Intimacy is Induced and Regulated Through Proxemic & Gaze Behaviour

A Study in Immersive Virtual Reality

J. KOLKMEIER

FACULTY OF EEMCS
MSC HUMAN MEDIA INTERACTION

GRADUATION COMMITTEE
prof.dr. D.K.J. Heylen
dr. G. Englebienne
J.H. Vroon, MSc

UNIVERSITY OF TWENTE.

19.06.2015

Acknowledgements

I would like to express my very great appreciation to my daily supervisor, Jered Vroon, for his support and guidance during this graduation project. I am also grateful to my supervisors Dirk Heylen and Gwenn Englebienne for their advice and assistance. My special thanks go to Lynn Packwood for reviewing grammar and spelling of my report, and for taking care of me throughout the past year. I wish to acknowledge the friendly and supportive environment in the Human Media Interaction group - special thanks to those that encouraged me to partake in the daily sportive activity. Thanks to my family for their support. Special thanks to my fellow students from the HMI master and Creative Technology bachelor. Especially to David Goedicke for his advice and for the extracurricular activities we organized, and to those that partook in them.

Jan Kolkmeier
Enschede, June 2015

Abstract

The goal of this study is to examine the relationship between gaze and proxemic behaviours during social interaction. Knowledge of this relationship could prove be beneficial for future design of artificial agents to better understand and employ these behaviours during social interaction, making the agents more believable and potent social actors. Existing theories on this relationship suggest that these behaviours subconsciously induce and compensate perceived intimacy in interaction partners. While the general validity of this claim has been shown, little work since has attempted to disentangle the single and joint effects of these behaviours more. In this work, we employ immersive virtual reality technology to simulate a meaningful social encounter, where virtual agents interact with participants in a dynamic fashion. Gaze and proxemic behaviours are manipulated dynamically, while participants gaze and proxemic responses are measured on-line. Participant showed strongest gaze and proxemic responses when agents manipulated both proxemic and gaze manipulations at the same time. More intimate manipulations such as standing closer and seeking more mutual gaze elicited gaze aversion and increase of personal distance from the participants. Less intimate manipulations such as increasing distance and averting gaze elicited more mutual gaze and reduction of personal distance from the participants. Agents that only manipulated gaze elicited weaker responses compared to agents that only manipulated proxemics.

Contents

List of Figures

List of Tables

1. Introduction

Artificial agents - such as robots or virtual characters - are becoming more pervasive in society. In the real world, we come in contact with robotic agents that have a mind of their own or are teleoperated by others. With head-mounted virtual reality displays, interaction with our own and other virtual selves happens from a perspective that is more immersive than ever before.

The space we act in, be it virtual or real, is shared with an increasing number of artificial actors. When acting in any social context, we exhibit a dynamic set of nonverbal behaviours, some more subtle than others. They are dynamic in that they are a constant back and forth between the involved social actors. We read and express nonverbal responses - often subconsciously.

As designers of artificially intelligent systems, we wish to understand these behaviours and use them in our agents to better grasp and act in social situations, making the agents more believable and potent social actors.

Figure 1.1.: The stereotypical uncomfortable-elevator-situation

In Figure 1.1, the stereotypical elevator situation is depicted. Why do we feel uncomfortable when using a crowded elevator, and how does this feeling change our behaviour during the experience? Passengers avoid looking each other in the eye, as - if we may anticipate - maintaining eye contact while being so physically close would be uncomfortable. In a less confined space however, the same group of people would spread out and eye contact would not be perceived as at all uncomfortable.

In this work we want to dedicate our attention to these two social phenomena that have been shown to have strong effect on social interaction in general, as well as each on other:

Regulation of eye contact and interpersonal distance.

A relationship between eye contact and interpersonal distance was first formalised by Argyle and Dean [1]. Their Equilibrium Theory states that in social interaction, actors attempt to keep a comfortable and contextually appropriate intimacy level. A social actor maintains this equilibrium by regulating interpersonal distance, amount of eye contact and topic of conversation. This theory has been tested and extended in various studies (e.g. Coutts and Schneider [2], Patterson [3], Cappella [4], Rosenfeld et al. [5]) with varying methodologies and results supporting its general validity. In later studies by Bailenson et al. [6, 7] and Wieser et al. [8], immersive virtual environment technology (IVET) was used to revisit this Equilibrium Theory.

Their IVET is a virtual space that can be accessed through a head-mounted virtual reality display. Movements of the user inside the physical world are tracked and translated into movements in the virtual world, allowing a sense of being present in this virtual world. The promise of using IVETs lies in the greater experimental control of computer simulated worlds. In their recent review on the use of IVET to study social interaction, Bombari et al. [9] emphasize the importance of *standardized interaction partners*, which IVETs can provide in the form of virtual embodied agents.

Bailenson et al., among others, found that in their IVET, participants behaved towards virtual agents in the way that psychological theories such as the Equilibrium Theory would predict.

While such findings give support to the validity of Equilibrium Theory, they did not contribute much to further disentangle what the single or joint effects of the examined behaviours are. In this work, we will create a simulation of a meaningful social encounter in an immersive virtual environment where virtual agents interact with participants in a dynamic fashion. In this simulation, we will be able to let agents change their behaviours dynamically, while participants responses are measured on-line - therefor not sacrificing experimental control. What is more, not only will we manipulate a combination of both gaze and proxemic behaviour during the social interaction, we will also use the technology to put behavioural measures in place that record user responses in these same two dimensions. This, to our knowledge, has not been part of an experimental design in the area so far.

The resulting contribution from our approach should give more insight on the relationship between gaze and proxemic behaviour, their single and joint effects on themselves and on each other - in the context of immersive virtual reality environments.

We formulate our hypotheses as predictions of behavioural responses to different gaze and proxemic behaviours exhibited by a virtual agent. The predictions of Argyle and Dean's Equilibrium Theory, which we will present in more detail in Section 2.3, were used to inform the following hypotheses:

- H1 Increasing proximity of the agent towards the user (moving closer) will be compensated for by the user by moving more away from the agent - compared to decreasing proximity of the agent to the user, where the user will move more towards the agent (proxemic compensation).

H2 Increasing gaze of the agent towards the user (more eye contact) will be compensated for by the user by looking more away from the agent - compared to decreasing gaze of the agent towards the user, where the user will look at the agent agent (gaze compensation).

H3 Besides proxemic compensation, gaze compensation will also be observed during changed proximity of the agent to the user.

H4 Besides gaze compensation, proxemic compensation will also be observed during changed gaze of the agent towards user.

H5 When non-contradicting behaviours are combined (increased gaze and increased proximity), users responses will 'add up':

a) increased gaze & increased proximity have greater effect on proxemic compensation than only increased proximity

b) increased gaze & increased proximity have greater effect on gaze compensation than only increased gaze

c) decreased gaze & decreased proximity have greater effect on proxemic compensation than only decreased proximity

d) decreased gaze & decreased proximity have greater effect on gaze compensation than only decreased gaze

In the following chapter we will examine the related work. First, we will review research on effects and simulation of gaze and proxemic behaviours to inform the design of our agent behaviours. Next, we will discuss the Equilibrium Theory and why it is a suitable starting point on the way to answering our research question.

To determine agent behaviours that could serve as baseline as well as 'increased' and 'decreased' variations of both gaze and proxemic behaviours we performed a pilot study. This study and the choices made as a result of it are documented in Chapter 3. In Chapter 4, we will present a framework of the relationship between gaze and proxemic behaviours and their effects. We will specify the behaviours based on the findings in our pilot study, and formulate how we can use these in an experiment to test our hypotheses. In Chapter 5, we will present the main material of the experiment, the IVET. We will then document and report the setup and results of the conducted experiment in Chapter 6. Lastly, we will present our conclusions in Chapter 7.

2. Related Work

In this chapter we will provide literature reviews on the topics related to our research. We will first introduce research on gaze and proxemics in Sections 2.1 and 2.2. Here, we are particularly interested in earlier studies that have examined the effects of gaze and proxemics on other behavioural attributes that could be measured using the virtual reality method.

In the context of this work, we are specifically interested in the interaction between gaze and proxemic behaviours. The Equilibrium Theory, which we will discuss in detail in Section 2.3, is a psychological theory on nonverbal regulative behaviours between individuals. We used the Equilibrium Theory generate our hypotheses on the effects of gaze and proxemic behaviours and to inform design choices for the behaviours of the virtual agents.

In the last section of this review, we will look at previous work on using Virtual Reality as a method to examine social behaviour and interaction in general.

2.1. Gaze

Gaze describes the visual attention of a human manifested in direction of the eyes and by extension the orientation of head and body, typically in a social context [10, 11]. In conversation, gaze is used to regulate the flow of conversation, turn-taking, and requesting listeners to provide backchannels or express emotions (see [12, 13, 14, 15] and [16] for a survey). There are a number of definitions and concepts related to different kinds of gaze, as summarised by Mutlu [17]: *One-sided gaze* describes the situation where one individual looks the other in or between the eyes, or, more generally, in the upper half of the face [13]. If gaze is reciprocal, it is referred to as *mutual gaze* where both individuals look into each others face, or eye region, thus acting simultaneously as sender and recipient [18]. When an individual exhibits *averted gaze*, he avoids looking at the other, especially if being looked at, and/or moves his gaze away from the other [18, 10]. Other concepts, such as *joint attention*, *shared attention* and *gaze following* relate to how interaction partners act in triadic constellations where attention shifts to objects or points in space. But what effects on behaviour do situations such as averted or mutual gaze have, and what other factors play a role?

The two recent surveys by Pfeiffer et al. [19] and Ruhland et al. [20] summarize research on gaze from a psychological and technical standpoint, respectively. It becomes apparent from both that a large body of research on social gaze deals with determining and describing intentions and attention during social interactions, but little research on behavioural effects of mutual or averted gaze is found outside the work that we will discuss in Section 2.3. On the technical side, the focus is on rendering and simulating realistic gaze behaviour in artificial agents - both virtual and robotic. Artificial agents have been shown to be able to communicate or elicit attention [21, 22, 23, 24, 17], express emotions [25, 26, 27, 28] and utilize nonverbal cues during conversations effectively [29, 30, 31].

Most of these studies use subjective or task performance measures for validation. Only in some cases physiological or behavioural effects of different (aspects of) gaze behaviour are examined [32, 33, 6, 7]. Ioannou et al. [32] employ a physiological measure in their study using a thermal infrared imaging. They measure changes in facial temperature of participants manipulating gaze of a virtual agent. During mutual gaze, increased temperatures were observed compared to the temperatures during averted gaze. Kuzuoka et al. [33] uses manipulates the orientation of their information-presenting robot to create joint attention with visitors to the exhibition piece. They found that this would result in spatial reconfiguration of the visitors, following the principles of Kendon's F-Formation [34]. Bailenson et al. [6, 7] revisited the Equilibrium Theory in their immersive virtual reality experiments with artificial humanoid agents. They manipulated the realism of a virtual agent's gaze behaviour, testing effects on participants' proxemic behaviour. Participants wore head mounted stereoscopic displays with positional tracking to navigate in the virtual environment without the need of additional input devices. In memory tasks that involved participants moving through virtual space to read something from the back of the virtual agent, participants kept a greater minimum distance from the agent when it was looking at them more realistically. These results coincide with previous sociological findings in proxemics and the Equilibrium Theory. In Bailenson et al. [7], effect of gaze was dependent on agency of the virtual human - an effect could be measured in the agent condition, however not when the virtual human was introduced as an avatar.

2.2. Interpersonal Distance

Interpersonal distance is the distance individuals keep towards each other in social situations. Hall's *proxemics theory* [35] approaches this distance by describing bubbles at different distances around individuals. These bubbles relate to the interaction that takes place in them, when implicit social norms are adhered to. As depicted in Figure 2.1, from inside out we have first the *intimate space*, with a radius of approximately 45 cm. In this space, couples and parents with their children interact. Next, in the *personal space* bubble (45-120 cm), interactions with groups, associates or with close friends are accepted. In the *social space* bubble (120-240 cm), individuals accept interaction with acquaintances and strangers, whereas the outermost bubble is reserved for public interaction, such as public speaking.

In more recent work, the proxemic theory is typically used to automatically infer rela-

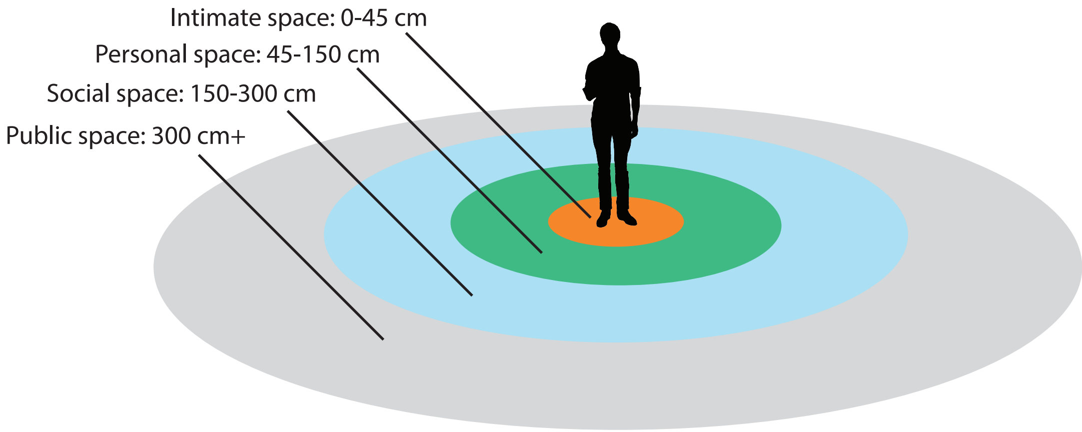


Figure 2.1.: Hall's model of personal space

tionships between humans, typically for surveillance, human-robot interaction purposes [36, 37, 38, 39, 40] and group or crowd simulation [41, 42, 43]. There is only little research where proxemics behaviour was intentionally manipulated to measure or predict behavioural responses in others [44, 45, 46, 47, 8].

Friedman et al. [44] used a Second Life[1] bot to observe other players proxemic behaviour, and found that they adhere to similar rules as suggested by Hall's personal space theory. Not a behavioural but a physiological measure was employed by Llobera et al. [45]. They measured skin conductance of participants that were approached by abstract objects, individuals and groups in virtual reality. They found heightened arousal at closer distances, but no significant difference between virtual objects and humans. Similarly, in the same study referred to in Section 2.1, Ioannou et al. [32] also measured facial temperature of participants when a virtual agent changed interpersonal distance. Increased temperatures were observed when interpersonal distance was reduced. In their experiment on perceived interpersonal distances in virtual and augmented reality, Obaid et al. [46] measured the loudness of participants' voices. They found that participants increased the loudness of their voice when the virtual agent was further away. Kastanis and Slater used a reinforcement learning method to train a virtual agent to move participants to a specified location [47]. The agent's valid actions in the learning process were idle, approach, retreat and 'waving', where the agent would ask the participant to come closer, accompanied by a waving animation. Based on proxemics, it was predicted that the agent could learn to move the participant backwards by approaching the participant closely, to which the participant would respond with retreating. In one condition, the closest allowed distance was 38 cm, whereas in the other condition, the closest allowed distance was 120 cm. In the condition where smaller distances were allowed, the agent could move most participants to the desired position in a short time, whereas in the other condition, the agent was only successful in just about half the cases, taking significantly longer.

[1] `http://en.wikipedia.org/wiki/Second_Life`

2.3. Interaction of Gaze and Proxemics: Equilibrium Theory

Based on their work on small scale non-verbal behaviours during social interaction between individuals, Argyle and Dean proposed the Equilibrium Theory [1]. This theory states that during co-located interaction an equilibrium of 'intimacy' develops. Their concept of 'intimacy' is a joint function of verbal and non-verbal behaviours such as eye contact, physical proximity or intimacy of the topic. The equilibrium state would be reached where none of the interaction partners feels the need to adjust any of these behaviours, that is to say, they feel *comfortable*. If, in one of its dimensions, the equilibrium is disturbed or cumbered, Argyle and Dean predict that participants will adjust their other behaviours to restore it.

In experiments with dyads, they supported their theory. In particular, interpersonal distance and amount of eye contact were shown to be inversely correlated. Individuals seated closer to each other exhibited more averted gaze, whereas those seated further apart exhibited more mutual gaze. Also, individuals regulated their interpersonal distance to other social actors.

Argyle and Dean also make suggestions about the underlying psychological motives for compensation of too low or too high intimacy. When intimacy is low, this motivation would be the desire for satisfying affiliative needs or desire for visual feedback, whereas fear of revealing inner states to fear of rejection by others is suspected to be the force behind compensation of high intimacy. This is similar to the motivation Hall gives to explain the existence of his personal space bubbles, reporting that individuals feel discomfort, anger or anxiety when social interaction falls outside these norms [35]. Relating Hall's model to the Equilibrium Theory further suggests that different equilibrium states exist for interpersonal distance which depend on the relationship between interacting partners.

Argyle and Dean's definition of the level of intimacy, from here on (ILS), is almost mathematical and gives intuitive predictions when combined with their explanation of the underlying motivations. The Equilibrium Theory is suitable for our purposes in that it makes clear predictions on the interaction between behaviours, and at the same time suggests a quality that these behaviours - which first have to be designed in the case of a virtual reality method - can be evaluated against: the *perceived intimacy* they elicit from an observer.

Argyle and Dean do not give an unambiguous definition of which behaviours should be included in the equilibrium. They only list verbal intimacy, gaze, proximity and "etc.". This has inspired various extensions to the Equilibrium Theory. Others such as Mehrabian and Patterson suggested lean, touch, body orientation and latency of response. Patterson [3] also provided further empirical support for the Equilibrium Theory and found that at close proximities, body orientation was also used to regulate intimacy. What is more, they found that only behaviours that mediated at least a minimum change in affect would also elicit compensatory adjustments from the interaction partner. Mehrabian [48] found that participants displayed more gaze aversion behaviour when being approached by an imaginary person they disliked rather than liked, suggesting that attraction also played a role in the equilibrium.

Patterson [49] further notes that there are also some counterintuitive findings. Some studies found that in some cases, intimate behaviour was not compensated for but reciprocated [50, 51], for example, when confederates touched subjects during experiments [50].

These extensions and remarks aim to explain more variance in observed behaviour. Our work however focuses on gaze and proxemic behaviour. When using the virtual reality method, selected behaviours can be manipulated while others are kept constant. This method is more robust against variance introduced by behaviours that have not been considered or controlled - which may be the case in observational experiments and experiments with human confederates. This is also what makes the Equilibrium Theory so attractive, as it predicts that when dimensions in the intimacy equilibrium are set constant, as is the case with deterministic animation of virtual humans, compensation follows in response to those behaviours that do change. However, we must also be aware that the response of a human to a virtual agent may still follow in any dimension, This needs to be registered in the measurements - which, of course, is not possible for all behaviours in great detail.

Concluding, Argyle and Dean's Equilibrium Theory is a suitable foundation for establishing hypotheses that can be tested using the virtual reality method. It further informs the requirements of the behaviours to be designed for the virtual agents. This enables us to make meaningful connections between observed responses and the psychological mechanisms that they were motivated by.

2.4. Behavioural Measures in Immersive Virtual Reality

A number of studies mentioned in the reviews above made use of virtual reality or immersive virtual reality technology to simulate gaze and proxemic behaviours on virtual humans. While many of these studies took subjective measures, physiological and behavioural measures were also employed successfully in studies examining the effects of gaze and proxemic behaviours. Most notably in the afore mentioned work by Bailenson et al. [6, 7], where immersive virtual environment technology (IVET) was used to revisit Equilibrium Theory successfully.

It stands to reason that the immersive virtual reality approach is a viable one for our purposes of examining the effects of using behavioural measures.

Presence One factor that is often mentioned when talking about virtual reality - particularly using technology beyond regular screens as means of experiencing the virtual environment - is *presence*. Witmer and Singer define presence as *the subjective experience of being in one place or environment, even when one is physically situated in another* [52].

It seems natural to assume that higher levels of presence are a desirable quality for virtual environments. One would expect that behavioural responses to cues in virtual environments correspond more to responses to similar cues in the physical world when a (high) feeling of presence is achieved in the user. Questionnaires such as the one of Witmer and Singer [52] aim to measure the level of presence in users after they have had a VR experience.

2.5. Conclusions

Concluding, a number of previous studies found that gaze and proxemic behaviours have measurable effect on others' behaviours during social interaction. The Equilibrium Theory and its extensions have suggested an intearaction between gaze and proxemic behaviour, in that they are both used during social interaction to continuously change and restore an equilibrium of intimacy. Empirical studies have supported this - to some extend even in immersive virtual reality experiments.

Considering the design of behaviour for virtual agents, few studies have specifically described and examined agent behaviours that are designed to mediate different levels of intimacy. We will address this in the following chapter in the form of a brief pilot study where we, based on qualitative evaluation, design behaviours that elicit different levels of perceived intimacy in the user of a prototype IVET.

What is more, earlier experiments in immersive virtual reality were limited to the manipulation of one behaviour in the agent, and the measurement of another in their participants. Our experiment will address that by manipulating combinations of gaze and proxemic behaviour in the agent, and look for both the gaze and proxemic responses in the participant. This way we want to disentangle the single and joint effects of these behaviour further. In Chapter 4 a framework is presented that illustrates this further, and explains how we can test our hypotheses.

3. Pilot Study on Intimacy-mediating Behaviour Design

In this chapter we will document a pilot study on the design of agent behaviours. We were interested in gaze and proxemic behaviours that would change the perceived intimacy when facing the agents in virtual reality. Based on the literature, some general rules are apparent: For gaze, a lot of eye contact means increased intimacy, whereas averted gaze elicits decreased intimacy. For proxemics, closer is more intimate, further away is more intimate, and some have suggested that body orientation has a role as well.

However, since we were aiming at a less robotic, more believable simulation of behaviour, we considered going further in our design. The findings from work that builds on the Equilibrium Theory typically do not go into more depth describing or even testing the dynamics of the involved behaviours. In the case in the body of work on artificial creation, there is little work that deals specifically with behaviours that mediate intimacy.

Therefore, the goal of this pilot study was to explore and evaluate qualitatively several variations of gaze and proxemics agent behaviours in terms of their intimacy-related qualities as well as their believability.

3.1. Approach

Two virtual agents were placed inside a virtual environment (see Figure 3.1), which could be experienced through an Oculus Rift DK2 HMD. This virtual environment was created in the Unity3D[1] game engine and editor, and acts as the prototype of the IVET that will be described in Chapter 5. The agents' gaze could by animated procedurally by means of setting a target in virtual space to look at and offsetting the gaze direction by an angle. Targets could be the user's head, the other agent's head, other objects in the scene, or an invisible point in front of the belly of the agent. The agents' proxemics towards the user could be changed by 'hovering' the agent forwards or backwards, letting the agent take steps forward or backwards as well as leaning towards the user or away from him.

In total, nine gaze and three proxemics related behaviour trees were tested and evaluated qualitatively by the researcher in terms of perceived intimacy-related qualities and realism. Behaviour trees were created using PlayMaker[2], a visual scripting editor to create Finite

[1] unity3d.com

[2] hutonggames.com

Figure 3.1.: Agents used during pilot study.

State Machines (FSMs). These FSMs control the functionality described above. They can be found in Appendix A.

3.2. Gaze

In the first nine implemented gaze behaviour trees, we examine differences between the use of different gaze targets, durations of maintained gaze, animation speeds and interaction rules. The *Random* tree was typically used as a baseline to compare against the other nine. We alternated which of the two agents would use the baseline, and which would use the other behaviour tree, to compensate for effects of appearance.

3.2.1. Random

In this behaviour tree, the agent alternates his gaze target between the user and the second agent. After each change in gaze target, the agent would wait a random amount of time would before he would change the gaze target again. Here, we experimented with the range from which the random amount of time could be selected.

We found that if the range was too small and the times were too short, the agent behaviour would look very unnatural, especially when both agents use this same behaviour, since gaze target changes would tend to synchronize and often overlap between both agents. Also the high frequency of change was found to be 'irritating'. Selecting the range to be wider - at least 3 but at most 8 seconds - yielded very believable behaviours where gaze changes were not consistently fast, and it would rarely happen that both agents would change gaze at the same time. We kept the random tree with this configuration as a baseline behaviour to compare others against.

Figure 3.2.: Averted gaze using a virtual gaze target.

3.2.2. Avoid Mutual

In this tree, the agent would randomly change between the following 'legal' targets: the user or other agent *that is currently not looking at the agent*, and a target in front of the agent's belly (averted gaze, see Figure 3.2).

This behaviour can be best described as 'creepy'. Especially so when the user is stared at when they are not directly looking, until they look directly at the agent, upon which the agent suddenly 'shies away'. While the staring part feels intimate if one is aware of it, once the agent looks away, perceived intimacy is much lower.

3.2.3. Avert using Offset

Here we implemented a gaze aversion behaviour where the agent does not change it's gaze target to the virtual point in front of his belly (as in Figure 3.2), but rather adds an angular offset to the direction towards the current gaze target.

This method feels much more natural than the first implementation. Just a 10 degrees angle in 'down-right' direction already give a good sense of averted gaze (see Figure 3.3). Also the animation to change the gaze are less outstanding, while still communicating the cue to the observer.

3.2.4. Reciprocate Max

In this tree the agent looks at the user with mutual gaze whenever it is detected that the user is looking directly at the agent. As long as the user is looking at the agent, mutual gaze is kept - but no longer than a certain *reciprocation time*. Then/otherwise, look at the other agent.

Figure 3.3.: Averted gaze by offsetting gaze from current target.

Changing the reciprocation time, mutual gaze felt most ‘comfortable’ when held for more than four seconds. The longer the gaze, the more intimate it feels, and at more than ten seconds of mutual gaze, if feels like staring. If the reciprocation time is shorter (around 2.5 s), it feels as if the agent averts his gaze, which feels distant, but not ‘creepy’ as in the previous case.

3.2.5. Reciprocate Prolonged

In this tree the agent looks at the user with mutual gaze whenever it is detected that the user looks directly at the agent. As long as the user looks at the agent, mutual gaze is kept. Once the user is looking away, the agent waits some *extra time* until he also changes gaze to a new target.

When being being gazed at, prolonged gaze time only feels natural between two and three seconds. It does feel noticeably more intimate when the prolonged time is much longer than that.

3.2.6. Eyes, Head & Chest Weight

In this tree we play with the animation of the gaze. The procedural animation allows us to also change to what extent only the eyes, head and/or chest rotate towards the gaze target.

Increasing the amount of rotation towards the target from chest to head to eyes, where chest is around 50%, head around 80% and eyes are 100% looks most realistic, at least for the gaze changes in the triadic setting. In terms of perceived intimacy, differences are not very striking, although it is more apparent with the agent that has wider shoulders and muscular chest.

3.2.7. Gaze Speed

Here we experiment with different animation speeds of gaze shifts, which could be set in degrees of head rotation per second.

Very contextual, but in general, 120 deg/s fits most cases well. It does feel a little slow when the agent is averting the gaze while not talking, but a little fast when the agent is talking. Higher or lower speeds however do not have a particular effect on perceived intimacy.

3.2.8. Match Dialog

Another experiment was to time gaze shifts in a meaningful way during the agent's turn of speech. From the lipsync module (see Section 5.1.5), start and end of dialog parts as well as silence moments were sent as events to the behaviour tree, and used as triggers to change gaze in different ways.

Averting at silence moments seems just unnatural. Avert when talking fits better. Gazing at the user during silence moments as well as at the beginning of dialog parts look natural, but it is also very dependent on the content of the dialog. Perceived intimacy increases when one feels directly addressed by the agent.

3.2.9. Follow Gaze, shared attention

For this behaviour tree, virtual targets such as a chair and a picture on the wall were incorporated. Whenever the user would look at one of these targets, the agent would first look at the user, and then look at the same target.

How natural this behaviour was perceived, was found to be heavily dependent on the spatial configuration between the user, the agent and the target. It could be very convincing if the agent was not required to assume a wrenched poses when alternating his gaze. This was due to the implementation of the procedural animation, which did not allow for rotating the entire body. The perceived intimacy was certainly low, when attention went to the object, and it was understood that the agent was observing the object as well. However, to exploit this further, more intelligent spatial reconfiguration behaviour would first be needed.

3.3. Proxemics

In these last three implemented gaze behaviour trees, we explore different animations, animation speeds and magnitudes of displacements that can be used to implement proxemic behaviours.

3.3.1. Hover

We displace the agent towards or away from the user, without any animation to explain this displacement, at different speeds[3] and with different magnitudes of the displacement in positive and negative direction.

If the displacement happens too fast, this behaviour draws immediate attention to the conflicting visuals (i.e. no foot movement). Only when very slow and subtle, it is not immediately apparent that the agent is approaching. From a certain closer distance on, even if the same speed is maintained as before, the approach becomes more and more apparent. Strong perception of intimacy is found when being very close to the virtual agent and perceived intimacy seems to increase faster the closer the agent becomes. A comfortable 'talking distance' to the agents seems to be between 75 and 90 cm. Perceived intimacy starts increasing *noticeably* when distance becomes smaller than 60 cm. Distances bigger than 100 cm were feeling too distant for regular conversation, although here, a contributing factor was that due to the resolution of the head mounted display, the agent's face became harder to 'read' at that distance as it was, due to perspective, rendered with far fewer pixels.

3.3.2. Lean

Instead of hovering, we attempted to use bend the agent procedurally forward and backwards to create a leaning animation.

For the leaning to be noticeable, the agent would have to be situated at an already close distance, say around 60 cm. Then, leaning forward would also change the perceived intimacy, although less so than moving the entire body. Leaning backwards did look a little unnatural. It should probably go in hand with changing posture, such as crossing arms. As noted before, the implementation of the procedural animation would also sometimes yield wrenched poses when the avatar was facing in one direction, while bending towards the user in a different direction.

3.3.3. Step

Lastly, we realised a behaviour to change interpersonal distance by using small step animations.

In terms of perceived intimacy, the same findings hold as for the *hover* approach, however now the the visuals are much less conflicting - although the foot placement is far from perfect. When the agent makes a step, the whole body - also including the hips - is animated accordingly. So when looking at the upper body, one can already understand the agent's behaviour.

[3] Speed was implemented as an arbitrary factor, hence no unit is provided.

3.4. Conclusions

In this pilot study some concepts around the realisation of dynamic agent behaviours related to gaze and proxemics were explored. Focus was both on what mediates different levels of intimacy, and what makes the behaviour more or less believable.

In terms of gaze, animation speed did not influence perceived intimacy. A value for animation speed was found that, while not perfect, fits most situations. Using an angular offset to produce averted gaze would stand out less than looking downwards, while still communicating well that the agent's gaze was not directed at the user anymore.

More intimacy was perceived the longer an agent would stare. However, a salient point was found where staring became 'creepy' and unnatural. Averted gaze was found to communicate less intimacy.

In terms of behaviours to change interpersonal distance, animating small steps on the agent when displacing him was more believable than simple hovering, and easier to implement reliably than bending.

We were able to have the agents mediate more or less intimacy through displacement towards or away from the participant. We relate the distance values we found to Hall's model (see Figure 2.1) and find that they roughly agree. We would have expected that intimacy would be perceivable halfway inside the 'personal space' (at around 80 cm), but this was only the case from 60 cm and closer. The tolerance for close behaviour seems to be bigger in our VR implementation than in the physical world. Consider however that we deliberately noted the distances where perceived intimacy would *change drastically*, whereas Hall's model presents general areas for different types of interaction, thus not necessarily related to the perceived intimacy that we report. In fact, perceived intimacy was already degrading from 100 cm+, but here the mentioned lower resolution that makes the face less easy to read was a contributing factor.

In the following chapter we will define a framework that will make the ties between the Equilibrium Theory, our hypotheses and how to test them in an experiment. There we will also explicitly define the required agent behaviours, drawing on the findings we have described in this chapter.

4. Framework

Based on our hypotheses, we have expectations on the interactions between virtual embodied agents and users that meet these agents in an immersive virtual environment. In particular on the users' responses to changed levels of intimacy as mediated by different behaviours. In this chapter, in anticipation of the experiment design, we will make explicit the relationship between behaviours and their effects, what is manipulated and what is to be measured, in order to test our hypotheses.

4.1. Agent Behaviours

Following the Equilibrium Theory, a compensation in the user would be expected after a *change in agent behaviour* that has impacted the intimacy level of the situation (ILS).

To be more explicit about this, we define a change in agent behaviour with intention to change the ILS as *manipulation*. The agent performing the manipulation is the *manipulating agent*. We consider changes in the user's gaze and proxemic behaviour following a manipulation to be the *user response*.

Each *manipulation* aims at affecting the ILS by either increasing or decreasing it. We consider three levels of intimacy: *Neutral*, *higher than neutral* and *lower than neutral*, which we simplify to *Neutral*, *High* and *Low*. As we have seen in our pilot study, we were able to produce agent behaviours that mediated intimacy at different levels. Based on these findings, the agent behaviours required to test our hypotheses are described in the following list. The behaviours marked as *High* and *Low* are the manipulations used by the agents. Note that the manipulations were deliberately chosen to be not just 'barely low' and 'barely high', but to depart significantly from their neutral counterpart. In short, during high gaze manipulations (G+) the agent will seek mutual gaze more, while during low gaze manipulations (G-), the agent will avert gaze more. During high proximity (P+) manipulations the agent will come closer, while during low proximity (P-) manipulations the agent will increase his distance. This is illustrated in Figure 4.1. In the following list, the behaviours and manipulations are specified further.

`Neutral Gaze` The agent switches gaze between the user and the other agent in random intervals of between 3 and 8 seconds, regardless of whether user gaze is detected or not. During gaze, in intervals between 2 and 5 seconds, the gaze is averted slightly by 10 degrees using the offset method, for 3.5 to 5 seconds.

`Low Gaze (G-)` The agent switches between gazing at the user and gazing at the other

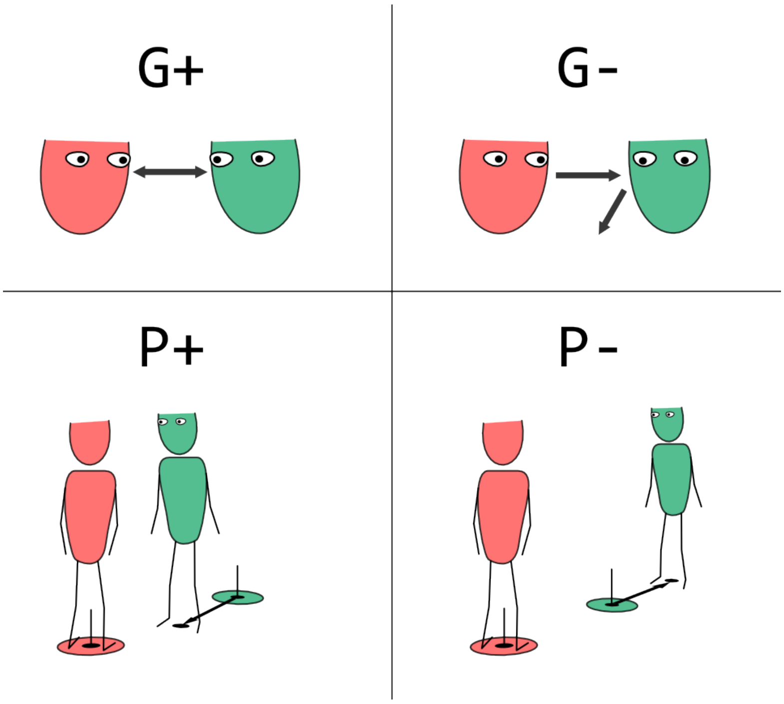


Figure 4.1.: Gaze and proximity manipulations of the agent (green) relative to the user (red).

agent in random intervals between 2 and 4 seconds. When mutual gaze is detected, the agent will avert its gaze to another target that is not the user (the other agent or the avert target). In intervals between 2 and 5 seconds, the gaze is averted slightly by 10 degrees using the offset method, for 3 to 6 seconds.

`High Gaze (G+)` The agent switches gaze infrequently between user and other agent. If the agent detects that the user gazes at him, he will always and immediately respond by gazing at the user - keeping the mutual gaze up as long as the user does, and then 1.5 seconds more. During mutual gaze, the agent will, in brief intervals, avert its gaze using the offset. Every 4 to 6 seconds, gaze will be briefly averted (1.5s to 3.5s) using the offset method.

`Neutral Proximity` The agent positions himself in such a way that the distance between the agent's and the user's face is around 75 cm in VR space.

`Low Proximity (P-)` The agent steps/leans away from the user, so that the distance between the agent's and the user's face is around 110 cm in VR space.

`High Proximity (P+)` The agent steps towards the user, so that the distance between the agent's and the user's face is around 40 cm in virtual space.

`Low Gaze & Proximity (G-P-)` The agent enacts both G- and P- at the same time.

`High Gaze & Proximity (G+P+)` The agent enacts both G+ and P+ at the same time.

4.2. User Response

We also consider gaze and proxemics in the users response, which we observe in the time during and after an agent manipulation. An illustration of the responses is given in ??.

Gaze Response The *Gaze Response* - or R_G - of a user is the change in angle towards the agent. This may be looking more towards the agent (smaller angle) or looking more away from it (larger angle).

Proxemic Response We call compensating displacement of the user's whole or upper body the *Proxemic Response* - or R_P - of the user. This may be moving away from the agent (positive response) or towards an agent (negative response).

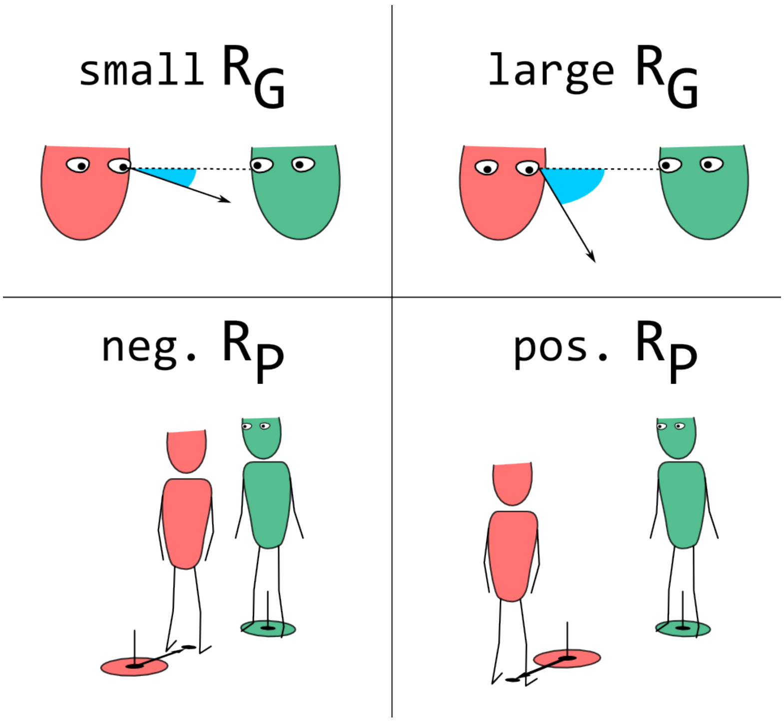


Figure 4.2.: Different values of gaze response R_G and proxemic response R_P of the user (red) relative to the agent (green).

More details on how R_G and R_P are computed so that we can use them in the data analysis of the experiment will be given in Section 6.1.4.

4.3. Conclusions

In this chapter, we have defined a framework that makes explicit the relationship between behaviours and their effects, what is manipulated and what is to be measured. We have defined the *manipulations* that the agents must be able to employ. Our hypotheses made predictions on the response of the user to the agents' manipulations. We can test our hypotheses by comparing the user response to the different manipulations. Before we discuss the experiment design in detail, we will first dedicate a chapter to the main material used in the experiment, the Immersive Virtual Environment Technology.

5. Immersive Virtual Environment

In this chapter technical and implementation details of the Immersive Virtual environment that was used to perform the experiment will be documented. We will first consider the software implementation in Section 5.1, including the chosen game engine, the virtual agents, the animation approach, some more details on the implementation of the required agent behaviours as defined in the framework and other agent capabilities. In Section 5.2, we present the scenario that was implemented in the virtual environment. This scenario was later used in the experiment to put the agent manipulations in context. Lastly, we will present the physical setup in Section 5.3. This includes the head mounted display, the tracking system, and the location.

5.1. Virtual Environment

5.1.1. Game Engine

To build the virtual environment, we used the Unity3D game engine and editor. Unity3D is currently a popular choice. We have used it in our previous work on mediated social touch (Huisman et al. [53, 54]), and it is used by other research platforms, such as the *Impulsion Engine* for simulating virtual crowds and small groups[1], and the *Virtual Human Toolkit*[2] uses it as well. It provides straightforward integration with the Oculus Rift, the head mounted display we used (see Section 5.3.2).

5.1.2. Virtual Agents

The virtual agents used in our IVET were generated using the Unity Multipurpose Avatar[3] system (UMA). UMA allows for dynamic creation and customisation of avatars. Each UMA avatar is created from a base mesh that can be deformed in several locations to change the shape of facial and body structures. It also comes with a pool of different attires that fit the base mesh and adapt to deformations of the body. The two agents generated for the experiment are shown in Figure 5.1. The advantage of this approach is that avatars created from the same base mesh can look very similar, yet discriminable if slight adjustments to face, hair and attire are made. In this experiment the outer

[1] `impulsionproject.tumblr.com`
[2] `vhtoolkit.ict.usc.edu`
[3] `github.com/huika/UMA`

Figure 5.1.: UMA Agents generated for the experiment, Mike (left) and Trevor (right)

appearance of the agents was not intended to be a variable. Therefor, by keeping looks similar, effects of appearance were kept minimal.

5.1.3. Animation

As already hinted at in the pilot study, we used procedural methods to animate the agents. The FinalIK[4] inverse kinematics plugin for Unity3D was used to have the agents gaze in a particular direction. FinalIK allows one to give different weights to rotation of chest, head and eyes towards a target, which we also experimented with during the pilot study. Walking and stepping animations are realised using Unity3Ds Mecanim animation blend-tree system[5]. A blend tree allows one to procedurally blend animations together. For example, it can be used to blend between forward and sideward step animations to generate a diagonal step. The *Impulsion Engine* mentioned before includes a complete blend tree that can produce walking animations in any direction using simple controls from a script. This was reused in our IVET.

5.1.4. Implemented Agent Behaviours

Several realizations of intimate agent behaviours were evaluated in the pilot study. Based on these findings, required agent behaviours for the experiment have been defined in Section 4.1 and were implemented in the IVET as described. The corresponding behaviour trees are shown in Appendix B. Screenshots of four manipulations are given in Figure 5.2.

[4] `root-motion.com/final-ik.html`
[5] `http://docs.unity3d.com/Manual/AnimationOverview.html`

(a) Right agent performs G+ manipulation.

(b) Left agent performs G- manipulation.

(c) Right agent performs P+ manipulation.

(d) Left agent performs P- manipulation.

Figure 5.2.: Screenshots of realized agent behaviours.

Note that some of the described gaze behaviours have an interactive element, as they respond to the user's gaze behaviour. To detect whether an agent is looked at by the user, we used a *ray-casting* implementation. An invisible ray or line is continuously projected from the head of the user in forward (i.e. looking) direction. If it intersects with a collider around the head of the agent, we consider the user to be looking at the agent. The collider is a capsule that is as wide as the agent's shoulders ($\tilde{4}$5 cm), and ranges from the agent's chest to just above his head ($\tilde{6}$0 cm).

5.1.5. Other Agent Capabilities

Agents use an idle loop of 30 seconds length that is offset by a random duration at the start of the application. This offset was added to prevent same-looking movements between agents that would make the idle loop more apparent. Further, lip-syncing was implemented. Facial blend shapes for several phonemes were created. For each audio clip that would be used by the agents, a phoneme detection was performed using CMU Sphinx[6]. Start and end time of detected phonemes in the audio were stored. When playing the audio clip, we blending between the facial blend shapes that correspond to

[6] `http://cmusphinx.sourceforge.net/`

Figure 5.3.: The virtual room. Note the transparent truss that was placed in correspondence with the truss in the physical room.

the detected phonemes in sync with the audio. Although the used phoneme-detection method yielded poor results in terms of accuracy, it was still sufficient for animation purposes, and significantly better than the amplitude based mandible animation method used in our previous work (Huisman et al. [53, 54]).

5.1.6. Virtual Location

The room used in the virtual environment is a generic large apartment asset[7] with a bigger empty space next to the living room area, which is mapped to the experiment space (see Figure 5.3). The t A transparent 3D model of the truss is placed in correspondence with its real-world position and dimensions to give users a reference in VR of where they are currently situated in the physical world.

Further, posters of persons and objects related to the scenario (see next section) were put in the room, which could also be hidden during the experiment.

5.2. Scenario

For the experiment design, which we will present in Section 6.1, we chose to use two agents that must maintain a conversation that is interesting for the participants to follow in the context of a listening task. The dialog should go back and forth between the agents with about equal pace. During each second turn, one of the agents performs the

[7] `https://www.assetstore.unity3d.com/en/#!/content/1899`

manipulation of his behaviour. Since manipulation takes a little time, and the participant response might also be delayed, dialog turns may not be too short (7 seconds was deemed to be the minimum). Further, to test each of the manipulations a number of times, the dialog must contain at least a certain number of turns (minimal 48).

A suitable source for this dialog was found with the 1957 movie *12 Angry Men* [8]. In this movie, 12 members of a jury have a discussion about whether or not they were presented sufficient evidence during the court case to sentence the defendant to death - a young man, standing accused of having killed his father. At first, only one member has doubts, but he manages to convince the others, one by one. This movie was chosen because it was dialog driven and takes place in the same room for its entire duration, with a dialog where most actors get turns regularly, and of similar duration. It is further suited for a listening task in that it presents a conflict where arguments are given for both sides, while leaving room for intuition and personal opinion. In total, 59 audio clips were extracted from this movie. Thirty clips with arguments from the 'against prosecution' side, and 29 from the 'for prosecution' side. On average, the clips lengths are 11.49 and 11.51 seconds, respectively. The clips were selected in chronological order and, when played in turn (against,for,against,for,...), make up a consistent conversation between the two groups. The entire conversation lasts just about 12 minutes. It should be noted that for each side, there are several different actors speaking, hence, when the scenario was enacted using two agents, their voices will change from time to time. Clips were intentionally selected from parts of the movie where the arguments were less heated, to prevent dominance mediated by voice to be a factor in the perception of the agents. More details on how the scenario is employed during the experiment will be given in Section 6.1.3.

5.3. Hardware & Location

5.3.1. Physical Location

The IVET is installed in a 4x5m experiment space in our lab. The area is roofed with a truss, as can be seen in Figure 5.4. On one of the long sides there are windows, and to the other long side it is open to the rest of the lab. On the two short sides there are walls. The area under the truss is empty.

5.3.2. Head Mounted Display

As VR display we use the Oculus Rift DK2. It has an OLED screen with 1920x1080 px resolution (960x1080 per eye) which can produce images at a rate of 75Hz. The diagonal field of view is 100 deg. For audio, we use a closed pair of stereo headphones, which shielded the user from outside noise. Together with the headphones, the HMD was tethered to a PC in the truss. This umbilical cord of approximately 2.6 meters in length

[8] `http://www.imdb.com/title/tt0050083/`

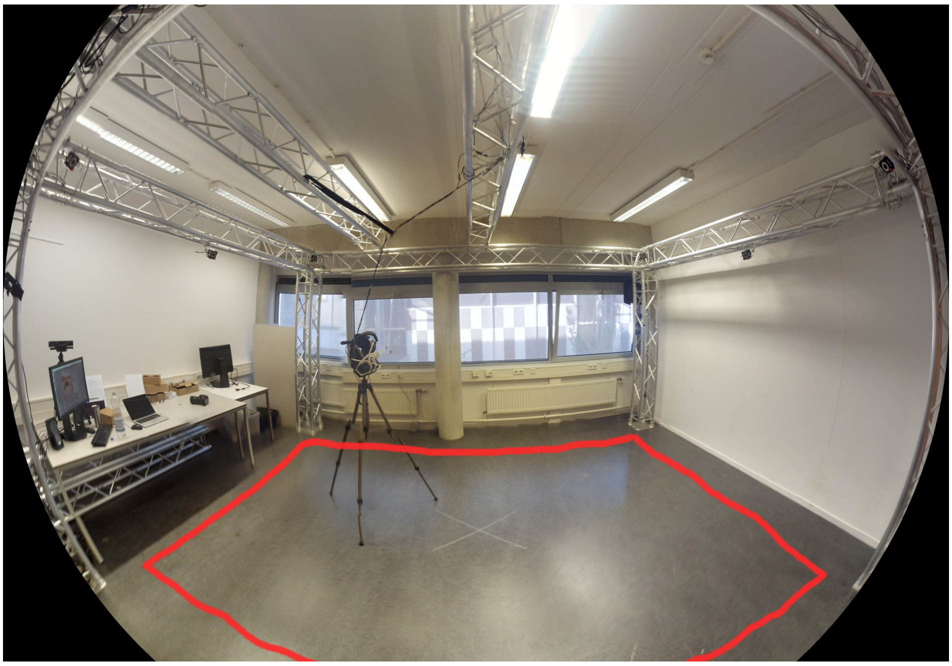


Figure 5.4.: The Physical Room, tracking area indicated with red outline.

was fixated at the centre-top of the experiment space, allowing the participants to walk freely up close to the edges of the room, although not entirely into the corners. This range also depended on the height of the participant. As can be seen in Figure 5.4, a rubber band was added to the umbilical cord to guide it behind the user's back.

5.3.3. Tracking

For positional tracking, we used the NaturalPoint OptiTrack IR-based tracking system. It provides position and rotation tracking of marker-equipped rigid bodies at up to 120 fps, with low latency and sub-millimeter accuracy. The six cameras were mounted on the truss frame at 2.3 m height, on the short sides of the experiment space. These cameras covered a sufficient area to reliably track the head of a single participant under the entire truss. In Figure 5.5, the Oculus Rift headset with attached retroreflective IR markers for the OptiTrack system is shown.

Using the data from this tracking system, we perform the behavioural measures used in the experiment, which will be described in detail in Section 6.1.4.

Figure 5.5.: The Head Mounted Display with retroreflective IR-markers on a plastic adapter, Headphones as worn by participants.

5.4. Conclusions

A virtual environment technology was created where the user was situated in a virtual space. A scenario was included, where two agents could act out a dialog with each other. They further could employ manipulations in the form of different gaze and proxemic behaviours.

6. Experiment

Based on our hypotheses, we have expectations on the interactions between virtual embodied agents and users that meet these agents in an immersive virtual environment. The agent behaviours were designed informed by the related work and the pilot study. To test our hypotheses, we wanted to compare the observed gaze and proxemic responses of the user to the manipulated virtual agent behaviours.

We chose a within subject design with two agents per participant. One agent was employing *high* manipulations, the other *low* manipulations (cf. Section 4.1). This choice was made so that we were able to ask participants how they perceived the two agent respectively, and in turn to see what qualities the different behaviours mediated, in the hope of being able to further disentangle the underlying mechanisms of the interaction.

6.1. Design

The two virtual agents positioned themselves to form a group with the user. The experiment design included one within subject variable: intimacy of agent, which was reflected both in gaze and proxemic behaviour. One agent had the *high* intimacy manipulations assigned, the other had *low* intimacy manipulations. They did not change their assigned role during the experiment. The agents would change their gaze and proxemic behaviour from neutral to a manipulation and switch back to neutral. When an agent performed a manipulation, he chose one of the three manipulations available to him: Either a single manipulation of gaze or proxemic behaviour, or a joint manipulation of both. Each manipulation by each of the agents was acted out four times, in randomised order.

The agents formed a group with the user by positioning themselves on the base corners of an equilateral triangle (see Figure 6.1(a)). The third corner was kept under the user's position, as determined by the head tracker, at the front side of the HUD (a bit in front of the user's nose). The triangle did not rotate with the user, but always faced the long side of the room. The length of the triangle's legs was 75 cm, which was the distance to be kept during the neutral proxemic behaviour. The angle of the user's corner is 60 deg, which was chosen such that when the user centres his view between the agents, both are in view.

Both agents had a slightly differing appearance to help discriminate them better during the questionnaire. To control for effects of the appearance, we randomised between subjects which agent had what role and what position (left or right of the participant).

Similarly, the roles in the dialog were counterbalanced to compensate for the effect the content of the used audio-clips that made up the scenario.

To conform to the voices in the scenario, both agents were chosen to be male. To prevent the size of the agents having intimidating (or belittling) effect, their height was adjusted in a calibration procedure to match the height of the participant.

Lastly, an unintentional between-subjects variable was introduced due to a logical mistake in the implementation of the procedure. On every second dialog turn, one of the agents manipulated their behaviour to their assigned role. On every other turn, both would employ the *neutral* behaviour. This means that within subjects, one of the agents changed his level of intimacy only when he is also the currently talking agent, whereas the other changed his level of intimacy only when he was not currently talking. This would have been prevented by interleaving two neutral episodes between each manipulation instead of only one. Whether it was the the high or the low agent that manipulated only during talking was still randomised between subjects. See Section 6.3 for more details on the implications of this oversight.

6.1.1. Materials

The only material used is the IVET as described in Chapter 5.

6.1.2. Participants

We convenience-sampled 35 participants from students and staff from the faculty of EEMCS (Electrical Engineering, Mathematics and Computer Science) at the University of Twente. They were between 19 and 30 years old ($m = 21.4$). Five were female. Of the 35 participants, one decided to stop the experiment early because of motion sickness, and another misunderstood the instructions, behaving in an unpredicted way. These two were discarded from the analysis.

6.1.3. Task and Deception

The responses we hoped to measure were the result of subconscious mechanics, rather than, for example, a conscious choice to satisfy expectations of what is 'correct' behaviour in the experiment. Therefore, participants were not told that the experiment was about examining their movement and gaze behaviour. Instead, we gave them a different task to focus on. The agents in the scenario (as described in Section 5.2) had opposing opinions about whether a defendant in a court case should be convicted or not.

It was suggested to the participant that the two agents would each attempt various 'strategies' (intentionally vague) in order to convince the participant of their side of the argument. The given task was to listen carefully and make up their own mind about what the right decision was. Lastly, it was announced that we would inquire in the questionnaire how the participant would decide and why.

6.1.4. Behavioral Measure

During the experiment we recorded the user's and agents' head positions and orientations in the virtual world using the Tracking system of the IVET. We continuously calculated the distance between the user's head and the respective agent's head as well as the angle of the user's gaze away from the respective agent.

Proxemic Response From these measurements, we calculated the proxemic response R_P of the participant as the difference between the distances of the user to the final agent position at the beginning and at the end of an episode, such that positive values indicated an increase in distance (stepping away), and negative values indicated a decrease in distance (stepping towards):

$$R_P = |P^A_{end} - P^U_{end}| - |P^A_{end} - P^U_{start}|$$

With P^A and P^U being the positions of the agent and user respectively, and the subscript indicating measurements at start or end of an episode. P^A_{end} is the position of the agent after the agent manipulation has been performed. The manipulation starts at the beginning of the current episode. The final position - if proximity is being manipulation (P-, P+) - is reached after about three seconds. This measure does not depend on agent movement. If the participant does not make an absolute displacement, the resulting R_P is zero. If proximity is not being manipulated by the agent, P^A_{end} equals P^A_{start}, so we can also measure proxemic responses during gaze-only manipulated episodes (G+,G-).

Gaze Response Eye contact was measured as the angle between two 3D vectors: The looking direction of the user and the vector between the user's head and the manipulating agent's head. The gaze response of the participant R_G is simply measured as the mean the vectors measured during the entire episode. Less eye contact should be reflected in a larger mean angle than more eye contact. Note that this is an approximation at best, since we do not know which part of the screen inside the head mounted display (HMD) the user's eyes are focussed on.

6.1.5. Questionnaire

While our hypotheses deal primarily with the behavioural responses of participants during the experiment, a post-experiment questionnaire was taken to support the measurements further by measuring the participant's perception of the individual agents. This questionnaire consisted of 14 items that have been successfully used before to measure perception of personality traits in both human and virtual human communication partners (see [55], with one extra item on politeness [56] and one for intimacy added by us). For each agent, we asked the participant's agreement with the questions given in Appendix D.2 on a 7-point Likert-scale. In the questionnaire, 'Agent' was replaced by the two male names, Trevor and Mike. Pictures of the agents were added to make identification possible. To measure the level of involvement and presence, we included 20 more items from Witmer and Singer's presence questionnaire [52], which are given in **??**.

6.2. Procedure

The participants were first given an oral introduction to the experiment. The technology and limitations were briefly discussed. The scenario was introduced and the participant's task explained. The participant was then asked to read and sign the consent form (see Appendix C), which also included the main points just discussed. After signing, the participant was reminded that he could decide to end the experiment at any moment, such as when he/she would feel discomfort in VR. The head-mounted display and headphones were mounted on the participant in the centre of the room. At this moment, the screen of the HMD was black. The participant was then rotated to face the front side of the room, and asked to hold still for a couple of seconds. The experimenter performed the calibration to align the HMD's internal and the external OptiTrack tracking system and to measure the height of the participant. The virtual environment then appeared on the participant's screen. It showed the virtual room with posters the of items and persons related to the court case in the scenario. To familiarise participants with the experience, we asked them to walk around and examine the posters, and explained that the experiment would start once the participant had 'explored the space enough'. If participants were hesitant to move, further friendly encouragement was given by the experimenter. The experimenter waited until the participant was situated in the front third of the room, and then started the experiment. The screen faded to black, the posters then disappeared, the two male agents appeared and the screen would fade back from black to show the scene again. The agents then approached the participant.

The agents positioned themselves in neutral position in the formation described above. Then, the dialog started. Each dialog turn formed an 'episode'. There were three types of episodes:

Neutral / Neutral Episode: Both agents keep the 'neutral' gaze and position.

When user moved, the agents adjust to the new neutral position (see Figure 6.1(b)).

Neutral / High Episode: Agents do not adjust when participant moves (see Figure 6.1(c)).

High agent changes proximity and/or gaze behaviour.

	Low agent stays neutral.
Neutral / Low Episode:	Agents do not adjust when participant moves (see Figure 6.1(d)).
	High agent stays neutral.
	Low agent changes proximity and gaze behaviour.

With each new dialog part, there was a new episode. The order of the episode-types was as follows:

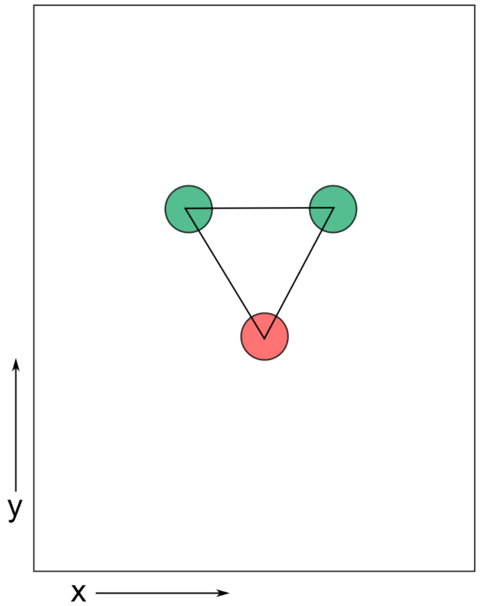
```
[Neutral/Neutral] -> [Neutral/{High/Low}] ->
  [Neutral/Neutral] -> [{High/Low}/Neutral] :/ repeat
```

To prevent measuring effects of 'surprise', an additional neutral/neutral episode was added at the beginning and end. To measure the participant's response to each of the six behaviours four times, with a neutral/neutral episode in between each other episode, we selected the first (6*4)*2 + 1 = 49 dialog clips created for the scenario. This ended up being just over ten minutes of agent dialog. The remaining dialog clips were not played to keep the experiment as short as possible. After this dialog was completed, the screen turned black again and the experimenter helped the participant out of the gear. The participant was lead to the questionnaire on a nearby computer.

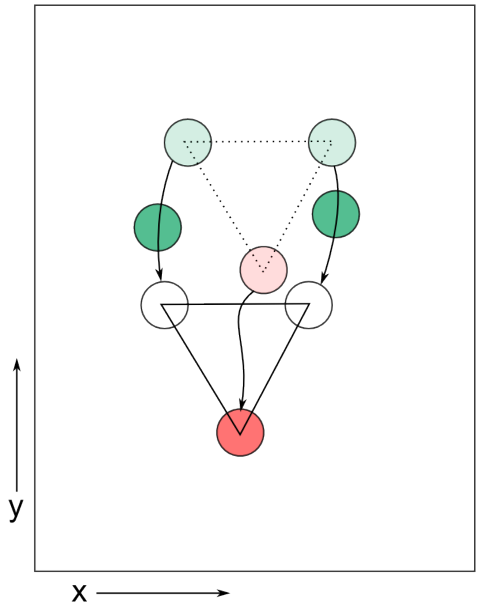
6.3. Data Analysis

The experiment was designed so that we could compare the effects of the six agent manipulations as a six level within-subject factor 'Agent Intimacy' on the two user measures R_G (gaze response) and R_P (proxemic response) using a repeated measures ANOVA. However, complications occurred since we - unintentionally - introduced a between subject variable that determined which agent's manipulations of behaviour also coincide with that agent's turn of speech. This means that within-subject, we could only compare participants' responses of a talking and non talking agent, when we compared effects of *high* manipulations (G+,P+,G+P+) with *low* manipulations (G-,P-,G-P-). Although we could argue that the act of 'talking' does not necessarily mediate 'intimacy', we must expect that the talking agent is the one that receives more attention, which will be manifested in a smaller angle towards that agent. While manipulations of proxemic behaviour will still be apparent to the participant even if he is focusing on the other agent, the more subtle changes in gaze behaviour are then less apparent, and as such can be expected to have less effect.

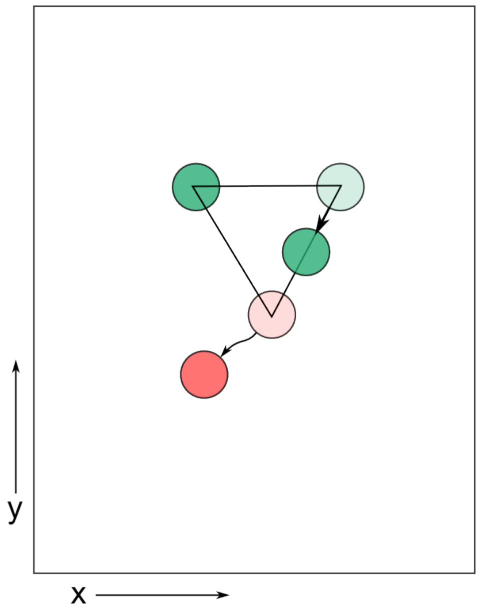
Consequently, since the turn of speech can be expected to have such a strong effect on the participant's own gaze direction, we would prefer to compare participants' gaze responses only inside the group of agents that manipulated their behaviour during their own turn of speech.



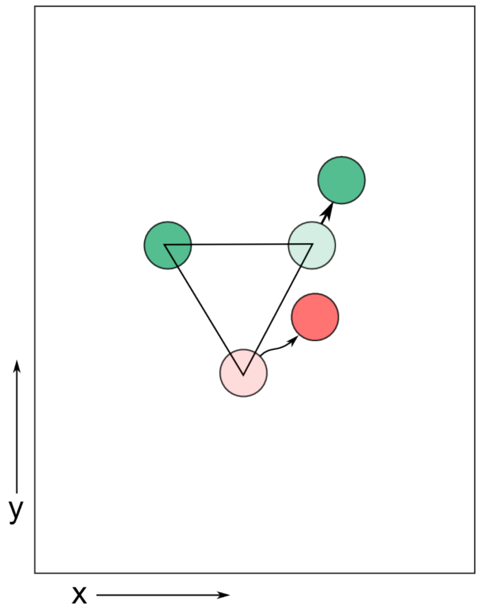
(a) Agents form a triadic group with the participant. Neutral formation.



(b) Neutral episode. If the participant moves, the triangle follows the participant and agents restore neutral formation.



(c) P+ manipulation. One agent comes closer towards the participant. The triangle does not follow the participant if he moves.



(d) P- manipulation. One agent increases distance towards the participant. The triangle does not follow the participant if he moves.

Figure 6.1.: The room from top-down perspective. Two agents (green) form a triadic formation with the user (red) (a), keeping the neutral formation when no manipulation is in place (b), but the formation-triangle does not follow the user during manipulations (c,d).

Our measurements violate the assumption of sphericity and normality for both measures at many levels (this seems unrelated to the previous problem). Brief attempts to reduce bias were unsuccessful. Under the assumption that the unintentional between subject variable does not represent a bias (for some of the measures), we instead used to the non-parametric Friedman test for statistical significance of the results.

6.4. Results

We will first present some overall results. An observation was made that most participants ended up at a different location of the experiment space at the end of the experiment compared to the start. It further seemed that sometimes they would end up on the left side, and sometimes on the right. The data supports this observation, and by highlighting on which side the agent with *high* manipulations was located, an explanation for the difference was found (Figure 6.2).

Looking at all interactions with high or low proximity (P+/P-), it also becomes apparent that in P+ episodes, participants stepped away from that agent (at the opposite angle of the agent's approach), while in P- episodes, participants stepped towards that agent - although the magnitude of R_P seems to be overall smaller here, which explains the general drift away from the agents seen in Figure 6.3.

Outliers We observe significant outliers in the proxemic responses. Upon inspection by reviewing video material and experiment notes, these outliers were caused by participants intentionally stepping around the agents to reach a position in the virtual space away from the agents. These are displacements of more than just one or two steps, but rather walking across the room. Although these changes in position seem motivated by the intimate situation, they diverge significantly from the typical proxemic response in other episodes, where participants would either lean or take one or two small steps. It was decided to identify all episodes where R_P was bigger than 50 cm. This way, out of the 800 episodes that are manipulations of any kind, 6 episodes were designated as outliers (see the 'outliers' column in Table 6.1). In the following analysis of the results, these outliers are not included anymore.

In the histogram in Figure 6.4, we can see that both for high and for low proxemic episodes, R_P peaks around a magnitude of zero (no displacement). However, for high proxemic episodes, a second peak occurs around +18 cm. In the responses to low proxemic manipulations, we see that the peak runs out asymmetrically, with higher frequencies on the negative displacement (towards the agent) side. We get a first impression that, indeed, high proximity results in a more positive proxemic response (away from the agent) than low proximity.

Figure 6.5 shows R_G in response to talking agents and non-talking agents. The distribution of R_G peaks at smaller angles (more eye contact) for the talking agent, confirming our

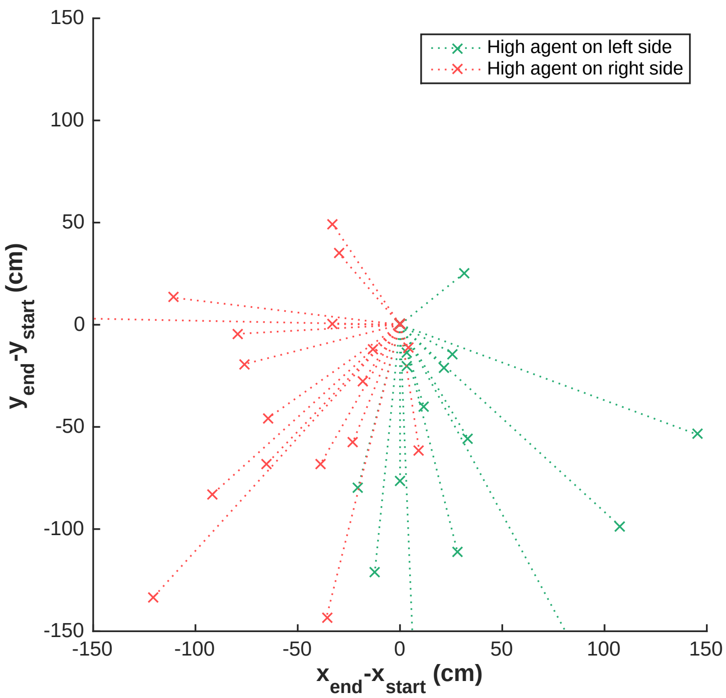


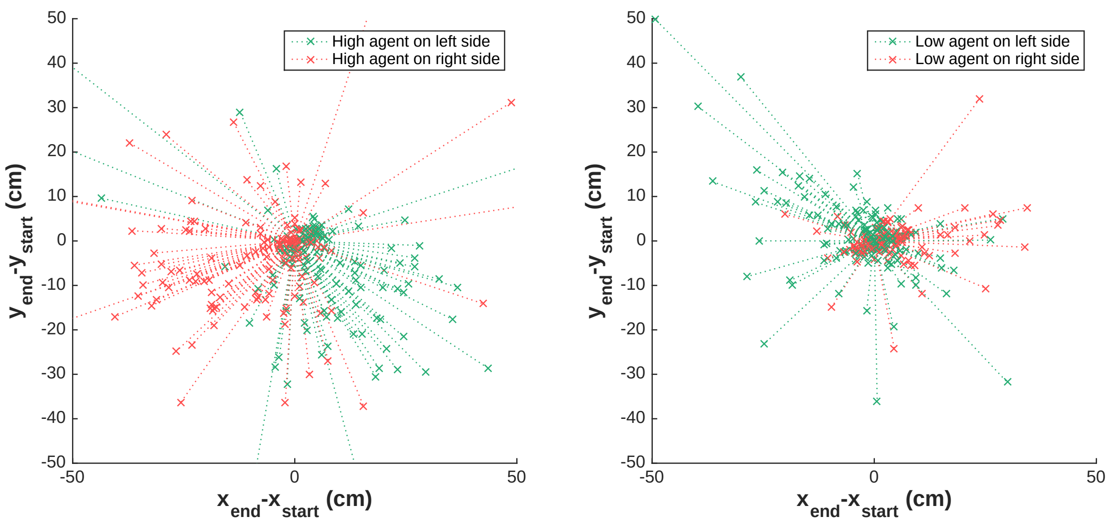
Figure 6.2.: Per participant, difference between position at start and end of experiment. Side of manipulating agent indicated by color

expectation that participants would look more at the agent that is talking. The implication would be that the subtle effect of changed gaze behaviour would only become apparent when the manipulating agent was also talking. We can see that this seems to be the case comparing Figures 6.6(a) and 6.6(b). There was a steady decrease in R_G when going from more to less intimate manipulations in Figure 6.6(a). The gaze response to manipulations of agents that were not talking however does not show such a trend (see Figure 6.6(b)) - rather, all means seem very similar. Comparing R_P in the same way (Figures 6.6(c) and 6.6(d)) does show very similar proxemic responses regardless of whether the manipulating agent was talking or not.

6.4.1. Tendencies

Here we examined whether the general tendencies were in line with the hypotheses. For this, we looked at the means of all participants and episodes where the manipulating agent was also the talking agent. An overview of the discussed results is shown in Table 6.1.

The first hypothesis states that displacement of the participant was more positive during episodes where the agents proxemic behaviour was high (P+), compared to the more negative displacement during low proxemic behaviour (P-). We selected all P+ and P- episodes where the manipulating agent was also the talking agent. The mean displacement



(a) Per episode, difference between participant position at start and end of P+(G+) episodes. Side of manipulating agent indicated by color.

(b) Per episode, difference between participant position at start and end of P-(G-) episodes. Side of manipulating agent indicated by color.

Figure 6.3.: Participants move away from manipulating agent in P+ and P+G+ episodes, and towards the manipulating agent in P- and P-G- episodes

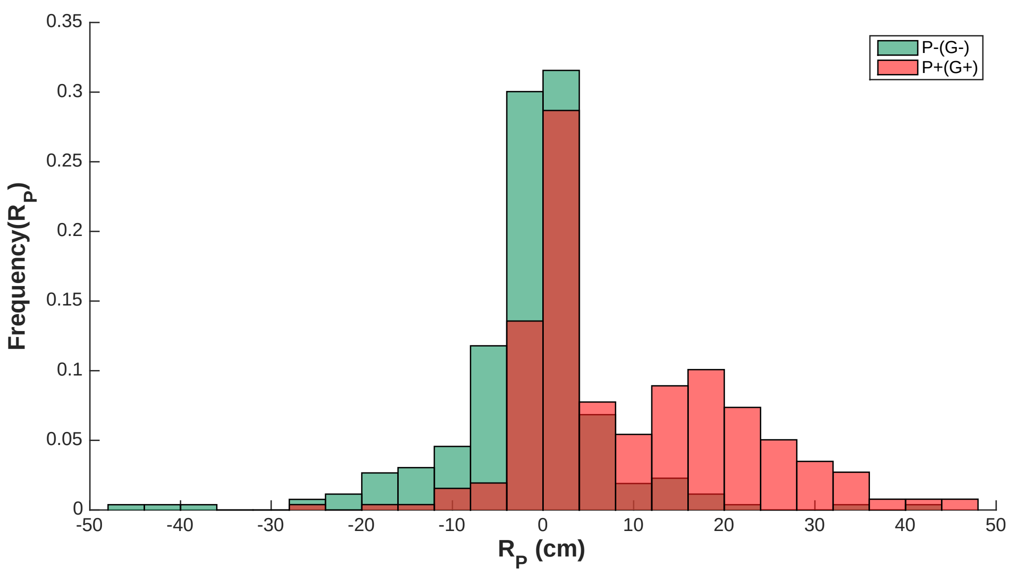


Figure 6.4.: Histograms of R_P for P+/P+G+ and P-/P-G- episodes.

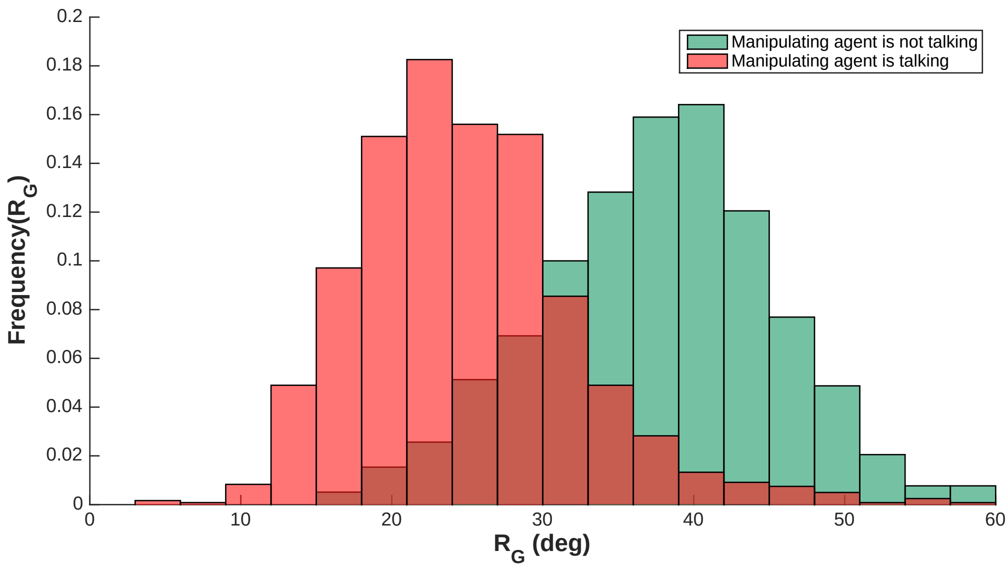
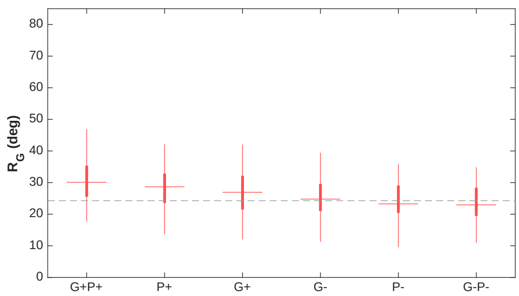


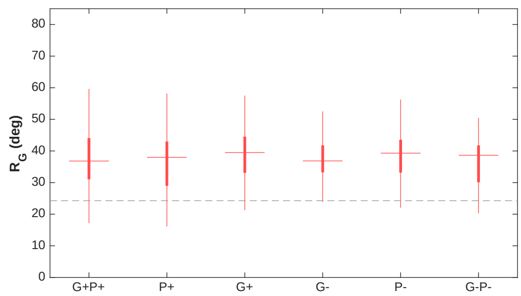
Figure 6.5.: Histograms of R_G for episodes where the manipulating agent *is* the talking agent and *is not* the talking agent

Manipulation	Mean R_G in $^\circ$	Mean R_P in cm	n	outliers
G+P+	30.17 ($SD = 6.41$)	8.56 ($SD = 11.70$)	55	1
P+	28.57 ($SD = 7.38$)	8.43 ($SD = 13.89$)	53	2
G+	27.01 ($SD = 7.73$)	0.36 ($SD = 9.50$)	56	0
G-	25.23 ($SD = 6.16$)	-0.37 ($SD = 5.79$)	75	1
P-	25.16 ($SD = 7.56$)	-2.97 ($SD = 8.89$)	76	0
G-P-	23.52 ($SD = 5.72$)	-3.48 ($SD = 6.51$)	74	2

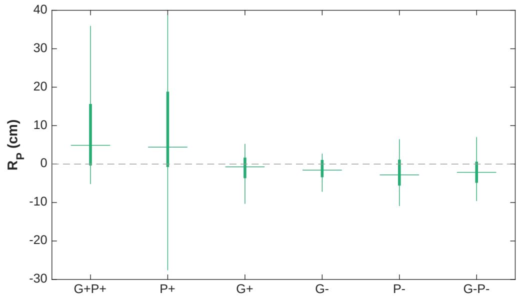
Table 6.1.: Mean gaze response R_G and proxemic response R_P per agent manipulation from all episodes where the manipulating agent was also the talking agent. The number of outliers that were not considerd is reported in the 'outliers' column.



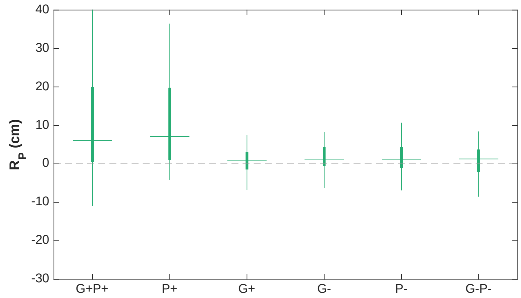
(a) Distribution of R_G per manipulation when manipulating agent was also the talking agent



(b) Distribution of R_G per manipulation when manipulating agent was not the talking agent

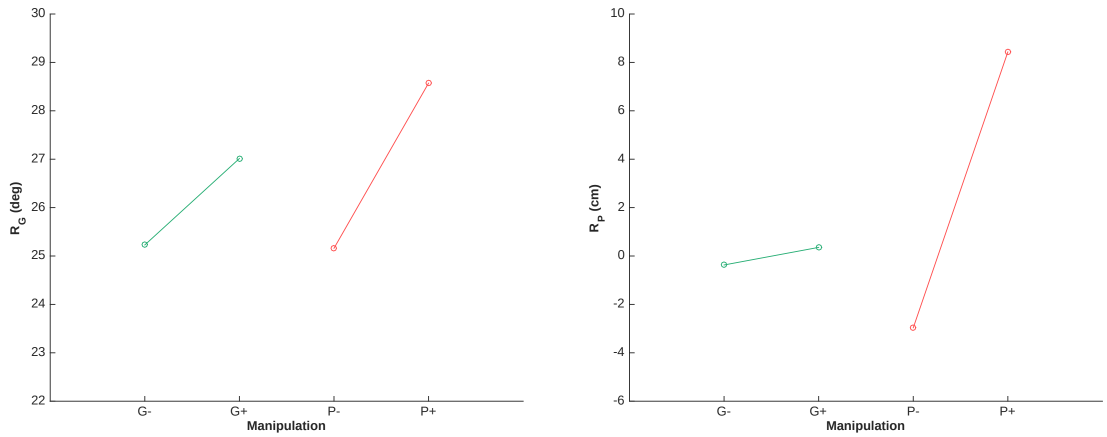


(c) Distribution of R_P per manipulation when manipulating agent was also the talking agent



(d) Distribution of R_P per manipulation when manipulating agent was not the talking agent

Figure 6.6.: Distributions of users' gaze responses R_G (a and b) and proxemic responses R_P (c and d) per manipulation, split by whether the manipulating agent was also the talking agent (a and c) or not (b and d).



(a) Mean effects of high and low gaze and proxemic manipulations on gaze response R_G.

(b) Mean effects of high and low gaze and proxemic manipulations on proxemic response R_P.

Figure 6.7.: Participant responses R_G (a) and R_P (b) to only gaze and only proximity manipulations, highlighting the differences between the respective high and low variants of each manipulations.

was $m = 8.43$ cm ($SD = 13.89$) during P+ episodes and $m = -2.97$ cm ($SD = 8.89$) during P- episodes. This tendency supports the first hypothesis.

The second hypothesis states that gaze angles of the participant towards an agent with highly intimate gaze behaviour (G+) are greater than towards an agent with low intimate gaze behaviour (G-). We selected G+ and G- where the manipulating agent was also the talking agent. The mean gaze angle was $m = 27.01$ deg ($SD = 7.73$ deg) during G+ episodes, and $m = 25.23$ deg ($SD = 6.16$) during G- episodes. This tendency supports the second hypothesis.

The third hypothesis states that high proximity does also have an effect on participant's gaze. We expected the participant's gaze angle towards the agent in P+ episodes to be be greater than in P- episodes. We selected all P+ and P- episodes where the manipulating agent was also the talking agent. The mean gaze angle was $m = 27.01$ deg ($SD = 7.73$) during high proximity episodes, and $m = 25.16$ deg ($SD = 7.56$) during low proximity episodes. This tendency supports the third hypothesis. It further appears that the agents' proxemic behaviour has a greater effect on participant gaze response than agents' gaze behaviour.

The fourth hypothesis states that manipulations of agent gaze also have an effect on participant's proxemic response. We expected that the participant's displacement in G+ episodes to be more positive (away from the agent) than in G- episodes. We selected all G+ and G- episodes where the manipulating agent was also the talking agent. The mean proxemic response was $m = 0.36$ cm ($SD = 9.5$) in G+ episodes and $m = -0.37$ cm ($SD = 5.79$ m) in G- episodes. This difference is too marginal to support the fourth hypothesis in any way. Figure 6.7 illustrates this well, here we can see that effects of manipulation of gaze are observed only in gaze response of the participant, and not in proxemic response.

The fifth hypothesis states that we expect effects to add up when exhibiting both high gaze and high proxemic behaviours at the same time (or low/low). Looking at Table 6.1, we see indeed that effects on both the participant gaze and proxemic response seem to have been be stronger in those episodes where an agent combined both high or low gaze and proxemic behaviours.

6.4.2. Satistical Analysis

As noted previously, the data violates several assumptions for using the repeated measures ANOVA. The Friedman test is the non-parametric alternative to the repeated measures ANOVA. This test was used to determine whether agent intimacy had significant effects on displacement magnitude and gaze angle towards the agents measured in the participants. If significance was found in the Friedman test, to examine where the differences actually occur, we ran separate Wilcoxon signed-rank tests on the different combinations of related pairs.

We performed three tests for each of the two measures R_G and R_P. First we assumed that whether the agent was talking or not had no effect on any measure, and therefore allowed comparison between low and high behaviours. Then we performed two more tests. In one we looked at effect differences between the three high behaviours in those cases where the talking agent was the agent with the *high* manipulations assigned ($n = 14$). In the other we looked at effect differences between the three low behaviours in those cases where the talking agent was the agent with the *low* manipulations assigned ($n = 19$).

We chose to compare the the third measurement for each agent behaviour. Upon inspection of all outliers, it appeared that not a single outlier (by the criteria previously described) happened during a third measure of any manipulation of any participant, making it a convenient choice. The other measures were not considered for statistical analysis.

Differences between all six manipulations The Friedman test revealed that there was a statistically significant[1] difference in displacement magnitude as a response to different levels of agent behaviour intimacy, $\chi^2(5) = 32.84$, $p < .001$. The Wilcoxon signed-rank test showed that in the 33 participants, the displacement magnitude in response to high proximity behaviours was significantly more positive (i.e.: moving away) than that to low proximity behaviours ($Z = -3.368$, $p = .001$).

[1]The Bonferroni-corrected significance level is 0.008, since we conpare the six relevant pairs to test the hypotheses: G-/G+, P-/P+, P+/G+P+, P-/G-P-, G+/G+P+ and G-/G-P-.

No significant difference in gaze angle was revealed $\chi^2(5) = 8.83$, $p = .259$, hence no further tests comparing the effect on participant gaze were performed.

Differences between high manipulations (G+, P+, G+P+) The Friedman test revealed that there was a statistically significant difference in response displacement magnitude between the high-intimacy behaviours, $\chi^2(2) = 7.00$, $p = .030$. The Wilcoxon signed-rank test showed that in the 14 participants where the *high* agent did manipulate his behaviours while also being the talking agent, the displacement magnitude in response to G+P+ episodes was significantly greater than the displacement magnitude in response to G+ ($Z = -2.542$, $p = .011$). In the same population, between the pair of G+ and P+ manipulations, we found that the former would elicit significantly[2] less positive (i.e.: moving away less) displacement magnitude ($Z = -2.229$, $p = .026$) than the latter. The difference between the pair of G+P+ and P+ behaviour was not found to be significant ($Z = -.910$, $p = .363$).

Again, no significant difference in gaze angle was revealed ($\chi^2(2) = 2.29$, $p = 0.319$), hence no further tests comparing the effect on participant gaze were performed.

Differences between low manipulations (G-, P-, G-P-) Between the low intimacy behaviours, the Friedman test did not reveal a significant[3] difference in displacement magnitude response of the 19 participants where the *low* agent did manipulate his behaviours while also being the talking agent ($\chi^2(2) = 2.95$, $p = 0.229$). No further tests comparing the individual pairs were performed.

The Friedman test, however, did reveal that there was a marginally significant difference in the participant gaze response between the *low* behaviours, $\chi^2(2) = 6.42$, $p = .040$. Upon inspection, it appears that difference is due to the asymmetry of the difference of the pairs, excluding it from further examination with the Wilcoxon signed-rank test. A sign test revealed no significant difference.

6.4.3. Presence Questionnaire

We computed the involvement and presence score following Witmer and Singer [52]. We found that of the 32 participants, 29 reported an involvement score of 4.0 or higher ($m = 5.22$, $SD = .99$). All 32 reported a presence score of 4 or higher ($m = 5.26$, $SD = .56$).

[2] Here, the Bonferroni-corrected significance level is 0.017, since we make comparisons only for the three pairs of low behaviours: G+/P+, G+/G+P+ and P+/G+P+.

[3] Here, the Bonferroni-corrected significance level is 0.017, since we make comparisons only for the three pairs of low behaviours: G-/P-, G-/G-P- and P-/G-P-.

Factor	Item	Factor loading
Warmth ($\alpha = .92$)	Friendly	.88
	Approachable	.83
	Warm	.83
	Likeable	.82
	Polite	.79
	Modest	.79
Trustworthiness ($\alpha = .87$)	Informed	.82
	Credible	.82
	Competent	.76
	Honest	.71
	Trustworthy	.58
	Sincere	.56
Intimacy ($\alpha = .57$)	Intimate	.78
	Interesting	.68
	Confident	.66

Table 6.2.: Three factors identified in PCA and their corresponding items with factor loadings. For each factor, consistency is reported.

6.4.4. Agent Personality Questionnaire

We performed a principal component analysis with Varimax rotation and Kaiser normalisation on all 35 responses to the 15 questionnaire items. Three factors were identified that together explain 69.15% of the variance. The factors and their loadings are shown in Table 6.2. Two of the three factors are in line with factors from earlier experiments using a similar set of items. We re-use the naming and call them 'Warmth' and 'Trustworthiness'. The items 'polite' and 'modest', which in previous work made up the 'Politeness' factor, shifted to the 'Warmth' factor in the current analysis. Instead, a new third factor emerged with the items 'intimate' (new item), 'interesting' (previously in 'Trustworthiness') and 'confident' (previously in 'Warmth'). We name this new factor 'Intimacy'.

For each respondent, we calculated factor scores given to the two agents by averaging out those items that were associated with the respective factors. We performed repeated measures ANOVA with the intimacy of the agent (*high* or *low*) as the within subjects variable and agent side, the talking agent, and agent appearance as between subject variables, and the three computed factor scores as measures.

We found a main effect for the intimacy behaviour of the agents on 'Warmth' ($F(1, 24) = 21.45$, $p < .01$) and 'Intimacy' ($F(1, 24) = 6.61$, $p < .05$). No interaction effects of agent appearance and agent side were found on either of the scores. There was however an interaction effect for the talking agent on 'Intimacy' scores ($F(1, 24) = 4.31$, $p < .05$).

Pairwise comparison revealed that participants scored the agent with low intimacy higher on ‘Warmth’ related items than the high intimate agent ($m_L^W = 4.97$ vs $m_H^W = 3.57$). Scores for ‘Trustworthiness’ follow the same trend ($m_L^T = 5.23$ vs $m_H^T = 4.88$, which was not significant.). ‘Intimacy’ scores align with the intimacy behaviour of the agents. Participants scored the agent with low intimacy lower ($m_I^L = 4.14$) than the agent with high intimacy ($m_I^H = 4.90$).

For the interaction effect of the talking agent, pairwise comparison revealed that the high and low agents score similarly on intimacy scores when they are not the talking agent. When talking during manipulation the high agent however scores significantly higher on intimacy ($m_I^{H\times T} = 5.25$) scores than the low agent ($m_I^{L\times T} = 3.86$).

7. Discussion & Conclusion

The goal of this work was to further disentangle the single and joint effects of gaze and proxemic behaviours in immersive virtual reality. In experiment we compared gaze and proxemic responses of participants to virtual agents that manipulate their own gaze and proxemic behaviour during a conversation.

The overall findings from this experiment are in line with the initial studies of Argyle and Dean [1]. The new contributions of our findings are the specifics of relationship of the behavioural responses different single and joint manipulations when individuals are not intentionally restricted in one of their behaviours, such as being forced to sit or stand still during the experiment.

Significant statistical evidence was found to support H1. We found that agents exhibiting higher proximity did cause participants to step away more than agents exhibiting low proximity, where participants tended to step more towards the retreating agent. Although this is the most straightforward hypotheses, it is also one that had not previously been tested experimentally in immersive virtual reality.

As for the predicted effects of manipulating gaze on gaze (H2), we did not find significant differences. While the tendencies are in line with the hypothesis, the approximation of gaze with head orientation might not be sufficient to reveal this effect appropriately. The data further suggests that joint effects of manipulating the intimacy of gaze and proxemic behaviours are stronger both on gaze and proxemic responses in the recipient of the manipulation (H5). Not all singular manipulations however appear to also have effects on both behavioural responses.

There was no notable effect of gaze manipulations on the proxemic response (H4). This is surprising, given the earlier results of Bailenson et al. [6]. It may be explained by their use of a more sensitive measure (minimum distance rather than the mean) and the different interaction between agent and participant (walking around rather than listening). For high proximity manipulations, we did observe that participants performed compensation of both their gaze behaviour (H3), which was not examined nor predicted before. A possible explanation could be ceiling effects of how comfortable individuals were with moving in the IVET - possibly also depending on whether they were already at the edge of the tracking area. But social norms could also introduce ceiling effects in such interactions. For example, it may not be appropriate to make huge displacements when someone comes closer, to not signal fear. If that smaller displacement was not sufficient to compensate intimacy, we would expect the remainder to be compensated with gaze. This interpretation is also in line with the personality scores of the high agent.

Scores were low on ‘Warmth’, which had loadings of the ‘politeness’ and ‘friendliness’ items, but high on ‘Intimacy’, which had loadings of ‘intimate’ and ‘confident’ items.

Our hypotheses were based on the equilibrium theory. Their explanation for the relationship between the regulating small scale behaviours such as gaze and proximity is the perceived intimacy. This informed the design of our behaviours, which generally performed as expected. This gives further support to the theory that intimacy is indeed induced and compensated by proxemic and gaze behaviours during social interaction.

Future Work & Recomendations The virtual reality approach was successful in testing our hypotheses. In the future, we would like to resolve the relationship between proxemics and gaze behaviours in even more detail, possibly with the existing data. Did those participants that compensate more with proximity compensate less with gaze and vice versa? Are there better, more dimensional measures that describe proxemic compensation, including direction and rotation?

The current measurement of gaze based on head-orientation of the participant is an approximation. Future experiments on the matter should consider using an eye tracker inside the HMD instead. Measuring upper body orientation in addition to head rotation might be worthwhile, as earlier researchers suggested that body rotation may be a part of gaze behaviour.

Researching the interaction between such behaviours in VR might benefit from a more iterative approach, where responses to behaviour changes are recorded - possibly even using motion capture - and can be used by the agents to respond to similar behaviour changes observed in new users.

As mentioned in the review, other studies found that during interaction, reciprocal responses could be found as well, for example where confederates touched the participants. Touch and other modalities are certainly interesting to incorporate into experiments in IVETs. Especially the conversational aspects of social interaction should be included as a modality. One of our findings was that intimacy as a personality trait was mediated stronger if the agent manipulated his behaviour while talking - both for high and low intimacy behaviours. The implications of this are not clear to us at this point, and deserve further attention in the future.

To conclude, we consider this work to be a successful first step in examining small scale social behaviours using immersive virtual reality technology. We made findings that support earlier work, got indications that previous related findings hold in IVETs, and our findings give additional insight that may have been difficult to obtain with the same amount of work in conventional experimental settings. We want to motivate more researchers in related fields to consider performing experiments in immersive virtual realty. Especially with immersive virtual reality hardware entering the consumer market, private and public VR Labs will also become more prevalent, making virtual reality research on the fields of telepresence, multimodal interaction, social signal processing and social robotics more accessible, and providing a new platform for novel virtual reality applications.

Bibliography

[1] M Argyle and J Dean. Eye-Contact, Distance and Affiliation. *Sociometry*, 28(3): 289–304, 1965. ISSN 00380431. doi: 10.2307/2786027.

[2] Larry M. Coutts and Frank W. Schneider. Affiliative conflict theory: An investigation of the intimacy equilibrium and compensation hypothesis. *Journal of Personality and Social Psychology*, 34(6):1135–1142, 1976. ISSN 0022-3514. doi: 10.1037/0022-3514.34.6.1135.

[3] Miles L. Patterson. Interpersonal distance, affect, and equilibrium theory. *The Journal of Social Psychology*, 101(2):205–214, 1977. ISSN 0022-4545. doi: 10.1080/00224545.1977.9924008.

[4] J N Cappella. Mutual influence in expressive behavior: adult–adult and infant–adult dyadic interaction. *Psychological bulletin*, 89(I):101–132, 1981. ISSN 0033-2909. doi: 10.1037/0033-2909.89.1.101.

[5] Howard M. Rosenfeld, Barbara E. Breck, Stephanie H. Smith, and Sara Kehoe. Intimacy-mediators of the proximity-gaze compensation effect: Movement, conversational role, acquaintance, and gender. *Journal of Nonverbal Behavior*, 8(4):235–249, 1984. ISSN 01915886. doi: 10.1007/BF00985981.

[6] Jeremy N. Bailenson, Jim Blascovich, Andrew C. Beall, and Jack M. Loomis. Equilibrium Theory Revisited: Mutual Gaze and Personal Space in Virtual Environments, 2001. ISSN 1054-7460.

[7] Jeremy N Bailenson, Jim Blascovich, Andrew C Beall, and Jack M Loomis. Interpersonal Distance in Immersive Virtual Environments. *Personality and Social Psychology Bulletin*, 2003. doi: 10.1177/0146167203253270.

[8] Matthias J Wieser, Paul Pauli, Miriam Grosseibl, Ina Molzow, and Andreas Mühlberger. Virtual social interactions in social anxiety–the impact of sex, gaze, and interpersonal distance. *Cyberpsychology, behavior and social networking*, 13(5): 547–554, 2010. ISSN 2152-2715. doi: 10.1089/cyber.2009.0432.

[9] Dario Bombari, M Schmid_mast, Elena Cañadas, and Manuel Bachmann. Studying social interactions through immersive virtual environment technology: Virtues, pitfalls, and future challenges. *Name: Frontiers in Psychology*, 6:869, 2015. ISSN 1664-1078.

[10] N. J. Emery. The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews*, 24:581–604, 2000. ISSN 01497634. doi: 10.1016/S0149-7634(00)00025-7.

[11] Alexandra Frischen, Andrew P Bayliss, and Steven P Tipper. Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological bulletin*, 133(4):694–724, 2007. ISSN 0033-2909. doi: 10.1037/0033-2909.133.4.694.

[12] Adam Kendon. Some functions of gaze-direction in social interaction. *Acta psychologica*, 26:22–63, 1967. ISSN 00016918. doi: 10.1016/0001-6918(67)90005-4.

[13] Mark Cook. Gaze and Mutual Gaze in Social Encounters: How longand whenwe look others" in the eye" is one of the main signals in nonverbal communication. *American Scientist*, 65:328–333, 1977. ISSN 0003-0996. doi: 10.2307/27847843.

[14] J Duncan and J Duncan. Selective attention and the organization of visual information. *Journal of experimental psychology. General*, 113(4):501–17, 1984. ISSN 0096-3445. doi: 10.1037/0096-3445.113.4.501. URL `http://www.ncbi.nlm.nih.gov/pubmed/6240521`.

[15] Evelyn Z. McClave. Linguistic functions of head movements in the context of speech, 2000. ISSN 03782166.

[16] Dirk Heylen. Head gestures, gaze and the principles of conversational structure. *International Journal of Humanoid Robotics*, 3(03):241–267, 2006. ISSN 0219-8436.

[17] Bilge Mutlu. Designing gaze behavior for humanlike robots. *ProQuest Dissertations and Theses*, 3367045(May):250, 2009. doi: TR-CMU-HCII-09-101. URL `http://ezproxy.net.ucf.edu/login?url=http://search.proquest.com/docview/304865504?accountid=10003$\delimiter"026E30F$nhttp://sfx.fcla.edu/ucf?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+&+theses&sid=ProQ:ProQuest+Dissertations+&+The`.

[18] Mario Von Cranach and Johann H Ellgring. Problems in the Recognition of Gaze Direction. *Social Communication and Movement: Studies of Interaction and Expression in Man and Chimpanzee.*, pages 419–443, 1969.

[19] Ulrich J. Pfeiffer, Leonhard Schilbach, Mathis Jording, Bert Timmermans, Gary Bente, and Kai Vogeley. Eyes on the mind: Investigating the influence of gaze dynamics on the perception of others in real-time social interaction. *Frontiers in Psychology*, 3(December):1–11, 2012. ISSN 16641078. doi: 10.3389/fpsyg.2012.00537.

[20] K Ruhland, S Andrist, and Jb Badler. Look me in the eyes: A survey of eye and gaze animation for virtual agents and artificial systems. *. . . 2014-State of the Art . . .*, 2014. URL `http://diglib.eg.org/EG/DL/conf/EG2014/stars/069-091.pdf.abstract.pdf;internal&action=action.digitallibrary.ShowPaperAbstract`.

[21] Sonu Chopra Khullar and Norman I. Badler. Where to Look? Automating Attending Behaviors of Virtual Human Characters. *Autonomous Agents and Multi-Agent Systems*, 4(1-2):9–23, 2001. ISSN 13872532. doi: 10.1023/A:1010010528443.

[22] C. Peters and C. O'Sullivan. Bottom-up visual attention for virtual human animation. *Proceedings 11th IEEE International Workshop on Program Comprehension*, 2003. ISSN 1087-4844. doi: 10.1109/CASA.2003.1199311.

[23] Laurent Itti. Realistic avatar eye and head animation using a neurobiological model of visual attention. *Proceedings of SPIE*, 5200(1):64–78, 2004. ISSN 0277786X. doi: http://dx.doi.org/10.1117/12.512618. URL `http://link.aip.org/link/?PSI/5200/64/1&Agg=doi`.

[24] Antoine Picot, Gérard Bailly, Frédéric Elisei, and Stephan Raidt. Scrutinizing Natural Scenes: Controlling the Gaze of an Embodied Conversational Agent. *Intelligent Virtual Agents*, 4722:272–282, 2007. ISSN 03029743. doi: 10.1007/978-3-540-74997-4_25. URL `citeulike-article-id:3667085$\delimiter"026E30F$nhttp://dx.doi.org/10.1007/978-3-540-74997-4_25`.

[25] Brent Lance and Stacy C. Marsella. Emotionally Expressive Head and Body Movements During Gaze Shifts. *7th International Conference on Intelligent Virtual Agents*, 4722:72–85, 2007. doi: 10.1007/978-3-540-74997-4_8. URL `http://www.ict.usc.edu/~marsella/publications/LanceIVA07.pdf`.

[26] Marcus Thiebaux, Brent Lance, and Stacy Marsella. Real-time expressive gaze animation for virtual humans. *Proceedings of The 8th International ...*, pages 321–328, 2009. URL `http://dl.acm.org/citation.cfm?id=1558057`.

[27] Cagla Cig, Zerrin Kasap, Arjan Egges, and Nadia Magnenat-Thalmann. Realistic emotional gaze and head behavior generation based on arousal and dominance factors. In *Motion in Games*, pages 278–289. Springer, 2010. ISBN 3642169570.

[28] Zheng Li and Xia Mao. Emotional eye movement generation based on geneva emotion wheel for virtual agents. *Journal of Visual Languages and Computing*, 23 (5):299–310, 2012. ISSN 1045926X. doi: 10.1016/j.jvlc.2012.06.001. URL `http://dx.doi.org/10.1016/j.jvlc.2012.06.001`.

[29] Goranka Zoric, Rober Forchheimer, and Igor S. Pandzic. On creating multimodal virtual humans-real time speech driven facial gesturing. *Multimedia Tools and Applications*, 54(1):165–179, 2011. ISSN 13807501. doi: 10.1007/s11042-010-0526-y.

[30] Binh H Le, Xiaohan Ma, Zhigang Deng, and Senior Member. Head-and-Eye Motion Generators. X(X):1–14, 2012.

[31] Stacy Marsella, Yuyu Xu, Margaux Lhommet, Andrew Feng, Stefan Scherer, and Ari Shapiro. Virtual character performance from speech. *Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation - SCA '13*,

page 25, 2013. doi: 10.1145/2485895.2485900. URL http://dl.acm.org/citation.cfm?doid=2485895.2485900.

[32] Stephanos Ioannou, Paul Morris, Hayley Mercer, Marc Baker, Vittorio Gallese, and Vasudevi Reddy. Proximity and gaze influences facial temperature: a thermal infrared imaging study. *Frontiers in Psychology*, 5(August):1–12, 2014. ISSN 1664-1078. doi: 10.3389/fpsyg.2014.00845. URL http://journal.frontiersin.org/Journal/10.3389/fpsyg.2014.00845/full.

[33] H. Kuzuoka, Y. Suzuki, J. Yamashita, and K. Yamazaki. Reconfiguring spatial formation arrangement by robot body orientation. *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*, pages 285–292, 2010. doi: 10.1109/HRI.2010.5453182.

[34] Adam Kendon. The F-formation system: The spatial organization of social encounters. *Man-Environment Systems*, 6:291–296, 1976.

[35] Edward Twitchell Hall. *The hidden dimension*, volume 1990. Anchor Books New York, 1969.

[36] Marco Cristani, Giulia Paggetti, Alessandro Vinciarelli, Loris Bazzani, Gloria Menegaz, and Vittorio Murino. Towards computational proxemics: Inferring social relations from interpersonal distances. *Proceedings - 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/SocialCom 2011*, pages 290–297, 2011. doi: 10.1109/PASSAT/SocialCom.2011.32.

[37] Ross Mead, Amin Atrash, and M Mataric. Automated analysis of proxemic behavior: Leveraging metrics from the social sciences. *...Robotics from the Human Sciences. ...*, 2011. URL http://robotics.usc.edu/~rossmead/docs/2011/2011MeadEtAl_RSS2011.pdf.

[38] Ross Mead, Amin Atrash, and Maja J. Matarić. Automated Proxemic Feature Extraction and Behavior Recognition: Applications in Human-Robot Interaction. *International Journal of Social Robotics*, 5(3):367–378, May 2013. ISSN 1875-4791. doi: 10.1007/s12369-013-0189-8. URL http://link.springer.com/10.1007/s12369-013-0189-8.

[39] Ross Mead and Maja J Matari. Probabilistic Models of Proxemics for Spatially Situated Communication in HRI. *The 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2014) Algorithmic Human-Robot Interaction Workshop*, pages 3–8, 2014.

[40] J. Rios-Martinez, a. Spalanzani, and C. Laugier. From Proxemics Theory to Socially-Aware Navigation: A Survey. *International Journal of Social Robotics*, 2014. ISSN 1875-4791. doi: 10.1007/s12369-014-0251-1.

[41] Dusan Jan and David R Traum. Dynamic movement and positioning of embodied agents in multiparty conversations. *Computational Linguistics*, (1968):1, 2007. doi: 10.1145/1329125.1329142. URL http://portal.acm.org/citation.cfm?doid=1329125.1329142.

[42] H Laga and T Amaoka. Modeling the spatial behavior of virtual agents in groups for non-verbal communication in virtual worlds. *Proceedings of the 3rd International Universal ...*, pages 1–2, 2009. URL http://dl.acm.org/citation.cfm?id=1667811.

[43] Claudio Pedica and Hannes Högni Vilhjálmsson. Spontaneous avatar behavior for human territoriality. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 5773 LNAI, pages 344–357, 2009. ISBN 3642043798. doi: 10.1007/978-3-642-04380-2_38.

[44] D Friedman, a Steed, and M Slater. Spatial Social Behavior in Second Life. pages 1–12, 2007. ISSN 03029743. doi: 10.1007/978-3-540-74997-4_23. URL http://discovery.ucl.ac.uk/190177/.

[45] J Llobera, B Spanlang, G Ruffini, and M Slater. Proxemics with Multiple Dynamic Characters in an Immersive Virtual Environment. 2010. ISSN 15443558. doi: 10.1145/1857893.1857896. URL http://discovery.ucl.ac.uk/1307851/.

[46] Mohammad Obaid, Radosaw Niewiadomski, and Catherine Pelachaud. Perception of spatial relations and of coexistence with virtual agents. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6895 LNAI:363–369, 2011. ISSN 03029743. doi: 10.1007/978-3-642-23974-8_39.

[47] Iason Kastanis and Mel Slater. Reinforcement Learning Utilizes Proxemics : An Avatar Learns to Manipulate the Position of People in Immersive Virtual Reality. *Transactions on Applied Perception*, 9:3:1–3:15, 2012. ISSN 1544-3558. doi: 10.1145/2134203.2134206. URL http://dl.acm.org/citation.cfm?id=2134206.

[48] Albert Mehrabian. Some referents and measures of nonverbal behavior. *Behavior Research Methods & Instrumentation*, 1(6):203–207, 1968. ISSN 1554-351X. doi: 10.3758/BF03208096.

[49] Miles L. Patterson. An arousal model of interpersonal intimacy. *Psychological Review*, 83(3):235–245, 1976. ISSN 0033-295X. doi: 10.1037/0033-295X.83.3.235.

[50] Sidney M Jourard and Robert Friedman. Experimenter-subject" distance" and self-disclosure. *Journal of Personality and Social Psychology*, 15(3):278, 1970. ISSN 1939-1315.

[51] George Breed. The effect of intimacy: Reciprocity or retreat? *British Journal of Social and Clinical Psychology*, 11(2):135–142, 1972. ISSN 2044-8260.

[52] Bob G. Witmer and Michael J. Singer. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators & Virtual Environments*, 7(3): 225–240, 1998. ISSN 1054-7460. doi: 10.1162/105474698565686.

[53] Gijs Huisman, Jan Kolkmeier, and Dirk Heylen. Simulated Social Touch in a Collaborative Game. In *Haptics: Neuroscience, Devices, Modeling, and Applications*, volume 8618, pages 248–256. Springer, 2014. ISBN 3662441926. doi: 10.1007/978-3-662-44193-0. URL http://link.springer.com/10.1007/978-3-662-44193-0.

[54] Gijs Huisman, Jan Kolkmeier, and Dirk Heylen. With Us or Against Us: Simulated Social Touch by Virtual Agents in a Cooperative or Competitive Setting. In *Intelligent Virtual Agents*, pages 204–213. Springer, 2014. ISBN 3319097660.

[55] Jeremy N. Bailenson, Nick Yee, Kayur Patel, and Andrew C. Beall. Detecting digital chameleons. *Computers in Human Behavior*, 24(1):66–87, January 2008. ISSN 07475632. doi: 10.1016/j.chb.2007.01.015. URL http://linkinghub.elsevier.com/retrieve/pii/S0747563207000040.

[56] M Ter Maat, KP Truong, and Dirk Heylen. How turn-taking strategies influence users' impressions of an agent. *Intelligent Virtual Agents*, 2010. URL http://link.springer.com/chapter/10.1007/978-3-642-15892-6_48.

A. Pilot Study Behaviour Trees

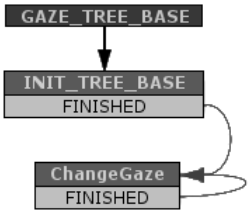


Figure A.1.: The baseline behaviour tree used for examining gaze in the pilot study. In random intervals, a new random gaze target is chosen.

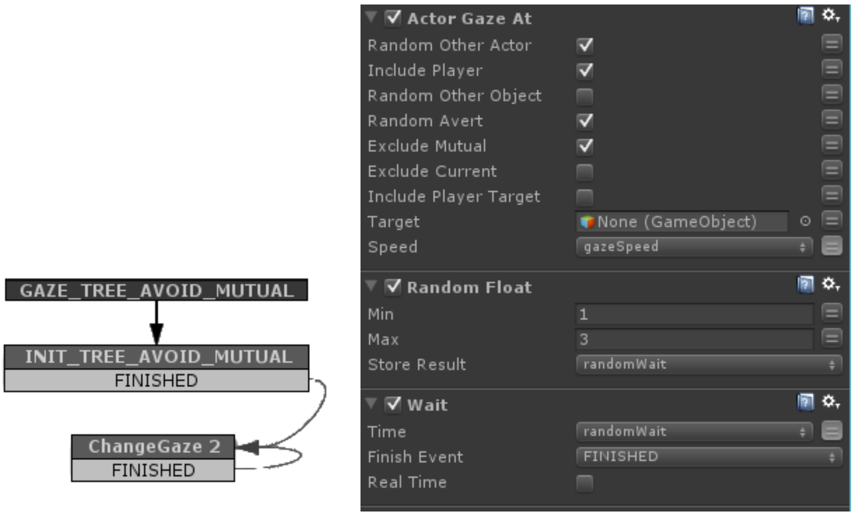


Figure A.2.: Behaviour tree used for examining *gaze aversion* in the pilot study. In random intervals, a now random gaze target is chosen. Other agents or users that currently look at the agent are excluded from the possible random targets.

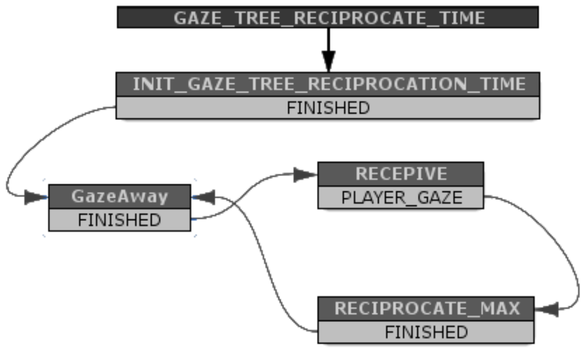


Figure A.3.: Behaviour tree used for examining *reciprocal gaze* in the pilot study. Once user gaze is detected, it is reciprocated, but after a certain interval, gaze is averted again.

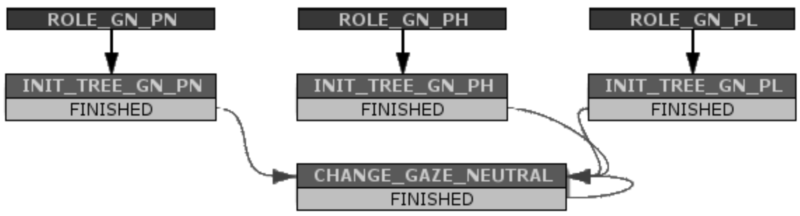


Figure A.4.: Behaviour tree used for examining *prolonged gaze* in the pilot study. Gaze is reciprocated for as long as gaze by the user is detected, and then some more.

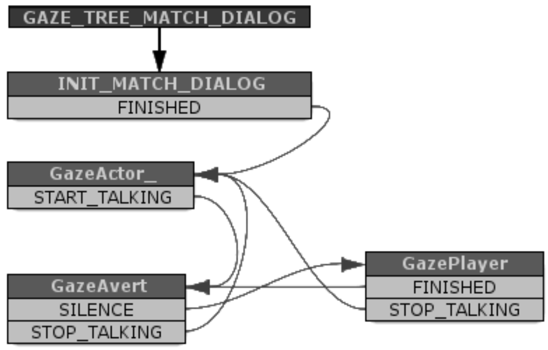


Figure A.5.: Behaviour tree used for examining *gaze matching the dialog* in the pilot study. Start and end of a dialog as well as silence are used as events to transition to different gaze targets.

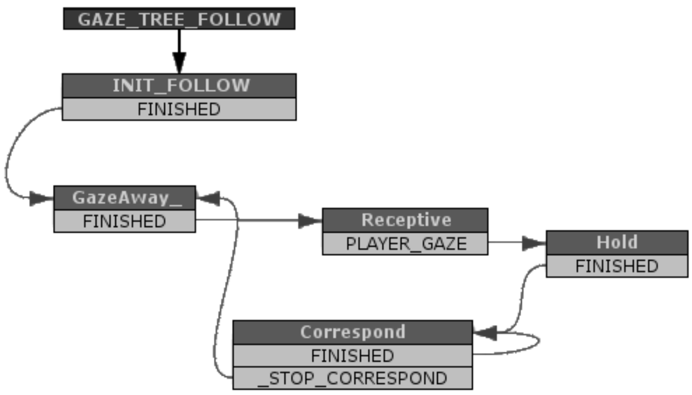


Figure A.6.: Behaviour tree used for examining *gaze following* in the pilot study. After mutual gaze was achieved ('Hold' state), the agent will check for some time (in intervals) where the user is currently looking at, and then change is gaze to that same target.

B. Experiment Behaviour Trees

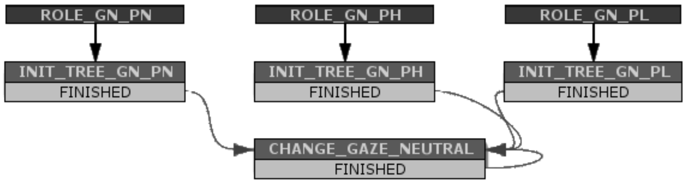


Figure B.7.: Behaviour tree used for neutral gaze behaviour during the experiment. In the 'INIT_' states proxemic behaviour is configured, as well as the intervals of *offset* averted behaviour.

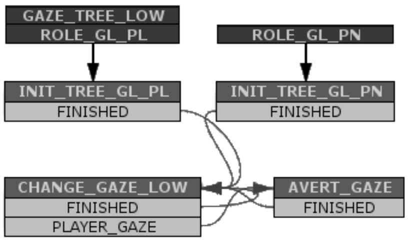


Figure B.8.: Behaviour tree used for low gaze behaviour during the experiment. In the 'INIT_' states proxemic behaviour is configured, as well as the intervals of *offset* averted behaviour.

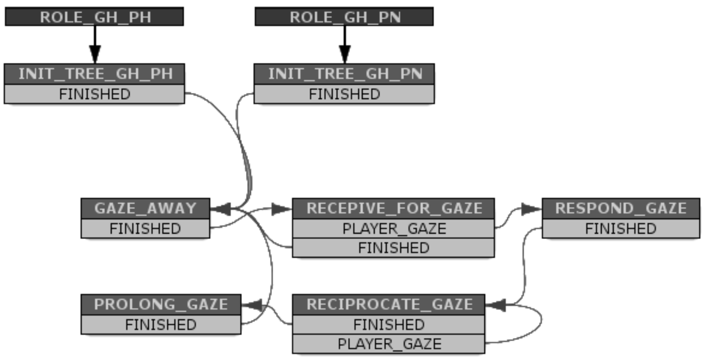


Figure B.9.: Behaviour tree used for high gaze behaviour during the experiment. In the 'INIT_' states proxemic behaviour is configured, as well as the intervals of *offset* averted behaviour.

C. Consent Form

Group	PP nr.

Consent form

UNIVERSITY OF TWENTE.

Contact information
Jan Kolkmeier, BSc (lead investigator)
Jered Vroon MSc
Dr. Gwenn Englebienne
Prof.dr. D.K.J. Heylen

Human Media Interaction group
Drienerlolaan 5
7522 NB Enschede
The Netherlands
http://hmi.ewi.utwente.nl/
053-4893740 (Secretary)

j.kolkmeier@student.utwente.nl

The University of Twente and the Department of EEMCS support the practice of protecting research participants' rights. Accordingly, this project was reviewed and approved by an Institutional Ethical Board. The information in this consent form is provided so that you can decide whether you wish to participate in our study. It is important that you understand that your participation is considered voluntary. This means that even if you agree to participate you are free to withdraw from the experiment at any time, without penalty.

The aim of this study is to collect data on how people interact with virtual humans in immersive virtual environments. The captured movement and questionnaire data thus collected will be used to create models that inform generation and recognition of behaviour.

During the experiment you will wear a *Head Mounted Display* (HMD) and headphones. Inside the HMD, you will see a virtual world. You will be able to navigate through this world naturally - by walking around. There are no obstacles in the real room, and visual helps in the virtual world indicate where the room's walls are.

In the experiment, you are member of a jury in a murder case. A young man is accused of having stabbed his father. You will find yourself in a room with other members of the jury after the main trial is over. Some of the pieces of evidence are on display. First, you will be able to examine the pieces of evidence. After some time, the other members of the jury will start a discussion. Each of them has a personal opinion about the defendant, and attempts to convince the others (including you) of it. It is your task to listen carefully to the facts, so you can make the right decision after the discussion is over.

A video is recorded only for review purposes of the research. The cameras of the tracking system are infrared, and record only the position of the markers attached to the HMD. Only the researchers will have access to identifiable data. This data will be carefully stored for at most five years (until November 2019). Non-identifiable data can be made available to other researchers in an anonymized dataset. This experiment poses no known risks to your health. If you have any questions not addressed by this consent form, please do not hesitate to ask.

Declaration of consent (please tick each checkbox if you consent)

◯ **1.** I agree to participate in this study

◯ **2**. I have read the instructions above and understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.

◯ **3**. I understand that my identifiable data is recorded for research purposes as described above, and can be stored until April 2019.

◯ **4.** I agree for my non-identifiable data to be made available to other researchers in an anonymized dataset.

____________________________ __________________

Name and signature participant — Date

____________________________ __________________

Name and signature researcher — Date

D. Questionnaires

D.1. Agent Personality Traits

1. I thought Agent was likeable
2. I thought Agent was honest
3. I thought Agent was competent
4. I thought Agent was warm
5. I thought Agent was informed
6. I thought Agent was credible
7. I thought Agent was modest
8. I thought Agent was approachable
9. I thought Agent was interesting
10. I thought Agent was trustworthy
11. I thought Agent was sincere
12. I thought Agent was friendly
13. I thought Agent was confident
14. I thought Agent was polite *
15. I thought Agent was intimate **

D.2. Presence & Involvement

1. How much were you able to control events?
2. How responsive was the environment to actions that you initiated (or performed)?
3. How natural did your interactions with the environment seem?
4. How much did the visual aspects of the environment involve you?
5. How natural was the mechanism which controlled movement through the environment?
6. How compelling was your sense of objects moving through space?
7. How much did your experiences in the virtual environment seem consistent with your real world experiences?

8. Were you able to anticipate what would happen next in response to the actions that you performed?
9. How completely were you able to actively survey or search the environment using vision?
10. How compelling was your sense of moving around inside the virtual environment?
11. How closely were you able to examine objects?
12. How well could you examine objects from multiple viewpoints?
13. How involved were you in the virtual environment experience?
14. How much delay did you experience between your actions and expected outcomes?
15. How quickly did you adjust to the virtual environment experience?
16. How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?
17. How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?
18. How much did the auditory aspects of the environment involve you?
19. How well could you identify sounds?
20. How well could you localise sounds?