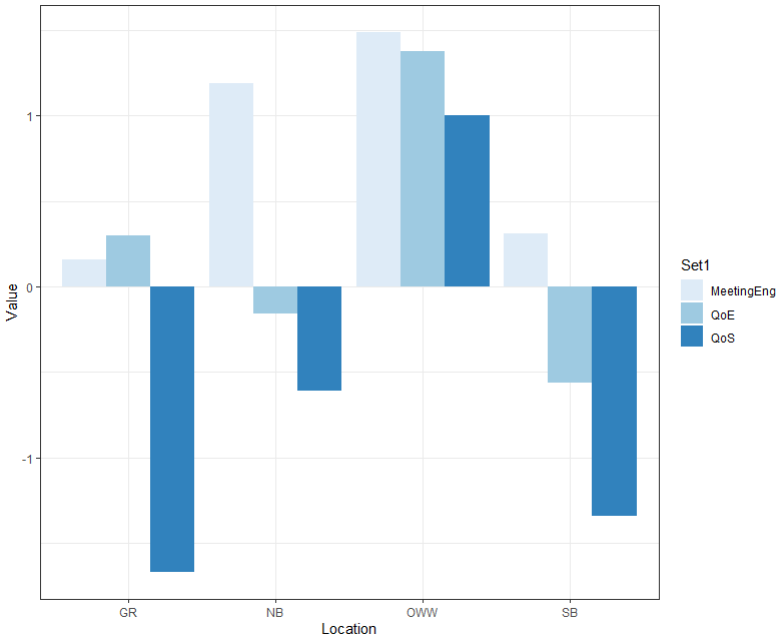


Figure 18: The scores for the QoS, QoE and MeetingEng per location

To further analyse the QoS, QoE and Meeting Engagement variables per office location, the bar chart presented in Fig. 18 was created. This visualization shows per location the results that the participants reported for the QoS, the aggregated QoE and the aggregated meeting engagement. The QoS scored the lowest in Groningen, while it scored the highest in Oude Waalsdorperweg.

However, it should be taken into account that there were only two participants (total: three meetings) who gave scores to the elements from Groningen. Moreover, there was only one participant (total: one meeting) who rated the QoS for Oude Waalsdorperweg. The data for New Babylon and Soesterberg shows that New Babylon scored somewhat better for the QoS than Soesterberg. However, all of the scores are still on the low side.

Moreover, the QoE element scored the highest for the Oude Waalsdorperweg location and the lowest for the Soesterberg location. While the QoS was rated somewhat better at New Babylon, this does not seem to have made much of a difference for the QoE variable. Moreover, at Soesterberg the QoS variable was rated poorly. Yet, again, this does not seem to impact the QoE variable too much. More surprisingly, the QoE variable is rated slightly positively at the Groningen location while the QoS was rated the poorest from all the locations.

Lastly, the meeting engagement element did not score negatively at any of the locations. However, the visualization shows that at Groningen, meeting engagement was experienced to be the lowest. At Oude Waalsdorperweg it scored the highest. If that office location is not taken into account due to the low amount of data, the next high-scoring office location would be New Babylon. The visualization also shows that meeting engagement does fall when the QoS at any of these locations is also low.

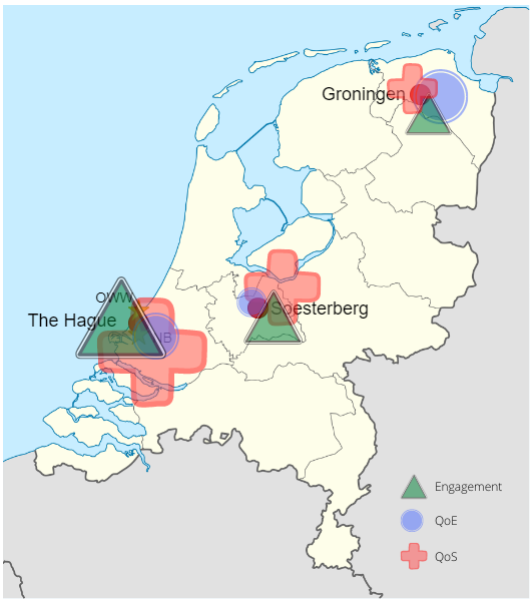


Figure 19: Mapping scores of locations GR, SB and NB

Fig. 19 also shows the scores mapped onto the different office locations in The Netherlands. The scores for the location Oude Waalsdorperweg are left out of this map. The different variables are represented by different shapes, as can be seen in the legend at the bottom. It is shown in the figure that the QoS and meeting engagement scored the highest results at New Babylon. So, for this location the QoS was highest of all locations, as was the score for meeting engagament. The QoE, however, was not the highest. While both meeting engagement and QoS were lowest at the Groningen location, QoE scored the highest here. Moreover, it can be seen that for the Soesterberg location, QoS was somewhat higher than Groningen, as was meeting engagement. However, this is not the case for QoE.

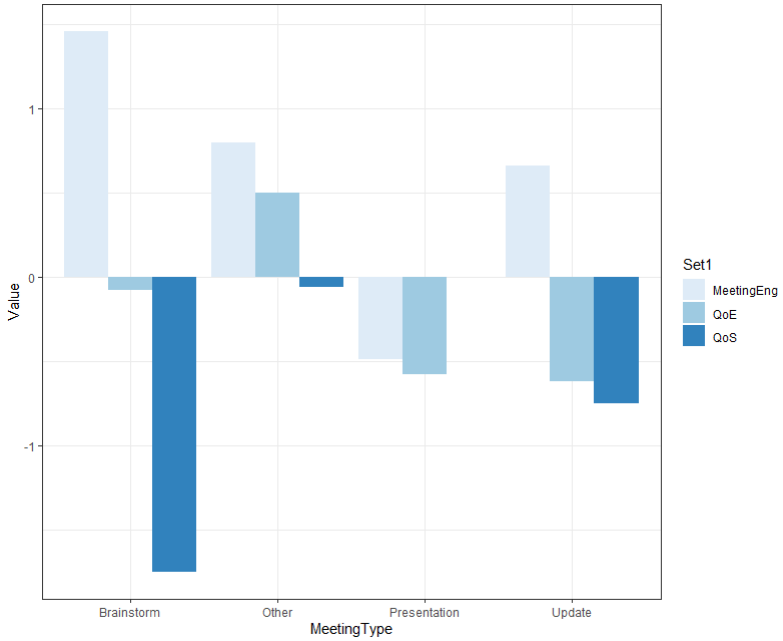


Figure 20: The scores for the QoS, QoE and MeetingEng per meeting type

Besides the office locations, a bar chart of the scores for QoS, QoE and meeting engagement has also been created based on the different types of meetings. This bar chart is presented in Fig. 20. QoS scored the lowest for brainstorm meetings. For the informal, or 'other' meetings, the score is still negative but it is higher than the rest. For the presentation meeting, there was no data on the QoS. Moreover, the QoE was highest for the informal meetings and lowest for the update meetings. Lastly, meeting engagement scored highest for brainstorm meetings and lowest for presentation meetings. However, both of these results are based on little data, as can be seen in Table 3.

4.3 Correlation analysis between the variables

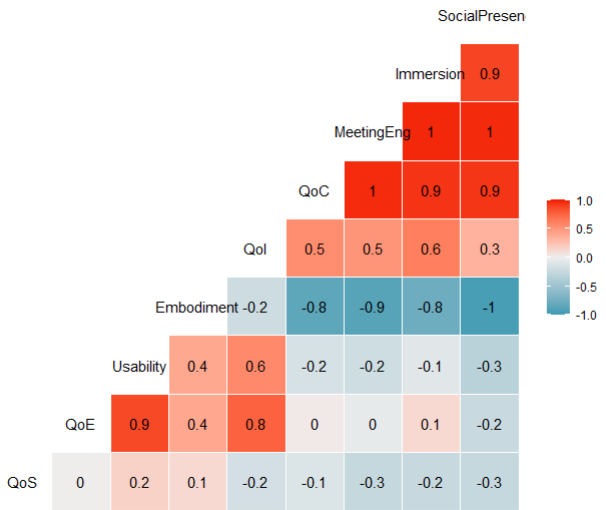


Figure 21: The correlations between all of the elements over all meetings

Fig. 21 shows a correlogram that depicts the correlations between the different elements over all of the meetings that were held using the system. The most notable relationships from these

findings are presented in the following:

- There is no significant linear relationship between quality of experience and meeting engagement.
- There is no significant linear relationship between quality of service and quality of experience.
- There is a weak negative linear relationship between quality of service and meeting engagement.

Although there are no strong relationships between these variables, the correlogram shows that there are moderate and strong correlations between some of the elements in the composite variables presented in the following:

1. There is a strong positive linear relationship between quality of experience and usability.
2. There is also a strong positive linear relationship between the quality of interaction and quality of experience.
3. There are strong negative linear relationships between embodiment and quality of communication, meeting engagement, immersion and social presence.
4. There is a moderate positive linear relationship between quality of interaction and immersion.
5. There are strong positive linear relationships between quality of communication, immersion, and social presence, and therefore also with meeting engagement.

(a) Correlations meeting 1

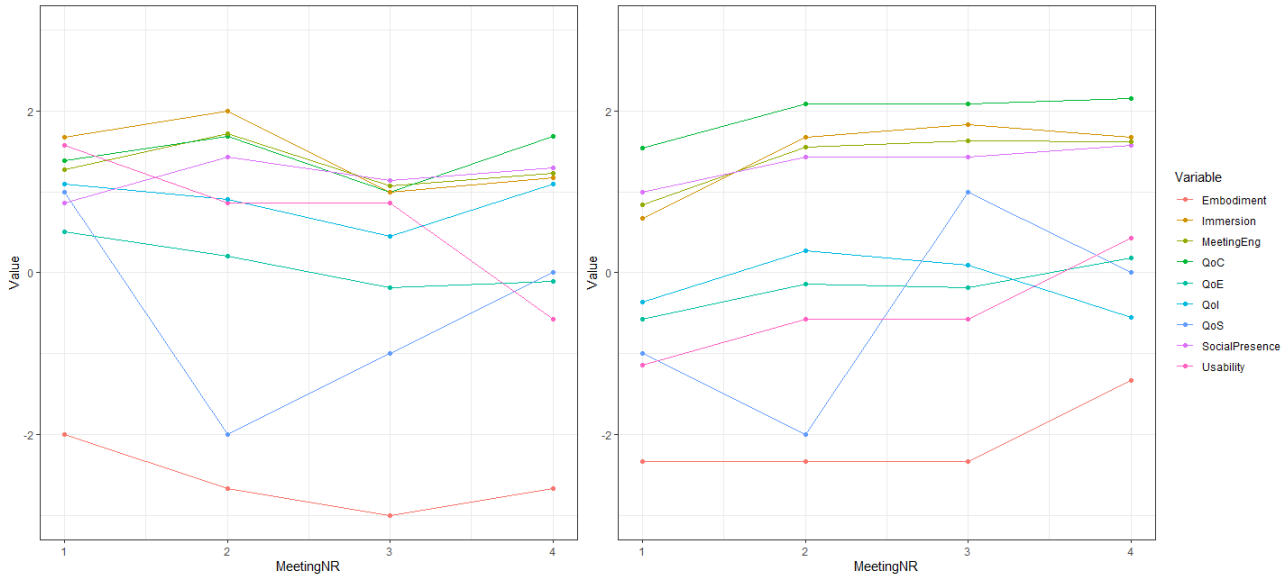
(b) Correlations meeting 2

Figure 22: Difference in correlations between meeting 1 and 2

Even though there are strong correlations found, Fig. 22b depicts a correlogram that only comprises the data for the second meeting of all participants. This correlogram shows that there are moderate and strong relationships between all of the elements. However, this might be the case because there is not enough data for this second meeting.

On the other hand, the correlogram in Fig. 22a is only comprised of the data from the first meeting. The data is only comprised of twelve participants, but the correlogram shows various different relationships.

4.4 Questionnaire results for participants NB1 and SB1



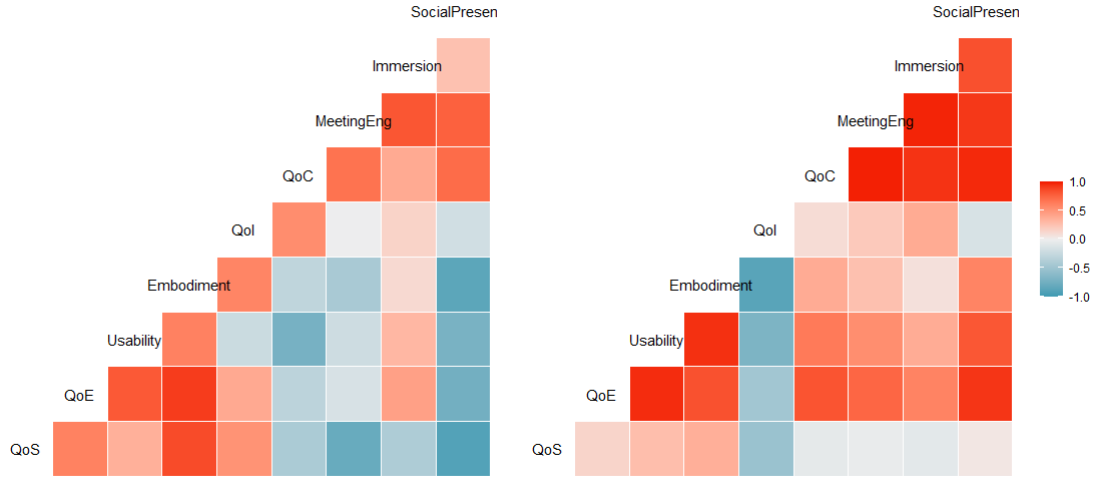
(a) Scores for participant NB1

(b) Scores for participant SB1

Figure 23: Variable scores of the pair who completed 4 meetings

Only one pair completed all of the four meetings. Their scores per meeting for each element are shown in Fig. 23a and Fig. 23b. It should be noted that the third meeting was done while both of the subjects were at New Babylon (The Hague). All of the other meetings were done while one participant was at New Babylon and the other was at Soesterberg. So, the results of the third meeting may be impacted by the change of location. QoC, immersion and social presence were all rated highly and positively in all of the meetings of these participants. Therefore, meeting engagement scored high as well. These scores remained relatively consistent for both of the participants. The scores for embodiment and QoE stayed relatively consistent as well, even though embodiment is the lowest scoring element in all of the meetings and QoE scored fairly close to zero for all of the meetings as well.

On the other hand, the scores for QoI, QoS and usability did not remain consistent. An element that changed significantly for both participants throughout their meetings was the QoS. Moreover, while for subject NB1 the usability of the system dropped towards the fourth meeting, the usability for SB1 increased. The QoI also shows different patterns for both participants.



(a) Correlations for participant NB1

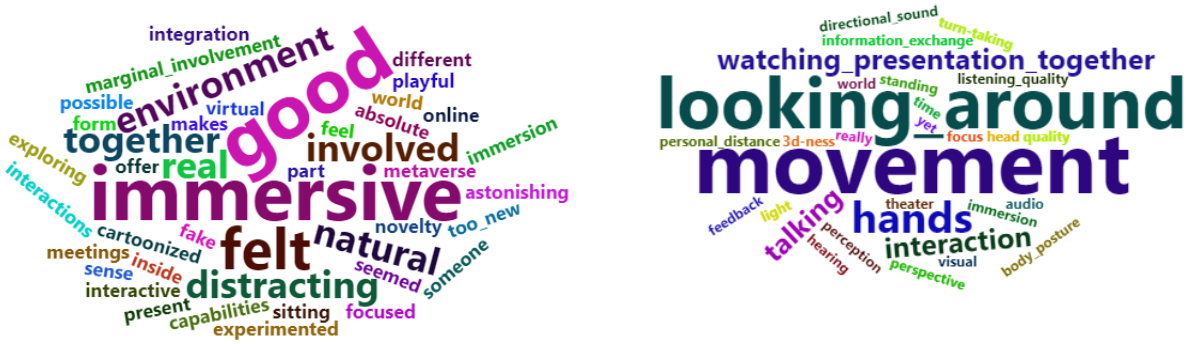
(b) Correlations for participant SB1

Figure 24: Correlations of variables of the pair who completed 4 meetings

As for the correlations of the meetings of these participants, which are shown in Fig. 24a and Fig. 24b, there are some big differences. While embodiment, usability, and QoE correlated negatively with QoC, immersion, social presence, and meeting engagement for NB1, they did not correlate negatively with these elements for SB1. Subject SB1's correlations are in this regard also different from the correlations over all of the meetings. This finding shows that results for correlations cannot be generalized over a small sample size.

4.5 Textual analysis for further insights

To summarize the answers to the open questions, word clouds were created that provide detailed insights of the experiences of the participants with the VR system.



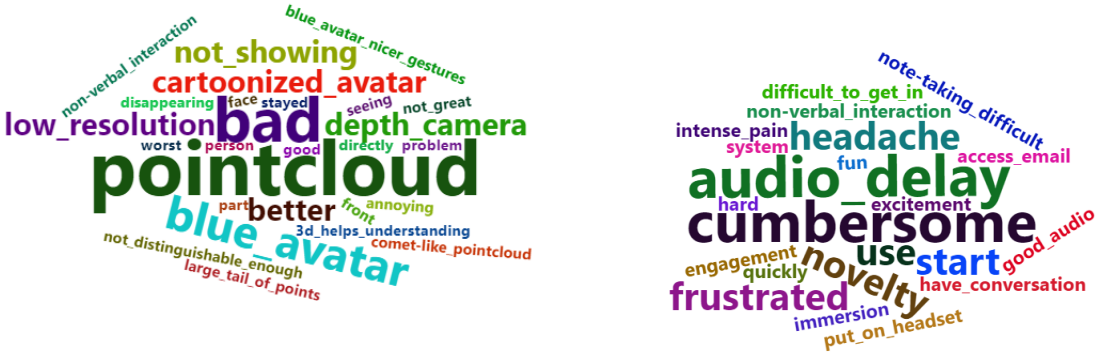
(a) Involvement in the VR environment

(b) Consistencies with real-world experiences

Figure 25: Word clouds for the involvement and feeling in the VR environment

The first word cloud in Fig. 25a shows that the VR environment was good and that users felt immersed in the environment. These words have been mentioned the most when answering

the question: 'How would you describe your involvement in the virtual environment experience?'. Moreover, the participants mention that they felt involved. However, they also sometimes described that they were distracted. Moreover, the following question was: 'Were there any experiences in the virtual environment that seemed consistent with real-world experiences?'. The word cloud in Fig. 25b shows that participants mentioned that they experienced movements in the VR environment that were consistent with real-world experiences. They expressed that looking around and moving your hands in the VR environment felt the same as when you are looking around and moving your hands in the real world. The same was said about watching a presentation together. The insights in Fig. 25a and Fig. 25b further explain why social presence, immersion and quality of communication were rated so highly in the closed questions of the questionnaire.



(a) User representation using point clouds

(b) Usability of the system

Figure 26: Word clouds on the user-representation and usability of the system

Furthermore, for the questions: 'Why would you use this system in the future for small project business meetings?' and 'Why would you not use this system in the future for small project business meetings?', there were mixed reactions regarding both the user representation as point cloud and the usability of the system. Fig. 26a shows that the point cloud did not receive positive reactions because it did not appear or had a low resolution. It was also mentioned that the standard, blue avatar of the system itself appeared more often than the point cloud. Therefore, some participants also commented that a cartoonized avatar would have worked as well. Moreover, the reactions to the usability of the system, shown in Fig. 26b, were quite mixed. The participants describe the system as cumbersome to use and frustrating. There were sometimes also audio delays. However, this last insight is surprising as the QoC was rated highest of all the elements. Some participants also did comment that the audio was good. Moreover, the participants did notice that the non-verbal interaction was conveyed well with the system and that they felt engaged.

(a) Technical issues experienced

(b) Improvements for the future

Figure 27: Word clouds on technical issues and improvements for the system

Lastly, as there were quite some technical issues with the system, it was necessary to create another question in the questionnaire that assessed the subjective QoS of the system. The participants were therefore asked whether they could elaborate on the following question: 'How did the system perform on a technical level?'. The answers to this question are shown in the word cloud in Fig. 27a. There were quite some set-up problems, the screen-sharing functionality did not always work and there was sometimes no WiFi connectivity. Because of these issues, the system had to be restarted multiple times. As for improvements regarding the system, which are summarized in the word cloud in Fig. 27b, the participants would like the equipment to be less heavy and to be at people's homes. They find it much more usable in that scenario. Furthermore, the point clouds need to be improved as well and the technical issues need to be solved. Participants also believe that the system is more valuable when meeting medium to large groups, where you can act more naturally with each other than with MS teams.

Chapter 5

Discussion

First, the main findings of the data analysis are discussed in this chapter. The findings are then compared and analysed to answer the sub-research questions of the study. Next, the limitations of the study are discussed, and lastly, suggestions for future work are presented.

5.1 Analysing the results

5.1.1 Embodiment

It is evident that embodiment has scored the lowest out of all elements over the 22 meetings that were held with the system. A reason for this can be provided from the textual analysis. The participants mentioned that the point clouds did not always appear, and the standard, blue avatar from the Connec2 platform would then be showed to the participants. Or when the point cloud did appear, the participants expressed that it was of low resolution quality or that the shape of the point cloud was distorted.

Moreover, the score for the embodiment element also decreased from the first to the second meeting. The results from NB1 and SB1 also show that over a period of four meetings, the score for the embodiment element stays relatively consistent. Yet, it does not only continuously decrease. For the fourth meeting of this pair, the score of the element increased slightly. Perhaps the participants knew by the fourth meeting how they could restart the system to be able to see each other's point clouds. By diving into the answers that the pair provides for the fourth meeting to the open questions, it was found that subjects NB1 and SB1 stated the following:

- *"On a technical level the only one problem that I face every time is with the depth camera. Image from my mate always sucks, to the point that I'm glad when it completely stops working and I just see the blue avatar. At least I can see his hands clearer, it's nicer for the gestures."*
- *"The point cloud of my partner did not show up."*

So, the higher score of embodiment for this meeting cannot be explained by the point cloud appearing and functioning. What is visible from Fig. 23a and Fig. 23b is that the participants rated the QoS of the system higher. So, the higher score for embodiment could be linked to the higher score for QoS.

5.1.2 QoS

Besides embodiment, the QoS also scored very low, especially for the second meeting of the participants. The QoS decreased the most out of all of the elements that have been studied. A reason for this decrease could be that the first meeting for all of the participants was guided. They asked for help whenever it was necessary through MS Teams and would then be helped with the problems they were facing through a video call. This guidance was not provided during the second meeting of the participants. So, when facing technical issues, the participants had to solve those by themselves.

Moreover, the second meeting was only done by six participants. So, the results cannot fully be generalized since there is a chance that for those specific meetings (only three), the participants faced more technical hick-ups than usual. Fig. 18 shows that the participants from the locations Groningen and Soesterberg gave the lowest QoS scores. For the second meeting, four participants (out of six) were from Soesterberg and one was from Groningen. So, it is possible that factors from the office locations influenced the rating of the participants.

Besides location, the type of meeting could have had an effect on the low QoS score as well. The second meeting was either an update meeting, presentation or brainstorm meeting. Both update and brainstorm meetings, overall, scored the lowest for the QoS. It could have been possible that the participants tried to make use of the interactive capabilities of the system, which possibly let them down due to technical hick-ups. Examining the textual data for the second meeting shows a summary of the experience by a participant:

- *"Set-up is still a complicated process but after 15 minutes we managed to do it. Shortly after the other had received the document in the metaverse on the screen, the image suddenly went black for me and I was shown a loading bar. At that moment we were already 20 to 25 minutes further and we gave up for the sake of time. We had important things to discuss."*

5.1.3 Usability

For usability, the most interesting finding can be seen in Fig. 23a and Fig. 23b. The patterns for usability of this pair are contradicting. Subject NB1's score for usability decreased as the meetings progressed while for their partner, subject SB1, the score for usability increased as the meetings progressed. However, from both figures it is also evident that the score for the QoS element for the first meeting was close to the score of the usability element for both participants. This is also the case for the fourth meeting. So, even though the correlogram in Fig. 21 shows that there is a very weak correlation between QoS and usability, this could in the future with more data be proven otherwise. Fig. 22a does show a strong correlation between the two elements for the first meeting of all participants.

Moreover, by analysing the textual data on usability and comparing it to the data for the technical issues, it is apparent that the word clouds are similar to each other. It is possible that while usability regards the feelings of the participants, i.e. they felt frustrated, the QoS (technical issues) data shows why they were frustrated. This data also suggests that the low scores for usability, QoS and embodiment demotivated the participants to have further meetings with the social VR system. The participants mentioned that the system was 'cumbersome' to use and that going to the office to use the system was not practical.

5.1.4 QoE

For the analysis of the QoE metric, it is important to consider the indicators that this metric was made up of in this research. The results for this metric were calculated by aggregating the results of embodiment, usability, QoI and QoC. While QoC scored the highest over all of the meetings, embodiment scored the lowest. So, these two indicators largely negate each other. This is why the score for QoE lies in between those of usability and QoI. QoE also scored the lowest at Soesterberg. However, overall, the score for the metric is quite consistent through all of the four meetings of subjects NB1 and SB1, as can be seen in Fig. 23a and Fig. 23b.

5.1.5 QoI

As for QoI, the scores from meeting one to meeting two, shown in Fig. 17, barely decreased. The scores also stayed relatively consistent throughout the four meetings of subjects NB1 and SB1. However, the word cloud in Fig. 27a shows that there were problems with the interactive capabilities of the VR environment, such as screen-sharing. Yet, even though there were some technical hick-ups, these issues did not lead to significantly lower scores for this metric. This shows that the participants did not only perceive the quality of interaction as direct interaction with the environment, but also as indirect interaction with the environment, which to them seemed real and in which they experienced their actions to be real.

This is also clear from the following contradicting statements expressed by participants:

- *"Sharing things was very hard."*
- *"It can be more interactive than a Teams meeting."*

5.1.6 Immersion

Feelings of immersion did not change much from the first to the second meeting and also remained relatively consistent throughout the four meetings of subjects NB1 and SB1. However, it is notable that immersion did not receive the highest score of all the elements, even though the word cloud in Fig. 25a shows that in most meetings, participants felt very immersed in the VR environment. They described their involvement in the VR environment as real and natural. However, feelings of immersion could have been negatively impacted by embodiment. The correlogram in Fig. 21 shows that there is a strong negative correlation between immersion and embodiment. This means that the stronger the feelings for embodiment were (in the current context with the point clouds), the weaker the feelings for immersion were. So, with the current system, embodiment impacted feelings of immersion negatively.

5.1.7 Meeting engagement

The scores for meeting engagement were calculated by aggregating the scores of immersion and social presence. Therefore, the results for engagement are in between these two indicators. Fig. 18 shows that scores of engagement were highest for participants from Oude Waalsdorperweg and New Babylon. However, if the results for Oude Waalsdorperweg are filtered out, due to limited data from only one participant from this location, New Babylon is the highest-scoring location for engagement from participants. The scores for QoS and QoE were also among the highest for this location. These scores might be higher because the network connectivity at New Babylon was the best out of all the locations due to a wireless access point being installed in one of the physical rooms where the system was placed.

Moreover, Fig. 20 shows that meeting engagement was also scored the highest by participants who had brainstorm meetings. However, participants also gave the lowest scores to the QoS of brainstorm meetings. As there were only three brainstorm meetings held with the system, the results are not generalizable. The engagement scores for update meetings and informal meetings were also quite similar.

5.1.8 Social presence

Fig. 17 shows that the score for social presence was somewhat higher than that of immersion for the first meeting. So, the participants felt more present with the other in the VR environment than feeling immersed. This is a sensible result as it was the goal to have a meeting together with another colleague. However, for the second meeting the score is almost similar to that of immersion. So, the score decreased more than the score of immersion did from the first to the second meeting. However, the scores for social presence did stay relatively consistent throughout all four meetings of subjects NB1 and SB1, which can be seen in Fig. 23a and Fig. 23b. The word clouds in Fig. 25a and Fig. 25b also indicate that the feeling of being together was strong since the word 'together' was mentioned several times and is present in both word clouds.

The following statements by two different participants from New Babylon and Soesterberg also indicate that feelings of social presence might have been higher if the quality of the point clouds was improved and the participants felt a better sense of embodiment:

- *"Interactions and integration with the environment was very good, but this time my colleague's avatar was showing a blue screen with an OBS logo and a scattered cloud of blue dots, totally nonsensical. We were forced to have the meeting that way. I could still feel connected to his presence because he was moving and interacting, he was "there", but in a different form. But that was annoying."*
- *"The potential to have a sense of 'nearness' of togetherness is still bigger than with conventional (video)conference meetings."*

5.1.9 QoC

Lastly, the QoC scored the highest out of all the elements. So, QoC was rated to be the best element on average in all the meetings. Fig. 17 shows that QoC scored highest for both the first and the second meeting. Also the scores over the four meetings from subjects NB1 and SB1 in Fig. 23a and Fig. 23b show that QoC is an element that remains relatively consistent, even when other elements are not consistent over the four meetings.

It is likely that QoC scored the highest because even when the point clouds did not appear, the participants were still able to communicate with each other. This also makes sense since the meeting would otherwise not take place at all using the social VR system. It is also likely that the scores for immersion and social presence were not affected too much by the quality of service because of the high score for QoC. By conversing with each other, the participants still felt the sense of being together in the same environment. This is clear from all of the previous findings listed above. The participants felt the communication between each other in several different ways, not only through the audio, especially when there were no interruptions with the point clouds. However, even when the standard, blue avatar was shown, nonverbal communication was still conveyed and received by the participants. This is highlighted by the following statements of two different participants, both from New Babylon:

- *"The audio was good, this felt quite natural and helped a lot with having a "normal" conversation, eg. information exchange, turn-taking, etc. Body posture helped a little bit, but we were sitting."*
- *"It gives a little more of nonverbal interaction (I was seeing the blue avatar) and it makes you feel like you're together with someone."*

5.2 Findings for the research questions

5.2.1 The relationship of QoS, QoE and meeting engagement

By analysing the results of the data collection, the following research questions are answered:

RQ7 How can the relationship between quality of service (QoS) and quality of experience (QoE) be defined for social VR business meetings?

RQ8 How can the relationship between quality of experience and meeting engagement be defined for social VR business meetings?

For RQ7 it is important to analyse the correlogram in Fig. 21. It has already been mentioned that this figure shows that there is no significant linear relationship between QoS and QoE based on all of the 22 meetings held with the system. However, when the correlogram is split into two correlograms based on all the first meetings and all the second meetings, it is noticeable that the correlations for QoS and QoE negate each other. This is visualized in Fig. 22a and Fig. 22b. Essentially, there are only two more meetings (of subjects NB1 and SB1) that have not been taken into account with these last two correlograms. So, the assumption that there is no relationship between QoE and QoS cannot be made. Yet, the first correlogram does contain much more data and based on that, it is likely that the relationship between QoE and QoS is not strong. The correlograms for subjects NB1 and SB1, in Fig. 24a and Fig. 24b also show that the relationship between QoE and QoS is not strong. However, there is also a possibility that the relationship between the two elements may change over time, as more and more meetings are held with the system by the same participants. A very interesting finding in Fig. 23a and Fig. 23b shows that for the fourth meeting, the scores of QoE and QoS were very close to each other for both participants, even though that was not the case in previous meetings.

Besides QoS, there is also no significant linear relationship between QoE and meeting engagement based on the correlogram in Fig. 21. However, here Fig. 22a and Fig. 22b also show that the results of the first and second meeting possibly negate each other. This is also shown in the correlograms of subjects NB1 and SB1 in Fig. 24a and Fig. 24b. The patterns for QoE and engagement are also quite similar over the four meetings that subjects NB1 and SB1 have done. This can be seen in Fig. 23a and Fig. 23b. The patterns between the subjects is different. However, in both

figures, it is noticeable that the pattern of QoE is similar to that of engagement for the participants themselves. So, for this relationship, it is also possible that there may be a weak or moderate relationship. However, it also seems that the type of relationship between the two is subjective to the participants.

Moreover, Fig. 18 and Fig. 19 show that the relationship between QoS and QoE was different for the participants from different locations. So, the location factors could have had an effect on the relationship between these elements as well. For the relationship between QoE and meeting engagement, not much can be stated based on this figure since no trend can be found. The only conclusion for the locations that can be made is that the scores from the participants at New Babylon are higher overall for all of these elements, if you do not include Groningen and Oude Waalsdorperweg due to the low number of participants. This is possible due to a wireless access point being placed in one of the rooms where the system was placed at this office location. The network connectivity for this system was thus improved.

Lastly, Fig. 20 shows whether there are trends between QoE and QoS, and QoE and meeting engagement for the different types of meetings. Informal and update meetings were held the most, so these will be analysed. For these meetings, it is apparent that the QoS does affect the QoE for the two types of meetings. When the QoS is lower, QoE also decreases. Moreover, for meeting engagement it also seems that when QoE is higher, meeting engagement is higher.

Overall, these results show that there may be relationships between QoS and QoE, and QoE and meeting engagement. However, this depends on several factors, such as the number of meetings that participants have with the system, the individual characteristics of the participants themselves, the office locations from where the meetings were held and the type of meetings that the participants had with each other.

5.2.2 The relationship of embodiment, QoI, QoC and meeting engagement

The following research questions are answered by analysing the results of the data collection and the literature study:

RQ9 How can the relationship between embodiment (using the photorealistic user representation) and meeting engagement be defined for social VR business meetings?

RQ10 How can the relationship between the quality of interaction (QoI) and the quality of communication (QoC), and meeting engagement be defined for social VR business meetings?

To identify and define the relationship between embodiment and meeting engagement, it is again necessary to look at the correlogram in Fig. 21. This shows that there is a strong negative linear relationship between the two elements. There are also no other contradicting results. This finding is also explained by analysing the word cloud in Fig. 26a, which describes the participants' experiences with the point cloud. Since the point clouds either did not appear, or were of low quality, meeting engagement was negatively impacted. If the embodiment was high with these low quality point clouds, the meeting engagement would not have been high. Since embodiment was also part of QoE, this impacted the results of QoE negatively as well. Thus, the relationships for both QoE and embodiment with meeting engagement might be different in the future as the user representation as point clouds constantly improves.

Furthermore, the correlogram shows that the relationship between QoI and meeting engagement is moderately strong. However, the correlograms in Fig. 22a and Fig. 22b show that there either is not, or is a strong relationship between the two for the first meeting and the second meeting. The pattern for QoI in Fig. 23a and Fig. 23b is also very contradicting for subjects NB1 and SB1. So, these figures all show that the results are inconclusive for the relationship between QoI and meeting engagement. The relationship between QoI and meeting engagement may also be very subjective for each participant. Additionally, it is evident that there is a strong linear relationship between QoC and meeting engagement.

5.2.3 The relationship of social presence, immersion and meeting engagement

The following research questions are answered by analysing the data and assessing the literature findings on social presence, immersion and meeting engagement:

RQ2 How can meeting engagement in social VR business meetings be captured and evaluated?

RQ11 How can the relationship for social presence and immersion be defined for social VR business meetings?

From the literature study, it was concluded that meeting engagement in social VR meetings could be evaluated using both immersion and social presence. There are high correlations between these two elements, shown in Fig. 21. However, the correlogram also shows that QoC was an important element as well. QoC highly correlated with immersion and social presence, which means that this element can be used to evaluate meeting engagement as well. This is also expressed by the participants who mentioned that they felt engaged, immersed, present together and could communicate well with one another. For RQ11 it can also be concluded that there is a strong linear relationship for social presence and immersion. Fig. 23a and Fig. 23b also show that both of these elements were also consistent throughout the four meetings of subjects NB1 and SB1. The patterns are also very similar. Yet, Fig. 16 shows that the scores for social presence are higher than those of immersion. So, the participants noticed the feeling of being together in the environment more than that they felt immersed by the environment.

5.2.4 Findings on the first timer effect

The following question on the first timer effect was posed and is answered using the results of the data collection and the literature study:

RQ12 Do feelings of engagement diminish over a longer period of time using social VR for business meetings when compared to using it for the first time?

The literature study showed that it may be the case for VR systems that a first timer effect occurs. Fig. 17 shows that all scores for the elements in the current study were lower for the second meeting. In particular, the QoS decreased to a greater extent compared to the other elements. It should be noted that all twelve participants had the first meeting and that only six participants attended the second meeting. Four of these participants were also from Soesterberg, which may have impacted the results of the second meeting due to location factors. No further statistical analysis was performed on these results due to the small sample size. It can be assumed that the results of the first meeting are, in general, higher than those of the second meeting. However, the results over the four meetings from subjects NB1 and SB1, shown in Fig. 23a and Fig. 23b, show that the first timer effect is not generalizable for all elements. The first timer effect is, especially, not displayed for feelings of immersion, social presence or quality of communication. This also leads to no first timer effect for meeting engagement. The scores for these participants either decrease or increase from meeting to meeting. So, it cannot be concluded that the scores continuously decrease and therefore, the previous meetings have an effect on subsequent meetings. The scores in Fig. 17 also showed only a minimal decrease for these elements for the second meeting over all participants.

Moreover, for the rest of the elements it can also not be concluded whether there is a first timer effect. The only element that decreases significantly for the second meeting is QoS. However, Fig. 23a and Fig. 23b show that the QoS can increase again, even though it decreased for a previous meeting. Lastly, as previously mentioned, the QoS could have been low because the location factors impacted this result.

5.2.5 Comparing usability and user experience findings to existing systems

The data analysis and the findings from the literature study on the usability and user experience are summarized below. The following questions were the last two sub-questions that were posed:

RQ13 How usable is the VR system to use for a longer period of time compared to other VR systems?

RQ14 What do users experience when working with a social VR system for a longer period of time compared to other VR systems?

The usability score was close to zero, which is neutral, for the first meeting and decreased to **-0.69** in the second meeting, which can be seen in Fig. 17. These scores indicate that the usability of the system was not rated well by the participants. The word cloud in Fig. 26b also indicates why the usability was not good enough. The system was cumbersome to use and the participants were in the end frustrated because of how many times they had to restart the system. However, because the experience of having a VR meeting together with another participant was novel and thus exciting and fun for the participants, the usability score was not too low. Moreover, the patterns for usability are also very different, and contradict with each other for the four meetings of subjects NB1 and SB1. Whereas for the first participant the usability score decreased, the usability score for the other participant increased. This can be seen in Fig. 23a and Fig. 23b. These results indicate that usability is subjective. The linear relationship between QoE and usability is also very strong, which can be seen in Fig. 21. So, it is clear that because of a low score for usability and embodiment, the score for the QoE is also low.

These results are not comparable to most other social VR meeting systems due to the photorealistic avatars. The system is comparable to other research systems, however these are only tested for a short amount of time. It is noticeable that other studies, also previous studies by TNO colleagues, have shown better usability findings, even though it was only for one meeting. However, that is because the system was usually run locally, and this was the first time it was run on the cloud, and by another company.

For RQ14 the verdict is that the system is comparable to other VR systems but the opinions of colleagues are all subjective. This is visible from the results of NB1 and SB1. This could be due to location factors of the systems. Employees stay engaged consistently throughout all meetings, because feelings of immersion and social presence stay relatively consistent. The word clouds in Fig. 25a and Fig. 25b also highlight the positive experiences that the users had with the system. These were all focused on the user experience itself. Moreover, the literature review showed that users can sometimes experience a first timer effect, and can also sometimes have some issues with using the systems for a longer period of time, especially for business meetings. However, the current study shows that users' experience regarding QoC, social presence and immersion stays relatively consistent. Moreover, they do see the potential in the system when all of the technical issues have been resolved regarding the point clouds. The results for the second meeting were bad due to a low QoS score, which can be explained by the location factors from which the participants held those second meetings. So, the current study shows that there is no first timer effect regarding user experience.

5.3 Limitations

The most important limitations of the current study, that have not yet been mentioned in the previous section, are presented below. Some insights and possible solutions on how their influence could have been limited are also presented.

5.3.1 Location-based limitations

Because the research was conducted in a real-world setting, it is naturally almost impossible to control all influencing factors. Some of these factors had an influence due to the different office locations and the differing conditions of the physical rooms in which the participants had their meetings. So, there were many location-based factors that impacted the experience of the users. At some locations, two systems were placed in the same room while at others a system had its own physical room. Furthermore, the physical rooms were also not easy to find, even though logos were placed on the doors. Some rooms at some locations were too hard to find. The participants lost quite some valuable time because of this before they could begin the first meeting.

Moreover, some participants reported that the equipment was misplaced or not charged because a previous participant had forgotten to put the charger in the HMD, which led to more waiting time for subsequent participants before they could use the system.

Next to that, the network connectivity at the locations may have affected the results of the QoS, usability, QoI and embodiment. It is also possible that the scores for immersion, social presence and QoC would have been higher if there were no hick-ups with the network connectivity. Sometimes, network connectivity on the HMDs would be lost as well, which was unexpected. Thus, participants had to reconnect the HMDs to the network. However, they did not know that the HMDs had to be connected to the Eduroam network instead of the TNO network. This information had been left out of the instructional guide because it was not expected that the connection would be lost. This all led to a loss of time as well for participants, leaving them unable to have a meeting or only having a short amount of time to meet. So, it would have been best to first test the network connectivity at all locations. It was already assumed that there would be different location-based factors influencing the experiences. However, it was not expected to be a big problem. With some locations being hard to reach, there was not enough time planned to test the network connectivity at the different locations beforehand.

5.3.2 Limitations regarding the hardware and software

The research team for the current project was small at TNO since there was a small budget for the research. This was not much of a problem and was already known beforehand but it led to some delays when issues occurred, such as falling out team members due to COVID-19. Many team members from both TNO and Connec2 were out of the running for a while because of the virus. Therefore, the completion of the system took much more time than expected and, in turn, there was much less time to gather the data even though the data collection period was also extended. There were sometimes also communication problems between both parties, which may have been caused by working remotely because of COVID-19 restrictions.

Some big limitations posed by Connec2 were related to the user limit of the platform and the absence of system data. It was mentioned that the platform would not be able to support more than three users simultaneously because the quality of the point clouds would then be low and the platform would respond slower to the actions of the users. However, it was not mentioned by the party that there was also a predetermined constraint of three users. So, when a team of four people wanted to use the system, this was not possible at all. Besides that, Connec2 also mentioned at a later moment that they would not be able to provide system (objective QoS) data. Because of the timing, it was difficult for TNO to implement their own methods for gathering QoS data.

Moreover, even though the participants did not set up the systems, they still experienced setting-up problems. This could have been a result of the hardware that was used. The laptops and PC's of the systems were being reused so there was already much software on them. So, this hardware could have caused some problems for participants. Moreover, navigation usability could also have had an effect on the current study as a number of different platforms and devices were needed to hold a meeting with the social VR system. There were many navigation steps; first navigating a laptop or PC to find the Uplink platform, then navigating the Uplink platform, then using the HMD and navigating to the correct application on it and eventually navigating in the VR environment to find a meeting room. In addition, the experience of the participants could also have been hindered because of the hassle of logging into the system with difficult login credentials. The email addresses were long and not simple because of limitations posed by TNO's IT helpdesk. Multiple participants made mistakes when typing in their login credentials. These steps may have had an influence on some of the elements that were studied.

Furthermore, the Connec2 Uplink platform itself could also have had an impact on some results, such as usability. There were some hick-ups with the platform that hindered the participant's experience. Sometimes the platform needed to be updated and there was no easy way for the participants to do that themselves. Moreover, participants mentioned that uploading a presentation was not easy to do either. Besides that, the use of the screen-sharing functionality

could also have had an impact on the participant's experience. This is somewhat evident from the results because for brainstorm meetings, the interactive capabilities of the environment may have been used and this decreased the QoS, QoI and usability for this type of meeting. However, in the end, it did not have a huge effect on social presence and immersion.

5.3.3 Limitations regarding the participants

It was best for this system to be used for meetings in which there are more than four people involved. This was also the aim for the system. However, this could not be realized due to the hardware and software limitations. Yet, it became apparent that teams at TNO are usually on the larger side as well. There were many bigger teams who responded to the initial emails for taking part in the study. There were almost no teams of three or four people who responded. It was either pairs who wanted to test the system, or groups of at least five people, many times even more. It was also apparent that most employees who wanted to participate, and eventually also the participants, did not read the full information regarding the research. So, they sometimes did not know that they had to do four meetings with the system.

Moreover, it was apparently cumbersome for the participants to come to the office to use the social VR system. This was not only expressed in the data, but also multiple times over email by many participants. This also explains why so many participants dropped out of the study or did not proceed with further meetings. The participants also did not know in the beginning that there was an expiration date for the research. Communicating an end date for the study from the beginning to the participants might have motivated them to try to do one meeting every week or every other week. The participants also did not know what the scope of the current study was. They, therefore, might have thought that they had more time or that it was not urgent to do all of the meetings.

Another important limitation that may have resulted in lower scores for the second meeting is that the participants were guided through the first meeting via MS Teams, while they were not guided through the second meeting. They could call if there were any technical problems but they only did so when there was a major problem. Moreover, the guide that was created to help the participants may also have been slightly too long. Most participants did not read it at all and others skimmed over it. Lastly, a big problem during the data collection phase was that all problems had to be solved remotely. Participants needed much more support than expected. Most of the participants also preferred being helped at the office, not remotely.

Lastly, the amount of planning that the participants had to do themselves may also have been a limitation in the current study. The number of meetings that they had to plan themselves may have been too many within the timeframe of this study. Also, the participants did not view the planning of meetings as a priority. This could have been streamlined in a better way in the research design, such as planning the meetings for the participants beforehand. They could then come at the mentioned pre-specified dates based on their availability. This would also have averted the use of the booking system, which not all of the participants used correctly.

5.3.4 Limitations regarding the study design

There were many limitations due to the study design of this research. It was, for example, suggested in the literature that the first meeting of the participants should not have been included in the data analysis of the study as the first meeting was needed to become familiar with the system. However, to research the first timer effect, it seemed sensible to include this data, especially with the current small dataset. In the end, a five to ten-minute meeting before the first long meeting should have been done by the participants as well to familiarize them with the system. Moreover, because of different types of meetings that can be done as business meetings, it may have been better to categorize the type of meeting that can be conducted with the system more than was done in the current study.

The meetings also needed more structure. Participants did not achieve much using the system on their own and most of the time only used it to converse with each other or be in the VR environment together. However, this means that they sometimes did not hold business meetings. So,

it may have been better to provide them with instructions for the meetings and let all pairs, for example, present a given slide set. Moreover, it should also have been taken into account beforehand that, due to the study taking quite some time for participants, a lot of participants would drop out. So, the study needed to take less time for participants so that more conclusive results on different aspects could be given.

In general, the research design of the current study was very broad. Many elements were examined, which led to not enough elements being thoroughly investigated. An example is the QoS element. Only one question is eventually used to measure this element in the questionnaire. This means that the score for this element is unreliable. In addition, the question was added later on to the questionnaire because it was communicated rather late by Connec2 that they did not collect system data. Lastly, regarding the body tracking software, a better plan needed to be created to ensure that analysing the data could be done faster. There were now many files of more than half an hour long and there was no pause, or play, or skip ahead function when watching these videos.

All in all, by reviewing these limitations, it is clear that the social VR communication system was not ready to move from a controlled setting to a full real-world setting. There are still many more aspects that need to be considered for employees to conduct real business meetings. The current study did not take all of these aspects into account. Most aspects of getting the system to work, etc. took much more time than expected, which most employees at TNO do not have. The busy lives of the employees was therefore, not taken into account enough. Therefore, these limitations resulted in a smaller sample size which prohibited the researcher from conducting hypothesis tests. The many limitations of the current study also hint to the assumption that the current study may have been ill-timed. Most of these limitations, especially regarding the technical issues of the system, should have been resolved before the study took place.

5.4 Future work

There are many future work directions that can be derived from the current study. These directions have been categorized below and the future work studies that could be conducted are presented.

5.4.1 Further analysing the diverse elements and the first timer effect

The current study only focused on analysing linear relationships between the elements of the study. However, future studies can also investigate monotonic or non-linear relationships. Moreover, even though there is no first timer effect found, based on the limited data that was available, more studies with more participants need to be done that focus on this aspect. In that case, it is also important that the relationship between usability, QoI and QoS is explored more since these can affect the results of the first timer effect at a later stage in the process. A future study should focus on teams of at least four people instead of pairs. Furthermore, it can also be investigated whether different group-dynamics lead to different results regarding the first timer effect and the scores for the different elements of this research. Groups could then be characterized first based on, for example, a personality scale. Overall, the current study needs to be conducted again with bigger project teams because it is evident that participants feel that the system is more valuable to use with bigger project teams.

5.4.2 Other (objective) methods to investigate engagement

One of the aspects that would have been interesting to analyse during this study, especially in the context of meetings, was turn-taking. Investigating turn-taking in future studies will provide more details on whether the meetings done with the social VR system are indeed similar to f2f communication or not. Turn-taking can also indicate the level of engagement that meeting participants feel. A future study could focus on an A/B test in which meeting participants have to conduct a meeting with each other in a setting using MS teams and in another setting in which they use the social VR system. Moreover, eye tracking software could be used for investigating engagement as well. Heatmaps could be generated using the data to assess where people look most often in the VR environment. For this to work, built-in eye tracking software is necessary

in the VR headset. It could then be investigated whether people are distracted during the meetings and by which elements they get distracted. However, this means that heavier HMDs would need to be used, unless there are smaller HMDs with built-in eye tracking software. Lastly, body tracking data could still be analysed in a future study to investigate whether there are engaged or disengaged states in users when they are having a meeting in a VR environment. The software could then capture even more details of the users and the interface of the software could allow for videos that were captured to be paused, played or skipped ahead.

5.4.3 Different settings for VR meetings

Regarding the evaluation of different types of meetings in VR, it is best to provide participants with a pre-specified task, before conducting the study, that is characteristic of the type of meeting. So, one team could hold a brainstorming session on a certain topic and another team could present a given slide-set to each other. A between-subjects study could be done with different types of meetings being done at the same time. Another element that can be investigated in a future study is navigation usability. It could be researched whether users are more immersed when they have to find a meeting room themselves in a VR environment, or whether they are immersed at the same level when they are immediately placed in a meeting room once they put their HMDs on, for example. Moreover, literature shows that social VR communication systems can allow for meeting each other in hybrid meetings to be easier. People joining in remotely and people joining in from a conference room can all put on the VR headsets and then meet each other in the VR world where they can speak more easily to each other and see each other better than using videoconferencing software. It would also make it more difficult for colleagues to work on their own tasks during important meetings, meaning that the engagement of the colleagues would be higher. However, for this to work, a higher budget has to be available so that every colleague can have a VR system at home. Future studies could also evaluate whether the motivation of using a social VR communication system is higher when participants have the system at home, compared to when it is at the office. The participants of the current study mentioned this many times. So, the next step is to test whether this assumption is true.

5.4.4 Improvement and evaluation techniques for user representation

A re-evaluation of the current photorealistic avatars in the Connec2 platform needs to be done to ensure that the quality has improved compared to the current study, especially when the issue of network connectivity has been resolved. This can even be done with only one meeting. Moreover, evaluating the value of the photorealistic avatars by doing A/B tests with photorealistic and cartoonized or 'Sims'-like avatars may also be an important next step to gather more feedback on the photorealistic avatars. It would also be interesting to investigate what the effects of self-view are on embodiment, both by using a mirror for seeing your full body and by seeing your own hands in the VR environment. Hand-tracking would also allow for many more research opportunities. The use of gestures for specific tasks could be investigated. It would be possible to investigate which gesture would be most intuitive/natural for walking around in the VR environment, for example.

As for the improvement of the photorealistic avatars, AI models could be trained to create higher quality point clouds. They could erase some of the random points at the back of the point cloud or try to create a fuller picture based on the image that they receive. The point clouds of these AI-based models could then also be evaluated with A/B tests. Point clouds could also be made by first fully scanning a person using a 'point cloud scanner' and then trying to use the ZED or Kinect to get real-time images onto the scanned image. Some companies already combine scanners and point cloud technology in other markets, such as 3Shape [1], a dentistry software company. Moreover, exchanging a camera for a 360 degrees camera to see whether a user could be captured 360 degrees and whether that would lead to a fuller point cloud, could also be evaluated in a future study. Again, an A/B test could be done here then as well. Lastly, HMD removal is a big future work direction and much research is being done on it already. Once the photorealistic avatars are improved, it could be researched whether various HMD removal techniques result in an uncanny valley effect.

[1] https://www.3shape.com/

Chapter 6

Conclusion

Social VR has shown promising results for providing people with a better sense of social presence and immersion due to better transmission of body language cues in remote meetings. Videoconferencing technologies lack these elements, which make them less engaging and more tiring than f2f meetings. To further research engagement in social VR business meetings, the current study had set out to conduct a longitudinal, exploratory study in the context of an informal, real-world setting. A social VR communication system created by TNO and Connec2 was therefore used to enable colleagues of TNO to communicate with each other over long distances in The Netherlands using photorealistic avatars as user representations. A secondary goal of the study was to investigate whether a first timer effect had occurred or not with the group of participants in the current study. This is why the study was conducted over a longer period of time. Other goals for TNO were to examine the usability and the user experience of the system for real-time remote collaboration.

A mixed-methods approach was taken to answer the multiple sub-questions that were formed to help answer the main research question. Participants of this study used the system for their own small-project business meetings. Both the quantitative and qualitative data was gathered from surveys that participants had to fill in after each meeting they had done with the system. Closed questions in the survey assessed various elements that were either related to meeting engagement or quality of experience. The subjective quality of service was also assessed using the survey. Social presence and immersion were the indicators of the composite variable meeting engagement. Usability, embodiment, quality of interaction and quality of communication were the indicators of the composite variable quality of experience. Moreover, open questions were included in the survey to gather textual data, which was then qualitatively analysed for further insights and explanations.

There were twelve participants, six pairs, who used the system for a total of 22 meetings. They held their meetings with each other from four different office locations in The Netherlands. Requirements for the system and the meetings were formed as well before deploying two systems per office location. The requirements were made so that it could be ensured that the conditions of the physical rooms in which the systems were placed were the same over all locations. However, this requirement could not be fulfilled as there were a number of limiting factors from the locations, such as the network stability at each location and the spaces they had for the systems. Moreover, the participants were not distributed equally over all locations since there was a much bigger interest to participate in the study by TNO employees from two of the four locations.

After the data was collected from all of the meetings, the data was cleaned and prepared for the analysis. Since there were not many participants who filled in the surveys, hypothesis testing could not be used to answer the research questions. Therefore, data visualizations were made to provide an answer to the questions. Moreover, Pearson correlations were done on the data of all the meetings to investigate the relationships between the different elements that were assessed with the survey. These relationships were explored to gain insights on which elements are most important for creating engagement in social VR business meetings. The answers to the sub-questions can be found in Table 4. This table is an updated version of Table 2.

Table 4: Sub-questions literature review findings and data analysis strategy

Sub-question	**Findings**
RQ2: How can meeting engagement in social VR business meetings be captured and evaluated?	This can be done by analysing body language cues such as body postures and gestures. Because of the VR setting, social presence and immersion are also important elements related to engagement. This was further proven by the strong relationships between immersion and social presence in this study. However, the correlations showed that QoC is also an important element to consider when evaluating engagement since this element also has a strong linear relationship with immersion and social presence.
RQ3: What were the conditions for meeting engagement in f2f, hybrid and online meetings?	In f2f meetings engagement was highest because of factors, such as body language cues and eye gaze. These were either missing or very hard to interpret in hybrid and online meetings.
RQ4: How can the quality of experience be defined for social VR business meetings?	It can be defined as as a multidisciplinary indicator that includes all the different aspects related to the experience that a user has in VR. It also includes the user experience.
RQ5: How can the quality of service be defined for social VR business meetings?	The quality of service comprises system performance measures that can impact the quality of experience.
RQ6: How can the first timer effect be defined?	The first timer effect can be defined as a wow effect because some experience with new technology is novel to its users. It happens when people experience VR for the first time and only spend a short amount of time in it. This effect is diminished when people again use VR technologies.
RQ7: How can the relationship between quality of service (QoS) and quality of experience (QoE) be defined for social VR business meetings?	Literature has shown that QoS can have an impact on QoE. However, this is not always the case. The Pearson correlation conducted in the current study also shows that there is no strong linear correlation. However, it cannot be assumed that there is no relationship between the two at all. Some of the data suggested that there may be a relationship, but that it is not strong. In addition, the relationship between the two is also different for the office locations from which the meetings were conducted. So, location factors had an impact on the relationship between these two. Moreover, the relationship between these two variables can also change over the course of meetings.
RQ8: How can the relationship between quality of experience and meeting engagement be defined for social VR business meetings?	Since QoE comprises all of the experiences that a user has in VR, it can be assumed that engagement is related to it as well. However, the current study shows that there is no strong relationship between the two variables. It is still possible that there is a weak relationship between the two and that the relationship between the two is subjective for different individuals.

RQ9: How can the relationship between embodiment (using the photorealistic user representation) and meeting engagement be defined for social VR business meetings?	An assumption is made in the literature that photorealistic avatars enhance engagement. However, because the photorealistic avatars in the current study were of poor quality and did not appear most of the time, the score for embodiment is very low. There was also a strong negative relationship for embodiment and meeting engagement. This shows that if the score of embodiment would have been high in the current study, meeting engagement would have been low. Thus, it can be concluded that the relationship for meeting engagement and embodiment is linear and that the photorealistic avatars need to be of high quality for a better score of embodiment and for it to enhance the score of meeting engagement as well.
RQ10: How can the relationship between the quality of interaction (QoI) and the quality of communication (QoC), and meeting engagement be defined for social VR business meetings?	Literature shows that these are important elements for social VR meetings. The linear relationship between QoI and engagement in the current study is moderately strong. However, again, this relationship may be subjective for each different individual. Moreover, it is proven that QoC is strongly related to social presence and immersion. It is therefore essential for high meeting engagement.
RQ11: How can the relationship for social presence and immersion be defined for social VR business meetings?	These two elements are both needed for creating successful VR experiences. The current study has also proven that these two elements have a strong linear relationship with each other and that these elements stay relatively consistent throughout several meetings.
RQ12: Do feelings of engagement diminish over a longer period of using social VR for business meetings when compared to using it for the first time?	Based on the literature findings, this is possible if a first timer effect occurs. In the current study, there was a slight decrease in the engagement score from the first to the second meeting. However, based on the consistent results throughout all meetings for social presence, immersion and QoC, for one pair that did four meetings, it cannot be concluded that feelings of engagement diminish over a longer period of time. Unfortunately, these results are not generalizable since this is solely based on the results of one pair.
RQ13: How usable is the social VR system to use for a longer period of time compared to other VR systems?	The literature review showed that there are no commercial social VR systems yet that support photorealistic avatars. These are supported in systems in a research setting but the systems are still in initial stages where people cannot efficiently communicate with each other yet over big distances. The current study has also emphasised the difficulty of allowing users to communicate with each other over big distances including the user representation as photorealistic avatars. Overall, the study does show that the usability of the system scored very low and that there is a relationship between QoS and usability.
RQ14: What do users experience when working with a social VR system for a longer period of time compared to other VR systems?	The literature review showed that users can sometimes experience a first timer effect, and can also sometimes have some issues with using the systems for a longer period of time, especially for business meetings. However, the current study shows that users experience regarding QoC, social presence and immersion stays relatively consistent. Moreover, they do see the potential in the system when all of the technical issues have been resolved regarding the point clouds. The results for the second meeting were bad due to a low QoS score, which can be explained by location factors. So, for the current study, the first timer effect does not have an effect on the user experience.

All in all, these findings are used to answer the main research question of this study, which was:

RQ1 *What elements lead to meeting engagement in a social VR environment in which business meetings take place over a longer timeframe?*

The elements have been ranked and linked as follows:

1. **Quality of communication, social presence and immersion**
   The study has shown that quality of communication is the most important element for creating engaging social VR meeting experiences. Simply communicating with another person in a social VR meeting leads to engagement, especially when there are minimal audio delays and nonverbal cues can be transmitted. Even when the photorealistic avatars were not shown, and the users had to look at each other's standard avatar, some nonverbal cues were still transmitted, such as gestures. Undoubtedly, social presence and immersion are also crucial elements for social VR experiences. The relationships between quality of communication, social presence and immersion are therefore high. In the current study, social presence scored higher than immersion. This is a sensible result as the goal for this study was to have a meeting together with another colleague. Moreover, all these elements also stay relatively consistent throughout multiple meetings that participants of this study had with each other. So, the first timer effect cannot be identified for these elements in the current study.

2. **Embodiment** Embodiment had a strong negative linear relationship with the quality of communication, social presence and immersion. These relationships show that embodiment is more related to meeting engagement than usability, quality of interaction and quality of service. While the point clouds did not show because of the bad network connectivity, the bad connectivity was not to blame for the visual appearance of the point clouds. It is therefore necessary that photorealistic avatars are used for social VR meetings once they are improved. The strong negative relationships may turn to positive strong relationships once users feel comfortable with the photorealistic avatars and there are no technical issues. Moreover, feelings of embodiment could also be enhanced with self-view, which was mentioned as a requirement in section 3.3.2 as well.

3. **Quality of interaction, usability and quality of service** Lastly, the quality of interaction, usability and the quality of service of the system are important elements to consider as well. These elements did not stay consistent for the pair that did four meetings. The results for these elements are also very different per participant. This shows that these elements are very subjective to individual participants and become more important for engagement when more meetings are done with the system. So, these elements may result in a first timer effect when the scores are low for a longer period of time. However, these elements are more related to the experience that a user has with the system and not necessarily to meeting engagement.

The quality of experience has been left out of this ranking purposely as it comprises all of the experiences that a user has in VR. So, it is made up of all of the elements that have been studied. These elements have been identified in the current study and have been ranked on their importance regarding lasting meeting engagement. However, there were multiple limitations in the current study that did not allow for more concrete analyses of the elements. This was due to the small sample size of the study and the lack of data on subsequent meetings by pairs of the study. In addition, the technical issues causing the disappearance of the photorealistic avatars in combination with the low quality of the photorealistic avatars suggests that the current study may have been ill-timed. These issues first need to be resolved before future work can be conducted. Eventually, future studies need to investigate these elements further to create the first guidelines for meeting engagement in social VR business meetings.

Bibliography

[1] ALLEN, J. A., BECK, T., SCOTT, C. W., AND ROGELBERG, S. G. Understanding workplace meetings: A qualitative taxonomy of meeting purposes. *Management Research Review* (2014).

[2] ARHIPPAINEN, L., PAKANEN, M., HICKEY, S., AND MATTILA, P. User experiences of 3d virtual learning environment. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (2011), pp. 222–227.

[3] BAILENSON, J. N. Nonverbal overload: A theoretical argument for the causes of zoom fatigue. *Technology, Mind, and Behavior 2*, 1 (2021).

[4] BAILEY, J. O., BAILENSON, J. N., AND CASASANTO, D. When does virtual embodiment change our minds? *Presence 25*, 3 (2016), 222–233.

[5] BIOCCA, F., AND HARMS, C. Defining and measuring social presence: Contribution to the networked minds theory and measure. *Proceedings of PRESENCE 2002* (2002), 1–36.

[6] BIOCCA, F., HARMS, C., AND GREGG, J. 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* (2001), pp. 1–9.

[7] BROOKE, J., ET AL. Sus-a quick and dirty usability scale. *Usability evaluation in industry 189*, 194 (1996), 4–7.

[8] BRUBAKER, J. R., VENOLIA, G., AND TANG, J. C. Focusing on shared experiences: moving beyond the camera in video communication. In *Proceedings of the Designing Interactive Systems Conference* (2012), pp. 96–105.

[9] CAMPBELL, A. G., STAIKOPOULOS, A., HOLZ, T., HARLICK, M., AND COSGROVE, J. Bank of ireland tests meetings in virtual reality. In *XR Case Studies*. Springer, 2021, pp. 123–128.

[10] CAO, H., LEE, C.-J., IQBAL, S., CZERWINSKI, M., WONG, P. N., RINTEL, S., HECHT, B., TEEVAN, J., AND YANG, L. Large scale analysis of multitasking behavior during remote meetings. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (2021), pp. 1–13.

[11] CHEN, Y., WU, K., AND ZHANG, Q. From qos to qoe: A tutorial on video quality assessment. *IEEE Communications Surveys & Tutorials 17*, 2 (2014), 1126–1165.

[12] CHEN, Y.-T., HSU, C.-H., CHUNG, C.-H., WANG, Y.-S., AND BABU, S. V. ivrnote: Design, creation and evaluation of an interactive note-taking interface for study and reflection in vr learning environments. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2019), IEEE, pp. 172–180.

[13] CHENG, A., YANG, L., AND ANDERSEN, E. Teaching language and culture with a virtual reality game. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (2017), pp. 541–549.

[14] COBB, S. V., NICHOLS, S., RAMSEY, A., AND WILSON, J. R. Virtual reality-induced symptoms and effects (vrise). *Presence: Teleoperators & Virtual Environments 8*, 2 (1999), 169–186.

[15] CROES, E. A., ANTHEUNIS, M. L., SCHOUTEN, A. P., AND KRAHMER, E. J. Social attraction in video-mediated communication: The role of nonverbal affiliative behavior. *Journal of social and personal relationships 36*, 4 (2019), 1210–1232.

[16] De Simone, F., Li, J., Debarba, H. G., El Ali, A., Gunkel, S. N., and Cesar, P. Watching videos together in social virtual reality: An experimental study on user's qoe. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2019), IEEE, pp. 890–891.

[17] Denstadli, J. M., Julsrud, T. E., and Hjorthol, R. J. Videoconferencing as a mode of communication: A comparative study of the use of videoconferencing and face-to-face meetings. *Journal of Business and Technical Communication 26*, 1 (2012), 65–91.

[18] Dijkstra-Soudarissanane, S., Gunkel, S. N., and Reinders, V. Virtual visits: Life-size immersive communication. In *Proceedings of the 13th ACM Multimedia Systems Conference* (2022).

[19] Eriksson, T. Failure and success in using mozilla hubs for online teaching in a movie production course. In *2021 7th International Conference of the Immersive Learning Research Network (iLRN)* (2021), IEEE, pp. 1–8.

[20] Fauville, G., Luo, M., Muller Queiroz, A. C., Bailenson, J. N., and Hancock, J. Nonverbal mechanisms predict zoom fatigue and explain why women experience higher levels than men. *Available at SSRN 3820035* (2021).

[21] Fauville, G., Luo, M., Muller Queiroz, A. C., Bailenson, J. N., and Hancock, J. Zoom exhaustion & fatigue scale. *Available at SSRN 3786329* (2021).

[22] Frank, M., Tofighi, G., Gu, H., and Fruchter, R. Engagement detection in meetings. *arXiv preprint arXiv:1608.08711* (2016).

[23] Garau, M., Slater, M., Bee, S., and Sasse, M. A. The impact of eye gaze on communication using humanoid avatars. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (2001), pp. 309–316.

[24] Goodwin, N. C. Functionality and usability. *Communications of the ACM 30*, 3 (1987), 229–233.

[25] Gunkel, S., Stokking, H., Prins, M., Niamut, O., Siahaan, E., and Cesar, P. 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* (2018), pp. 233–238.

[26] Gunkel, S. N., Dohmen, M. D., Stokking, H., and Niamut, O. 360-degree photorealistic vr conferencing. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2019), IEEE, pp. 946–947.

[27] Gunkel, S. N., Hindriks, R., Assal, K. M. E., Stokking, H. M., Dijkstra-Soudarissanane, S., Haar, F. t., and Niamut, O. Vrcomm: an end-to-end web system for real-time photorealistic social vr communication. In *Proceedings of the 12th ACM Multimedia Systems Conference* (2021), pp. 65–79.

[28] Gunkel, S. N., Stokking, H. M., Prins, M. J., van der Stap, N., Haar, F. B. t., and Niamut, O. A. Virtual reality conferencing: Multi-user immersive vr experiences on the web. In *Proceedings of the 9th ACM Multimedia Systems Conference* (2018), pp. 498–501.

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

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

[31] Jonas, M., Said, S., Yu, D., Aiello, C., Furlo, N., and Zytko, D. Towards a taxonomy of social vr application design. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (2019), pp. 437–444.

[32] Kuzminykh, A., and Rintel, S. Classification of functional attention in video meetings. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), pp. 1–13.

[33] KUZMINYKH, A., AND RINTEL, S. Low engagement as a deliberate practice of remote participants in video meetings. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), pp. 1–9.

[34] LAI, C., CARLETTA, J., AND RENALS, S. Modelling participant affect in meetings with turn-taking features. In *Proc. Workshop of Affective Social Speech Signals* (2013), Citeseer.

[35] LATOSCHIK, M. E., ROTH, D., GALL, D., ACHENBACH, J., WALTEMATE, T., AND BOTSCH, M. The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology* (2017), pp. 1–10.

[36] LI, J., KONG, Y., RÖGGLA, T., DE SIMONE, F., ANANTHANARAYAN, S., DE RIDDER, H., EL ALI, A., AND CESAR, P. Measuring and understanding photo sharing experiences in social virtual reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), pp. 1–14.

[37] LIU, Y., CASTRONOVO, F., MESSNER, J., AND LEICHT, R. Evaluating the impact of virtual reality on design review meetings. *Journal of Computing in Civil Engineering 34*, 1 (2020), 04019045.

[38] MANUEL, M., DONGRE, P., ALHAMADANI, A., AND GRAČANIN, D. Supporting embodied and remote collaboration in shared virtual environments. In *International Conference on Human-Computer Interaction* (2021), Springer, pp. 639–652.

[39] MATTHEWS, B., SEE, Z. S., AND DAY, J. Crisis and extended realities: remote presence in the time of covid-19. *Media International Australia 178*, 1 (2021), 198–209.

[40] MCVEIGH-SCHULTZ, J., AND ISBISTER, K. The case for "weird social" in vr/xr: A vision of social superpowers beyond meatspace. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (2021), pp. 1–10.

[41] MCVEIGH-SCHULTZ, J., MÁRQUEZ SEGURA, E., MERRILL, N., AND ISBISTER, K. What's it mean to" be social" in vr? mapping the social vr design ecology. In *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems* (2018), pp. 289–294.

[42] MYSTAKIDIS, S. Metaverse. *Encyclopedia 2*, 1 (2022), 486–497.

[43] NOWAK, K. L., AND BIOCCA, F. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments 12*, 5 (2003), 481–494.

[44] OJALA, J., SELIN, J., PARTALA, T., AND ROSSI, M. Virtual construction: Interactive tools for collaboration in virtual reality. In *Future of Information and Communication Conference* (2020), Springer, pp. 341–351.

[45] OP DEN AKKER, R., HOFS, D., HONDORP, H., OP DEN AKKER, H., ZWIERS, J., AND NIJHOLT, A. Supporting engagement and floor control in hybrid meetings. In *Cross-Modal Analysis of Speech, Gestures, Gaze and Facial Expressions*. Springer, 2009, pp. 276–290.

[46] PEJSA, T., KANTOR, J., BENKO, H., OFEK, E., AND WILSON, A. Room2room: Enabling life-size telepresence in a projected augmented reality environment. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing* (2016), pp. 1716–1725.

[47] PERRY, T. S. Virtual reality goes social. *IEEE Spectrum 53*, 1 (2015), 56–57.

[48] REIDSMA, D., OP DEN AKKER, R., RIENKS, R., POPPE, R., NIJHOLT, A., HEYLEN, D., AND ZWIERS, J. Virtual meeting rooms: from observation to simulation. *Ai & Society 22*, 2 (2007), 133–144.

[49] ROTH, D., LUGRIN, J.-L., GALAKHOV, D., HOFMANN, A., BENTE, G., LATOSCHIK, M. E., AND FUHRMANN, A. Avatar realism and social interaction quality in virtual reality. In *2016 IEEE Virtual Reality (VR)* (2016), IEEE, pp. 277–278.

[50] ROUPÉ, M., JOHANSSON, M., MAFTEI, L., LUNDSTEDT, R., AND VIKLUND-TALLGREN, M. Virtual collaborative design environment: Supporting seamless integration of multitouch table and immersive vr. *Journal of Construction Engineering and Management 146*, 12 (2020), 04020132.

[51] RUAN, J., AND XIE, D. A survey on qoe-oriented vr video streaming: Some research issues and challenges. *Electronics 10*, 17 (2021), 2155.

[52] SANDER, L., AND BAUMAN, O. reasons why zoom meetings are so exhausting. *The Conversation 5* (5).

[53] SCHÄFER, A., REIS, G., AND STRICKER, D. A survey on synchronous augmented, virtual and mixed reality remote collaboration systems. *ACM Computing Surveys (CSUR)* (2021).

[54] SHIN, D. Empathy and embodied experience in virtual environment: To what extent can virtual reality stimulate empathy and embodied experience? *Computers in Human Behavior 78* (2018), 64–73.

[55] SHIN, D.-H. The role of affordance in the experience of virtual reality learning: Technological and affective affordances in virtual reality. *Telematics and Informatics 34*, 8 (2017), 1826–1836.

[56] SHOCKLEY, K. M., GABRIEL, A. S., ROBERTSON, D., ROSEN, C. C., CHAWLA, N., GANSTER, M. L., AND EZERINS, M. E. The fatiguing effects of camera use in virtual meetings: A within-person field experiment. *Journal of Applied Psychology 106*, 8 (2021), 1137.

[57] SMITH, H. J., AND NEFF, M. Communication behavior in embodied virtual reality. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (2018), pp. 1–12.

[58] STEINICKE, F., LEHMANN-WILLENBROCK, N., AND MEINECKE, A. L. A first pilot study to compare virtual group meetings using video conferences and (immersive) virtual reality. In *Symposium on Spatial User Interaction* (2020), pp. 1–2.

[59] TOET, A., MIOCH, T., GUNKEL, S. N., NIAMUT, O., AND VAN ERP, J. B. Holistic framework for quality assessment of mediated social communication.

[60] WANG, P., BAI, X., BILLINGHURST, M., ZHANG, S., ZHANG, X., WANG, S., HE, W., YAN, Y., AND JI, H. Ar/mr remote collaboration on physical tasks: a review. *Robotics and Computer-Integrated Manufacturing 72* (2021), 102071.

[61] WIEBE, E. N., LAMB, A., HARDY, M., AND SHAREK, D. Measuring engagement in video game-based environments: Investigation of the user engagement scale. *Computers in Human Behavior 32* (2014), 123–132.

[62] WITMER, B. G., AND SINGER, M. J. Measuring presence in virtual environments: A presence questionnaire. *Presence 7*, 3 (1998), 225–240.

[63] YASSIEN, A., ELAGROUDY, P., MAKLED, E., AND ABDENNADHER, S. A design space for social presence in vr. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society* (2020), pp. 1–12.

[64] YOERGER, M., CROWE, J., AND ALLEN, J. A. Participate or else!: The effect of participation in decision-making in meetings on employee engagement. *Consulting Psychology Journal: Practice and Research 67*, 1 (2015), 65.

Appendix A: VRDays Questionnaire

What is your gender? *(Please cross one)*

◯ Male ◯ Female ◯ Other ◯ No answer

What is your age? ☐☐

What is your nationality? ☐

What is your level of VR experience? *(Please cross one)*

◯ Never used VR ◯ Used VR less than 5 times ◯ Used VR more than 5 times

How would you rate the overall experience?

Very poor 0 ◯ 1 ◯ 2 ◯ 3 ◯ 4 ◯ 5 ◯ 6 ◯ 7 ◯ 8 ◯ 9 ◯ 10 ◯ Very good

How would you rate the overall video quality?

Very poor 0 ◯ 1 ◯ 2 ◯ 3 ◯ 4 ◯ 5 ◯ 6 ◯ 7 ◯ 8 ◯ 9 ◯ 10 ◯ Very good

How would you rate the overall audio quality?

Very poor 0 ◯ 1 ◯ 2 ◯ 3 ◯ 4 ◯ 5 ◯ 6 ◯ 7 ◯ 8 ◯ 9 ◯ 10 ◯ Very good

How natural was the interaction with the live participant in the VR?

Very unnatural 0 ◯ 1 ◯ 2 ◯ 3 ◯ 4 ◯ 5 ◯ 6 ◯ 7 ◯ 8 ◯ 9 ◯ 10 ◯ Very natural

How restricted did you feel in your interaction with the live participant in the VR?

Very restricted 0 ◯ 1 ◯ 2 ◯ 3 ◯ 4 ◯ 5 ◯ 6 ◯ 7 ◯ 8 ◯ 9 ◯ 10 ◯ Not at all restricted

If you felt restricted, what kind of restrictions did you experience?

1

To respond: ☒ *or* ● papersurvey.io MWPL 0001

What did you like most about the VR experience?

What did you like least about the VR experience?

How can we make the VR experience more realistic? (What would you like to see, hear, feel, do, etc.)

Did touch enhance the quality of the VR experience? *(Please cross one)*

◯ Yes ◯ No

How important is it to be able to touch someone in VR? *(Please cross one)*

◯ Not at all ◯ Somewhat ◯ Very much

Do you have any other remarks or suggestions to improve this VR experience?

Appendix B: TNO Metaverse flyer

TNO Metaverse

TNO is introducing a new communication system using virtual reality for doing meetings with colleagues remotely. A 3D sensor will be used to capture your body and render it as a 3D point cloud in the VR environment. By putting on your VR goggles you will be able to see the point clouds of your colleagues in the VR environment and have meetings with them as if you are really together in the same room.

Functionalities that can be found in Teams will be available as well, with the added bonus that non-verbal communication cues can more easily be seen and processed. The system will first be tested until May to fine-tune details on the user experience. For these tests we need duo's or teams of 3 people who have weekly brainstorm or update meetings from the following locations:

- **TNO New Babylon**
- **TNO Oude Waalsdorperweg**
- **TNO Soesterberg**
- **TNO Groningen**

TNO innovation for life

We are now looking for teams who are willing to test out the system from the week of the **1st of April** and suggest any improvements. You will be able to book the systems at your TNO location through the Outlook Calendar for each system. The Calendar is available through the Sharepoint sites of the systems or you can link it to your own Outlook Calendar. Once you decide to join, the details of booking will be shared.

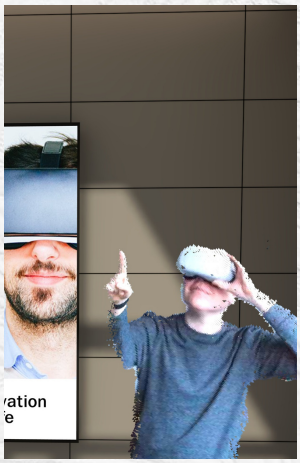
If you want to be a part of this pilot study with your team and are willing to:

- Do at least 4 meetings of up to an hour each with the system,
- Do the meetings at least once a week,
- Suggest any feedback using 15-minute surveys,

you are more than welcome to get in touch and test out the system.

Warning: you will not be able to take part in the study if you need to use external objects in your meetings or if your meetings involve heavy note-taking. The system can also not be used for classified work.

The pictures below show a representation of a pointcloud and the VR environment that has been built by our partner Connec2.



Get in touch

Dr. S.S. (Sylvie) Dijkstra-Soudarissanane sylvie.dijkstra@tno.nl
Simardeep Singh simardeep.singh@tno.nl

TNO innovation for life

Appendix C: Recruitment email

Dear colleagues,

We hereby invite you to test out the new TNO VR telecommunications systems that capture users using a 3D sensor and represent them as 3D point clouds in a shared VR environment. Using VR goggles you can see your colleagues in 3D in the VR environment, and experience the feeling that you are actually physically together in the same room.

This novel way of having remote meetings presents more immersive and engaging experiences. We call our platform the TNO Metaverse and it empowers users with a better sense of eye gaze with their colleagues, a better sense of mobility in a shared environment and a larger field of view on colleague's body postures. In turn all of these positive aspects decrease the unnecessary cognitive loads and so called 'zoom fatigue' that people who have a lot of remote meetings experience. Functionalities that can be found in Teams meetings are also included such as screen-sharing and note-taking using a virtual keyboard. The system will first be tested until May. For these tests, we need duo's or teams of 3 people who have regular (weekly) brainstorm or update meetings, from the following locations:

- TNO New Babylon
- TNO Oude Waalsdorperweg
- TNO Soesterberg
- TNO Groningen

During this pilot phase there are 2 systems at each location. Later down the line more systems will be opened up and more users will be supported. For now, we need your feedback in the pilot phase! Please see the attached flyer for more information on the systems and the pilot study.

If you are interested, please contact:

- Sylvie Dijkstra-Soudarissanane sylvie.dijkstra@tno.nl
- Simardeep Singh simardeep.singh@tno.nl

Appendix D: Informed consent

D.1 Information on the Social VR communication system

First and foremost, thank you for showing an interest in this study. We hope that by using the Social VR communication system you are able to engage in more immersive business meetings. In turn we hope to gain valuable feedback from you to improve the VR systems! The study is meant to evaluate whether VR systems using photorealistic avatars can truly create immersive and engaging meeting experiences.
This form is meant to inform you on what you will be able to do with the system and what you will not be able to do with the system, so please read this carefully. Moreover, we also need your consent for taking part in this study.
The system has been developed together with The Virtual Dutch Men using their Connec2VR platform. You will be able to do your meetings in this platform. As there are two systems at each of the 4 locations participating in the study, there will also be 2 emails. These emails can be used to log into the system. Further details will be revealed to you once you agree to take part in this study. The VR rooms are accessible to everyone. All of the equipment that is used for the system will be available in those rooms. A detailed guide on how to use the system will also be sent later on and will also be placed in the rooms.
Your body will be captured using a 3D sensor, which will then be rendered in the VR environment as a point cloud. The environment also enables you to connect your own laptop to it and use a remote desktop in the environment so that you are able to share presentations, share your screen or take small notes. Heavy note-taking will not be possible as you will not be able to use a physical keyboard.
Moreover, there are certain criteria to the meetings that can be done in the VR environment.

- You are willing to use the system at least 4 times, preferably more times.
- You are willing to use the system at least once a week. It would be best if you have a weekly meeting with a team of up to 3 people and use the system with the whole team. The system will be able to handle more users later down the line.
- You will not be able to use external objects during the meetings such as 3D models.
- You will not be able to hold meetings with the system about classified projects.
- Your meetings can only last an hour max.
- You will only be able to do short brainstorms, ongoing project meetings, update meetings or presentations with the system.

D.2 Data collection

The data that will be gathered with this system is purely meant for research purposes so that we can fine-tune details to create a better experience. The data will be collected using the system itself and using surveys.
The data that will be collected with the system is the following:

- Body-tracking data using the ZED camera, which captures your body so that the 3D point cloud can be made. Body positions and gestures will hereby also be tracked and logged. However, videos will not be made.
- Speech time, which means that the system will log when one user is speaking and when another user is speaking. Speech itself will not be recorded.
- The total time of the meeting in VR.
- System performance measures, such as framerate.

The survey will only be used to collect feedback on the system and to evaluate the user experience.
Furthermore, please be aware that we will always take into account TNO's privacy statement as mentioned on the following page: Privacystatement | TNO.
If there are still any questions on the criteria for using the system or on the data that will be collected, you can send an email to either:

- Sylvie Dijkstra-Soudarissanane, sylvie.dijkstra@tno.nl
- Simardeep Singh, simardeep.singh@tno.nl

D.3 Consent

Finally, thank you for agreeing to participate in the study. Before we continue, we also need your consent on the following:

- It is clear to you that your participation in this study is voluntary.
- The purpose of this study is clear to you.
- The data collected in this study is clear to you.
- It is clear to you that your data will be anonymized.
- It is clear to you that the data collected in this study will be stored for research purposes.
- It is clear to you that you can request any of the data that will be collected during the study.
- It is clear to you that you can request to delete any of the data that will be collected during the study.
- It is clear to you that you can stop at any moment during the study.
- It is clear to you that the data collected can be used for any secondary research within TNO.
- It is clear to you who you can reach out to for additional questions.

Appendix E: Participation email

Dear . . . .,

Thank you for filling in the consent form! I have added you to the Sharepoint sites of the systems at your locations. Please make sure that you both book the system that you will be using at your location in advance through the calendar that you can find on the Sharepoint site. You can also easily connect this calendar to your own Outlook calendar using the Outlook application on your laptop. For a detailed instruction you can take a look at the guide that I have attached to this email. This guide also includes more information on how to log into the system and how to enter and leave the TNO Metaverse.

From next week on you can start using the systems at [Location] and [Location]. First you need to know the password of the PC's/laptops, this is VRpilot1!. Only some systems require a PIN, this is 2012. To log into the Uplink platform on the machines you need the following information:
System 1 at [Location] [Room]

- Email =
- Username =

System 2 at [Location] [Room]

- Email =
- Username =

System A at [Location] [Room]

- Email =
- Username =

System B at [Location] [Room]

- Email =
- Username =

Furthermore, the password to all the Connec2 Uplink accounts is VRpilot1! as well. The workspace that you need to fill in is tno.

If you want to use your own laptop in the VR environment, please look at the instructions in the guide to see how that works. You can download the Uplink application for your laptop here: Download software – Connec2. You can download it on a TNO machine as well.

After you have had a meeting you will need to fill in the 15 minute survey. Here is a link to the survey: https://research.sc/participant/login/dynamic/882C729A-04AF-4845-93F4-9AF64D251009. You will also be able to scan a QR code in the room that directly leads to the survey as well on your mobile phone. The participant ID for [Name] is [ID] and for [Name] it is [ID].

Please let me know at what date/time your first meeting will be so that I can stay available for a Teams call if there is anything that you need help with!

Best regards, Simardeep Singh

Appendix F: Questionnaire

Demographics

1. What is your age?
   20-29 30-39 40-49 50-59 60-69
2. What is your gender?
   Male Female Other NA

Meeting characteristics

3. What is your participant ID?
   (Text entry)
4. How many meetings have you had with the system up until now?
   1 2 3 4 5 6 More
5. At what TNO location did you have your last meeting?
   New Babylon Oude Waalsdorperweg Soesterberg Groningen
6. Which account did you use for that meeting?
   (Text entry)
7. At what time did your meeting start?
   (Text entry)
8. What type of meeting was that last meeting?
   Update/project meeting Brainstorm meeting Presentation
9. How long did you think that meeting was?
   (Text entry)
10. To what degree did you already have meetings or talk in general with the other person(s) in the last meeting?
    Never Very rarely Rarely Neutral Occasionally Very frequently Always

The VR system

Please rate the following statements based on your own experience.

11. I think that I would like to use this system frequently.
    Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree
12. I thought that the system was easy to use.
    Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree
13. I would imagine that most people would learn to use this system very quickly.
    Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree
14. I found the system unnecessarily complex.
    Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree
15. I felt discouraged while using the system.
    Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree
16. I felt annoyed while using the system.
    Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

17. I felt that using the system was mentally taxing.
    Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

18. Please rate 'Strongly Agree' on the scale below.
    Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

Avatars & presence

19. I was often aware of all the other person(s) in the environment.
    Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

20. The other person(s) were often aware of me in the environment.
    Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

21. I felt the presence of the other person(s).
    Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

22. The other person(s) appeared to feel my presence.
    Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

23. The appearance of the other person(s) felt normal.
    Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

24. My appearance seemed normal to the other person(s).
    Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

25. The other person(s) in the VR environment seemed 'real'.
    Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

26. I felt an emotional and intellectual connection with the other person(s).
    Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

27. The other person(s) appeared to feel an emotional and intellectual connection with me.
    Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

28. I had a real sense of personal contact with the other person(s).
    Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

Communication in VR

29. The conversation held my attention for the whole meeting.
    Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

30. I lost track of time during the meeting.
    Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

31. My behavior was in direct response to the other person's behavior.
    Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

32. The behavior of the other person(s) was in direct response to my behavior.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

33. While communicating, my behavior felt normal.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

34. While communicating, the behavior of the other person(s) felt normal.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

35. I understood what the other person(s) meant.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

36. The other person(s) understood what I meant.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

37. While communicating, my reasoning felt normal.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

38. While communicating, the reasoning of the other person(s) felt normal.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

39. I could readily tell when the other person(s) was listening to me.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

40. I was able to take control of the conversation when I wanted to.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

41. It was easy for me to contribute to the conversation.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

42. The conversation seemed highly interactive.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

43. This felt like a natural conversation.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

44. Please rate 'Disagree' on the scale below. Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

Interaction in VR

45. I felt in direct contact with the environment.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

46. I was aware of my display and control devices.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

47. My sensations were consistent and agreed with the environment.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

48. The information coming from my various senses was inconsistent or disconnected.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

49. My interaction with the virtual environment felt realistic.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

50. The environment appeared real.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

51. I was consciously aware of being in a virtual world during the meeting.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

52. I was unaware of what was happening around me in the real world.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

53. I felt the urge to stop the meeting and see what was happening around me.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

54. My sense of being in the virtual world was stronger than my sense of being in the real world.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

55. I could not do some of the things I needed to do in the meeting.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

56. The virtual environment was responsive to interactions that I initiated.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

57. My interaction with the virtual environment seemed natural.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

58. I experienced delay between actions and expected outcomes.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

59. This meeting was like a face-to-face meeting.
Strongly disagree Disagree Somewhat disagree Neither Somewhat agree Agree Strongly agree

Please answer the following questions in 1-3 sentences.

User experience

60. How would you describe your involvement in the virtual environment experience?
(Text entry)

61. Were there any experiences in the virtual environment that seemed consistent with your real world experiences?
(Text entry)

62. Why would you use this system in the future for small project business meetings?
(Text entry)

63. Why would you not use this system in the future for small project business meetings?
(Text entry)

64. How did the system perform on a technical level?
Extremely poor Very poor Poor Neither Good Very good Excellent

65. Can you elaborate on your previous answer?
(Text entry)