UNIVERSITY OF TWENTE.

**Faculty of Electrical Engineering,
Mathematics & Computer Science**

HejVR: a Virtual Reality online cultural learning system

Zhuowen Fang
M.Sc. Thesis
November 2020

Supervisor:
Dr. ir. Dennis Reidsma
External Supervisors:
Alvin Jude (Ericsson)
Gunilla Berndtsson (Ericsson)

Master of Interaction Technology
Faculty of Electrical Engineering,
Mathematics and Computer Science
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

Abstract

Video conference technology has been developing tremendously during the past decades and the ongoing Covid-19 pandemic has further boosted its growth and adaption. One of the emphasized areas is online learning. However, this is insufficient as compared to face-to-face learning in many ways. This thesis presents a Virtual Reality (VR) system - HejVR - for online cultural lectures that aims to improve the learning experience of students in higher education.

The system focuses on two features: the first one is augmenting lecturer's gaze on the learner's side to provide learner an illusion of being looked at by the lecturer most of the time, and the second feature is hand raising gesture embodiment. Two Swedish language and culture lectures were created and integrated into HejVR. A user study was conducted with 13 participants combining quantitative measurements and qualitative interviews, to compare the augmented condition and a baseline condition where lecturer's gaze is equally distributed among students inside the scene and hand raising feature adopts common design in video conference platforms.

Our findings do not prove augmented gaze and embodied hand raising features to improve the sense of presence and engagement of the learners but indicate cultural learning a field that benefits from VR. The participants found their learning experience interesting and enjoyable and were in general more in favor of the augmented condition. In addition, we applied the concept of pre-recorded actors into our system to provide a multi-student scenario. The actors were only convincing for a small amount of people but the majority of the rest have a neutral to positive attitude towards the actors. Other behavior patterns of VR students were observed, such as nodding, shaking heads, touching surrounding objects, which can be useful features to consider in designing future systems.

Acknowledgements

The two years of EIT Digital master school study is an extraordinary experience. Attending two wonderful universities in two distinct but beautiful cities, having the opportunity to explore my field of interest, it has been a chance hard to come by. This thesis project is a remarkable ending to this journey. It guided me into the world of VR, provided me the opportunity to practice knowledge gained in my previous years of study, and challenged me in many ways. There were obstacles along the way but fortunately I was not alone.

I would like to give my deepest appreciation to my supervisors Alvin, Dennis and Gunilla for their time, guidance and inspiration. Greatest thank to Jakob and Emil for their generous help in the hardware. Thanks so much to all participants who joined the experiment voluntarily and everyone who filled in the questionnaire. I really appreciate the help from EIT Digital Stockholm CLC manager Joanna in recruiting participants. Finally, I would like to thank my family and friends who consistently gave me love and support. This thesis would not have happened without any of your help.

I wish this thesis can be inspiring and valuable to you who are reading in some ways.

Contents

Abstract **ii**

Acknowledgements **iii**

List of acronyms **vii**

1 Introduction **1**

1.1 Motivation 2
1.2 Research questions 4
1.3 Report organization 6

2 Background **7**

2.1 Distance education and avatar-based lecture 7
2.2 Avatar-based remote embodiment 8
2.3 Gaze awareness in telecommunication 9
2.4 Raising hand in video conference software 11
2.5 Summary 12

3 System Design **14**

3.1 Technologies 14
3.1.1 Hardware 14
3.1.2 Software 15
3.2 Avatars 15
3.2.1 Avatar Appearance 16
3.2.2 Basic Behaviors 16
3.3 Research Features 16
3.3.1 Hand Raising Feature 16
3.3.2 Augmented Gaze Feature 17
3.4 Lecture Scenarios 19
3.4.1 Supermarket Scene 20
3.4.2 Berries Scene 20
3.5 System Objective 20

3.6 Architecture 22

4 Implementation 23

4.1 Networking System 23
4.1.1 Network Manager 23
4.1.2 Network Objects 24
4.2 System Menu 26
4.3 Lectures 26
4.3.1 Lecture Environment 27
4.3.2 Avatars and Animated Behaviors 27
4.4 Research Features 29
4.4.1 Hand Raising Feature 29
4.4.2 Gaze Feature 30

5 Evaluation 32

5.1 Study Design 32
5.1.1 Procedure 33
5.1.2 Measurements 33
5.1.3 Special measures regarding COVID-19 34
5.2 Pilot study 35
5.2.1 Results 36
5.2.2 Bugs and improvements 36
5.3 Main Study 37
5.3.1 Hypotheses 37
5.3.2 Participants 38
5.3.3 Data Processing 38
5.3.4 Quantitative Results 39
5.3.5 Participant Behaviors 41
5.3.6 Interview Results 43

6 Conclusions and recommendations 45

6.1 Conclusions 45
6.2 Discussion 47
6.3 Limitations 48
6.4 Future Work 49

References 51

Appendices

A Teaching Script **60**
A.1 Supermarket Lecture 60
A.2 Berries Lecture 63

B Information Brochure **67**

C Consent Form **70**

List of acronyms

VR	Virtual Reality
AR	Augmented Reality
HMD	Head-Mounted Display
MOOC	Massive Open Online Course
IVBO	Illusion of virtual body ownership
CMC	Computer Mediated Communication
TSI	Transformed Social Interaction
RPC	Remote Procedure Call
UI	User Interface
FoV	Field of View

Chapter 1

Introduction

From smoke signal to calling a friend anywhere in the world by some simple taps on your smartphone, many problems have been tackled to enable affordable long-distance, real-time communication with rich information and modalities for the general public. Video conference tools have long provided a cost-effective and efficient way for business to communicate and collaborate with remote branches, partners, customers in a low emission way. Additionally, they greatly support daily team calls, meetings, presentations, all types of communication under the ongoing Covid-19 pandemic. Platforms such as Google Meet, Zoom and Microsoft Teams among other have become the new norm, allowing us to live as normal as possible in this difficult times. There is also significant surge of video conference in education. More teaching is undertaken remotely and on digital platforms although it was already rising before the pandemic like the massive use of Massive Open Online Course (MOOC).

Over the past few decades, technologies behind video conference including powerful CPUs, high-speed internet broadband, advances in video compression and high definition webcams have developed tremendously to be capable of providing a smooth user experience. However, this is still far from face-to-face communication, on account of its limited 2D view in a small screen space and a lack of physical interaction. The newly emerged VR technology brings new possibilities in a computer-simulated immersive environment, where a user is placed inside the environment, able to interact with the surrounding using multi-sensory modalities including visual, auditory, haptic, etc. It has witnessed a significant growth and widespread adoption, and is believe to change the way we live. This master thesis explores the use of virtual reality in online cultural lectures, one area we believe will benefit greatly from the adoption. This introduction chapter presents motivation behind and resulting research questions, and further describes overall structure of the report.

1.1 Motivation

This thesis project is in collaboration with Ericsson Research, Digital Representation and Interaction department in Stockholm, Sweden. As a global corporation, online business meetings across region make up a large part of Ericsson's daily work routine. Meanwhile, Ericsson provides plenty personal development opportunities to its employees through learning portals, training programs. We would like to use the chance of this thesis to explore opportunities in VR for above and more use cases.

In the beginning of the research, 10 user interviews regarding usage of current video communication technologies were conducted. The participants of the interviews include corporate employees, teachers and university students. Complete result of the user interviews can be found in the research topics report [1]. In brief, the result shows that participants are satisfied with the current technology in general, think they are easy to use, flexible, and the video quality has been greatly improved. Nevertheless, the following common issues were reported: 1. Difficulties in focusing on the subject as a listener or getting people's attention as a speaker; 2. Getting tired easily from bad audio quality resulting in losing attention from the subject; 3. Not easy to collaborate on creative work, brainstorm; 4. Lower physical engagement. The interviewees reported communication to be slower than in reality, which may be a result of less social cues, and one needs to be extra careful to not speak simultaneously with others. For presenters especially teachers, they sensed that direct interaction is missing, because they are not able to walk around among audiences or get as much feedback, consequently they cannot react accordingly.

To simulate a realistic face-to-face communication, holographic telepresence systems became a hot topic for scholars. Researches in this area was investigated and summarized in [1]. The following three paragraphs are rephrased from Chapter 3.1 in [1] to demonstrate our motivation. In holographic telepresence systems, remote people and surrounding objects are captured, compressed, transmitted, decompressed and finally projected in hologram. Substantial amount of researches were unfolded in human performance capture and dynamic reconstruction. We summarize several well-known projects as follows. Holoportation system by Microsoft [2] is a prominent example that provides a complete solution for end-to-end real-time 3D interaction in AR or VR. However, the system requires very complicated hardware setup and suffer from over-smoothing. To address some of the problems, the same research group brought out Motion2Fusion [3] which improves speed, generates more faithful reconstruction, more robust to fast motions and topology changes and supports single RGBD camera. Other single-camera attempts also exist, such as HybridFusion [4] which uses one depth camera and sparse inertial measurement units. LiveCap [5] even supports a single monocular RGB camera but this system cannot handle large topological changes and requires accurate template acquisition beforehand.

Another issue in a real-time capturing telecommunication is that the Head-Mounted Display (HMD) usually occludes a large portion of the user's face, hence blocking important social cues such as facial expressions and eye contact. This problem is addressed as well in Holoportation [2]. Many researches dedicated to tackling this problem integrate infrared cameras into HMDs with markers attached outside for calibration and accurate head pose tracking, and make us of an external RGB or RGBD camera for facial performance capturing [6], [7]. Frueh et al. [8] eases the use of infrared cameras by creating a database of user's face model in relation to gaze direction and blinks in advance and dynamically synthesis headset removed frames by aligning above face model onto HMD in real-time video stream. Meanwhile, instead of completely removing the headset, a translucent headset is added which reminds the viewer of the presence of the headset to mitigate uncanny valley effect.

Tremendous progress has been made, nevertheless relatively complicated setting, advanced hardware requirements and demand of significant amount of computational power show that there is still a long way to go in holographic telepresence systems until widespread adoption. Moreover, bandwidth required for transmitting volumetric data is still a concern in real-time telecommunication.

On the other hand, VR has been widely used for communication and collaboration purposes on the market, and we summarize three major use cases in existing commercial platforms. Below content in this paragraph is partly presented in [1] Chapter 3.3 but restructured for better explanation, and some additional materials are added. One of the first and most prevailing types of VR applications is social communication. Platforms like VRChat [9], Mozilla Hubs [10], AltspaceVR [11], Facebook Horizon [12] allow users to socially connect and explore, build creative environment, customize avatars and even gather for large scale events like concerts, conferences. This type of applications are also utilized for the second use case - business meetings and remote collaboration. But owning to specific business needs, specialized software for instance MeetinVR [13], Glue [14] and ENGAGE [15] emerged. These systems usually allow users to share and display media file in various formats to support work. With VR's immersiveness, users are no more restricted in the computer screen space so that more information is accessible at a time. Tools like sticky notes and whiteboards are also common in these platforms which intend to simulate office environment and improve the shared difficulty of brainstorming in video conferences. Lastly, many industrial specific platforms appeared. For example, The Wild [16], Resolve [17] and IrisVR [18] in architecture and construction industry use the immersive advantage of VR to enable users to view and actually experience a design work. In medical industry, Surgical Theater [19] intends to educate patients and their families in surgical plan, and Health Scholars [20] aims to provide first responders and clinicians repeatable, deliberate practice needed to close skill gaps and deliver confident, life-saving care. These types of education or training are domain specific, require elaborate implementation for each scenario, and highly focus on objects or environment rendered inside VR. With regard to more general lectures for wide public, some of the aforesaid business conference platforms attempts to target it as one use case. But we believe some features can be specially designed to bring more benefits to online live lectures, in regard of student behaviors, teacher behaviors in classroom, teaching strategies [21] [22] [23], and classroom interaction patterns whether teacher controls topic and or activity rule [24] etc..

Beyond that, VR is already used inside classrooms, to provide access to things or locations that are physically inaccessible in real-life like a peek inside the human body [25] [26], a cruise through galaxies, a walk on the Moon [27] [28]. But that is usually a part of the offline lectures instead of a remotely connected session.

Furthermore, cultural learning, having a global awareness and inclusive mindset is increasingly important, especially in region like Europe where culture converges and collides constantly over thousands years of history. The use of VR technology should aim more than digital accessibility but the experiential aspect of cultural heritage [29], and the Covid-19 pandemic is ushering numerous museums, cultural institutions to make use of virtual media to better deliver knowledge and create amazing experience online. For instance, Google Arts & Culture Virtual reality tours [30] serves a very affordable way to visit museums with merely a Google Cardboard and a mobile phone.

We believe these advantages also apply to the delivery of enterprise and school culture likewise, therefore we imagine a future VR to be used for online cultural lectures, remote school orientation, employee on-boarding, external branding and much more scenarios.

1.2 Research questions

Based on the above discussion, we conclude that application of VR in distance education is relatively immature, and challenges of distance learning lie in sense of engagement, the difficulty of stay focused, and the ability to interact between lecturer and students. There is still a long way to go in realistic and reliable telepresence. Apart from that, cultural education is a field that we see the many benefits from the advantage of VR. The goal of the thesis is to design, implement and evaluate a Virtual Reality system called HejVR for online cultural lectures. Naming of the system is explained in 3.4. Hence, the following research questions arose:

RQ1 How to create a VR system beneficial for avatar-based online cultural lectures?

- **RQ1.1** How to create a VR system for online lectures?
- **RQ1.2** What do avatars look like and what kinds of behaviors should they have?
- **RQ1.3** How to create VR environment for cultural lectures?

Here we refer to avatar as a digital representation of user's alter ego that facilitates interaction with other users, entities, or environment, in contrast to directly teleported self as holographic telepresence [31].

RQ2 What features can we design for avatars in online VR lectures to improve learning experience of the students?

A lecture is similar to a conference setting that only one person or a few people, either lecturers or presenters speak for the majority of the time, with occasional interactions with the audience and live exercises. As mentioned in Chapter 4 in my proposal [1], Sherblom discussed five variables and their moderating effect on Computer Mediated Communication (CMC) classroom: the medium, the social presence, the amount of student and instructor effort involved in classroom interaction, the student's identity as a member of the class, and the relationships developed among the instructor and students [32]. Although the intention of the essay is to support instructors to strategically develop positive learning conditions, it brings new ideas in implementing features into software system to facilitate study. The use of Virtual Reality directly falls into the variable *the medium* with VR's unique ability to provide rich information all-around the user compared to desktop interface. Below listed are features we brainstormed that may help engage the students together with their corresponding variables:

- Transformed eye gaze, illusion of obtaining high attention from the lecturer [33] [34] [35] (social presence)
- The ability to raise hands (the medium)
- Grouping effect [36] using colors or boundaries for students belonging to the same group in group discussions or who provide the same answer for in-class exercises (the student's identity)
- Keeping the lecturer in the student's Field of View (FoV) [37] (social presence)

- Highlight teaching material such as slides, images, objects according to lecturer's speech (the medium)
- Adjust the pitch of the lecturer's utterance into a more excited level to gain attention [38] (social presence)

Due to the limited scope of the study, we decide to focus on the first two features which suits more general audience and lecture scenario and construct the following research question:

RQ2.1 How do augmenting gaze at the learners' side and embodying hand raising gesture influence learners experience?

1.3 Report organization

The overall structure of the report is as follows: Chapter 2 covers related work in the field of distanced education, avatar based remote embodiment, gaze awareness and hand raising feature in video conference. In Chapter 3, design choices and reasons behind different system aspects are explained and their implementation details are revealed in Chapter 4. Chapter 5 describes method, procedure, measurements and results of user study. Finally, in Chapter 6, conclusions and recommendations of this study are given.

Chapter 2

Background

This chapter presents existing literature and state-of-art design in the market on relevant topic of the study including distance education, avatar-based remote embodiment and our two focused features: gaze awareness and hand raising.

2.1 Distance education and avatar-based lecture

This section is directly copy-pasted from Chapter 3.4 in my research topics report for this thesis [1]. Distance education is widely adopted in recent years with the emergence of MOOCs, which falls in the mode of asynchronous learning. Lectures are often provided in video recordings or in the form of video podcast [39] by combining audio recordings with slideshow images. [40] and [41] proved that distance education stands out in its flexibility in time and location, but students still prefer live lecture for its higher engagement, the ability to ask questions, easier to be focused or get motivation for learning. Similar deficiencies were mentioned during our interviews that even if in synchronous learning, like live lectures organized on video conference platforms, students' perceived engagement is much lower and less likely to ask questions although they have the ability to interact with lecturers in real-time. Many of the live online lectures are not very different from video podcast hence teachers are unable to gauge understanding.

Several studies dedicated on avatar-based online learning. Participation intention is a very important feature in learning activity, Chae et al. conducted a questionnaire based study and discovered that social presence has significant influence on participation intention through avatar trust, and attractive avatar type has stronger positive influence than an expert type [42]. Gesture is also considered to have a beneficial effect on learning and Cook et al. demonstrated hand gesture facilitates children's mathematics learning [43]. In the experiment, they used pedagogical agent so that the verbal and non-verbal behavior are very well controlled by the computer.

However, the mechanism of how gesture enhance learning remains unknown. Another study about student disclosure to pedagogical agents [44] in sensitive settings indicates positive correlation in emotional engagement and experience, and the emotional engagement can be influenced by appearance and body language. Besides, we may not forget that learners' preference differs, hence the same avatar can result in different levels of engagement and experience. Having a variety of choices of avatar types may serve a benefit in e-learning.

2.2 Avatar-based remote embodiment

The upcoming two paragraphs were written in Chapter 3.2 in my proposal [1]. The types of avatar representation can be very broad. Previously, scholars have focused on how the different avatar forms can influence user's perception of oneself and others. Nowak & Fox summarized researches around avatar anthropomorphism that more human-like representations are more favorable, lead to higher involvement, social presence, communication satisfaction and are more persuasive [31]. Latoschik et al. compared avatar realism in the form of abstract wooden mannequin and photogrammetry 3D scan and found the later one evoke stronger Illusion of virtual body ownership (IVBO), and the appearance of the others' avatars additionally influences the self-perception [45]. But visual fidelity (low: iconic, high: photo realistic) does not significantly influence user experience and task performance under stressful situation in VR, yet partial embodiment can greatly improve IVBO [46]. They further explored impact of avatar personalization and degree of immersion and verified both in increasing IVBO and virtual presence [47]. However, an experiment in comparison of real imagery reconstructed self and cartoon-like version avatar found later one exhibits higher sense of body ownership and presence [48]. This could be the result of poor reconstruction quality and Uncanny Valley. In the perception of others, people may feel realistically reconstructed avatar to be more reliable and trustworthy than character-like avatar [49].

Technology also enables novel interactions which are impossible in real life, one of such examples is Mini-Me [37]. Apart from a life-size avatar of remote collaborator, an adaptive mini version with redirected gaze and gestures consistently remain in the local AR user's field of view. This implementation is proved to convey the non-verbal communication cues necessary to improve the performance in an asymmetric remote expert setting, and improve social presence. The redirected gaze and gestures register the relationship of the person and surrounding objects, provide additional information for local viewer. SpaceTime also developed techniques for motion adaption but through spacial and object-level matching that avatars appear in reasonable position and their motion can match with the environment on remote site [50].

In the existing commercial VR communication systems, the choice of avatar types tend to align with the targeting industry and purpose of the platform. Platforms for industry specific purposes such as The Wild [16] for architecture and construction in the left most image of Figure 2.1 often present users only in abstract human-shaped forms as the focus is on the 3D models, design and self-experience of the environment. For VR social platforms, their goal is to allow users to connect socially, explore and express themselves, which can vary greatly from the true self of the user. Avatars are often customizable, in full-body shape and have creative looking like VRChat [9] in the middle of Figure 2.1. Users can additionally send continues emojis on top of their avatars to express emotions in AltspaceVR [11]. VRChat further supports lip sync, eye tracking, and selectable hand gestures and emotes. In applications targeting on general business meetings like Glue [14] in the right most of Figure 2.1, avatars are either in upper-body or full-body with ordinary human looking. They may have a series of gestures to communicate, controllable head gaze direction and hand positions. Other hand gestures are designed to conform to functionalities such as sketching, moving an virtual object. Figure 2.1 presents avatars in the three different types of platforms.

Figure 2.1: Commercial VR communication platforms: The Wild [16], VRChat [9], Glue [14] (from left to right)

2.3 Gaze awareness in telecommunication

Eye gaze is long regarded to be important in conversation. Some of the functions of gaze direction includes: regulate interaction, to facilitate communicational goals, and to express intimacy and social control etc [51]. But most importantly, perhaps, other individuals' directed gaze signals their direction of attention [52]. Gaze is generally used to refer to one person looking at the face of another person, particularly the region around their eyes. Mutual gaze denotes two people doing this to each other, often called *eye contact* [53].

Various kinds of ways to express gaze were experimented by former researchers. The following studies in this paragraph was summarized in Chapter 3.2 of my proposal [1]. Early attempts like [54] express gaze information by rotating participant's image and using light spots placed on table and documents to indicate their point of attention as shown in Figure 2.2. Similarly [36] uses floating bubbles to gently express eye contact, but inside VR environment, and further implemented joint attention and grouping augmentation displayed in Figure 2.3a. The result shows an improvement in social presence. CoVAR shows boundary for field of view cue and gaze cue in ray beam (Figure 2.3b), and supports eye-gaze-based interaction including gaze and gesture combination and collaborative gaze [55].

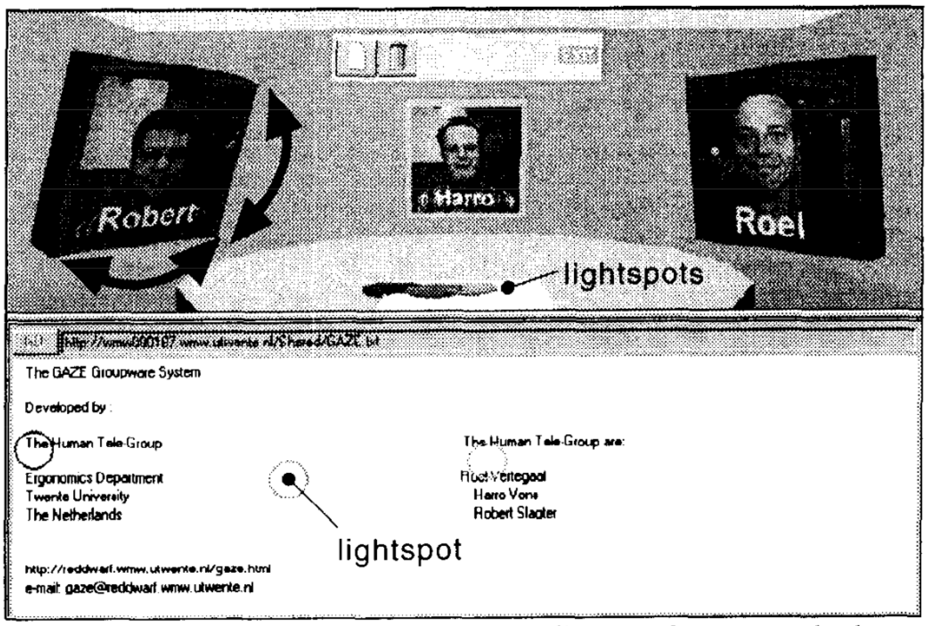
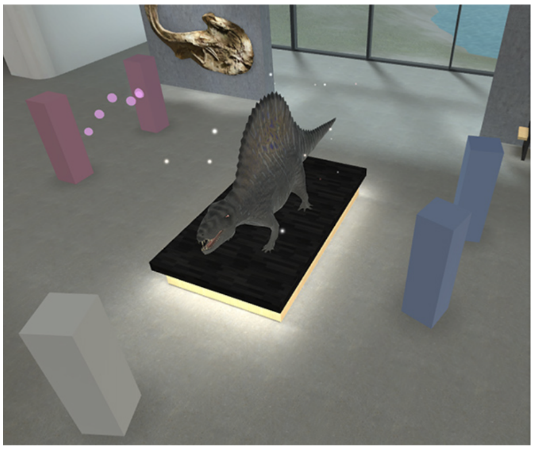
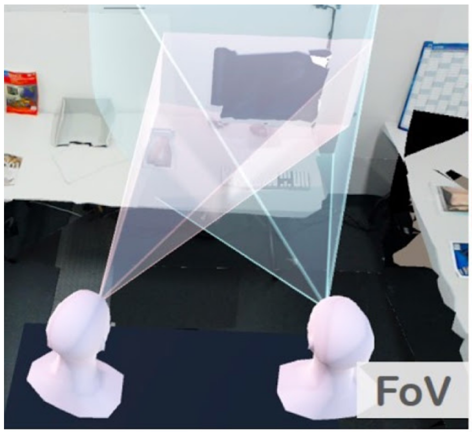


Figure 2.2: Rotating users' image according to where they look. Taken from [54]

Several studies were also carried out to experiment augmented gaze in CMC, belonging to the self-representation dimension of Transformed Social Interaction (TSI). These transformations decouple the rendered appearance or behaviors of avatars from the actual appearance or behavior of the human driving it [56]. Figure 2.4 illustrates a system where the presenter is rendered differently for each other interactant, giving the illusion that they are being looked at even if the reality is not. Separated projects [33] [34] [35] in augmented gaze produced common findings that: (1) participants never detected that the augmented gaze was not in fact backed by real gaze; (2) participants returned gaze to the presenter more often in the augmented condition than in the normal condition; and (3) participants were more persuaded



(a)



(b)

Figure 2.3: (a) Condition with transformations for eye contact (floating bubbles), joint attention (particle highlights on object) and grouping (avatar colors). Taken from [36]; (b) Virtual awareness cues, the FoV and Gaze cues. Taken from CoVAR [55]

by a presenter implementing augmented gaze than a presenter implementing normal gaze. They also believe augmented gaze would have a high impact in distance learning.

2.4 Raising hand in video conference software

Hand raising is a prominent way of interaction in classroom, meetings and conferences. Sahlström et al. pointed out that hand raising involves certain procedure: raising one's hand while at the same time directing one's gaze and torso toward the teacher, while being silent [57]. It is usually done with only one hand, occurs in teacher turns at or in projection to turn-transition-relevance places. Raising one's hand is a way of displaying that you know the answer, a willingness of a public turn at talk, or plies a recognizable display of participation [57]. Hand raising is seen to be associated with cognitive engagement [58].

In Zoom, Microsoft Teams, and the most recent version of Google Meet, participants can click a button to raise their hands and everyone in the meeting can see hand icons next to the name of whom raise their hands in participant list as shown in Figure 2.5. Additionally the meeting host or in other places all participants will receive notification of the action. This is a helpful indication to the speaker that they can choose to stop at an appropriate time to let the listeners express their opinions without much interruption. Language teachers suggested Zoom is better suited for distance learning than many other video conferencing applications and the option to

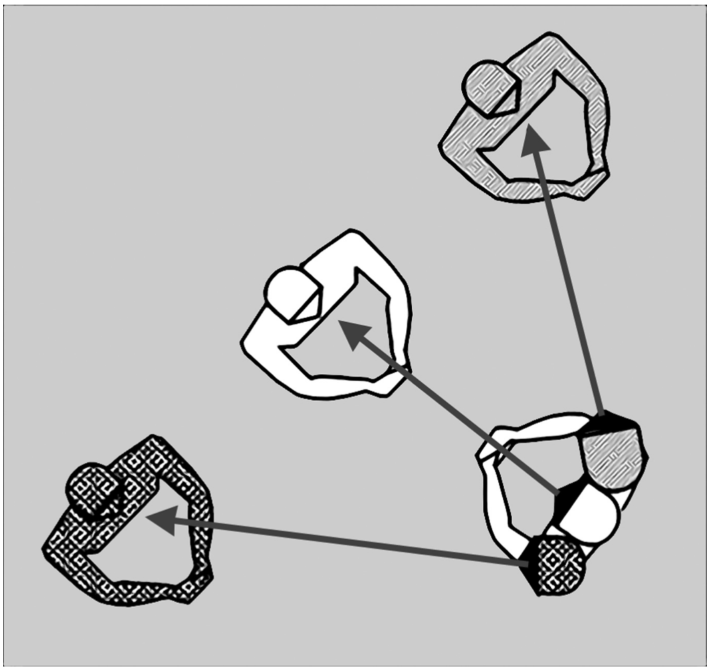


Figure 2.4: A schematic illustration of non-zero-sum gaze. All interactants on the left perceive the speaker on the right gazing directly at themselves. Taken from [56]

raise one's hand is mentioned as one of its advantages [59].

2.5 Summary

This chapter presented researches related to different aspects of our system. The available literature compared distance education to face-to-face lectures, designed and evaluated features that positively impact participation, learning, emotional engagement in avatar-based online learning system. We also looked at avatar-based mixed reality systems in general to draw a relation between avatar representations and user experience.

In terms of our research features, different techniques to express eye gaze in both 2D and VR systems are explored. We further discovered the potential to augment gaze in order to build a stronger positive relation with the listeners. Behavior and function of hand raising in classroom are discussed and we presented how this feature is currently implemented in popular video conference applications.

In the next sections, knowledge gained in this chapter is applied into HejVR system design.

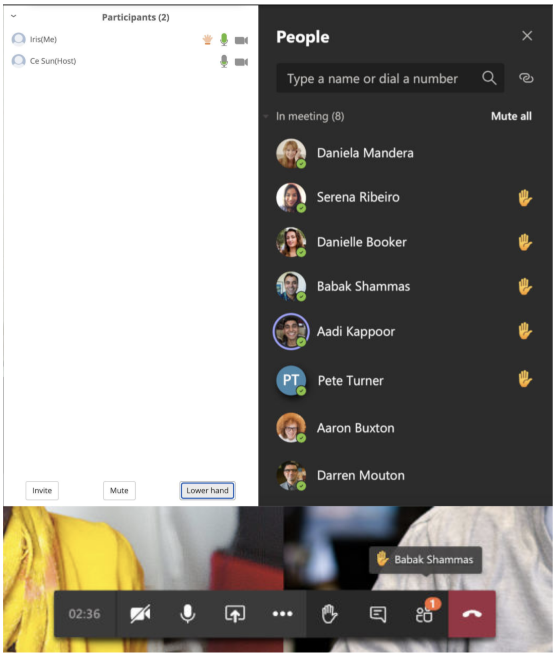


Figure 2.5: Common hand raise feature on video communication platforms. Top left: Zoom. Top right & bottom: Microsoft Teams (Taken from [60])

Chapter 3

System Design

Previous chapters identified opportunities in online cultural education. In this chapter, different aspects of the system and reasons for certain design choices are provided.

3.1 Technologies

This section gives an overview of hardware and software technologies used to develop and evaluate the system.

3.1.1 Hardware

HTC Vive Pro: HTC Vive Pro [61] is a popular tethered VR headset that requires connection to a personal computer in order to run but is expected to have lower latency and better processing power for more complex, smooth VR experience compared to standalone VR headsets. Two base stations (BS 1.0) positioned diagonally at the corners help to determine the location of headset and controllers in the space by emitting timed infrared pulses. The headset has 110 degrees FoV and 2880 x 1600 pixels resolution both eyes combined. Two wireless controllers with 6 degrees of freedom are used to control events.

Apple AirPods 2nd Generation: Although the HTC Vive Pro headset comes with Hi-Res certified headphones that supports 3D spatial sound, severe sound distortion was frequently experienced during the implementation, which appears to be a common issue of the headset. In order to provide consistent audio throughout the experiments, a pair of AirPods [62] were used instead. They only provide stereo sound instead of spatial sound but owing to the experiment setting, the audios only include speeches from lecturer and students around, which come out relatively in the horizontal plane of testers' ears, we consider this substitution does not have much impact on the experiment result.

Figure 3.1: Hardware used: A. HTC Vive Pro with controllers and 2 base stations. Taken from [61]; B. Apple AirPods. Taken from [62]; C. A participant wearing VR headset and AirPods

3.1.2 Software

Unity: Unity [63] is a game engine which is able to create 2D, 3D, Augmented Reality (AR) and VR systems. SteamVR plugin [64] by Valve Corporation provides easy HTC Vive VR development support in Unity. The Unity version used on the host side of the system is 2019.3.14 and on client side is 2019.3.15.

Blender: Blender [65] is an open-source 3D creation suit, the model created by which can be directly imported into Unity as 3D objects. It is used in this study to create lecture specific objects for the experiment.

3.2 Avatars

The goal of this study is to identify features on avatars that are able to enhance the experience of students in online VR cultural related lectures. This section breaks down the avatars into their appearances, basic behaviors and the design of two features we think that might benefit multiple dimensions of students' learning experience.

3.2.1 Avatar Appearance

Taking into account the study topic as online lectures, which is regarded as a professional scenario similar to commercial platforms for business meetings, we decided to use anthropomorphic avatars dressed in ordinary garments to simulate a study excursion, so that the students can focus more on the lecture content and environment than avatar appearance. We also limited the fidelity of the avatars to low poly (polygon meshes with relatively small number of polygons) for four reasons: 1. Computational limitation. The surrounding VR environment has considerable amount of objects to render and real-time response is desired in an online lecture, hence we choose to lower down avatar fidelity to reduce system delay. 2. Lower the tendency of uncanny valley. 3. Allow participants to focus on research features, reduce distractions from appearance and other behaviors while providing basic channel for communication. 4. As Nowak and Biocca pointed out in [66], higher anthropomorphic sets up higher expectations that lead to reduced presence when these expectations were not met. With the given hardware, we are not able to have full body tracking of the participant, hence we want to maintain the level of expectation compatible to what the system can offer.

3.2.2 Basic Behaviors

The avatars have neutral facial expressions and basic animations for idle standing and the ability to rotate the entire body to another direction. The lecturer avatar has more expressive behaviors including turning head without turning entire body, pointing to objects, different talking gestures to match with the teaching content and get the students' attention. Besides, a name tag is displayed at chest height in front of each avatar so that players can easily tell who the person is.

3.3 Research Features

3.3.1 Hand Raising Feature

The ability to raise one's hand either to react to the lecturer's questions or to provide a sign of the willingness to speak is very important in an online lecture. It is one of the most predominant way of interaction from the students besides gaze and speaking. Lugrin et al. mentioned in their findings that partial embodiment can greatly improve

IVBO [45]. Therefore, for the first feature we choose to enable partial embodied hand raising which consists of the following two parts:

1. Hand raising animation on all avatars, so that the VR player can see surrounding people performing hand raising and also this behavior applied on his or her own avatar while looking at the shadow.

2. Augmenting the sense of hand raising by displaying an additional arm in front of the VR player when the player raises hand. Due to the limited FoV of the VR headset, the player is not able to see hand raising animation from first-person perspective except for from the shadow, hence we attached an arm in front of the VR camera in a reasonable position to enhance the sense of raising one's hand. This arm only consists of a hand and forearm and is not attached to the torso, but that cannot be observed from the player. The arm is also hidden from other clients connected to the scene so that they can only see the first part of this feature.

In comparison, the baseline condition adapts the common design from video conference platforms as shown in Chapter 2.4. In our VR system, a hand icon is displayed next to the name tag in front of the avatar once a hand raising behavior is triggered. Since all avatars are placed in a immersive 3D space facing inwards to form a circle and the amount of people is low that one can easily see each other's front face, we do not use an additional participant list. The trigger of the hand raising and hand down is the same as the augmented condition in order to remove the influence of interaction modality but focus on the interface design.

Chapter 4.4.1 explains the implementation detail and provides run-time screenshots as illustrations for both conditions.

3.3.2 Augmented Gaze Feature

Gaze and eye contact in conversations have been an interesting topic to scholars and various attempts of expressing and augmenting gaze in computer-mediated communications has been made as presented in Chapter 2.3. To design gaze rules for our system, studies on actual gaze behavior of people in conversation are reviewed as a reference. A frequently used measure of individual and mutual gaze is the percentage of time spent looking [53]. Studies of conversations between two strangers found that on average people spend about 50% of time looking at the other person, ranging from 30% - 70%. Length of looking is usually short, at about three seconds in average for individual gaze. Vertegaal et al. [67] measured gaze at conversational partners during four-person conversations and found that when listening or speaking to individuals, a person has 88% chance looking at the person listened to and 77% chance at the person spoken to. When addressing all partners, the speaker gazes at each individual 19.7% of the time.

Based on above information, we summarize that there are two cases of the lecturer's gaze in a lecture scenario: 1. Gaze when the lecturer speak to all students for providing teaching content; 2. Gaze when the lecturer speak or listen to an individual student. In design of our system, the lecturer's gaze focuses on the student 100% of the time in the second case while speaking or listening that person in both augmented and baseline conditions. The amount of time spent in gaze is slightly higher than in reality but since the avatars have low fidelity, we assume it would not result in unnatural feeling. This type of gaze also assists as a signal to encourage a participant to speak.

When the lecturer provides teaching content and speaks to all students, we split the lecturer's gaze into three-second time slots that the lecturer switches gaze subject every three seconds. In the augmented condition, the lecturer gazes at the VR participant every other time slot, resulting in 50% of gaze at the participant, and the lecturer randomly gazes at one of the four actors in other time slots to reduce pre-programmed feeling so that each actor receives approximately 12.5% of gaze. Consequently, the participant gets 4 times gaze that of each other student actor. In baseline condition, the lecturer picks a random student including both VR participant and actors at each time slot as gaze subject hence the participant receives 20% of gaze, the same as the actors. The gaze process is visualized in Figure 3.2 and Figure 3.3.

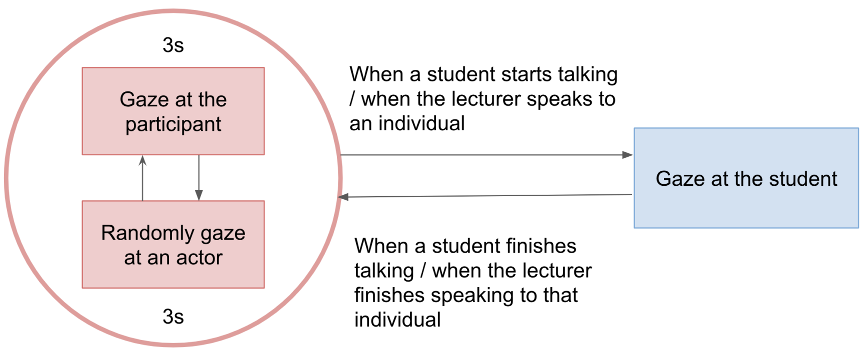


Figure 3.2: Lecturer's gaze in augmented condition

As for the student actors, because there is no direct interaction between the students, all actors gaze at the lecturer except when the lecturer signals to look at a specific object.

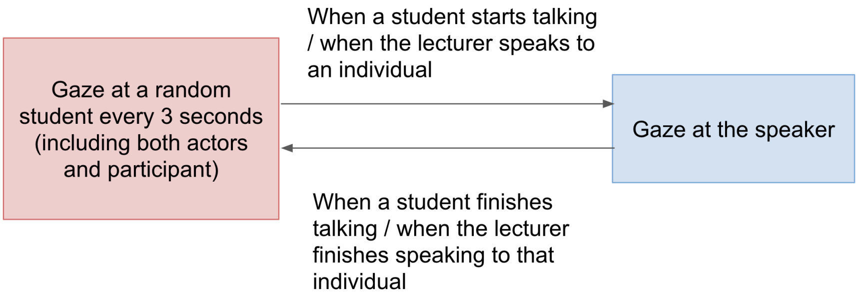


Figure 3.3: Lecturer's gaze in baseline condition

3.4 Lecture Scenarios

The detailed scenario of VR cultural lecture is further explored in this section. Many European universities have language and culture introduction courses targeting international students. For example, Dutch A1 - Dutch is fun! course in University of Twente [68] teaches simple Dutch words and phrases, everyday language which the students can easily put into use in their daily life. Some Dutch culture is also introduced in this course. Similarly, KTH Royal Institute of Technology has Introductory course: Swedish language and culture [69] with the goal of teaching "survival" skill language in the meanwhile introduce history, culture and everyday life to international students. Upon interviewing responsible teachers for above courses, we found that besides in-class lectures, excursions and field work are often involved to better provide context with surrounding environment and objects, and give the students a chance to practice words phrases in places like museums, supermarkets, etc.. Under current time when museums and non-essential business are closed and social gathering is limited, normal lectures can still carry out using video conference platforms, however these types of excursions have more difficulties to carry out. We see excursions benefit more from the immersiveness and unrestricted context a VR system can provide compared to in-class lectures and hope to tackle this niche field with the use of VR,

Considering the study was going to be conducted in Sweden, two 12-minute long Swedish language and culture lecture scenes were created based on Introductory course: Swedish language and culture course material from KTH [69]. The lecturer who is previously an exchange student to Sweden first gives a short introduction of herself and then starts giving the lecture in a friendly, sharing way. The students are internationals who recently arrived in Sweden for study, work or travel purposes and are interested in knowing more about Sweden. We briefly present contents of each scene below. The complete teaching script can be found in Appendix A. We also chose to name the system ”HejVR” as ”Hej” is one of the most used greeting word in Swedish meaning ”Hi”, being used by the lecturer inside our system as well, and the lecturer further teaches this word in Supermarket Scene.

3.4.1 Supermarket Scene

This lecture introduces several basic Swedish phrases for daily interaction and shopping specific situations. Besides it discusses about the special PANT recycling machine common in Swedish supermarkets. PANT refers to a legal term for packaging that one receives financial compensation upon return of the packaging which is responsible by AB Svenska Returpack (Pantamera) [70]. The scene is located inside a virtual supermarket close to the checkout counter, and to match the teaching content, a supermarket staff and PANT machine must be visible from the VR user's perspective. There is also a canvas in the scene next to the lecturer in order to display texts and pictures to assist teaching.

3.4.2 Berries Scene

This lecture brings students into a virtual forest, presents various types of berries rich in Sweden, their usage in Swedish cuisine, berry picking tool and the special right ”Allemansrätten” which makes berry picking possible in almost every non-private nature land. A table stands in the middle of the group is used to display the different berries and berry picker. The scene contains a canvas for displaying texts and pictures as well.

3.5 System Objective

The objective of the system is to host an online VR lecture inside a pre-designed lecture related environment. There are two types of user of the system: lecturer and students. The students are fully immersed in VR at all time but the lecturer is connected from a computer which acts as host (server with a local client) for the clients in the network. All participants first enter a menu scene, where the lecturer decides on the lecture scene and the students can choose an avatar of preference. By joining the lecture, the students are spawned into the lecture scene selected by the lecturer. The VR participants are free to move around and speak via headphone at any time in the virtual environment. Another possible interaction from the students' side is to raise hand by raising the controller. The lecturer can speak via headphone and move the avatar position with keyboard. Besides seeing all attendees inside the environment, the lecturer can also see a participant list at the top-right corner of display. Additionally, the lecturer can choose to play different talking gestures by clicking buttons displayed on the screen.

The original idea of the experiment was a between-subject design, where multiple student testers would join at the same time, each of them would be connected via a VR device with half of them in augmented condition and half in baseline condition. The lecturer would teach the same content to all students but the testers would perceive different behaviors and feedback of the lecturer‘s gaze and hand raising effect. However, due to the ongoing pandemic, we have to be cautious in introducing extra in-person contact, in the meanwhile limit the amount of participants, a within-subject study was implemented. Nevertheless, so as to simulate a group excursion setting with only one active participant at a time, we introduced pre-recorded student actors (hereafter being referred as actors) into the lectures, and therefore the system was fine-tuned to control the behavior of the avatars of actors. The points of time when a certain actor raises his or her hand were marked out on the teaching script A and their responses were recorded into individual audio files. During the experiment, the lecturer in the meantime acted as a wizard to control the actors to raise up or put down their hands, speak by playing corresponding audio on the matching actor, and turn around to specific objects when the lecturer mentions them. Furthermore, in order to maintain speed, intonation of speech and overall quality of teaching across experiments, the general teaching script was also recorded into audio files that the lecturer can play or pause the audio during the experiment to interact with students accordingly. Figure 3.4 demonstrates setup of a lecture.

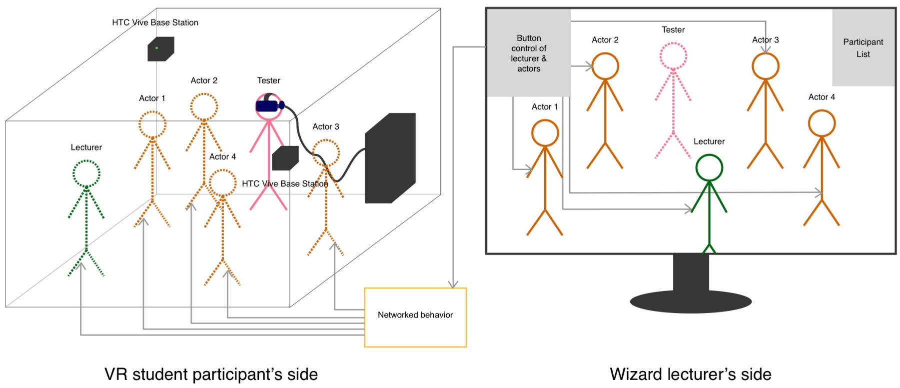


Figure 3.4: Basic setup overview with a VR student and a lecturer

3.6 Architecture

The system architecture is founded on above design aspects. A schematic overview of the system can be found in Figure 3.5. The system is split up into two types of endpoints: host for the lecturer, and client for a student. At each lecture session there is only one host but multiple clients can exist simultaneously, however in our experiment only one client is used as there is only one participant in each session. A complete description of each component is given in the Chapter 5.

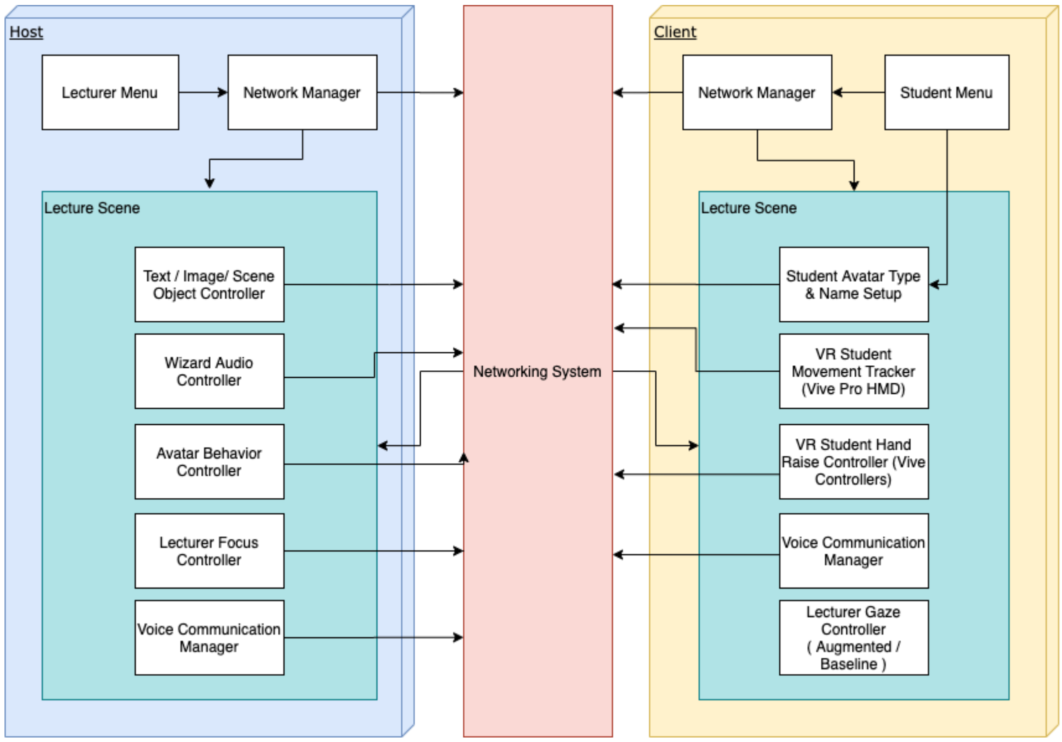


Figure 3.5: An overview of the system architecture

Chapter 4

Implementation

In this chapter we explain in detail how the different components introduced in previous chapter are implemented and integrated using aforementioned software and hardware.

4.1 Networking System

The networking system is the middleware of the lecturer and students' side of the system, connects them over internet and manages data streams between them. It is built on top of Mirror [71], an open source high level networking API for Unity. Mirror is easy to use, and can support a large scale of players.

4.1.1 Network Manager

Network manager is the core of Mirror multiplayer setting which manages game scene and object spawning. It exists on all endpoints (both host and client) and persists through scenes in our system until the game session is shut down. Starting from the offline scene - Menu Scene, the lecturer selects lecture scene which is either Supermarket scene or Berries scene, and the network manager sets it as the online scene for all upcoming clients. Once the lecturer confirms the selection, *StartHost* function of the network manager is triggered so that the lecturer is switched into the online lecture scene and the host keeps listening for incoming connection on the specified port. At the client side, when the student selects to join lecture, the network manager tries to connect to the host with specified IP address and port by calling *StartClient* function, spawns in the corresponding avatar game object based on student's choice in menu scene, and switch to online scene as soon as network session successfully starts.

4.1.2 Network Objects

Network objects refer to objects in scene that have networked behaviors, including all lecturer and student avatars, canvas for text and image display, several lecture specific objects (various berries and berry picker in Berries scene) and a *ServerWizard* object. *ServerWizard* is a server only game object with *WizardCanvas* as a child object, so that only the lecturer is able to see it and can control audios and behavior of the lecturer and actors using buttons placed on top. The other network objects exist in both host and client sides. All networked objects have a unique network identity to make the networking system aware of and differentiate each object.

On the host side, lecturer controls the following behaviors by clicking buttons on *WizardCanvas* as shown in Figure 4.1. *Wizard audio controller* controls a certain avatar's speaking behavior whether it is the lecturer or an actor by playing or pausing an audio file. Meanwhile, when a student actor or the participant is speaking, the lecturer's gaze focuses on the speaker until *Unfocus* button is clicked. In *Avatar behavior controller* of baseline condition, to raise hand or put hand down of the lecturer or an actor, *SetActive* function of the hand icon in front of the avatar is called with value true or false. Similarly, *Image and scene object controller* displays or hides an image or a lecture specific object by calling *SetActive* function of the game object with a boolean argument. For controlling Swedish text on canvas, a string is passed to the Text game object and the text value can be set equal to the string. To hide texts, an empty string is passed.

Figure 4.1: Lecturer's view with *WizardCanvas* visible

On the client side, the VR student participant can trigger the same hand raising behavior to display or hide the hand icon via VR controllers in baseline condition.

All above controls are done by Mirror Networking's Remote Procedure Call (RPC)s, that objects in the local client of the host or the VR participant client will make a command call to the server. For security reason, the command calls can only be made by local objects with client authority. The host owns the lecturer, actors and the canvas objects and the VR client has ownership for the student's avatar merely. Upon receiving the client command actions, the server triggers a clientRpc call to all clients, including local client of the host. In some cases an *excludeOwner* option is set to be *true*, so that the client who made the command call will not receive the corresponding clientRpc call. The RPC action flow is illustrated in Figure 4.2.

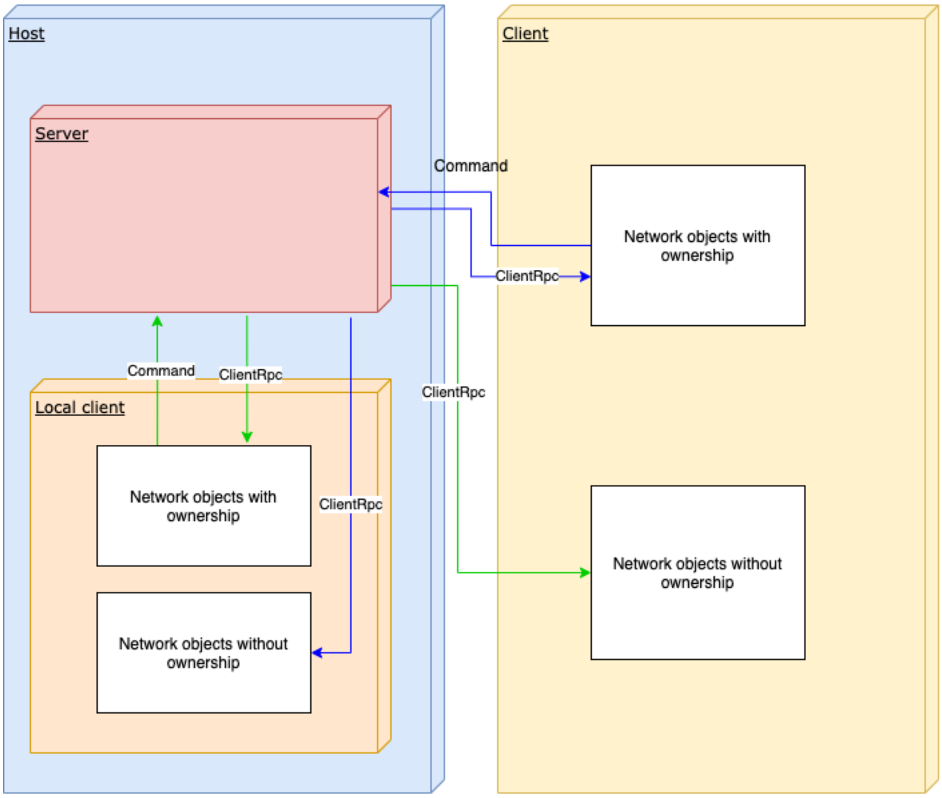


Figure 4.2: RPC actions in network system

Live voice communication is done by streaming selected microphone audio inputs. Once the user turns on the feature, Unity's *Microphone.Start* function is called, to record audio clips from the microphone. The loop parameter is set to be true so that the recording will continue from the beginning if the audio clip length is reached. At each frame, the newly collected audio clip samples will be transmitted to all clients in byte array inside *Update* function by using above-mentioned RPC calls, and transferred back on clients side into audio clip to play on the corresponding avatar.

Furthermore, all avatars contain *Network Transform* components to synchronize the position, rotation, and scale of the object and *Network Animator* components that handles animation states across network. For the VR student participant, the player game object owns a *Network Transform Child* component in addition so as to sync the object's position and location with its VR HMD with the aim of maintaining a first-person perspective. These are pre-existing components handled by Mirror and no additional configuration needs to be done.

4.2 System Menu

Upon entering the system, all joiners are placed inside the Menu Scene to select their role as either lecturer or student as displayed in the first image in Figure 4.3. After that, the lecturer sees the menu in second image in Figure 4.3 where she can choose the lecture scene and start hosting. This must happen before the VR client joins so that the correct IP address could be found. The VR participant can then select a preferred avatar and join lecture as shown in the third image in Figure 4.3. The participant's name and host's IP address are also displayed here but to simplify the process the data is preset in script.

The lecturer's menu is in full-screen view and can be interacted via mouse click. Inside VR view, the menu is displayed like a projection screen and a laser pointer is emitted from controllers that the player can press trackpad on the controller for selection.

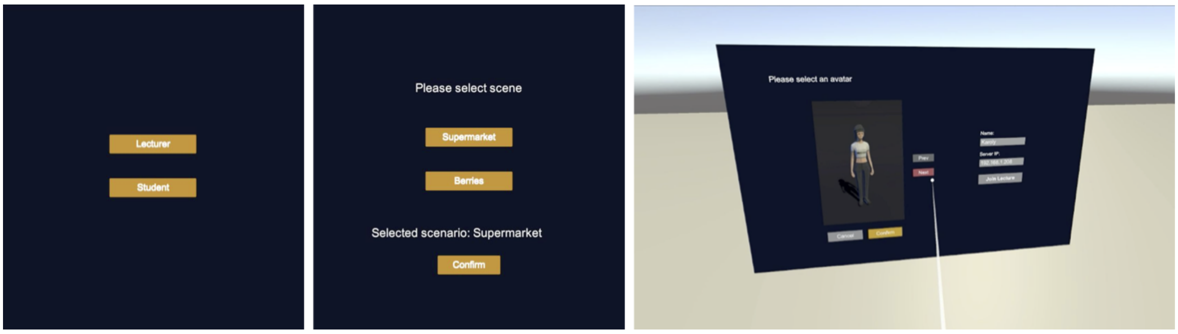
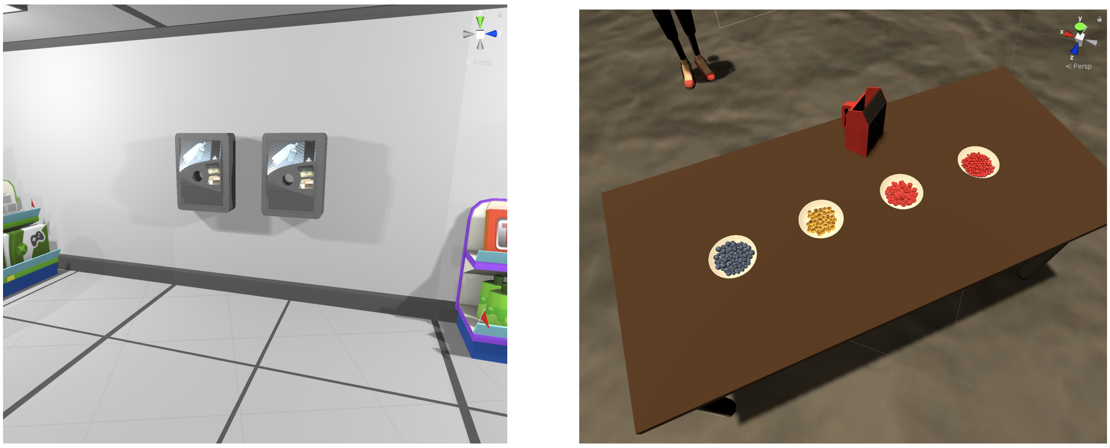


Figure 4.3: From left to right: Menu initial page; Lecturer's menu; VR student participant's menu

4.3 Lectures

A lecture scene is made up of environment background and human avatars. The environment is specially built up for the lecture to simulate an excursion situation. We also added several content specific objects to help students understand teaching context. Avatar implementation involves building their appearance and behaviors. In this section, we describe these aspects in detail.



(a) PANT machines *(b)* Berries and berry picker

Figure 4.4: Scene specific objects imported into Unity 3D from Blender

4.3.1 Lecture Environment

For the two lecture scenarios, two different environment scenes are built up using existing libraries in Unity Asset Store and original models self-made in Blender. In the Supermarket Scene, Snaps Prototype — Office [72], Ultimate Low Poly Supermarket, Shop Shelf Pack [73] and Modular Railing Set [74] are used to construct the overall supermarket hall and the counters were created inside Blender. A Swedish bottle recycling machine was also built in Blender that the lecturer later points to the machine and explains its functionality and history (Figure 4.4.a). The Berries Scene adapts the sample scene in Green Forest library [75] and picks a relatively flat location as excursion spot. Various types of berries, a berry picker and a table to place them on top were made in Blender to help visualize the lecture content (Figure 4.4.b). Figure 4.5 presents environment of the two scenes, in which light bulbs signify light sources inside scenes and the blue squares indicate the belonging objects are network objects as explained in Chapter 4.1.2. These symbols are hidden from users in actual play mode.

4.3.2 Avatars and Animated Behaviors

A series of avatars were created from Advanced People Pack Unity assets [76]. This library allows creation of male and female avatars with different skin tones, haircut, facial features and clothing. One lecturer, four pre-recorded student actors (two male and two female), four supermarket staff (two male and two female) were created and placed in the scene, and eight student avatars made to be chosen by the participants (four male and four female) as shown in Figure 4.6.

Lecturer and students perform standing idle animation in loop by default. When

Figure 4.5: Left: Supermarket Scene; Right: Berries Scene

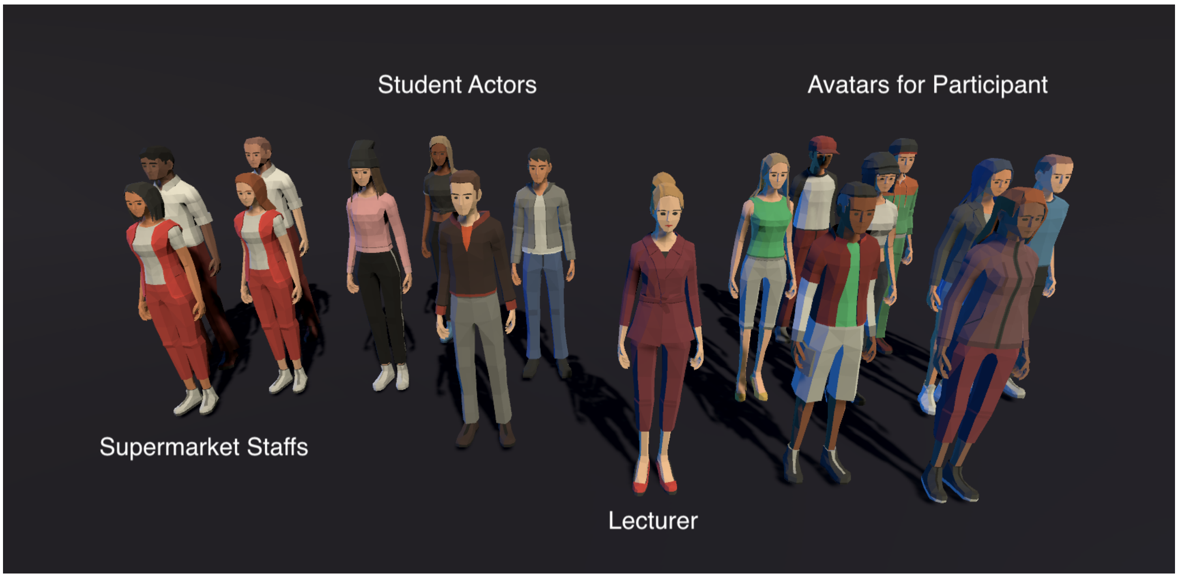


Figure 4.6: All avatars created for lectures

hand raising is triggered whether from a button or VR controller, a hand lifting animation arises and leave the hand and arm in the air. The hand loops over an animation of slight movement in the air until putting hand down event is triggered. The lecturer is able to perform three different talking gestures to emphasize the sentence and attract students' attention that are controlled by a button. While the lecturer is presenting a specific object, she turns to the object and performs a presenting gesture, that is pre-programmed into a canvas button as well. Inside the Supermarket Scene, three of the supermarket staff sit at checkout counters and one staff stands in front of a storage rack as if he is stacking the shelf. All animations are sourced from Mixamo [77] in FBX format under 24 frames per second frame rate that can be directly applied on humanoid avatars in Unity. Figure 4.7 below shows an example of animator controller of the student avatars that controls the animation flow, transition and trigger for the avatars.

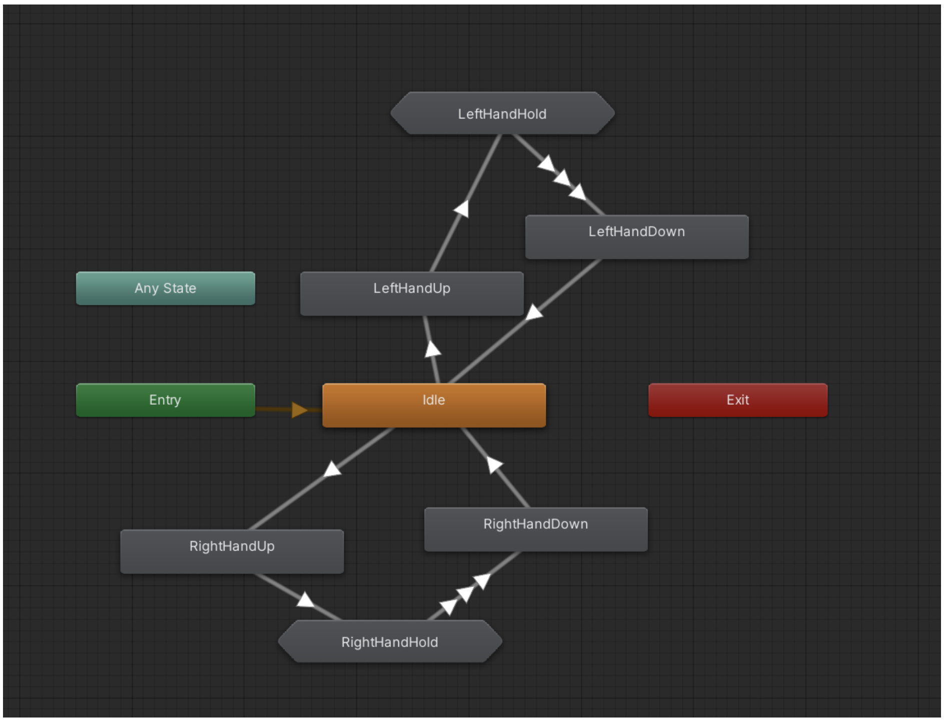


Figure 4.7: Animator Controller of student avatars

Moreover, the actors can turn their body towards a specific object when it is mentioned by the lecturer. This is done by rotating the avatar along y axis so as to point their forward transform to the object. To make it more natural, a random short delay between one to two seconds was introduced so that not all actors turn simultaneously. The VR player can also turn their body around as orientation of the avatar is at all time aligned with VR camera.

4.4 Research Features

4.4.1 Hand Raising Feature

The VR student participant is able to raise hand to give response to lecturer's questions or as a signal of a desire to speak. The heights of both VR controllers are constantly detected when in the lecture scene, and once a controller's height exceeds a preset threshold, we regard the player raises hand.

In augmented condition, it immediately triggers corresponding hand raising animation and displays the augmented arm as shown in the left image in Figure 4.8. The image on the right side provides view from another angle to better illustrate the actual positioning of the additional arm and actual arm. Again the additional arm is not visible by other system endpoints. Note that the top right participant list is not visible in VR. The threshold for controller height is set at 85% of the participant's height. Once the height comes below the threshold, we consider the player puts hand down, hence trigger hand down animation and hide the augmented arm. The hand raising feature applies to both hands.

In baseline condition, hand raising is indicated by an icon in front of the avatar at chest height, next to the name tag as in Figure 4.9.

Figure 4.8: Left: Hand raising effect in augmented condition from VR player's perspective; Right: Arm positioning of augmented hand raising

The lecturer and student actors can carry out both forms of hand raising behavior controlled by canvas User Interface (UI) buttons.

4.4.2 Gaze Feature

The gaze direction of the lecturer while *addressing all* students is controlled by a script running only on the client's side that is not synced across network. In other words, lecturer's gaze behavior is not backed up by real gaze of the lecturer. Once the VR student is spawned into the lecture scene, the script starts running at the VR client's side locally to control lecturer's gaze target following the rule defined in Chapter 3.3.2. When the lecturer listens or speaks to an individual, *WizardCanvas* UI buttons are used to set a focused target across network. The buttons can be a single function button only used for this or together with controlling an actor to speak. The lecturer stays focused until an *Unfocused* button is clicked to signal the client to return back to "addressing all" type of gaze.

Figure 4.9: Hand raising effect in baseline condition

The lecturer avatar's animator *SetLookAtPosition* function is used to allow the lecturer to rotate head without rotating the whole body. The *LookAt weight* is set to 1.0 which means the avatar looks fully at the targeted direction.

The lecturer may point to a certain object in the scene to help visualize lecture content. This behavior is also controlled by buttons, and once triggered, the lecturer turns her body towards the pointed object while performing a pointing gesture. Meanwhile LookAt weight is lowered down to 0.6 so that the lecturer gazes in between target and pointed object. Once a *Finish Pointing* button is clicked, the lecturer turns back to original standing position and *LookAt weight* is set back to 1.0. Figure 4.10 shows a first-person perspective of the VR participant when he or she is gazed at by the lecturer.

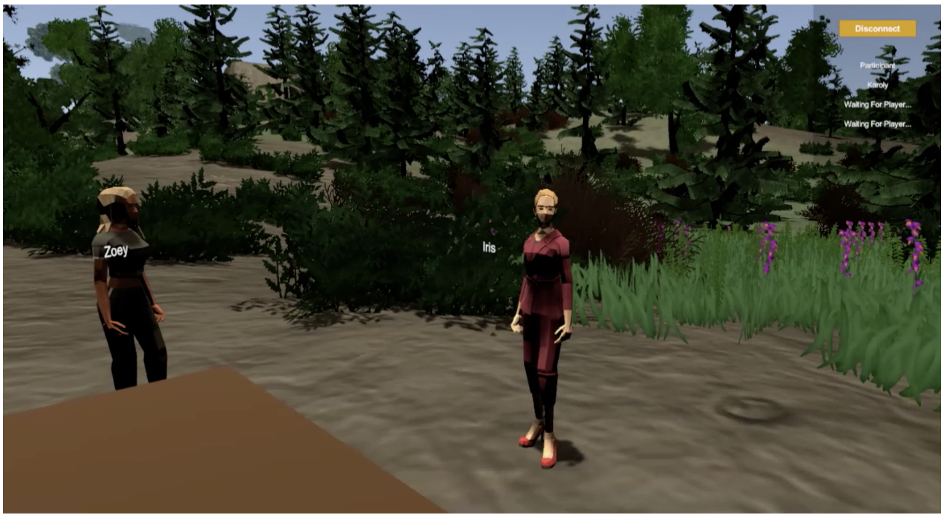


Figure 4.10: VR participant's first-person perspective while being gazed by the lecturer

Chapter 5

Evaluation

This chapter presents evaluation method and obtained results for HejVR regarding research questions raised. The evaluation consists of two parts: a preliminary pilot study and a main study. Both studies were carried out in the same form but several changes were made in the system and evaluation method after the pilot study.

5.1 Study Design

The study consists of a quantitative experiment and a qualitative interview. The experiment adopts a within-subject design comparing the augmented condition and the baseline condition. Each participant was asked individually to experience both of the conditions in two separate virtual lectures followed by a post-study questionnaire immediately after exposure to each lecture. The VR lectures are counterbalanced in order to avoid the effect of differences in lecture content and the order of exposure, resulting in four experiment scenarios as shown in Table 5.1.

Scenarios	Order of conditions	
Order of lectures	Supermarket scene augmented + Berries scene baseline (Scenario A)	Supermarket scene baseline + Berries scene augmented (Scenario C)
	Berries scene augmented + Supermarket scene baseline (Scenario B)	Berries scene baseline + Supermarket scene augmented (Scenario D)

Table 5.1: Experiment scenarios

The interviews are semi-structured with following core set of questions:

1. How was your experience?
2. Do you feel any differences in the 2 lectures?
3. Which lecture do you prefer? Why do you prefer this lecture?
4. How do you feel about the people around you?
5. Have you ever had such lectures online in video form? How do you feel about them?

If any unexpected behaviors were observed during the experiment, we also asked them to explain the reason behind.

5.1.1 Procedure

Before the experiment took place, participants first read an information brochure with an introduction of the study as given in Appendix B, and a instruction of how to put on HTC Vive pro HMD. They also gave their informed consent (original form can be found in Appendix C) at this stage. Then participants were equipped with the HMD, controllers and received oral instructions of logging in, menu selection, and to raise their hands if they have any questions at any time.

The lectures started once the participant was ready and each lecture session lasted about 10-15 minutes. After each lecture the participants were asked immediately to fill out the post-experiment questionnaire. A short break was given between two lectures. The lectures were recorded with screen-recording software from both the server and client sides and additionally a camera was used to record user in the actual physical experiment space.

After completing both lectures and questionnaires, the participant was given another short break and followed by the interview recorded in audio. The participant received chocolates and candies in separate packaging after the entire study procedure.

5.1.2 Measurements

In order to get an indication of users' experience in terms of presence and engagement, we chose to use existing validated questionnaires. The concept of presence has yet not to have an agreed definition but two commonly referred definitions are "the sense of being there" in the provided virtual environment [78] and "the perceptual illusion of non-mediation" [79]. The two definitions align and presence has been recognized as an important aspect for virtual reality systems. Besides, feeling oneself being in the lecture environment is desired in our use case which we believe to help the students to better focus and absorb the information with related environment. Three main measures for presence include subjective means using post-immersion questionnaires and objective means by behavioral and physiological measures [80]. Taking into account that this research is still in very early stage and the environment is not fear or stress-inducing, we decide to use questionnaire to measure presence. Presence Questionnaire (PQ) by Witmer and Singer [81] and ITC-Sense of Presence Inventory (ITC-SOPI) [82] questionnaire both found themselves reliable by large scale user studies and commonly used among researchers.

Student engagement is positively linked to their achievement, persistence and retention. Although there lacks a definition of student engagement but three widely accepted dimensions of student engagement contains affective, cognitive and behavioral [83]. Higher degrees of attention, interest, passion are regarded as indications for higher engagement. Many measures for student engagement are based on self-reported behaviors over an entire course period such as [84] and [85], which are not suitable in our case. But ITC-SOPI also considered engagement as one contributing factor to presence. They measure engagement as the tendency of a user to feel psychologically involved and to enjoy the content which we think is a suitable measurement as student engagement in the lecture.

In our study, 4 pilot tests were done with PQ as post-experiment questionnaire and 1 pilot participant used ITC-SOPI. We decided to use ITC-SOPI for our main study due to the fact that the participants raised doubts to some of the questions in PQ, many questions are not able to differentiate between our two test conditions and we are interested in the engagement factor that ITC-SOPI is able to measure.

Apart from the questionnaire data, the number of times participants raise their hand and a total amount of time they speak in a certain lecture were extracted from video recordings as an objective indication of their behavioral engagement.

5.1.3 Special measures regarding COVID-19

For the health of our participants and experimenter, special measures were taken to minimize the potential spread of COVID-19

Air circulation

The entire experiment was placed in one room. To provide fresh air, windows in this room should remained open as long as possible. They were opened from preparation until right before the experiment started. During the experiment, windows were closed. When the experiment was finished, windows were re-opened immediately.

Experimenter

While coming from outside, experimenter first washed hands with hand soap and water. The experimenter disinfected hands with a hand sanitizer regularly throughout the session. The experimenter kept distance to the participants.

Participants

When called for an appointment, participants were asked if they or anyone in their close network had any Covid-19 symptoms during the last 14 days, whether they had contact with someone with a confirmed case or whether they had stayed in an area for which an official travel warning exists. If the answers were positive, they were not invited to the experiment.

In this call, participants were informed about the safety measures. In the experiment room, they signed a confirmation sheet indicating that they were symptom free. They were also asked to inform us in case they observe symptoms in the next week after their visit.

Participants were picked up from the main entrance at the appointed time. They were offered hand sanitizers at the entrance of the building. Additionally, they were given the option to wear a face mask that covers their mouth and nose. They may wear their own mask or get one from the experimenter. They were provided with soap and hand sanitizers inside the experiment room as well. One hand sanitizer was placed close to the participant area.

Distance

1.5 meters distance between people in the room was maintained at all times. In the room, there were areas designated for experimenters and participants. The segmentation was marked with tapes on the ground.

Cleaning and disinfecting

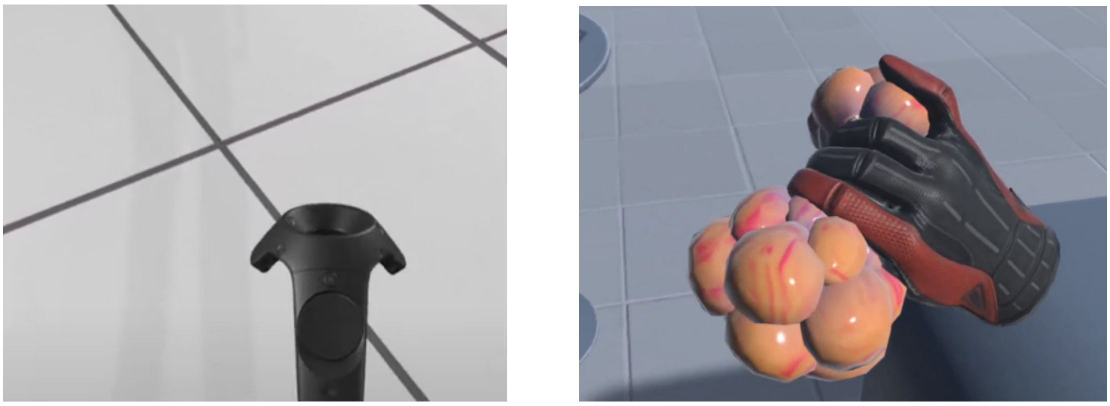
VR headsets were disinfected after each session. Keyboards, mouses, tables, pens, all the materials that were touched by participants were disinfected.

Time in between testings

To ensure enough time for cleaning and air circulation in between testings, at least 30-minute interval was allocated.

5.2 Pilot study

The purpose of the pilot study is to evaluate the operability and feasibility of the experiment design, assist system design decisions and identify issues in the system and procedure. 5 participants with previous VR experiences from KTH Royal



(a) Controller displayed as controller

(b) Controller displayed as hand (Taken from [86])

Figure 5.1: Different ways to present controllers in the scene

Institute of Technology were invited in this stage. The study went through the same procedure as the main study. The first two participants experienced an additional phase before the first lecture of two different versions of VR controller presentation. This section first presents overall results of the pilot study and then explains bugs and improvements identified and being changed in the system for the main study.

5.2.1 Results

In the first two sessions, we compared two different ways to present controllers in the scene. One way is to display them just as controllers, the same look they are in reality as shown in Figure 5.1a. The other way is to present them as hands as shown in Figure 5.1b, since the controllers are able to track thumb and index fingers of the player if they are held properly. In this way the movements of a player's hands are better presented. However, both participants preferred the first design because the avatar has arms and having an extra pair of hands was felt bizarre. Because the whole experience does not involve much interaction from the controllers except for raising hand, we chose to present them just as controllers and tell the participants to relax their hands during the experiment.

Additionally, we also tested the hand raising threshold at 85% of a participants' height and all desired hand raising were successfully triggered and no unwanted behavior happened.

5.2.2 Bugs and improvements

One major issue discovered was from the integrated headphone of HTC VIVE pro HMD. The sound became distorted after starting SteamVR which is a commonly reported issue for this type of headset. On account of the inability of access to other hardware, we used a pair of AirPods instead for audio output as presented in Chapter 3.1.1.

Some game objects were not correctly synced across network which was fixed before main study. Speed of the lecturer's recorded teaching script audio was also increased to be more natural. Name tag canvas position in front of avatars was adjusted so that name text and hand raising icon are clearly visible from both first and third person perspective. And lastly, a script was added to simplify the process of changing participant's name and avatar's height before every test session.

5.3 Main Study

The main study focused on answering the research questions: **RQ1** *Are avatar-based online VR lectures suitable for cultural education?* and **RQ2.1** *How do augmenting gaze at the learners' side and embodying hand raising gesture influence learners experience?* A series of within-subject experiment following the procedure as defined above was carried out in Stockholm.

5.3.1 Hypotheses

We constructed the following hypotheses with the given expectations.

- **H1.** Virtual Reality is beneficial for online cultural education.
- **H2.** Augmenting gaze at the learners' side and embodying hand raising gesture can increase the sense of presence and engagement of learners.

For Hypothesis 1, we expected a positive feedback to be reported during interviews regarding the participants' experience in both lectures as we provided a learning environment closer to face-to-face compared to video lectures and the scenery is highly related to learning content.

For the two augmented features, hand raising was expected to contribute to *spatial presence* with partial embodiment and more prominent hand raising behavior from other avatars in the scene. Augmented gaze was expected to make the participants sense higher attention from the lecturer, pay more gaze back to the lecturer and hence pay more attention to the lecture content, resulting in higher engagement. But due to the time and resources constrains of this project, we did not conduct separate experiments for the two features, in other words, we only measured their collective contribution.

5.3.2 Participants

Due to the lecture content and the nature of the study, there was some screening requirements for the participants. The participants should not have a history of motion sickness or epilepsy because VR has the possibility to cause unpleasant feelings, nausea or even severe side effects. The participant should have a reasonable sight (or with the help of glasses) and hearing to be able to experience the lectures fully. The participants need to be proficient in the English but know little about Swedish and Sweden so that they could understand the lecture content while maintaining the novelty of the lectures.

A total of 13 participants, 3 female and 10 male, who are all new students in KTH Royal Institute of Technology as part of EIT Digital masters program joined the study. The mean age is 24.6 (sd=2.2, max=29).

5.3.3 Data Processing

This section explains how different sources of data was processed after the experiment. There are three main sources of data: ITC-SOPI questionnaire data, experiment video recordings and audio recordings taken during the interview sessions.

The ITC-SOPI is a questionnaire that is intended to reliably measure cross-media presence. It consists of 3 main sections: 1. background variables related to demographics, education, self-rated experience with computers, computer games VR, TV/film production etc.; 2. 6 questions about the user's thoughts and feelings once the displayed environment was over; 3. 38 questions about thoughts and feelings while the user was experiencing the displayed environment. The completed questionnaire reveals 4 factors: spatial presence, engagement, ecological validity / naturalness and negative effects. All 44 questions in section 2 and 3 have a five-point Likert scale (1 = strongly disagree; 5 = strongly agree) response option and the 4 revealing factors are calculated by a mean of all completed items contributing to each factor.

The ITC-SOPI is used for testing **H2**. Firstly, Cronbach's Alpha value was calculated for each factor with all their contributing items to measure the internal consistency reliability of the data we collected. Since the sample size is not large enough, a Shapiro Wilk normality test was then performed on the data. On account of the within-subject design nature of the experiment, a paired t-test was used if the factors can be regarded as normally distributed to compare the two experiment conditions - baseline and augmented - of the same participant. All questionnaire data are processed using RStudio Version 1.3.1093 with R version 4.0.3.

The number of times participants raise hands was manually counted by re-watching experiment recordings. We extracted audio from the video recordings, marked out the time a participant spoke by looking at sound wave and summed up across the entire lecture. Lastly, interview recordings were transcribed into text. For data analysis, analytical coding was applied onto the transcription and themes were defined based on the codes.

5.3.4 Quantitative Results

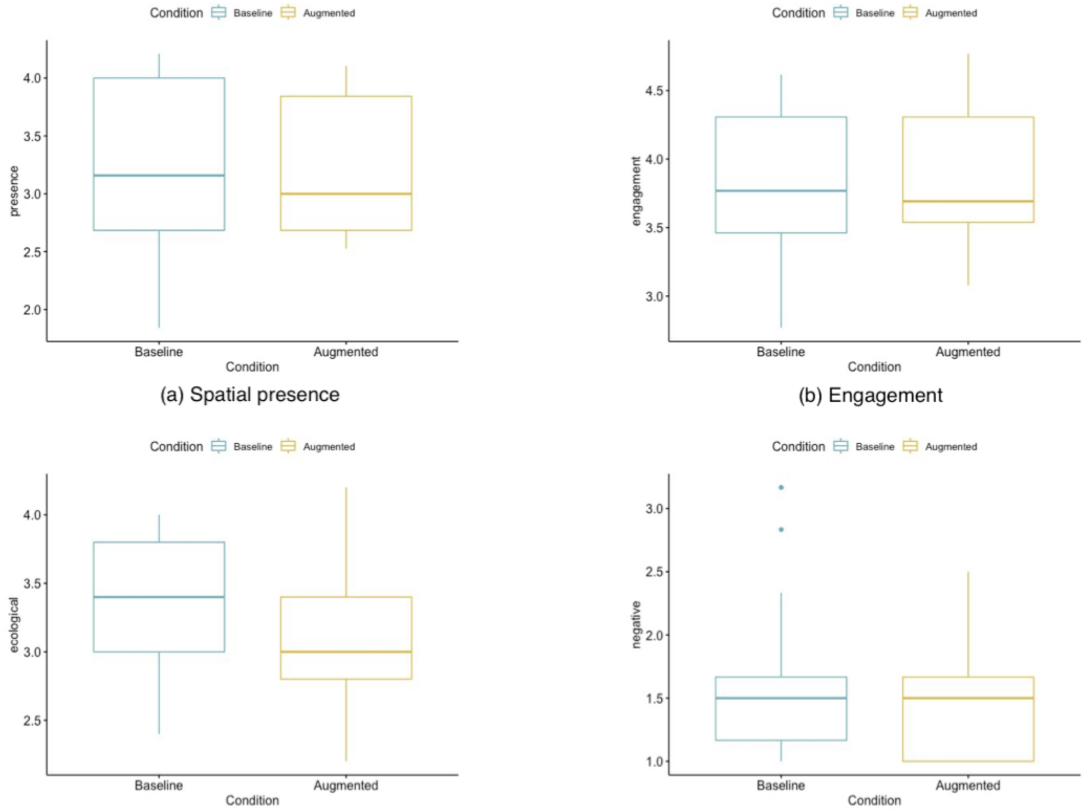
The results of Cronbach's Alpha presented in Table 5.2. Spatial presence, engagement and negative effects all gives Cronbach's Alpha value higher than 0.70 meaning that we can accept the data we collected for these three factors are reliable as our research is a exploratory research. But the data for ecological validity dimension is not acceptable reliable.

Factor	Alpha reliability	Standardized alpha
Spatial Presence	0.890	0.890
Engagement	0.815	0.820
Ecological Validity/ Naturalness	0.524	0.505
Negative Effects	0.781	0.769

Table 5.2: Cronbach's Alpha result for different factors

In Shapiro Wilk normality test, spatial presence ($p = 0.097$), engagement ($p = 0.624$) and ecological validity ($p = 0.438$) factors result p-values greater than 0.05, which indicates that we can assume these factors are normally distributed. Factor negative effects has a p-value of 0.003 suggesting that the sample is not normally distributed. We think it is a good indication that the perceived negative effects are quite low for our system. The corresponding boxplots for each factor in augmented and baseline conditions are presented in Figure 5.2. For both spatial presence and engagement, students' scoring are more densely distributed in augmented condition but the difference is not significant. Reported negative effects factor is quite low in both conditions, with two outliers in second order lecture in baseline condition. According to the participants' report, the reason for the negative feelings are eyestrain and standing too long respectively, which are not directly affected by our system design.

Since spatial presence and engagement can be accepted as normally distributed, paired t-test was conducted on both of them, giving the result $t(19) = -0.028$, $p = 0.978$ for spatial presence and $t(13) = -0.352$, $p = 0.731$ for engagement factor. The p-value given from both factors are way greater than 0.05 indication the difference between the two paired samples are not significantly different. This suggests that our augmented condition does not prove to increase self-reported sense of spatial presence and engagement.



(a) Spatial presence (b) Engagement

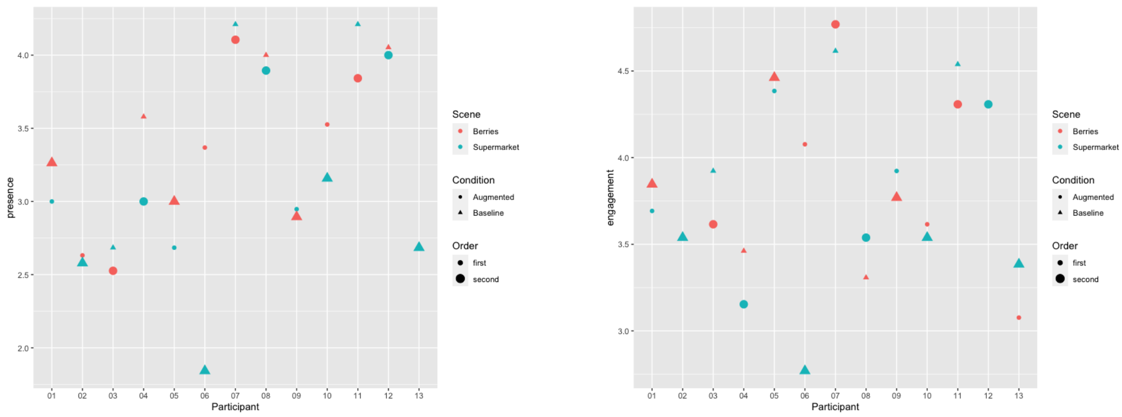
(c) Ecological validity / naturalness (d) Negative effects

Figure 5.2: ITC-SOPI four factor scores in different test conditions

We then took a closer look at the data to identify if other independent variables including lecture order and lecture scene contribute to the result. Figure 5.3 presents the individual response of all participants under different lecture scenes, conditions and orders.

In terms of spatial presence, 1 participant gave equal score to both lectures, only 4 out or 13 tests have higher score in augmented condition but as the paired t-test result already explained the difference is not significant. For the lecture content, 8 out of 13 participants gave a higher presence score in berries scene. For lecturer orders, 10 of 13 first order lectures have a higher score. A paired t-test was also executed on these two dimensions giving the result for scene difference: $t(19) = 1.522$, $p = 0.154$; and lecture order: $t(19) = 1.674$, $p = 0.120$. Although the p-values are still greater than 0.05, they are much lower than the difference between different conditions and the lecture order does have the most prominent impact for spatial presence, indicating our experiment has a strong order effect.

As for engagement, 2 participants gave equal score, the higher scores for different conditions (augmented over baseline), lecture scene (berries over supermarket)



(a) Spatial presence *(b)* Engagement

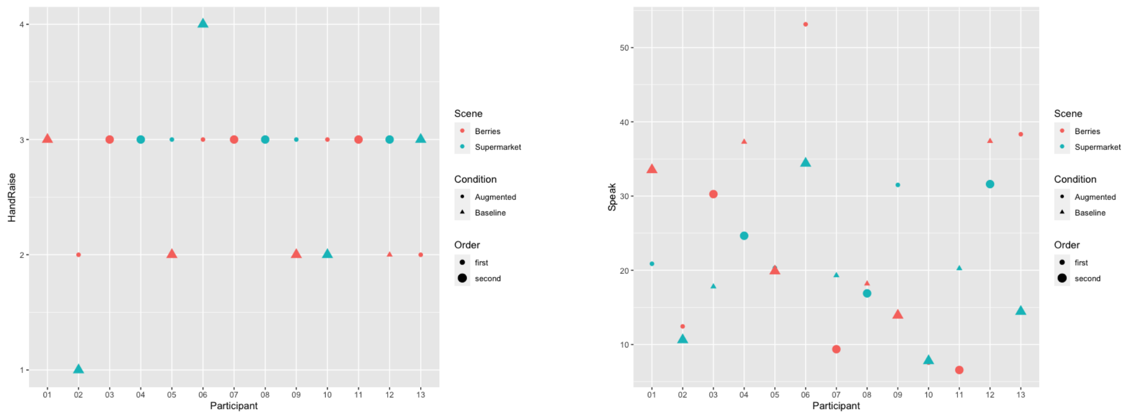
Figure 5.3: Individual response of spatial presence and engagement in different lecture scenes, conditions and orders

and lecture order (first over second) are 5, 6, 6 out of 13 participants, respectively. The engagement score is evenly distributed across all three dimensions, and the paired t-test result also proves that (lecture scene: $t(13) = 0.557$, $p = 0.587$; lecture order: $t(13) = 0.988$, $p = 0.346$).

Moreover, the amount of hand raising behavior (Figure 5.4a) does not manifest correlation with any of the conditions. This is likely because our lecture content is mainly informative, and the participants only raise hands when they conforms with or know the answer of the lecturer's question. None of the participants raised hand to ask additional questions in the entire experience. 10 out of 13 participants talked more in the first order lecture they experienced, although the difference is not significant. This might result from nervousness, unfamiliarity of the experiment procedure, environment, lecturer and fellow students. Total amount of time a participant speak ranges from 8 to 55 seconds, see Figure 5.4b.

5.3.5 Participant Behaviors

We observed a number of positive behaviors from the participants while reviewing video recordings. The majority of the participants looked at other student actors when they talked and gazed at the lecturer most of the time. This is considered as a good indication of them focusing on the lecture and paying attention to their fellows. However we did not program the actors to gaze at the participant which may have broken the sense of presence and their feelings about the actors. Many people performed nodding and head shaking during the lectures, which also indicates that they were paying attention and engaging. This could be an interesting feature to consider in future design because it provides valuable feedback to the lecturer. Furthermore,



(a) Number of hand raising *(b)* Amount of time speak

Figure 5.4: The number of hand raising behavior and amount of time each participant speak in different lecture scenes, conditions and orders

a few people repeated Swedish words after the lecturer during the experiment which is an additional sign of engagement.

Whereas, a few unexpected behaviors were discovered as well. The experiment setting was confusing for several participants in the beginning that after they raised hands, the lecturer indicated them to answer the question but they were unclear about how to interact with the lecturer. “Do I need to answer”, “What should I say”, “I have to answer I guess” were said by some participants. But they adjusted quickly and the rest of the experiment went smoothly. Many people tried to touch objects around and swung their arms (controllers). As we did not implement any collision feedback, the sense of presence may be negatively influenced. Some participants spent a certain amount of time looking around after entering the lecture scene, some also did that later. This might reduce the amount of time they were able to sense the gaze of the lecturer thus influence the experiment result. A few people kept their hand up in the air for quite some time when other actors were speaking or even when themselves were speaking. That was due to the fact that the wizard did not put hands of actors down until a certain actor finished speaking and the participants kept their hands raised following the actors. But this behavior is unnatural in a real world classroom. The control of actors needs to be adjusted in the future. And lastly several people first looked around for a few seconds and after finding that other actors have not give reactions, they raised their hands to answer a question. We think the use of actors helped the participant to involve more in the lecture in some ways.

5.3.6 Interview Results

In the interview, we first asked the participants how their experiences were. The overall feedback was positive, they described the experience as interesting, fun, enjoyable, immersive, they felt ”drowned” in the scene, the experience was interactive. They additionally enjoyed the graphics and visualisation of objects even though they are low poly.

As displayed in above quantitative results, engagement score is evenly distributed across lecture scenes, conditions and order of the lectures, meaning that none of the dimensions shows strong correlation to the engagement score. What could possibly influence the engagement score? During the interview, we asked the participants which lecture they prefer, and the result suggests 8 participants’ preferences align with the higher engagement score they gave for a lecture, with 2 equal scores resulting in only 3 people’s preferences do not corresponds to the higher engagement score. This suggests that engagement could be a very personal sense. Reasons behind the preference were also inquired. For the 6 participants who preferred lecture about Swedish berries, their given reasons include the background being more dynamic because the grass and branches were moving slightly, in fond of outdoor environment, the content being unique and inspiring, or just more interested in culture than language in general. For the 7 participants preferring lecture inside supermarket, the most frequently mentioned reason was the content being useful, relevant to their life status, they learned something that can be used daily. One person mentioned it being more interactive because the lecturer asked them to turn around to see an object located behind, causing more movement for the participant than the berries scene.

Moreover, we asked their feelings of other people in the scene. All participants had a positive attitude towards the lecturer, they thought of her being real, helpful and responsive, ”like real person I was interacting with”. The supermarket staff were thought to be weird because they were static in the scene. As for the actors, the majority thought they are not real, described them as AI or characters in games. One person explicitly stated he would prefer 1 to 1 class without the actors because he does not know how to treat them. However the rest did not mind having them in the scene to provide a classroom vibe, and the audio clips actually made them feel kind of real. “It was nice hearing their answers,” one participant said. In contrast, 3 people felt the actors quite real just like normal people, one person said he thought they are students connect remotely and he would like to connect with his fellow students. Another participant noticed that while being asked the reason for coming to Sweden, one actor answered ”study, as an exchange student ” and another actor followed with ”same”, as if they understood each other. This specific event greatly improved the realness of the actors.

Last but not least, we asked the participants their opinions of the two research features. For the hand raising feature, 4 people thought it did not make a huge difference for them. One person who had previous experience with VR and kinematic projects preferred the icon (baseline version) because he thought kinematic tracking is tricky, he would prefer just using a simple icon unless the alignment is perfect. But 8 participants, the majority of people preferred the augmented version giving the reason that seeing the arm rising felt more real, natural and more involved. For gaze, 7 people did not notice the difference. It may cause by them paying too much attention on the surroundings, but on the other hand it implies our augmented gaze implementation being quite natural. One person stated the augmented condition was a bit disturbing, thought “why does she keep looking at me” and said baseline was more comfortable. But 5 other people had quite positive attitude to augmented gaze, some felt it more natural and involving than baseline condition. One said it did not make him nervous as being looked at by a teacher in real life. Another said although it made him feel not as free as in baseline condition, it helped him to pay attention to the lecture.

Chapter 6

Conclusions and recommendations

We presented HejVR, a a Virtual Reality online cultural learning system. The system has been designed to simulate cultural study excursions and improve students' learning experience with augmented features on avatars. A preliminary user study was conducted to compare the augmented condition to baseline condition. This chapter consists of four sections. First we conclude answers to the research questions: **RQ1** *How to create a VR system beneficial for avatar-based online cultural lectures?* and **RQ2.1** *How do augmenting gaze at the learners' side and embodying hand raising gesture influence learners experience?* Then we discuss interpretations from this research, its indication to industry and learnings from research process. Limitations of the study are then presented and finally we conclude recommendations for future work.

6.1 Conclusions

RQ1 *How to create a VR system beneficial for avatar-based online cultural lectures?*

RQ1.1 *How to create a VR system for online lectures?*

The system was created entirely on top of Unity making use of Mirror Networking open source API as a base for the networking system and SteamVR plugin for VR support. The lecturer (host) side is completely platform independent that can be built for all Unity supporting platforms. The student (client) side supports headsets that is supported by SteamVR including Valve Index, HTC Vive, Oculus Rift and Windows Mixed Reality. Using only commodity hardware makes the system accessible and easy to setup.

The system has been made easy to use with simple interactions including button clicking for menu selection and raising controllers for raising hand. The student are free to move around in the space and talk via headphone.

RQ1.2 *What do avatars look like and what kinds of behaviors should they have?*

Low poly anthropomorphic avatars dressed in ordinary garments are used in HejVR. There are a few different avatars free to be chosen by the participants. All avatars have neutral facial expressions and basic animations for idle standing and the ability to rotate the entire body to another direction. The lecturer avatar has more expressive behaviors including head gaze, pointing to objects and talking gestures.

We did not compare with other types of avatar representations but the participants did not show negative attitudes towards avatars being used in the system.

RQ1.3 *How to create VR environment for cultural lectures?* To create VR environment for two designed lectures, scenes and objects from public libraries in Unity Asset Store were used and manually placed in appropriate setup for our excursion simulation. Several objects mentioned by the lecturer were created in Blender for teaching illustration.

RQ2.1 *How do augmenting gaze at the learners' side and embodying hand raising gesture influence learners experience?*

We proposed two features on avatars in our system: partial embodied hand raising with animations and augmented gaze for lecturer. From the quantitative result of the user study, augmenting lecturer's gaze at the learners' side and embodying hand raising gesture do not prove to increase the sense of presence and engagement of learners. The augmented condition scores slightly lower in both spatial presence and engagement factor in ITC-SOPI although the difference is not significant.

However as we have seen, the presence factor has a strong order effect that most participants feel more present in the first lecture they experienced. As the two lecture settings are relatively similar and exposure time is quite long, the participants may feel bored after some time or the second VR experience felt less novel, thus resulting in lower score. On the other hand, differences in our research conditions may not be significant enough to influence spatial presence. The engagement factor aligns highly with participants' preference in the lecture. The two lectures have many differences in content and environment, which lead to the inability to know for a fact if the result is due to the augmented features we introduced. Besides, although not unanimous, people in general gave a more neutral to positive attitudes towards both features and their given reason corresponds to our intention behind these features. Further study needs to be conducted to better identify the individual contribution and direct influence of the two features.

6.2 Discussion

Above result suggests Virtual Reality is likely to be beneficial for online cultural education. Based on interview feedback after the experiment, all participants were positive about their experience, felt it interesting and learned something useful. We also asked about their experience in online education that everyone is going through. Aligned with our previous interview while defining the research topic, they further confirmed the flexibility as a benefit but found it difficult to focus, needed more effort to pay attention. They thought their VR experience during the experiment was more interactive and they were able to pay more attention to the lecture.

A few participants also took the Swedish cultural course in KTH where the teaching content is based on. The course is at the moment fully online for both language and cultural content, hence no excursion or field study is taken place. For the language part, the lecturer teaches Swedish words and phrases through video calls and ask the students to repeat after her. But the participants reported this being awkward because they felt very distanced and the lecture was too silence since all students were muted. On the other side, several participants repeated the phrases after the lecturer and even used some phrases to interact with the lecturer during the experiment without the lecturer even explicitly asking that. We think this is a great indication of them being engaged in the lecture. For the cultural part, these students tend to skip the lectures because they do not find the motivation to join these online. We do not know if VR can solve the motivation, but according to their feedback they found the more cultural oriented Swedish berries lecture interesting and provides a dynamic natural environment with good graphics. As a result, we believe our system is able to provide a more attractive learning environment for more informative cultural related content, and an opportunity for teachers to conduct excursion lectures remotely.

This research not only reveals new possibilities in remote online lectures, but also opens up new use cases for industries. There is a great potential to use this system for enterprise internal training, employee on-boarding, and external branding, since a brand is not just a logo or the products or services it provides. The entire character a brand stands for, the combination of vision, culture, and images needs to be correctly communicated with both internal customers (employees) and external stakeholders, to assist a company's success [87]. Especially at the time when in person contact, site visit are difficult, virtual reality can play an important role in communicating the abundant information with its various interaction modalities.

Besides, it was noticed that large scale online lecture may reduce motivation of the students where the listeners are often muted and have their cameras off. In the interviews, participants mentioned that some teachers tried to use breakout rooms function in video communication platforms to divide students into small groups. However the solution is not very effective because every switch to those rooms needs an adaption period and when conversations become more intense they are forced to return back to the main room. Our system can be used to solve this problem by optionally rendering a limit amount of students, so that on each client side the VR student see the lecturer and only a handful of fellow students that the possibility of them interact with each other becomes higher.

Furthermore, this system can help our research partner, Ericsson, to discover business requirements. The immersive experience enabled by VR devices comes with further technical challenges. As a network provider, Ericsson can use the system to identify edge computing possibilities to support graphical rendering processes that lower the on-device workloads on VR devices [88], and network requirements with 5G for fast and reliable data transmission to achieve low latency, high quality experiences.

The measures we took to conduct face-to-face experiment during the pandemic successfully prevented virus spreading among experimenter and participants. None of the people involved in the experiment experienced any symptoms or was tested positive in the following week. These measures can be taken as reference for other user studies.

Lastly, making use of pre-recorded actors to simulate a multi-participant scenario can be a useful method to consider in conducting user research. However, we must keep in mind that the "realness" is easily breakable. If the actors are supposed to be in the same position as the participant, their behaviors must be kept equal to the real participants to a high extent. We can potentially improve the realness by carefully design actor script to make their speech and behaviors connected.

6.3 Limitations

There are a few deficiencies in both technological and experimental aspects. First of all, avatars are constrained in standing position with limited behaviors. Avatars also do not have facial expressions or lip movements when they talk. Students can only perform raising hand gesture and the lecturer has additional talking gestures but there is a lot more a person would do in a face-to-face communication. The students are able to turn around the whole body but not able to change gaze direction without turning their body. One potential solution is to combine tracking from HMD with the controllers to decide if a VR player is shifting gaze direction or turning the whole body.

Secondly, the VR controllers are not aligned with avatar's hands that when the player move hands by moving controllers, avatar remains in idle position unless we regard the participant raises hand. This misalignment can cause a reduced virtual body ownership.

In the experimental aspect, the scale of the user study is very limited. The quantitative data failed to show significant difference and we are not able to provide decisive conclusion on top of that. There also lacks a direct comparison against video conference platforms.

In this research we designed, implemented and tested HejVR system as a whole under two conditions, augmented and baseline. According to participants' feedback, the following potential improvements are summarized. To start with, although the lecturer explicitly gazed at the participant and said his or her name to indicate a permission to speak, she did not do that to all other student actors. This might have led to the confusion of the participants that who is supposed to speak. The lecturer could provide additional indication such as pointing gesture and say the name of all students as a signal to make the process clearer and more consistent.

Besides, the surrounding staff in the supermarket were static throughout the experiment. This was reported to be weird by the participants. Actions that conform to the role of the avatars needs to be added. If living creatures are introduced in a scene, movements must be considered as well.

Furthermore, spatial presence factor seems to be influenced by order effect, which can possibly be solved by adding a pre-exposure session in the beginning of the experiment to let the students get familiar with the environment and interaction modalities. This may as well help the students to focus on the lecture in the actual experiment stage and reduce risk in the procedure.

6.4 Future Work

Current experiment design blends hand raising and gaze control as one influential factor, however it is possible that they are influencing user experience in opposite direction. Future study should identify their individual contributions in respective factors.

It can be seen that the lecture content and VR environment also affect students' preference and engagement. To eliminate this effect, larger scale between-subject study under the same lecture needs to be conducted.

At present, the system is fine tuned for the experiment design and for the HTC Vive hardware we are using. A next step would be to modulate features according to general lecture needs to simplify designing new lectures in other scenes. Beyond that, with the various choices of VR headsets on the market, we must provide support to different platforms to reach a larger audience and achieve the intention of making remote communication easier.

Lastly, our research features, and observed behaviors of the participants provide valuable design considerations in developing future remote communication systems. It must be emphasized that we are not trying to replace all means of remote communication by VR, but to assist and improve the experience when possible.

Bibliography

[1] Z. Fang, “Master thesis proposal - extended reality in immersive telecommunication,” 2020, unpublished, University of Twente.

[2] S. Orts-Escolano, C. Rhemann, S. Fanello, W. Chang, A. Kowdle, Y. Degtyarev, D. Kim, P. L. Davidson, S. Khamis, M. Dou, V. Tankovich, C. Loop, Q. Cai, P. A. Chou, S. Mennicken, J. Valentin, V. Pradeep, S. Wang, S. B. Kang, P. Kohli, Y. Lutchyn, C. Keskin, and S. Izadi, “Holoportation: Virtual 3d teleportation in real-time,” in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, ser. UIST ’16. New York, NY, USA: Association for Computing Machinery, 2016, p. 741–754. [Online]. Available: https://doi.org/10.1145/2984511.2984517

[3] M. Dou, P. Davidson, S. R. Fanello, S. Khamis, A. Kowdle, C. Rhemann, V. Tankovich, and S. Izadi, “Motion2fusion: Real-time volumetric performance capture,” *ACM Trans. Graph.*, vol. 36, no. 6, Nov. 2017. [Online]. Available: https://doi.org/10.1145/3130800.3130801

[4] Z. Zheng, T. Yu, H. Li, K. Guo, Q. Dai, L. Fang, and Y. Liu, “Hybridfusion: Real-time performance capture using a single depth sensor and sparse imus,” in *European Conference on Computer Vision (ECCV)*, Sept 2018.

[5] M. Habermann, W. Xu, M. Zollhoefer, G. Pons-Moll, and C. Theobalt, “Livecap: Real-time human performance capture from monocular video,” *ACM Transactions on Graphics (TOG)*, vol. 38, no. 2, pp. 1–17, 2019.

[6] J. Thies, M. Zollhöfer, M. Stamminger, C. Theobalt, and M. Nießner, “Facevr: Real-time facial reenactment and eye gaze control in virtual reality,” *arXiv*, pp. arXiv–1610, 2016.

[7] G. Song, J. Cai, T.-J. Cham, J. Zheng, J. Zhang, and H. Fuchs, “Real-time 3d face-eye performance capture of a person wearing vr headset,” in *Proceedings of the 26th ACM International Conference on Multimedia*, ser. MM ’18. New York, NY, USA: Association for Computing Machinery, 2018, p. 923–931. [Online]. Available: https://doi.org/10.1145/3240508.3240570

[8] C. Frueh, A. Sud, and V. Kwatra, “Headset removal for virtual and mixed reality,” in *ACM SIGGRAPH 2017 Talks*, ser. SIGGRAPH ’17. New York, NY, USA: Association for Computing Machinery, 2017. [Online]. Available: https://doi.org/10.1145/3084363.3085083

[9] Vrchat. [Online]. Available: https://vrchat.com/

[10] Hubs - private social vr in your web browser. [Online]. Available: https://hubs.mozilla.com/

[11] Altspacevr — be there, together. [Online]. Available: https://altvr.com/

[12] Facebook horizon — oculus. [Online]. Available: https://www.oculus.com/facebookhorizon/

[13] Meetinvr - business meetings & collaboration in vr. [Online]. Available: https://meetinvr.net/

[14] Glue — universal collaboration platform. [Online]. Available: https://glue.work/

[15] Home — engage - vr education & corporate training platform. [Online]. Available: https://engagevr.io/

[16] The wild - vr collaboration for architecture & design teams. [Online]. Available: https://thewild.com/

[17] Resolve - virtual collaboration for complex bim. [Online]. Available: https://www.resolvebim.com/

[18] Vr for architecture, engineering, and construction — irisvr. [Online]. Available: https://irisvr.com/

[19] Medical virtual reality — virtual surgical planning — surgical theater. [Online]. Available: https://www.surgicaltheater.net/

[20] Vr medical training & simulation — health scholars. [Online]. Available: https://healthscholars.com/

[21] T. Gray and L. Madson, “Ten easy ways to engage your students,” *College Teaching*, vol. 55, no. 2, pp. 83–87, 2007.

[22] J. McLeod, J. Fisher, and G. Hoover, *The key elements of classroom management: Managing time and space, student behavior, and instructional strategies*. ASCD, 2003.

[23] A. Roehl, S. L. Reddy, and G. J. Shannon, "The flipped classroom: An opportunity to engage millennial students through active learning strategies," *Journal of Family & Consumer Sciences*, vol. 105, no. 2, pp. 44–49, 2013.

[24] C. J. Kramsch, "Classroom interaction and discourse options," *Studies in second language acquisition*, pp. 169–183, 1985.

[25] M. Mahmood, V. Rodwell, and T. Bird. (2018) Can virtual reality clinical scenarios improve medical students' clinical learning? leicester: Discovering teaching excellence at leicester conference. [Online]. Available: https://www2.le.ac.uk/departments/medicine/people/staff-profiles/terese-bird

[26] J. Pottle, "Virtual reality and the transformation of medical education," *Future healthcare journal*, vol. 6, no. 3, p. 181, 2019.

[27] Visit the moon with me! virtual reality in the classroom. [Online]. Available: https://ctl.byu.edu/tip/visit-moon-me-virtual-reality-classroom

[28] C. Kwon, "Verification of the possibility and effectiveness of experiential learning using hmd-based immersive vr technologies," *Virtual Reality*, vol. 23, no. 1, pp. 101–118, 2019.

[29] E. Ch'ng, Y. Cai, and H. Thwaites, "Special issue on vr for culture and heritage: the experience of cultural heritage with virtual reality: guest editors' introduction," *Presence: Teleoperators and Virtual Environments*, vol. 26, no. 03, pp. iii–vi, 2018.

[30] Virtual reality tours — google arts & culture. [Online]. Available: https://artsandculture.google.com/project/virtual-tours

[31] K. L. Nowak and J. Fox, "Avatars and computer-mediated communication: a review of the definitions, uses, and effects of digital representations," *Review of Communication Research*, vol. 6, pp. 30–53, 2018.

[32] J. C. Sherblom, "The computer-mediated communication (cmc) classroom: a challenge of medium, presence, interaction, identity, and relationship," *Communication Education*, vol. 59, no. 4, pp. 497–523, 2010. [Online]. Available: https://doi.org/10.1080/03634523.2010.486440

[33] J. N. Bailenson, A. C. Beall, J. Loomis, J. Blascovich, and M. Turk, "Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments," *Presence: Teleoperators & Virtual Environments*, vol. 13, no. 4, pp. 428–441, 2004.

[34] R. E. Guadagno, J. N. Bailenson, A. C. Beall, A. Dimov, and J. Blascovich, "Transformed social interaction and the cyranoid: The impact of non-verbal," *Environments*, vol. 13, no. 4, pp. 428–441.

[35] A. Beall, J. Bailenson, J. Loomis, J. Blascovich, and C. Rex, "Non-zero-sum mutual gaze in collaborative virtual environments," in *Proceedings of HCI international*, 2003.

[36] D. Roth, C. Klelnbeck, T. Feigl, C. Mutschler, and M. E. Latoschik, "Beyond replication: Augmenting social behaviors in multi-user virtual realities," in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, March 2018, pp. 215–222.

[37] T. Piumsomboon, G. Lee, J. Hart, B. Ens, R. Lindeman, B. Thomas, and M. Billinghurst, "Mini-me: an adaptive avatar for mixed reality remote collaboration," in *CHI 2018 - Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, A. Cox and M. Perry, Eds. United States of America: Association for Computing Machinery (ACM), 2018, international Conference on Human Factors in Computing Systems 2018, CHI 2018 ; Conference date: 21-04-2018 Through 26-04-2018. [Online]. Available: https://chi2018.acm.org/

[38] I. R. Murray and J. L. Arnott, "Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion," *The Journal of the Acoustical Society of America*, vol. 93, no. 2, pp. 1097–1108, 1993.

[39] J. Copley, "Audio and video podcasts of lectures for campus-based students: production and evaluation of student use," *Innovations in Education and Teaching International*, vol. 44, no. 4, pp. 387–399, 2007. [Online]. Available: https://doi.org/10.1080/14703290701602805

[40] B. E. Schreiber, J. Fukuta, and F. Gordon, "Live lecture versus video podcast in undergraduate medical education: A randomised controlled trial," *BMC Medical Education*, vol. 10, no. 1, pp. 1–6, 2010.

[41] R. H. H. Groenwold and M. J. Knol, "Learning styles and preferences for live and distance education: an example of a specialisation course in epidemiology," *BMC medical education*, vol. 13, p. 93, 2013. [Online]. Available: https://europepmc.org/articles/PMC3702471

[42] S. W. Chae, K. C. Lee, and Y. W. Seo, "Exploring the effect of avatar trust on learners' perceived participation intentions in an e-learning environment," *International Journal of Human–Computer Interaction*, vol. 32, no. 5, pp.

373–393, 2016. [Online]. Available: https://doi.org/10.1080/10447318.2016.1150643

[43] S. W. Cook, H. S. Friedman, K. A. Duggan, J. Cui, and V. Popescu, "Hand gesture and mathematics learning: Lessons from an avatar," *Cognitive Science*, vol. 41, no. 2, pp. 518–535, 2017. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/cogs.12344

[44] G. Tombs, R. Bhakta, and M. Savin-Baden, "'it's almost like talking to a person': Student disclosure to pedagogical agents in sensitive setting," *Proceedings of the 9th International Conference on Networked Learning 2014,*, pp. 296–304, April 2014, proceedings of the 9th International Conference on Networked Learning, Edinburgh, 07/04/2014-09/04/2014, Edited by: Bayne S, Jones C, de Laat M, Ryberg T & Sinclair C. [Online]. Available: http://eprints.worc.ac.uk/3552/

[45] M. E. Latoschik, D. Roth, D. Gall, J. Achenbach, T. Waltemate, and M. Botsch, "The effect of avatar realism in immersive social virtual realities," in *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '17. New York, NY, USA: Association for Computing Machinery, 2017. [Online]. Available: https://doi.org/10.1145/3139131.3139156

[46] J. Lugrin, M. Wiedemann, D. Bieberstein, and M. E. Latoschik, "Influence of avatar realism on stressful situation in vr," in *2015 IEEE Virtual Reality (VR)*, 2015, pp. 227–228.

[47] T. Waltemate, D. Gall, D. Roth, M. Botsch, and M. E. Latoschik, "The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 4, pp. 1643–1652, 2018.

[48] D. Jo, K. Kim, G. F. Welch, W. Jeon, Y. Kim, K.-H. Kim, and G. J. Kim, "The impact of avatar-owner visual similarity on body ownership in immersive virtual reality," in *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '17. New York, NY, USA: Association for Computing Machinery, 2017. [Online]. Available: https://doi.org/10.1145/3139131.3141214

[49] D. Jo, K.-H. Kim, and G. J. Kim, "Effects of avatar and background types on users' co-presence and trust for mixed reality-based teleconference systems," in *Proceedings the 30th Conference on Computer Animation and Social Agents*, 2017, pp. 27–36.

[50] ——, “Spacetime: adaptive control of the teleported avatar for improved ar tele-conference experience,” *Computer Animation and Virtual Worlds*, vol. 26, no. 3-4, pp. 259–269, 2015. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cav.1645

[51] C. L. Kleinke, “Gaze and eye contact: a research review.” *Psychological bulletin*, vol. 100, no. 1, p. 78, 1986.

[52] J. K. Hietanen, “Affective eye contact: an integrative review,” *Frontiers in psychology*, vol. 9, p. 1587, 2018.

[53] M. Argyle and M. Cook, “Gaze and mutual gaze.” 1976.

[54] R. Vertegaal, H. Vons, and R. Slagter, “Look who’s talking: the gaze groupware system,” in *CHI’98*. United States: Association for Computing Machinery (ACM), Apr 1998, pp. 293–294.

[55] T. Piumsomboon, Y. Lee, G. Lee, and M. Billinghurst, “Covar: A collaborative virtual and augmented reality system for remote collaboration,” in *SIGGRAPH Asia 2017 Emerging Technologies*, ser. SA ’17. New York, NY, USA: Association for Computing Machinery, 2017. [Online]. Available: https://doi.org/10.1145/3132818.3132822

[56] J. N. Bailenson, N. Yee, J. Blascovich, and R. E. Guadagno, “Transformed social interaction in mediated interpersonal communication,” in *Mediated Interpersonal Communication*, T. M. U. S. . L. A. Konijn, E., Ed. Lawrence Erlbaum Associates, 2008, ch. 5, pp. 77–99.

[57] J. F. Sahlström, “The interactional organization of hand raising in classroom interaction,” *The Journal of Classroom Interaction*, pp. 47–57, 2002.

[58] R. Böheim, T. Urdan, M. Knogler, and T. Seidel, “Student hand-raising as an indicator of behavioral engagement and its role in classroom learning,” *Contemporary Educational Psychology*, vol. 62, p. 101894, 2020.

[59] G. R Nurieva and L. M Garaeva, “Zoom-based distance learning of english as a foreign language,” *Journal of Research in Applied Linguistics*, vol. 11, no. Proceedings of the 7th International Conference on Applied Linguistics Issues (ALI 2020), Saint Petersburg, 13-14 June 2020, 2020.

[60] Raise your hand in a teams meeting - office support. [Online]. Available: https://support.microsoft.com/en-us/office/raise-your-hand-in-a-teams-meeting-bb2dd8e1-e6bd-43a6-85cf-30822667b372

[61] Vive pro starter kit — the professional-grade vr headset. [Online]. Available: https://www.vive.com/eu/product/vive-pro-starter-kit/

[62] Airpods (2nd generation) - apple. [Online]. Available: https://www.apple.com/airpods-2nd-generation/

[63] Unity real-time development platform — 3d, 2d vr amp; ar engine. [Online]. Available: https://unity.com/

[64] Steamvr plugin — integration — unity asset store. [Online]. Available: https://assetstore.unity.com/packages/tools/integration/steamvr-plugin-32647

[65] blender.org - home of the blender project - free and open 3d creation software. [Online]. Available: https://www.blender.org/

[66] K. L. Nowak and F. Biocca, "The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments," *Presence: Teleoperators & Virtual Environments*, vol. 12, no. 5, pp. 481–494, 2003.

[67] R. Vertegaal, R. Slagter, G. van der Veer, and A. Nijholt, "Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '01. New York, NY, USA: Association for Computing Machinery, 2001, p. 301–308. [Online]. Available: https://doi.org/10.1145/365024.365119

[68] Dutch a1 - dutch is fun! [Online]. Available: https://www.utwente.nl/en/courses/366889/dutch-a1-dutch-is-fun/

[69] Introductory course: Swedish language and culture — kth. [Online]. Available: https://www.kth.se/en/larande/sprak/for-internationella-studenter/masterstudenter/introduktion-till-sverige-sprak-och-kultur-1.813214

[70] Välkommen till pantamera. [Online]. Available: https://pantamera.nu/

[71] Mirror networking – open source networking for unity. [Online]. Available: https://mirror-networking.com/

[72] Snaps prototype — office — 3d environments — unity asset store. [Online]. Available: https://assetstore.unity.com/packages/3d/environments/snaps-prototype-office-137490

[73] Ultimate low poly supermarket amp; shop shelf pack — 3d props — unity asset store. [Online]. Available: https://assetstore.unity.com/packages/3d/props/ultimate-low-poly-supermarket-shop-shelf-pack-84356

[74] Modular railing set — 3d — unity asset store. [Online]. Available: https://assetstore.unity.com/packages/3d/modular-railing-set-17090

[75] Green forest — 3d fantasy — unity asset store. [Online]. Available: https://assetstore.unity.com/packages/3d/environments/fantasy/green-forest-22762

[76] Advanced people pack — 3d humanoids — unity asset store. [Online]. Available: https://assetstore.unity.com/packages/3d/characters/humanoids/advanced-people-pack-154474

[77] Mixamo. [Online]. Available: https://www.mixamo.com/

[78] W. Barfield, D. Zeltzer, T. Sheridan, and M. Slater, "Presence and performance within virtual environments," *Virtual environments and advanced interface design*, pp. 473–513, 1995.

[79] M. Lombard and T. Ditton, "At the heart of it all: The concept of presence," *Journal of computer-mediated communication*, vol. 3, no. 2, p. JCMC321, 1997.

[80] G. Riva, F. Davide, and W. IJsselsteijn, "7 measuring presence: Subjective, behavioral and physiological methods," *Being there: Concepts, effects and measurement of user presence in synthetic environments*, pp. 110–118, 2003.

[81] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence*, vol. 7, no. 3, pp. 225–240, 1998.

[82] J. Lessiter, J. Freeman, E. Keogh, and J. Davidoff, "A cross-media presence questionnaire: The itc-sense of presence inventory," *Presence: Teleoperators & Virtual Environments*, vol. 10, no. 3, pp. 282–297, 2001.

[83] M. Bond, K. Buntins, S. Bedenlier, O. Zawacki-Richter, and M. Kerres, "Mapping research in student engagement and educational technology in higher education: A systematic evidence map," *International Journal of Educational Technology in Higher Education*, vol. 17, no. 1, p. 2, 2020.

[84] M. M. Handelsman, W. L. Briggs, N. Sullivan, and A. Towler, "A measure of college student course engagement," *The Journal of Educational Research*, vol. 98, no. 3, pp. 184–192, 2005.

[85] S. Ahlfeldt*, S. Mehta, and T. Sellnow, "Measurement and analysis of student engagement in university classes where varying levels of pbl methods of instruction are in use," *Higher Education Research & Development*, vol. 24, no. 1, pp. 5–20, 2005.

[86] Introduction — steamvr unity plugin. [Online]. Available: https://valvesoftware.github.io/steamvr_unity_plugin/tutorials/Skeleton-Poser.html

[87] M. J. Hatch and M. Schultz, "Are the strategic stars aligned for your corporate brand," *Harvard business review*, vol. 79, no. 2, pp. 128–134, 2001.

[88] How 5g and edge computing can enhance virtual reality. [Online]. Available: https://www.ericsson.com/en/blog/2020/4/how-5g-and-edge-computing-can-enhance-virtual-reality

Appendix A

Teaching Script

The bellowing content contains lecturer's teaching script for two lectures created for HejVR system, together with wizard control instructions and actors' utterances.

A.1 Supermarket Lecture

Hej Hej, Välkommen. Welcome to the virtual Swedish supermarket.

My name is Iris, I am from China, studying my master's degree in collaboration of KTH and University of Twente, and I have been living in Sweden for 2.5 years. If you have any questions or comments during the entire session, feel free to raise your hand and ask me directly.

Who of you are the first time in Sweden? Raise your hand if it's your first time. You can do that by just raising your hand **[Click button to raise hand of lecturer]**, like this **[Pause audio] [Click the same button to put hand down of lecturer]**.

[Actor 0, 1, 3 raises hands] [Continue audio] What brings you here? **[Pause audio]**

[If participant raises hand, focus on participant.]Yes, (name)? (Interact)

[Actor 1-1: "Study, as an exchange student." (lecturer focus on A1). Actor 3-1: "Same." (lecturer focus on A3)] [Actor 0, 1, 3 put hands down]

[Continue audio] Nice. **[Unfocus lecturer]** Which of you have been to Sweden before? Raise your hand! **[Pause audio]**

[Actor 2 raises hand] [Continue audio] How many times? Where else have you been to besides Stockholm? **[Pause audio]**

[If participant raises hand, focus on participant.]Yes, (name)? (Interact)

[Actor 2-1: "Once in Lapland, Kiruna." (lecturer focus on A2)][Actor 2 puts hand down]

[Continue audio] Wow nice! The very north. **[Unfocus]** If you have been to Sweden already, I hope you can still have some takeaways from this lecture.

[Talk animation] In the next 10 minutes, I would like to share with you some common Swedish phrases and some phrases especially useful for going shopping in Sweden, and I will also tell you something special about Swedish supermarkets.

As you might have known, most people in Sweden speak very good English. Despite this, it is still good to know some basics, and make yourself feel more at home. Let's start with the essentials, which you probably already have heard a lot.

To start with, **Hej**! Does anyone know or want to guess what that means? Raise your hand! **[Pause audio]**

[If participant does not react, or after the participant if wrong, Actor 1-2: "Hi? For greeting?" (lecturer focus on A1)]

[Continue audio, Unfocus] Yes! It's the "Hi", "Hello" in Swedish. It sounds like a chill word huh, but it's a perfectly fine greeting to be used as a start in business emails. And you can say "hej hej" to make it sound cuter, friendlier. **[Pause audio]**

[Actors 3-2 raises hand and asks: "What is goodbye?" (lecturer focus on A3), hand down]

[Continue audio] Good question. **[Unfocus]** "Goodbye" is **Hej då**.

So we are shopping in the supermarket, but there is something on the shopping list that we can't find, and we want to ask the staff over there for help. **[Lecturer points to staff Actors turn to staff and turn back]** Now you can say - **Ursäkta**. It means "excuse me". When you need someone's attention for asking a question or favor, you can say "ursäkta".

And now the staff offered you help, you could say - **Tack**, "thank you", or **Tack så mycket**, thank you very much. But tack is more than a thank you. Swedish people really say it a lot. It's very common to offer your gratitude and add 'tack' at the end of the sentence. For example, "En kaffe, tack (En kopp kaffe tack)". May I have a cup of coffee please.

[Talk animation] Next, **Varsågod**. Another important multi-purpose word. It can mean "Here you go/Please/You're welcome" depending on the context. For instance, when you hand over someone something, open the door for someone, invite someone for food, you could say "Varsågod".

If you accidently bump into someone, you would say "**Förlåt**", meaning "sorry", "forgive me".

[Lecturer focus on actor 0] Ok Jane, may I ask you a question? Pratar du svenska? **[Pause audio]**

[Actors 0-1 says: "Em, no."]

[Continue audio] No? **[Focus on participant]** And (Participant's name), Pratar du svenska? **[Pause audio] [Wait for response, interact] [Continue audio]** Pratar du engelska? **[Pause audio] [Wait for response, interact]**

[Continue audio] [Unfocus] So the sentences I just said, **Pratar du svenska?**

And **Pratar du engelska?** Means simply: Do you speak Swedish? And Do you speak English. You can add any other language at the end.

[Talk animation] And this is a general (yes-no) question. So how to answer in Swedish? Yes is **Ja**, no is **Nej**. But there is sometimes a third option - **Nja**. It's a blend of nej + ja, used to give a hesitant or uncertain answer or (dis)agreement. But what they really mean is usually (wait a few seconds) - more of a no.

Now we've got everything we need and move to checkout, here are some more general questions often asked by cashiers.

1. **Är du medlem?** - Are you a member?
2. **Vill du ha påse?** - Do you want a bag?
3. **Vill du ha kvitto?** - Do you want the receipt?

In Sweden, bags are often charged at the supermarket and many other shops in order to reduce waste. So always remember to bring a reusable bag to the store, to save you some money and help with the environment.

Have you seen the machine over there? **[Lecturer points to machine Actors turn to machine and turn back]** Does anybody know what it is? Anyone wants to make a guess? Raise your hand! **[Pause audio] [Wait for response]**

[If participant does not react, or after the participant if wrong (Any other guess?), Actor 2-2: "For bottle recycling?"(lecturer focus on A2)] [Continue audio]

Yes! **[Unfocus]** It is a plastic bottle and can recycling machine, a part of the Swedish PANT scheme - empty bottles deposit-refund system. You can find this type of machine in many supermarkets (ICA, Coop, Hemköp etc.) This is not exclusive in Sweden, it is also commonplace in other Scandinavian countries as well as Germany etc..

Is there such a system where you are from? Raise your hand! **[Pause audio] [Wait for response, interact] [Continue audio]**

Sweden recycles over 80% of all its used plastic bottles and metal drinks cans thanks to the PANT scheme. The word PANT is not an acronym – it refers to a legal term whereby someone is owed back money at a later date. It also has a quite long history, started in 1984 for aluminium cans, and since 1994 for plastic bottles. There is even a verb - **panta**, which usually refers to the process of putting bottles in the supermarket's magic machines and getting money back.

[Talk animation] If you buy a can of Coca-Cola, for example **[show cola image]**, the price might be advertised at 11 SEK, but when you go to checkout, you will be charged 12 SEK. Upon looking closely at the bottle, you'll see it says PANT 1 kr. So, effectively, you're paying a deposit every time you purchase. To redeem the money, you just need to collect the empty bottle, find one of these machines, put them in, and you will get the receipts of the amount of the refund which can be used in the shop. There is also another option, by pressing the yellow button, you can donate the money to charity.

That's all for this lecture, I hope you learned something useful. Practice these phrases in your next shopping! Do you have any questions? **[Pause audio] [Wait for response, interact] [Continue audio]**

Thank you all for attending, I hope you enjoy your time in Sweden!

A.2 Berries Lecture

Hej Hej, Välkommen. Welcome to the virtual Swedish forest.

My name is Iris, I am from China, studying my master's degree in collaboration of KTH and University of Twente, and I have been living in Sweden for 2.5 years. If you have any questions or comments during the entire session, feel free to raise your hand and ask me directly.

Have any of you been to some nature, forests in Sweden? Raise your hand if you have! You can do that by just raising your hand, **[Click button to raise hand of lecturer]** like this. **[Pause audio] [Click the same button to put hand down of lecturer]**

[Actor 1, 2 raises hands] [Continue audio] Where have you been to? **[Pause audio]**

[If participant raises hand, focus on participant.]Yes, (name)? (Interact)

[Actor 1-1: "The forest behind Lappis, the student accommodation." (lecturer focus on A1)]

[Actor 2-1: "I have been to Skansen, but not real forests." (lecturer focus on A2)]

[Actor 1, 2 puts hands down]

[Continue audio] Nice! Did you see any berries? **[Pause audio]**

[Actor 1-2: "Yes! There are lots of blueberries in Lappis." (lecturer focus on A1)]

[Continue audio] [Unfocus] Anybody who has not got the chance to see Swedish nature yet? **[Pause audio]**

[Actor 0, 3 raise hands] [If participant raises hand, focus on participant. (Interact)]

[Actor 0-1: "I just arrived a few days ago then school started, I didn't have time."]

[Actor 3-1: "I was quarantining myself..."] [Actor 0, 3 put hands down]

[Continue audio] Ah okay. **[Unfocus]**

[Talk animation] Well in the next 10 minutes, I will tell you something about a sweet sweet topic - berries. As you might have seen in the supermarket now, there are lots of berries selling at the moment. Swedish forests are also abundant in berries at this time of the year.

What are your favorite berries? Or what kind of unique berries do you have in your country? Anyone wants to share? Raise your hand! **[Pause audio]**

[If participant raises hand, focus on participant.]Yes, (name)? (Interact)

[Actor 2 raises hand] [Actor 2-2: "I like Mulberry."] [Actor 2 puts hand down]

[Continue audio] Of their bags? I'm just kidding. Despite there being a fashion brand with the same name, mulberry is actually a kind of berry. I just learned the English word myself recently. I've never seen them here in Europe though. Who else? **[Pause audio]**

[Actor 0 raises hand] [Actor 0-2: "Strawberries and raspberries."] [Actor 0 puts hand down]

[Continue audio] Ah yummy. In summer, Swedes invade the forests with sturdy boots and mosquito spray to fill their baskets with delicious berries. In old times it was important to have a source of vitamins and minerals for the long winter, but today, with the grocery's plentiful assortment of fruits and vegetables from all over the world, berry picking is mostly for the pleasure of the outdoor experience and DIY your own jam.

So what can we find in Swedish forests? **[Talk animation]** As you've mentioned before, **Blåbär** - Blueberries! Nordic blueberries are actually different from the ones in North America. They are richer in vitamin C/D and antioxidants which have been found to prevent cancer, diabetes. And in fact, the correct English name for the Nordic blueberry is bilberry, although people tend to call them blueberries anyway. They are the most abundant among all, and can be found everywhere in Sweden, below pine trees in low bushes. They're small and sweet, not very sour, can be paired with pancakes and whipped cream, made into jams, cakes, and even sweet soup.

Next up, **Hjortron** - cloudberries. Have you ever heard of it? **[Pause audio] [Wait for response, interact]**

[Continue audio] They are often called, 'King of berries', or 'forest gold', high in vitamins C and A, as well as iron and other essential minerals. They prefer swampy areas in the arctic region like Scandinavia (Norway and Finland), Russia, Canada and Alaska. It is difficult to cultivate cloudberries and so it's primarily a wild plant. In Sweden, cloudberries and cloudberry jam are used as a topping for ice cream, pancakes, and waffles.

[Talk animation] Next is my favorite - **Hallon** - Raspberries. Red, juicy, sweet, with a very interesting texture. They are the best to eat fresh, with some yogurt or ice cream, or on top of cakes. Wild raspberries are much smaller than the ones you find in the supermarket, but sweeter and have stronger aroma, like all wild berries. Fun fact, there is a Canadian raspberry shaped candy called Swedish berries, like this. **[Show Swedish berries]**

Last but not least, the most common on Swedish table, the staple **Lingon** - lingonberries. Look a bit like cranberries, but much smaller and more round.

Have you ever been to IKEA? Raise your hand if you have. **[Actor 0, 1, 2, 3 raise hands]** Ever ordered meatballs there? **[Actor 0, 1, 2, 3 keep hands up for a few sec and put down]**

[Pause audio]Interact (everyone has/ No?)[Continue audio] Lingonberry jam is the essential, indispensable part of authentic Swedish meatballs. Lingonberries are bright red, small, taste a bit sour and bitter, so it is more common to be made into drinks or jam.

A very typical way to make lingonberry jam is to just stir with some sugar, no boiling required. This is called "**rårörda lingon**", literally "raw stirred lingonberries". Lingonberry jam is like a ketchup, mayonnaise or mustard in Swedish cuisine, for which I don't mean the taste, but it's so widely used to accompany a variety of dishes. Besides meatball, it also goes well with bacon and potato pancakes, blood pudding, cabbage rolls, fried herring or normal sweet stuff like cheesecake or oatmeal porridge.

Let me show you something. **[Click to show berry picker on the table]** Does anybody know what it is? Anyone wants to make a guess? Raise your hand! **[Pause audio] [Wait for response]**

[If participant does not react, or after the participant if wrong, Actor 2-3: "For storing berries?"] Ahem, sort of [Wait for response] [Continue audio]

This is a berry picker, and is used to speed up the harvesting process. It's very suitable for blueberries and lingonberries those kinds of round berries, just go through the bushes with the fork and the berries can be easily collected inside.

So where can we go to pick up berries? The answer is - almost anywhere! **[Talk animation]** Thanks to the rule, **Allemansrätten** – "All Men's Right", or the Right of Public Access, everyone is allowed to roam freely in the countryside, camp, swim, travel by boat and pick up mushrooms and berries in the forest. With this right comes the responsibility to tread carefully and to show consideration for landowners and others. As long as the land is not cultivated, and as long as no damage is caused, most of Sweden's nature is yours to explore.

Allemansrätten has become a part of the national identity of Sweden over generations. Many Swedes grew up picking berries from the forest and making the most fresh delicious jam at home.

That's all for this lecture. There are a lot more types of berries in the magical Swedish forest and a lot more to explore, so grab your basket and enjoy the last bit of summer. Thank you all for attending, I hope you enjoy your time in Sweden!

Appendix B

Information Brochure

Information Brochure Department Human Media Interaction (HMI)
Faculty of EEMCS
University of Twente

Project Title: Master Thesis - XR in Immersive Teleconferencing
Researcher: Zhuowen Fang, z.fang-1@student.utwente.nl, +46706978473
Supervisor: Dr.ir. D. Reidsma, d.reidsma@utwente.nl, +31534893718

Purpose: You are being asked to participate in a master thesis study under the topic of eXtended Reality (XR) in immersive teleconferencing. The purpose of this user study is to evaluate the current design of a Virtual Reality (VR) system for online cultural lectures in an explanatory sequential approach, with a quantitative questionnaire based experiment followed by an interview. The information gathered from this study will be analysed and used to verify our design and search for new solutions for VR based online lectures.

Procedures: Your participation in the study will take up to 60 mins. After reading this information brochure and signing the consent form, we will play an instruction video about how to put on the Vive Pro Head Mounted Display (HMD) which you will use to experience A written manual about putting on the VR HMD is also placed on the wall.

You will then participate in two lectures related to Swedish culture and language in VR as a student. Each lecture will take about 15 minutes and after each lecture you will be given a questionnaire related to your lecture experience. Between the two lectures, you will be given a 5-minute break after completing the questionnaire. After finishing both lectures and questionnaires, you will also be given a short break and then we will move on to a semi-structured interview which will last about 15 minutes.

If at any time you experience motion sickness or any other discomfort and feel that you cannot continue the study, you can tell the experimenter and we will stop the system immediately and you can take off the VR HMD.

Data: All data collected in this study will be processed anonymously, and it will be destroyed after 12 months. It will only be accessible by the researcher and her supervisors.

Participants:
Participants must be 18 years or older. Participation remains at all times voluntary and that without giving any reasons, subjects may refuse to participate in the research. Participants may also end their participation at any time and may also refuse afterwards (within 24 hours) to allow their data to be used for the research. All of this may not at any time have any adverse consequences for the subject.

Contact Person: Any questions or comments about the study may be directed to Zhuowen Fang at z.fang-1@student.utwente.nl.

Contact Information for Questions about Your Rights as a Research Participant: If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee of the department of EEMCS, mail ethics-comm-ewi@utwente.nl or call +31534892085.

Before participating in this study, you will be given a consent form.

Please read it carefully and ask any questions you might have before signing it!

Thank you!

Appendix C

Consent Form

Participant Number (filled out by the researcher): ___________________________

Consent Form Department Human Media Interaction (HMI)
Faculty of EEMCS
University of Twente

Project Title: Master Thesis - XR in Immersive Teleconferencing
Researcher: Zhuowen Fang, z.fang-1@student.utwente.nl, +46706978473
Supervisor: Dr.ir. D. Reidsma, d.reidsma@utwente.nl, +31534893718

Data:
By participating in this study, you will be asked to share limited personal information including age and gender and the experiment will be recorded in video. The interview after the experiment will be recorded in audio.

- Video recording will only be taken for the time you are in the VR lectures, where you will be wearing a VR HMD that covers most part of your face.

- Audio recording will only be taken for the interview where you will be asked a series of semi-structured questions regarding your experience in the VR lecture.

Risks: Some people may experience motion sickness in Virtual Reality and if that happens to you we will stop the system immediately. This type of sickness can be recovered in a few minutes or hours after taking off the headset. The study will not cause any other physical risks to you.

Confidentiality:

- For the duration of the study, your data will be treated confidentially. You will be assigned a participant number that is linked to all data we collect from you. In case you choose to withdraw your consent, this will allow us to delete all your data.
- After completing the data analysis, we may store some of the collected (video and audio) data. However, this will be anonymized and any links between recordings, your personal and contact information will be destroyed.
- We will attempt to analyze the questionnaire and interview results, identify user behavior, problems and potential improvements on the current system design. The result itself will not contain or produce any information that can be linked to your personal identity.
- Your video and audio recordings will not be shared with any third party. If your individual results are discussed, presented or included in the report, no reference can be made with regard to your personal identity.
- All data collected in this study will be destroyed after 12 months. It will only be accessible by the researchers and their supervisors.

Subjects' Rights: Your participation in this study is voluntary and you are free to withdraw at any time without any given reason. We will stop collecting your data and delete any data collected from you from the dataset, as soon as you withdraw consent. You may also refuse afterwards (within 24 hours) to allow your data to be used for the research. All of this may not at any time have any adverse consequences for you.

Participant Number (filled out by the researcher): ___________________________

Regarding the ongoing COVID-19 pandemic, if the experimenter shows symptoms or has a positive PCR test result in the next 14 days, or we hear that from the other participants, we will inform you promptly. If you experience any COVID-19 symptoms or receive a positive PCR test result within the next 14 days, please notify us as well. Your health is our top priority.

Consent:
"I hereby declare that I have been informed in a manner which is clear to me about the nature and method of the research as described in the aforementioned information brochure. My questions have been answered to my satisfaction. I agree of my own free will to participate in this research. I reserve the right to withdraw this consent without the need to give any reason and I am aware that I may withdraw from the experiment at any time. If my research results are to be used in scientific publications or made public in any other manner, then they will be made completely anonymous. My personal data will not be disclosed to third parties without my express permission. If I inform the experimenter afterwards that I show symptoms or a positive PCR test result of COVID-19, I agree that this information can anonymously be shared with other participants without including my identity. If I request further information about the research, now or in the future, I may contact Zhuowen Fang (z.fang-1@student.utwente.nl)."

If you have any complaints about this research, please direct them to the secretary of the Ethics Committee of the Faculty of Electrical Engineering, Mathematics and Computer Science at the University of Twente, P.O. Box 217, 7500 AE Enschede (NL), email ethics-comm-ewi@utwente.nl or call +31534892085.

Signed in duplicate:

....	....	....
Name subject	Signature	Date

I have provided explanatory notes about the research. I declare myself willing to answer to the best of my ability any questions which may still arise about the research.'

....	....	....
Name researcher	Signature	Date