AALBORG UNIVERSITY
STUDENT REPORT

Title:
Examining the Effects of Eye-tracking on Dyadic Conversations in VR

Project:
Medialogy Master Thesis

Project Period:
Spring 2022

Project Group:
MTA10 Gr04

Participants:
Razvan Constantin Andrei
Oliver Benjamin Engermann
Christian Glerup Sørensen

Supervisor:
Markus Löchtefeld

Page Count: 12

Date of Completion:
May 24, 2022

Abstract:

This paper examines the effects of eye-tracking-driven social gaze on the quality of communication in a virtual environment between two participants engaging in dyadic conversation in immersive VR. While the importance of social gaze on quality of communication has been well documented, determining if its effects carry over into VR has not. This paper will outline the design and implementation of a multi-user VR experiment, focusing on developing the humanoid avatars and their gaze interactions and the findings of a comparative between-subjects study on three conditions: eye-tracking, without eye-tracking, and simulated gaze. The findings of this study indicate that for the specified implementation, eye-tracking-driven social gaze did not yield a statistically significant effect on the quality of conversation. However, it provided a more realistic conversation than simulated social gaze.

Examining the Effects of Eye-tracking on Dyadic Conversations in VR

Razvan Constantin Andrei
randre17@student.aau.dk

Oliver Benjamin Engermann
oenger17@student.aau.dk

Christian Glerup Sørensen
cgsa16@student.aau.dk

ABSTRACT

This paper examines the effects of eye-tracking-driven social gaze on the quality of communication in a virtual environment between two participants engaging in dyadic conversation in immersive VR. While the importance of social gaze on quality of communication has been well documented, determining if its effects carry over into VR has not. This paper will outline the design and implementation of a multi-user VR experiment, focusing on developing the humanoid avatars and their gaze interactions and the findings of a comparative between-subjects study on three conditions: eye-tracking, without eye-tracking, and simulated gaze. The findings of this study indicate that for the specified implementation, eye-tracking-driven social gaze did not yield a statistically significant effect on the quality of conversation. However, it provided a more realistic conversation than simulated social gaze.

KEYWORDS

Virtual Reality, Eye-tracking, Gaze-tracking, Embodiment, Involvement, Social Gaze, Multi-Modal Interaction, Online Collaboration, Online Meetings, Partner Evaluation, Co-presence, Quality of Communication

1 INTRODUCTION

Due to COVID-19 and the many restrictions, working remotely from home has been the only option for many. Applications for online meetings with voice and webcam support, such as Zoom [8] and Microsoft Teams [4] have been the standard for conducting meetings during the restriction times. However, communication over a webcam, compared to communicating in person, lacks several aspects of non-verbal communication, such as gaze, proximity, and gestures. These aspects have a significant effect on the extent to which the conversational subjects positively evaluated their partner and the extent to which the conversation was enjoyed [23, 44], and help in communicating interest [19] and emotions [17, 28, 32]. Despite restrictions lessening in most parts of the world, the work-from-home culture is now seen as a more legitimate way of working, even after lock-down has ended.

Using webcams to feel present with other meeting participants can be difficult as the physical webcam does not align with the view-ports of participant's camera feeds. Sometimes it is placed on a secondary monitor resulting in a constant state of gaze-aversion. A study by Steinicke et al. found that immersive Virtual Reality (VR) using head-mounted displays (HMD) seems like a good alternative for online meetings and other webcam interactions. HMDs allow for better non-verbal communication, which leads to a higher sense of presence [47]. Simultaneously companies like Meta released Horizons Workroom [2] which places meeting participants in the same virtual reality environment (VRE), allowing for proximity and head-based non-verbal communication. However, they still lack gaze-based communication outside of what is possible by only moving one's head or what can be simulated. With Tobii [7] allegedly in talks with Sony about implementing their eye-tracking technology for PlaystationVR2 [20], it seems eye-tracking will continue to become a more affordable and widely available technology for use in VR, be it PlaystationVR2, binocular-addons [6], or the new HTC VIVE Pro Eye HMD [3]. Accurately depicted eye movements of a virtual avatar could provide an additional collocated modal of communication. It could more accurately represent attention, interest, and signal turn-taking than purely head-and-hand-based motion, which we suspect will affect how participants perceive the quality of communication with their conversation partner in VR.

In this paper, we investigate the effect of eye-tracking-driven gaze behavior of virtual avatars in VR on the indicators of face-to-face involvement, co-presence, and partner evaluation. It will be done in a meeting scenario between two participants, comparing eye-tracking-driven eye movements to no-eye-movement as is currently the standard in VR-meeting applications. Furthermore, this paper will investigate whether simulating the eye movement of avatars based on gaze data from previous non-verbal-communication experiments yields similar results to actual eye-tracked motions.

2 BACKGROUND RESEARCH

In the following section, research, supported by related works, which has formed the base of design and implementation, will be outlined, such as non-verbal communication, gaze, and social gaze. Furthermore, this section will outline the theory regarding measuring the quality of communication and partner evaluation.

2.1 Non-verbal Communication

Non-verbal communication is the process of sending and receiving messages without using words, neither spoken nor written [24]. It is possible in several ways, such as eye contact, facial expressions, gestures, posture, and body language. Non-verbal communication is used consciously and unconsciously and is a part of every social interaction to regulate speech through turn-taking signals, emphasize points, and signify interest [9]. Therefore, to perceive an interaction as *'real'* or *'correct'*, it must include forms of non-verbal communication. Facial expressions are an important element for expressing emotions and personality, as a great amount of information is interpretable during interactions. A study by Gu et al. examines the impact of social gaze and different types of facial expressions of virtual avatars. Their study suggests that virtual avatars should have not only proper human attention attributes but also convey appropriate distractions and engagement behavior while communicating [25]. However, creating realistic facial expressions and meaningful visual speech is a complicated task due to the complex structure of the human face. There are plenty of factors that have to be taken into account when trying to understand the meaning behind non-verbal communication. Features such as eye movement, lip movement, gestures, emotional facial expressions, and body orientation can give a clear insight into what a person is trying to express. All these features are important parts of non-verbal communication. When developing humanoid avatars, lip synchronization of the facial models is crucial, as it can easily convey a close representation of real human face movements in a VE [27].

Maloney et al. explored different communications dynamics in VR [1]. Their research is a collection of findings from two other studies. The first study collected data through observations to explore which non-verbal interactions users used naturally in VR. These interactions are similar to what one would expect from a real-life face-to-face conversation, such as nodding, head movement, gestures, and pointing. The other study was an interview that looked into how users perceive and understand various non-verbal communication interactions in VR and how it affects interaction outcomes. They found that interacting in VR, by allowing forms of non-verbal communication, immerses users more than interacting through a computer screen. They found that this enables participants to connect more with those they communicate with within the VE. Participants highlighted that more realistic facial expressions and movements would likely heighten this effect.

This paper suggests several non-verbal communication methods, such as facial expressions, movements, and eyes. These non-verbal-communication allowances lead to a communicative experience more congruent to real life. It relates well to the focus on investigating the effect of eye-driven gaze behavior during conversations in these settings. Therefore, in this paper, we use the findings as a starting point for our design of the virtual avatars to make sure they can replicate good non-verbal behavior in all conditions. It is necessary to examine if eye movements provide enough difference to sufficiently elicit participants' responses.

2.2 Social Gaze

Gaze is of central importance in social behavior; paying attention to the way someone gazes can have meaningful signals, such as the amount of interest they have for the conversation partner. In Bodily Communication, Argyle separates gaze into ten different aspects [9]:

Amount of gaze at others: Measured as a percentage of time an interactor spends looking at another in the area of the face, generally done with fixations lasting one-third of a second, focusing on facial aspects like eye and mouth [53].
Mutual gaze: Percentage of time interactors spend looking at each other in the region of the face; while eye contact would be better to measure, it is not easy to detect.
Looking while talking and looking while listening: Measured separately, can be used to determine a ratio, which reflects status.
Glances: Gaze consists of glances, lasting typically two to three seconds. One aspect of glance is the eye flash used to give emphasis.
Mutual glances: Average length of mutual glances can be measured, typically lasts one second.
Pattern of fixation: Requires special equipment to measure to record the precise patterns of fixation.
Pupil dilation: Affects others, though they may not be aware of it.
Eye expression: The eyes are expressive in other ways, by how open they are and the amount of white showing above and below the pupil. This may be seen as *'staring'*, *'looking 'intently'*, *'looking daggers'*, or *'looking through'* the other person, i.e. fixating beyond them, etc.
Direction of gaze breaking: When not looking at others, depressed people tend to look downwards; interactors may break their gaze by shifting to the left or right.
Blink-rate: The rate at which one blinks varies, for example, with anxiety and concentration.

There is a balance between individual gaze and mutual gaze. People tend to look more closely at those they like. Exline and Winters found that when speaking with multiple individuals, participants tended to look at their preferred confederate 2.7 times more than the other [19]. People also tend to look more when cooperating than competing [21]. This difference in the level of gaze is noticeable: Argyle and Williams found that when the difference of the level of gaze is between 15% and 80%, people interpreted that higher levels of gaze signals interest and a wish to interact [13]. Just like people tend to look more at those they find favorable, people who look more are also perceived more favorably and self-confident [12, 28]. In a study by Bente et al. on computer-simulated gaze in avatar-based conversations, they found that even when the difference in levels of gaze is not consciously perceived, participants still preferred humanoid avatars with longer gaze duration [15]. From data gathered by Argyle [11], Argyle & Kendon [10], and Beattie [14], Argyle derived some averages regarding gaze while listening and speaking in dyadic conversations at a distance of 6 ft between interactors, see table 1. It is worth noting that there is a wide variation depending on the personalities and affiliations of those involved [9]. That there is more gaze and mutual gaze when interactors are further apart [10, 11] and interaction between strangers is more likely to take place when inside a room.

Individual gaze	60 %
While listening	75 %
While talking	40 %
Length of glance	3 seconds
Eye-contact (mutual glance)	30 %
Length of mutual glance	1.5 seconds

Table 1: Basic statistics for two people in conversation, on an emotionally neutral topic, at a distance of 6 feet [9].

These results can be regarded as the representation of normal gaze behavior, though it varies significantly from person to person. When people deviate from the norms, for example, with prolonged gazing or staring, it can be interpreted as a threat signal or bizarre, leading to heightened discomfort of the observer [12, 49]. Therefore, if one would attempt to program a basic form of simulated gaze, these are values to consider. When examining the impact of eye-gaze on communication using humanoid avatars, Garau et al. determined that when the avatar's gaze was directly related to the conversation, participants tended to look at the avatar while speaking 46% of the time and while listening 61% of the time [23]. These results seem congruent with other studies, suggesting that as long as the implementation of the social gaze of a humanoid avatar is sophisticated enough, the current literature and findings regarding social gaze should apply in VR as well.

Previous work found that gaze varies with the intensity of emotion and with different emotions. It is evident from studies that people look less and downward when sad [18] and that the level of gaze is lowest when embarrassed [17]. Higher levels of gaze signal feelings of warmth and joy, lower levels of anxiety, submission, and depression. The findings suggest that higher levels of gaze reflect positive moods [28]. Another way to interpret gaze concerning emotions is that gaze is greatest when the object of emotion is outside the self and lowest when the self is the object of the emotion. Laljee determined this when asking subjects to roleplay emotions and found that gaze was highest for surprise, excitement, joy, and scorn and lowest for despair, annoyance, rage, and anxiety [32].

Gaze plays an essential role in conversations. To mimic conversations in VR as close as possible, it will also be necessary to mimic an avatar's gaze correctly, either by accurately tracking a user's gaze or by implementing a form of simulated gaze.

2.3 Meetings and Conferencing in VR

In an attempt to maintain a social presence during virtual meetings, new studies have sought to investigate the benefits of having meetings in a VE compared to traditional video-based conferencing. Gunkel et al. conducted a study where participants held meetings in VR around a round table, had stand-up meetings, and had remote meetings between offices. Overall, participants agreed that the concept of meetings and conferences in VR is a promising concept, with 54% of participants believing VR conferencing is an excellent alternative to video conferencing. However, they wished for better interaction with the environment and note-taking functions [43].

A pilot study by Steinecke et al. compares traditional video-based meetings with meetings conducted in VR with and without HMDs to investigate the difference in the feeling of presence. Findings from the pilot study suggest that participants feel a higher sense of presence when partaking in the meeting in VR, rather than through video, but only if an HMD is used [22]. Steinicke et al.'s and Gunkel et al.'s papers were some of the first to compare video conferences to meetings in virtual reality. The findings suggest that meetings in a VE could be a suitable alternative to video meetings if done in immersive VR with an HMD. However, they did not explore some of the interaction possibilities available in VR where participants are collocated, such as gaze-tracking. As the lack of eye contact and gaze behavior is noticeably present in video conferences, it could be interesting to investigate how noticeable its effect on meetings is in immersive VR.

2.4 Evaluating Communication Methods

One way to determine the effectiveness of social gaze on conversations in VR is to examine how participants evaluate their partner differently when the colonial gaze is possible. One can examine how gaze changes the extent to which the interaction feels real, how involved in the conversation the participant feels, and whether the participant feels as if they are interacting with another person rather than a virtual avatar. It is the approach Garau et al. used when investigating the impact of eye gaze on communication with a humanoid avatar. Their evaluation consisted of 100 participants randomly assigned to one of four conditions representing different communication methods. Their evaluation had two sections. The first section determined if participants preferred conversing with a humanoid avatar rather than a purely auditory scenario. The second section investigated the impact of an avatar's informed and random gazes on a conversation. Garau et al. developed a questionnaire examining four indicators: Face-to-face, Involvement, Co-presence, and Partner Evaluation based on research by Sellen [42], and Straus & McGrath [48]. Their results suggest that communicating with an avatar with a random gaze does not outperform audio-only communication, but conversing with an avatar with an informed gaze, significantly outperforms random gaze and also outperforms audio-only based communication on several responses indicators [23]. It may be interesting to develop avatars with informed gaze through eye-tracking to correctly display where participant's conversation partners are gazing at.

Garau et al. and Straus & McGrath's experiments were not in immersive VR, meaning gaze was the only form of non-verbal communication perceivable to the participants. The difference in response for each condition is elevated compared to if the gaze behavior of the virtual avatars were unclear among other non-verbal behavior. The findings by Garau et al. in particular, however, seem indicative of what they derived from background research. Their questionnaire examines how one's conversation partner is perceived, which is usable for this paper.

3 DESIGN

This section will present the design choices and criteria of the avatars used in the virtual environment. The goal is to achieve avatars resembling human beings in their visual representation and the avatar's movement, lips, and eyes. To properly investigate this, the design criteria are as follows:

(1) It must incorporate eye-tracking enabled gaze to represent the participants' gaze accurately.
(2) It must employ realistic collocated humanoid avatars to allow for various forms of non-verbal behavior [1, 9].
(3) It must allow for verbal communication between participants.
(4) It must allow participants to perceive the lip movements of the other's avatar [25].
(5) It must occur in a realistically modeled environment to promote immersion. This environment should be indoors as background research suggests that is where conversation between strangers is most likely or natural to take place [9].

(6) It must allow two participants to be present in the same virtual environment while being physically isolated.

3.1 Avatar Models

Realistic-looking avatar models were needed to simulate honest conversations with real people as closely as possible [40]. Furthermore, a large selection of avatars is beneficial to ensure that users receive an avatar that reflects them to some extent, such as the color of their skin and gender. The avatar's faces need to be realistic, as the social gaze only pertains to human beings. Therefore, users would need to associate the virtual avatar with a real human being to the highest degree possible.

The virtual avatars and their functionality used Microsoft's Rocketbox avatar library, which contains fully 3D designed and rigged characters [34]. There were several types of avatars available in this library. It allowed people with different characteristics, such as gender and skin color, to be somewhat the same in the VE as in real life, see 1.

Figure 1: Selection of some of the avatars participants could select, with varying hair, skin, clothing, and eye colors.

The virtual avatars were collocated to the actual participants' movements. Because of this, when participants hold the controllers in their hands, the virtual avatar mimics the movements of the participant's body, see 2. Normal maps, height maps, and specular maps were also added to the avatars' visual representation to make them look even more realistic. The avatars also had lip-sync integrated through Oculus Lipsync [27].

Figure 2: Left: User gesturing. Right: Collocated avatar mimicing the gesture.

3.2 Movement

Like with the avatar's visual representation, the physical movement of the avatars needs to be collocated as closely to the users' actual movement as possible. Background research related to non-verbal communication states that realistic simulation of body movement helps achieve the best forms of communicating non-verbally [40]. To properly mimic the users' gestures and pose using a 3-point tracking setup, inverse kinematics properly pose the avatar's arms and hands, as hand and arm movements are common ways one expresses themselves non-verbally.

3.3 Virtual Environment

The setting of the VE was an in-doors office building. Several objects were present here, such as a table, chairs, plants, paintings on the walls, and smaller decorative items, see figure 3. The VE uses Unity's HDRP to develop realistic lighting settings. The VE also had an ambient sound of what seemed like an office space sound. The participants were placed in two chairs with a table to precisely simulate this office meeting setting.

Figure 3: Image of the virtual environment where conversation takes place, rendered in Unity 3D with their High Definition Render Pipeline

3.4 Lip-syncing

Another relevant feature for communication in VR is matching the movement of the lips every time the user talks [40]. Developing lip-syncing made it possible to have the avatar's lips move according to the pronounced words. It is the easiest way to realistically animate an avatar's face and give it a semblance of life without delving into affection-recognition-programming.

3.5 Eye-movements

To properly implement the avatar's gaze, informed by eye-tracking, first, the system needed to be able to track when users were blinking and have the avatar do the same. Secondly, the fixation point of the virtual avatar should be at the point in space the user is currently gazing at in the HMD. This way, users can adequately perceive each-others gazes, and eye interactions, inferring the intention behind the gaze behavior.

To simulate gaze behavior without eye-tracking data. Blinking should be done at randomized intervals within a set range as informed by background research. The avatar's gaze point should alternate between gazing at and averting the conversation partner's gaze at randomized intervals within a set range. The point of gaze-aversion should also be randomized between several points to make the avatar appear less robotic in its gaze behavior. Finally, the avatars should have a static direction pointing forward if no gaze behavior is wanted.

3.6 Experiment Conditions

The experiment had three conditions. The first one was with eye-tracking enabled. This condition required a calibration process of the eyes to correctly measure where they looked in the VE and, therefore, accurately present it in their avatar. The second condition was with eye-tracking disabled. The third condition simulates eye-tracking by having the avatars blink and move within a set range informed by background research (see table 1).

3.7 System Limitations

The experiment used two computers simultaneously. They did not have the same hardware, but both maintained a framerate above 60, so participants did not get disrupted by clear lag spikes. HTC lists in their specifications that the HTC VIVE has a field of view of 110 degrees [26]. However, due to many factors such as the headset's fit, facial geometry, and the distance between the lens and the eyes, Lange et al. have identified a more appropriate field of view of 65 degrees [33]. With the added binocular addons, the field of view further decreases to about 58 degrees.

4 PILOT STUDIES

This section will present two pilot studies and their findings. They were conducted to understand better what should be included in the final evaluation. The first pilot study was conducted to understand how participants felt with their eyes being realistic in their gaze and movement. The second pilot study was conducted to understand how far away participants should be sitting from each other and whether or not gaze and eye movements affected the virtual meeting.

4.1 Pilot Study | Mirror Test

Before any significant experimentation, it is crucial to investigate whether participants will find the eye and gaze movements of the virtual avatars realistic. This is done to ensure findings are not drawn from a potentially flawed system that does not adequately represent the user's gaze to conclude how gaze behavior affects the feeling of embodiment, presence, and partner evaluation in VR. Therefore, a pilot study was developed where participants could report their feeling of avatar embodiment after inspecting their virtual avatar, its eyes, and eye movements based on live eye-tracking data in VR. The test was conducted on 9 participants, two females and seven males, aged between 22 and 28 years old.

Participants were first asked to select the avatar they identified with the most from a selection of avatars with differing gender, skin, hair, and eye colors, a selection of which can be seen in figure 1. The variety of avatars was implemented to lessen the difference in the feeling of avatar embodiment between a participant who identified with an avatar and one who didn't. After selecting an avatar and having the gaze-tracking calibrated in VR, participants were placed in a simple room in VR, where the only point of interest was a mirror.

The mirror was implemented using a planar reflection probe, accurately reflecting depth, the avatar, and the virtual environment. Participants were asked to closely examine their avatar's eyes and eye movements, as seen in figure 4. Once they felt they had adequately studied their gaze, they were asked to fill out an avatar-embodiment questionnaire.

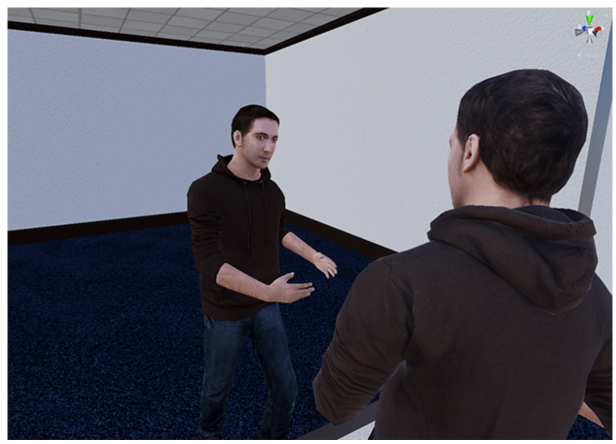

Figure 4: Participant investigating their avatar using a mirror in VR

The experiment was then repeated with a different condition, of which there were two. Condition A had realistically modeled and textured eyes, whereas condition B had eyes resembling ping-pong balls with no shading and flat colors. see figure 5. Each participant had the order in which they experienced flipped conditions to prevent order bias.

Figure 5: Eye appearance of condition A (left) and B (right)

Condition A, which has realistically textured and modeled eyes, scored 13.39/25.4, whereas Condition B, scored 12.94/25.4. See table 2.

While the total avatar embodiment score for Condition A is 13.39/25.4, it was reported that the latency between the participant's eye movements and the reflected movement of the avatar was noticeable,

	A	B
Ownership Average	Mean = 2.77 (SD = 6.51)	Mean = 2.33 (SD = 7.46)
Agency Average	Mean = 13.88 (SD = 1.83)	Mean = 13.44 (SD = 1.81)
Location Average	Mean = 3.55 (SD = 1.87)	Mean = 3.22 (SD = 1.71)
Appearance Average	Mean = 11.88 (SD = 4.85)	Mean = 12.55 (SD = 3.32)
Total Embodiment Average	Mean = 13.39 (SD = 4.73)	Mean = 12.94 (SD = 4.28)

Table 2: Results from avatar embodiment questionnaire, comparing condition A and B

with some remarking it harmed the experience. The latency is a result of hardware and software limitations. According to Pupil Labs' documentation, there is 8.5ms camera latency, and around 4ms processing latency, [5]. In a dyadic conversation, when perceiving someone else's avatar and not your own collocated avatar in a mirror, this small latency would probably not be noticeable. Results from the pilot study indicate that the implementation of gaze should be sufficient enough to use it as a basis to investigate the effects of gaze in VR in dyadic conversations.

4.2 Pilot Study | Partner Evaluation

The second pilot study was to determine whether there were any significant oversights in the implementation. It also examined whether eye-tracking in a dyadic conversation likely provided a noticeable difference in partner-evaluation co-presence, involvement, and face-to-face indicators.

The experiment was conducted on 10 participants, consisting entirely of men (due to availability) aged between 24 and 31 years old, with a mean age of 27. The experiment consisted of two five-minute conversations in an immersive virtual environment, modeled to look like an office, as research suggests people are more likely to engage in conversation while indoors [9]. The participants were seated a comfortable distance apart, at around 180 cm, as a conversation is more likely to occur when interactions are further apart (see section [9]. The experiment took place in two different locations at Aalborg University, so participants could be in separate rooms to ensure they would not be able to hear each other during the pilot study. Both setups required an HTC Vive headset (one with eye-tracking cameras), two controllers which functioned as the user's hands in the virtual environment, and two towers to track said HMD and controllers. Each conversation partner was placed in a chair as the virtual environment took place in a meeting setting where the avatars were also sitting down. One conversation partner was an author of this paper, while the other was a recruited participant. The pilot study had two conditions; one condition had eye-tracking enabled, meaning the participant would be able to see their conversation partner's eyes move, tracked to their real-life gaze. The other had eye-tracking disabled, meaning the participant would not know any gaze movements from their partner but a static forward-facing gaze.

The study was a within-subject study, where each participant experienced both conditions in different orders to avoid order bias. Participants would first fill out a demographics questionnaire before being guided to their seats and explained what would occur. Then participants would freely talk to one of the researchers for 5 minutes, located physically in a different room but across from the table in the virtual environment, as seen in figure 6.

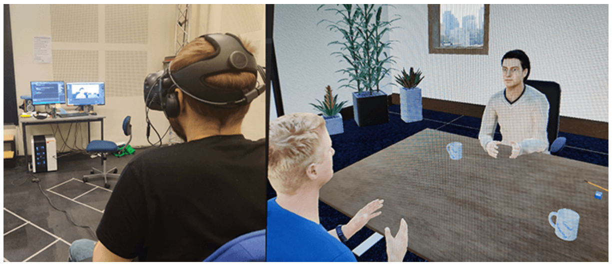

Figure 6: Participant alone in test room, though conversing with partner in an immersive virtual environment over network

The author partaking in the study was unaware of which condition was currently active and would calibrate their gaze for eye- and non-eye-tracking conditions. Topics would range from work, studies, hobbies, favorite foods, etc. Once the five minutes were up, participants would fill out a partner evaluation, co-presence, face-to-face, and involvement questionnaire for the condition. Participants would then return to their seat for another 5-minute conversation with a different condition before filling out the questionnaire again.

Looking at the results from the second experiment, there is very little difference. Condition A (eye-tracking) did slightly worse than condition B (no eye-tracking) on Face-to-Face and Involvement while doing slightly better on Co-presence and Partner Evaluation. (see table 3)

Condition	A	B
Face to Face	Mean = 37.6 (SD = 5.44)	Mean = 37.4 (SD = 6.93)
Involvement	Mean = 15 (SD = 2.10)	Mean = 15.4 (SD = 2.17)
Co-presence	Mean = 16 (SD = 2.26)	Mean = 15 (SD = 2.86)
Partner evaluation	Mean = 31.9 (SD = 3.69)	Mean = 31.8 (SD = 3.70)

Table 3: Average scores of condition A and B

Interestingly, the second condition always did significantly better than the first condition, despite it being condition A or B. This could be that participants have built a rapport with the examiner during the first 5-minute conversation round and are, therefore, more relaxed and open during the second conversation round. (see table 4 and 5)

	A first	A second
Face to Face	Mean = 35 (SD = 5.52)	Mean = 40.2 (SD = 4.38)
Involvement	Mean = 14.2 (SD = 2.16)	Mean = 15.8 (SD = 1.92)
Co-presence	Mean = 14.4 (SD = 2.07)	Mean = 17.6 (SD = 0.89)
Partner evaluation	Mean = 30 (SD = 4.47)	Mean = 33.8 (SD = 1.30)

Table 4: Average score of condition A as first and second condition.

	B first	B second
Face to Face	Mean = 35.4 (SD = 8.20)	Mean = 39.4 (SD = 5.54)
Involvement	Mean = 15 (SD = 2.82)	Mean = 15.8 (SD = 1.48)
Co-presence	Mean = 13.8 (SD = 3.49)	Mean = 16.2 (SD = 1.64)
Partner evaluation	Mean = 30.4 (SD = 4.72)	Mean = 33.2 (SD = 1.92)

Table 5: Average score of condition A as first and second condition.

4.3 Changes to Design from Pilot Studies

Our pilot studies brought several changes to our prototype before conducting the final experiment. A calibration room was added to ensure that the participants did not calibrate their gaze in front of each other, potentially disturbing their calibration. The second pilot study had participants point out that the distance between their conversation partners was too far and therefore had trouble seeing their partner's eyes due to the visual limitation of the HMD. Because of this, the table's length has been changed from 180cm to 120cm. Participants seemed to be surprisingly focused on the cups present in the virtual environment of the second pilot test with which they could interact. Most feedback was about these cups and how interaction with them could be improved. Participants would spend excessive time playing with the cups rather than focusing on their conversation partner. Therefore, interactable items will be removed for the final experiment. Participants also noted sounds from the real world distracting them while in VR. Therefore, ambient sound will be placed in the virtual scene to make the outside sound less obtrusive. The conversation time also increased from five to eight minutes since five seemed too short while ten proved too long in our second pilot study.

5 IMPLEMENTATION

This chapter contains a brief outline of the different technologies used to implement the experiment.

5.1 Game Engine & Eye-tracking Software

The experiment was developed in Unity 3D [50] version 2019.4.21f1 using its High Definition Render Pipeline [52]. To implement Virtual Reality functionality with 3-point tracking from HTC Vive HMD and controllers [26], the Steam VR package [16] was downloaded and implemented into the virtual scene to interface with Steam VR [45].

The HMD available with eye-tracking had eye-tracking cameras from Pupil Labs [31] fitted. The binocular add-on consists of two clip-on 200hz cameras with infrared illuminator rings [6]. Pupil Labs maintains a Unity package called HMD-eyes [30], which receives eye and gaze-tracking data from their Pupil Capture executable program [29], which can then be accessed in Unity.

5.2 Characters

The humanoid avatars used for the experiment were gathered from the Microsoft Rocketbox avatar library, which consists of multiple fully modeled and rigged characters, supplied for free by Microsoft for research and academic use [34]. Aside from being fully rigged and modeled, the avatars also have multiple blend-shapes of visemes, the visual counterparts of phonemes, the most minor units of sound that distinguishes one word from another, which can be used for lip-syncing.

Lip-syncing was achieved through the use of a package for Unity called Oculus Lipsync, which by processing an audio stream, matches visemes to detected phonemes, blending between the different viseme blend-shapes of the character to have the surface appear to be speaking [27, 35].

To have the virtual avatars follow the user's movements, FinalIK VR [41], a package for Unity, was used. By setting the avatar's head and hands to follow the position of the Vive HMD and controllers, FinalIK then inferred the real-life position and pose of the user with inverse kinematics, moving the body parts of the model to correspond to its estimate. FinalIK also allowed for limb-stretching to ensure that the hand position of the avatar always corresponded to the tracked position of the Vive controllers.

5.3 Gaze-, Simulated Gaze- & No Gaze-tracking Versions

To have the virtual avatar properly reflect the user's gaze, eye-tracking and gaze-tracking data from Pupil Capture was accessed by using the Eyes-HMD package [30]. Pupil Capture [29] solved for gaze direction and depth, which could then be used to calculate the position of the user's gaze in 3D space. Using Unity's LookAt() function, [51], the eyes of the virtual avatars were then directed to look at that point in space.

Blinking was achieved by blending two blend-shapes, one with eyes open and another with eyes closed. Checking for when Pupil Capture detected a blink, the eyes would be closed and not opened until the eyes could be confidently tracked again.

To simulate gaze behavior when no eye-tracking data was available, blinking was set at random intervals within a set range, and gaze points were alternated between the conversation partner and three aversion points; one to either side and one downward. The gaze-timings were set based on research by Argyle outlined in section **??**, table 1.

5.4 Multiuser functionality

Multiuser functionality was implemented using the packages Photon Pun 2 [38] and Photon Voice [37, 39], which uses Unity's built-in networking. Photon offers a free server, which was used that seats up to 20 users [36]. The packages allow one to connect clients to the network, synchronize game object transforms and game object ownership, and enable audio broadcasting from appropriate speakers in the scene.

6 EVALUATION

The evaluation was a between-subjects experiment to determine the effect social gaze has on the indicators of face-to-face, involvement, co-presence, and partner evaluation reported after a dyadic conversation in immersive virtual reality with collocated virtual avatars. Measurement was score-based and self-reported by participants using a questionnaire developed by Garau et al. [23] focusing on the quality of communication. Following the score-based part of the questionnaire, participants were free to answer some open-ended questions relating to gaze and their experience, in general, to discern the user's immediate thoughts following the experiment.

$H1_0$: Eye-tracking driven social gaze will not affect the quality of communication. **Hypotheses 1** Eye-tracking driven social gaze will affect the quality of communication.

$H2_0$: Simulated social gaze will not affect the quality of communication. **Hypotheses 2** Simulated social gaze will affect the quality of communication.

$H3_0$: Eye-tracking driven social gaze will not lead to a greater quality of communication than simulated social gaze. **Hypotheses 3** Eye-tracking driven social gaze will lead to a greater quality of communication than simulated social gaze.

6.1 Participants

Forty-five participants took part in the experiment: Fifteen for each of the three conditions. The age and gender distribution was as follows:
Condition A (eye-tracking): Age range 18-28, with a mean age of 24.9 (SD =2.52). 13/15 were male, while 2/15 were female.
Condition B (no eye-tracking): Age range 21-28, with a mean age of 23.4 (SD=1.96). 9/15 were male, while 6/15 were female.
Condition C (simulated): Age range 21-26, with a mean age of 23.13 (SD=1.54). 12/15 were male, while 2/15 were female., with one participant not wishing to disclose their gender.

Before the experiment, participants gauged the amount of time they spent playing video games per week and how many times they had tried VR before as it might have an influence on how they perceived the experiment as some might have no or few similar experiences, while others have many (see worksheet page X).

To ensure proper gaze-tracking: For all participants of condition A, their eye confidence (how accurately pupil capture determines it can estimate their gaze) was above 90%, well over what Pupil Labs deem acceptable, which is a confidence of 80% [30].

6.2 Experimental Setup

For each experiment, two setups were needed as participants would need to be present in the same virtual environment to converse but physically separated so as not to meet each other beforehand or hear each other in the same physical room.

Each setup consisted of two VIVE base stations, two HTC VIVE controllers, and one HTC VIVE HMD fitted with a binocular eye-tracking add-on from Pupil Labs and a computer capable of running the Unity 3D scene and pupil capture simultaneously at 60 frames per second. A researcher was present in each room to assist the participants with fitting the HMD and calibrating the pupil capture for each participant. An image of the setup in one of the rooms can be seen in figure 7.

Figure 7: Image of the experimental setup in one of the rooms.

6.3 Procedure

For each condition, two participants were recruited from the premises at AAU Create. Participants were brought down one by one to either of the two rooms reserved for experimentation to limit any prior interaction between the two; for the same reason, participants were from different semesters and fields of study.
While participants answered a brief demographics questionnaire, the researcher selected their virtual avatar among a selection of different gender, ethnicity, clothing, hair color, and eye color, which they thought resembled the participant best. After filling out the demographics questionnaire, the participants were seated and assisted with fitting the VIVE HMD and had the eye-tracking software calibrated to their gaze.

Being told to treat the experience as a first-time meeting between colloquies, a friend of a familiar friend, or a friendly stranger at a gathering, and to introduce themselves and discuss/talk about whatever they desired and found appropriate. Once both participants were ready, the researcher moved their virtual avatar from a separate calibration room into a shared room resembling an office. Their virtual avatars were present in chairs on opposite sides of a table. Here they were given eight minutes to converse. Following the experiment, participants filled out Garau et al.'s quality of conmmunication questionnaire and, afterward, some open-ended questions regarding their experience.

6.4 Measures

The main measure for this experiment was perceived quality of communication. Using the questionnaire developed by Garau et al., as mentioned in section 2.4. The questionnaire divides quality of communication into four broad indicators, briefly explained as the following:

(1) **Face-to-face -** The extent to which the conversation was experienced as a real face-to-face conversation.
(2) **Involvement -** The extent to which participants experienced involvement in the conversation.
(3) **Co-presence -** The extent to which it felt as if participants communicated with another human being rather than a computer interface.
(4) **Partner evaluation -** The extent to which the conversation was enjoyed and how positively participants evaluated their partner.

Questions relating to these indicators would be scored on a Likert scale from 1-9 (strongly disagree to strongly agree), such that a significance test can be run on the quantitative results. As there are three conditions with one changing variable, a One-way-ANOVA could be used to determine the significance of the results. However, due to the relatively low sample size, a One-way-ANOVA on ranks or a Kruskal-Wallis H Test should be used for non-parametric data. As well as a Mann-Whitney U test for pairwise significance.

7 RESULTS

Hypothesis 1 and **Hypothesis 2** was tested with a Kruskal-Wallis H significance test on all four indicators across the three conditions; however, no statistically significant difference could be proven. Furthermore, after doing a Mann-Whitney U significance test, comparing the conditions pairwise, no statistically significant difference could be proven.
Condition A outperformed the other conditions on all but one indicator, being involvement, while condition C had the worst response on all but one indicator, being partner-evaluation, as can be seen in table 6.

Hypothesis 3 was tested with a pairwise Mann-Whitney U test on conditions A and C for all indicators. For one indicator, face-to-face, condition A did perform better than condition C, with a mean value of 7.81 to 7.055 and a statistically significant difference. The significance values for the pairwise Mann-Whitney U test can be seen in figure 7.

7.1 Additional Results

Following the quality of conversation questionnaire, the participants had to answer three additional multiple-choice questions with some open-ended follow-up questions. It was to determine a clear and more rudimentary difference in experience between conditions. The questions and responses were as follows:
When asked if participants found their partner's gaze realistic: 62.5% answered yes for condition A, 37.5% for condition B, and 43.8% for condition C.

When asked if participants felt their conversation partner looked too much at them: 87.5% from conditions A and B reported their conversation partner looked at them a normal or acceptable amount. In contrast, 6% reported they looked at them too much, and 6% reported they looked at them too little. For Condition C, participants reported 93.8% looked at them a normal or acceptable amount, while 6.3% reported they looked at them too much.

Participants were free to elaborate on their reasoning for these answers and some of their responses will be shown and discussed in the following section.

8 DISCUSSION

Gaze-tracking-driven social gaze had no statistically significant effect on the quality of communication; therefore, both $\mathbf{H1_0}$ and $\mathbf{H2_0}$ could not be rejected. The indicator for a face-to-face response did have a higher mean for condition A than B and C (as seen in figure 6 and 8) and close to being statistically significant when considering all conditions in a Kruskal–Wallis H significance test. Perhaps with more participants, a statistically significant effect could have been proven. When considering the four indicators: face-to-face, involvement, co-presence, and partner-evaluation, face-to-face is the indicator where the most significant difference in response is expected, as it directly relates to how close to reality the interaction felt.

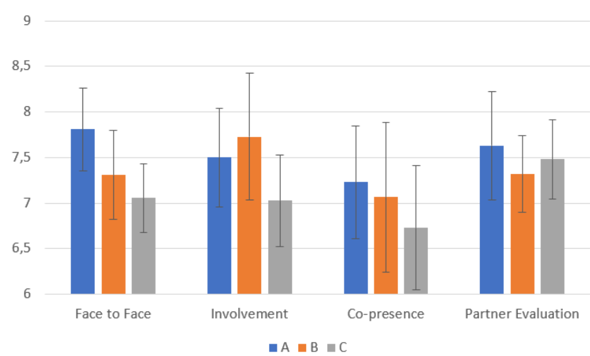

Figure 8: Comparison of mean responses between conditions.

$\mathbf{H3_0}$ could be rejected as comparing the indicator face-to-face between conditions A and C; the difference was statistically significant. Looking at the difference in means, 7.81 to 7.055, we can conclude that accurate eye-tracking driven gaze of an avatar outperforms simulated social gaze of an avatar.

There is value in discussing why $\mathbf{H1_0}$, and $\mathbf{H2_0}$ could not be rejected, as background research details how critical social gaze is to the quality of communication. One reason could be that in dyadic conversations, it is not as essential to signal intent, interest, and where one's attention is focused as it would be in a conversation with three or more individuals. In a conversation with multiple individuals, one's gaze can signal to others precisely to whom one's interest lies at the moment. Another reason could be the experiment itself; the gaze behavior was 'obscured' behind other factors like collocated movements, lip-syncing, and the virtual environment. Therefore, participants might not consider gaze as heavily when answering a questionnaire as they would if it had been evident that gaze was the experiment's focus. For example, in Garau et al.'s experiment, from which the questionnaire was sourced, they

	Condition A	Condition B	Condition C	Kruskal–Wallis H Test
Face to Face	Mean = 7.81 (SD = 0.907)	Mean = 7.31 (SD = 0.97)	Mean = 7.055 (SD = 0.757)	**p = 0.07**
Involvement	Mean = 7.5 (SD = 1.085)	Mean = 7.73 (SD = 1.385)	Mean = 7.03 (SD = 1.005)	**p = 0.12**
Co-presence	Mean = 7.23 (SD = 1.235)	Mean = 7.035 (SD = 1.635)	Mean = 6.73 (SD = 1.36)	**p = 0.54**
Partner Evaluation	Mean = 7.626 (SD = 1.186)	Mean = 7.32 (SD = 0.834)	Mean = 7.48 (SD = 0.874)	**p = 0.8**

Table 6: Table showcasing the Kruskal–Wallis H results for each condition of the Garau et al. quality of conversation questionnaire, as well as the mean and SD for each group.

	A-B Comparison	A-C Comparison	B-C Comparison
Face to Face	p = 0.19	p = 0.02	p = 0.38
Involvement	p = 0.3	p = 0.2	p = 0.06
Co-presence	p = 0.96	p = 0.23	p = 0.52
Partner Evaluation	p = 0.95	p = 0.73	p = 0.47

Table 7: Table showing the significance of results from a pairwise Mann-Whitney U significance test.

compared inferred gaze to random gaze, with only an avatar's face visible to participants, which means the gaze-behavior was in no way obscured by other factors, such as the virtual environment, the collocated avatars, and lip-syncing. [23].

8.1 Discussion of additional results

When examining the qualitative data gathered at the end of the questionnaire, it is clear that participants from group A did notice what can be considered realistic gaze behavior: Participant **A6 remarked:** *"Yes, I think it (gaze) moved naturally with the conversation."*, **A5 remarked:** *"I noticed that he was looking at me while I was talking, and looking around more while he was talking."*, and **A8 remarked:** *"The eyes moved and the head followed, which gave you the feeling of being really present."*.

When examining responses from other conditions, while positive, it focused more on the head movement than their gaze. **B3 remarked:** *"I could follow when he moved his head; that made the conversation feel more real."* **B12 remarked:** *"I noticed the head movement as we looked around the room. However, I did not pay attention to the eye movement."*. Participants from Condition C generally noticed the eye movement. However, judging from their answers, they could tell it was not the actual gaze behavior of their conversation partner but rather the simulated gaze. **C7 remarked:** *"It (the gaze) followed me often, but also sometimes wandered off a bit to look at the surroundings, which seemed very much like how it would be in a real conversation."* **C1 remarked:** *"I would say it's realistic enough as not to be distracting."*

When examining comments with a negative connotation, most comments from participants from condition A related to how uncanny the experience felt. **A2 stated:** *"A bit uncanny, the gaze did not seem to matter as much as head rotation did in the conversation."* they further commented: *"It was not super realistic, but that might be because of the model."* Most comments from participants from condition B worded that eye movements were missing (B1), that it felt like a deadpan stare (B2), or, as participant B8 put it: *"They did not blink, and eye movement was not present. This is a tough thing to simulate, I know, but also very important in a conversation. The eyes are the window to the soul."*. Participants **C4 and C2** mirrored many of the same opinions from the C group, namely that a person's gaze is hard to simulate and the eyes were *'weird,'* displaying the same gaze behavior regardless of the topic of conversation or who was talking. Interestingly while condition C scored the worst on all indicators but one, the highest ratio of participants who felt their conversation partner gazed at them an appropriate amount was from condition C; this is likely since averages from Argyle's work determined the gaze behavior (see table 1).

It is fair to conclude that participants generally felt the experience could feel uncanny for condition A. However, they recognized it was the reflected gaze of their conversation partner. While for condition B, due to the noticeable lack of eye movement, it did not feel as uncanny. Participants of condition B instead paid attention to the head movement of their conversation partner. For condition C, participants recognized that the gaze was simulated and instead relied on head movement to gauge interest and intent, the eye movement instead proving a somewhat distracting factor.

9 FUTURE WORK

Repeating the experiment with more test subjects could lead to a statistically significant distance between conditions, as there seems to be a trend in the data of condition A outperforming the other conditions. The subject count for this project only just met the bare minimum recommended amount, and had ties which complicates calculation of a P-value.

In order for participants to more actively utilize their gaze, one or two more conversational subjects could be included in the experiment to make it a conversation between multiple subjects. This way, participants are forced to use their gaze to signify intent and attention within the group while also being forced to more actively observe the gaze of others to infer their intent and attention.

In discussion, we theorized that the lack of a statistically significant effect could be due to too many confounding factors in the experiment. Meaning participants did not pay close attention to their partner's gaze, experiencing too many factors within the short timeframe of the experiment. One approach could be to increase the time spent conversing, allowing more time for participants to consider every factor. Another approach is to cut down dramatically on the scope of the experiment, making it similar to other experiments by Garau et al. [23] and Steed et al. [46] where the conversational subject is only visible from the shoulders up. However, we feel this negates the motivation for the experiment, which was to investigate how eye-tracking could affect conversation in immersive VR.

Finally, it could be useful to examine the results where more stylistic avatars were used, as is the case in Horizon Workrooms, which can more clearly communicate gaze and gestures with bigger eyes and hands, also negating any chance of an uncanny-valley effect.

10 CONCLUSION

This paper aimed to examine whether implementing gaze-tracking-driven social gaze in a dyadic conversation in VR would result in a statistically significant difference in the quality of communication compared to no social gaze implementation and simulated social gaze. While supported by the theory outlined in this paper and suggested by examining additional qualitative data, the results were inconclusive. However, it was proven that accurate eye-tracking-driven social gaze did make participants feel as if the conversation was closer to real-life than a naïve simulated gaze. For future works, perhaps a statistically proven effect between eye-tracking driven social gaze and no social gaze could have been proven on the indicator of face-to-face, which is the extent to how real the conversation felt compared to real life.

REFERENCES

[1] [n.d.]. "Talking without A Voice" : Understanding Non-Verbal Communication in Social Virtual Reality. ([n. d.]).
[2] 2022. *Horizon Workrooms Landing Page.* https://www.oculus.com/workrooms/.html
[3] 2022. *HTC Vive Pro Eye Landing Page.* https://www.vive.com/us/product/vive-pro-eye/overview
[4] 2022. *Microsoft Teams Landing Page.* https://www.https://www.microsoft.com/en-us/microsoft-teams/group-chat-software.html
[5] 2022. *Pupil-Labs.* https://docs.pupil-labs.com/developer/core/network-api/#latency
[6] 2022. *Pupil Labs Binocular Addon for VR.* https://docs.pupil-labs.com/vr-ar/htc-vive
[7] 2022. *Tobii.* https://vr.tobii.com/developer/
[8] 2022. *Zoom Landing Page.* https://https://www.zoom.us.html
[9] Michael Argyle. 2013. *Bodily communication.* Routledge.
[10] Michael Argyle and Janet Dean. 1965. Eye-contact, distance and affiliation. *Sociometry* (1965), 289–304.
[11] Michael Argyle and Roger Ingham. 1972. Gaze, mutual gaze, and proximity. (1972).
[12] Michael Argyle, Luc Lefebvre, and Mark Cook. 1974. The meaning of five patterns of gaze. *European journal of social psychology* 4, 2 (1974), 125–136.
[13] Michael Argyle and Marylin Williams. 1969. Observer or observed? A reversible perspective in person perception. *Sociometry* (1969), 396–412.
[14] Geoffrey Beattie. 1983. *Talk: An analysis of speech and non-verbal behaviour in conversation.* Open Univeristy Press.
[15] Gary Bente, Felix Eschenburg, and Nicole. C. Kramer. 2007. Virtual Gaze. A Pilot Study on the Effects of Computer Simulated Gaze in Avatar-Based Conversations. In *Virtual Reality*, Randall Shumaker (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 185–194.
[16] Valve Corporation. 2022. *Unity Asset Store: Unity SteamVR Plugin.* https://assetstore.unity.com/packages/tools/integration/steamvr-plugin-32647
[17] Robert J Edelmann and Sarah E Hampson. 1979. Changes in non-verbal behaviour during embarrassment. *British Journal of Social and Clinical Psychology* 18, 4 (1979), 385–390.
[18] RV Exline, I Gottheil, A Paredes, and D Winklemeier. 1968. Gaze direction as a factor in judgment of non-verbal expressions of affect. In *Proceedings 76th Annual Convention of the APA*, Vol. 3. 415–416.
[19] Ralph V Exline, Lewis C Winters, et al. 1965. Affective relations and mutual glances in dyads. *Affect, cognition, and personality. New York: Springer* 1, 5 (1965).
[20] Jon Fingas. 2022. *Tobii says it's in talks to provide eye tracking for Sony's PlayStation VR2.* Retrieved February 7, 2022 from https://www.engadget.com/tobii-playstation-vr2-eye-tracking-talks-155325972.html
[21] Margaret Foddy. 1978. Patterns of gaze in cooperative and competitive negotiation. *Human relations* 31, 11 (1978), 925–938.
[22] Nale Lehmann-Willenbrock Frank Steinicke, Annika Meinecke. 2020. A First Pilot Study to Compare Virtual Group Meetings using Video Conferences and (Immersive) Virtual Reality. *N/A* 1, 1 (2020), 1–2.
[23] Maia Garau, Mel Slater, Simon Bee, and Martina Angela Sasse. 2001. The impact of eye gaze on communication using humanoid avatars. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 309–316.
[24] Vijai N Giri. 2009. Nonverbal communication theories. *Encyclopedia of communication theory* (2009), 690–694.
[25] Erdan Gu and Norman Badler. 2006. Visual Attention and Eye Gaze During Multiparty Conversations with Distractions. *Lecture Notes in Computer Science* 4133, 193–204. https://doi.org/10.1007/11821830_16
[26] HTC. 2022. *VIVE - VR Headsets, Games, and Metaverse Life | United States.* hhttps://www.vive.com/us/
[27] Mohammed Hazim Alkawaz Itimad Raheem Ali, Hoshang Kolivand. 2017. Lip syncing method for realistic expressive 3D face model.
[28] Chris L Kleinke. 1986. Gaze and eye contact: a research review. *Psychological bulletin* 100, 1 (1986), 78.
[29] Pupil Labs. 2022. *Pupil Capture | Pupil Labs.* https://docs.pupil-labs.com/core/software/pupil-capture/
[30] Pupil Labs. 2022. *Pupil Lab HMD-Eyes Repository.* https://github.com/pupil-labs/hmd-eyes
[31] Pupil Labs. 2022. *Pupil Labs About.* https://pupil-labs.com/about/
[32] Mansur Lalljee. 1978. The role of gaze in the expression of emotion. *Australian Journal of Psychology* 30, 1 (1978), 59–67.
[33] Daniel Lange, Tim Stratmann, Uwe Gruenefeld, and Susanne Boll. 2020. HiveFive: Immersion Preserving Attention Guidance in Virtual Reality. https://doi.org/10.1145/3313831.3376803
[34] Microsoft. 2022. *Microsoft-Rocketbox.* https://github.com/microsoft/Microsoft-Rocketbox
[35] Oculus. 2022. *Oculus Lipsync Unity Integration Guide.* https://developer.oculus.com/documentation/unity/audio-ovrlipsync-unity/
[36] Photon. 2022. *Photon Engine PUN.* https://www.photonengine.com/pun
[37] Photon. 2022. *Unity Asset Store: Photon Voice 2.* https://assetstore.unity.com/packages/tools/audio/photon-voice-2-130518
[38] Photon. 2022. *Unity Asset Store: PUN 2 - FREE.* https://assetstore.unity.com/packages/tools/network/pun-2-free-119922
[39] Photon. 2022. *Voice For PUN2.* https://doc.photonengine.com/en-us/voice/current/getting-started/voice-for-pun
[40] Shane L. Rogers, Rebecca Broadbent, Jemma Brown, Alan Fraser, and Craig P. Speelman. 2022. Realistic Motion Avatars are the Future for Social Interaction in Virtual Reality. *Frontiers in Virtual Reality* 2 (2022). https://doi.org/10.3389/frvir.2021.750729
[41] RootMotion. 2022. *Unity Asset Store: Final IK.* https://assetstore.unity.com/packages/tools/animation/final-ik-14290
[42] Abigail Sellen. 1995. Remote Conversations: The Effects of Mediating Talk With Technology. *Human-Computer Interaction* 10 (12 1995), 401–444. https://doi.org/10.1207/s15327051hci1004_2
[43] Hans Stokking Omar Niamut Simon N.B Gunkel, Marleen D.W. Dohmen. 2019. 360-Degree Photo-realistic VR Conferencing. *Netherlands Organisation for applied scientific research* 1, 1 (2019), 1–2.
[44] John W Stass and Frank N Willis. 1967. Eye contact, pupil dilation, and personal preference. *Psychonomic science* 7, 10 (1967), 375–376.
[45] Steam. 2022. *SteamVR on Steam.* https://store.steampowered.com/app/250820/SteamVR/
[46] Anthony Steed, Ye Pan, Fiona Zisch, and William Steptoe. 2016. The impact of a self-avatar on cognitive load in immersive virtual reality. In *2016 IEEE virtual reality (VR).* IEEE, 67–76.
[47] Frank Steinicke, Nale Lehmann-Willenbrock, and Annika Luisa Meinecke. 2020. A first pilot study to compare virtual group meetings using video conferences and (immersive) virtual reality. In *Symposium on Spatial User Interaction.* 1–2.
[48] Susan Straus and Joseph McGrath. 1994. Does the Medium Matter? The Interaction of Task Type and Technology on Group Performance and Member Reactions. *The Journal of applied psychology* 79 (03 1994), 87–97. https://doi.org/10.1037//0021-9010.79.1.87

[49] Stephen Thayer. 1969. The effect of interpersonal looking duration on dominance judgments. *The Journal of Social Psychology* 79, 2 (1969), 285–286.
[50] Unity. 2022. *Unity Development Platform.* https://unity.com/
[51] Unity. 2022. *Unity Documentation: Transform.LookAt.* https://docs.unity3d.com/ScriptReference/Transform.LookAt.html
[52] Unity. 2022. *Unity High Definition Render Pipeline.* https://unity.com/srp/High-Definition-Render-Pipeline
[53] Alfred L Yarbus. 1967. Eye movements during perception of complex objects. In *Eye movements and vision.* Springer, 171–211.