Eye Tracking and Avatar-Mediated Communication in Immersive Collaborative Virtual Environments

William Arthur Hugh Steptoe

A dissertation submitted in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy

of the

University College London.

Department of Computer Science

University College London

September 17, 2010

To Peter and Patrick

I, William Steptoe, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The research presented in this thesis concerns the use of eye tracking to both enhance and understand avatar-mediated communication (AMC) performed by users of immersive collaborative virtual environment (ICVE) systems. AMC, in which users are embodied by graphical humanoids within a shared virtual environment (VE), is rapidly emerging as a prevalent and popular form of remote interaction. However, compared with video-mediated communication (VMC), which transmits interactants' actual appearance and behaviour, AMC fails to capture, transmit, and display many channels of nonverbal communication (NVC). This is a significant hindrance to the medium's ability to support rich interpersonal telecommunication. In particular, *oculesics* (the communicative properties of the eyes), including gaze, blinking, and pupil dilation, are central nonverbal cues during unmediated social interaction. This research explores the interactive and analytical application of eye tracking to drive the oculesic animation of avatars during real-time communication, and as the primary method of experimental data collection and analysis, respectively.

Three distinct but interrelated questions are addressed. First, the thesis considers the degree to which quality of communication may be improved through the use of eye tracking, to increase the nonverbal, oculesic, information transmitted during AMC. Second, the research asks whether users engaged in AMC behave and respond in a socially realistic manner in comparison with VMC. Finally, the degree to which behavioural simulations of oculesics can both enhance the realism of virtual humanoids, and complement tracked behaviour in AMC, is considered.

These research questions were investigated over a series of telecommunication experiments investigating scenarios common to computer supported cooperative work (CSCW), and a further series of experiments investigating behavioural modelling for virtual humanoids. The first, exploratory, telecommunication experiment compared AMC with VMC in a three-party conversational scenario. Results indicated that users employ gaze similarly when faced with avatar and video representations of fellow interactants, and demonstrated how interaction is influenced by the technical characteristics and limitations of a medium. The second telecommunication experiment investigated the impact of varying methods of avatar gaze control on quality of communication during object-focused multiparty AMC. The main finding of the experiment was that quality of communication is reduced when avatars demonstrate misleading gaze behaviour. The final telecommunication study investigated truthful and deceptive dyadic interaction in AMC and VMC over two closely-related experiments. Results from the first experiment indicated that users demonstrate similar oculesic behaviour and response in both AMC and VMC, but that psychological arousal is greater following video-based interaction. Results from the second experiment found that the use of eye tracking to drive the oculesic behaviour of avatars during AMC increased the richness of NVC to the extent that more accurate estimation of embodied users' states of veracity was enabled. Rather than directly investigating AMC, the second series of experiments addressed behavioural modelling of oculesics for virtual humanoids. Results from the these experiments indicated that oculesic characteristics are highly influential to the perceived realism of virtual humanoids, and that behavioural models are able to complement the use of eye tracking in AMC.

The research presented in this thesis explores AMC and eye tracking over a range of collaborative and perceptual studies. The overall conclusion is that eye tracking is able to enhance AMC towards a richer medium for interpersonal telecommunication, and that users' behaviour in AMC is no less socially 'real' than that demonstrated in VMC. However, there are distinct differences between the two communication mediums, and the importance of matching the characteristics of a planned communication with those of the medium itself is critical.

Acknowledgements

My research was supported by the EPSRC funded projects: EQUATOR (GR/N15986/01) and Eye Catching (EP/E010032/1).

Firstly, I would like to thank Dave Roberts, who supervised my undergraduate project at University of Reading, which was my first exploration into Virtual Reality. During that project I worked with the DIVE system, for which I have my primary PhD supervisor, Anthony Steed, to thank. I could not have asked for a better supervisor than Anthony, whose knowledge, enthusiasm, and support have been invaluable in developing my ideas and research skills. I would also like to thank my secondary supervisor, Simon Julier, with whom I shared teaching duties for the Virtual Environments course at UCL.

I have worked with a number of talented people (and good friends) while carrying out my PhD. Robin Wolff, Alessio Murgia, and Aitor Rovira deserve medals for their patience when acting as confederates during the three major experiments appearing in this thesis. Thanks also go to John Rae and Estefania Guimaraes, who introduced me to Conversation Analysis and gave me a deeper understanding of social interaction research. I would like to thank John in particular for teaching me lashing as we attempted to build a bamboo stand for some heavy cardboard cubes. The stand was rickety, but just about did its job. Finally, thanks to Oyewole Oyekoya, who continues to be a close colleague.

Thanks to everyone, both past and present, in the VECG group at UCL. In particular, David Swapp for managing the CAVE lab and providing support during experiments, Vijay Pawar for the mutual moaning when times got tough, Vinoba Vinayagamoorthy, Marco Gillies, and all those in room 6.22 and beyond.

Finally, and most importantly, I want to thank my family and friends for their love and support. Laura, with whom I have shared these years. Lucy, has recently brought a new addition into the family. Matt for the good times. Ed, who comfortably pipped me to the post. Nick, who will go on for great things. Love also to my grandparents. Immense gratitude goes to my father, who is an inspiration to me, and to my mother, whose love knows no bounds.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"Mr. Watson, come here, I want to see you."

These were the words, spoken by Alexander Graham Bell, and heard by Thomas A. Watson, on March 10^{th} 1876, that comprised the first telephone communication. Since then, audio telecommunication between geographically-remote people has become a ubiquitous part of life throughout the world. More recently, the rapid advancement of computing performance and computer networks has allowed real-time video-mediated communication (VMC) to enjoy a similar status. Thus, when AT&T introduced the Picturephone in 1964 [Mol69], Bell's wish to see Watson, as well as speak to him, may have been granted.

Interpersonal communication is a continuous process, in which both verbal and nonverbal information is imparted by a sender, and subsequently decoded and responded to by a receiver. Nonverbal communication (NVC) relies on the sending and receiving of signals via wordless expressive channels, and hence, mediating these signals is the primary goal of visual telecommunication systems. VMC portrays users' actual visual appearance, so is optimally-suited to the task of capturing many channels of NVC, particularly the subtle nature of facial expression. However, the three-dimensional (3D) nature of social interaction in reality ensures that many nonverbal cues, such as movement, gesture, and gaze direction are inherently spatial. It is in this regard that Bell's intermediate request, for Watson to "come here", is still some way from being achieved convincingly with VMC, being as it provides a 'window' into a user's surroundings, rather than creating the illusion of a shared space.

To this end, avatar-mediated communication (AMC) in immersive collaborative virtual environments (ICVEs) has recently, and rapidly, emerged as an alternative mode of visual telecommunication. Spatiality is a central feature of such immersive virtual interaction, and users populate a shared virtual environment (VE) with humanoid graphical embodiments, known as *avatars*. The idea of being immersed in a VE, generated by computer displays, dates back to Ivan Sutherland's work during the 1960s [Sut68]. Sutherland's system comprised a formidable structure suspending a head-mounted display (HMD), capable of generating simple wire-frame graphics. It was the advent of multi-user virtual reality (VR, a term credited to Jaron Lanier in the early 1980s) systems which introduced the paradigm of user embodiment within a VE, implying that users "puppeteer" their avatars via natural bodily movement. Thus, based on natural sensory input and proprioception, ICVE users are able both to transmit and observe nonverbal signals in a shared and spatial setting.

The term “avatar” finds its origins in the religious tradition of Hinduism, in which the term refers to an Immortal Being’s descent and incarnation, or embodiment, into mortal realms. In the domain of computer science, the semantics of the term are similar, but the embodiment that takes place is between people, in reality, and graphical humanoids, in VR. Released in 1986, Lucasfilm’s online role-playing game, Habitat [GL86], provides perhaps the first example of the term as it is now used in today’s online culture. Habitat may also be seen as a forerunner to the online virtual worlds of today, which enable millions of remote users to engage in AMC, catering to both small-group and large-scale social interaction. Currently, the vast majority of users enjoying this emerging form of mediated interaction are doing so via non-immersive computing hardware, comprised of standard displays and input devices. However, driven by the video games and film industries, immersive hardware devices, such as motion tracking interfaces and large field-of-view stereoscopic displays, are emerging for both home and commercial use. Historically, such hardware has been the domain of VR labs, in which much of the research and development into AMC continues to be carried out. However, as consumer technology evolves and is adopted, it is certain that over the coming years, forms of AMC which are technically comparable to that which is investigated by the research presented in this thesis will become commonplace.

1.1 Research Problem

A major hindrance to high-quality interpersonal AMC is the fact that many channels of users’ nonverbal expression are not transmitted by the medium. Unlike VMC, which transmits a user’s actual visual appearance, AMC generally relies upon specialist tracking devices to capture and replicate a user’s expressive behaviour in their embodied avatar. Hence, an avatar’s capacity for nonverbal replication is dependent on the tracking devices available to the system operating the ICVE. Studies into small-group interaction in ICVE systems typically employ minimal bodily tracking devices to drive avatar animation, generally consisting of head tracking to update position, gross posture, and head orientation, and a single hand tracker to infer arm gesture [BBH+90, BBF+97, SSA+01, RWOS03]. Thus, while AMC in ICVEs may offer a spatial and immersive context for interaction, the relative lack of nonverbal information exchange is clearly lacking compared to VMC. A significant challenge to the maturation of AMC is the capture of critical channels of users’ natural NVC, which may then be used to drive the expressive behaviour of an embodied avatar in real-time, thereby developing avatars that more faithfully exhibit their embodied user’s actual nonverbal behaviour.

The complexity of human behaviour ensures that holistic reproduction of users’ behaviour in AMC is not a realisable goal at the present stage of maturity in the research area. Indeed, Schroeder suggests that AMC will never provide completely realistic ways of interacting or communicating with others because a number of features of face-to-face interaction will always be lacking [Sch02]. However, the bidirectional flow of nonverbal information is considered to be a critical requirement in order to support avatar-mediated interactions that are more similar to those which take place in reality, between humans [Sch02, SRSH05]. This argument is derived from the importance of the visual, nonverbal, element of collocated communication, where people draw impressions of one another regarding intention, mood, and emotional state [AT79]. It is important, then, to focus on critical behavioural cues, which, unless they are tracked and replicated, are likely to be a hindrance to communication.

Oculesics, or the use of the eyes during communication, are of central importance to social behaviour and NVC, and the eyes are considered to be the most intense social signallers on the human face [Arg88]. Gaze is perhaps the most salient oculesic cue, providing a bidirectional signal used both to monitor and indicate focus of attention, as well as a resource to manage the unfolding interaction. More subtle, but providing rich insight into an individual's emotional and cognitive states, are pupil dilation and blinking. The capture of oculesics may be achieved using eye tracking.

The research presented in this thesis investigates the use of eye tracking, with the parallel aims of enhancing and understanding AMC in ICVE systems. The first challenge, to enhance AMC, aims to enable oculesics as a nonverbal resource during the virtual social interaction, similar to how oculesics are both transmitted and observed in face-to-face meetings. This implies that the *interactive* capability of eye tracking must be exploited to capture users' oculesic cues, and replicate them in their embodied avatars, in real-time. The second challenge, to understand AMC, aims to make use of the *analytical* capability of eye tracking, to gather data describing users' behaviour and response during the mediated interactions.

1.2 Research Questions

The overarching goal of the research presented in this thesis has been to investigate how eye tracking may be used to both enhance and understand AMC. The research extends earlier studies in the VE literature, by conducting a series of controlled experiments designed to observe social interaction performed by users engaged in AMC. The main experimental research, presented in Chapters 4–6, investigates small-group interaction between users of state-of-art visual telecommunication systems. The three chapters are each concerned with a specific collaborative scenario, and document the associated experiments. Chapter 4 investigates a three-party conversation scenario in both AMC and VMC, Chapter 5 investigates three-party object-focused interaction in AMC, and Chapter 6 investigates truthful and deceptive interaction in both AMC and VMC. In addition to these telecommunication studies, three experiments investigating behavioural simulations of oculesics are presented, in Chapter 7. These studies investigate factors influential to AMC, rather than AMC itself. Section 7.1 presents a preliminary experiment assessing how behavioural fidelity is seen to impact observers judgements of an avatar. Section 7.2 presents and investigates parametric models of eyelid motion, which generate realistic blinks and lid saccades. Finally, Section 7.3 presents and investigates a gaze model, which aims to generate realistic and meaningful gaze behaviour by computing the saliency of the objects within a VE.

The experimental work was guided by, and addressed, the following overall questions:

1. *Can eye tracking be used to increase the nonverbal information transmitted during AMC, and does this improve quality of communication between users?*

2. *Measured by eye tracking, do users engaged in AMC behave and respond in a socially realistic manner compared to users engaged in VMC?*

   These two questions are addressed by the telecommunication experiments presented in Chapters

4–6. The work addresses the central premise of whether eye tracking may be applied both interactively, to enhance the nonverbal richness of AMC (Question 1), and analytically, to further understand peoples' social behaviour and response during AMC (Question 2). In Chapters 4 and 6 VMC provides the base class of visual telecommunication with which AMC is compared, whereas in Chapter 5, only AMC is investigated.

3. *Can simulations of oculesic behaviour have a positive impact on observers' perceptions of virtual humanoids, and can such models be used to complement tracked behaviour in AMC?*

   The final question (Question 3) is addressed by the behavioural simulation work presented in Chapter 7. The work is secondary to the central focus of the research, and presents models of oculesics, together with associated experimental evaluations, which assess the models' ability to generate realistic oculesic animation for avatars and agents. The combined use of tracking and modelling to drive embodied avatars is also investigated.

1.3 Contributions

The main contribution of this thesis is the evaluation of the use of eye tracking to enhance and understand AMC in ICVE systems. While the work's driving motivation lies in the aspiration to enhance AMC, insight into the understanding of how people behave and respond when engaged in AMC is a no less fundamental goal. The research covers multiple collaborative scenarios, comparisons with state-of-art VMC, methods for representing gaze, blinks and pupil dilation data in AMC, models of gaze and eyelid movement, networked VE system design and development, and analytical techniques. The contributions of this thesis can be classified as methodological and substantive:

1.3.1 Methodological Contributions

1. Methods of processing the tracked oculesic cues of gaze, blinks, and pupil size, and mapping them onto an embodied avatar for real-time display of replication of behaviour in AMC (Chapter 3).

2. A method for collecting multimodal interaction data in a single log-file (Chapter 3). Data comprises of oculesics, head movement, hand movement, verbal signals, and additional experimental markup information.

3. Experimental task designs for use in studies on interpersonal and object-focused AMC (Chapters 4–6), and experimental frameworks for studying perceptual aspects of virtual humanoids (Chapter 7).

1.3.2 Substantive Contributions

1. A review of the problematic use of simulation, as opposed to tracking, to control avatar behaviour during AMC (Chapter 2). Focusing on gaze, the review argues that, due to the unpredictable, interpersonal, and idiosyncratic nature of human social interaction, it is not feasible to simulate a user's nonverbal behaviour without distorting the semantic content of the human communication. Hence,

for faithful telecommunication mediated by avatars, tracking and replication must be implemented wherever possible.

2. Research findings that address the impact of the interactive and analytical use of eye tracking in AMC (Chapters 4–6). These findings have implications for the design of future AMC systems.

3. Oculesic models which generate realistic animation of eyelid kinematics and gaze behaviour for use in real-time AMC (Chapter 7).

1.4 Scope of Thesis

This thesis is concerned with the operation and evaluation of eye tracking in AMC that takes place between users of ICVE systems. The focus of the research is therefore not on the technology itself, but rather on the use of the technology to support both interpersonal and object-focused collaboration. However, the software platforms used throughout are bespoke, and have been developed with the experimental evaluations in mind. Where appropriate, key phases of development, and system overviews are provided.

The three telecommunication experiments are concerned with small-group collaboration performed by users of highly immersive state-of-art VR technologies such as CAVE™ systems [CNSD93]. Chapters 4 and 5 investigate three-party collaboration between three networked CAVE systems in conversational and object-focused scenarios, while Chapter 6 investigates truth and deception between dyads performing between a CAVE system and a semi-immersive VR system, the WALL. Thus, the research presented in this thesis does not consider HMDs, which hinder natural proprioception, or non-immersive systems, such as desktop displays or mobile devices.

There is little work in the VE literature investigating avatars featuring a non-humanoid form [FSS07], and the research documented here is no exception. Reeves and Nass proposed that if an agent or avatar looks or behaves like a human, observers will have an automatic social response to it, similar to how they would respond to another person [RN96]. Findings in the VE literature unambiguously support this theory, which is detailed Chapter 2. Thus, the avatars appearing in this work exhibit high levels of anthropomorphism, and cartoon-like or semi-realistic avatars are not investigated. This is reflected in the behavioural simulation work, which presents physiologically-accurate models of eyelid movement, and simulation of human gaze behaviour.

Finally, as covered in Chapter 2, there are several potential applications of eye tracking in VEs, some of which could be used or adapted to benefit collaboration in ICVEs. However, this work is explicitly concerned with real-time AMC, and the use of eye tracking to enhance NVC through replication of oculesics, including gaze, eyelid movements, and pupil dilation. In accordance to the realistic appearance of the avatars embodying each user, this research is concerned only with faithful replication of behaviour to achieve high-quality AMC, as opposed to "transformed" social interaction, as studied by Bailenson et al. [BBL+05]. From an analytical standpoint, eye tracking is used as the primary data source to describe the captured interactions, but questionnaire and performance metrics, as commonly used in VE studies, are also applied throughout.

1.5 Structure

Chapters 2 and 3 are introductory and cover relevant research and methods. Chapters 4–6 present the design and findings of three studies investigating eye tracking in AMC. Chapter 7 presents simulations of oculesics and associated experiments. Chapter 8 draws conclusions from the findings and propose directions for continuing research.

Chapter 2 contextualises the research by expanding upon the motivation, the central problem addressed, and the general approach taken. Although ICVEs are able to support spatial and fully-immersive interaction, a significant deficiency of the medium, as a form of interpersonal telecommunication, is the limited expressivity of users' avatar embodiments. The communicative properties of the eyes, their relevance to visually-mediated interaction, and previous work on avatar oculesics inform the general approach taken in this research. Finally, the theoretical arguments surrounding issues related to reproducing users' gaze with eye tracking, versus simulating eye movement through gaze modelling, are awarded detailed discussion.

Chapter 3 covers both technical and methodological aspects of the research. Firstly, the design and key features of *EyeCVE* are presented. EyeCVE is the ICVE system which was developed (in collaboration) throughout the course of this research, and was used as the primary platform supporting the experimental AMC. The eye tracking and avatar subsystems are presented in detail, followed by a performance evaluation of an early version of the system. A multimodal data collection system that is closely coupled to EyeCVE is then presented. During an EyeCVE session, a user performance is captured using several tracking devices, each monitoring a separate channel of natural communication. These data input streams include eye tracking, head tracking, hand tracking, and microphones which detect verbal communication. The data collection system is able to collate the synchronous data streams in an holistic manner, thereby preserving the temporal characteristics of the multimodal action. Together with the traditional approaches to measuring user experience and quality of communication in VEs, this data collection system is used to analyse the experimental interactions presented in Chapters 4–6.

Chapter 4 presents the first of the three telecommunication experiments which form the main empirical research contribution of this thesis. The first experiment investigates three party conversation in tracked gaze AMC and gaze aware VMC. In the VMC setting, gaze awareness is realised by careful alignment of video displays and camera positions, while the AMC setting uses an early version of EyeCVE, networking three CAVEs, to drive avatar gaze. Eye tracking is combined with conversation analysis as the primary means of evaluation and analysis of the interactions.

Chapter 5 presents the second telecommunication experiment, again performed between three CAVEs, networked with EyeCVE. The experiment investigates the impact of varying methods of avatar gaze control on quality of communication during object-focused multiparty AMC in ICVE systems. An experimental puzzle scenario was devised to emphasise the operational importance of gaze during the collaborative interactions. The experiment compared three forms of AMC: tracked gaze using eye tracking, static gaze featuring no eye movement, and a simple gaze model. Eye tracking data, performance timing, and questionnaires were used to assess the interactions in terms of task performance, subjective user experience, and interaction analysis.

Chapter 6 presents the final telecommunication study, consisting of two closely related experiments investigating truthful and deceptive interaction in AMC and VMC. The first experiment in the chapter aims to compare participants' behaviour during truthful and deceptive conversation between dyads. Alongside gaze, blinking and pupil size are also captured for real-time representation in AMC, and also for post-experimental analysis. Following the interactions, a questionnaire collects data describing participants' psychological arousal and mood state. The second, follow on, experiment aims to assess the impact of bestowing avatars with faithful reproduction of their embodied users' tracked oculesic behaviours of gaze, blinks, and pupil dilation. The AMC recorded from the first experiment are replayed to a new set of participants, who assess the original participants in terms of veracity and engagement, together with confidence levels relating to the two judgements. These ratings are performed over three stimuli conditions: avatars exhibiting oculesics, avatars featuring no oculesics, and audio-only.

Chapter 7 presents models of oculesics and associated experiments. The first of three sections in the chapter investigates the impact of varying the fidelity of an avatar's oculesic behaviour on observers' ability to identify its direction of gaze, together with subjective perceptions of realism. The second section presents two parametric models, based on physiological data, that generate lid saccades and blink animation. The models are detailed algorithmically, before being validated against motion captured data, and then being assessed in terms of general impact on realism. The third section focuses on gaze modelling. The gaze model computes the saliency of objects in the VE, and animates avatar gaze accordingly. The model is assessed in terms of realism against tracked gaze, the basic gaze model presented in Chapter 5, and static gaze.

Chapter 8 draws conclusions and gives suggestions for future work.

Chapter 2

Background

Driven by the potentials and demands of an increasing global market, and fed by advances in information and communication technology, one of the trends in the modern workplace is for more distributed team working. Within the domain of computer supported cooperative work (CSCW), telecommunication systems that enable synchronous interaction amongst geographically-separated users are referred to as *remote interaction systems* [Joh88]. *Visual telecommunication systems* comprise that subset of remote interaction systems which aim to transmit visual information, usually alongside the aural component of an unfolding communication. State-of-art VMC systems remain the most favoured form of visual telecommunication in terms of supporting NVC between users. This thesis focuses on AMC performed by users of ICVE systems as a maturing form of visual telecommunication, but one which has superlative potential for fostering the perceptual illusion of being collocated with others in a shared space [Sch06].

This chapter aims to contextualise the research presented in this thesis, by discussing the literature that has shaped its motivation, the problem it aims to address, and the approach it takes. The chapter is comprised of four main sections, which narrow down the focal area of research in a top-down manner. The first section explores natural human nonverbal communication in collocated (face-to-face) small-group interaction, with a particular focus on the communicative properties of the eyes, known as oculesics. The second section discusses visual telecommunication systems, and covers the relevant literature in small-group CSCW, with a particular focus on VMC and AMC. The issues of gaze communication, termed *gaze awareness*, in VMC systems is covered in detail, together with topics of immersion, place illusion (PI, also known as *presence*) and spatiality in ICVEs. The third section explores the research's central topic of AMC. The importance, and methods, of expressing nonverbal behaviour are presented. Current failings of these methods, which limit the ability of avatars to act as the visual mediators for interpersonal telecommunication, are presented. The theories of social presence, media richness, and avatar fidelity are introduced. A review of oculesic animation, including gaze, eyelid dynamics, and pupil dilation is then presented. The fourth section motivates the research problem by discussing the problematic implications which arising from the use of behavioural simulation to drive avatar animation during AMC, as opposed to tracking and replicating users' actual behaviour. By reviewing analyses of social interaction, this section identifies common gaze practices that add information that is critical to those engaged in an unfolding communication, but which cannot be predicted or determined by a behavioural simulation, nor indeed by the participants themselves in the unfolding interaction. Hence, the motivation for the use of eye tracking to capture users' oculesics for real-time replication by embodied avatars is developed, drawing into the focal area of the research.

2.1 Communication and the Eyes

Communication in general occurs in order for some number of parties to perform in four different task contexts [WPKD02]: *cooperation* to perform a task together, *coaction* to exist in the same vicinity, *competition* to perform a task at the expense of another, and *conversation* to entertain or pass on information. The act of communication is a continuous process in which discrete packages of information are imparted by a sender, and subsequently decoded and responded to by a receiver. This bidirectional feedback loop must be supported by at least one channel or medium, and hence requires both semantic and syntactic commonality between all parties. In collocated communication, the process of decoding messages is a part of *intrapersonal* communication, or internal and private cognition [Lan03], while the sending and receiving of messages is a process of *interpersonal* communication, in which participants have access to the full gamut of sensory channels [Bor04]. Collocated interaction is classified as *unmediated communication*, as it involves direct exchange of information between interactants, with no intervening agents or technologies.

Mediated communication relies on a process by which messages are transmitted via some form, or medium, external to direct face-to-face interaction [PM04]. Parties involved in mediated communication may still be collocated, and have partial access to unmediated expressive cues. For instance, the task of linguistic translators is to decode, recode and send verbal messages between parties sharing a common language barrier, but the nonverbal channels such as body language, touch, and eye contact may generally still be observed directly [OA02]. In the case of translation, a by-product of the mediation is that the two major components, visual nonverbal and aural verbal, of a communication are separated by some temporal measure throughout the exchange. Thus, the verbal signal reaches the receiver at an offset delay from the time of the original utterance by the sender, and with that delay, any concomitant nonverbal signalling has already occurred.

Generation of verbal and nonverbal cues are highly interrelated, both temporally and semantically [AT79], so alteration of these characteristics has the potential to influence a communication from a subtle to drastic manner. Unless a telecommunication system is "perfect", transmitting and displaying all communicative channels and spatial context in a timely manner, alterations arising from mediation are always present. This situation holds true for all forms of mediated communication, whether human or technological. Thus, mediated communication may be viewed as applying a filter to direct, unmediated, collocated interaction. Transmitting high quality and synchronous verbal and nonverbal signals is the central goal of visual telecommunication systems. Bodily movement and expression is the primary source of the visual component of human interaction, and is briefly reviewed in the following section. In accordance to the aims of this research, the subset of NVC related to the communicative and physiological properties of the eyes is then explored in detail. The behaviours of gaze, eyelid movement, and pupil dilation in common social interaction scenarios are discussed, followed by a review of how oculesics are influenced by deceptive situations.

2.1.1 Nonverbal Communication: an overview

Nonverbal communication (NVC) is a subset of human communication which relies on the sending and receiving of signals via wordless expressive channels. NVC can give tone, accent, and sometimes even override verbal messages [AT79]. Channels of NVC include, but are not restricted to, gaze facial expression, gesture, bodily movement, posture and contact, physical proximity and appearance [Arg88]. NVC provides information regarding beliefs, desires, and intentions of an individual, and also provides indicators regarding various psychological states, including cognitive, emotional, physical, intentional, attentional, perceptual, interactional and social [Duc86]. The corresponding verbal component of communication is similar in that every message contains insight into the psychological state of the individual and the relationship between the individuals in the conversation [WBJ+67]. Nonverbal signals are able to manipulate the verbal component of face-to-face communication, and even though verbal and nonverbal content might not always indicate the same message, what they convey is almost always compatible [GRA+02].

Kinesics is the study of the visual aspects of nonverbal communicational behaviour related to movement: either of any part of the body, or the body as a whole [Bir60]. Burgoon et al. consider kinesics to be one of the richest nonverbal codes [BBW96], providing people engaged in collocated interaction with a wealth of information. This is due, in part, to the range of kinesic cues, which includes posture, gesture, and facial expression. Argyle classifies three main human postures: standing, sitting (including squatting and kneeling), and lying down [Arg88]. Each of these has further variations corresponding to different positions of the arms and legs, and different angles of the body. In any given culture, there are a finite number of postures commonly used, and the ways in which attitudes and emotions are expressed vary accordingly. Observation of interactants' posture during collocated communication is able to provide insight, both with regards to how they may feel about one another, and information of their emotional state [Meh07].

Gestures, while not exclusively related to emotional expression, are a ubiquitous part of human communication. Gestures are voluntary bodily movements, usually made by the head or hands, which are intended to independently communicate, or correlate with, some verbal utterance [Arg88]. Ekman and Friesen classify the structural elements of gesture as emblems, illustrators, affect displays, regulators, and adaptors [EF69]: emblems have a direct verbal translation, and are most often used with concious intent to transmit a message, such as lifted shoulders and upturned palms to indicate "I don't know"; illustrators aid in the description of what is being said; affect displays are used to display emotion; regulators aid in the turn-taking of conversations; and adaptors are those behaviours which are essentially private reactions to stimuli, such as fidgeting when nervous, and are generally thought of as being involuntary [Mas96].

Facial expressions displaying the basic emotions of anger, disgust, fear, joy, sadness, and surprise, take specific forms that are universal across human cultures [Ekm92]. The answer to why this is the case can be traced back to Darwin, who argued that facial expressions are primarily vestiges of behaviours that, previously in our evolutionary history, had specific and direct biological function [Dar04]. As exemplified by Krauss et al. [AS96], for a species that attacked by biting, baring the teeth was a necessary prelude to an assault, and wrinkling the nose reduced the inhalation of foul odours. The reason that these formally functional behaviours have persisted as facial expressions (people bare their teeth when they are angry, and wrinkle their noses in disgust despite absence of an unpleasant odour) is because they have acquired communicative value: they provide observers with external evidence of an individual's internal state [Hin72].

Before the literature regarding oculesics is presented, the communicative modalities of *proxemics* and *chronemics*, related to space and time respectively, are detailed. Management of the exocentric environment, in particular with regards to the physical proximity of the communicating parties, is fundamental practice during collocated communication. Hall defines proxemics as the study of how humans unconsciously structure microspace in terms of the interpersonal distance during daily transactions, and extends the theory to the organisation of space in buildings, and ultimately, the layout of towns [Hal63]. Table 2.1 shows Hall's four concentric, but not necessarily circular, ranges. The study of proxemics is particularly useful when determining interpersonal relationships and current purpose of the communication. Embracing, touching or whispering is generally reserved for intimate distance, interactions among good friends generally occurs in the personal range, interactions among acquaintances is performed at social distance, and public distance is used for public speaking.

Zone	*Phase*	*Range*
Intimate	Close phase	< 15 cm
	Far phase	15 - 45 cm
Personal	Close phase	45 - 75 cm
	Far phase	75 - 120 cm
Social	Close phase	1.2 - 2.1 m
	Far phase	2.1 - 3.6 m
Public	Close phase	3.6 - 7.5 m
	Far phase	≥ 7.5 m

Table 2.1: *Proxemic zones, phases and ranges [Hal63].*

Chronemics refers to how people perceive, structure, and react to time, and the messages that may be interpreted from such usage [Bru80]. Chronemics is studied on two scales: those related to social and cultural organisation, and those that are managed on a micro-level during interpersonal communication. There is growing interest in chronemics in the CSCW literature [Tid95, KR05]. The use of time as a communicative channel can be a powerful, if subtle, force in collocated interaction, where it is used for regulation, expression, and affect management [Bru80]. Regulation refers to managing the orderly transition of conversational turn-taking. For instance, when the current speaker "opens the floor" for a response, they will pause, or when no response is desired, they may talk at a faster pace with minimal pauses [Cap85]. Expression refers to changes of chronemic behaviour depending on relationship type. For instance, people who are part of an intimate relationship will often maintain mutual eye contact for a length of time beyond what would be the norm in a non-intimate relationships [Pat90]. Affect management is related to the onset of powerful emotions which may cause a stronger affect, such as joy, sorrow, or embarrassment. Some of the behaviours associated with negative affects include decreased time of gaze and awkwardly long pauses during conversations. When this happens, it is common for the individuals to try and decrease any negative affects and subsequently strengthen positive affects [EI87].

2.1.2 Oculesics

The eyes have been characterised as the most significant area of the body for communicating messages [RM95], and the most intense social signallers in the human face [Arg88]. As Heron states, "The most fundamental *primary* mode of interpersonal encounter is the interaction between two pairs of eyes and what is mediated by this interaction. For it is mainly here, throughout the wide ranges of social encounter, that people actually *meet* (in the strictest sense)" [Her70].

As was documented in the preceding kinesics literature regarding facial expression, Darwin argued that an intimate relationship exists between almost all human emotions and their outward manifestations [Dar04]. Subsequently, Ekman and Friesen have demonstrated universal recognition of basic emotions [EF71, EF75]. This is a clear indication that symptoms of a cognitive mental state are observable in the face. Extending emotional communication by facial expression, Baron-Cohen et al. have shown that a range of mental states beyond these basic emotions can be read from facial expressions, and particularly from the eyes [BCWJ97]. This theory is aided by normal adults' ability to consistently interpret both basic and complex mental states in the face. Baron-Cohen et al. suggest that, regarding complex mental states, the eyes convey as much information as the whole face, and significantly more information than the mouth. This nonverbal communicative channel of the eyes is what Baron-Cohen terms the "language of the eyes" [BC95]. More commonly, it is termed *oculesics*, and in this thesis, the term is employed to refer to the communicative properties of the eye, including gaze, pupil dilation, and eyelid movement.

Gaze

Human capacity for information processing is limited, so inspection of a visual scene is performed by paying particular attention to selected stimuli of interest. In natural scenes, attentional selections are influenced by complex interactions between top-down goal driven control, and bottom-up stimuli driven control [IKN98]. Understanding how visual targets compete for attention is critical in understanding how gaze is allocated to targets within a visual scene. While attention has been shown to be influenced heavily by top-down components, referring to visual inspection with a task in mind (e.g., looking for John Doe in a crowd), it has also been shown to be strongly influenced by the bottom-up approach guided by the contents of the visual input [Yar67].

Gaze is of central importance in social behaviour, where it acts both as a nonverbal signal, showing, for example, direction of attention, while also opening a channel so that another person's nonverbal signals, and particularly their facial expression, can be received [AF67]. Hence, gaze is both signal for the person being looked at, and a channel for the person doing the looking. We use our eyes to study the behaviour and appearance of others, and we look particularly in the region of the eyes. For instance, when looking at a photograph or painting of a person, we will typically spend 60% of the time looking at the eyes, and 15% at the mouth [PW96]. *Mutual gaze*, or eye contact, refers to when two people are looking at each other's eyes at the same time, and is used as a bond-forming social signal very early in life: mutual gaze between mother and child first occurs at the age of three or four weeks [BB01]. Several types of gaze behaviour in human communication have emerged, and Cranach offers the following definitions as presented in Table 2.2 [Cra71]. While some of Cranach's definitions may appear ambiguous or unnecessary, it is nevertheless an attempt to clarify the inconsistent use of gaze terms in the psychology literature; the semantics of which are often only possible to decode within the context of a particular study's methodology [HWM78]. This thesis, consistent with the modern literature, uses 'gaze' similarly to 'look' [Rut84], and uses explicit phrases such as 'gaze at the face', as opposed to 'face gaze' as defined by Cranach.

Term	*Definition*
Onesided look	Gaze by one person in direction of another's face.
Face gaze	Directing of one person's gaze at another's face.
Eye gaze	Directing one person's gaze at another's eyes.
Mutual look	Two persons gaze at each other's face.
Eye contact	Two persons look into each other's eyes and are aware of each other's eye gaze.
Gaze avoidance	Avoidance of another's eye gaze.
Gaze omission	Failure to look at another without intention to avoid eye contact.

Table 2.2: *Definitions of gaze behaviour by Cranach [Cra71].*

Argyle et al. specify six different functions of gaze in social interaction: information seeking, signalling, regulating conversation flow, initiating intimacy, avoidance, and limiting distraction [AIAM73]. Information seeking may be used to obtain feedback of others' reactions by observing their nonverbal behaviour. Signalling is used to express interpersonal attitudes and emotions. For example, people tend to gaze more at those we like than those we dislike [EW65], competitors may gaze intently at one another [EGS65], and couples in love spend more time engaged in mutual gaze [Rub74]. Regulating conversation flow aims to control and synchronise speech, and depends on sending and receiving gaze-shift signals. Kendon found that shifts of gaze are systematically coordinated with the timing of speech, and for instance, if a speaker does not establish mutual gaze at the end of an utterance, there is a longer pause before the other replies [Ken67]. The initiation of intimacy employs mutual gaze as a main cue, and this increased with close proxemic range, which is itself another cue for intimacy [AIAM73]. Avoidance involves inhibition of gaze to avoid situations of undue intimacy, revealing one's inner feelings, or perceiving negative feedback. Similarly to initiating intimacy, the influence of gaze avoidance decreases as interpersonal distance increases [AI72]. Limiting distraction is used by speakers to avoid excess input of information when they wish to maintain focus. In particular, a speaker will look away from listeners during the initial planning phase of an utterance when cognitive demand is high [Ken67].

These functions of gaze behaviour operate, during an interpersonal encounter, to determine an equilibrium level of intimacy, what Argyle and Dean term 'equilibrium theory' [AD65]. The theory states that, as one behaviour increases to disturb the equilibrium level of intimacy, another may decrease to restore that equilibrium. Testing the theory, they compared the proximity from various stimuli that participants chose to position themselves. Participants positioned themselves closest to a photograph of Argyle with his eyes closed, next closest to a similar photograph but with his eyes open, next closest to Argyle himself with his eyes closed, and furthest from Argyle himself but with his eyes open. This is a simple demonstration of the increased intimacy caused by gaze, and the compensatory response to signal avoidance via increased interpersonal proximity. This result has been reproduced using immersive VR, featuring avatars exhibiting various gaze behaviours [BBBL01]. This study, and others, are documented in Section 2.3.

There has been much work regarding the gaze behaviour of dyads engaged in conversation. A common metric used for evaluation involves measuring the time people *look at* and *look away from* their partner when speaking and when listening. This approach to gaze behaviour analysis enables interaction to be classified into discrete blocks. Consequently, this structure is often adopted, algorithmically, by avatar gaze models [PLBB02], which is further documented in Section 2.3.6. When two people are talking, they look at each other between 25% and 75% of the time [AF67]. Critically, they look nearly twice as much when listening compared to when speaking. Individual glances generally last for anything up to seven seconds, and mutual gazes are rather shorter. Relating to the gaze behaviour of limiting distraction, if the conversational topic is cognitively difficult, people tend to look at each other less in order to avoid distractions which, if present, will draw gaze and may become a focal point especially if relevant to the conversation [AG77]. Table 2.3 presents mean gaze duration proportions from an experiment by Argyle and Ingham, investigating dyads talking on an emotionally neutral topic at a distance of 1.8 metres (close phase of social distance as defined by proxemics grammar [Hal63]) [AI72]. *Individual gaze* refers to one conversational partner looking at the other's eyes, and *mutual gaze* indicates that this condition is true for both partners.

Type of Gaze	*Percentage / Time*
Individual Gaze	61%
While Listening	75%
While Speaking	41%
Length of Glance	2.95 seconds
Mutual Gaze	31%
Length of Mutual Gaze	1.18 seconds

Table 2.3: *Gaze distribution during dyadic conversation at social distance [AI72].*

There is significantly less existing work that investigates gaze behaviour during group interactions consisting of three people or more. The structure of multiparty conversation is more complicated than that between dyads, and the complexity of gaze behaviour increases accordingly. When a third participant is introduced, the next turn is no longer guaranteed to be the initial speaker, and when the number of parties rises beyond three, it becomes possible to have side-conversations between subgroups [VSvdVN01]. Exline studied gaze distribution during conversation between same-sex triads, the results of which are presented in Table 2.4 [Exl63]. The values demonstrate less individual gaze in triads compared to Argyle and Ingham's study on dyads, previously discussed in Table 2.3, and proportion of mutual gaze is also low in comparison.

Vertegaal studied the synchronization between auditory and articulatory attention during four-person conversations by measuring gaze directed at partners' faces when speaking and listening [Ver99]. Results suggest that when someone is listening to an individual, there is a high probability (88%) that the person they look at is the person they are listening to. Likewise, when someone is speaking, the probability is 77% that the person they look at is the person they are speaking to. The authors also found

Type of Gaze	*MMM*	*FFF*
Individual Gaze	23.3%	37.3%
While Listening	29.8%	42.4%
While Speaking	25.6%	36.9%
Mutual Gaze	3%	7.5%

Table 2.4: *Gaze distribution during male (MMM) and female (FFF) triads [Exl63].*

that when a person starts speaking to all three listeners, they will typically distribute gaze over all parties. In this condition, the total percentage of mutual gaze for the speaker rises to 59%. Hence, similarly to dyadic conversation, gaze is an excellent predictor of conversational attention in multiparty discourse, where it also provides a central resource for turn management, and heightens the influence of shifts in gaze in relation to speech. A common practice employed by speakers in dyadic conversation is to give a lengthy glance to their partner at the completion of an utterance or thought unit, thereby yielding a speaking turn to the listener, which persists until the speaking role is assumed. In multiparty scenarios, gaze becomes a primary communicative resource to maintain and regulate turn-allocation. The range of such communicative signals and their associated common gaze signals are presented in Table 2.5 [GB06].

State	*Signals*	*Gaze Behaviour*
Speaker	Turn yielding	Look toward listener
	Turn-claiming suppression signal	Avert gaze contact from audience
	Within turn signal	Look toward audience
	No turn signal	Look away
Audience	Back channel signal	Look toward speaker
	Turn-claiming signal	Seek gaze contact from speaker
	Turn-suppression signal	Avert gaze contact from speaker
	Turn-claiming suppression signal	Looking toward to suppress speaking
	No response	Random

Table 2.5: *Turn-taking and associated gaze practices [GB06].*

The work in this thesis performs analysis of gaze behaviour during AMC and VMC. Therefore, it is essential to understand how gaze is utilised in the types of social scenarios that the experimental work, presented in Chapters 4–6, aim to capture. To summarise, some of the uses of gaze which are particularly relevant to this research are: indicating focus of attention, and likewise, determining the direction of the visual attention of others [AG77]; determining actions according to the gaze direction of those listening to you [Goo00]; addressing and prompting another speaker [Ler03]; when handling objects [Str96]; and proposing courses of action [Rae01]. The following section outlines the physiological properties of eye movement, before the two other major oculesic cues of eyelid movement and pupil dilation are presented.

Physiological Properties of Eye Movement

Each eye has six muscles that control its movements: the lateral rectus, the medial rectus, the inferior rectus, the superior rectus, the inferior oblique, and the superior oblique. This structure is illustrated in Figure 2.1. When the muscles exert different tensions, a torque is exerted on the eyeball, causing an almost pure rotation, with only about one millimetre of translation [Car88]. In addition, it is generally accepted that, under the influence of vestibular impulses, the eyes can rotate around the sagittal axis,

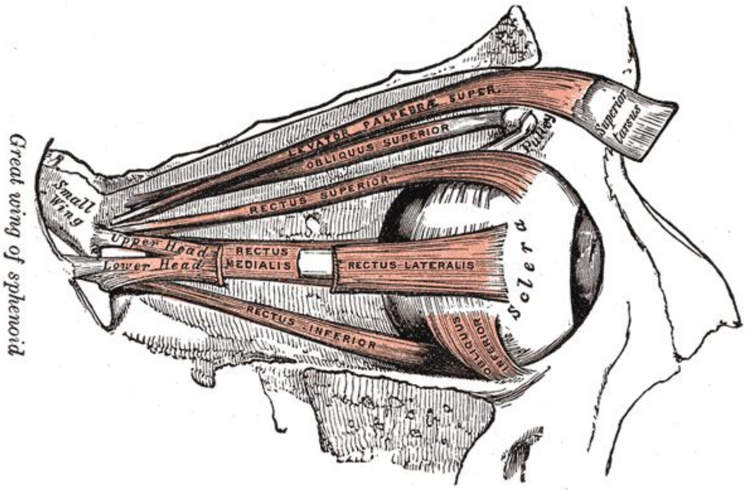


Figure 2.1: *External eye muscles illustration from Gray's Anatomy [Gra18].*

causing a compensatory "vestibulokinetic" reflex, with the purpose of stabilising the visual target on the retina during movements and changes of head posture [CEH75].

The major eye movement that the human eye performs are *saccades*. Saccades are rapid movements of the eyes from one fixational point to another, with velocities as high as 500° per second [Ray98a]. During saccadic motion, the eyes are moving so quickly across the stable visual stimulus that only a blur is perceived [US68]. There is still much debate as to whether cognitive processing activities are suspended during a saccade, but certainly, sensitivity to visual input is reduced: a phenomenon termed *saccadic suppression* [Mat74]. Between saccades, the eyes remain relatively still for durations of about 200-300 milliseconds (ms). These periods of reduced motion are termed *fixations*. Fixations stabilise the fovea (responsible for sharp central vision) over a stationary object of interest, and this is when virtually all visual input occurs [Mat74]. During fixations, miniature movements of no more than 5° occur. Such movements include *microsaccades*, *ocular drifts* and *microtremors* [Car88]. By continuously stimulating neurons, these movements serve to maintain visibility during vision.

A secondary form of voluntary eye movement is *smooth pursuit*, which serves to keep the image of a moving target centred on the fovea by matching the velocity of the eye's rotation with the target's velocity [KSJ00]. Smooth pursuits are a conjunctive eye motion, but only one eye needs to be focused on the target. However, in order to per perform a smooth pursuit, a target must exist, otherwise a series of small saccades will be generated [Kra04]. A sub-class of smooth pursuit motion is *vergence*, which refers to movement of the eyes in opposite directions toward a central, focal, point [Toa74]. Hence, as a target is brought near, the eyes turn inwards and as it is taken away they rotate outwards, but no more than parallel, in order to remain directed at the target.

In summary, humans use a combination of saccades and smooth pursuits, in addition to micro movements, to stabilise the retinal image of selected objects within the high-acuity region near the fovea [Kra04]. When determining the current gaze target, also referred to as point-of-regard (POR), visual attention as informed by the visual stimulus acts in combination with a contextual cognitive process describing voluntary intent to focus on a particular point [Duc03]. Thus, taking visual stimuli and cognition as input, the eyes will scan from the most important point to the least in repeated cycles when confronted with a scene, representation, object or human [PW96]. Finally, eye movement should also be considered in the broader context of bodily movement, and in particular, head movement. Horizontal gaze shifts greater than 25° and vertical shifts greater than 10° typically produce combined head and eye movement, and naturally occurring saccades rarely have a magnitude greater than 15° [KB01].

Pupil Dilation

Atropa belladonna, commonly known as deadly night-shade, is a perennial herb that has been valued for its medicinal properties for over five centuries. One of the first widespread uses of the herb was purely cosmetic, and sixteenth-century Italian women reportedly applied belladonna solutions to their eyes to dilate their pupils and achieve a dreamy, and supposedly more desirable appearance [Lee07]. Stass and Willis experimentally confirmed the efficacy of the beautification technique, measuring both male and female responses to photographs of people with and without artificially-dilated pupils [SW67]. Atropine, an alkaloid of belladonna that blocks the muscarinic receptors in the muscles of the eye, is still used by some ophthalmologists today to dilate the pupils for eye exams. Figure 2.2 illustrates dilated and constricted pupils.

Hess and Polt were among the first researchers to study the effect of visual stimuli on the size of people's pupils [HP60]. During the experiments, participants were shown erotic photographs of both males and females. Both males' and females' pupils dilated to images of the opposite sex. The authors concluded that pupil dilation was a reliable index of positive emotional arousal and interest, and also that pupil constriction was a sign of emotional aversion. While the former finding is undisputed and supported by subsequent work, the latter theory has been shown to be false, with evidence countering the idea that the pupils constrict with aversive stimuli: Partala and Surakka demonstrated that pupil size is significantly larger during both emotionally positive *and* negative stimuli than during neutral stimuli [PS03].

The communicative connotations of pupil size are substantial. Stass and Willis performed an experiment in which participants were asked to converse with two confederates, one of which had artificially-dilated pupils, and choose one of them to be their partner [SW67]. They found that the confederate with dilated pupils was chosen significantly more often. In a variation of Milgram's seminal obedience studies [Mil63], Kidd asked participants to administer shocks to a confederate [Kid75]. The experimental conditions varied oculesic behaviour of the confederate, with directed or averted gaze, and who had artificially-dilated or constricted pupils. The number of shocks given to the confederate was unaffected

Figure 2.2: *Patient suffering from tonic pupil. Dilated right pupil (left of image) displays minimal reaction to light, while normally-constricted left pupil according to ambient light level [oO10].*

by gaze, but participants witnessing confederates with dilated pupils administered fewer shocks and also gazed more at them during the experiment.

During vision, the size of our pupils is in constant flux. Along with emotional arousal, there are two other main effectors of pupil size: light and cognitive load. The light reflex occurs when the eye is exposed to light: the pupil constricts in order to regulate the amount of light admitted, and the brighter the light, the smaller the pupil becomes [LF42]. Although varying between individuals and decreasing with age [WWEP94], pupil diameter under extreme constriction is around 2 mm, and 8 mm when fully dilated [AC76]. Pupil constriction velocity is approximately three times faster than dilation velocity [Ell81]. Cognitive load, such as problem solving and memory tasks [KB66], and verbally answering questions [BO69] has been demonstrated to increase pupil size. In the latter study, pupil dilation was observed to be greatest just before individuals responded to a question, and pupil size increased linearly with question difficulty, indicating increased cognition.

In summary, pupil dilation is a subtle yet powerful signal in communication that is a reliable indicator of both psychological arousal and cognitive state of an individual. Observers are seen to respond appropriately and empathetically to variations in pupil size. A final effector of pupil size is eyelid closure, and in particular, blinking motion, which produces short constriction and redilation, assuming no other influencing factors, according to the ambient light.

Eyelid Kinematics

During vision, the human eyelids are in near-continuous motion, exhibiting what Evinger et al. term *lid saccades* [EMS91] and blinks. Lid saccades are bilateral shifts of the eyelids initiated by changes in the vertical rotation of the eyes, so that the eyelids follow the direction of gaze to ensure they do not obscure the pupils [Hal36]. Blinking is the act of extremely brief closure of the eyelids, and may be voluntary or involuntary [Gor51]. Physiologically, the eyelids serve to protect the eye from debris and rays of harmful intensity [Hal36], and also to regularly spread the tears and other secretion on the eye surface to keep it continually moist [Bri89]. Hence, while subtle to an observer, eyelid kinematics are a ubiquitous element of human behaviour. Figure 2.3 illustrates the anatomy of the human eyelid muscles.

Lid saccades describe the rising and falling of the eyelids with an associated upward or downward saccadic gaze motion [BF88]. Lid movements exhibit characteristic trajectories and amplitude-maximum velocity relationships. Upward lid saccades follow a smooth trajectory to final lid position, and the maximum velocity increases linearly with lid saccade amplitude (generally measured in degrees $^{\circ}$). During downward lid saccades, the maximum lid velocity is best described by a power function showing a soft saturation with increasing amplitude [EMS91]. On average, lid saccades start some 5 ms later than the concomitant eye saccades, but reach peak velocity at about the same time as the eye saccade [BF88]. Concurrent lid and eye saccades in the downward direction have similar amplitudes and velocities, but lid saccades in the upward direction are often smaller, and usually slower, than the concomitant eye saccades. However, the reverse appears to be true at large amplitudes. On average, downward saccadic lid movements have a shorter duration than upward lid movements at all amplitudes [EMS91].

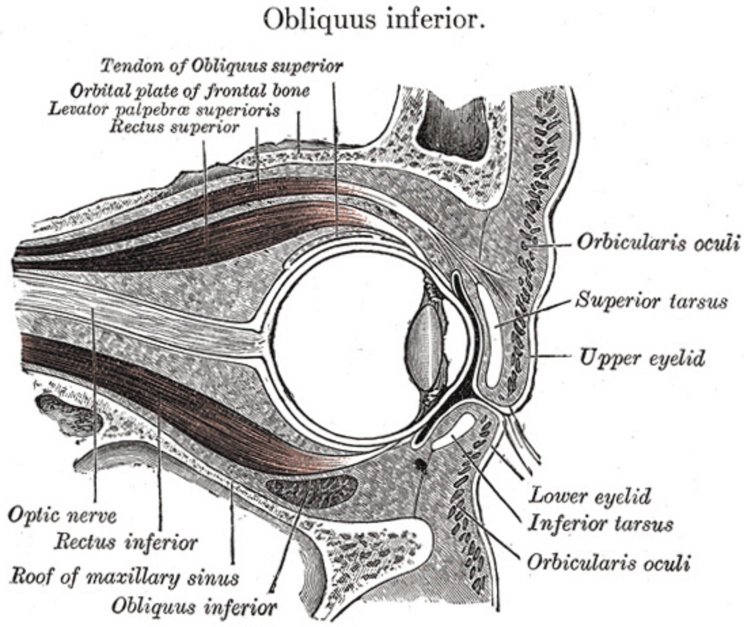


Figure 2.3: *External eyelid muscles illustration from Gray's Anatomy [Gra18].*

Downward lid saccades can exhibit an overshoot that takes the eyelid transiently lower than final lid position [KG64]. Overshoots occur more frequently at larger target amplitudes, but there appears to be no consistent relationship between the target amplitude and the magnitude of the transient overshoot or its appearance [WVdBS+95]. The eyes are also observed to overshoot during saccadic movement [VGRG81], but they are considerably less frequent than those in the lids, and transient overshoots of lid and eye saccades occur independently of one another [WVdBS+95]. Similar to the eyes, which often exhibit microsaccades during fixations, the eyelids also undergo small idiosyncratic movements of up to 5° when the eyes are stationary [BF88]. As lid saccades are coupled with vertical shifts of eye rotation, they also serve to highlight the communicative behaviour of gaze.

A blink is defined as the extremely brief closing and opening of the eyelids, and can be initiated by voluntary or involuntary stimuli [Gor51]. Involuntary reflex blinks are solely for the protection and efficient action of the eyes themselves [Hal45], whereas the onset of voluntary blinks may be caused by a number of factors such as current activity, psychological state, communicative intent, and personality [VBR+03]. However, regardless of their origin, all blinks exhibit a similar pattern: 10-12 ms after the onset of orbicularis oculi activity (the facial muscle that closes the upper eyelid), the lid rapidly lowers to close the eye. The superior tarsal muscle muscle then raises the upper eyelid more slowly to nearly its starting position. The maximum velocity observed during the down-phase and up-phase of a blink is a linear function of blink amplitude [EMS91]. This is defined by the current opened position of the eyelids, which amongst other factors, is due to the characteristics of the recent lid saccade.

Similarly to both gaze and pupil dilation, blinking behaviour is indicative of an individual's psychological state: including emotional mood and cognitive load. Bentivoglio et al. studied changes in how often people blink, termed *blink rate*, depending on their current activity. The authors established 17 blinks per minute as the average blink rate for a person at rest, an increased blink rate of 26 blinks per minute when engaged in conversation, and a dramatically decreased blink rate of 4.5 blinks per minute when reading [BBC+97]. Increased blinking during conversation correlates with communicative stimuli including speech, and response to others in an unfolding interaction, and is thus used as a nonverbal communicative signal. The reduced blink rate observed when reading is a strategy aiming to maintain the open-state of the visual input channel during reception of visual information [PK27]. Blinking has been shown to be influenced by cognition, with blink rate decreasing as cognitive demand increases, followed by a flurry of blinks after cognitive demand drops [HT72]. Regarding emotional state, Harris et al. found that blinks occur particularly during states of anxiety or tension, and decreased rate during concentrated thinking [HTS66]. As well as being an indicative of an individual's current activity and psychological state, eyelid kinematics are also revealing to more general conditions, such as slower blinks as an indicator of fatigue [SBS94].

2.1.3 Communication of Deception

The final section in the review of oculesics details deception, which is an ever-present element in human social interaction. Oculesic cues to deception are detailed, followed by an overview of how people detect deception in others. This section aims to provide the relevant background for the final telecommunication experiment in this thesis, presented in Chapter 6, which investigates truth telling and deception in AMC and VMC.

Deception, or lying, is a deliberate attempt to mislead others, and can be defined as an act intended to foster in another person a belief or understanding which the deceiver considers false [ZDR81]. Lying is a fact of everyday life, and it is estimated that people tell an average of one or two lies a day [DK98]. People's motivation for lying is varied, but the rewards that liars seek are typically psychological: to make themselves appear more sophisticated, to protect themselves from disapproval, and from getting their feelings hurt [DLM+03]. As DePaulo et al. state, the realm of lying is one in which identities are claimed and impressions are managed, and is not a world apart from non-deceptive discourse, where truth tellers edit their self-presentations, often in pursuit of similar goals [DLM+03]. Most lies are unremarkable, but some, such as betrayals of intimacy or trust or those told in influential situations such as a job interview, can have serious implications. For instance, estimates of costs due to applicant dishonesty and employee misconduct range from $6 billion to as much as $200 billion annually in the United States of America alone [BTG03].

Amongst the first investigations of gaze behaviour in deceptive situations was by Exline et al., who aimed to explore Machiavellianism (the employment of cunning and duplicity) [ETHG70]. Experimental participants initially worked with a confederate on a decision-making task. Part-way through the session, the confederate cheated by looking up the answers while the experimenter was out of the room, thus implicating the participant in deception. Later, the participant was interviewed by the experimenter and confronted with what had happened. The experimenter gazed continuously at the participant, and at the start of the interview, both low and high Machiavellian participants gazed back around 30% of the time. However, at the point when the accusation was made, both groups responded by looking much less, and this was especially true for the low Machiavellian participants. Even by the end of the interview, participant gaze had still not recovered to the original state, although high Machiavellian participants were now almost looking at the confederate as much as in the beginning.

Despite Exline et al.'s positive correlation between gaze and deceptive state, it still remains controversial as to whether gaze is a direct correlate of lying. In their meta-analysis collating results of previous studies, DePaulo et al. present a comprehensive catalogue of the extent to which 158 nonverbal and verbal cues occur during deceptive communication [DLM+03]. Perhaps the main finding of the work was that very few cues are specifically related to deception; pupil size is among the few. However, the review did demonstrate that certain psychological responses, including emotional arousal and high cognitive demand, are commonly generated by deceptive situations. The current work is concerned with oculesic cues of gaze, blinks, and pupil size that can be captured by eye tracking, and all of which have been revealed, in the prior section, to be influenced by an individual's psychological state. Of these cues, pupil size is perhaps the most revealing, at it generally acts as an automatic response to mental stimuli (and environmental luminance). In contrast gaze and blinking may be regulated to a greater degree, thereby potentially concealing a subject's psychological state [DeP92]. Both cognitive load [DGHP01], and emotional response to both positive and negative stimuli [PS03] are correlated with pupil dilation. A person's pupils dilate when they communicate deceptive messages [Hei76], and the amount of dilation has been shown to increase linearly with deceptions of greater magnitude [WSC06]. Gaze behaviour is not classified so clearly, and is highly dependent on the social scenario, idiosyncrasies, and culture. Indeed, studies have shown liars to gaze at their conversational partner both for less [KHD74] and more [BK76] time than truth tellers. Regarding blinks, Leal and Vrij demonstrated blink rate as an indicator of high cognitive load in a deceptive scenario in which participants were told to lie and tell the truth over a series of stages. Participants' blink rate decreased when lying, followed by a compensatory effect in the form of increased blinking directly after the lie was told and cognitive demand dropped [LV08].

Spence et al. demonstrated that lying is generally a more cognitively-demanding process than truth telling [SFH+01]. The Activation-Decision-Construction Model (ADCM) of lying developed by Walczyk et al. [WSC+05] proposes that people go through three cognitive stages when lying: activation (concerning the encoding of a question and retrieval of episodic and semantic memories), decision (the process whereby the person chooses to answer deceptively), and construction (the generation of a plausible lie). Using this model, response time has been shown to differ significantly between lying and truth telling. The final telecommunication experiment presented in Chapter 6 borrows the experimental question-answer framework described in by Walczyk et al. [WSC+05] to place participants in states of relatively high (lying) and low (truth telling) cognitive demand.

Detection of Deception

In many situations, people question the credibility or accuracy of statements that another person makes. The field of deception detection still harbours an active research community, as the process of detection relies on a complex set of interactions and cues. However, the general consensus among deception scholars is that people's ability to distinguish truths from lies tends to be around 57%, which is significantly, but only slightly, better than chance levels [DeP80]. An explanation for this poor performance is that 'detectives' seem to focus on the wrong behaviours when trying to distinguish lies from truths [ZKD81]. It has been shown that providing people with knowledge about cues to deception can improve detection accuracy [Vri94]. Also, the theory of truth bias or the "veracity effect", referring to peoples' tendency to judge more messages as truths than lies is an influential factor [LPM99].

In visual telecommunication, the richness of a medium defines its ability to transmit nonverbal and verbal cues, and in turn, defines both the potential for users to detect the veracity of fellow interactants, and to manage their own cues to deception. Hence, media communicating fewer channels becomes preferable for the deceiver [CGB+04]. For instance, high-quality VMC allows rich exchange of verbal and nonverbal cues, including facial expression, oculesics, and upper-body language, and it has been shown that peoples' ability to detect deception is statistically identical when assessing subjects via video or in reality [FUS97, HGSV04]. The degree of social presence engendered by a system is also key to the elicitation of behavioural cues in the first instance. Hence when interacting via VMC, it is a reasonable prediction that interacting via video will be more arousing and socially 'real' than interacting via virtual avatars [GSBS01]. Hence, cues to deception may not be elicited as prominently during the latter. However, while user representation is still crude compared to VMC, an AMC system that is able to faithfully represent users' oculesic behaviours, including gaze, blinking and pupil dilation has the potential to foster a high degree of social presence, and also to communicate behavioural cues that are influential in detection of deception. Such factors of visual telecommunication are explored in the forthcoming experimental chapters. The following section in this literature review contextualises the technical aspects of the research, detailing state-of-art in AMC and VMC systems.

2.2 Visual Telecommunication

The fact that face-to-face, collocated, contact is almost always the most satisfactory form of communication has been a fundamental constraint on society. Meier's theory argues that cities evolved primarily as a means of facilitating interpersonal communication [Mei62]. While this may be something of an over-simplification, there is wide agreement that the physical and social environments of modern man have been conditioned to a very large extent by his apparent need to travel to acquire information [SWC76].

A result of today's ecologically- and financially-conscious global society is an increasing reliance on telecommunication systems. A goal of such systems is to enable remote users to naturally interact to such a degree that, for certain situations, their use may be considered an alternative to actual travel. *Computer Supported Cooperative Work* (CSCW) is concerned with supporting multiple individuals working together with computer systems [Gre88]. It can be considered an umbrella-term incorporating application domains which have cooperative work at their core. This includes videoconferencing, public wikis employing collaborative authorship, and multi-user virtual reality applications. CSCW is born out of the need to address the following specific requirements of cooperative work: articulating cooperative work, sharing an information space, and adapting the technology to the organisation [SB92].

Johansen introduced the *CSCW Matrix*, which has become a popular tool in the community to categorise *groupware* systems according to degree of physical proximity of the group members and the degree of synchronicity [Joh88]. The definition of synchronous communication varies with context: Noel

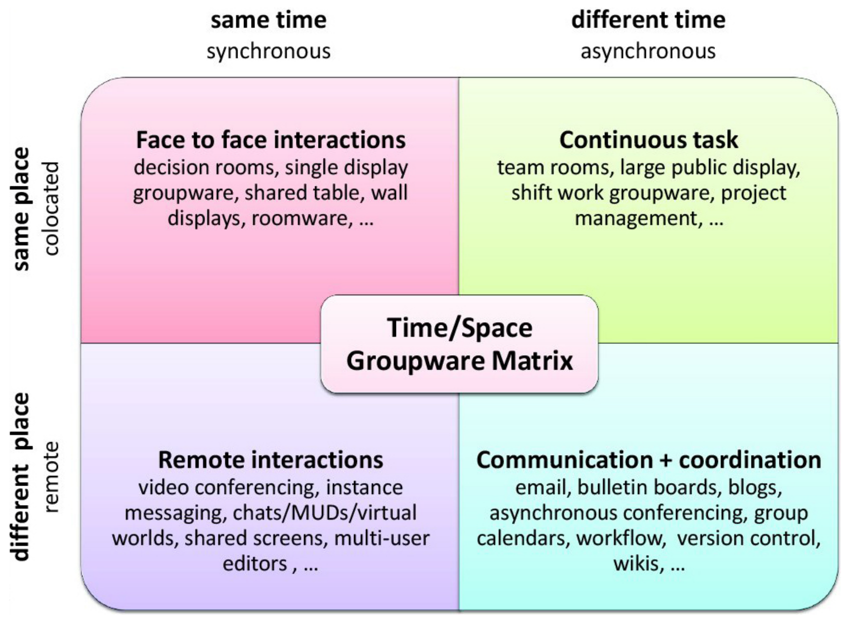


Figure 2.4: *Johansen's CSCW Matrix as represented in [SMIRM07].*

uses the term to denote *concurrent access* [NR04], while for Ellis, the term implies *real-time interaction* [EGR91]. The latter definition is appropriate when discussing visual telecommunication systems, which aim to support real-time interaction between two or more remote users. (In this context, asynchronous communication refers to users contributing to a collaborative task at different times.) Groupware represents a subset of CSCW systems, and is label for tools explicitly designed to support groups of people working together [IKA94]. The term may be used analogously to 'collaborative' software and its related hardware. A representation of the the CSCW matrix as it appeared in [SMIRM07] is illustrated in Figure 2.4.

While over twenty years old, and despite problems such as exponents belonging in multiple categories due to the strict categorisation, the CSCW matrix remains a useful tool for classification and indicating the types of systems within the groupware domain. The category that this thesis is primarily concerned with is *remote interactions*, which are both synchronous and remote. This category includes ICVEs (a subset of "virtual worlds" as denoted in the matrix) that support AMC, VMC (denoted "video conferencing") and hybrid solutions including shared screens. These systems are similar in that the mediated communication they support features a visual component displaying real-time representations of fellow, remote, interactants. These representations are critical in order to support natural communication of nonverbal behaviour between remote users, and their form adopts specific paradigms depending on the medium. The following sections detail VMC and ICVEs, which are the two mediums of visual telecommunication that the work in this thesis is concerned.

2.2.1 Video-Mediated Communication

VMC is the most accessible and established form of audio-visual remote interaction for dyadic and small-group communication. It has been shown to improve (over audio-only communication), users' ability to show understanding, forecast responses, give nonverbal information, enhance verbal descriptions and manage pauses and express attitudes [IT94]. The first video communication system, developed by Bell Laboratories, demonstrated the use of closed-circuit television to transmit video between New York and Washington in 1927 [Ive27]. Through subsequent research and development, several technological tiers of VMC systems have emerged, from common webcam chat and video-calling on mobile devices, through to high-end commercial "telepresence" solutions (e.g. Cisco, LifeSize, Polycom, Telris, BrightCom, Telanetix) supporting high-definition (HD) video and audio streams.

The word "telepresence" has taken on a multitude of meanings, which are often adapted to a particular system or application. The term originates from Minsky who, in 1980, introduced the concept of telepresence to describe the feelings that a human operator might experience when interacting through a teleoperator system in which the operator sees through the eyes of the remote machine, and uses their own limbs to manipulate its effectors [Min80]. As elucidated in the following section on ICVEs, this idea seems closer to VEs, in which a sense of being in a different place might develop, and the user's tracked natural body movements having an immediate impact within the system. This state is also considered to be conducive to effective task performance in the remote environment. Hence, to use "telepresence" to describe devices which are ostensibly state-of-art video conferencing systems may seem a rather curious misuse of the term. Figure 2.5 presents a depiction of Cisco's *TelePresence* system in use.

Such commercial systems are able to provide a richer collaborative experience than common desktop videoconferencing. Such capabilities include a wide field of view of remote users' environments, life-sized video representations of users, and critically, *gaze awareness*, which this section will now focus on. Faithful representation of user gaze has long been recognised as a requirement to enable natural communication in visual remote collaboration and conferencing systems [AL87]. Comparing small-group communication in VMC systems supporting gaze awareness with those that do not support it, Isaacs and Tang observed that people using the latter needed to address each other through explicit use of each other's names, and control the turn-taking process by this means in the latter [IT94]. Contrastingly, in collocated and gaze aware interaction, they observed many instances when people used gaze to indicate whom they were addressing and to suggest a next speaker. It was frequently observed that when more than one person started speaking at the same time, the next speaker was determined by the gaze of

Figure 2.5: *Cisco's* TelePresence *system, supporting gaze awareness, and life-size representations of users in a visually-seamless workspace [Sys10b].*

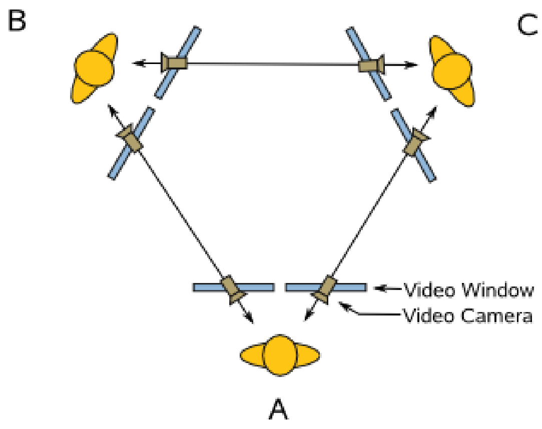


Figure 2.6: *Aligning cameras and video displays to users natural line of sight in VMC to achieve gaze awareness.*

the previous speaker without the need for conventions or explicit verbal intervention. Gaze awareness implies that direction of gaze is preserved, so users can signal and observe eye movement in a natural manner, thus enabling the channel to be used similarly to collocated interaction. Hence, the cameras and displays of a gaze aware VMC system must be positioned so that users can rely on natural sight cues to facilitate gaze interactions with remote users [CDO+00a]. An illustration of this technique in a three-party VMC interaction is illustrated in Figure 2.6.

Developed by Okada et al., the design of the MAJIC system defines the standard method for supporting communicational gaze over VMC, in which relative position, head orientation, and gaze are all preserved during multiparty meetings [OMIM94]. In MAJIC, the cameras are placed behind a thin half-transparent curved screen, on which life-size videos of participants are seamlessly projected to simulate a round-table meeting. Gaze awareness is achieved by positioning cameras behind the screen at the location of each participant's head. Hence, when two remote users gaze at each other's video representation, they both perceive mutual gaze to have been established. Recalling Argyle's conviction that gaze is of central importance in social behaviour and NVC [AC76], the importance of gaze awareness is a natural extension, enabling the bidirectional channel to be used to monitor, initiate, maintain, and terminate messages. Correspondingly, collaborative problem solving performed via gaze aware VMC systems has been shown to improve both performance and sense of collocation over VMC not supporting the cue [IKG93].

It has only been relatively recently, in 2002, that the camera-to-display offset-angle threshold in order to achieve gaze awareness in VMC was established [Che02]. Chen's initial experiment, measuring how accurately people perceive mutual gaze, determined that sensitivity to direction of gaze is asymmetric. Results indicate that people are an order of magnitude less sensitive to perceiving mutual gaze when the person being observed is gazing at a location below the eyes of the observer compared with gazing to the left, right, or above their eyes [Che02]. Figure 2.7 presents the experimental results. For the up, left and right cases, the camera-to-display offset threshold is only 1° away from the camera before perception of eye contact is lost, while for the down case, observers are much less sensitive to POR, with

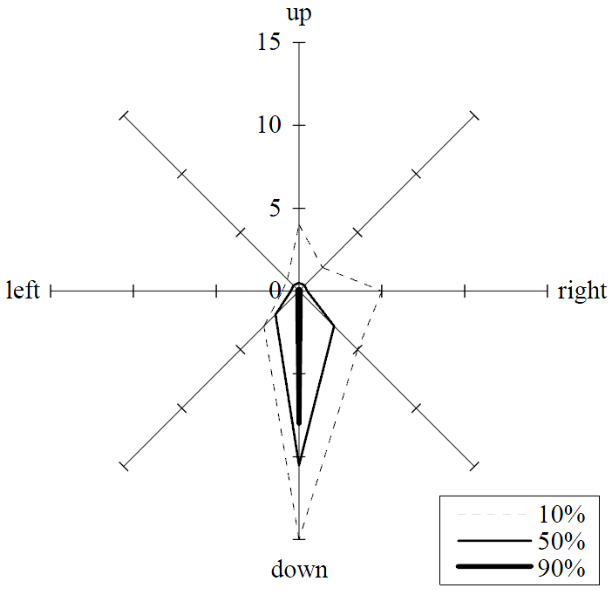


Figure 2.7: *Sensitivity to gaze direction as presented by Chen [Che02]. The contour curves mark how far away, in degrees of visual angle, the looker could look above, below, to the left, and to the right of the camera without losing perception of eye contact. The three curves indicate where eye contact was maintained more than 10%, 50%, and 90% of the time. The percentiles are the average of sixteen observers. The camera is at the graph origin.*

90% of participants still perceiving mutual gaze at an offset of 8°.

Chen suggests that the anatomical characteristics of human eyelids may explain why prediction of gaze direction is asymmetric: when people rotate their eyes to the left or right, their eyeballs rotate within the socket, causing a noticeable change in the position of the iris within the sclera (the whites of the eyes). This is also the case for when people look up, as the upper eyelid performs lid saccades, thus rising accordingly to maintaining the iris' position relative to the eyelid, while the lower eyelids remain approximately stationary. However, when people look down, the upper eyelids perform downward lid saccades, increasing the closed state of the eye. Thus, there is not a very noticeable change to the position of the iris with respect to the sclera, which is also less visible. Deriving from the results, Chen offers the *snap to contact* theory, suggesting that, when people cannot judge gaze direction accurately, they will bias their perception toward mutual gaze unless they are certain that it is not the case [Che02]. In VMC, when people look below the camera, the resulting change in their appearance to a remote user is less pronounced than if they look in other directions (horizontal or upwards), and therefore, more mutual gaze will be perceived in the down direction. Finally, Chen also notes that the limited resolution of VMC displays serve to make gaze direction estimation more difficult, which increases the bias toward assuming eye contact. In summary, Chen's work is critical to the design of state-of-art VMC systems, providing physical parameters on which to base design specifications. Chapter 7 details an experiment which measures peoples' ability to estimate the POR of a virtual humanoid, uncovering similar results to Chen [Che02].

While enabling the nonverbal channel of gaze is clearly beneficial to the quality of communication in visual remote interaction, there are varying degrees of the fidelity of gaze awareness preservation which must be considered. This notion was introduced by Ishii and Kobayahi with their *ClearBoard-2* system, for which the key metaphor "talking through and drawing on a transparent glass window" was devised [IK92]. Full, high-fidelity gaze awareness implies that users know what their partner is looking at, including their face or anything else on the shared workspace. This preservation of gaze allows the channel to be used similarly to collocated communication, and supports awareness of mutual gaze and intuitive identification of objects in the workspace. Incorporating full gaze awareness into VMC systems for collaborative problem solving has been shown to improve performance, as success depends heavily on the perspectives of the users, and thus was deemed highly advantageous for others to be aware of their partner's gaze [IKG93]. Depending on the technical setup, some VMC systems may only support partial gaze awareness, which is defined as the ability to determine the general direction someone is looking, for instance, up or down, left or right [MG02]. Low fidelities of gaze awareness may preclude the ability to distinguish gaze directed at objects in a scene, and allow mutual gaze to be established with specific users only. Hence, the properties and requirements of a specific interaction are significant when determining the operating fidelity of gaze awareness. For instance, while the *video tunnel* system, presented by Buxton and Moran [BM90], provides the ability to maintain mutual gaze in two-party conversations, when the scenario is complicated by a third user or another feed for a shared document, gaze awareness is only partially supported. Hence, the intended usage of a VMC system must act as a parameter towards determining that system's support for gaze awareness.

Limitations of Video-Mediated Communication

VMC can provide a rich mode of visual remote interaction, in which users can see and hear each other in real-time, and communicate using both verbal and nonverbal cues such as speech, gaze and facial expression. However, an inherent disadvantage of VMC is the compressed representation of 3D space, which constrains rich cues available in collocated collaboration such as depth, resolution, and field of view, thereby limiting awareness and ability to look and point at users and the environment [HRBC06]. In the regard of spatiality, VMC has proven to be more similar to audio-only conferencing than to unmediated face-to-face collaboration [Wil77]. A major drawback even of state-of-art VMC systems is the inability for users to move in their environment, or significantly shift posture in their chairs, and still remain in-frame and thus visible to remote users. Gaze awareness is also not robust, and parallax between the camera position and video display of a user, due to a posture shift for instance, may result in the loss of meaningful gaze information. Despite technical advances and novel system design, these limitations have remained an element of VMC.

There have been some notable examples of novel VMC systems which attempt to lessen the medium's spatial limitations. Such systems often involve elaborate arrangements including movable cameras ([NMK09]), arrays of separate displays ([SBA92]), and hybrid VMC-VE systems operating in immersive ([GWN$^+$03]) and desktop ([VWSC03]) displays. Gross et al.'s immersive *Blue-C* system will be addressed in the following section reviewing ICVEs. A system that is of particular interest here is Vertegaal et al.'s second iteration of the *GAZE Groupware* system [Ver99], *GAZE-2* [VWSC03]. GAZE-2 takes a hybrid VMC-VE approach to support collaborative scenarios in which gaze, together with some sense of spatiality, is preserved. The system uses desk-mounted eye tracking to control three (centre, left and right) mounted cameras, thus providing parallax-free gaze awareness in group conversations [VWSC03]. Multiple cameras are placed in video tunnels behind the on-screen video displays of users, similarly to the MAJIC system. Based on gaze direction captured by the eye tracker at each site, the system automatically alternates the currently active camera, streaming this data to the other sites. Video displays of users are rendered on billboards in a shared VE. The billboards then rotate towards or away from other users' billboards, depending on where each user is looking, thereby conveying attention. Despite angular shifts caused by camera alternation in the transmitted video, the system is able to both preserve a high-fidelity of gaze awareness, and positions users in a shared, albeit non-immersive, space.

Using GAZE-2's novel approach, VMC-based systems may overcome some of common spatial limitations of the medium. However, users must still interact via flat two-dimensional (2D) 'windows' into each other's environment, rather than sharing a perceptually unified 3D space. Restriction of movement and NVC that relies on spatiality remains the critical limitation of VMC, and is contrary to collocated interaction in the real world, in which the surrounding environment, and the use of its space, is a ubiquitous feature. The growing interest in preserving spatiality in telecommunication systems is represented in the CSCW research community, and has led to increased development of immersive VR systems [BBH+90], which are by definition, able to circumvent this limitation of VMC.

2.2.2 Immersive Collaborative Virtual Environments

ICVEs provide a platform in which to conduct AMC, connecting remote or collocated users of immersive projection technology systems such as the CAVE within a spatial, social and informational context, with the aim of supporting high-quality interaction [RWOS03]. A typical ICVE system will rely on a combination of specialist hardware and software to deliver a real-time graphical simulation, which is responsive to users' tracked body movement and behaviour, and operates simultaneously and consistently at multiple sites via a distributed database. The key concept behind ICVEs is that they are shared virtual worlds: computer generated spaces whose occupants are represented to one another in graphical form, and can control their own viewpoints and can interact with each other and with various representations of data and computer programs [BBRG96]. User embodiment and avatars will be discussed in the Section 2.3. The current section presents an overview of key features of ICVEs.

Immersion

Barfield and Furness define a VE as a representation of a computer model or database which can be interactively experienced and manipulated by the users [BFI95]. This definition is broad, and could be used to describe everything from a home video game through to a multi-user CAVE-based VR application. The critical difference between these two examples is each system's level of *immersion*, which defined as the degree to which a user's sensory input channels to all modalities are stimulated by the VE interface [DKU98]. Thus, the distinction between an immersive and non-immersive VE depends on the capabilities of the operating hardware, particularly regarding displays and input devices.

Steed notes that the key differences between immersive and non-immersive VEs are questions of both interaction style and user embodiment within the environment [Ste96]. In an immersive system, the perspective of the graphical display is coupled to the user's head, through use of head tracking, so that the rendered displays match as closely as possible the changes that would be expected in reality when the same motions are made. Correspondingly, the movement and behaviour of a user's embodiment in an immersive system may be coupled to bodily tracking devices, forming a natural interaction metaphor that maintains the user's sense of proprioception in the surrounding environment. As discussed in the following section, this control metaphor becomes critical in a multi-user scenario, as user representation acts directly as a communication mediator [CWS99]. In contrast, the lack of tracking and non-surrounding displays in a non-immersive system means that neither the rendering viewpoint or the behaviour of a user's embodiment are coupled to the user's motion, but rather are controlled indirectly through standard input devices. Slater summarises the difference between an immersive and a non-immersive system by stating that, in an immersive system, it is possible in principle to fully simulate what it is like to use a non-immersive system, but not vice versa [Sla09].

A VR system is defined by Cruz-Neira et al. as one which provides real-time viewer-centred head-tracking perspective with a large angle of view, interactive control, and binocular (stereoscopic) graphics [CNSD93]. This positions the CAVE at the apex of what constitutes a VR system, while a standard desktop computer would not be encompassed by the definition, despite its likely ability to run the same VE software. A critical point here is that both hardware and software components must be considered when 'ICVE systems' are discussed, and that the two components are may only be loosely-coupled. This implies that two remote users (one located in a CAVE, and one using a mobile device) can conceivably be connected via the same VE software, but interact according to the individual characteristics of their hardware device: full-body tracking may feature in the CAVE and a touch-screen input metaphor may operate on the mobile device. Such an interaction would be therefore be asymmetric: the CAVE user would class the system as an ICVE, while the mobile user would perceive the communication to be taking place using a CVE. Chapter 3 presents the platform developed and used throughout the research in this thesis, *EyeCVE*, which operates on a variety of technologies, including fully immersive, semi-immersive, and standard desktop hardware. Figure 2.8 shows two users engaged in AMC supported by EyeCVE operating between a semi-immersive VR system, the WALL, and a fully immersive CAVE system.

The impact of such asymmetric interaction has been investigated in Schroeder et al.'s [SSA+01] and Roberts et al.'s [RWOS03] work on object-focused collaboration. Object-focused tasks in multi-user VEs involve the manipulation of virtual objects, for example to solve a logic puzzle or organise the virtual elements into a certain arrangement [HFH+00]. This process is analogous to product engineering design in the real world, but also provides a paradigm for visual programming techniques [OAS+02] and construction of new VEs [Ste96]. In an experimental scenario, the collaborative elements of the task are emphasised, so that progress toward the goal-state is either dependent on, or significantly aided by, the participants working together. Schroeder et al. [SSA+01] compared pairs of participants collaborating between networked CAVE systems (symmetric) with collaboration between a CAVE and a desktop com-

Figure 2.8: *EyeCVE users engaged in AMC. The local user is located in the semi-immersive WALL system, featuring perspective rendering on a single projection wall. The cointeractant, represented as an avatar, is located in a fully immersive CAVE system featuring perspective stereo rendering within four surrounding display walls.*

Figure 2.9: *Screen captures provided courtesy of Ralph Schroeder and Dave Roberts.* Left: *Two participants collaborating to solve a simplified Rubik's cube puzzle as presented in [SSA+01].* Right: *Three participants working together to build a virtual gazebo as documented in [RWOS03].*

puter (asymmetric), and also with the same task performed in the real world. Taking puzzle-completion time as the primary metric of task performance, participants took longer in the CAVE-to-desktop condition than the CAVE-to-CAVE and collocated conditions, between which no significant difference was found. These results are echoed in Roberts et al.'s work on constructing a virtual gazebo, where participants working together in CAVE systems were able to complete the task in significantly less time than those who performed the task between asymmetric systems, or using non-immersive systems [RWOS03].

Figure 2.9 shows screen captures depicting the collaborative interaction scenarios investigated in two experiments, reported in [SSA+01] and [RWOS03]. Both studies revealed interesting findings regarding collaboration and group dynamics. When asked, post-collaboration, to judge their contribution to the task, desktop users evaluated their share as less than that of CAVE users. In the Roberts et al. study, users of immersive systems were considered by all (themselves and desktop users) to contribute more than desktop users. Furthermore, where a team comprised of two immersed and one desktop user, the latter was left out of most of the activity. Similar results are observed in [SSUS00] and [WAS+00], in which users reported that they contributed unequally (favouring immersed users) despite being unaware of what type of system their partner was using. Another implication of a system's level of immersion is on leadership, which is seen to be dynamically, and automatically, assigned to the more immersed user during a collaborative session [SSS+99, SSUS00]. Thus, similarly to the different forms of VMC systems, which support varying quality of communication, there also exist several technological tiers of multi-user VE systems, of which, degree of immersion is the primary variable.

Presence

The central feature of users' response to VR is *presence*, which may also referred to as *place illusion* (PI) [Sla09]. Immersion describes a system's hardware, and also provides the boundaries within which presence can occur. Presence is best defined as a user's psychological response to patterns of sensory stimuli, resulting in the user having the impression of "being there", in a computer-generated space [SUS94]. Slater explains that immersive VR systems can be characterised by the sensorimotor contingencies (SCs) that they support, referring to the actions that a user can carry out in order to perceive the VE [Sla09]. For example, moving one's head and eyes to change gaze direction, or bending down in order to see underneath something are SCs of a typical immersive VR system. The set of SCs supported by a system are analogous to the system's level of immersion, and define the set of valid actions that are meaningful in terms of perception within the VE. When immersed within a VE, sense of presence maintained through SCs that provide synchronous correlations between the act of moving and the concomitant changes in the images that form perception. For instance, Pan describes an experiment in which a virtual representation of a woman interacts with participants through verbal and nonverbal behavioural cues [PS07]. Due to the head tracking that featured in the CAVE system in which participants performed the experiment, when participants moved, the image of the virtual woman in their visual field updated as it would be expected in reality. This is an example of a SC. During those experimental sessions, all participants found themselves automatically responding to this illusory woman, by talking and behaving comparably to as they would be likely to in reality. When immersed in a VE, such personal and plausible feedback is critical in maintaining the illusion of reality. Anecdotal evidence from post-experimental interviews strengthen evidence of this automatic response of presence: "The idea of cheating on my partner with this virtual woman caused a real physical and emotional response; this was the strongest and most surprising aspect of the experience" [PS07].

There is no single method for measuring a user's sense of presence. Perhaps the most common approach is to use questionnaires which aim to elicit subjective responses regarding the VE experience. This was first demonstrated in [SU93], and a refined questionnaire later appeared in [SW97]. The questionnaires generally feature ordinal Likert scales that anchor responses between two extremes, and have been shown to be effective in eliciting meaningful responses in many cases. However, there are a number of criticisms to questionnaire-based measurement of presence: they been shown to be unstable, in that prior information can influence the results [FAPI99], they may be unable to discriminate between presence in a VE and physical reality [UCAS00], and they may be prone to methodological circularity (asking questions about PI may foster the very phenomenon that the questionnaire is supposed to be measuring) [Sla04].

More robust and compelling measures of presence involve analysis of behavioural response to the VE stimuli. If participants in a VE behave as if they are in an equivalent real environment, this is a sign that they are experiencing presence. For example, the *pit room* is a classic VE which has evolved through several iterations to assess aspects of presence [UAW+99, Ins01, MIWBJ02]. The experiment recalls Walk and Gibson's classic studies investigating responses to a 'visual cliff' [GW60]. The visual cliff consists of a narrow platform supported by vertical sides that drop a few inches to a large plate of glass. Both human infants, and rats [WGT57] have been shown to demonstrate avoidance behaviour when confronted with the visual cliff. The VE that forms the stimuli for the presence experiments consists of two adjacent rooms: the starting room is unremarkable, populated with a few chairs and blocks, and has the purpose of familiarising the user with the use of the VR system and learning procedures for interaction and navigation. Upon walking into the second room, the user is confronted with a three-metre drop surrounded by a narrow walkway. When given the task of navigating to the opposite side of this room, almost all individuals do so by carefully edging themselves around the ledge of the room, even though they know that there is no real danger. In an experimental scenario, a participants reactions to the pit have been quantified through physiological measurements such as heart rate, respiratory rate, and galvanic skin response. When compared to the baseline level of the first room, people experience significantly heightened physiological response. Meehan and Razzaque investigated the influence of passive haptics on response to the pit, in which real (but small) ledges were aligned to their virtual counterparts [MIWBJ02]. This added to the effect of standing over a real pit, and significantly increased the heart rate of participants compared with when the ledge was absent. Thus, if the normal physiological response of a person to a particular situation is observed in a VE, this is a sign that the user is experiencing a high level of presence. The obvious drawback of such measurement methods is that they are limited to situations in which there is a significant physiological response in reality, so are less useful in mundane situations [SVS05].

The issue of *social presence*, also referred to as *copresence* arises when discussing multi-user VEs. Copresence is highly relevant to user representation and embodiment, and therefore is discussed in the following section on avatars. Spatiality, a feature of ICVEs that renders them particularly germane to natural remote social interaction forms the final discussed in the current section.

Spatiality

In an early evaluation of teamworking in AMC supported by non-immersive CVEs, Hindmarsh et al. suggested that some of the limitations to natural collaboration that had been observed in such non-immersive systems could be alleviated by their immersive counterparts [HFH+98]. A subsequent study by the same authors confirmed the predicted benefit of intuitive bodily tracking (head and hand) and surrounding nature of the displays, that provided users with a field-of-view similar to that which is available in natural perception in reality, to enable more natural interaction [HFH+00]. Investigating

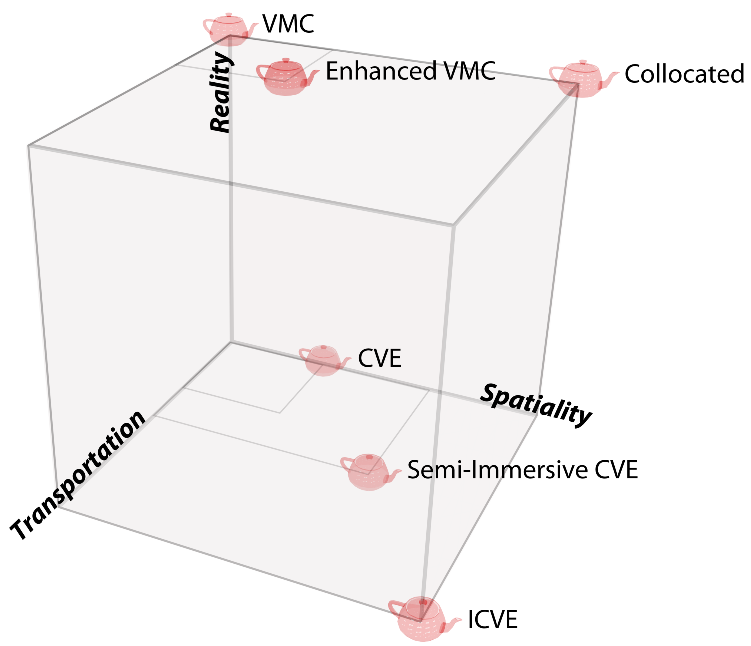


Figure 2.10: *Comparison between visual telecommunication systems in terms of Benford et al.'s three dimensions of spatiality [BBRG96]. Note that to ensure consistent polarity of the axes, artificiality has been positively renamed as reality. System types include desktop VMC, "telepresence" VMC (e.g. MAJIC, telepresence systems), non-immersive CVEs, semi-immersive CVEs, and ICVEs. Face-to-face collocated interaction is also included.*

various collaborative tasks with the same participants over extended periods of time (a situation that is closer to the practice of actual collaborative work), Steed et al. conclude that interaction in ICVEs is intuitive, and particularly suited to highly spatial and interactive tasks [$\text{SSH}^{+}03$].

Benford et al. suggest that the increased interest in spatial approaches to CSCW might be viewed as a shift of focus towards supporting the context within which work takes place, rather than the process of the work itself [BBRG96]. The authors suggest two dimensions representing the fundamental properties of spatiality by which visual telecommunication systems may be classified: *transportation* is analogous to presence, and is concerned with the extent to which users perceive that they have left behind their local space and have entered into some new remote space, while *artificiality* concerns the extent to which the space is either synthetic or is based on the physical world [BBRG96]. In addition, the authors state that *spatiality* is the primary dimension on which system types may also positioned, and is the degree to which a system supports key spatial properties such as containment, topology, distance, movement, and a shared frame of reference. Figure 2.10 positions desktop VMC, "telepresence" VMC, non-immersive CVEs, semi-immersive CVEs, and ICVEs along Benford et al.'s [BBRG96] spatial dimensions.

As indicated by the high dimension of reality in Figure 2.10, VMC systems are likely to remain superior in terms of presenting the truthful appearance of fellow interactants in their remote environments. However, even state-of-art VMC struggles to provide high levels of immersion and presence, as denoted by the low transportation dimension, in a perceptually shared space, denoted by the slow spatiality dimension. High dimensions of spatiality and transportation are inherent to highly-immersive VR systems, enclosing users in wholly synthetic environments at the price of faithful replication of reality.

This implies that, rather than observing others via flat 2D windows, ICVE users are able to share views of a unified and shared workspace, consisting of the same objects and artefacts, and populated with other users' avatar embodiments. These embodiments grant access both to other users and to their actions and orientations in the shared VE. Therefore, unlike VMC, ICVE users are visibly 'embodied in', and 'connected to', the common environment and objects within it [HFH+98].

Garau exchanged Benford et al.'s dimension of artificiality with the synonym of *fidelity*, arguing that, in the context of interaction in ICVEs, realism hinges on a system's capacity to portray a convincing context and process for collaboration [Gar03]. This definition is instructive with regards to the technological evolution of ICVEs, implying that operational practicality should be the focus of the medium's development, with the goal of maximising quality of communication. Thus, ICVEs are not restricted along the dimension of fidelity as VMC is along the dimension of spatiality. The fidelity of user appearance and behaviour has been the focus of much research in the AMC literature. This focus is a logical induction from the importance of our own bodies in collocated social interaction, where the full gamut of communicational cues by which to exchange knowledge is available. The correlate of our own bodies in the virtual domain is an avatar, with which the research in this thesis, and the following section, is concerned.

2.3 Avatars

Before the central topic of avatars is explored, the import of the two previous sections covering human nonverbal communication and visual telecommunication systems on the current work will be stated.

Section 2.1 covered a range of nonverbal cues that are used during collocated social interaction. These cues include facial expression, gesture, posture, proxemics, and chronemics. Oculesic cues, including gaze, eyelid movements, and pupil dilation were focused on in particular, as the impact of their transmission in AMC is the focus of this research.

Focusing on VMC and ICVEs, Section 2.2 detailed approaches to CSCW that aim to support synchronous visual telecommunication between remote users. Despite development of highly sophisticated systems, there is no substitute for the richness afforded by collocated meetings in reality. Benford et al. suggest the three dimensions of transportation, artificiality, and spatiality to define characteristics of telecommunication systems [BBRG96]. Illustrating the relative strengths and weaknesses of various system types along these dimensions, Figure 2.10 exposed deficiencies in VMC with regards to representation of 3D space which limit the ability for interactants to be aware and make use of the shared environment.

ICVEs present an alternative paradigm to VMC, in which users are transported to a perceptually-unified shared virtual space, and represented by graphical avatar embodiments. Schroeder considers immersive technology, as an "end state", in the sense that synthetic and multi-user environments that are displayed to users senses cannot be developed further than fully immersive systems [Sch06]. ICVEs are able to overcome some of the limitations of VMC and non-immersive CVEs with regards to framing interaction in a navigable and 3D environment. However, in comparison with a VMC, which transmits a users' actual appearance and behaviour, avatars exhibit more primitive fidelity: particularly with regard to faithful replication of users' nonverbal behaviour. This lack of capture and display of many NVC channels is considered a critical hindrance to the support of high-quality interpersonal AMC in ICVEs [Gar03]. In particular, gaze is identified as an important cue not captured or represented faithfully by the medium [HSS+05].

Studying collocated interaction, Kendon established that shifts of gaze are systematically coordinated with the timing of speech [Ken67]. Subsequently, Argyle et al. found that managing and synchronising verbal negotiation was made more difficult if the interactants' eyes were concealed by dark glasses [ALC68]. Analogous observation in AMC was reported by Steed et al., in which participants found negotiation tasks difficult due to the absence of avatar eye movement, which hindered the ability to comprehend the intentions and activities of others [SSH+03]. Hence, in order to support collaborative interaction in ICVEs that is more similar to non-mediated, collocated scenarios, the capture and representation of nonverbal behaviour is considered to be essential [SRSH05].

The work in this thesis investigates immersive AMC in which oculesics are captured and transmitted. The following section defines the state-of-art in AMC, framing the central research problem. The theory of *social agency* is firstly introduced, followed by approaches to user embodiment and tracking in ICVEs. The theories of *social presence* (often termed *copresence* in the VE literature [Gar03]) and *media richness* are then detailed. Work defining aspects of avatar fidelity, and the impact of nonverbal behaviour exhibited by virtual humans in VEs is discussed, introducing the concept of 'transformed social interaction' that arises in AMC. Finally, the current paradigms of animating the oculesic behaviour of avatars is detailed, including a review of gaze modelling (the predominant method of gaze control), and methods of animating eyelid kinematics and pupil dilation.

2.3.1 Social Agency

In their meta-review of social responses to computers, Nass and Moon state: "individuals mindlessly apply social rules and expectations to computers" [NM00]. People generally require minimal encouragement to view computer systems and applications as *social agents*, reading far more understanding than is warranted from symbols and graphical displays [Hof95]. This was unexpectedly observed, and first documented, by Weizenbaum, when performing user studies with ELIZA: a computer program for the study of natural language communication between man and machine [Wei66]. During the purely text-based interactions between participants and the system, ELIZA simulated a Rogerian psychotherapist, by rephrasing input statements from the user, and returning them as questions. Weizenbaum observed many examples of people becoming emotionally engaged when "communicating" with ELIZA, and some even asked to be left alone with the system. ELIZA and contemporary conversational-agents, often referred to as *chatterbots*, take advantage of the tendency for humans to unconsciously equate programmed computer behaviour as analogous to conscious human behaviour: despite conscious knowledge to the contrary. This phenomenon has become known as the *ELIZA effect* [Bod97], and may be considered a precursor to many observations of immersion and presence reported in the VE literature covered in Section 2.2.2.

In the specific context of software-based virtual humanoids, *agency* describes their method of control or interaction, with *avatars* and *agents* occupying either end of the agency-spectrum [NB03]. The actions of an agent are dictated entirely by software, while an avatar's actions are controlled by a human in real-time. Definitions that acknowledge the subjective nature of the VE experience are provided by Blascovich [Bla02] and Schroeder [Sch02], who similarly state that agency is the extent to which a virtual human is perceived by individuals to be a representation of other individuals in the physical world. Agency can be conceived as a continuum, with agents perceived to be entirely autonomous at one end, and avatars perceived to be completely controlled by human means at the other end. Hybrid control schemes in which an avatar is perceived as being partially autonomous and partially user-controlled lie between these ends.

In the VE literature, it is common to find examples of virtual humans which are avatar-agent hybrids. For instance, Garau et al.'s study on gaze modelling used participants' tracked body movement and posture to drive their embodiments, while gaze behaviour was controlled by a simulation [GSBS01, VGSS04]. Similarly, Bailenson et al.'s work on "augmented gaze" altered the attentional cues of head orientation and gaze in order to influence perceived direction of interest [BBL^{+}05]. Experimental work in this thesis primarily concerns virtual humans that approach pure avatars on the agency scale, and follows Bailenson and Blascovich's use of the term "avatar" to refer to a virtual human that is controlled partially or entirely by a user, and "agent" to refer to a virtual human that behaves entirely by simulated autonomous means [BB04], for instance as demonstrated in [VSS05]. Section 2.4 argues that, for the application of telecommunication, behavioural simulation is inappropriate, and has the potential to alter an unfolding communication. However, modelling may be cautiously considered when tracking of a kinesic cue is impractical, or may not deliver significant benefit. This is explored in Section 7.2, which investigates the modelling of eyelid kinematics.

2.3.2 User Embodiment and Tracking

In shared VEs, users' avatar embodiments act as the fundamental mediators of the visual component of an interaction [CWS99]. Hence, avatars function both to identify users and to communicate nonverbal behaviour including position, identification, focus of attention, gesture and action [Tha99]. Avatars generally exhibit generic humanoid form, which reflects their status as a representation of a human user, and critically, enables a natural mapping between a user's bodily movement and the corresponding virtual behaviour. Avatars that exhibit humanoid form and behaviour have been shown to evoke a richer sense of copresence in observers [CB01].

An 'ideal' ICVE system would precisely track a user's full range of nonverbal behaviour, synthesising them in real-time on a personalised avatar to be observed at other users' sites. Aural, haptic, and olfactory signals would also be spatially reproduced. The complexity of human behaviour and state of real-time tracking and graphics ensures that such holistic reproduction of users' behaviour in AMC is not a realisable goal at the present stage of maturity in the research area. Indeed, Schroeder suggests that AMC will never provide completely realistic ways of interacting or communicating with others, as a number of features of collocated interaction will always be lacking [Sch02]. Even personalised and high-fidelity avatars are unable to exactly depict either the behavioural or the representational idiosyn-

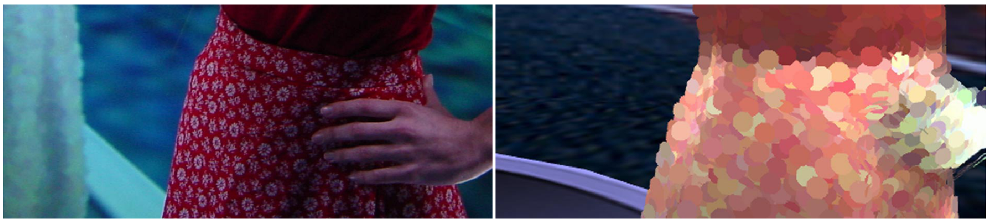


Figure 2.11: *Images of the* blue-c *system.* Left:*Close-up view of the real user's body.* Right: *Corresponding view of the synthesised 3D video display (courtesy of Gross et al.) [GWN+03].*

crasies of its controlling user. Hence, rather than complete and individualised behavioural reproduction, avatar embodiments should be likened to a 'marionette' featuring a series of strings which the user is continuously 'pulling' as smoothly as possible [BBF+97].

Holistic and intricate replication of behaviour is clearly an advantage of video-based technology, and the realisation of this orthogonal paradigm of 3D video reconstruction of remote users and environments is foreseeable in the future. Currently, the only example of this new generation of fully-immersive video-based telecommunication systems is *blue-c*, which is capable of operating between CAVE-like systems, and computes 3D video representations (using 3D irregular point samples as video primitives) of users from multiple cameras positioned around the displays [GWN+03]. However, several technical and performance issues limit the system as a practical telecommunication platform: at 15,000 point reconstruction the system operates at 9 frames-per-second (FPS), and round trip latency (i.e. the time taken between a user's motion and display at a remote site) is 330–560 ms, while at 25,000 point reconstruction frame-rate is 9 FPS and delay is between 600–1000 ms. These values do not satisfy the threshold of a 200 ms round trip delay in order to support high quality voice telecommunication that was established during the late 1960s by Brady [Bra71]. Brady also found that delays up to 600 ms were tolerable, but communication was adversely affected. The author is unaware of any similar investigations into latency tolerance in VMC, but one can reasonably assume that the highly temporal nature of the nonverbal component of communication will exacerbate communicational difficulties given a similar round trip delay. Finally, the blue-c cameras are also low-resolution, capturing content at a resolution of 640×480 pixels. Figure 2.11 illustrates the display fidelity of the blue-c system as it appeared in [GWN+03].

Schroeder distinguishes the two approaches of avatar and video representation respectively as 'computer-generated 3D environments' and 'video 3D environments' [Sch06]. Both are considered as end-states, supporting remote people completely immersed in mediated communication environments and interacting with each other. However, the two have quite different capabilities. Video 3D environments aim to capture the appearance of real users and real places, while computer generated 3D environments generate user embodiments (avatars) and virtual places. The two technologies also grant different activities: video environments are realistic and are constrained by this realism, while virtual environments allow manipulation of (virtual) objects and scenes. While the two approaches may be mixed, such as rendering video-based users in a VE as demonstrated in [GWN+03], they represent two quite different end-states toward the realisation of immersive 'copresent' communication.

During AMC in ICVEs, the capture, synthesis, and display of users' nonverbal behaviour from

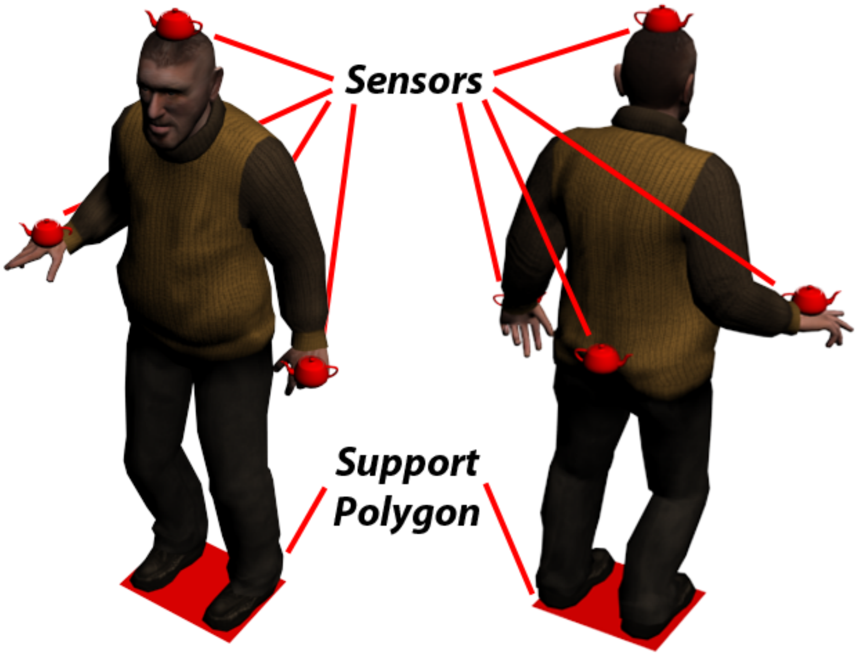


Figure 2.12: *Sensor placement and support polygon required to approximate a human user's standing pose as specified by Badler et al. [BHG93]*

reality to virtual reality is an approximation rather than precise replication. Cues that are captured are driven using tracking devices worn by the user [BBF+95]. Various types of tracking exist, including magnetic, ultra-sonic, and optical, and all variants aim to capture particular components of NVC with minimal latency in order to animate an avatar. Badler et al. demonstrated that four six-degree-of-freedom (DOF, three position, three orientation) sensors are required to create a good approximation of a human user's standing position and posture, and that this arrangement may be used to drive an articulated avatar in a VE [BHG93]. The positioning of the sensors is illustrated in Figure 2.12. The unsensed joints are positioned by an inverse kinematics (IK) algorithm, while a support polygon is compared against the sensor-computed centre of mass to activate stepping motions to balance the avatar.

While full-body articulation is attainable using optical tracking [Wor10], the most common tracking arrangement in academic ICVE research is the simplified solution of a head tracker and one hand tracker [GB95, SSA+01, SSH+03, GSV+03, RWOS03, SKMY09]. This two-tracker arrangement is identical to that used in single-user VE applications, where head tracking is used for perspective rendering, and the hand tracker is used for input. This two-tracker setup allows general kinesics to be approximated by avatars, but fails to capture more subtle nonverbal cues. In terms of tracking, the novel aspect of the research contained in this thesis is the introduction of mobile eye trackers, which are used both interactively for real-time oculesic replication, and analytically for data collection and subsequent analysis.

2.3.3 Copresence (Social Presence)

The theory of *social presence* in telecommunication systems was introduced by Short et al. as the degree of salience of another person taking part in the interaction, with a particular emphasis on how the transmission of nonverbal cues is supported by the medium [SWC76]. Similar to presence, social presence is subjective, and concerns a medium's perceived capacity to transmit the cues available in collocated communication. In the VE literature, the terms *social presence* and *copresence* are often used interchangeably, but as Garau summarises, what all definitions aim to capture is the sense of being in the company of another person during the course of mediated interaction [Gar03].

The idea of copresence in the field of ICVEs, may be defined as the extent to which the system interface becomes transparent and there is a sense of being present with other people in the environment specified by the displays, and there is a direct working with the other people [SHS+00]. The term is parallel to the established usage of presence that entails the sense of being present in a simulated virtual environment [Zha03]. ICVEs often combine a high degrees of presence and copresence, because the sense of being in another place and of being there with other people reinforce each other [Sch02]. Measurement of copresence can be performed by questionnaires [CB01], and also through physiological measurement as discussed in Section 2.2.2. That section revealed that a user's bodily movement relates to the degree of presence that they may be experiencing in a single-user VE. It follows that, during AMC between users of ICVE systems, the higher a user's sense of both presence and copresence, the more natural the behaviour of an avatar will be as observed by fellow users [FAM+00]. Hence, the degree of copresence, experienced by all interactants, is influential to the quality of the AMC.

The assumption in social psychology that humans do not need to experience the actual physical presence of others in order to influence or be influenced provides a firm basis for assuming that social influence can occur within a VE [Bla02]. Fostering a high degree of copresence during AMC is essential, and it has been suggested that an embodied user's level of social influence increases as a positive function of copresence [BLB+02]. The capability for an individual to experience copresence with fellow users of an ICVE system rests on their willingness to perceive avatars as representations of real persons [Bla02]. There have been several studies suggesting that, similarly to presence, people's response to virtual representations of humans is automatic, and leads to copresence. Jeremy Bailenson and colleagues working at the Virtual Human Interaction Lab at Stanford University have been pioneers of such work.

In an observational study of Second Life [Res10b], a large-scale online community, Yee et al. explored whether social norms of gender, proxemics, and gaze transfer into (non-immersive) CVEs, despite others being represented by avatars [YBU+07]. Results suggest that the established norms of proxemic and gaze behaviour are indeed preserved into VEs: male-male dyads maintain greater interpersonal distance than female-female dyads, male-male dyads maintain less eye contact than female-female dyads, and decreases in interpersonal distance are compensated with gaze avoidance, echoing Argyle et al.'s equilibrium theory specifying an inverse relationship between mutual gaze and interpersonal distance [AD65]. Six years earlier, Bailenson et al. established that people also behave according to social science theory in immersive VEs, by investigating interpersonal distance participants kept when faced with a humanoid agent compared with a similarly shaped and sized non-humanoid object [BBBL01]. All participants maintained more space around agents than they did around the objects, and female participants positioned themselves further from agents that engaged them in mutual gaze than from agents who did not.

Yee and Bailenson demonstrated that people infer their own expected behaviours and attitudes from observing their avatars appearance: a phenomenon termed the *Proteus Effect* [YB07]. For example, participants in attractive avatars were found to become friendlier and to increase self-disclosure to confederate strangers than participants in unattractive avatars, and users given taller avatars negotiated more aggressively than users given shorter avatars [YB07]. Applying the effect to avatars featuring photographic facial textures of their embodied participants, Fox and Bailenson found that the appearance of participants' avatars influenced the amount of exercise participants performed following the experiment [FB09]. When viewing an avatar who gained or lost weight in accordance with their own physical exercise, participants performed significantly more exercise in a voluntary phase than those viewing avatars exhibiting no appearance change, or no avatar at all. Subsequently, participants were asked to run on a treadmill while viewing an avatar in one of three states: running and featuring their own facial appearance; loitering and featuring their own facial appearance; or running but not displaying a visual likeness. Follow-up surveys performed 24-hours after the experiment revealed that participants in the likeness and running condition demonstrated significantly higher levels of exercise in the intervening period than those in other conditions. The Proteus Effect suggests that neither the virtual nor the physical self can ever truly be liberated from the other. While avatars are entities that are created and directed in VEs, this process can also occur in the opposite direction: avatars are able to recreate and direct people, influencing behaviour and interactions with others in the real-world long after the virtual experience has ended [YB09].

As the visual and behavioural properties of fellow users' avatars directly influence the experience of presence and social presence during AMC, the VR system's interface characteristics are critical. Imagine two ICVE users, John and Sue, each physically located in remote CAVE systems, but interacting via avatar embodiments in a shared VE. Now imagine that John sees Sue's avatar and decides to wave to her. John's hand is tracked, so he is able to perform this action in an identical manner to as he would if he was waving in the real world. Sue sees John's avatar performing the characteristic greeting gesture, and waves back by similar means. They approach each other and begin conversation. This is an example of a successful unit of nonverbal communication common to small-group interaction in ICVEs. It matters little if John's waving motion is not exactly replicated by his avatar's corresponding animation. Rather, the critical measure of success for a unit of communication is that the semantic properties of the original signal are preserved with high enough fidelity to be recognised by an observer. Following the exchange of waves, both users are convinced that they are interacting with another person in a shared environment.

2.3.4 Media Richness

A closely-related theory, which owes its origins to social presence, is Daft and Lengel's media richness theory [DL84]. Media richness theory describes the ability of a medium to transmit and reproduce information about the individuals who are communicating, with particular focus on transmission and display of users' natural nonverbal behaviour. The theory proposes that task performance will improve when the task needs are matched to a medium's ability to convey information. Hence, media capable of sending "rich" information, such as collocated interaction or visual telecommunication, are better

Figure 2.13: *Low-, medium- and high-fidelity avatars respectively from left to right. Left: Avatars used by Schroeder and colleagues in 2001 (courtesy of Ralph Schroeder) [SSA+01]. Middle: Avatars used in experiments covered in Chapters 4 and 5. Right: Avatar used in experiments covered in Chapter 6.*

suited to equivocal tasks where there may exist multiple interpretations of the available information, while media that are less rich, such as text-based chat, are best suited to tasks of uncertainty where there is framework for interpreting the information [DV99]. In the VE literature, media richness has been measured by the ability to communicate with and interpret others [GSBS01].

2.3.5 Fidelity

The *fidelity* of an avatar describes its sophistication. Vinayagamoorthy argued that the fidelity of avatars can vary along three dimensions: *anthropomorphism* (exhibiting humanoid form), *photorealism* (detailed texturing), and *truthfulness* (resemblance to controlling user) [Vin06]. While Bailenson and colleagues' work on truthfulness in VEs has been covered above, this section concentrates on the former pair of anthropomorphism and photorealism. These two aspects relate generally to avatar design and development, and are concerned with technical characteristics which give rise to the ability of an avatar to display nonverbal elements of human behaviour.

The classification of a "high-fidelity" avatar is a moving target that evolves with state-of-art real-time computer graphics, and expectations are influenced by exposure to media such as video games, online social worlds, and computer-generated films and artwork. Suggested characteristics of the current state-of-art of high-fidelity avatars are: facial expression capable of conveying emotion and some of the subtleties of NVC; freely-rotating eyeball geometry and mechanics to animate the eyelids to support blinking and lid saccades; phoneme-based lip-synchronisation; photo-realistic texturing; and a skeletal structure approaching that of humans. Avatars exhibiting characteristics of higher levels of visual and behavioural fidelity can potentially communicate subtleties of nonverbal communication more successfully than those of lower realism, and thereby enhance the quality of AMC [VSS05, Gar06]. A consistent interaction effect between visual and behavioural realism has been found, indicating that the effect of identical behavioural traits change in relation to the avatar's appearance: higher visual fidelity benefits from consistently realistic behaviour, while same rule applies for lower fidelities [Gar03]. Figure 2.13 aims to clarify this relationship.

Figure 2.13 shows three avatar models that may be considered low, medium and high fidelity. The leftmost avatars were used in Schroeder et al.'s Rubik's cube study circa 2001, the middle avatars are used in experiments covered in Chapters 4 and 5, and the rightmost avatars represent avatars used in experiments covered in Chapter 6. Taking the example cue of gaze, these three avatars demonstrate the interrelated nature of behavioural and representational fidelity: it is clear that the geometric (visual) fidelity of the leftmost avatars fundamentally preclude the ability to display gaze (behavioural), as the eyes are simply two black squares. According to the theory of consistency between components of fidelity, observers are unlikely to expect animated gaze in this case [NB03]. However, the middle and the right avatars exhibit higher levels of anthropomorphism and photorealism, and accordingly, observers are likely to expect more human-like behaviour to be exhibited, including gaze [Sch02]. In the absence of such expected behaviour, the avatars are likely to be perceived as lifeless [TBS+98].

Investigating the impact of the behavioural fidelity of eye movement on how avatars of various visual fidelities are perceived, Garau et al. argued that more realistic gaze behaviour had a positive impact on the perception of more visually-realistic avatars, but in the case of a lower visually realistic avatar, the more complex gaze model had a negative effect on participant response [Gar03]. A further iteration of this principle may be expected when the middle and rightmost avatars in Figure 2.13 are considered: while the geometry of the middle avatar does not allow for eyelid animation, the increased anthropomorphic fidelity of the rightmost avatar is likely to give rise to the expectancy for this behaviour. Hence, a failure of the rightmost avatar to display blinks is likely to bestow it with a quality of lifelessness.

As avatars increase in anthropomorphic realism, able to more successfully communicate nonverbal behaviour and thereby increasing the richness of AMC, a tension regarding control and agency arises. The increased expectancy for avatars to exhibit a complete and natural human behaviour relies on more complex methods of behavioural control, which is ideally achieved through tracking. However, tracking in ICVEs is technically challenging, as the VR systems in which they operate allow users to move freely, and feature low ambient light levels. This is a driving force behind the emergence of hybrid avatar-agents featuring a combination of tracking and model-based behaviour. Gaze is a particular cue that has warranted much interest in the VE literature. A review of gaze modelling is presented in the following section, and have the general aim of generating plausible eye movements during an unfolding interaction.

2.3.6 Avatar Oculesics

The work presented in this thesis investigates the addition of eye tracking to the gamut of bodily tracking devices used in ICVEs. The primary interactive use of eye tracking is to enhance the behavioural fidelity of avatars with faithful reproduction of their embodied users' oculesic cues during AMC. Gaze is investigated throughout the main experimental work presented in Chapters 4–6, and an avatar that features gaze behaviour that is replicated from their embodied user may be referred to as a *tracked gaze* avatar. Eyelid movement and pupil dilation are also investigated in Chapter 6, and an avatar featuring the three oculesic cues of gaze, eyelid movement, and pupil dilation may be referred to as an *oculesic* avatar. Tracked gaze and oculesic AMC has the potential to support superior gaze awareness than that possible in VMC, as the spatial context of ICVEs allows users to move freely within a perceptually-unified shared environment. Avatars generally exhibit the attentional cue of head orientation, driven by the prerequisite head tracking. Head orientation is a significant contributor to how observers estimate an individual's attention, but provides lower fidelity than the combination of both head orientation and gaze direction channels.

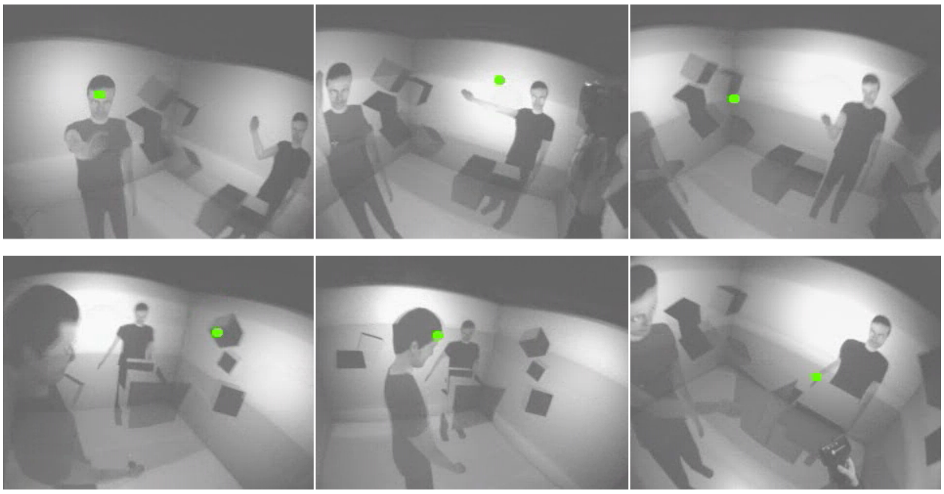


Figure 2.14: *Sequences of immersive AMC illustrating the changing focus of attention inferred from direction of gaze (represented by the green circle) despite similar head orientation during* greeting *(top sequence) and a* puzzle task *(bottom sequence). The addition of tracked gaze enables communicational gaze rather than just attentional head orientation.*

Figure 2.14 presents two sequences of interaction, taken from the experiment documented in Chapter 5. The two sequences are displayed horizontally and aim to illustrate the operational difference between the use of tracked gaze avatars, and avatars featuring just head movement. The two sequences each show three images captured from the perspective of a participant engaged in three-party object-focused AMC. Each avatar's eye movement is driven by an eye tracker worn a user, and similarly, each avatar's head is updated based on user movement from head tracking. While both sequences of images show similar views of the virtual scene, and corresponding head orientation can be used as a general indicator of attention, the user's direction of gaze indicated by the green circle varies dramatically throughout the interaction. Thus, in a VE populated with objects and avatars, the benefit of tracked gaze avatars is likely to become apparent.

In summary, the richness of attentional signalling and observation during AMC between users of ICVE systems corresponds to tracking capability. Due to the shared virtual space provided by ICVE systems, tracked gaze AMC allows gaze to be used similarly to collocated interaction, thereby overcoming VMC's restriction on gross movement and indicating objects with eye direction. Work prior to that presented in or associated (such as [MRS+07]) with this thesis, avatar and agent gaze has been driven by simulation, or appeared static. The following sections review work in the agent and avatar literature, together with studies found in the general computer graphics and animation literature, to present the current state-of-art in oculesic behaviour of virtual humans. Corresponding to the prior review of human oculesics, presented in Section 2.1, gaze, eyelid kinematics, and pupil dilation are covered.

Gaze Models

As the complexity of VEs and avatar behaviour increases, so does the difficulty in maintaining a direct correlation between the user's wishes and the avatar's actions [PSS+01]. Control of the full range of human nonverbal cues via tracking is impractical, and also too complex and temporal to be directed by means of manual input. Consequently, models directing various behavioural channels have emerged

[GB04]. Gaze models aim to generate naturalistic eye movement for a given interactional state and scenario in order to enhance the realism of a virtual humanoid. Other than VEs, application areas of gaze models include video games, computer-generated films, and general models of attention or gaze prediction.

Gaze models typically exhibit several types of characteristic behaviours which are often inferred by the current state of an unfolding interaction. In the case of conversation, this has included who is speaking and who is listening [PLBB02]. Further input to such analytical models often include parameters corresponding to behavioural components of gaze such as fixation time, angular velocity and saccade magnitude. These values implement statistical generalisations about human gaze behaviour derived from empirical studies of saccades [GB06] and/or statistical models of eye tracking data [PLBB02]. Manipulation of such parameters can dramatically influence how an avatar's psychological state is perceived, including being excited or sleepy [DLN05], and dominant or submissive [KG08]. Critically for AMC, results from associated user studies indicate that avatars exhibiting gaze behaviour that is directly related to the current interactional state are able to significantly improve subjective quality of communication compared to static gaze (no eye movement) or random gaze (eye movement is not inferred from any interactional state) [GSBS01, GSV+03, DLN05]. Such findings support Vilhjalmsson et al.'s assertion that, in order for avatars to meaningfully contribute to communication, their animation needs to reflect some aspect of the interaction that is taking place [VC98].

Lee et al.'s *Eyes Alive* model is based on both empirical studies of saccades and statistical models derived from empirical eye tracking data captured from dyadic conversation [PLBB02]. Eye trajectory kinematics were extracted from the eye tracking data, which was further segmented according to whether the wearer was speaking or listening. The model takes into account the dynamic characteristics of eye movement, including saccade magnitude, direction, duration, velocity, and inter-saccadic interval. An autonomous virtual agent head was used to exhibit various methods of gaze control: static, random, and model-based. On a standard display, experimental participants were then asked to give feedback relating to the perceived naturalness of the agent's gaze. Results indicated that model-generated gaze was perceived as more natural, friendly and outgoing, while stationary gaze was perceived as lifeless, and random gaze gave an unstable element to the agent.

Garau et al. [GSBS01] presented a parametric gaze model which took timings from the classic social science research on collocated dyadic conversations including Argyle and Cook [AC76], Argyle and Ingham [AIAM73], and Kendon and Cook [KC69]. Similarly to Lee et al.'s Eyes Alive model, gaze animations varied between speaking and listening states. For the speaking state, mean saccade fixation was 1.8 seconds when looking towards the conversational partner, and 2.1 seconds when looking somewhere else in the visual field. A mean frequency of 14 "at partner" glances per minute was programmed. For the listening state, mean saccade fixation was 2.5 seconds when looking at the conversational partner, and 1.6 seconds when looking away. A mean frequency of 17 "at partner" glances per minute was programmed. When looking at the conversational partner in both speaking and listening states, the avatar's eyes focused directly ahead, assuming that this was the partner's location. Values for vertical and horizontal angles of “away” gaze were chosen randomly from a uniformly distributed range between 0–15°. The associated experiment investigated impact of gaze on perceived quality of communication by comparing different gaze behaviour. Pairs of participants were asked to conduct a conversational role-playing task over a non-immersive video-tunnel link, on which an virtual humanoid representing the partner was displayed. The avatar exhibited either modelled or random gaze. Results indicated that an avatar whose gaze behaviour was directly related to the conversation consistently and significantly improved the quality of communication compared to random gaze.

Vinayagamoorthy et al. later combined the statistical elements of the Eyes Alive model with the timing data implemented in Garau et al.’s model [VGSS04]. Consequently, the model presented an approach to generating natural gaze motion which took theoretical information from social psychology studies to infer the gross gaze behaviour, and augmented this with the spatio-temporal eye trajectories derived from eye tracking data. Led by Garau, the user study extended those associated with the informing models [GSV+03]. While the same role-playing task in [GSBS01] was used, AMC was performed in an ICVE between pairs of participants using either a CAVE system or wearing an HMD. As well as gaze behaviour (a component of behavioural fidelity), the experiment also investigated representational fidelity of avatars, embodying users either with ‘cartoonish’, or photo-real avatar representations. These full-body avatars appeared life-size to participants, and exhibited either random or model-based gaze behaviour. Speaking and listening states were both divided into sub-states of “at partner” and “away”, determined by head-orientation derived from head tracking data. Findings relating to the photo-real avatars concurred with the previous studies, showing that gaze models inferred from interactional states can simulate behaviour that significantly enhanced the perceived quality of communication. However responses to the lower-realism avatars were adversely affected by the more realistic inferred gaze, supporting the theory of a significant interaction effect between appearance and behaviour as subsequently addressed in [Gar06]. It was also noted that experimental participants stood facing each other and maintained appropriate personal space in accordance to Hall’s social proxemics classifications [Hal68]. Subsequently, Yee and Bailenson’s work into the Proteus Effect observed this phenomenon in non-immersive AMC [YB07].

Gaze behaviour is a significant indicator of level of engagement, and gaze models have also been implemented as modules in broader simulations of human attention for agents, illustrating an alternate paradigm for simulating eye movement. Gu and Badler’s model [GB06] aims to provide agents with human-like responses to environmental stimuli by modelling aspects of human vision, memory and attention. The model implements low-level motor control of saccades and smooth pursuits, together with high-level gaze patterns that consider multi-party turn taking practices that include next speaker selection, engagement, cognitive workload, and distractions. The model views cognitive resources of agents as a finite resource. Thus, as an agent is assigned more demanding tasks or conversational situations, their mental workload increases and more attention is devoted to the most salient feature in a scene, consequently increasing the likelihood of missing an unexpected event or environmental distraction. Similarly, Peters and O’Sullivan introduced a gaze model as part of a broad agent animation system that infers interest from other agents in a scene, based on gaze, head orientation, body posture, and locomotion [PPB+05]. Subsequently, the observing agent makes the decision to continue speaking or take another action. The model takes into account external occurrences in the environment for both the speaker and the listener, and generates appropriate gaze to adjust engagement levels accordingly.

While the Gu and Badler [GB06] and Peters et al. [PPB+05] models do not have associated user studies, and are presented in rather theoretical manners, their approach to gaze modelling highlights the complexity and unpredictability of human social interaction that models based purely on statistical and empirical observations are unable to consider. Previous work on visual attention modelling on static images and dynamic video scenes provides a saliency-based approach to gaze modelling that, combined with statistical properties of eye movement, is investigated in Chapter 7. The remainder of this section covers work related to development of the model.

The aim of modelling visual attention is to focus computational resources on a specific, salient region within a scene. Koch and Ullmans framework for simulating human visual attention focuses on the idea that the control structure underlying visual attention needs to represent such locations within a topographic saliency map [KU85]. Multiple image features such as colour, orientation, and intensity may be combined, forming a saliency map that reflects areas of attention. Similarly, the intrinsic saliency of an object within a scene can be derived from parameters such as proximity, eccentricity, orientation, and velocity [FW99]. Computation of this intrinsic saliency can then be used to determine the spatial coding of gaze fixations in the virtual scene.

The extrinsic saliency of an object determines the duration of fixations [PLN02], and is concerned with coherent fixation distributions during inspection of a scene. Henderson and Hollingworth [HH99] review this area of high-level scene perception research further, which concerns the role of eye movements in scene perception, focusing on the influence of ongoing cognitive processing on the position and duration of fixations in a scene. Their review speculates whether ongoing perceptual and semantic processing accounts for the variability of fixation durations, which range from less than 50 ms to more than 1000 ms in a skewed-distribution with a mode of 230 ms. The average fixation duration during scene viewing is also stated to be 330 ms, with a significant variability around this mean. The review suggests that the fixation positions are not random, rather that they cluster on both visually and semantically informative regions. Spatial distribution of the first few fixations in a scene seems to be controlled by both the visual features in the scene and global semantic characteristics. As viewing progresses and local regions are fixated and semantically analysed, positions of later fixations come to be controlled by both the visual and semantic characteristics of local regions. The length of time the eyes remain in a given region is immediately affected by both characteristics.

Henderson and Hollingworth's research [HH99] leads to the hypothesis that the eye will be attracted to regions of a virtual scene that convey the most important information for scene interpretation. The intrinsic saliency of an object in a scene determines the spatial distribution of fixations, inferred from continuous interaction within the virtual scene. The extrinsic saliency drives the temporal coding of fixations, determining their duration. The gaze model presented in Chapter 7 takes head tracking as input, which defines a user's current view of a virtual scene. The extrinsic saliency of objects and other avatars is then calculated, and gaze is distributed accordingly. The approach also implements a plausible linear interpolation algorithm for the dynamics of the human eyeball.

2.3.7 Animating the Eyelids

When compared to eye gaze, animation of the eyelids has been largely overlooked in the computer graphics literature. Although subtle and deceptively simple, the kinematic behaviour of eyelids is non-trivial. As covered in Section 2.1, the eyelids have two major motion components: lid saccades that follow the saccadic motion the eyes, and blinking. Section 2.1 also established the role of eyelid movement, both in regard to providing information as to where people are looking and as indicators of an individual's psychological state.

As eyelid motion is a critical component of human facial expression, facial animation engines have often incorporated some form of eyelid dynamics in order to enable richer expressivity during human-agent interactions in VEs [VC98, BBP01]. The overall success of these systems depends on their ability to communicate recognisable emotion to observers rather than to generate accurate eyelid kinematics. Hence, quantification of physiologically accurate human eyelid motion is generally not an explicit concern. Optical marker-based facial performance capture techniques have been successful in recording the subtleties of an actor's facial movements [Wil90, CET+01]. This performance data can then be composited or learned to subsequently animate a character's eyelids. Similar methods have also been successful when used in combination with gaze motion data, building on the high correlation of movement between the two [BB09, DLN05]. However, such data is often closely coupled to a particular character rig, and thus is not easily parametrized to create a reusable procedural model. Also bypassing accurate dynamics, Tateno et al. implement "enhanced" (exaggerated more than double) eyelid deformation based upon an artist's sketches of the eyes [TTO05].

There has been some work investigating the impact of the frequency of blinking, or *blink rate*. Takashima et al. [TOY+08] presented a study investigating the effect on observers' subjective impressions of agents. The stimuli included humanoids of both sexes with generic reality, cartoon-style humanoids, animals and unidentified life-forms. Animations of agents performing blinks at various rates (9, 12, 18, 24 and 36 per minute) were judged on a seven-point Likert scale. Results revealed that agents' frequency of blinking has a dramatic impact on how they are perceived by observers. A rate of about 18 blinks per minute was seen to give a friendly impression, while a higher rate reduced the potency of the avatar, and a lower blink rate gave a more intelligent impression. Finally, the impact of varying blink rate was greater with the humanoid representations than the cartoon-style avatars [TOY+08].

Itti et al. [IDP03] incorporated blinks into their engine for animating avatar eye and head motions using a simple heuristic to determine frequency of blink rate. Blinks lasted for a fixed period of 150 ms, and no further details regarding dynamics are provided. Bitouk and Nayar implement a similar method, taking 200 ms for blink duration [BN08]. Although they are not specific about lid saccades, the gaze model presented by Peters and O'Sullivan [PO03] implements blinking correlated with gaze shifts by parametrising the vertical angle of gaze (presumably inferring a starting position of the eyelids) as an

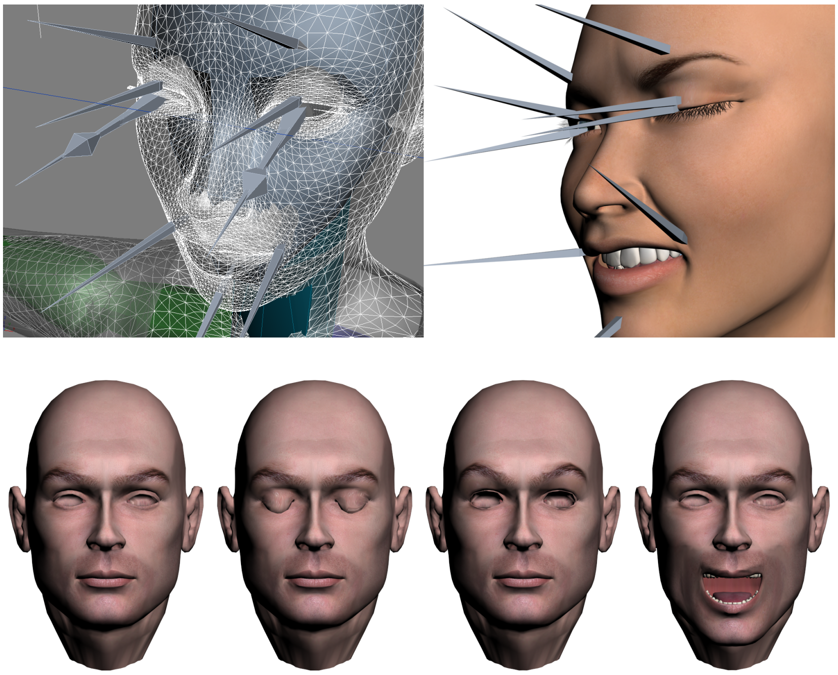


Figure 2.15: *Examples of technical methods to achieve facial animation, and in particular, eyelid movement. The models are used at various throughout the research presented in this thesis.* Top: *character with bone rig. The 'skeleton' of the character is mapped the geometry, so when the bones rotate, the weighted vertices deform accordingly.* Bottom: *character with morph target sub-meshes. Sub-meshes are blended between by the animation system to achieve facial expression.*

input to the magnitude of a blink. While this last study may hint at the complexity of eyelid movements, details of kinematic behaviour have so far been left undefined.

The computer graphics literature on virtual character animation has not presented physiologically-based models of human lid saccades or blinks. This is certainly not to say that current real-time and pre-rendered virtual characters have static eyelids, but rather that the problem of eyelid animation has generally been solved in an ad hoc way, using motion captured data or unrealistic linear motion. While animation generated from motion captured data can generally be considered physiologically-accurate, the process of capture and encoding is expensive, and often performed for specific implementations and character rigs. Thus, whilst a character's eyelid animation may be sophisticated, often based on blend shapes or linear blend skinning as illustrated in Figure 2.15, the characteristics of human eyelid kinematics have not been presented in the literature. This is addressed in Chapter 7, which presents models for lid saccades and blinks based on the eyelid dynamics of normal human subjects. A preliminary experiment, also presented in Chapter 7, demonstrates that the inclusion of lid saccades is a significant factor in enhancing the perception of realism of virtual humanoids.

2.3.8 Pupil Dilation

Even more-so than eyelid animation, it has only been relatively recently that VE and graphics researchers have begun to address the issue of pupil dilation in virtual characters. The temporal progression representing the synthesis, into graphical animation, of the gamut of visual human behaviour can be seen to follow a cost-benefit analysis that is coupled to the gradual advances of graphics and display technologies. Hence, critical elements of character animation, such as skeletal motion, facial expression, and gaze are generally well-researched, while the more subtle cues such as pupil dilation, tears, and wrinkles have only recently been considered. The evolution of real-time computer graphics is approaching a stage in which it is both technically-feasible to depict subtle characteristics of human physiology and behaviour, and, with the proliferation of high-definition displays, these subtleties are able to be represented and observed with clarity.

To date, there is only a single study in the VE literature that includes pupil dilation in the range of nonverbal expression exhibited an agent. Addressing the difficultly of robust gesture-recognition as an input modality to VE applications, Kotranza et al. [KJC+09] present a method in which a a single hand-held and tracked interaction device acts as a surrogate for multiple virtual tools including the user's hand, tools, and other objects. The authors evaluate the method in a medical education scenario, training users to conduct a neurological exam of a virtual human agent. An eye exam is among the tests that participants were asked to perform on the virtual patient, and this includes asking the agent to follow a moving object, asking the agent to blink or wink its eyes, and testing pupillary reflex to light. In the 'healthy' agents, changes in pupil size were animated according to the presence or absence of light. However, the animation dynamics were not defined, and so are unlikely to be based on physiological characteristics of pupillary response to light.

Pamplona et al. introduce a physiologically-based model for pupil light reflex and an image-based model for iridal pattern deformation [POB09]. The light reflex model expresses pupil diameter as a function of the environmental lighting, and is described by a delay-differential equation, naturally adapting the pupil diameter even to abrupt changes in lighting conditions. Derived from measured data, the reflex model correctly simulates the actual behaviour of the human pupil, producing high-fidelity appearance effects, and can be used to produce real-time predictive animations of the pupil and iris under variable lighting conditions. Images of Kotranza et al.'s and Pamplona et al.'s work on pupil dilation are presented in Figure 2.16.

While these studies are concerned with pupillary responses to light, work presented in this thesis is concerned rather with pupil dilation as an indicator of the psychological state of users engaged in immersive ACM. Similar to the other oculesic cues of gaze and blinking, the data is employed both for real-time animation of avatars and for analytical use. While the current literature has not directly approached this problem, there has been some interesting work with regards to using pupil dilation as an input to VE applications. Ekman et al. present an eyes-only computer game, Invisible Eni, which uses gaze, blinking and pupil size to affect game state [EPM08]. Pupil size can be indirectly controlled by physical activation, strong emotional experiences, and cognitive effort, and Invisible Eni maps users' pupil size variations to the game mechanics, allowing players to control game objects by use of "willpower" (manipulating their own cognitive and emotional effort). Specifically, the game uses pupil size to model magic, allowing the player to open flowers by staring at them with dilated pupils. Thus, the study implements a fantastical means of visualising a user's pupil size together with a novel interaction technique.

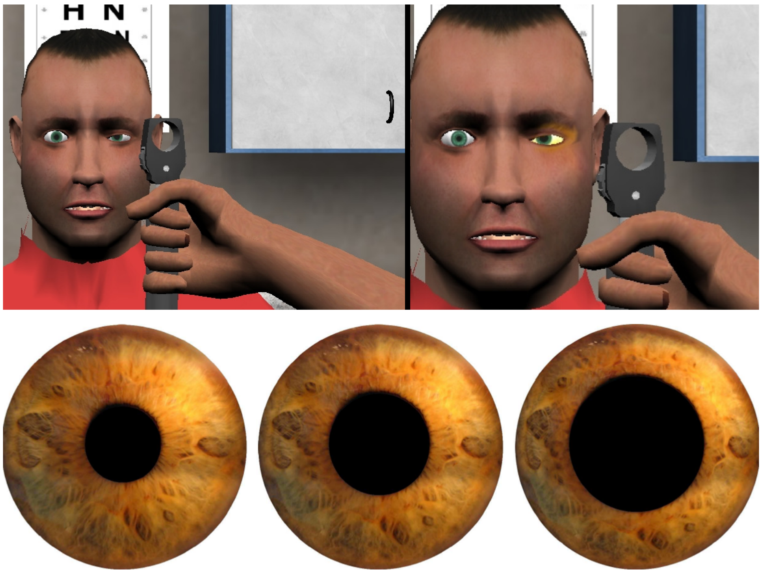


Figure 2.16: Top: *Virtual eye exam as presented in [KJC+09]. By shining the light of the virtual ophthalmoscope into the virtual patient's eyes, the right eye constricts, while the damaged left eye remains dilated.* Bottom: *Simulated results of pupillary reflex to light as presented in [POB09].*

2.4 Tracking versus Simulation

As examined over this chapter's previous sections, VMC is generally restricted by spatial orientation because users have a flat window onto each other's space rather than sharing perceptually unified environment. The spatial characteristics afforded by ICVE systems, in which remote participants are embodied as avatars, avoids this limitation, but introduces the challenge of transmitting users' nonverbal behaviour. Prioritising gaze as a critical nonverbal cue in social interaction, prior research in AMC has often relied on automated simulations that aim to generate eye motion that is relevant to the current state of unfolding interaction. Simulating gaze is an attractive solution for avatar gaze control, able to both enhance the realism of avatars and avoid the technical challenges associated with integrating eye tracking into an ICVE system.

This following section reviews analyses of collocated social interaction and identifies common gaze practices that cannot be determined or predicted by any practical means, including by the interactants themselves. The extent to which such practices constrain the prospects for faithful simulation of user gaze in AMC is discussed. It is argued that displaying unfaithful gaze modifies the communicational properties the AMC, thereby potentially changing the communication itself. This argument motivates the central problem of the research contained in this thesis, suggesting that tracked gaze AMC has potential to enable superior quality of communication. Finally, methods for better designing avatars and agents that inspire subsequent work presented in this thesis are discussed.

2.4.1 Limitations of Gaze Models

Gaze models have been shown to be capable of generating autonomous eye movement that achieves certain levels of perceived realism and task-relevant interactional cues. Associated user studies have demonstrated that modelled gaze is able to enhance copresence and believability beyond that achieved by static or random gaze avatars [CCD00, PLBB02, FOM+02, VGSS04, DLN05, KG08, MD09]. However, in the context of real-time AMC, there are a number of limitations and implications of modelling, as opposed to tracking, gaze. The empirical or statistical data on which gaze models are based has, overwhelmingly, been informed from dyadic collocated interaction. Accordingly, evaluation of the models has largely taken place using standard non-immersive displays and dyadic conversational scenarios. Consequently, it is unclear how well they can operate in multiparty or task-based interactions in ICVEs that feature a wide field-of-view and a high degree of spatial awareness. While models, such as Gu and Badler [GB06] and Peters et al. [PPB+05], aim to support richer, multiparty interaction between autonomous agents, it remains to be seen how they may be adapted for use in real-time immersive AMC.

A general issue hindering the use of behavioural modelling in AMC is that aspects of users' NVC are likely to vary: both across different moments in the interaction, and also across individuals and combinations of interacting individuals. Therefore, while inferred models may generate more interactionally-relevant behaviour than models generating random behaviour, action displayed by a model-driven avatar does not necessarily match the action performed by the user it represents. Thus behavioural models alter the semantic properties of the nonverbal component of communication between users engaged in AMC. An avatar which displays behaviour that contrasts to that of the user it claims to embody should be regarded as a subtle example of *transformed social interaction*. Thus, in the case of modelled gaze AMC, users' actual eye movements are not transmitted, and are instead replaced with autonomous behaviour. Throughout an avatar-mediated interaction there is likely to be a certain proportion overlap, in terms of gaze direction at specific moments in time, between gaze generated by a model and the actual gaze performed by a user. However, this is due to statistical probability rather than communicational similarity, and holistically, the nonverbal information available to interactants is altered significantly. In contrast, tracked gaze AMC, in which avatars reproduce their controlling user's actual gaze, enables the real-time communicational practices afforded by the cue, including signalling and observing others.

Gaze models infer the behaviour they generate from other actions and events. Therefore, a particular problem for such models is posed by interactional practices that are accomplished through the deployment and observation of gaze itself, rather than via other resources, such as talk or gesture. The following section reviews research into such practices. Findings emerging from conversation analysis (CA) are presented, which aim to provide detailed technical analyses of the resources out of which participants build social interaction. CA is an inductive process for analysing how human interaction is organised into sequences of action or systematic practices [Sac95]. CA is also employed as an analysis technique in the experimental work presented in Chapters 4 and 5.

2.4.2 Gaze Practices in Collocated Social Interaction

Underlying the implementation of behavioural modelling for avatars and agents is the aim of making their behaviour realistic, perceived in terms of similarity to experience of collocated interaction. Consequently, simulation requires knowledge of human practices such that they can be modelled. This section draws on CA research, demonstrating that in social interaction, humans habitually use gaze practices which are consequential for the interactions in which they occur, yet which cannot be predicted or inferred from any other observable behavioural state, including talk. These practices are significant, as many gaze models are based upon such classifiable, explicit, states. There are several practices accomplished with the deployment and observation of gaze that challenge the suitability of using behavioural simulation during AMC. These are addressed in turn.

Speaker Selection in Multiparty Interaction

One important use of gaze that poses serious difficulties for simulation is its deployment for speaker selection. In a multiparty setting featuring three or more interactants, a speaker may sometimes address all those present, and sometimes address just one person. There are number of resources that speakers can use to accomplish such unique addressing. One resource is to include a unique identification term such as the addressee's name. However, gaze appears to be an important and widely used practice to this end [Dun74]. Commonly, a speaker will gaze at the party whom they are addressing, and in particular, bring their gaze to them as their turn reaches its end. Often, talk will include the pronoun "you", thereby indicating that someone is being addressed, but whom that someone is may only be revealed by the speaker's gaze [Ler03].

The CA fragment presented in Figure 2.17 details an example of this phenomenon. Four friends are talking and having dinner when one participant (Vivian) asks a question: "Have you been watching it [referring to the film Rocky 3] a lot?" (Line 12). Evidently, there is no vocal clue about who is being addressed by "you", and following this utterance, the listeners' monitor Vivian's gaze to determine who is being questioned and selected as next speaker. This occurs firstly on Line 13, annotated by *Nv* and *Mv*, indicating that '*N*ancy gazes at *V*ivian' and '*M*ichael gazes at *V*ivian' respectively. However, as indicated by *Vs* on Line 13, it is apparent that Vivian is in fact gazing toward Shane during the production of her question. It is only after 1.2 seconds, indicated on Line 14, that Shane finishes chewing his food and looks towards Vivian, on Line 15, that he recognises that the question is directed at him, and begins to produce a response, at Line 16.

In this case, the utterance of "you" in the question is used to address a single person, thus selecting that person to speak next. However, all those listening to Vivian could assume that they were possibly being addressed. It is only through the use of gaze, when Nancy and Michael look up at a point when a response may soon be due, that they can determine from Vivian's gaze direction that the question is being visibly directed to Shane, and neither responds during the 1.2 seconds before Shane speaks at Line 16.

This human social practice has two important implications for the experiments investigating tracked gaze AMC presented in Chapters 4–6. Firstly, it demonstrates the importance of gaze during natural collocated communication, and hence suggests the benefit of capturing and representing gaze in a faithful and meaningful form during visual telecommunication. Secondly, it reveals an important limitation of gaze simulation. Given that there is nothing in the speaker's talk that listeners can use to determine who is being addressed, the listeners must monitor the speaker's gaze, which allocates the next speaker. Accordingly, there is no information embedded in the speaker's talk or kinesic behaviour that a gaze model could use to infer where the speaker is in fact gazing. Hence, if the interaction scenario described

[Chicken dinner]

```
 1          (1.5)
 2 Nancy:   Let's watch Rocky Three.
 3          (0.7)
 4 Shane:   Yhheahh.
 5          (0.8)
 6 Michael:°'M gunna be s:[ick.°
 7 Shane:                  [Um (.) always up f'tha:t
 8 Michael:'M gunna be sick.
 9 Shane:   huh ha h[oh haa-aa-heh
10 (Vivian):        [mm-hm-mm-hm-mm.
11          (1.6)
12 Vivian: Have you been watching it a lot?
13          Vs-------------------- NvOMv--
14          (1.2 )
15          -----Sv------- ((here each "-"5 0.1))
16 Shane: Ner-nahwuh- (.) Well
17         (.)
```

Figure 2.17: *Extract from Lerner [Ler03]*

.

in Figure 2.17 were to be performed in AMC, the communication is likely to be significantly altered if gaze is modelled rather than reproduced faithfully.

Promoting Sequence Expansion

In collocated interaction, gaze is often employed in particular sequential positions, such as following a response (such as an answer) to an initiating action (such as a question) in order to pursue further talk. For example, Rossano [Ros05] shows that, following an answer to a question, a questioner may maintain their gaze at the answerer, indicating that they do not consider the provided answer as sufficient for the sequence to be complete. More generally, interactants display to one another that they take a sequence to be possibly (and actually) closed at that point by withdrawing from mutual gaze. By maintaining their gaze, people can show themselves not to be treating a sequence as being finished, so the sequence is often extended as their gaze behaviour pursues further talk from an interlocutor. Moreover, people can withdraw gaze from an interlocutor at a point of possible sequence-completion, and then look up to see if their interlocutor has also withdrawn their gaze, indicating whether the potential sequence-completion has indeed been treated as actual completion or not [Sch07]. This careful monitoring of participants' gaze in interaction is only understood in terms of talk as the place in which a sequence reaches its *possible* completion, but what is treated as *actual* completion is only decided locally by the interactants, and is accomplished by gaze signalling and observation. Again, this communicational exchange would presumably fail in a majority of instances in a simulated gaze scenario in AMC, as gaze itself is the conductor of the interaction.

Characterising a Suspension of Talk

The examples of speaker selection and sequence expansion presented above show how interactants can engage in 'meaningful' gaze practices that can only be understood from gaze itself, and not from talk or other interactional elements that it is associated or coincident with. There are more communicational practices that can be achieved by gaze (in conjunction with other factors) that cannot be easily inferred just by the talk or other kinesic action during a conversation. The practices examined below involve interruptions to talk with different gaze practices that are deployed to achieve very different actions.

Kidwell shows how very young children understand the importance of different looks produced by their caregivers, distinguishing between 'mere looks' and 'the look' which can be used for sanctioning a child [Kid05]. An interruption to talk associated with fixed and relatively long gaze at a child produces sanctioning of some kind of bad behaviour, and is often successful in stopping the child's activity. For example, in Figure 2.18, a caregiver (CG) reads a story book to two children. At the beginning of the exert, the caregiver looks towards the children while she reads from the book. During this 'mere look', one child pulls the shirt of the other child, Heather (H), who cries out as the caregiver's gaze returns to the book at the utterance of "hop". On reaching the end of her sentence, the caregiver's gaze then returns to the children, this time interrupting the process of reading the story until the child being sanctioned lets go of Heather's shirt: this is 'the look'. The children are shown by Kidwell to be both attentive and responsive to both *whether* they are being looked at and *how* they are looked at.

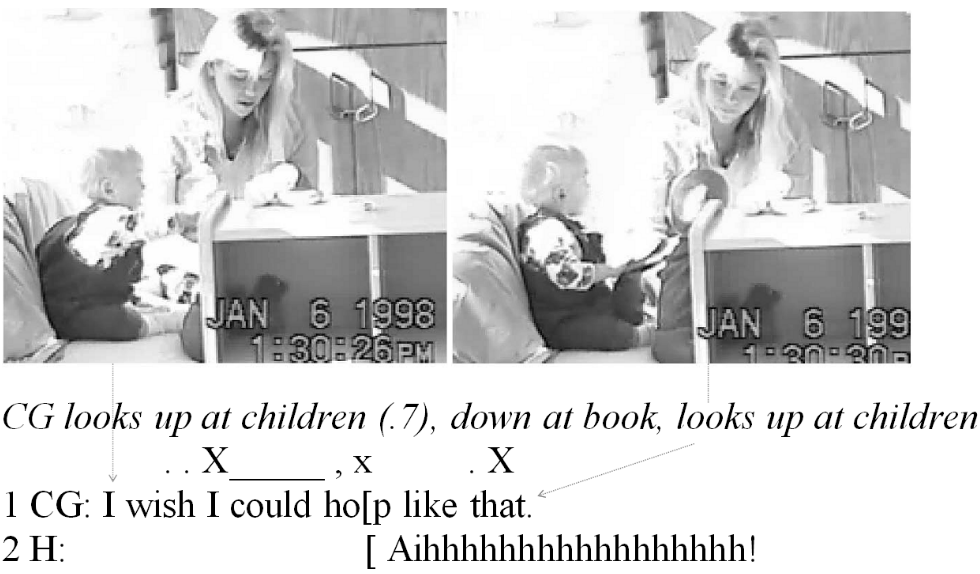


Figure 2.18: *Extract from Kidwell [Kid05].*

Not every interruption to talk is, however, a sanctioning activity. Goodwin and Goodwin show that when engaging in word-searches (thinking about the correct word to use), and thereby interrupting the progression of their talk, a speaker will look away from the listeners of their talk [GG86]. Such interruption of talk can be understood in those cases as being performed as the course of activity of finding the right term, name or word in order to then progress with the activity that was in course. Other interruptions and restarts while the speaker looks at their interlocutor can be used when speakers seek to attract gaze toward them [Ros05]. It can be seen from such cases that similar vocal practices, such as silences and the interruption of talk, can be associated with different actions. As Kidwell states, "the import of a gaze is a locally contingent, interactionally achieved matter" [Kid05]. Another relevant import from the examples shown above is the fact that some activities are, in effect, implemented with the use of gaze rather than via other verbal or nonverbal means. This practice has been demonstrated above with regards to speaker selection, to pursue an action, and to sanction bad behaviour. Another example of an activity featuring sequences of interaction which cannot be inferred by the talk amongst interactants, and is instead achieved by nonverbal resources, are those involving objects.

Projecting Involvement with a Different Activity

As shown by Rae [Rae01], participants can project an involvement with some other activity and/or object (for example a telephone and the action of ringing someone) via nonverbal resources. The extract presented in Figure 2.19 makes evident that although a participant does not make any verbal reference to phoning someone, their nonverbal signalling implies entry into a phone call.

Case 4 (UAA:14)7

```
A: I know that there should be >in principle<
                ((head turn))
                 |           |
   mid                        phone
   I don't know whether (.) its been remembered
       ((swivels chair))
               |
   to be done on this occasion
                    ((picks up phone))
                                |
   um let's see if Sue Cummins is there
```

Figure 2.19: *Extract from Rae [Rae01].*

Two particular body movements are especially relevant here. First, during his talk, and well before he picks up the receiver, Anthony (A) turns his head toward the phone. This redirection of gaze involves a head turn of about 90°, and subsequent rotation of the chair toward the phone. These nonverbal actions, together with talk, clearly project involvement in phoning. In this fragment, the speaker's head-turn can be seen as the initial movement in a trajectory that leads to making contact with the telephone [Rae01]. Then, by picking up the receiver while his talk is still in progress, the speaker further projects that the relevant action on the cessation of his talk will be making a phone call. Again, the employment of gaze in this fragment cannot be inferred from talk, but rather from the specific use of meaningful head orientation and gaze direction.

Another important issue to be raised is the role of head orientation in understanding gaze direction and participants' attention. As shown in the extract above, the listener does not have to see the eyes of the speaker in order to understand that a head-turn towards the phone, and a subsequent repositioning of seating arrangement towards the phone, means 'looking at the phone' and projects the making of a phone call. Participants and analysts, often do infer what/who is being looked at by their interlocutor's head orientation. It should be pointed out that in general, research in human interaction using CA, the methodological approach to data analysis considered here, focuses on naturally occurring, spontaneous social interaction and uses video recording to capture it. Therefore, it does not use eye tracking in the analysis of gaze in social interaction, and consequently the focus of participants' gaze sometimes must be inferred from head orientation.

2.4.3 Discussion

The human practices associated with gaze reviewed over the previous sections have shown the importance of the cue in natural collocated interaction and, consequently, demonstrate the need to capture and represent gaze in visual telecommunication systems. Gaze is "part of the interactional machinery by which participants sustain and regulate their conjoined activities" [Rae01]. Consequently, there are profound problems in attempting to simulate the gaze of embodied avatars during real-time interaction, as gaze itself, rather than talk or other nonverbal signals, is most commonly used to accomplish certain interactional practices including speaker selection, to pursue an action, to sanction a bad behaviour, and to project an involvement with a different activity.

The representation of users' gaze behaviour poses particular challenges to AMC in ICVEs. Whilst certain aspects of human behaviour, such as talk, can easily be captured and transmitted, gaze is technically challenging, particularly in immersive VR systems where users are granted free movement and ambient light levels are low. Clearly, in case of technical failures and other shortcomings, there are practical benefits to at least some degree of simulation of avatar gaze. Simulation of remote users' gaze at each local site appears an attractive strategy, as it would circumvent the need for the network transmission of gaze information, and also adapts to failures at remote sites and asymmetric hardware configurations. In addition, when remote users' gaze has been tracked, transmitted and represented, smoothing or filtering may be required due to eye tracker noise or losses in data transmission. Nevertheless, with the exception perhaps of certain degrees of low-level filtering, simulation requires knowledge of human gaze practices. However, based on CA research presented in this section, there are certain classes of gaze behaviour where the interactional work is done by gaze itself. This review identified four cases: selecting next speaker in multiparty interactions, showing that a sequence has not yet reached completion, sanctioning, and proposing a course of action. These behaviours cannot be inferred from any verbal or nonverbal behaviour but gaze.

The distinction between head orientation and gaze has been discussed in Section 2.4.2. When a person's gaze displays what they are doing, it is possible that head orientation may be sufficient to communicate the relevant information. Indeed, it was noted that sometimes when carrying out interactional analysis without eye tracking, that gaze is inferred from head orientation. Whilst it appears to be the speaker's gaze that is the critical point of this issue, there is still the possibility that head orientation might be enough of a clue to infer gaze direction. However this seems unlikely: as documented by Lerner, although gaze is an explicit form of addressing, the success of this practice depends upon the separate gazing practices of co-participants [Ler03]. So, the addressed participant must recognise that they are being addressed, and other participants must understand they are not being addressed. This demonstrates that gaze placement is carefully used and monitored by participants. Moreover, when observing agents in semi-immersive VR systems, gaze has been shown to be of vital importance for the correct identification of gaze direction [MRS+07].

The case of speaker selection is particularly interesting as it has received attention both within the VMC literature [VSvdVN01] and within CA literature [Sch07]. In collocated interaction, the success of speaker selection through gaze depends on the careful attention of listeners as to the speakers' gaze behaviour. As such, it seems likely that mutual gaze is relevant at this juncture because it enables a speaker who is allocating the floor to another interactant through gaze to see that this party has indeed seen their gaze. Likewise, it also puts the allocated party in the position of seeing that this is the case. Such junctures are sometimes referred to as "eye contact", however this is misleading for two reasons. Firstly, when concerned with a participant seeing that they have been allocated a turn at talk, the issue is whether or not they can see that they are being gazed at. Consequently, the term "eye contact", with its connotations of "holding" each other's gaze, introduces extraneous ideas. Secondly, in collocated interaction, the use of gaze as a resource for speaker selection is not just a matter of the selected party seeing that they have been addressed: it is also a matter of other parties seeing they have *not* been selected, and subsequently using their gaze to identify who has been selected. So, it is not enough to know when gaze is directed at oneself; it is also crucial for interactants to see who else the speaker is gazing at.

Therefore, whilst speaker selection is clearly important activity during visual telecommunication, it should be remembered that there is also the more general requirement for all participants to be able to be aware of, and to be able to determine, the gaze of others. Sometimes, addressing is done not by gazing at the addressee but by gazing at certain objects. For instance, "Can you pass that?" is a common phrase heard in a collocated multiparty setting, where gaze shows what "that" refers to and "you" can then be inferred by proximity to the object. Incidentally, in some cases, it seems very plausible that hand gesture would be implicated.

A major aim of work involving virtual avatars and agents is providing behavioural and representational realism such that observers feel that they are collocated with other humans, who may be embodied by an avatar. In Chapter 7, simulation of human gaze behaviour with a model developed from captured eye tracking data is investigated. The potential to model gaze presents itself as a solution to various challenges that arise in using AMC as a form of remote interaction. However, work on avatar realism is partially orthogonal to accurate display of users' nonverbal behaviour. A particular problem is that inferred gaze may be rated as "realistic" [PLBB02, VGSS04], yet fail to replicate crucial communicational gaze behaviour. Therefore, a gaze model may receive high ratings for perceived realism or authenticity, yet not actually be faithful to the behaviour of the embodied user. In the context of autonomous agents, the aim is to present gaze behaviour that is perceived to be realistic. However, in the case of real-time AMC, there are actually human participants communicating, whose gaze behaviour could be quite different from that displayed by a gaze model. Indeed, in some contexts, ethical issues might arise if a participant does not know how their behaviour is being represented.

Finally, it should be noted that an avatar exhibiting modelled behaviours it is no longer a pure avatar measured on the dimension of agency detailed in Section 2.3.1. Such avatar-agent hybridisation is perhaps evident in Vinayagamoorthy and colleagues remark: “By embodying an avatar with behaviour, emotion or personality skills, we provide the participant with a virtual character in the full sense of the word” [VGSS04]. Although some problems of using models to represent human behaviour in AMC have been emphasised here, this does not remove from the fact that such studies have their value as empirical demonstrations of how certain aspects of avatar behaviour are perceived. Avatars and agents have strong potential in educational and entertainment applications, where they may be used to display manipulable behavioural traits and interactional styles that may differ from the actual behaviour of a controlling user [BBL+03].

2.5 Chapter Summary

This chapter has been divided into four main sections. The first section presented an overview of human NVC, covering kinesics, gesture, facial expression, proxemics, chronemics, and oculesics. Oculesics were covered in particular detail, identifying the importance of gaze, eyelid kinematics, and pupil dilation in social interaction, along with the physiological properties of the cues. The particular scenario of deceptive communication was introduced, demonstrating the importance of oculesics during the psychologically- and cognitively-arousing interaction of communicating and detecting deception.

The second section focused on visual telecommunication systems, establishing the importance of gaze awareness, and positioning state-of-art VMC systems as the optimal form of high-quality interpersonal remote interaction. VMC’s inherent problems with regards to representation of 3D space, along with novel but imperfect approaches aiming to alleviate this problem were then detailed. AMC in ICVEs were then presented as a maturing medium, able to overcome the spatial limitations of VMC, locating users in a navigable and interactive shared graphical environment populated with objects and avatars embodying users. The topics of immersion and presence, central to VE systems, were also covered.

The third section introduced avatars as the primary mediator of the visual component of telecommunication in ICVEs. The influence of social agency, user embodiment and control, copresence, and fidelity were established as critical factors influential to the success of AMC. However, the primitive nature of avatar fidelity, particularly with regards to faithful replication of nonverbal behaviour, was established as a critical hindrance to the support of high-quality interpersonal AMC. Reflecting the review of human oculesics, work regarding avatar gaze, eyelid dynamics, and pupil dilation was then presented, with a particular focus on gaze models: a predominant method of avatar gaze control.

The fourth and final section constructed the argument that gaze modelling has profound problems for use in real-time AMC, and that gaze must be tracked in order to preserve the semantic properties of communication. CA and social interaction research was reviewed, exposing four classes of gaze behaviour (selecting next speaker in multiparty interactions, showing that a sequence has not reached completion, sanctioning, and proposing a course of action) where interactional work is performed by gaze itself. Therefore, while simulations of gaze may be able to generate observably believable behaviour, they are incapable of inferring the actual eye movements of users from any verbal or nonverbal cue but gaze itself.

The major experimental research presented in this thesis investigates the use of eye tracking in AMC. The experiment presented in Chapter 4 is preliminary, and compares tracked gaze AMC and gaze aware VMC in a multiparty conversational scenario. Analysis is performed using eye tracking data in combination with CA. Chapter 5 investigates the influence of varying methods of avatar gaze control during collaboration in a multiparty object-focused scenario. Objective measures of performance are combined with eye tracking analysis and subjective questionnaires. Resuming the interest in truth and deception, the final telecommunication experiment, presented in Chapter 6, compares 'full' oculesic AMC (featuring gaze, blinking, and pupil dilation) with VMC. Eye tracking data and questionnaires are used to perform analysis of the interactions, and a follow-up experiment determines the impact of bestowing avatars with full oculesic behaviour on the ability of observers to detect truth-telling and lying. Chapter 7 presents work involving behavioural modelling: a preliminary experiment investigating the impact of varying oculesic properties of an agent on how they are perceived by observers; two models describing the kinematic properties of eyelid motion for blinks and lid saccades, with associated experimental evaluations; and a saliency-based gaze model with an associated experimental evaluation.

The following methodology chapter is divided into two sections. Firstly, the ICVE platform, *Eye-CVE*, which is developed and used throughout the following experimental work is presented. Secondly, a method for multimodal data collection and interaction analysis is presented. This method is employed in the experimental work in Chapters 4–6, and enables collection of oculesic behaviour (including gaze, blinking, and pupil dilation), head and hand motion, and vocal signals into a single log-file.

Chapter 3

Technical and Analytical Methods

This chapter details technical and analytical methods used to address the research questions in the forthcoming experimental work. Section 3.1 documents the architecture and performance of *EyeCVE*, which acts as the primary ICVE platform supporting the experimental AMC investigated over the following chapters. A particular focus on the technical integration of eye tracking, and on the avatar subsystem are presented. As the research questions presented in Chapter 1 suggest, one aim of the experimental work is to compare state-of-art AMC and VMC systems. Explication of the VMC system setups, which require only minor technical implementation effort when compared to that involved with EyeCVE, are presented in the relevant experimental chapters (Chapters 4 and 6) rather than here. Section 3.2 focuses on methods of data collection and analysis. Throughout the forthcoming experimental work, a variety of approaches to both data collection and analysis are employed, some, such as presence questionnaires and CA, have been documented in Chapter 2's literature review. The section focuses on a novel approach to multimodal data collection and analysis. The technical architecture of the method is detailed, which operates alongside EyeCVE to capture multiple components of a user's action in a single output file, thus preserving the temporal and interrelated nature of the multiple tracking streams. Finally, the chapter is summarised with a reading guide to the forthcoming experimental work.

3.1 *EyeCVE* System Overview

This section reports on key technical aspects of the ICVE application, EyeCVE, which was developed and investigated throughout the experiments in this thesis. The development of the system was a collaborative effort. The initial system design and implementation was mainly performed by Robin Wolff at University of Salford, but also included Alessio Murgia at University of Reading, and Oyewole Oyekoya at UCL, together with the author. The advanced avatar and eye tracking subsystems were implemented solely by the author.

EyeCVE is the first ICVE system to support AMC in which the oculesic behaviour of avatars is driven in real-time using eye tracking technology. These oculesic cues include gaze, blinking, and pupil dilation. This replication of users' oculesic behaviour augments the standard tracking modes of head and hand movement, which are standard to ICVE systems [SSA+01, RWOS03]. Hence, EyeCVE allows people in different physical locations to not only see what each other are doing, but also to follow each other's gaze, even when physically moving around the space provided by a CAVE system. Projected into each users' display is a VE populated with avatar embodiments of remote users, providing a spatially- and temporally-aligned shared workspace.

The temporal challenge of the system is to reproduce users' gaze for observation on remote displays both quick enough and often enough to preserve context and meaning during multiparty interaction. Gaze is the primary oculesic behaviour captured by the system, and hence is the primary focus of this section. However, the same requirement stands for the other nonverbal signals of blinks, pupil dilation, and head and hand movement. Additionally, the verbal component of communication must also be transferred with minimal delay. The spatial challenge of the system is to maintain these channels of nonverbal communication while granting users free movement within the shared VE.

Derived from properties of human gaze behaviour, the minimum performance requirements of a tracked gaze AMC system are firstly presented. The distribution architecture of EyeCVE's approach is then presented, with a particular focus on ensuring concurrency of nonverbal messages. The two commercial eye tracking systems used to integrate oculesic communication into EyeCVE are then covered, together with calibration details. EyeCVE's avatar subsystem is then documented. Two distinct avatar types, *basic* and *advanced*, with varying capabilities for nonverbal expression, are used throughout the experimental work. Their relative components of behavioural animation, which visually synthesise users' tracked oculesic and bodily movement, and also verbal signals, are introduced. An evaluation of an early version of EyeCVE in a two-CAVE link over the Internet is then presented, establishing the system's ability to support tracked gaze AMC within the specified temporal and spatial requirements.

3.1.1 Requirements

Mediating gaze, and other nonverbal cues, across a distance using a computer graphics approach to visual telecommunication involves stages of acquisition (tracking), interpretation, simulation, distribution, and representation. Each stage introduces a delay and, perhaps, loss of information. Linking remote sites across public switched networks such as the Internet is particularly prone to delay, jitter, disorder and loss of packets. In such Human-in-the-Loop (HITL) systems, a common misconception is that networks are responsible for the majority of delay during the remote interaction. However, computation usually causes as much, and sometimes, more. The distributed architecture must include consistency management to balance between responsiveness, synchronisation, and causality in the face of network and computational delay. Responsiveness to viewpoint updates is a key requirement for immersive graphics that, if insufficient, may reduce the sense of presence [MR03] and result in motion sickness [CNSD93]. Responsiveness of objects and avatars during interactions does not need to be as high, but low response levels can impact on the interpretation of NVC, and are likely to hinder both conversational and object-focused interaction. While the responsiveness of viewpoint updates may be increased through a loosely-coupled replicated database approach, this may reduce synchronisation of observed events, especially when they originate from distinct sites. For example, a loss of temporal synchronisation between a user's head and eye movements as represented by an avatar may change the semantic intention of the original communication, and at the least, present a less faithful replication of user behaviour.

While humans read, visually search, or perceive a scene, our eyes move on average every 200–350 ms to rotate the fovea to an area of interest in order to process it in greater detail [LZ99]. Fixation duration is task dependent: when reading it is about 225–250 ms before the eyes jump over 8–9 characters to the next fixation point [Ray98b]; when viewing web pages, mean fixation duration is around 350 ms [PHG+04]; and when playing first-person shooter video games, mean fixation duration lies between 350–750 ms [KKD+05]. The work presented in this thesis investigates communicating eye gaze in multiparty conversations and collaborative tasks around shared objects. These tasks are likely to induce eye movements with fixations falling on the objects under discussion, areas of interest, and mutual gaze between interactants. For these scenarios, relatively long fixation durations are expected, such as 1000 ms up to a few seconds [PLBB02]. This implies that in order to support the attentional properties of gaze in a visual telecommunication system, an update rate of at least 2 Hz (500 ms) is required. Moreover, in order to distinguish fixations from saccades, this requires an update rate of at least 10 Hz. In terms of latency, a global requirement for high-quality telecommunication is 200 ms or less [Bra71].

3.1.2 Distribution Architecture

In a distributed VE system aiming to support real-time telecommunication, the combined delay from acquisition, computation, communication, and display are perceptible, and can impact the interpretation of nonverbal cues. EyeCVE adopts the replicated database approach, which is common to distributed and real-time simulation and visualisation, and allows responsive user feedback. The architecture of EyeCVE, showing a simplified unidirectional view of information flow, is presented in Figure 3.1.

Based on input from dedicated tracking nodes, each client simulates the current state of the local user's avatar, and sends updates to peers via a server. Upon receiving such updates, remote clients then render avatar behaviour within the VE. Each client holds a replication of two tightly-coupled databases: the *simulation object model*, which describes the behaviour of the virtual world; and the *scene graph*, which describes its appearance. For example, head and eye movement are updated in the simulation object model which, on remote sites, ties to respective parts of the scene graph, which visualises the database. The scene graph hierarchy follows that of the simulation object model, but adds specific data defining object appearance such as texture maps and geometric models. Rendering of the virtual scene is performed using OpenSG [RVB02], which supports a wide range of immersive display types including tiled, panoramic, L-shaped, and cubic. Audio communication between users interacting via EyeCVE is handled externally by Google Talk [Goo10b], which maintains low-latency *Voice over Internet Protocol* (VoIP) communication.

Time management is necessary to balance local responsiveness with consistency across client replications, both in terms of synchronisation and causality. Like many HITL simulations, the majority of update messages in EyeCVE describe discreet absolute movement as opposed to the difference from the previous state. This implies that the system can recover from a lost update as soon as a subsequent update for the same object arrives. It also means that preceding updates become obsolete and will not be further processed. Due to the temporal nature of real-time communication, it is preferable to transmit, process, and display updates as quickly as possible rather than to ensure that every update is received. However,

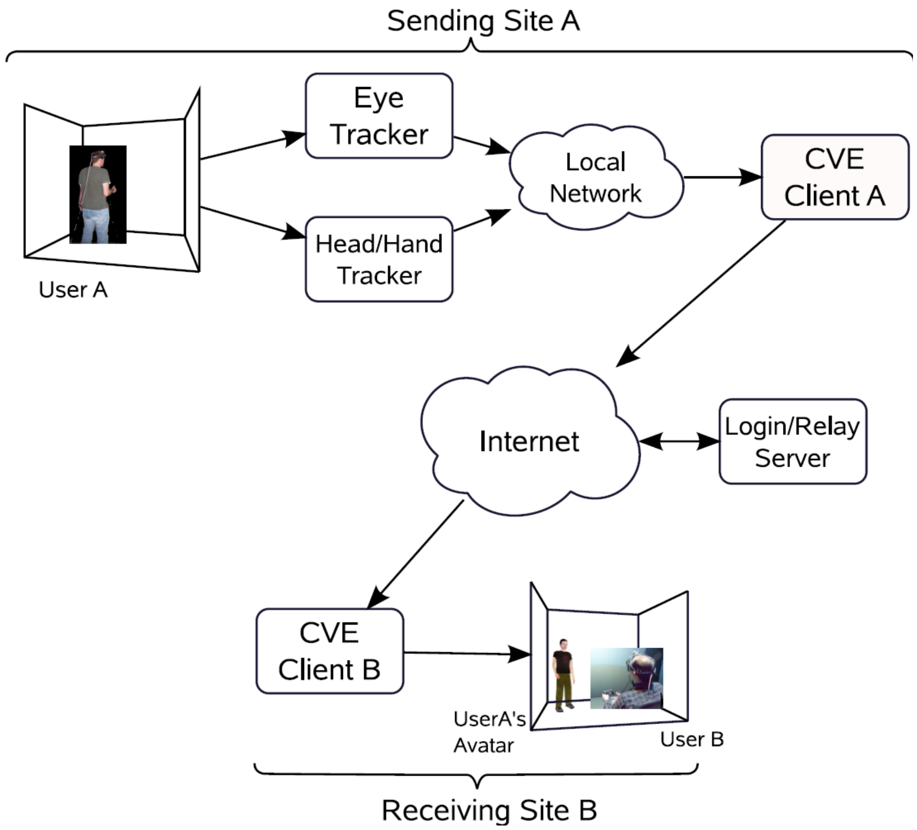


Figure 3.1: *Distribution architecture of EyeCVE. Two sites are shown. Site A tracks a CAVE user's gaze, head, and hand motion, and distributes updates via the client software to Site B, where the tracked motion is mapped onto an avatar. The relay server facilitates login and message passing across a group of distributed users.*

this must be balanced with the need to ensure both the delivery and sufficient causal ordering of critical events such as a user picking up an object and passing it to another user, or clients joining and leaving. These vital events are expected to occur far less than non-vital events, so overhead from bandwidth and ordering of reliable transmissions is minimal. EyeCVE's network layer was implemented using RakNet [Sof10a], which uses the User Datagram Protocol (UDP) for transmission, but employs additional mechanisms to assure reliable transmission of vital update messages. Ordering is applied to vital events in a separate ordering channel so as to maintain their causal sequence and to avoid discarding updates which have yet to be applied. Hence, EyeCVE transmits critical events using a reliable and ordered channel and all non-critical events, such as the majority of body and eye tracking updates, unreliably.

To ensure concurrency of nonverbal signals such as eye, head and hand movement, EyeCVE bundles all updates into a single packet. Assembly and disassembly is done after and before the simulation cycle at the sending and receiving site respectively. Figure 3.2 illustrates how discrete eye, head, and hand tracking data is combined before transmission over the network, and applied to a remotely-rendered avatar. The data packet contains 8 floats for each head and hand tracking node describing 3D (x, y, and z) position, quaternion rotation (x, y, z, and w) and scale; two floats for the tracked vertical and horizontal eye angle; and 4 bytes overhead from the network layer for sequencing. Using floats with 64-bit precision (8 bytes), this leads to a payload of 148 bytes for a single avatar update message. With a typical maximum

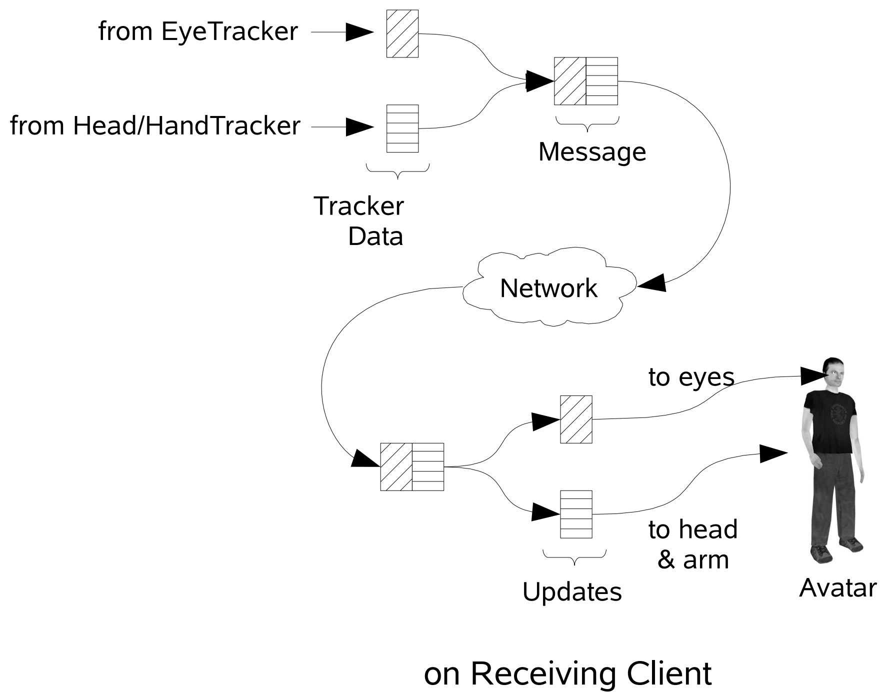


Figure 3.2: *Bundling discrete eye and head and hand tracker data in one network message.*

transmission unit (MTU) of 1500 bytes for the Internet, this fits easily into a single Ethernet packet, and critically, minimises the likelihood of delays caused by fragmentation and reassembly from intermediate underlying protocols. To reduce network traffic and thus the impact of buffer overload and computation on latency, EyeCVE only sends an update when a configurable spatial-temporal threshold has been reached. The use of spatial thresholds alone, such as typically used in dead reckoning, have been shown to induce potentially uncapped inconsistency [RMM+05].

Motion tracking is acquired and processed on two dedicated PC's at each site: one for head and hand tracking, and another for eye tracking. This follows a common approach that decouples acquisition rate from render and simulation load, and provides the ability to filter unnecessary and outlying data from the client. For instance, as described in the following subsection, the initial eye tracker produced jitter and many erroneous outliers, which necessitated the use of a low pass filter. Therefore, the decoupled approach allows filtered data, rather than the original noisy data delivered from the eye tracker, to be transmitted across the local area network at a configurable update rate to the local client. Body tracking was performed using a VRCO Trackd™ server coupled with InterSense IS-900 [Int10] head and hand sensor devices. The hand tracking device also enables navigation of the VE using a joystick and offers other interactions such as grasping virtual objects and triggering the eye tracker calibration procedure. The client reads and combines the latest tracking data from both PCs every simulation cycle.

In summary, EyeCVE's system requirements of persistence and robustness as sites join and leave, server-side logging, sender side scalability, and ease of firewall maintenance have led to the use of a central server that relays updates and maintains a local replication of the object model. A problem with this approach which is ignored in favour of performance, is that updates are not synchronised across remote sites due to varying network characteristics. For example, updates for animating the head, hand, and eyes of an avatar that were sent out by site *A* may arrive marginally earlier at site *B* than at site *C*. Thus, users at sites *B* and *C* may perceive slightly different stimuli at a specific time. However, the approach to message bundling, which combines all body tracking data into a single update, ensures that users at sites *B* and *C* will view identical stimuli, albeit at approximately, rather than precisely, the same time. In any case, this is unlikely to have an operational impact on communication.

An alternative approach to synchronisation of temporal characteristics across all participating sites involves common reference to a centralised 'time server', with a programmed delay that is long enough to assure timely delivery on each site. However, due to the coupling of avatar head position and viewpoint rendering, this reduces responsiveness of the local system, as viewpoint updates from local head tracking can only be applied after they have been distributed to remote sites. Additionally, this approach cannot ignore the impact of differing delay induced by varying computation and projection technology between sites.

3.1.3 Eye Tracking

Throughout the course of the forthcoming experimental work, two different eye tracking systems were used. During initial development, the evaluation presented later in this chapter, and the experiment presented in Chapter 4, the *MobileEye* eye tracker from Applied Science Laboratories (ASL) [Lab10] was used. The MobileEye eye tracker is head-mounted and features two cameras: one monitoring the right eye (eye camera) and one recording the scene as viewed from the perspective of the wearer (scene camera). Video feeds from both cameras are interleaved and recorded using a battery-operated digital video camera recorder (DVCR). The tapes have a typical recording duration of 60 minutes. Together with recording the video streams, DVCR also outputs gaze tracking data, in real-time, in machine-readable form on an IEEE 1394 (FireWire) cable connected to the capture PC. The MobileEye eye tracker operates with the dark-pupil principle based on the corneal-reflection of three infrared light-emitting diodes (LEDs) mounted alongside the eye camera. Image analysis determining positions of the pupil and the reflections produced by the LEDs on the pupil/corneal boundary are used to determine eyeball rotation. This eye-in-head direction is mapped to x/y coordinates with respect to the scene camera video, on which the coordinate is also superimposed, thus indicating where the wearer is looking. The coordinates are also made available on a serial port for external processing. The MobileEye eye tracker outputs data at a rate of 30 Hz (33 ms). The system was mounted on the CAVE's shutter glasses, as illustrated in Figure 3.3.

During the early stages of this research, several technical and practical issues with the MobileEye eye tracker, when operating in combination ('through') the CAVE's shutter glasses as shown in Figure 3.3, were observed: the low-level of ambient lighting typical in CAVE systems severely limited tracking ability; the eye tracker failed to robustly track people with light (blue or green) irises; contact lenses caused distortion in the corneal reflection points as viewed by eye camera; and the always-connected DVCR hindered the wearer's physical movement.

Further research into the head-mounted eye tracker market led to the purchase of the *ViewPoint*

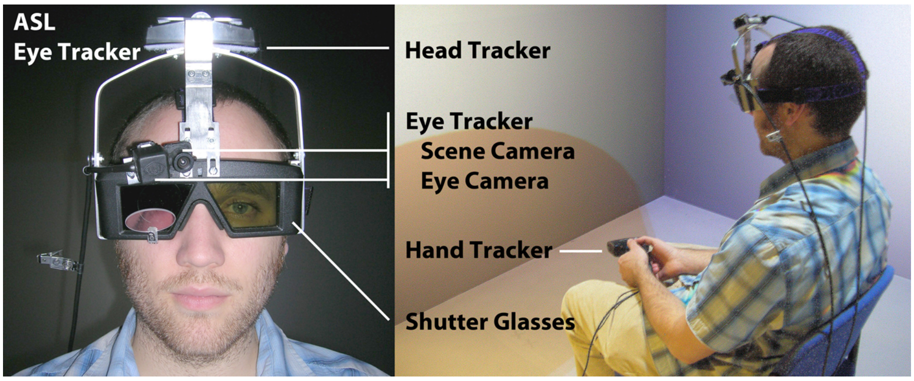


Figure 3.3: ASL MobileEye *eye tracker, and also head tracker, mounted on CAVE's shutter glasses to combine head and eye tracking. The MobileEye system is used in the initial evaluation of EyeCVE appearing in this section, and in the conversational experiment presented in Chapter 4.*

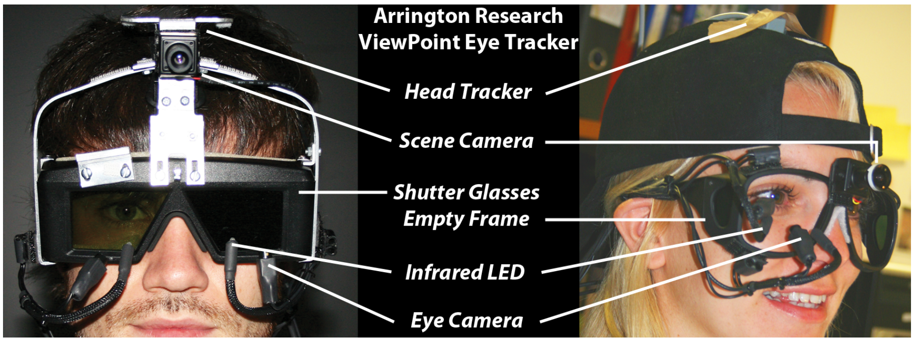


Figure 3.4: Arrington Research ViewPoint EyeTracker *eye tracker mounted on the CAVE's shutter glasses (*left*) and the WALL's empty frame (*right*). The CAVE setup is used in the object-focused experiment in Chapter 5, and for data capture in the gaze and eyelid simulation work in Chapter 7. The WALL setup is used in the truth and deception experiment in Chapter 6.*

EyeTracker system by Arrington Research [Res10a]. The ViewPoint eye tracker provided a more robust and lightweight solution than the MobileEye eye tracker, and also provided a greater update rate of 60 Hz (16 ms). Rather than using corneal reflection to track the eyes, the ViewPoint eye tracker uses infrared LEDs to illuminate the eye. The ViewPoint software then processes the eye camera's video-stream to identify the location of the pupil. The pupil's position is used to calculate x/y gaze position based on prior calibration data, and this gaze position is superimposed onto the output from the scene camera, which is streamed directly to the connected PC rather than recorded on a DVCR. Upon request, Arrington Research provided 160° wide-angle lenses for the scene cameras. This FOV approaches that of humans, which is considered to be around 180° [Cos95]. A critical feature of the ViewPoint eye tracker is its ability to output a detailed range of oculesic data including saccade velocity, fixation duration, pupil aspect ratio, and pupil size. This facilitated further integration of oculesic behaviour beyond just gaze, including blinks and pupil dilation. The ViewPoint eye tracker was used in all of the forthcoming experiments presented in this thesis, apart from those stated above. The system, mounted on the CAVE's shutter glasses, and also mounted on the WALL's lenses-free frame is illustrated in Figure 3.4.

The following three subsections detail the calibration procedures that are required before a user enters the shared VE of EyeCVE. The gaze calibration procedure is common to both the MobileEye and

ViewPoint eye trackers, while the latter calibration pair of blink and pupil dilation are only relevant to the ViewPoint system.

Gaze Calibration

In order to determine precisely what a user is looking at, some prior calibration procedure is required. During this process, the wearer looks at a series of points while the eye tracker records the value that corresponds to each eye position. An accurate and reliable calibration is essential for obtaining valid and repeatable eye movement data. Before calibration in the ViewPoint system, the pupil, as viewed by the eye camera, must be isolated with appropriate threshold settings. In the MobileEye system, the corneal reflections, in addition to the pupil, must be isolated. Both trackers start up in a coarsely-calibrated state that provides precise timing of raw eye movements. This is sufficient for many applications that utilise relative eye movements, such as quadrant-wise "preference of looking" tasks. However, EyeCVE requires more precise gaze information in order to map to avatar gaze in a meaningful way, so further calibration is required. Raw pupil and corneal reflection locations, captured by the eye camera, do not indicate where the wearer is looking in reality, as viewed by the scene camera. Hence, the purpose of calibration is to provide a mapping between the raw data and the gaze direction in the scene.

Figure 3.5 illustrates the graphical user interface (GUI) for the ViewPoint eye tracker. During calibration in both the MobileEye and ViewPoint systems, a number of points are sequentially presented to the user, who is instructed to foveate each point in turn and keeping their head still for the duration. The software then determines appropriate coefficients for the mapping. Completing the initial gaze calibration, the accuracy of a calibration should be tested by asking the user to look at particular locations in space, and verifying that the location superimposed on the scene camera video matches. The final stage of calibration is then performed in the 3D VE of EyeCVE. This process is presented in Section 3.1.4 covering the avatar subsystem, and provides a further mapping to enable avatar gaze replication. Gaze was tracked, displayed, and analysed in all three telecommunication experiments presented in Chapters 4, 5, and 6.

Blink Calibration

The ViewPoint eye tracker is capable of outputting a range of oculesic data. This includes pupil aspect ratio: a dimensionless value, with 1.0 indicating a perfect circle. When humans blink, the eyelid descends to cover the eye before rising again to approximately its original position. Correspondingly, as viewed by the eye tracker's eye camera, the elliptical fit to the pupil becomes increasingly flat before it disappears for a brief time, and subsequently reappears and returns to its circular outline following the blink. This characteristic change in the aspect ratio of the elliptical fit to the pupil can be used to detect blinks, which can be classified as the pupil aspect ratio crossing below a particular threshold. Rather than dynamically monitor eyelid position during a blink, which would involve significant image processing analysis, the pupil aspect ratio signal is then used to initiate a simulation of human blink dynamics, which is the focus of Section 7.2. Figure 3.6 shows views captured from the eye camera as the pupil aspect ratio crosses below the threshold specifying a blink. Experience with the system established that a threshold of around 0.6 provides robust blink detection in the majority of wearers. Blinking was tracked, represented, and

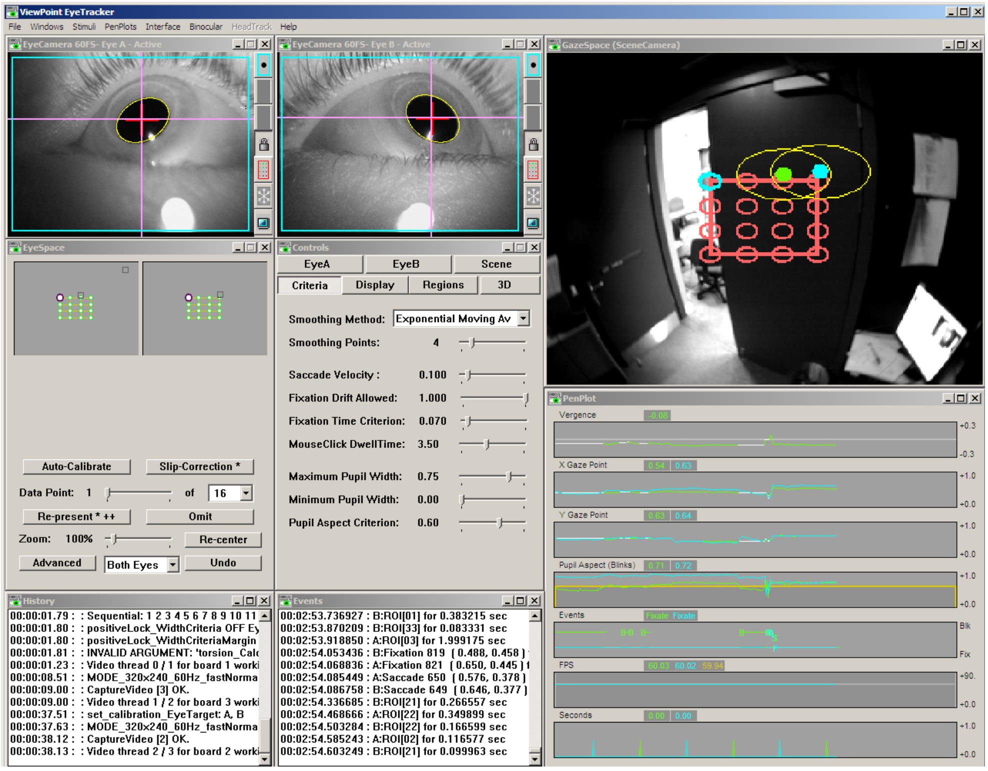


Figure 3.5: *ViewPoint GUI. The calibration procedure maps the eye position as observed by the eye camera to the gaze direction superimposed on the scene camera view.*

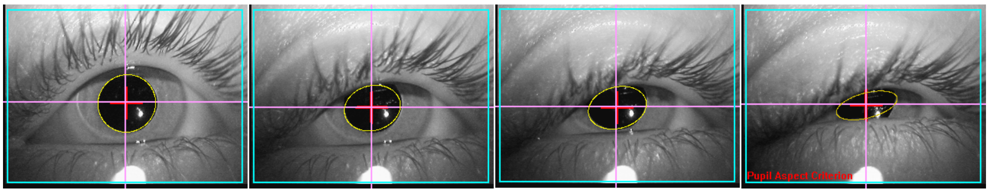


Figure 3.6: *Views captured from the ViewPoint eye camera during the down-phase of a blink. From left to right, the tracked pupil aspect ratio gradually flattens until a threshold is met, signalling a blink.*

analysed in the truth and deception experiments presented in Chapter 6.

Pupil Dilation Calibration

The primary input to gaze calculation is the position of the pupil within the eye camera's stationary view. Along with position, for gaze, and aspect ratio, for blinks, pupil size is also monitored by the ViewPoint eye tracker. The ViewPoint eye tracker delivers x and y values, normalized with respect to the eye camera's video dimensions which have a 4:3 aspect ratio, to represent pupil size. Thus, the x- and y-values are anisotropic, and the y-value must be are initially rescaled by 0.75 (due to the difference in aspect ratio) before computing the area of the pupil ellipse using the formula: $area = \pi XY$.

The above calculation defines a user's current pupil size, but does not provide information revealing the extent of constriction or dilation with regards to a particular individual's natural pupil size range. In order to position the measured pupil size in this meaningful context, a wearer's natural range, between extreme constriction and extreme dilation, must first be established. These responses are elicited by triggering the pupillary light reflex. First, the wearer is placed in a high-luminance environment to trig-

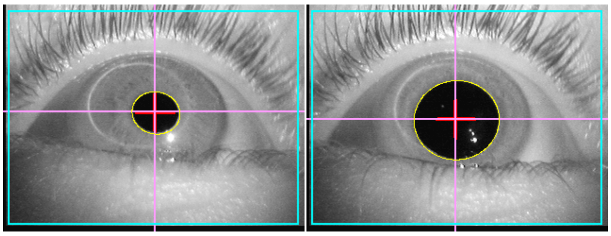


Figure 3.7: *Views captured from the ViewPoint eye camera during pupil size calibration.* Left: *extreme constriction in high luminance lighting.* Right: *extreme dilation in darkness.*

Figure 3.8: *Basic avatar model. Avatar is used in the conversation and object-focused experiments in Chapters 4 and 5.*

ger constriction. This is followed by placing the wearer in dark surroundings, triggering full dilation. During this calibration procedure, wearers are instructed to maintain a forward gaze direction, and to keep blinking to a minimum. During exposure to both light and dark surroundings, the wearer's pupil size is monitored for 30 seconds. The minimum (in light) and maximum (in darkness) pupil sizes are used to establish the individual's range. Hence, following this calibration procedure, practical minimum, maximum, and current pupil sizes are known. These measures may be used to determine the current dilated state of a wearer's pupil, which acts as the primary input to the avatar pupil animation system, detailed in the following section. Figure 3.7 shows views captured from the eye camera during pupil dilation calibration. Pupil size is tracked, represented, and analysed in the truth and deception experiments presented in Chapter 6.

3.1.4 Avatar Subsystem

In EyeCVE, each user occupies a distinct VR system which are all connected across the Internet, and in which, avatars of remote users are displayed. In the conversation and object-focused experiments presented in Chapters 4 and 5, three CAVE systems are connected, while in the truth and deception experiment presented in Chapter 6, a CAVE system is connected to a WALL system. These technologies immerse users in a VE while allowing them to see and move their own bodies, thereby maintaining natural proprioception. In contrast to looking into an environment in which users see themselves as an avatar from either a first- or third-person perspective, this reduces motion sickness [HR92] and increases

Figure 3.9: *Male and female advanced avatars. Avatars are used in the truth and deception experiment presented in Chapter 6 and the simulation work presented in Chapter 7.*

sense of presence. Furthermore, social use of gaze is largely irrelevant to anything but an embodied first-person perspective. For these reasons, EyeCVE does not render a user's own avatar locally, but rather, tracks the user's head, hand, eyes, and voice to drive a simulation of their activity and attention in an avatar that is displayed at fellow interactants' sites. There are two distinct classes of avatars that have been developed and utilised throughout the reported work. The classes are termed *basic avatars* and *advanced avatars*, and they are separated by their capacity for nonverbal expression.

Basic avatars are capable only of gaze behaviour, while advanced avatars have additional facial animation capability to simulate eyelid dynamics, pupil dilation, and mouth movement. Figure 3.8 illustrates the basic avatar. The geometric model is implemented as polygonal 3D mesh resembling a human male and is mapped with photographic textures. The eyes are separate sphere objects, which allows for independent rotation. The eyes also feature a high-contrast texture map to increase visibility of gaze direction at a distance.

The advanced avatars include both male and female versions, and are illustrated in Figure 3.9. These avatars are capable of eyelid animation, mouth movement, and pupil dilation. The facial animation is achieved using blend shapes (also known as morph targets) as illustrated in Figure 3.10. The animation system linearly interpolates between the available meshes by translating vertices in order to provide accurate temporal control of expression. This is critical in particular for rapid eyelid motion.

The following five sections detail EyeCVE's avatar animation system. The first three sections echo the previously-documented eye tracking calibration procedures, detailing how tracked gaze, blinking, and pupil dilation data is processed to animate avatars in real-time. Avatar head and hand kinematics,

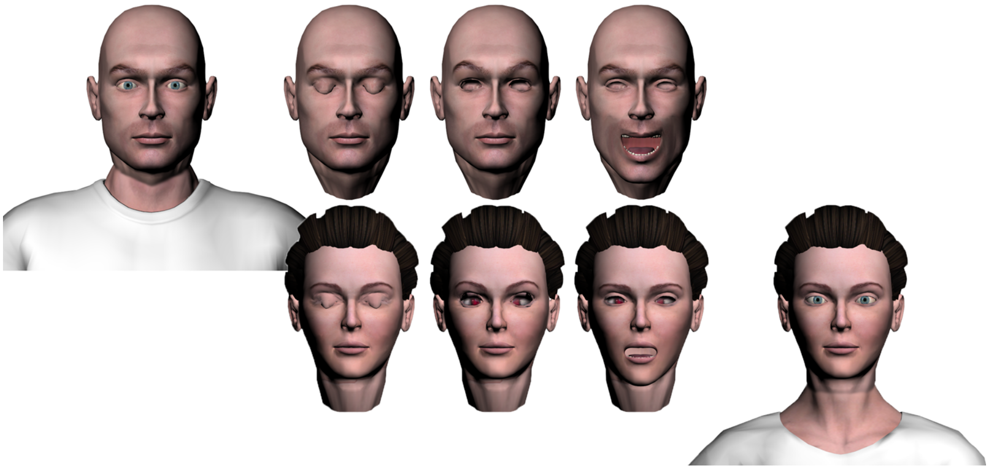


Figure 3.10: *Examples of blend shapes for male and female advanced avatars. Eyelid motion, pupil dilation, and mouth movement is animated by passing tracking data to the avatar animation system, which blends between poses to approximate users' nonverbal behaviour.*

driven by tracking, is then outlined, followed by mouth movement which is synchronised with a user's vocal signals. Gaze, head, and hand components of avatar animation are relevant to both the basic and the advanced avatars, while animation of the eyelids, pupil dilation, and mouth movement are relevant only to the advanced avatars.

Avatar Gaze

Following initial eye tracker calibration, detailed in Section 3.1.3, the rotation of the tracked eye, relative to the position and rotation of the head, is recorded. The eye tracker outputs 2D, x and y, gaze direction at a specified refresh rate to a client running EyeCVE. In order to be represented correctly during an EyeCVE session, in this 2D gaze data must be mapped to corresponding angles of avatar eye rotation. Thus, while users are granted free movement, the position of the head and eye trackers must remain stationary, relative to one another, in order to provide meaningful data. This is achieved by mounting both trackers on the user's head as illustrated in Figures 3.3 and 3.4. The geometric offset between the position of the origins of the avatar's eye and head models in the VE, relative to the position of the tracked user's eye and head must also be known. This information is specified in a configuration file when executing EyeCVE.

The calibration procedure within the 3D VE of EyeCVE is similar to the initial eye tracker calibration. A plane, featuring a 5×4 array of 20 calibration points, as illustrated in Figure 3.11, is presented, and the user must foveate each point in turn. However, unlike the initial gaze calibration, the user is not required to keep their head stationary during this procedure, as the calibration plane is 'attached' in front of the user's tracked head position, and moves with their head. The user gazes at each point as it is highlighted, and the x and y angles are used to rotate the avatar's eye geometry accordingly. As only single eye, usually the right, is tracked, vergence of the non-tracked eye is performed by casting an intersection ray from the origin of the tracked eye into the VE to establish the virtual hit-point. Subsequently, the horizontal rotation angle required for the non-tracked eye to focus on the same position in space is calculated. The hit-point delivered by the intersection ray is also used for post-interactional analysis of

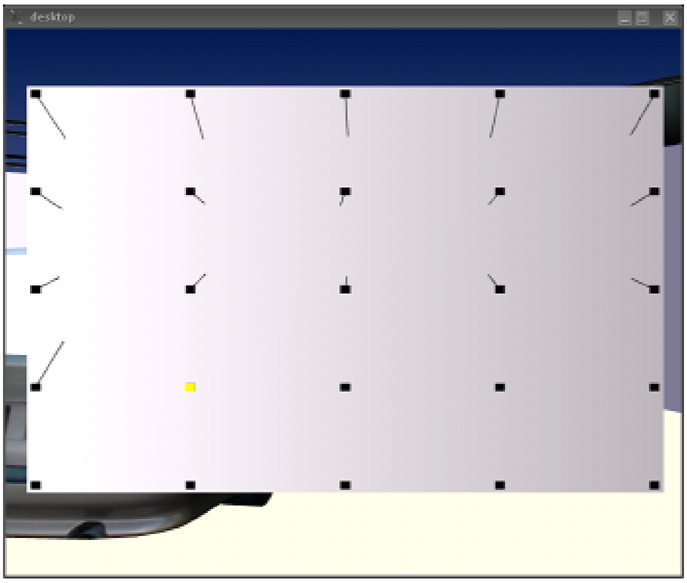


Figure 3.11: *Calibration plane in EyeCVE which maps the initial eye tracking calibration to the avatar eye movement in the VE, including vergence.*

gaze in the forthcoming experimental chapters.

Avatar Eyelid Kinematics

Chapter 7 details parametric models of lid saccades and blinks, and presents two associated user studies designed to validate and assess the impact of bestowing avatars with physiologically-accurate eyelid dynamics. While the blink and lid saccade models treat these two characteristic motions of the eyelids separately to an extent, the high interrelation between the two behaviours, and critically, the fact that they both control identical components of an avatar, stipulate that they must act concurrently and with knowledge of one another. A state-diagram of the combined models as implemented in EyeCVE is presented in Figure 3.12. Data may be streamed from an eye tracker or, as the diagram suggests, from a potential variety of sources, acts as primary input to the two eyelid models.

In the absence of a blink signal, the lid saccade model remains in operation, animating the eyelids according to vertical shifts in gaze. At each iteration of the lid saccade mode, the input source is checked for the presence of a blink signal, indicated by the specified pupil aspect ratio threshold as discussed in Section 3.1.3. When a blink is detected, the current state of the lid saccade model is saved, and the blink model takes control of eyelid animation. Following completion of the blink, the state of the lid saccade model is restored before resuming control. Blink motions are informed by the open-angle of the eyelids when a blink is initiated, which, in turn is dependent on the final position of the most recent lid saccade motion. Hence, while the blink and lid saccade models are independent, they must also act with knowledge of one another to provide realistic combined behaviour. Figure 3.13 presents a sequence of images taken during a sample modelled blink motion.

Avatar Pupil Dilation

As documented in Section 3.1.3, during initial eye tracker calibration, the pupillary light reflex is used to establish the natural range of an individual's pupil size: from constriction due to bright light to extreme

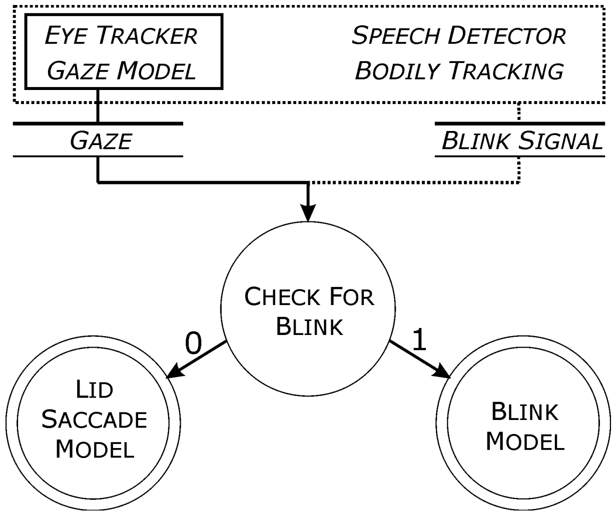


Figure 3.12: *Sample implementation of combined lid saccade and blink models using multimodal detection methods. The blink and lid saccade models operate with the advanced avatars, and are studied in Chapter 7.*

Figure 3.13: *Images taken during a model-generated blink animation displayed by the female advanced avatar.*

dilation in darkness. These two extremes are mapped to two avatar eye geometry models representing constriction and dilation, as illustrated by the female advanced avatar in Figure 3.14. During an EyeCVE session, the size of each user's pupil is continually tracked. This size will lie within the recorded range, and acts as input to the avatar animation system, which blends accordingly between the models to represent the current dilated-state of a user's pupils. In general, some intermediary size between the two extremes will be displayed. Pupil dilation is investigated in the truth and deception experiments presented in Chapter 7.

Avatar Head and Hand Movement

In the CAVE systems used in the majority of the experimental work, head and hand tracking was performed using InterSense IS-900 [Int10] sensor devices and VRCOs Trackd library. Typical update rates of the devices is around 100 Hz (10 ms). In the WALL system, a Polhemus Fastrak [Pol10] magnetic tracking system coupled with the Virtual Reality Peripheral Network (VRPN) library [TRH+01], was used. The Fastrak device has similar update rates to the InterSense IS-900.

The approach, popularised by Badler et al. [BHG93], of estimating the orientation of an avatar's body and arm positions, from those of the tracked head and hand positions was adopted. A user's head position and orientation is mapped directly to the motion of their avatar's head, while body rotation is performed once a defined heading threshold is met. Similarly, a user's hand position and orientation has a direct mapping to their avatar's hand, while an inverse kinematics (IK) algorithm is used to articulate appropriate upper- and lower-arm positions and orientations. Example synthesised arm poses,

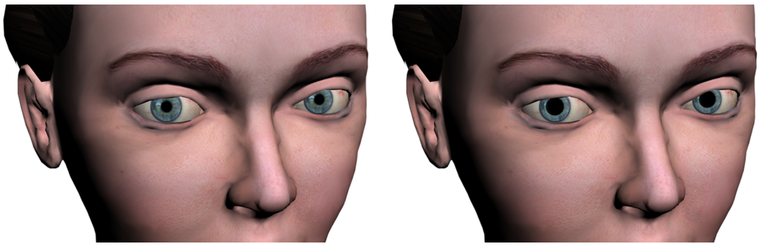


Figure 3.14: *Extremes of pupil constriction (left) and dilation (right) as displayed by the female advanced avatar. An individual's current pupil size acts as input to the avatar animation system which blends accordingly between the two extremes.*

Figure 3.15: *Arm poses generated by an IK algorithm, which uses hand tracker position and orientation as input.*

represented by the female advanced avatar are presented in Figure 3.15.

Avatar Mouth Movement

Mouth movement, synchronised with a user's vocal signals, is an additional feature of the advanced avatars animation. To ensure high performance for real-time display, simple generic mouth motion was used. A more computationally-complex alternative would be to generate mouth movement based on a mapping between phonemes, a unit of audible speech, to visemes, the visual positions of the mouth. Vocal utterances of users are detected by wireless microphone headsets. A volume-thresholding algorithm, implemented using OpenAL [Spe10], is used to filter audio signals that are unlikely to be intentional vocal utterances emitted from local users. The animation system then generates avatar mouth movement by manipulating blend shapes as illustrated in Figure 3.10. Experience with the system established the most visually-natural duration of a mouth open/close cycle to be around 0.2 seconds, which approximately mapped to syllabic utterances from users. Figure 3.16 shows screen captures of the male advanced avatar's mouth during mouth animation.

3.1.5 Evaluation

Detailed measurements and results of EyeCVE's performance in a two-CAVE link are now presented. The data provides insight into the resulting system performance from a minimum-load setup. It should be noted that this evaluation was performed with an early version of EyeCVE, and that the initial (lower refresh rate) MobileEye eye trackers were used rather than the high-performance ViewPoint eye trackers. Additionally, the test was performed using basic avatars rather than advanced avatars featuring facial expression. Subsequent development of EyeCVE's rendering and network components have significantly increased performance: particularly regarding the number of updates in avatar animation frames per

Figure 3.16: *Opening phase of mouth movement performed by male advanced avatar upon detection of audio input from microphones worn by users. Closing phase reverses the sequence at a faster rate.*

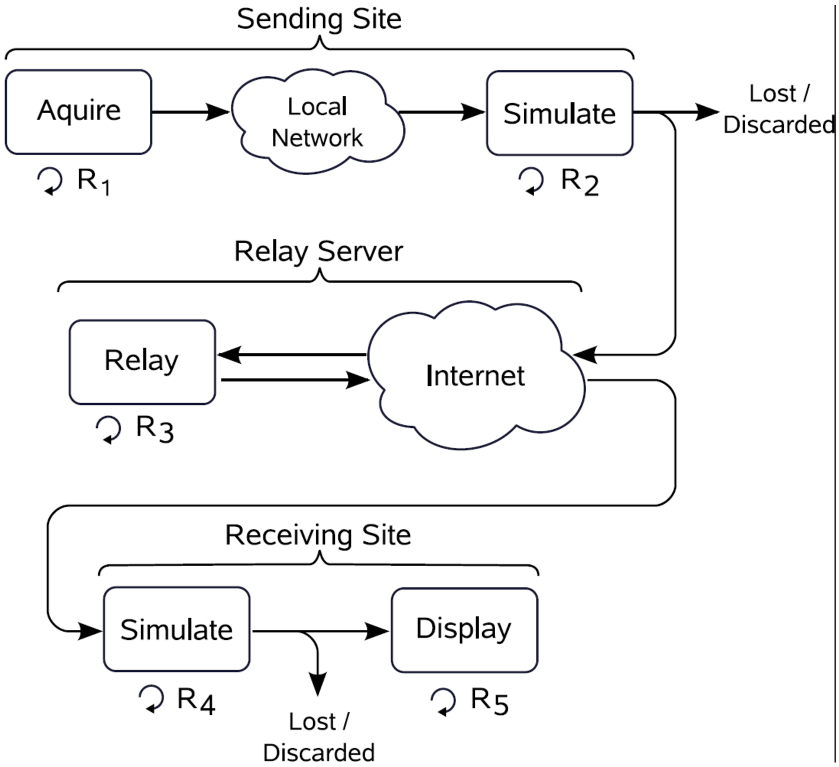


Figure 3.17: *Processing pipeline from sending to receiving site. $R_1 - R_5$ represent the various simulation cycle rates at each stage.*

second between remote sites. Nevertheless, the following evaluation of the system is still relevant, providing a baseline performance measure that satisfies the temporal and spatial requirements of a tracked gaze AMC system specified in Section 3.1.1.

This evaluation is particularly concerned with the time taken to process and distribute gaze data, as well as characteristics of loss and discontinuity of the synthesised information. To analyse the data, avatar update traffic was logged at various stages within the processing pipeline when sending avatar updates from one site to another. Figure 3.17 illustrates the processing pipeline of the system. Gaze data is firstly acquired by the eye trackers and sent over the local area network to the local client. The data is then applied to the local object simulation, calculating the local avatar's eye angles, including vergence, with respect to head orientation. The resulting angles are sent as an update message over the Internet and passed via the relay server to the receiving client's object simulation. The message is then applied to the respective avatar, before rendering at the receiving site's display.

Acquisition and simulation processes run on separate machines, and thus may run at different speeds. The filtered eye tracking acquisition rate, R_1, is variable, and for this evaluation, was set to the following rates over a series of sessions: 5 Hz (200 ms), 10 Hz (100 ms), 20 Hz (50 ms), and 50 Hz (20 ms). The MobileEye eye tracker supports a maximum update rate of 30 Hz, so in the case of 50 Hz, the effects of driving the device above its specification are investigated. R_2 represents the object simulation rate of the sending client, and R_4 represents that of the receiving client. The actual display is refreshed at rate R_5, independent of the object simulation cycle. In this evaluation, both clients operate on very similar machines, so only minor performance differences are expected. Hence, no investigation into varying object simulation and display rates is performed. R_3 was also left constant.

The procedure of each of the five test sessions was as follows:

1. Configure the specified update rate on eye tracker driver;
2. Connect the two EyeCVE clients and calibrate gaze;
3. Users asked to walk around and look at a set of virtual objects laid out on a virtual table;
4. Users take turns looking and pointing towards specific objects while partner who follows direction of gaze;
5. Log all incoming and outgoing updates locally at each site.

Occurrences of avatar updates at time of acquisition, sending simulation, and receiving simulation were logged. Each log entry included a time-stamp, object ID, and current position coordinates. In order to identify individual updates within the log files, each entry was tagged with an incrementing frame number. During the tests, threshold filtering was activated, which discarded tracking updates featuring less than 3 cm or 1° difference from the last transmitted update. The ordering mechanism discarded obsolete updates, as described in Section 3.1.2. This implies that packets may be discarded due to redundancy filtering or ordering, making the measurements dependent on the specific task and activity of the interactants. However, these preliminary results are sufficient to demonstrate typical system characteristics.

The hardware environment consisted of two CAVE systems with dimensions of $3m^3$. The stereo projection screens have a resolution of 1024×786 and a refresh rate of 96 Hz. Computation was provided by an *SGI Prism* system at each site. The eye trackers were connected to *Intel Pentium D* or *Core Duo* PCs, equipped with 1Gbit/s Ethernet cards. The relay server hosting the EyeCVE sessions ran on the SGI Prism at one site. The test VE consisted of a single room with nine objects arranged in three rows over a table. The objects were positioned approximately 1.5 metres from the CAVE floor (i.e. at approximately eye height of both users and their avatars). The complexity of the entire scene was approximately 24000 polygons.

Results

Four sessions were conducted, each with a specific update rates set on the MobileEye eye tracker driver. The duration of each session, from connection to disconnection varied between 3–5 minutes due to differences in time taken for setting up and calibrating the eye tracker in EyeCVE. The measurements presented in this section refer to values collected over a 100 second period during each session in which the gaze task was ongoing.

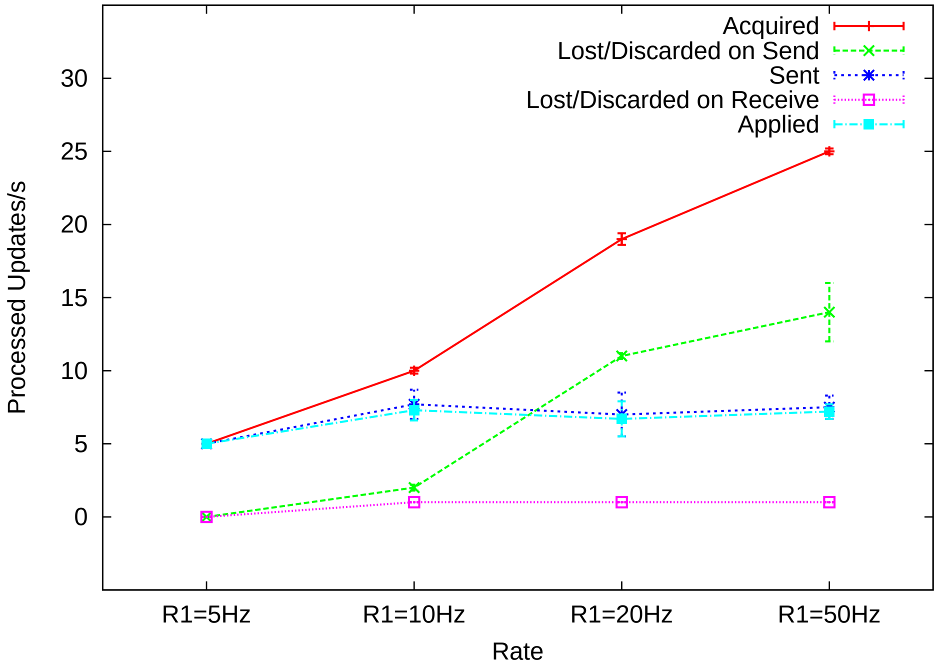


Figure 3.18: *Frequency (occurrences) of updates that passed through the system at various stages.*

Figure 3.18 shows the frequency of updates that passed through the processing pipeline at acquisition, distribution and representation with varying R_1. The graph also shows the frequency of occurrences where updates that were lost in transit over a network, or discarded due to redundancy or ordering during object simulation, at either the sending or the receiving site. The figure demonstrates how new updates were generated at the selected acquisition rates. However rate of updates that were actually sent and applied did not increase beyond 8 Hz. An acquisition rate that is higher than the simulation rate results in updates being queued, and subsequently superseded and discarded. An acquisition rate close to that of the simulation rate yields a throughput of 80% of gaze updates, while approximately 15% were lost or discarded after acquisition, and 5% were lost during, or discarded after, transmission over the Internet. Almost every update message that was sent from the sending client was received in the correct order at the receiving client.

Figure 3.19 shows an example of latency measurements taken during the session in which R_1 was set to 20 Hz. It shows how the resulting end-to-end latency of the system is composed of the delays induced by the individual modules within the processing pipeline. The measured times are as follows: T_1 is the time it takes to deliver a message containing new eye tracker data from the MobileEye to the sending client over the local intranet, including the time for sending and the time that new data resides in the receiving buffer before it is read by a simulation call; T_2 is the time it takes the simulation process on the sending site to calculate gaze, plus other minor management tasks; T_3 is the latency for sending update messages over Internet; T_4 is the time it takes the object simulation on the receiving site to read new updates from the receiving buffer, apply the update to the avatar object, and update the scene graph just before the renderer is called. It is evident from Figure 3.19 that the time taken to acquiring eye tracker data and apply the simulated data to the avatar on the receiving site comprises a smaller part of

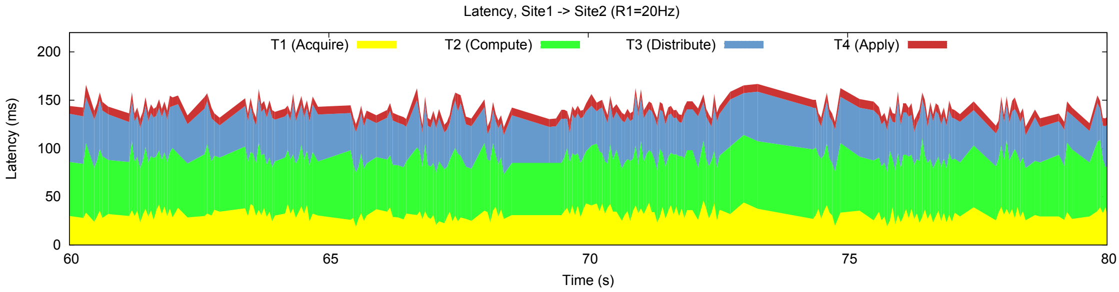


Figure 3.19: *Extract of latency measurements in condition R_1 = 20 Hz. End-to-end latency (from acquisition to application) is composed of delays within the various stages of the processing pipeline.*

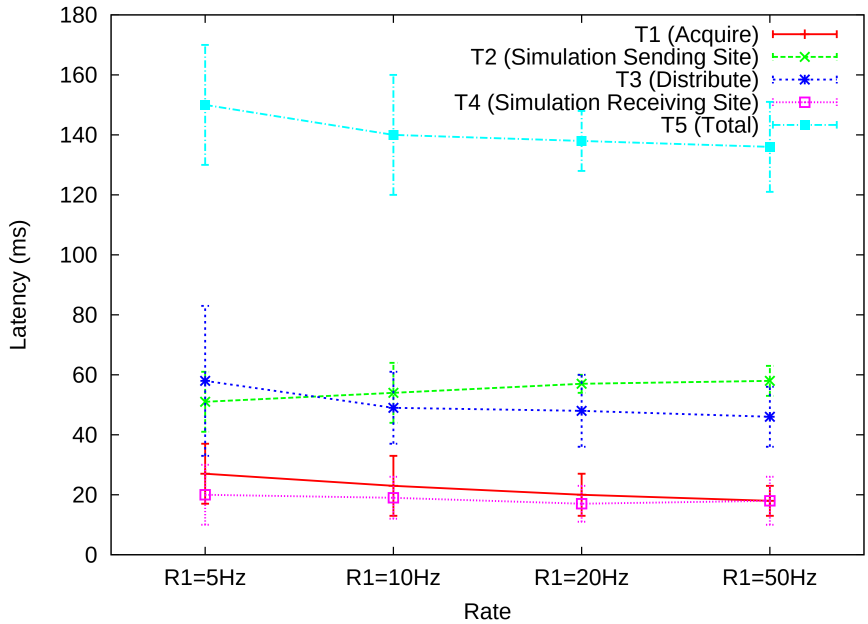


Figure 3.20: *Mean latency caused by stages of the processing pipeline.*

the total latency than computation of gaze and distributing it over the Internet.

Figure 3.20 shows the mean latency occurring throughout the processing pipeline. Computation during simulation on the sending site appears to increase slightly with higher acquisition rates, and thus, with higher data load. In contrast, average latency for transmitting data over a network, both in the local network during acquisition, and across the Internet when distributing to remote sites, reduces slightly with increasing acquisition rate. This stage appears to have the greatest effect on the resulting total end-to-end latency, which ranged between 136–150 ms.

3.1.6 Summary

From a technical perspective, a key benefit of AMC in ICVEs as a telecommunications platform is the ability to transmit the visual and nonverbal, component of communication with very low bandwidth requirements in comparison to video-based systems. The implementation of EyeCVE demonstrates that the communication of users' natural kinesic behaviour, such as gesture and gaze, requires only the continuous transmission of various tracked coordinates: a fraction of the data required for live streaming video or dynamic point-based video reconstruction such as the blue-c system [GWN+03].

Another aspect of managing bandwidth is related to locating computation strategically. EyeCVE adopts the approach of computing the gaze of each local avatar at the local site, and transmitting updates describing eye rotation to remote sites. An alternative approach would be to transmit the raw eye tracker data to remote clients in order for the resulting avatar eye angles to be calculated at each receiving site. However, in a multiparty collaborative session, this method could conceivably become a limitation which drastically increases computational demand at all sites. Hence, the approach described in this section locates the computationally-intensive gaze and hit-point calculation at each sending site, distributing only key values and thereby minimising overall computation and network load.

The advanced avatars introduced in this section are able to express the major human oculesic behaviours, including gaze, pupil dilation, and eyelid dynamics, together with mouth movement based on vocal input. The animation computation of the advanced avatars is designated to the sites displaying a particular avatar, rather than the locally-controlling site, which is responsible only for transmitting signals which may initiate a particular animation such as a blink, lid saccade, or mouth movement. This decision was made due to the nature of the parametric animation models, which demand rapid response in order for the motions to be displayed correctly (for instance, the duration of human blink is around 200 ms [PK27]). Therefore, to ensure correct performance of the animation models, following an initiating signal from a sending site, the computation involved with actually displaying the animation animation is performed exclusively at the displaying site.

EyeCVE satisfies the spatial aim of enabling communicational gaze while allowing free movement of participants within a shared immersive VE. This was achieved by integrating a head-mounted eye tracker with the existing head tracking and shutter glass technology. The quality of free movement is particularly investigated in the object-focused experiment presented in Chapter 5. The temporal requirement of EyeCVE was to reproduce eye movements quick enough ($<$200 ms) and often enough ($>$10 Hz) in order for users to be able to both signal with their own gaze, and interpret the focus of others' gaze, along with communicating other verbal and nonverbal cues. EyeCVE's object simulation rate was measured at 10–15 Hz. However, the changes in eye movement above the level of filtering did not exceed 8 Hz. An end-to-end latency around 140 ms was achieved.

Once again, it should be noted that this evaluation was performed with an early version of EyeCVE. Subsequently, the integration of the ViewPoint eye trackers, featuring high refresh rates of 60 Hz, and improvements to both the rendering and object simulation model, has significantly increased performance at all processing stages. While a thorough investigation of performance has not been repeated with the mature version of EyeCVE, a conservative estimate of gaze update rate, made during the truth and deception experiment presented in Chapter 6, is 20 Hz. In the current evaluation, synchronisation of concurrent nonverbal signals for each avatar was achieved by bundling all movement information into a single packet during each object simulation cycle: whenever any movement exceeded the threshold of 3 cm or 1°. Subsequently, with the improved performance of the object simulation rate (now around 30 Hz), this movement sensitivity has been increased dramatically to 0.1 cm and 0.1° respectively. Extensive use of the system with these highly responsive tracking thresholds has produced no ill effects with regards to network congestion or overall system performance.

The following section addresses the capture of interactional data, describing a users' natural nonverbal and verbal behaviour during interaction in EyeCVE.

3.2 Multimodal Data Collection and Analysis

When engaged in collocated social interaction, people use a wide gamut of nonverbal behavioural cues to express themselves, respond to others, and emphasise the verbal component of communication. Originating in the fields of anthropology and sociology, *interaction analysis* is a powerful tool in the investigation of human activities, including verbal and nonverbal interaction, and the use of artefacts and technologies. Interaction analysis aims to identify routine practices and problems and the resources for their solution [JH95]. Typically, interaction analysis is performed through review of video recordings, allowing for general and, depending on video fidelity and camera position, detailed observation of nonverbal cues. It may be argued, however, that this 'observer's perspective' method of data collection fails to record interactants' behaviour in a quantifiable manner, thereby limiting the ability for analysts to measure and explain the unfolding interaction. This method of data collection in unlikely to change when investigating collocated interaction in the real-world, as the capture of peoples' natural behavioural channels require specialised wearable equipment that would be considered intrusive and impractical. However, when engaged in AMC in ICVE systems such as EyeCVE, devices used to track users' natural behaviour are a prerequisite of system usage. Tracking devices used in immersive VR systems are generally used for real-time input, such as head tracking to inform both perspective-correct stereo rendering [CNSD93] and avatar position [BBF+95], and hand tracking to interact with the VE [BH97] and drive avatar gesture [BHG93]. However, an orthogonal application of user tracking, which is central to the development of both single- and multi-user VEs as a medium, is precise data collection, which may be subsequently used to analyse users' behaviour, response, and performance during AMC.

Experimental analysis of user behaviour in VEs is well-represented in the literature, much of which has been detailed in Chapter 2. A typical study will employ one or more means of data collection: *questionnaires* assessing users' subjective sense of presence [SU93] and experience of a VE [RWOS03]; observation of characteristic *behaviour*, which is indicative of a user's level of involvement in a VE, and includes body posture [PS07], verbal signals [PSB02], proxemics [YBU+07], and in the forthcoming experimental work, gaze and blink rate; *performance* measures which aim to provide objective data describing interaction in VEs, and are particularly suited to assessing the success of collaborative tasks [SSA+01, SRSH05]; and *physiological* measurement of response to stimuli in the VE, which include heart rate [MIWBJ02], galvanic skin conductance [HEHR+08], and in the forthcoming experimental work, pupil dilation. Over the course of the experimental work in this thesis, metrics from each of these categories are employed at various stages.

While the objective methods (behaviour, performance, and physiological response) described above are able to elucidate particular elements of the use of VEs, they are often collected and analysed independently of one another, with imprecise or no temporal alignment. Conversely, human behaviour, and particularly social interaction, is an intricate and multimodal activity, in which a range of highly interrelated verbal and nonverbal communicative signals are produced and responded to in both temporal and spatial dimensions.

In this section, the problem of multimodal data capture describing a user's natural nonverbal and verbal behaviour during interaction in ICVE systems is addressed. In EyeCVE, a user's performance is captured using several tracking devices, each monitoring a separate nonverbal or verbal cue. These data input streams include eye tracking to measure oculesics, including gaze, blinks and pupil size; head and hand tracking which measure gross body movement including head orientation, gesture, and interaction with objects; audio input detecting vocal signals; and additional arbitrary markup data critical to an experimenter in an experimental scenario. The central challenge of capturing such multimodal behaviour so as to be meaningful for analysis is collating the data in a manner which preserves the temporal characteristics of the synchronous input streams. To this end, the approach described in this section considers eye tracking to be critical, and is used both as a central component of the interactional data, and as a device able to record precise tracking and timing data in a readable format. Consequently, the method capitalises on the logging ability of eye tracker software, by directing all data output streams from other tracking devices to the eye tracker logging process, which collates the data in a single log file at a high update rate. In this way, multiple components of users' natural nonverbal and verbal behaviour during AMC in EyeCVE are able to be recorded, in a manner which preserves the intricate temporal characteristics of original action.

This section follows with a brief discussion of the predominant video-based approach to interaction analysis, which leads to requirement specification of the proposed approach to multimodal data capture in EyeCVE. The process and technical architecture, supporting collation of an individual's natural behavioural channels through EyeCVE, are then described. Two data extracts, recorded from the truth and deception and object-focused experiments, presented in Chapters 6 and 5 respectively, are then explored. These analyses emphasise the interrelation between the various components of captured interaction, and demonstrate how, through combined, multimodal analysis of these behavioural components, additional explication of human activity may be revealed.

3.2.1 Requirements

Goodwin and Heritage consider social interaction as the primordial means through which the business of the social world is transacted, the identities of its participants are affirmed or denied, and its cultures are transmitted, renewed, and modified [GH90]. Shared meaning, mutual understanding, and the coordination of human conduct are achieved through processes of social interaction. Face-to-face collocated interaction is recognised both as the central form of social interaction, and as a strategic site for the analysis of human action, as interactants have at their disposal a full range of verbal and nonverbal resources. The general domain of interaction analysis is defined by Jordan and Henderson as an interdisciplinary method for the empirical investigation of the interaction of human beings with each other and with objects in their environment [JH95]. Interaction analysis examines human activities, such as talk and nonverbal behaviour, in an attempt to describe routine practices, structure, and sequential patterns throughout a unfolding interaction.

Only electronic recording produces the kind of data corpus that allows close interrogation required to perform the process of interaction analysis [JH95]. Hence, development of video technology has been instrumental to the evolution of a field which depends on audiovisual recording for its primary records, and on playback capability for analysis. It is rare to find examples, even in the contemporary literature, in which the analysis of collocated interaction has been performed with any other technology apart from video: Vertegaal et al. report on gaze behaviour recorded by eye tracking in round-table conversation [VSvdVN01], and Keysar et al. analyse people's tracked eye and head movements when following a confederate's instructions to manipulate objects [KBBB00]. Compared to these approaches of tracking communicational channels with specialist devices, there are several drawbacks to the traditional video-based methods of recording and analysing interaction. Firstly, interpreting video is an intensely manual process which must generally be performed in real-time (as the video plays), and incorporating frequent pauses and replay of action. This process results in analysis which is clearly imprecise in both temporal and objective terms: recording and playback devices are prone to minor timing errors, while observation and comprehension of action is difficult to quantify. Furthermore, a critical drawback of video-based approaches is the generally static positioning of the camera amidst the highly dynamic and spatial nature of human interaction [JH95]. Although this issue may be alleviated somewhat by the use of additional cameras, this has the compounding effect of escalating the amount of data required to be analysed by this highly manual process.

While video recording must take an exocentric perspective of interactants and their behaviour in order to preserve the context of the social interaction, wearable tracking devices common to VR systems able to record components of human behaviour in a quantitative and egocentric manner. Cues that such devices are able to record include bodily movement (either partially, using a single sensor, or in full, using multiple sensors or motion capture), oculesic behaviour (through eye tracking), vocal signals (using audio microphones), and physiological response (including heart rate and galvanic skin response). In a scenario in which an individual's multiple behavioural channels are recorded synchronously and from several sources, the central requirement of multimodal data collection is the processing and collation of the data streams in a manner which preserves the temporal and interrelated nature of the action.

A typical ICVE system distributes computation and tracking interfaces over multiple computers and locations. Hence, a central process that is accessible to remote machines over a network, that is responsible for both collation and output of all data streams is required in order to support such distributed architecture. To be conducive to subsequent processing and analysis, this output log file must also be human readable, preferably with data elements appearing with consistent formatting, using rows and columns. In this way, an individual's natural behaviour in an ICVE system can be preserved with high-fidelity and temporal consistency to enable multimodal interaction analysis.

3.2.2 Capture Architecture

Figure 3.21 presents an overview of the approach to the collation of multiple data streams, and will be the focus of this explanatory section. The diagram features six interrelated columns, each representing a stage of the information processing pipeline. This pipeline progresses from left to right, and accordingly,

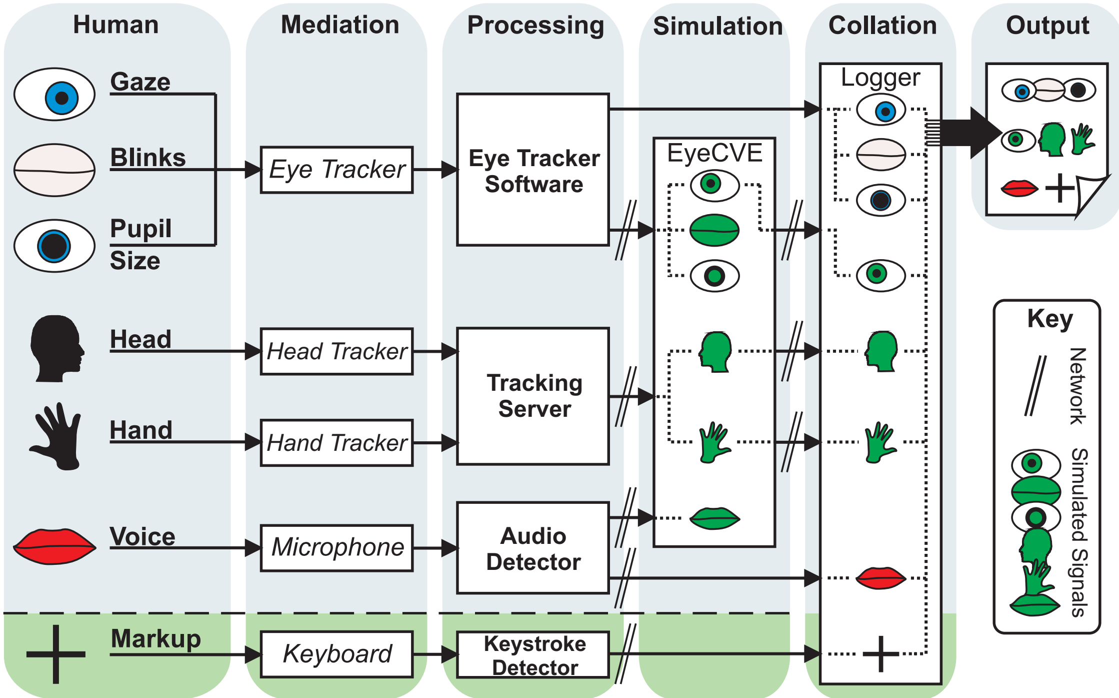


Figure 3.21: *Multimodal data collection pipeline. The six columns describe the flow of information, from left to right. Initial expression of* Human *behaviour is captured by* Mediation *devices. These signals are synthesised and* Processed *by software running on machines local to each device. Input channels relevant to EyeCVE's real-time avatar animation system are passed to EyeCVE for* Simulation *and display at remote sites. The specific simulated data streams (green icons) of gaze, head, and hand movement are then transferred across a network to the logging system running on the machine local to the eye tracker. This data is* Collated, *along with raw gaze coordinates, blinks, pupil size, verbal signals, and additional markup data. These raw data streams all have inherent meaning outside of the 3D context of EyeCVE's VE, so do not require simulation. Finally, all data is* Output *in human-readable form, together with timing information, to a single line of a log file, at a rate of 60 Hz (16 ms).*

from initial communication performed by a human user, to the final collated output. Each stage of the pipeline will now be discussed in turn.

The act of communication is a continuous process in which discrete packages of information are imparted by a sender, and subsequently decoded and responded to by a receiver [WPKD02]. Hence, whether collocated or visually mediated, the initial form of human communication is always expressed through natural verbal or nonverbal channels. Telecommunication in EyeCVE is both visual and aural. Hence, data collection must reflect these synchronous modes. The first column, headed *Human*, of Figure 3.21 illustrates the particular cues able to be tracked available to EyeCVE. These include oculesic behaviour of gaze, blinks, and pupil dilation, head and hand movement, and vocal signals. Additionally, in an experimental scenario, it is often critical for an experimenter to be able to input markup data alongside the tracking streams in order to delineate, for example, experimental stages and timing, or add miscellaneous notes. Figure 3.21 indicates this divide between data recorded from an experimental participant using EyeCVE, and that input by an experimenter.

The second column illustrated in Figure 3.21 is concerned with *Mediation*, which is the initial capture and conversion of natural human behaviour to electronic signals ready for processing. Head-mounted eye tracking is used to monitor oculesic behaviour including gaze, blinks, and pupil size. Head and hand movement is inferred from tracking sensors attached to a user's head and hand respectively. In an immersive VE, head tracking has the primary function of updating graphical rendering perspective, and hand tracking is generally used as an input device. While EyeCVE does employ the devices for that purpose, the data is also recorded, measuring head and hand movement, posture, and interaction input during system use. Alongside these nonverbal elements, a wearable wireless microphone monitors the verbal component of the social and collaborative interactions supported by the system. Finally, input of additional markup data by an experimenter is performed using a standard keyboard.

When discussing the third column in Figure 3.21, entitled *Processing*, it should be noted that each of the illustrated software processes may be distributed over various machines that are connected on the same local network. This design is important in order to distribute the significant computation that arises from the synchronous processing of the mediated tracking streams which may result performance bottlenecks should they all be collocated. In the approach described here, the eye tracker software is critical to both processing of oculesic data, and system-wide data collation and logging. This is indicated by the direct (i.e. non-networked) link from processing to collation stages. The ViewPoint eye trackers feature exceptional tracking, logging, and API capabilities. Calibrated for each individual prior to system use as documented in Section 3.1, the eye tracker software outputs 2D direction of gaze, blink signals, and pupil size at a rate of 60 Hz. Trackd is used to track both head and hand sensors, filtering and processing the streams into human-readable format, featuring six-DOF (three position and three rotation) output at a rate of 96 Hz. Voice detection is implemented using OpenAL [Spe10]. Rather than the actual speech, a binary value indicating the presence of speech is recorded. (As noted in the summary of this section, the original aural component of the interaction is also recorded, which embeds timing data consistent with the multimodal log file. However, this process is not central to the data collection method currently under discussion. The audio detector software has a cycle rate of 100 Hz, and uses a simple volume-based threshold to filter verbal signals which are likely to have been emitted from the monitored individual, from those likely to have been emitted from other sources. Finally, a keystroke detector operating on a machine accessible to the experimenter is used to input additional markup data. Individual key presses may be assigned to specific text strings, which are transmitted to the data logger at a rate of 100 Hz. For example, at the start of an experiment, the experimenter may input 's', which is processed by the keystroke detector, and output in the log file as "START".

The 3D spatial and synthetic environment in which interaction takes place in ICVE systems implies that, in order to preserve the context in which an immersed user performs a specific action, the processed tracking data must be transposed in some way before logging takes place. This stage is the concern of the ICVE system itself, and is indicated by the fourth column in Figure 3.21, which is dedicated to *Simulation*. As the diagram illustrates, all six user-tracked signals (gaze, blinks, pupil size, head movement, hand movement, and voice) are transferred over a network, and simulated by the EyeCVE process, locating the actions semantically within the context of the VE. These simulated versions of the data are represented by altered green-coloured icons. The primary use of the simulated data streams is to animate avatars in real-time: gaze is simulated and replicated in an avatar's gaze, and, when operating with advanced avatars, is also used as input to the lid saccade model; blink signals instigate the blink model when using the advanced avatars; changes in pupil size are approximated by an advanced avatar's pupils; head movement is replicated by an avatar, and is also used to update body movement and posture; hand and arm movement is simulated in an avatar using an IK algorithm; and vocal signals cause the avatar's mouth to move.

When considering the analytical use of user tracking, it is essential to record data in a manner which preserves the semantic context of the original action. When determining whether an input stream requires simulation to provide this context before the data is logged, the answer depends on the dimensionality of the signal. Data featuring more than one dimension must generally be simulated prior to logging, while data that can be represented on a single dimension generally does not require simulation. Consider, for example, pupil size and gaze. Figure 3.21 indicates that, while both pupil size and gaze are simulated (for avatar animation), only the simulated gaze data is output to the data logger, while the simulated pupil size is not. The reason for this lies in the dimensionality of the signals. Pupil size may be represented by a single scalar value: for instance as a percentage, with 0% representing full constriction and 100% indicating full dilation. Similarly to blinks and vocal input, which can be further simplified by binary values (i.e. blinking or not blinking, and speaking or not speaking), this scalar value has sufficient intrinsic meaning that is present at the processing stage of the data collection pipeline, which occurs prior to the simulation stage. Hence, pupil size requires neither the context of, or calibration in, the 3D spatial environment of EyeCVE in which interaction takes place. The same may partially be said for gaze. It is correct to state that, to some degree, gaze does have intrinsic meaning at the processing stage of the pipeline, defining a user's 2D direction of gaze relative to the viewpoint of the eye tracker's scene camera. However, while useful for general gaze analysis, these 2D coordinates do not relate explicitly to the immediate visual stimuli being looked at in the VE. In the case of gaze, the simulation phase consists of casting an intersection ray from the position and orientation of a user's tracked eye, into the VE. The ray hits the virtual geometry (for instance another user's avatar or an object) that is currently being looked at, and returns the object name, along with its 3D location in the VE. Analytically, this data is clearly more meaningful than 2D gaze coordinates, which, as the user is able to physically move in the VR system, are recorded from a varying positions and viewing perspectives. Similar simulation is performed with head tracking data, which returns the hit-object from the position and orientation of a user's head, which is useful as an indicator of attention. Hand tracking simulation records the names of the virtual objects and avatars with which a user interacts. Thus, through simulation in EyeCVE, multidimensional tracking data may be recorded and enriched, semantically, beyond the prior stage's processed tracking data, thereby enhancing subsequent analysis. In contrast, scalar data streams for which no simulation is required, including additional markup data, may be sent directly to the logging process. As a side-note, such data is not possible to be recorded in during collocated action, as information describing properties of objects and people within a scene generally do not exist.

The penultimate stage of the data collection pipeline concerns *Collation*, and is represented in the fifth column of Figure 3.21. This stage is responsible for combining all separate data streams prior to final output. As the diagram indicates from the unbroken connection from the *Eye Tracker Software* at the processing stage to the *Logger* at the collation stage, this computation is performed by the computer local to the eye tracker. All other data streams originate from other processes operating on different computers, and are streamed over a network to the logger process. While the network architecture potentially allows for any process to collate and log data in from multiple sources, the decision to locate the task at the computer running the eye tracker is well-founded. Fundamentally, the Viewpoint software is designed to output a multitude of oculesic and timing data to a single human-readable log file at a high rate of 60 Hz (16 ms). Additionally, the ViewPoint API allows external data to be streamed from multiple network connections, and output on the subsequent write-cycle to specific tab-separated columns in the log file. The final reasoning behind tasking logging to the eye tracker computer is one of latency. While the cycle-rate of eye tracking data acquisition (60 Hz) may be less than other tracking processes, the natural speed and frequency of human oculesics ensures that the amount of discreet and consequential data that is recorded by eye tracking exceeds all other tracking sources combined. Also due to their speed and frequency, oculesics are likely to be more sensitive to temporal misalignment than other less rapid channels of tracked behaviour. While gaze, blink signals, and pupil size data is sent for simulation in EyeCVE, this data is also streamed directly to the local logger. Of these streams, only simulated gaze is logged, while simulated blink and simulated pupil size is not logged.

Finally, as indicated in the sixth column of Figure 3.21, the collated data is *Output* to a single line of a tab-separated log file. As all computers streaming data to the logger process are connected via the same reliable local network, and due to the highly temporal nature of human behaviour, logger process outputs data as fast as possible, leading to a maximum of delay of 16 ms from the time that data is received until it is written to the log file. It should be acknowledged that the varying cycle rates and data transfer throughout the mediation, processing, and simulation stages of the pipeline are likely to result in some temporal variability between the different modes of input as measured from the initial stage of human action to the final stage of log file output. However, this delay is in the order of milliseconds, and is both unavoidable, and in practice, negligible to the intended application of interaction analysis. In summary, the data collection pipeline has been designed to operate alongside EyeCVE's distributed system architecture, strategically locating computation to enables collation and output of rich multimodal data describing a user's actions when engaged in AMC.

3.2.3 Multimodal Interaction Analysis

Two sequences of interaction, performed in EyeCVE and captured using the multimodal data collection method detailed above are now presented. The sequences are extracted from data-sets collected during the forthcoming experimental work. The first analysis presents a dyadic conversational scenario, and features data extracted from the truth and deception experiment detailed in Chapter 6. The second analysis examines an object-focused task featuring three users, and is extracted from the data collected from the experiment presented in Chapter 5. The two examples are typical to the type of collaborative interaction performed in EyeCVE. The segments are analysed similarly to traditional interaction analysis: by identifying significant moments of interaction, looking for causal relationships between observed behaviour, and, through verifiable observation, making judgements of how the interaction may have occurred [JH95]. For clarity, each analysis only considers metrics that are insightful with regards to describing the unfolding interaction. These metrics vary accordingly between the conversational and object-focused cases.

Conversational Scenario

Table 3.1 shows a sequence of interaction drawn from Chapter 6's truth and deception experiment data. Six lines from the collated log file are included in the table, each one pertaining to a significant moment during the short interaction. Each line's time of occurrence is stated in the leftmost column, which, for ease of reading, has been reset to 0 at the first line, and finishing 7.33 seconds later at the sixth. The remaining columns refer to: markup data (input by the experimenter as the action unfolds); head direction (the object in the VE directly in front of the participant's head); gaze (the object in the VE at which the participant is looking); pupil size (ranging between 0 indicating full constriction, and 1 indicating full dilation); whether the participant is currently blinking or not (1 indicates blink, 0 indicates no blink); and whether the participant is currently speaking or not (1 indicates that they are talking, 0 indicates that they silent).

Table 3.1: *Selected data from an interaction sequence taken from the truth and deception experiment documented in Chapter 6. Particular lines taken from the log file have been chosen to highlight significant moments during the interaction. In this sequence, the participant is required to lie to questions issued by a partner. In this case, they are asked "What is your first name?"*

Time	*Markup*	*Head*	*Gaze*	*Pupil Size*	*Blink*	*Voice*
0.00	General/Lie/Q01	Partner-Body	Grid-E2	0.08	0	0
1.81	Question issued	Partner-Face	Partner-Face	0.27	1	0
2.65	Answer start	Grid-D4	Grid-B4	0.39	0	1
5.17	Answer end	Partner-Body	Partner-Face	0.20	0	1
5.65	-	Partner-Body	Partner-Arm-R	0.31	0	0
7.33	General/Lie/Q02	Partner-Face	Partner-Face	0.13	1	0

The sequence begins at *T=0.00*. The "General/Lie/Q01" markup data indicates the experimental stage, in which the participant is about to be asked the first of a series of general questions, to which they must respond to deceptively, by lying. At this time, the participant is gazing a little away from their partner (the questioner), but their head direction indicates that general attention is focused on the questioner. The participant's pupil size is close to normal given the environmental luminance levels, indicating that they are relaxed, and neither cognitively or emotionally aroused. The participant is not talking at this time. At *T=1.81*, the markup data indicates that a question has been issued to the participant (in this case the question is "What is your first name?"). Both head and gaze direction indicate focus of attention on the questioner's face, and a slight increase in pupil size may indicate arousal. At this time, the participant also begins to blink. At *T=2.65*, the participant begins to answer the question, directing both gaze and head direction away from the questioner. The participant's pupils dilate significantly as the talk begins, suggesting cognitive load. At *T=5.17*, the participant finishes delivering the verbal answer, and returns head direction and gaze to the questioner. The participant's pupils are now less dilated. Shortly after, at *T=5.65*, the participant again averts gaze from the questioner's face, fixating downwards on the right arm. Tellingly, pupil size increases again, suggesting that the prior eye contact following the lie-telling evoked the participant's arousal, possibly negatively, due to social discomfort. Finally, at *T=7.33*, the markup data indicates that the current question is over, and the next is soon to be issued. The participant's pupil size has almost returned to normal, indicating a more relaxed state. A blink occurs, and gaze indicates that the participant's attention is again focused on the questioner.

Object-Focused Scenario

Table 3.2 shows a sequence of three-party object-focused interaction drawn from the experiment documented in Chapter 5. The task revolves around constructing a cube from eight smaller cubes, so that each face of the finished cube consists of a single colour. In an experimental scenario, the collaborative interaction generally consists of working together to find specifically-coloured cubes, picking up the cube, and positioning the cube in its correct position. Seven lines from the log file are presented, each one pertaining to a significant stage during the interaction. Each line's time of occurrence is indicated in the leftmost column, reset to 0 at the first line, and finishing 9.32 seconds later at the seventh. The key difference between this and the above analysis segment is the inclusion of hand tracking data, used to signify the name of any currently-grabbed object, while pupil size and blink data are omitted.

Table 3.2: *Selected data from an interaction sequence taken from the object-focused experiment documented in Chapter 5. In this sequence, the tracked participant is solving a simplified Rubik's Cube puzzle with two partners.*

Time	*Markup*	*Head*	*Gaze*	*Hand*	*Voice*
0.00	Search	Wall-01	Cube-04	-	0
2.17	Grab and Query	Cube-04	Partner-Body	Cube-04	1
4.81	Search	Wall-03	Cube-02	-	0
5.65	Search	Cube-02	Cube-05	-	0
6.17	Query	Cube-05	Partner-Face	-	1
7.48	Grab	Cube-05	Cube-05	Cube-05	0
9.32	Position	Cube-05	Cube-05	-	1

At *T=0.00*, the participant is searching for a particular cube, and appears to be examining "Cube-04". Soon after, at *T=2.17*, the participant has grabbed "Cube-04". The participant's head is oriented toward the grabbed object, but gaze is directed at their partner's body. Voice data indicates that the participant is speaking. In the context of the experiment, this is likely to be querying the correctness of the currently-grabbed cube for the intended placement position in the puzzle. This initial choice appears to have been incorrect as, at *T=4.81*, the participant has released "Cube-04", and is again searching the VE, presumably for the correct cube. At *T=5.65*, the participant's gaze falls on "Cube-05", and this time, caution is exerted before grabbing as, at *T=6.17*, gaze is directed to their partner's face, and another vocal query is uttered. At *T=7.48*, following what appears to have been an affirmative response, "Cube-05" is grabbed, and subsequently positioned at *T=9.32* where the segment ends.

3.2.4 Summary

In summary, this section presented an approach to multimodal data collection, and presented two sample analyses from conversational and object-focused experimental scenarios. During analysis of the conversation segment, the participant's state of arousal and cognitive load was inferred from behavioural data, particularly from pupil size in combination with gaze. The collation of several data streams in a single log file preserves the temporal interrelationships between components of recorded behaviour, and is critical in preserving the context of interaction. In this case, the behaviour of establishing mutual gaze, and the response of increased pupil dilation to the, perhaps uncomfortable, situation of lying to another, is observable in the data. Analysis of the object-focused task centred on hand action in combination with both verbal and gaze behaviour to uncover the logistic process of the task. In this case, the benefit of being able to reference tightly-synchronised elements of both the visual and aural telecommunication is evident when attempting to explicate sequences of interaction. For instance, the verbal and hand tracking data can elucidate how, following an initial grabbing error, the participant learns not to grab a cube until the correctness of that action is confirmed, thus deploying a successful repair strategy in order to remedy a prior mistake.

A limitation of the documented approach to data collation for interaction analysis is that a log file relates only to a single user. Thus, data collation must be generated on a per-user basis at each local site. There is certainly scope for remote users to be represented in a site's log file, but the included data must be processed wisely, taking into account latency and subsequent synchronisation with the local user's recorded data. For instance, a low-cost and high-benefit addition to the recorded interaction would be the binary state of mutual gaze between two interactants. In contrast, it would be unwise to stream full gaze data from a remote user for logging.

It must be noted that the logging process writes 60 lines per second to the output file. Hence, in order to distil an interaction as presented in the two examples above, a significant amount of processing must be performed prior to interaction analysis. However, due to the consistent format of the log file, the majority of this computation may be automated. Additionally, in order to be certain of the context by which a recorded interaction took place, it is often necessary for the log file to be referenced against an audio or visual replay of the performance. To this end, two solutions are available when using EyeCVE. Firstly, the aural component of interaction is recorded in the Ogg Vorbis [Mof01] bitstream format, which supports multiplexing of a number of separate codecs including audio and text. Alongside the multiparty audio communication, a textual time-stamp, matching that of the main collated log file, is embedded. Secondly, the suite of applications related to EyeCVE does include a log file player, which is able to reconstruct and replay the virtual action recorded in a log file. The log file player allows a session to be replayed, paused, and randomly accessed, and is also capable of visualising additional data, including gaze targets as the interaction proceeds. Finally, the player application operates on standard desktop displays, and also in immersive CAVE systems, the latter enabling a free first-person viewpoint and perspective rendering in the spatial VE in which the original interaction took place. In this way, an analyst can be a bystander to a pre-recorded interaction similarly to a video replay, but with the critical advantage of an adjustable camera viewpoint.

Throughout the forthcoming experimental research, a variety of analytical methods, with varying aims, are employed. The multimodal approach as documented in this section provides precise timing of a range of data sources, preserving the holistic temporal characteristics and emphasising the causal interrelationships between the various tracking streams. More so perhaps than other methods of analysis, an objective understanding of users' intent and state of arousal may be gained, giving a broad picture of how the interaction unfolded. More generally, through this kind of analysis, communicational failures related to the medium of AMC in ICVEs may emerge. Whether arising from technical or human-centred factors, critical bottlenecks restricting high quality virtual telecommunication may then be addressed.

3.3 Chapter Summary

The research presented in this thesis consists of three telecommunication experiments, and a collection of smaller-scale investigations relating to behavioural modelling and associated experiments. The telecommunication experiments investigate the impact of eye tracking, both to replicate and analyse oculesic behaviour in AMC, while the behavioural modelling work is focused on simulations of oculesics.

The first part of this chapter detailed the ICVE system, EyeCVE, which acts as the primary experimental platform for the forthcoming work. Alongside the standard head and hand tracking common to ICVE systems, EyeCVE integrates head-mounted eye tracking capable of capturing a wearer's oculesic behaviour, including gaze, blinks, and pupil size. The evaluation of EyeCVE demonstrated its ability to support tracked gaze AMC while allowing physical movement within an immersive VR system such as the CAVE. The second part of this chapter documented an approach to multimodal data collection in a single log file. Using all available tracking devices in EyeCVE, the logging ability of the Viewpoint eye tracking system employed in order to collate an array input streams in a temporally accurate format that is both human-readable and conducive to post-processing and analysis. Demonstrating the technique, interaction analysis relating to two sequences of AMC recorded in EyeCVE were then presented. Analysis emphasised the ability of collated input streams to elucidate interrelationships between components of captured behaviour, providing insight into an individual's state of attention and arousal, together with logistic elements of interaction and repair strategy.

3.3.1 Reading Guide to Experimental Chapters

Table 3.3 is designed as a reading guide to the upcoming experimental chapters. The three telecommunication experiments are presented, together with two oculesic model experiments. The chapter in which each may be found is shown in the rightmost column. The *System* column refers to the VR system (CAVE or WALL) used in the experiment, and the presence of a VMC comparison condition. The number listed in this column indicates the number of interactants in an experimental session. For instance, in the truth and deception experiment, two-party AMC between a user located in the CAVE, and a user in the WALL is investigated, together with a VMC comparison condition. The column headed *EyeCVE* indicates the degree of maturity of the system in terms of performance, eye tracking ability, and avatar subsystem in accordance to Section 3.1. Performance of *Basic* maturity is comparable to the evaluation documented in this chapter, while the system's core performance is significantly improved in

Mature and *Mature++* levels. In particular, Mature++ transmits avatar update messages at significantly higher frequency and sensitivity. This maturity is also reflected in the following two columns entitled *Eye Tracker* and *Avatar*, which refer to the eye tracking device and avatar type employed during the experiment respectively. Implications of these components of the EyeCVE system have been documented in this chapter. Finally, the *Analysis* column refers to the methods of data analysis performed in each experiment. A variety of approaches are used, including pre- and post-experimental questionnaires, eye tracking data alone (i.e. not multimodal), task performance, conversation analysis, subjective ratings of visual stimuli, and multimodal analysis as documented in this chapter.

In summary, the three telecommunication experiments are reported in chronological order, with increasing capability with regards to both performance of EyeCVE as a telecommunications medium, and to data analysis methods. An initial experiment on three-party conversation is presented in Chapter 4, followed by a three-party object-focused experiment in Chapter 5, and finally a highly interpersonal scenario studying truth and deception in Chapter 6. The behavioural modelling work and associated experiments are presented in Chapter 7, and represent work that was carried out, chronologically, between the object-focused and truth and deception experiments.

Table 3.3: *Overview of experimental chapters. Information includes VR system in use, maturity of EyeCVE, eye tracker in use, avatar type in use, and analysis methods.*

FOCUS & CHAPTER	SYSTEM	EYECVE	EYE TRACKER	AVATAR	ANALYSIS
EyeCVE Evaluation Section 3.1	CAVE CAVE	Basic	MobileEye	Basic	- System performance
Conversation Chapter 4	CAVE / *VMC* CAVE / *VMC* CAVE / *VMC*	Basic	MobileEye	Basic	- CA - Eye tracking
Object-Focused Chapter 5	CAVE CAVE CAVE	Mature	ViewPoint	Basic	- Performance - Questionnaires - Multimodal
Truth & Deception Chapter 6	WALL / *VMC* CAVE / *VMC*	Mature++	ViewPoint	Advanced	- Multimodal - Questionnaires - Subjective rating
Oculesic Behaviours Section 7.1	HD display	N/A	N/A	Advanced	- Performance - Subjective rating
Eyelid Model Section 7.2	CAVE	Mature++	ViewPoint	Advanced	- Subjective rating
Gaze Model Section 7.3	CAVE	Mature++	ViewPoint	Advanced	- Subjective rating

Chapter 4

Experiment: Three-Party Conversation

The overarching goal of the telecommunication experiments documented over the following three chapters is to investigate the use of eye tracking in AMC for both interactive and analytical applications. Each chapter addresses a particular scenario that represents a common usage of, or issue central to, remote work and collaboration. The current chapter investigates multiparty conversation, Chapter 5 examines object-focused interaction, and Chapter 6 explores truth and deception.

The experiment presented in this chapter investigates three-party conversation in tracked gaze AMC and gaze aware VMC. In the VMC setting, gaze awareness was realised by careful alignment of video displays with respect to camera position, while the AMC setting used an early version of EyeCVE to replicate users' gaze, head, and hand movements in their avatar embodiments. The goal of the experiment was to compare how people are observed to interact and behave when using these two visual telecommunication mediums. Gaze data, recorded from eye trackers worn by participants interacting with confederates in both systems, acts as the primary data source for evaluation and analysis. CA is coupled with the eye tracking data to assess participants' ability to employ strategies commonly observed in collocated interaction, in order to successfully manage the mediated multiparty conversation. The theme of social agency is also addressed, assessing the extent to which participants engaged in AMC are observed to distribute their own gaze, and monitor others' gaze, in a manner that is similar to when they are confronted with actual video representations of fellow interactants.

4.1 Experimental Aims and Expectations

The aim of this experiment was to investigate tracked gaze AMC and gaze aware VMC, in terms of how participants' behave and use the capabilities of the systems in a multiparty conversational scenario. Of particular interest was the investigation of the role, and importance, of gaze as a communicational resource during both information exchange and management of the interaction. This included the ability to interpret attention from observing others' gaze, signalling with gaze to select the next speaker, and establishing mutual gaze at critical moments of interaction. The basic hypothesis was that triads engaged in both AMC and VMC will be able to conduct conversation successfully, and will utilise gaze, measured by eye tracking, similarly in both mediums for purposes of both nonverbal expression and management of the unfolding interaction. Also, behaviour observed during the mediated communication is expected to follow established social norms observed in collocated interaction.

4.2 Experimental Design

A between-groups one-way design was employed, with the main two-level factor being the classes of AMC and VMC. Ten male participants were recruited from the student and staff population at University College London (UCL). Experimental sessions involved mediated interactions between one participant and two male confederates. Five of the interactions were performed in the AMC condition, and five in the VMC condition. One confederate was located at University of Reading (UoR), and the other at University of Salford (UoS). The same confederates performed all ten experimental sessions in both AMC and VMC conditions. The main part of the experiment involved the group engaging in a conversational scenario, in which the confederates subjected the participant to an informal interview. Participants were asked a total of nine questions, which the two confederates took turns in issuing. The questions related to academic background, and were designed to stimulate conversation amongst the group, rather than to simply be answered. In both AMC and VMC conditions, the visual and aural interactions were recorded from the perspective of the eye tracker's scene camera, mounted on the participant's head. Gaze direction is also embedded in the video. The following sections detail the experimental design. For additional materials, see Appendix C.

4.2.1 Independent Variables

Two conditions of mediated telecommunication were investigated:

1. *AMC:* Interactants engaged in the triadic communication were represented by avatars in a shared VE. Each avatar embodiment replicated their controlling user's tracked head, hand, and gaze motion in real-time. Audio communication allowed users to converse with each other.

2. *VMC:* Triadic gaze aware video-conferencing enabled interactants to communicate both visually and aurally. The large-format displays granted head-and-shoulders displays of each user, thereby allowing both signalling and observation of gaze and head movement.

4.2.2 Apparatus

A total of six laboratories, one for the AMC condition and one for the VMC condition at each of the three sites, were set up prior to the experiment. AMC was supported by EyeCVE, which operated between three CAVE systems, while VMC was supported by *Access Grid* [CDO+00b]. In both mediums, ASL MobileEye eye trackers were used to record the gaze of participants and confederates, and also to control avatar gaze in EyeCVE. The following sections present the arrangement of the technical apparatus.

EyeCVE Setup

The CAVE systems located at each site were networked over the Internet though EyeCVE. At all sites, *ReaCTor* CAVE-like systems, originally made by Trimension, were used. The displays consisted of three rear-projected 3×2.2 m walls (3×3 m at UoS) and a front-projected 3×3 m floor. The operating resolution of all projectors was 1024×768 pixels, with a refresh rate of 96 Hz in active stereo. *SGI Prism*

Figure 4.1: *A user located in a CAVE system, prepared for three-party AMC supported by EyeCVE. The user is fitted with eye, head, and hand tracking devices.*

computers ran EyeCVE at each site. Head and hand tracking was achieved at UCL using an InterSense IS900 [Int10] acoustic tracking system, at UoS using VICON [Gro10] optical tracking, and at UoR using Ascension MotionStar [Cor10a] magnetic tracking. Eye tracking at each site was achieved using ASL MobileEye units. As illustrated in Figure 4.1, confederates and participants sat on chairs placed in the centre of each VR system, wore shutter glasses with mounted head and eye tracking, and held a hand tracker in their right hand. Wireless microphones were worn by all interactants, and a Skype [Sky10] conference call supported audio communication, which was output through speakers located in each CAVE.

The VE in which the AMC took place consisted of a basic room with the same volume ($3m^3$) as the physical space enclosed by the CAVE systems. In order to maximise the tracking ability of the MobileEye eye trackers, the ambient luminance level was increased by colouring the walls of the VE white. In order to minimize distractions, which may have influenced gaze behaviour during the conversational task, the only objects in the room, together with the avatars, was a round table and three virtual chairs. The real chairs on which the interactants sat were aligned to their virtual counterparts, while no real tables were installed. The VE is illustrated in Figure 4.2.

Avatars' gaze, head, and right arm were animated based on users' tracked physical movement. The basic avatars, as described in Chapter 3 were used in the experiment. While EyeCVE allows free physical movement within the CAVE, this was informally restricted by the seated positioning of interactants around the virtual table. This restriction acknowledged the inherent limitation of movement in the VMC condition, and aimed to promote the use of upper-body kinesics during use of both mediums. The round-table scenario aimed to simulate a three-party meeting arrangement that commonly takes place in the real-world. Figure 4.3 shows the EyeCVE setup and positioning of interactants in the virtual meeting room.

Figure 4.2: *The VE in which the experimental interactions took place, consisting of a white 3×3 m room with a round meeting table in the centre surrounded by three chairs.*

Figure 4.3: *AMC in progress.*

Gaze Aware Access Grid Setup

The Access Grid (AG) project, led by Argonne National Laboratory, is an ensemble of resources including large-format displays, presentation, and interactive environments aiming to support room oriented group-to-group collaboration [CDO+00a]. The vision of the AG project is to provide a mixed-media visual and aural telecommunication environment with ambient video and audio, large-scale displays and with software to enable the relatively transparent sharing of ideas, thoughts, experiments, applications and conversation. The system aims to support both formal and informal interaction between remote groups and individuals.

In the current experiment, VMC was achieved by linking AG nodes, located at each site, to enable high-quality video and audio communication. The AG nodes at each site consisted of 2×1.3 m rear-projected displays operating at a resolution of 1024×768 pixels and at a rate of 60Hz. Audio was captured by microphones, either installed in the ceiling or placed in front of the local user, and played back at remote sites through speakers aligned to appropriate displays of remote users. Interactants were fitted with MobileEye eye trackers, which recorded the unfolding interaction from the perspective of each scene camera, with overlaid gaze direction.

The displays and cameras at each AG node were aligned as optimally as possible to achieve gaze

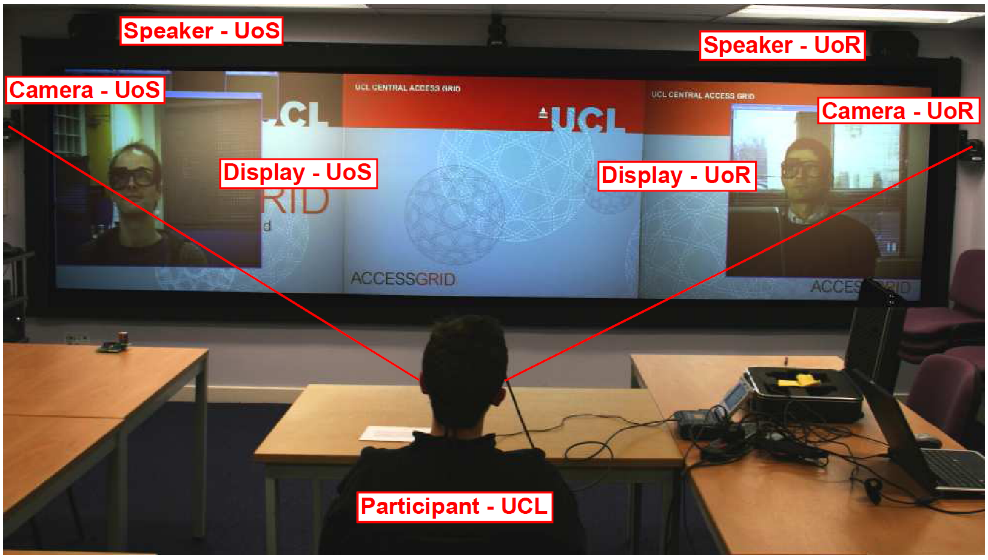


Figure 4.4: *Arrangement of Access Grid displays, cameras, and speakers to achieve gaze aware VMC at UCL. Note the head and shoulders views of fellow interactants.*

awareness with regards to Chen's angular offset threshold of 5° vertical and 1° horizontal [Che02]. Ideally, a camera would be located behind the display of each remote user, but this was not an option at any of the sites. Hence, the video displays were centred either at a vertical or horizontal offset to the appropriate camera. Figure 4.4 presents the arrangement at UCL, with annotated video display position, audio output, and camera direction. In this setup, interactants are able to perceive the general gaze direction of others, including at who they are looking, together with mutual gaze. Despite the physical differences in size between the VMC and AMC labs, the AG arrangement at each site intended to resemble the round-table meeting room as more easily established, due to the surrounding shared environment, in the EyeCVE setup. Depending on the physical characteristics of the AG room at each site, the distance from the local seated user to the displays was defined so that the viewing angles between video displays of the two remote users was between 30–40° from centre, similar to the AMC setup. The video displays at each site were separated horizontally so that the local user could see the head and shoulders of both remote others, peripherally, at once. However, redirection of gaze and head orientation was required in order to focus on either one. Interactants sat at a table facing the displays, with one video display and one camera required for each connection between each of the remote sites. Thus, each site had two video displays recorded from two cameras, with one streaming to and displayed at remote site A, while the other sent to remote site B. This specification, supporting natural line-of-sight between remote interactants is visualised in Figure 4.5, which also illustrates the equivalent setup in the AMC between CAVE systems.

Eye Tracking

During both AMC and VMC experimental sessions, the MobileEye eye trackers record the gaze of participants and confederates. In the AMC condition, gaze data was also used to drive the gaze of avatars. The eye tracker requires calibration for every wearer in order to process gaze. The video recorded by the scene camera is output to a DV tape, together with audio and embedded gaze point, relative to the scene.

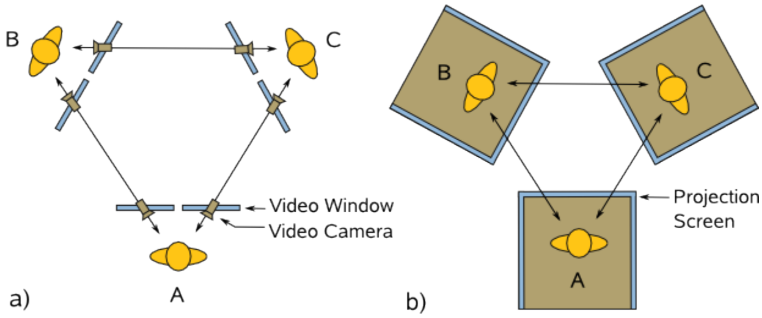


Figure 4.5: *Visualisations of gaze awareness in VMC supported by display and camera alignment in AG (*left*) and tracked gaze AMC achieved in EyeCVE using mobile eye tracking (*right*).*

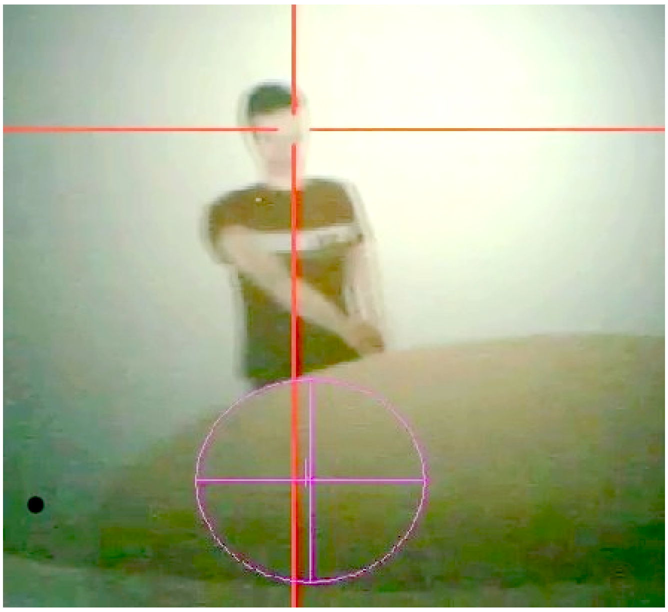


Figure 4.6: *Frame captured by eye tracker scene camera. Gaze direction overlay marked by red crosshair.*

An example frame recorded from the scene camera is shown in Figure 4.6.

During the VMC experiments, the eye trackers were mounted on lightweight frames which aimed to minimise obscuring the wearer's eyes to those observing via the AG video streams as shown in Figure 4.4. In the AMC experiments, eye tracking was made possible by mounting the MobileEye on the CAVEs' CrystalEyes 3 [Rea10] shutter glasses as shown in Figure 3.3. The attenuation of the infrared LED beams was measured to be about 70% every time the beams traversed the shutter glass lens. Hence, the total attenuation of the bidirectional path was about 50%. However, this still allowed a sufficiently clear image of the LED reflections to be captured by the tracker's infrared camera for reasonable tracking ability. When using both types of frame, the eye tracker software's pupil contour detection algorithm was found to be more robust when tracking dark irises than on lighter, blue or green, eyes. This effect was emphasized when the tracker was mounted on the shutter glasses. Thus, to ensure high-quality data, only participants with dark eyes were recruited.

4.2.3 Population

Ten participants with normal or corrected-to-normal vision and no previous experience of VR were recruited from UCL's campus through an advertising poster campaign. They were paid £5 for the hour-long study. In order to minimise possible effects of gender, and as both confederates were male, only male participants were recruited. The ten participants were divided equally over the VMC and AMC conditions, with five taking part in each.

4.2.4 Procedure

UCL was established as the main experimentation site. One experimenter was responsible for organising and running the experiments as well as local technical setup and participant organisation. The two confederates located at UoS and UoR maintained their local site's technical setup, and played the roles of interviewers during the experimental interactions. Experimental sessions performed using the two mediums were interspersed over the ten participants, aiming to balance the impact of confederates' habituation to the procedure across the two conditions. The experimental procedure was similar for both AMC and VMC sessions, and consisted of four phases: briefing upon participant arrival, technical setup, eye tracker calibration, and finally the experimental interaction itself.

Phase 1: Briefing

During recruitment, participants were assigned either to the AMC or VMC condition, and were instructed to arrive at the location of either the CAVE or AG lab. Upon arrival, participants were welcomed in a reception area at each lab, and were given an information sheet and ethical consent form. The information sheet explained the general experimental scenario, setting the scene for the forthcoming interview in the triadic interaction. Participants were left in private to fill out the forms, and were informed that they may ask any questions not directly relating to the forthcoming experimental procedure or technical setup.

Phase 2: Technical Setup

While the participant read and completed the forms provided to them in the briefing stage, the telecommunication system was initialised. In the AG labs, this involved establishing a three-party video and audio conference using UoR's shared AG room, with UCL and UoS connecting to the room. The displays and cameras were aligned at each site to optimise gaze awareness, and audio input and output levels were set appropriately. The established VMC at each site was checked to ensure both correctness of the physical arrangement, and performance of the streaming video and audio between sites. Finally, the eye trackers at each site were prepared for use, and a new DV tape was inserted into each DVCR.

The technical setup in the CAVE labs involved initialisation of EyeCVE. Firstly, a Skype [Sky10] conference call was established between the three sites. Wireless microphones were initialised, and talk from remote users was output from speakers positioned in each CAVE. The MobileEye eye trackers at each site were initialised, with a new DV tape inserted into each DVCR. UoS ran the EyeCVE server, which loaded the meeting room VE as displayed in Figure 4.2. Each site then connected EyeCVE clients to different ports at the IP address of the listening EyeCVE server. Once the local client completed loading, the confederates at UoS and UoR and the experimenter at UCL navigated their rendering perspective in the VE (which in turn repositioned the local avatar as displayed at remote sites) to the appropriate chair around the shared virtual table, which was aligned to the physical chair positioned in the local CAVE.

Phase 3: Eye Tracker Calibration

When performing the VMC condition, the confederates at UoR and UoS donned the empty frame with mounted eye tracker and completed the single-phase gaze calibration procedure using the MobileEye software. This completed the pre-experimental VMC setup. In the AMC condition, confederates wore the CAVE shutter glasses with mounted eye and head tracking. In addition to calibration in the MobileEye software, calibration within the VE was performed, mapping raw 2D gaze to the appropriate 3D rotation of the embodied avatar's eyes (see Section 3.1.4), thus enabling tracked gaze avatars. In the AMC condition, the confederates held the hand tracker in their right hand, and ensured correct positioning of their wireless microphone.

Following confederate calibration, the participant entered the lab at UCL and was seated. The eye tracker was fitted to the participant, whose gaze was then calibrated using the same one- or two-phase procedure. The experimenter checked the accuracy of the calibration and made adjustments if required. In the AMC condition, the wireless microphone was attached to the participant, who also held the hand tracker in their right hand.

Finally, completing the pre-experimental phase, the experimenter at UCL issued a verbal signal to the confederates to commence data collection. The experimenter then left the lab for the duration of the experimental interaction.

Phase 4: Experimental Interaction

The AMC or VMC began with the confederates greeting and introducing themselves to the participant. The confederates briefly outlined the informal interview scenario, and explained that they would alternate the issuing of questions between them, and that the participant should answer as they wished. The confederates commenced to issue the nine questions is presented in Table 4.1. The questions were often followed up or expanded on naturally, engaging all interactants in multi-way conversation, rather than a strict question-answer protocol. Following the question period, which lasted for a duration of around 15 minutes, the confederates thanked the participant for their time, concluding the experimental session.

Table 4.1: *Questions issued to participants during the mediated interviews.*

Question	*Issued by*
Tell me briefly about your academic background.	UoR
Do you consider yourself successful and why?	UoS
Are you a team player? Give me an example.	UoR
What is your philosophy towards work?	UoS
What irritates you about co-workers?	UoR
What is your greatest strength?	UoS
What kind of person would you refuse to work with?	UoR
What would your supervisor say your strongest point is?	UoS
Describe your work ethic.	UoR

4.3 Analysis

The following sections present analyses of participants' behaviour during the experimental interactions, comparing AMC and VMC. Qualitatively, CA, combined with eye tracking data, is the primary method of analysis used, and aims to identifying systematic practices used during the social interaction. Each line of the CA extracts shows a reference number, an abbreviation identifying the speaker, and the transcribed talk. Various symbols and spellings are used to capture aspects of the sound of the talk as introduced by Atkinson and Heritage [KF73, AH84]. The participant's direction of gaze is indicated as a labelled line below the transcribed talk, parallel to the talk to which it relates, and gaze transitions are shown by a row of dots. Quantitatively, eye tracking data is used to assess gaze distribution. The analysis focuses on explication of participant behaviour, rather than that of the confederates. However, due to the social context that the data describes, it is necessary, at times, to consider all parties. Informational exchange and participants' involvement, including their ability to interpret attention from gaze and to establish mutual gaze with confederates, during the experimental interactions is considered. The analysis is divided into three subsections: support of gaze practices, management of speaker transition, and gaze distribution. For additional materials, see Appendix C.

4.3.1 Supporting Gaze Practices

Certain gaze practices commonly employed in collocated interaction may also be observed in the recorded AMC and VMC. In collocated conversation, a common gaze practice is for a listener to look directly at a speakers's facial region, particularly at the eyes, while being questioned. Once the question has been issued, response formation and delivery begins with the respondent looking away from the questioner, which signals a hold of the conversational floor, and ends by redirecting gaze back to the questioner as the answer reaches completion, returning control of the floor [Dun74]. This "at-away-at" behavioural sequence was frequently observed during remote interactions in both VMC and AMC, and examples of which are transcribed in Figure 4.7 and Figure 4.8 respectively.

Figure 4.7 presents a CA extract from a VMC interaction featuring a confederate, Rob, issuing a question, and the subsequent answer from the participant, Wal. At lines **06–07** of the transcript, Rob

```
#1.1 AccessGrid #4 (simplified)

06  Rob:   Can you tell us briefly about your
           Rob________________________________

07         professional academic background °please°.
           Rob______________________________________..

08  Wal:   Er:m >sure< em I'm a PhD
           Centre_____________________

09         student at (names institution)(.)
           ___.Rob_______________________.
```

Figure 4.7: *VMC: simplified CA transcript illustrating "at-away-at" gaze behaviour of the participant when being asked, and when responding to, a question. Names have been changed to ensure anonymity.*

```
#1.2 CAVE#06 (Simplified)

04  Rob:    Okay thanks er a what kind of person:
            Rob

05          (.) would you refuse to work with
            Rob                         Paul

06          (.)
            Centre

07  Ric:    Er:: (.) someone stubborn (.) that doesn't
            Centre     Rob      Paul

08          listen to others  (2.0)
            Paul        Rob
```

Figure 4.8: *AMC: simplified CA transcript illustrating "at-away-at" gaze behaviour of the participant when being asked, and when responding to, a question. Names have been changed to ensure anonymity.*

asks Wal to tell them (the confederates) about his professional or academic background. As indicated by the solid underlining of the text, during delivery of the question, Wal is gazing directly at Rob. As Rob finishes issuing the question with "please", Wal shifts his gaze to the centre of the AG display, away from his two interviewers as illustrated in Figure 4.4. Wal begins his vocal reply at line **08**, maintaining his gaze direction between the location of the confederate video displays until he reaches the semantic point, "student", of his answer at line **09**. This utterance coincides with his return of gaze direction back to the questioner, Rob.

The CA extract from an AMC interaction presented in Figure 4.8 describes similar "at-away-at" gaze behaviour employed by a participant, Ric, during a question issued by confederate, Rob. The non-questioning confederate is Paul. Over lines **04–05**, Rob asks Ric to explain the characteristics of someone that they would refuse to work with. Ric gazes toward Rob, before a brief glance at Paul as the question comes to an end. There is a brief silence, marked by *(.)*, at line *06*, and Ric's gaze is directed at the central space between the confederate avatars. This away-gaze strategy continues on line *07* while Ric utters "Er", signalling that he is still formulating a verbal response. Following another brief pause, Ric starts to talk, returning his gaze first to the questioner, Rob, and then to Paul at line *08*, and finally back to Rob as his response ends. The *(2.0)* at line *08* indicates a significant pause of two seconds before any further action takes place, during which Ric maintains his gaze on Rob, signalling the end of his talk.

4.3.2 Management of Speaker Transition

The semi-formal structure of the conversation, which involved scripted questions issued between two interviewers to a single interviewee who answered all questions, has clear interactional constraints with respect to ownership of the conversational floor. However, speaker transition is less constrained immediately following a participant's answer, at which point, the two confederates must decide, often nonverbally, whether to issue a follow-up response to the answer just produced, or to simply advance to

```
#2.1   AccessGrid #02 (simplified)

01  Mac:     uh: (.) and I am working as a: uh: research
             Centre          Rob          Paul       Rob

02           fellow here at UCL(.) uh in computer


03           science (.) er primarily character animation


04           (0.8; Rob directs gaze at Paul; Mac follows gaze)
             Rob                               Paul

05           (0.3)
             Paul

06  Paul:    But do::: do you c- consider yourself
             Paul

07           successful.
             Paul
```

Figure 4.9: *VMC: simplified CA transcript demonstrating the deployment of gaze for conversation management. Gaze is used to signal speaker transition between confederates, which is subsequently observed and recognised by the participant.*

the next question. In the former case, the answer is likely to have been deemed inadequate, possibly due to brevity, or a misunderstanding by either or both confederates, who may subsequently produce further talk. In case of simple advancing to the next question, this decision is likely to be indicated through a familiar nonverbal signal such as mutual eye contact coinciding with a nod of the head between confederates. Thus, between questions in both conditions, there is a turn-allocation situation that requires local management. This management of speaker transition, including general awareness of conversational state was sub-optimal in both forms of mediated communication.

VMC

In those moments in which turn-allocation had to be locally managed by the interactants, the faithful representation of users provided video seemed to support more effective nonverbal signalling than AMC. Particularly during decisive moments following a participant's answer to a question, the confederates at UoS and UoR (who were also personal friends) were able to employ and recognise gaze, head movement, and characteristic facial expression to ease transition. The import of gaze is illustrated in the CA extract presented in Figure 4.9. At line *03* of the extract, the participant, Mac, finishes his vocal answer to a prior question issued by Rob, who he then continues to gaze at. At line *04*, there is a silence of 0.8 seconds, during which Rob directs his gaze at the next questioner, Paul. This redirection of gaze is recognised by Mac as a turn-taking strategy, who follows Rob's gaze to also fixate on Paul. Mac maintains his gaze on Paul, the identified next speaker, during a brief pause at line *05* before Paul begins the next question at line *06*.

Despite the VMC system's effectiveness of gaze as a management technique, the spatial fragmentation of the visual displays, as illustrated in Figure 4.4, led to difficulties in conversational management.

```
#2.2 AccessGrid #03  (simplified)

continued...

07  Jon:    is something I could be accused of
            Paul            Rob

08          sometimes so: .hh youkno:w huh huh
            Rob                        Paul

09          (0.0)    (1.0)    (2.0)     (3.0; R->P) (4.2)
            Paul  Centre  Paul Centre   Rob       Paul

10  Paul:   And my last question is uh:m what would
            Paul
```

Figure 4.10: *VMC: simplified CA transcript highlighting difficulties in conversational management due to fragmentation of the shared workspace.*

For example, the transition between questions sometimes involved a relatively long period of silence, indicated by gaze performed by the participant following completion of an answer. During these silences, the participant often alternated gaze between the confederates, monitoring the unfolding of the speaker transition. This process was often concluded with the confederate who issued the prior question shifting his gaze away from the participant and toward his fellow confederate. Due to the fragmentation of the displays showing the remote confederates as viewed by the participant, gaze awareness is only able to provide general information regarding attention, as opposed to precise direction of gaze. Figure 4.10 shows how, following his answer, a participant, Jon, witnesses the unfolding of speaker transition over a long period of 4.2 seconds of silence, at line *09*. Jon uses gaze to monitor which confederate is likely to speak next. Initially he looks at Paul, expecting the next question. However, in the absence any talk from Paul, Jon moves his gaze to the centre, between Rob and Paul. After three 3 seconds, John directs his gaze at Rob, who subsequently looks at Paul (indicated by R->P), prompting for the next question. Finally, Jon again looks at Paul at the end of the silence, sustaining his gaze while Paul issues the next question at line *10*. The high degree of fragmentation is contextualised in Figure 4.11, showing the physical separation of confederate video displays.

AMC

Deficiencies of the current version of EyeCVE's avatar behaviour and animation system created specific challenges to participants during AMC. The basic avatars used in this experiment did not feature mouth movement corresponding with embodied users' vocal utterances. Combined with the lack of spatialised audio output, participants were often confused about which confederate was currently talking. In an attempt to elucidate who was speaking, the confederates adopted a system of naming and pointing to the next questioner, in order to explicitly allocate the conversational floor. This gestural approach to turn-taking was combined with gaze and head movement, which generally provided participants with sufficient transition management information. This strategy is elucidated in Figure 4.12, which shows how, after completion of his answer at line *02*, the participant, Sim, alternates his gaze between the two confederates at line *03* in an attempt to monitor who is going to speak next. The transition of the

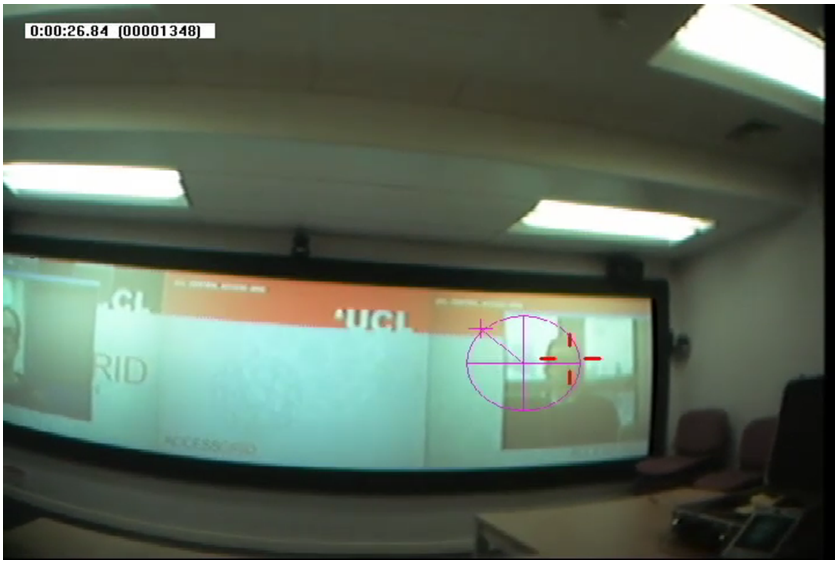


Figure 4.11: *Participant view during VMC, illustrating the fragmentation of the shared workspace.*

questioner at line *04* is performed by Rob's arm gesture towards Paul, eliciting a shift in Sim's gaze from Rob to Paul. Sim's gaze at Paul is maintained, as Paul performs a corresponding 'receive' arm gesture and begins to issue the next question at line *06*. This gestural sequence to management of turn-taking is illustrated from the participant's perspective in Figure 4.13.

Although such gestures are somewhat artificial in comparison to how turn-allocation is performed in collocated interaction [SSJ74], their success highlights the importance of the shared and perceptually unified space provided by EyeCVE. In contrast to most VMC systems, the surrounding nature of the CAVE displays allows full-body representations of users, thus allowing gesture to be used as a communicational resource, and attention to be naturally signalled and observed by this means.

4.3.3 Gaze Distribution

To support the CA, a quantitative measure of gaze distribution was performed using eye tracking gaze data as recorded from the scene camera mounted on the CAVE and AG frames worn by participants. Six academics from departments of psychology and computer science located at UCL, UoR, UoS, and Roehampton University were asked to classify participants' direction of gaze for each period of time during which a question was being issued by a confederate. During the experimental interactions, the ability for confederates to ask follow-up questions in order to elicit further information from participants resulted in a varying number of questions that were available to be rated over the two conditions, with 56 over the AMC sessions and 52 in the VMC sessions over all ten participants.

An application, which embedded videos and recorded user ratings, was developed, and is presented in Figure 4.14. The video panel positioned to the right of the GUI plays a clip captured from the experimental interactions, and allows for pausing and re-watching. The panel to the left of the video enables raters to classify gaze direction into four categories: at the questioner (*Speaker*), at the bystanding confederate (*Other*), away from both confederates (*Neither*), and gaze deemed ambiguous, and hence difficult to classify (*N/A*).

```
#3.1 CAVE#03 (Simplified)

01   Sim:     Uh:m: (.) before coming to UCL
              ____________________  Paul____

02            to:: run the tech la:b
              Paul____  Rob_  Paul_____

03            (0.0) (1.0) (2.0)
              Rob__ Paul_ Rob__

                         (Rob extends arm to Paul)
04   Rob:     O:ka:y (0.2) So: to yo:u
              Rob__________  Paul_____

05            (0.4; Paul extends arm too)
              Paul_______________________

06   Paul:    Oka:y=uh:: do you consider yourself up till
              Paul_______________________________________

07            no::w: (.) successful::And why:
              Paul_____________________________
```

Figure 4.12: *AMC: simplified CA transcript highlighting gestural strategies to conversation management. Figure 4.13 shows the corresponding view from the participant's viewpoint.*

To ensure inter-rater reliability, Fleiss' kappa coefficient was calculated to be 0.67, constituting substantial agreement between the six response sets. Subsequently, judgements recorded from the six raters were then classified according to a majority vote, with a threshold of ≥ 5 agreeing votes ensuring the according classification, and <5 identical responses being classified as ambiguous. For instance, a clip in which five out of the six raters judged the participant's gaze to be directed at the questioning confederate would be classified as such. However, a clip judged identically by only four raters would be deemed unsuitable for confident categorisation, resulting in an 'ambiguous' classification. Table 4.2 presents participants' gaze distribution over all questions in both AMC and VMC conditions. Participant gaze *At Speaker* is less during AMC compared to VMC. There are two explanations for this: firstly, recalling the confusion caused due to lack of avatar mouth movement synchronised with talk, participants occasionally had problems identifying which confederate was talking, thus accounting for increased proportions of *At Other* and *Away* gaze; secondly, eye tracking was less robust when operating in the low-luminance environment of the CAVE and through the tinted shutter glass lenses. This is likely to contribute to the increased *Ambiguous* ratings, referring to cases in which gaze was deemed too erratic for broad categorisation.

Table 4.2: *Participant gaze direction while being asked questions in AMC and VMC conditions*

Condition	At Speaker	At Other	Away	Ambiguous
AMC	66.7%	5.6%	16.6%	11.1%
VMC	80.2%	2.9%	10.4%	6.5%

Figure 4.13: *Gestural management of conversation corresponding to Figure 4.12.* Left: *Rob extends arm to Paul, thereby 'giving' him the conversational floor (line* 04 *in Figure 4.12).* Right: *Paul gestures to 'receive' the floor, prompting his speaking turn (line* 05 *in Figure 4.12).*

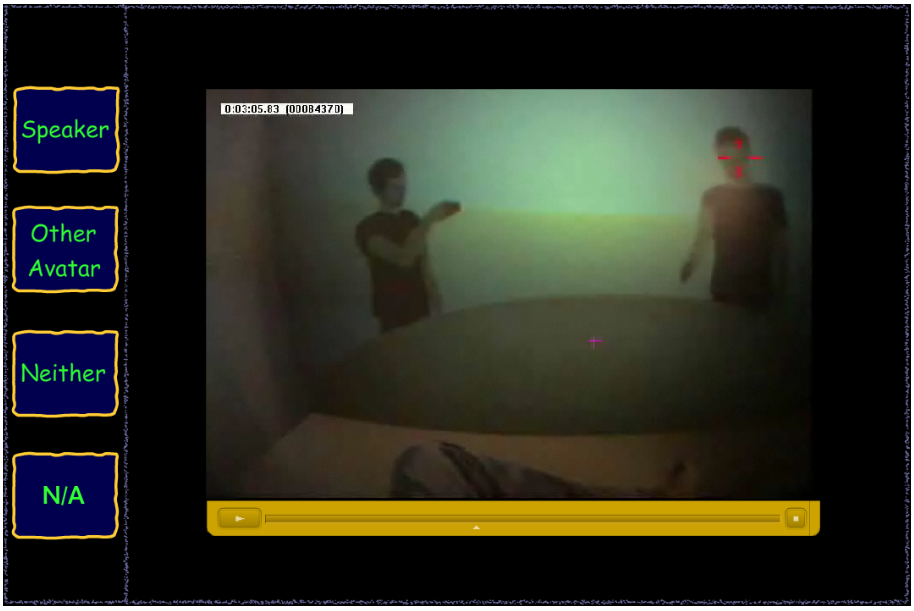


Figure 4.14: *Gaze distribution analysis application. Participants' gaze while being asked a question was classified into four groups: at speaker, at non-speaking confederate, away from both confederates, and ambiguous.*

4.4 Discussion

In both AMC and VMC conditions, interactants were able to signal using their own gaze and observe others' gaze. Gaze was also recorded to act as the primary data source for analysis. Analysis of the experimental interactions concerned three broad behavioural categories: support of gaze practices, management of speaker transition, and gaze distribution. The following section discusses the results, exploring behavioural differences observed between users of the two telecommunication systems. Varying characteristics of spatiality, user representation, and emergent gaze behaviour provide particular focus for discussion.

4.4.1 Spatiality

In the three-party conversational scenario, participants were free to choose how to distribute gaze between the two confederates when responding to questions, issued by one party. CA presented in Figure

4.7 detailed the common tendency in VMC for participants to direct gaze more exclusively towards the questioning confederate when answering a question. In contrast, Figure 4.8 presented practice more common to AMC, detailing how participants alternated their gaze between both the questioner and the bystanding confederate when responding to a question. This characteristic difference in gaze behaviour is likely to arise from the way in which the two technologies present the shared workspace.

Studying a non-immersive CVE, Hindmarsh et al. [HFH+98] observed that fragmentation, due to the narrow field of vision provided by the displays, of content in the shared workspace led to problematic interaction. In particular, the authors noted that users found it difficult to re-assemble the relation between another user's avatar and where it was indicating, thereby undermining the ability to assemble the coherence of the scene. In the current experiment, the particular VMC setup significantly separated the video displays of remote confederates, which limited participants' to only be able to direct attention to a single party at any one time. Accordingly, participants' ability to observe both confederates simultaneously, by means of peripheral vision, was hindered due to the high degree of separation of the displays. This limitation was demonstrated during the experimental interactions by sequences of problematic conversational management. As presented in Figure 4.10, such situations were particularly prone to occur following the end of a participant's response to a question that one or both confederates deemed to be ambiguous as to whether it should indeed be treated as the end. In contrast, the AMC supported by Eye-CVE was performed in a shared and seamless workspace that did not limit positioning of users' avatar embodiments. This allowed the three parties engaged in the AMC to adopt a spatial formation that was perceptually similar to a round-table meeting that may take place in reality. As a result, participants were able to attend to, and observe, both confederates more efficiently, with minimal shifts in head orientation. This allowed, for instance, head orientation to be directed toward one confederate, while directing gaze towards the other. This unified shared workspace, together with full-body representation of users that allowed for observation of gross gesture, suggest the potential of ICVEs as a platform that is highly conducive to the support of multiparty conversation and gaze practices.

4.4.2 User Representation

The spatial advantages of the immersive AMC over the fragmented display characteristics of the VMC were somewhat diminished, however, by the primitive nature of the visual and behavioural fidelity of the avatar embodiments compared to live video. VMC presents users actual appearance, enabling rich signalling and observation of facial expression, oculesics, and mouth movement according to vocal utterances. In contrast, AMC embodied users as generic virtual humanoids featuring faithful replication of gaze behaviour as the single dynamic characteristic of facial expression. Hence, whilst the tracking of single arm allowed for both gross and subtle gesture, and head tracking enabled posture shifts while maintaining communicational gaze, much of the richness of users' expressive facial behaviour, including mouth movement, was not captured or transmitted during AMC. The impact of this disparity in fidelity between VMC and AMC surfaced in the experimental interactions in the form of problematic sequences of conversational management during AMC. In particular, confusion arose from the lack of avatar mouth movement, resulting in moments of participants being uncertain with regards to which confederate was

Figure 4.15: *Participant engaged in AMC deploying gaze, marked by crosshair, generally toward confederate avatars in early stages of interaction (*left: after 10 seconds*), to explicit targeting of confederate avatars' eyes, seeking to establish mutual gaze or seek information regarding attention and action in later stages (*middle and right: after 52 seconds and 1:43 minutes respectively*).*

currently talking. As described in Section 4.3.2, confederates compensated for this deficiency with exaggerated use of arm gesture to both pass control of the conversational floor, and to hold the floor. Confederates often employed such gesture in combination with the additional compensatory technique of talk, intending to make turn-taking explicit, and to guide participant gaze. While these techniques were generally successful as means to substitute action and visual conduct that are recurrently implicit in collocated conversation, they reveal deficiencies in the avatars' display of nonverbal cues.

4.4.3 Emergent Gaze Behaviour

In both the AMC and VMC, participant gaze behaviour was similar, and comparable to collocated interaction. Participants did not solely look away from confederates, nor gaze fixedly at them, and gaze was dependent on the current state of the ongoing interaction. However, a particular longitudinal phenomenon was observed in the AMC sessions with regards to how participants utilised gaze as the experiment progressed. During the early minutes of interaction, it was observed that two of the five AMC participants directed their gaze *generally* towards the confederate avatars, as opposed to focusing on the face or attempting to establish mutual gaze. However, as the interview progressed, these participants began to consistently target confederate avatars' eyes. This emergent behaviour is illustrated in Figure 4.15, and may be explained by participant's initial inexperience of system, which, as it increased, led to the realisation of the relevance of other avatars' gaze behaviour, and subsequently elicited more natural gaze behaviour from the participant themselves. Participants had no previous experience of immersive VR, and hence, the novelty of the scenario of interacting with remote people via life-size avatar embodiments was likely to elicit different attentional and gaze behaviour during acclimatisation to the medium. From analysis of eye tracking videos of the two participants who displayed the emergent gaze behaviour, the change from general targeting of confederate avatars to specific gaze directed at the eyes and face occurred towards the end of the first minute of the interview, as marked in Figure 4.15. This, more socially normal gaze behaviour, particularly of attempting to establish mutual gaze, remained in practice for the remainder of the interactions, and was comparable to the other three AMC participants, who demonstrated such gaze strategy throughout.

As discussed in Section 4.2.2, the VMC setup allowed only head and shoulders views of confederates. Hence, when looking towards confederates, the potential for participants to look anywhere apart from their faces was limited in comparison to the full-body and unified environment presented during

AMC. Hence, while direct comparison of longitudinal trends in gaze behaviour between AMC and VMC would not be valid, all VMC participants sought to establish mutual gaze with confederates, throughout the entire course of the interactions. This was implied quantitatively in the gaze distribution analysis in Section 4.3.3, which demonstrated significantly higher proportion of gaze directed at the questioning confederate during VMC than during AMC. Combined with observations of emergent gaze behaviour in AMC, this suggests that participants fostered a higher degree of social presence when confederates were represented by live video than by avatar embodiments. The limitations of the basic avatars used in this experiment are likely to have influenced participant gaze behaviour during AMC. In particular, a lack of mouth movement, facial expression, and eyelid movement were likely to be detrimental to the social realism of the embodied virtual humanoids.

4.5 Chapter Summary

The experiment presented in this chapter aimed to compare the behaviour of participants engaged in a triadic conversational scenario performed using tracked gaze AMC and gaze aware VMC. An early version of EyeCVE supported the AMC, while AG nodes at the three interacting sites were setup as optimally as possible to enable gaze awareness between interactants. A semi-structured role-play was formulated, featuring two confederates interviewing a single participant. Qualitative CA combined with eye tracking indicated that interaction in both mediums is influenced by technical characteristics of the system, particularly with regard to workspace fragmentation and user representation. However, interactants were seen to adopt strategies to successfully manage conversational turn-taking in both mediation types: in VMC, confederates generally employed gaze and facial expression to signal transfer of the conversational floor, while in AMC, this was usually handled through hand gesture combined with gaze and head orientation.

The findings support the hypotheses that multi-party conversation is able to be conducted adequately, and that gaze behaviour will be employed similarly during both AMC and VMC for purposes of both nonverbal expression and management of the unfolding interaction. Gaze behaviour in both telecommunication mediums approximately followed established social norms observed in collocated interaction. This experiment was purposefully less formal, in terms of both design and analysis, and more exploratory, in terms of the general issues regarding gaze-enabled telecommunication systems, than the experiments detailed in the forthcoming two chapters. Discussion made both general and specific observations regarding the advantages and deficiencies of the two mediums. In particular, limitations of EyeCVE's avatar system were uncovered, and are subsequently developed and explored over the following chapters. Regarding data analysis, a critical limiting factor in this experiment was the MobileEye eye tracker's low tracking robustness and the lack of detailed logging ability. The units are replaced by the ViewPoint eye tracker in the remaining work presented in this thesis.

Chapter 5

Experiment: Object-Focused Scenario

This chapter presents an experiment designed to evaluate the impact of varying methods of avatar gaze control in a three-party object-focused collaborative task. The experimental design takes Schroeder et al.'s [SSA+01] simplified Rubik's cube study as a primary influence. In that experiment, performance and subjective experience was compared over varying technologies (reality, immersive AMC, and desktop AMC) in a collaborative task performed by two naïve participants. The participants worked together to solve the 3D puzzle, and no structure was imposed on the interaction, which progressed based upon the action of, and communication between, the two participants. Accordingly, in their post-experimental analysis of the AMC, Schroeder and colleagues sought to explore their paper's titular question: *is collaboration in ICVEs as good as being there together?*

In contrast, the experiment presented in this chapter imposes a structure on the experimental AMC, defining a sequence of interactional subtasks performed between two confederates and single participant. Informed by initial piloting of an object-focused scenario, the structured interaction was designed to emphasise the communicative importance of gaze as an interactional resource in AMC, and also to enable rich analysis of the recorded data. Hence, while Schroeder and colleagues were concerned with how varying technologies may influence collaboration, the current experiment focuses on the impact of varying methods of avatar gaze control on quality of communication in immersive AMC. The goal of the experiment was to compare tracked gaze, a simple gaze model, and no eye movement during an object-focused task. The methods of analysis are extended from those employed in Chapter 4, and include task performance, subjective user experience, and interaction analysis using CA coupled with eye tracking data.

5.1 Experimental Aims and Expectations

The aim of this experiment was to assess the benefit of tracked gaze AMC in a multiparty task that required physical movement and object manipulation. Gaze aware VMC systems preclude users' ability to move freely within their local environment while maintaining meaningful communicational gaze with remote others. Hence, in contrast to the experiments presented in the preceding and subsequent chapters, the current experiment does not investigate VMC, and instead focuses on varying conditions of avatar gaze control in immersive AMC. The ability to both manipulate and share virtual objects in a

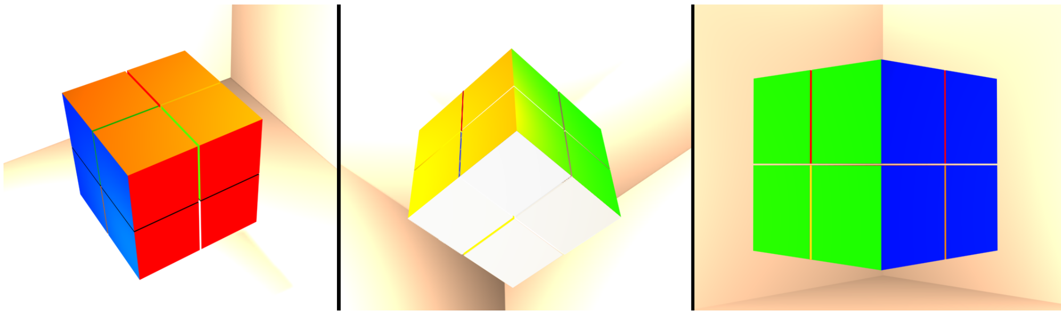


Figure 5.1: *Views of the completed 'Rubik's cube' puzzle, formed by arranging eight small cubes to form one large cube in which each side displays exactly one colour.*

unified shared workspace is a feature unique to ICVEs, and forms an integral part of the experimental interactions. The overall goal of this experiment was to investigate differences between varying methods of avatar gaze control, as measures of task performance, subjective user experience, and interactional behaviour.

The primary hypothesis was that tracked gaze AMC would be able to support a higher quality of communication, measured by task performance, subjective user experience, and interaction analysis, than both static and modelled gaze AMC. Specifically, when all interactants are embodied by tracked gaze avatars, this form of AMC is expected to outperform the two other classes of avatar gaze control in the following ways: action will be performed faster and with fewer errors; participants will rate overall communication as more natural, and instructions issued by interactional partners as less ambiguous; during interaction, participants will demonstrate gaze practices that are more comparable to those observed during collaboration in the real-world; and finally, when a mistake is made during the collaborative task, participants will be able to more effectively employ repair strategies to remedy the situation.

5.2 Experimental Design

A within-groups repeated measures design was used, with twelve participants performing three consecutive experimental sessions in each of the three avatar gaze conditions of tracked gaze, simulated gaze, and static gaze. Similarly to the previous conversation experiment, participants took part in three-party interactions comprising of themselves and two confederates: one located at the University of Reading (UoR) and the other at the University of Salford (UoS). The same two male confederates performed interactions with all twelve male participants. The main component of interaction involves the two confederates alternately issuing instructions to the participant, who performs the requested action. Once the instruction has been completed successfully, the next instruction may be issued. This process continues until the entire sequence of instructions has been issued, and the goal state of the puzzle is achieved; at which point the current experimental interaction concludes. Figure 5.1 shows the puzzle's goal state, rendered from various angles, illustrating the arrangement of eight smaller cubes to form a single, larger cube, with each side displaying exactly one colour.

The core of the collaboration is the interactional sequence of confederates issuing instructions to a participant, and the participant performing the requested action. Each experimental session comprises a total of ten instructions toward achieving the goal state of the puzzle. These ten instructions are logically paired, resulting in five pairs per experimental session. Each instruction pair comprises one *grab* instruction, and a follow-on *position* instruction, both of which are issued by the same confederate. Grab instructions request the participant to take action to identify and pick up a specific cube being indicated by the instructing confederate. Subsequently, position instructions require the participant to identify and place the grabbed cube in the specific position, as indicated by the instructing confederate, within the cube puzzle. Following the participant's successful grab and position action, the bystanding confederate (i.e. the confederate who did not issue the previous instruction) assumes control of the floor, and issues the next instruction pair. This process continues until all five instruction pairs have been issued, and the goal state of the puzzle is achieved.

In summary, this study does not present a symmetrical collaborative task as seen in Schroeder et al.'s experimental design [SSA+01]. Rather, the use and importance of gaze as a bidirectional resource in AMC, for both indication and observation of attention and action, are investigated over three varying methods of avatar gaze control. The following sections detail the experimental design. For additional materials, see Appendix D.

5.2.1 Independent Variables

The three methods of avatar gaze control are listed below in ascending order of their expected contribution to quality of communication. Participants performed a total of three experimental sessions (one in each condition), in which both confederate avatars and their own avatar's gaze was controlled by identical methods described by the conditions. Hence, the following independent variables are termed *static gaze AMC*, *modelled gaze AMC*, and *tracked gaze AMC*. The order of conditions were resequenced over the twelve participants to negate the effect of learning and experience. Additionally, three VEs with different starting configurations were also built, and presented to all participants in the same order. The VE design is described in Section 5.2.3.

1. *Static gaze AMC:* In this condition, avatars' eyes are centred and do not exhibit any movement. This is the most basic form of AMC investigated in the experiment, and aims to provide a baseline from which the latter two conditions may be judged. When operating with static gaze, the attention of others must be inferred from alternative, tracked, nonverbal cues such as head orientation and gesture.

2. *Modelled gaze AMC:* In this condition, a basic simulation of gaze behaviour controls avatars' eye movement. The characteristics of the model are described in Section 5.2.3. A central feature of modelled gaze avatars is that they display gaze behaviour that is different to their embodied user's actual eye movement.

3. *Tracked gaze AMC:* In this form of AMC, avatars' gaze mimics that of their controlling users. Thus, users engaged in tracked gaze AMC will be able to both signal using their own gaze, and observe others' gaze behaviour. The head-mounted eye tracking setup used to achieve this high-fidelity form of AMC is outlined in Section 5.2.3.

5.2.2 Piloting

Prior to running the experiment as it is documented in this chapter, the experimental design was refined over several iterations. The experiment aimed to assess the benefit of tracked gaze AMC in a multi-party task requiring physical movement and object manipulation. Initially, pilot sessions involving a three-party version of Schroeder et al.'s [SSA+01] symmetrical collaborative task, were carried out, investigating the same independent variables of gaze control described above in Section 5.2.1. However, while thorough interaction analysis would have likely uncovered subtle differences in communicative behaviour, it quickly transpired that task performance as a whole seemed relatively unaffected by avatar gaze condition when users were engaged in such unstructured interaction. Two main reasons may explain this finding: the puzzle task is extremely involving, and while the three participants often verbally communicated, the success of the object-centric action did not necessarily require those involved to pay close attention to the visual presence of their fellow collaborators. This was compounded by the approximately hip-level height of the puzzle base (see Figure 5.1) displayed in the CAVE systems, leading to participants generally adopting a downward-gazing posture. Secondly, three collaborators appeared to be one too many for the particular task, and one or two members generally emerged as dominant, thus reducing the cooperative nature of the task. A subsequent version of the experiment was also trialled, with two confederates acting as aides to a single participant, suggesting and helping them toward solving the puzzle. This design also suffered from the aforementioned problems.

Experience with pilot experiments suggested that, in order to investigate the operational benefit of using tracked gaze avatars in object-focused scenarios, the experimental task must emphasise the communicational importance of gaze during the collaborative interactions. Preliminary analysis of pilot experiments also suggested that some form of structure should be imposed on the interactions in an attempt to minimise the intervening variable of participant ability and approach, and also to enable data collection of specific and repeated sequences of action. The paradigm of an instructional scenario was developed, transforming the initial symmetrical and open-ended collaborative puzzle solving into a series of sub-tasks acting as components of a whole, and which featured the same eventual goal state. Pilot studies involving a single participant being instructed by two confederates were conducted successfully, and evolved into the final experimental design as described in Section 5.2.5. The role of confederates was established as a key element of interaction, serving multiple purposes: the communicative importance of fellow users' visual presence was emphasised; participant behaviour was controlled and constrained by a defined sequence of instructions; and the time taken from beginning to goal state of the puzzle was significantly reduced by the instructional design and expertise of confederates. From the practical perspective of experimental duration, which dictated the amount of time participants would be required to spend within the CAVE, the redesigned task enabled a within-subjects design in which participants performed experimental sessions in each of the three avatar gaze conditions. Finally, and as explained in Section 5.2.3, the difficulty of the task was increased by the addition of several "dummy" cubes, alongside the "puzzle" cubes, into the VEs. The dummy cubes acted as decoys for the participant when carrying out grab instructions issued by confederates.

5.2.3 Apparatus

EyeCVE supported the object-focused AMC between the three CAVE laboratories located at UCL, UoR, and UoS. At all sites, ViewPoint eye trackers were fitted to the local user, recording gaze. The following sections present the arrangement of the technical apparatus.

EyeCVE Setup

The EyeCVE setup for this experiment was similar to that described in Section 4.2.2. A more mature version of EyeCVE was used, featuring improved network performance, and, due to the integration of the ViewPoint eye trackers, more frequent gaze updates to the avatar subsystem when tracked gaze AMC was in operation. The basic avatars documented in Section 3.1.4 were used, capable of displaying the oculesic behaviour of gaze, but not supporting eyelid kinematics or changes in pupil size. Similar to the previous conversation experiment, EyeCVE ran on SGI Prism™computers at UoS and UoR. However, UCL migrated to a Microsoft Windows®XP cluster, which operated with identical display properties: 1024×768 pixels per wall, projected at 96-120 Hz in active stereo. Head and hand tracking remained identical at all sites to that described in Section 4.2.2. UoS ran the centralised server, which maintained the state of the shared VE and distributed changes to all clients. Wireless microphones were worn by all interactants, and audio communication was achieved using a Skype [Sky10] conference call, and output through speakers located in each CAVE.

Three VEs featuring different starting configurations for the three experimental sessions were built. As the independent variables were resequenced over the twelve participants, the presentation order of the VEs remained static in order to negate the impact of any variation in difficulty of starting configuration. The spatial volume of the VEs was approximately equal to that of the CAVE systems (3××3×2.2 m), allowing participants and confederates to navigate to any position within the VE naturally, through physical movement, as opposed to requiring the use of an additional input device. Each VE was populated with a configuration of eight puzzle cubes and five dummy cubes. The puzzle cubes were coloured in order to have specific positions within the overall 2×2×2 goal state, while the dummy cubes appeared to be similarly coloured, but would not fit correctly into the finished puzzle. The purpose of the dummy cubes was to increase the difficulty of the grab instructions, by presenting participants with a greater number of potential choices. The quantity of five dummy cubes was informed by the piloting work described in Section 5.2.2. Critically, the dummy cubes appeared differently in the participant display at UCL compared to in the confederate displays at UoR and UoS: to participants, dummy cubes appeared coloured, and thus indistinguishable from the puzzle cubes; however, to confederates, dummy cubes appeared black on all sides. This intended to aid confederates when issuing grab instructions, by reducing the number of possible cubes (i.e. non-black) they may indicate, towards completing the puzzle.

Each VE starting configuration featured three of the puzzle cubes positioned correctly to form the base of the solution, while the remaining five puzzle cubes were scattered around the VE. The five additional dummy cubes were also scattered around the VE close to the puzzle cubes. Views of the three starting VEs, from perspectives of both the participant and the confederates, are shown in Figure 5.2.

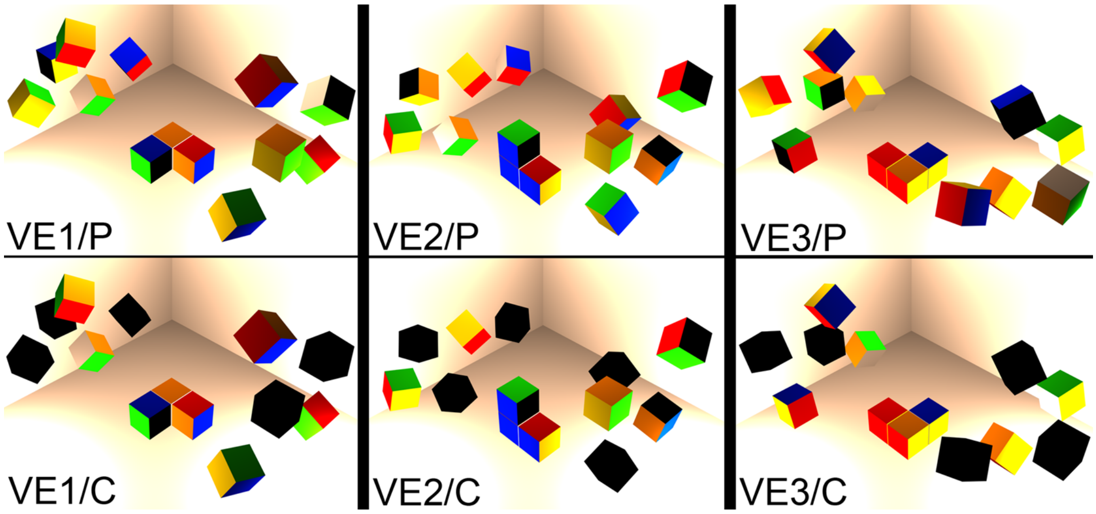


Figure 5.2: *Renders of the three starting configuration VEs. The top row of images are marked* P, *and show the VEs as they appeared in the participant display. The bottom row of images are marked* C, *and show the VEs as they appeared in the confederate displays. In the former, dummy cubes are indistinguishable from puzzle cubes, while in the latter, dummy cubes are displayed black on all sides.*

Eye Tracking

ViewPoint eye trackers were mounted on the CAVEs' CrystalEyes 3 [Rea10] shutter glasses as illustrated in Figure 3.4. The eye trackers provided robust data capture in varying lighting conditions with all wearers, including when they physically walked and moved within the CAVE. Scene cameras recorded action with a wide 150° FOV, granting peripheral view of the scene close to the 180° of humans [Cos95]. The wearer's foveal fixation point is overlaid on the scene camera video, allowing for real-time observation and post-analysis of gaze behaviour once synchronised with the separately recorded audio.

Gaze Model

Given the widespread use of gaze models for avatars and agents alike, it was important to integrate an autonomous simulation of eye motion into the conditions of gaze control in the current experiment. The gaze model developed for this experiment is basic, and aims to provide a baseline performance measure for future work into gaze modelling, covered in Chapter 7.

For a gaze model to generate eye movement which has potential to be interactionally-relevant, the particular application and scenario of the VE must be considered. The current scenario features triadic object-focused collaboration, including object manipulation, conversation, and free movement. Hence, users' action is highly dynamic and varied, precluding the option for an approach simply based on interactants' talking and listening state. Rather, in the scenario presented in the current experiment, a user's current FOV of the virtual scene, and positions of objects contained within, is a more appropriate form of input to gaze generation. The approach of the current gaze model firstly calculates a user's current FOV, determined by head tracking, and the considers objects and avatars lying within that region, to which gaze is distributed. Figure 5.3 illustrates top- and side-views of the gaze model's input frustum, the origin of which lies at the centre of the avatar's eyes.

Boff and Lincoln specify the human eye's maximum ideal horizontal and vertical foveal rotation to

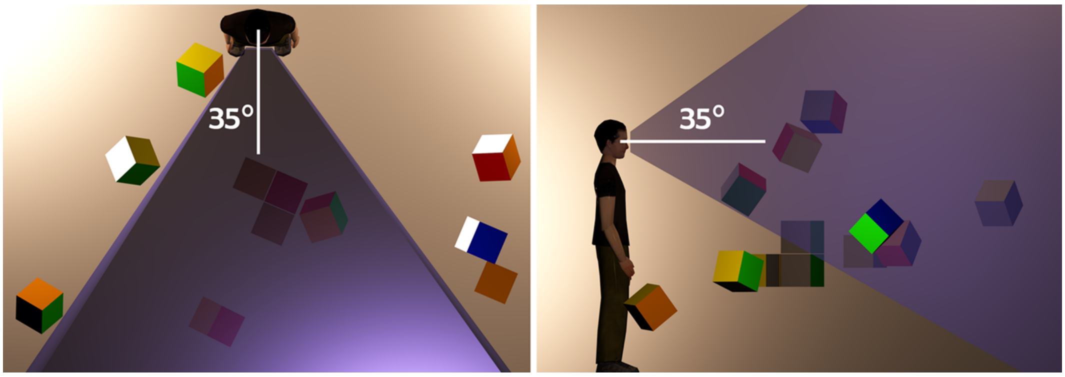


Figure 5.3: *Gaze model input frustum.* Left: *Top view.* Right: *Side view. A user's current head position and orientation is used to calculate a FOV frustum (blue shaded area) specified at 70° vertical and horizontal. Objects and other avatars lying within the frustum are targets for gaze distribution, which the algorithm chooses at random.*

a fixation point at 35° [BL88]. The frustum approximates this general limit of eye rotation, defining a FOV of 70° both horizontally and vertically from the head-centric vector, leading to a maximum avatar eye rotation of 35° on each axis. Algorithm 1 details the operational behaviour of the gaze model. During action in EyeCVE, saccades and fixations are randomly distributed between targets (cubes and other users' avatars) within the current FOV. The position and orientation of the FOV is dynamic, and is inferred from head tracking data. Thus, potential targets enter and exit the FOV, and saccades and fixations are generated. The duration of fixations are determined by a random-sampling method which varies with head motion. The temporal coding of fixations is therefore dependent on timing and velocity of head movement: reduced activity generates fewer saccades with longer fixations, while rapid motion results in greater numbers of saccades and shorter fixations. A uniformly distributed random number was generated every second, which determined which target in the current FOV was foveated. In summary, the gaze model takes a user's current FOV, inferred from head position and orientation, as input, and distributes gaze randomly to objects located within. The model is naïve with regards to how gaze is distributed between objects, and does so randomly. The performance of the model is addressed in the forthcoming experimental analysis, in Section 5.3.

5.2.4 Population

Twelve participants with normal or corrected-to-normal vision and no previous experience of VR were recruited from UCL's campus through an advertising poster and email campaign. They were paid £10 to perform the study, the duration of which was between one and two hours depending on the participant's ability. Similar to the prior conversation experiment, both confederates were male, so in order to minimise possible effects of gender, only male participants were recruited. The twelve participants performed three experimental sessions: once in each avatar gaze condition. As depicted in Table 5.1, exposure was resequenced over the twelve participants to negate learning bias in the within-subjects design. In each session, all avatars exhibited the same method of gaze control. Confederates were not informed of their own avatar's gaze condition, nor were they informed of others', and were also told that each avatar in the interaction may feature a different gaze condition. In reality, all avatars operated with

Algorithm 1 *Random gaze model* computes target object (o_x, o_y, o_z)

```
Require: scene database of objects {O_1, O_2, O_3, ...O_n}
Require: field of view, fov = 70° (i.e. eccentricity, θ ≤ 35°)
for each frame do
    seed random to 1 second {determines fixation duration}
    compute avatar's eye location in world coordinates
    for each object in the scene database do
        determine object's location in world coordinates
        compute eccentricity, θ {equation 7.1}
        compute vertical angle, θ_v {equation 7.2}
        if (θ < 35°) and (−35° < θ_v < 35°) then
            add to list of objects within field of view
        end if
    end for
    pick randomly from objects within field of view
    aim avatar's eyes at centre of selected target object
end for
```

Table 5.1: *Order of independent variables presented to the twelve participants.*

Order	Participant Number					
	1 & 7	2 & 8	3 & 9	4 & 10	5 & 11	6 & 12
1st	Static	Static	Model	Model	Tracked	Tracked
2nd	Model	Tracked	Static	Tracked	Static	Model
3rd	Tracked	Model	Tracked	Static	Model	Static

the same gaze condition during an experimental session. These measures aimed to minimise differences in confederate behaviour due to knowledge of their own avatar's gaze capability. Participants were not informed of the significance (or insignificance) of other avatars' eye movements, nor the ability (or inability) to communicate using their own gaze. Finally, the setup procedure preceding each of the three sessions was identical, aiming to keep the participant's experience prior to each session consistent, and also so confederates would have no information with regards to participant gaze condition.

5.2.5 Procedure

The experimenter at UCL was responsible for organising and running the experiments as well as local technical setup. The two confederates at UoS and UoR maintained their local site's technical setup, and took part in the object-focused interactions supported by the EyeCVE. The experimental procedure consisted of seven major phases: briefing upon participant arrival; technical setup; eye tracker calibration; the three experimental sessions, each followed by a questionnaire; and finally a post-experimental interview.

Phase 1: Briefing

Upon arrival at the CAVE lab, participants were welcomed in a reception area, and were given an information sheet and ethical consent form. The information set the scene for the forthcoming object-focused tele-collaboration scenario. The participant was left in private to fill out the forms, and was informed that they may ask any questions not directly relating to the experimental procedure or technical setup.

Phase 2: Technical Setup

While the participant read and completed the forms, EyeCVE was set up between the three sites. A conference-call hosted using Skype [Sky10] was established, wireless microphones were initialised, and talk from remote partners was output from speakers positioned in each lab. The ViewPoint eye trackers at each site were initialised. UoS ran the EyeCVE server, specifying the first VE starting configuration as illustrated in Figure 5.2. Each site connected EyeCVE clients to the server, which loaded the graphical scene and positioned all avatars appropriately in the VE.

Phase 3: Eye Tracker Calibration

Confederates entered their local CAVE, wearing the shutter glasses with mounted eye and head tracker. Initial gaze calibration in the ViewPoint software, and subsequently in EyeCVE, was performed. Depending on the preconfigured script, which was unknown to confederates, avatars would exhibit static, modelled, or tracked gaze. Confederates held the hand tracker in their right hand, and ensured correct positioning of their wireless microphone.

The CAVE displays and audio communication were muted in the UCL lab in order to eliminate visual distraction and overhearing any talk between confederates. The participant entered the lab and was fitted with the shutter glasses with mounted eye and head tracker. The two-phase gaze calibration procedure was performed, and the participant avatar's gaze control method was set according to the experimental script. The experimenter checked the accuracy of the calibration and made adjustments if required. The wireless microphone was attached to the participant, who also held the hand tracker in their right hand. Finally, completing the pre-experimental phase, the experimenter at UCL re-enabled the audiovisual displays, and issued a verbal signal to the confederates. The experimenter at UCL then left the lab and started the data logging processes.

Phase 4a: Experimental Interaction 1

As the first experimental session began, participants were greeted by the confederates within the VE. Confederates performed a brief training session, teaching the participant how they may naturally move within the virtual space, and how to grab, move, and rotate cubes using the hand tracker. Once the participant was comfortable with system use, the confederates introduced the experimental task, explaining that they needed help from the participant to finish the semi-completed puzzle. The puzzle's goal state was defined as forming a larger cube from eight smaller cubes so that each side, including the bottom, consisted of a single colour, in a similar manner to the classic Rubik's cube. It was then made clear to the participant that they were not required to solve the puzzle themselves; rather, they would be guided by the confederates over a series of instructions, which, if followed correctly, would result in the correct solution to the puzzle. Therefore, the success of the interactions depended on the participant's ability to efficiently follow instructions, rather than to solve the puzzle themselves.

Confederates commenced to issue a defined sequence of instructions (which varied between each of the three VE starting configurations). Instructions were defined as a pair, comprising one grab instruction and one position instruction, for each of the five missing cubes. Grab instructions required the participant to identify and pick up a specific cube from within the assortment of puzzle cubes and dummy cubes scattered around the VE. Position instructions then required the participant to place the correctly grabbed cube into a specific position in the puzzle. Confederates took turns in issuing instructions, with the confederate at UoR starting the sequence, and hence issuing three instructions per session, while the confederate at UoS issued two.

During the instructional period of the interaction, both the verbal and nonverbal communicational behaviour of the confederates was restricted. In periods between issuing instructions, confederates were allowed to move freely. However, when issuing instructions, their movement was restricted to just head and eye movement, thereby limiting gesticulation to the channels of NVC under experimental investigation. Confederates were not allowed to point at or grab cubes. Verbally, no information describing the colour or specific location of cubes was allowed, thereby eliminating information likely to influence the recorded metrics. Hence, the verbal component of grab instructions involved variations on the generic phrase, *"pick up this cube"*. Accordingly, the verbal component of position instructions were issued with a term similar to *"place the cube here"*. In summary, both the verbal and nonverbal behaviour of confederates was tightly controlled during the instructional phase of the experiment in order to emphasise the use of gaze direction and head orientation. Confederates were free to answer questions that did not reveal the indicated cube or position, and participants willingly took to their role, becoming highly engaged in the task.

During a typical instruction, a confederate would issue a grab instruction with a verbal utterance as described above, while indicating the intended cube with gaze and head orientation. The participant would use this information, and take action to grab a cube. If the participant identifies an incorrect cube, the confederate would inform him, and continue to indicate the intended cube as before, with gaze and head direction. Once the correct cube had been identified and grabbed, the same confederate would then issue a position instruction with a verbal utterance, together with gaze and head orientation, to indicate the cube's correct placement within the puzzle. If the participant's response to the placement instruction incorrectly identifies the cube's intended position, the confederate would inform him, and continue to indicate the correct position with the defined cues. Once the cube had been correctly positioned, the current instruction came to an end, and the bystanding confederate began the next in the sequence. This process would continue until all five instruction pairs had been completed successfully, and the puzzle's goal state had been achieved. Figure 5.4 shows three typical interactional states during an experimental session.

Phase 4b: Questionnaire 1

Following the completion of the first experimental interaction, data logging from both EyeCVE and the eye trackers were stopped. The experimenter at UCL paused the local audiovisual displays, and entered the CAVE to stand opposite the participant. The participant's comfort was checked, and the experimenter proceeded to verbally issue a questionnaire. Due to the lengthy setup and eye tracker calibration procedure, and in order to maintain their position within the CAVE, participants did not remove the shutter glasses. The questionnaire was performed orally, as the participant's ability to read and write was hindered by the shutter glasses. The questionnaire is presented in Section 5.2.6 and took

Figure 5.4: *Screen captures of an experimental session.* Top*: The confederates (wearing UoS and UoR shirts) greet the participant (wearing a UCL shirt) before commencing the task. The initial puzzle base is located between two groups of loose dummy and puzzle cubes. Users may navigate in the VE with natural physical movement.* Middle*: The participant responds to a grab instruction by reaching for an indicated cube.* Bottom*: The participant responds to the final position instruction by positioning the cube to complete the puzzle.*

approximately two minutes to complete. During this time, the confederate at UoS loaded the second VE starting configuration on the EyeCVE server, and the experimental script altered avatar gaze condition accordingly. The experimenter at UCL then left the lab and data logging was resumed.

Phase 5a: Experimental Interaction 2

The second experimental interaction, featuring a different avatar gaze control method, and located in the second VE starting configuration began. Action followed identical to as described in Phase 4a, but the initial training was naturally not performed.

Phase 5b: Questionnaire 2

When the second experimental interaction had concluded, the same questionnaire and accompanying technical procedure described in Phase 4b was performed. The third VE starting configuration was loaded, ready for the final session.

Phase 6a: Experimental Interaction 3

The third and final experimental interaction, featuring the remaining avatar gaze control method, and located in the third VE starting configuration was performed. Action followed identical to as described in Phase 4a.

Phase 6b: Questionnaire 3

The questionnaire was issued by the experimenter at UCL. All data collection was stopped, and devices worn by the participant were removed.

Phase 7: Interview

The experimenter at UCL performed an informal interview with the participant. Initial discussion revolved around the general experience of the AMC, which gradually addressed specific details of other avatars' behaviour and any observed differences between the three experimental sessions.

5.2.6 Data Collection

With ethical clearance, the following data was recorded at all sites in all sessions:

- *Eye Tracking:* Video and audio streams, recorded from the three eye tracker scene cameras. The wearer's gaze direction is overlaid on the video. Eye tracker log files, recording gaze behaviour were also recorded.

- *EyeCVE:* Log files recording changes to the EyeCVE scene graph, including avatar movement and object interaction, were recorded at each site. These log files allow both for offline analysis and replay of sessions using the tool mentioned in Section 3.2.4. The replayer application operates in the CAVE, and provides identical visualisation to the original captured EyeCVE sessions, free camera movement, and additional data such as gaze heat-maps.

- *Gaze Model:* Gaze generated by the gaze model was recorded. This log file is similar to the main EyeCVE log, but explicitly records data regarding the behaviour of the simulated gaze.

Following each experimental session, the questionnaire presented in Table 5.2 was issued. The questions were designed to elicit responses regarding subjective experience, copresence, and self-performance ratings. Responses were recorded on a 1..7 Likert scale.

5.3 Analysis

Analysis of experimental data was approached from three perspectives, each aiming to explore components of the object-focused AMC under the varying gaze conditions. *Task performance* aimed to classify the success of the collaborative interactions objectively, exploring metrics related to timing and errors during the experimental task. *Interaction analysis* explored differences and trends in participants' behaviour, focusing particularly on gaze patterns, bodily movement, and repair strategies. Finally, *user experience* considered questionnaire responses and interviews, aiming to elucidate subjective perceptions of the AMC. This analysis focuses on the action of the participant. However, the multiparty nature

Table 5.2: *Questionnaire issued to participants following completion of each of the three experimental sessions. Responses were scored on 1..7 Likert scale.*

#	*Question*
1	I could readily tell when my partners were concentrating on what I was doing.
2	I felt that my partners and I were in a shared space.
3	This felt like a natural interaction.
4	I was often confused about which cube my partners were indicating.
5	I was often confused about where my partners wanted me to position a cube.
6	The way my partners addressed me appeared natural.
7	It was as if I had people in front of me.
8	My partners and I frequently made eye contact.
9	I had a sense of being in the company of my partners.
10	I had a real impression of personal contact with my partners.
11	My partners seemed to be responsive to where I was indicating.
12	My partners were very expressive.
13	I was able to follow instructions efficiently.
14	The way my partners managed the task between each other appeared natural.

of the interactions ensure that confederate behaviour must also be addressed when appropriate. For additional materials, see Appendix D.

5.3.1 Task Performance

The core of the experiment's collaborative task was subdivided into a total of five paired grab and position instructions. The following analysis of task performance capitalises on this structure, separately addressing the two instruction types. The dataset from all twelve participants was analysed, and the repeated measures within-groups design resulted in twelve experimental sessions for each of the three avatar gaze conditions. The metrics of *number of errors* and *time taken* were used to assess task performance. Regarding grab instructions, an error is defined as a participant identifying and taking (or indicating) action to grab a different cube to that indicated by the instructing confederate. The time taken for a grab instruction is defined by the elapsed period from when the instruction is issued by a confederate, until when the correct cube is identified or grabbed by the participant. Regarding position instructions, an error is defined as a participant identifying and taking (or indicating) action to position a currently-grabbed cube into a different location to that being indicated by the instructing confederate. The time taken for a position instruction is defined by the elapsed period between when the instruction is issued by a confederate, until when the participant either identifies or places the currently-grabbed cube into the correct position in the puzzle. Overall puzzle completion time, defined as the elapsed period between the issue of the first grab instruction until the completion of the fifth position instruction, is also analysed, together with the impact of learning as participants become more experienced at the task.

Errors

Two-factor (3×5) repeated measures analysis of variance (ANOVA) calculations were performed on the number of errors per grab instruction and per position instruction. Input factors to the calculations were the three avatar gaze conditions, and the five grab or position instructions. Regarding grab instruction errors, a main effect of gaze condition ($F(2,4) = 5.30$; $p < 0.01$) was found. Post-hoc Tukey tests determined that significant differences lay between the gaze model condition and the other two conditions of

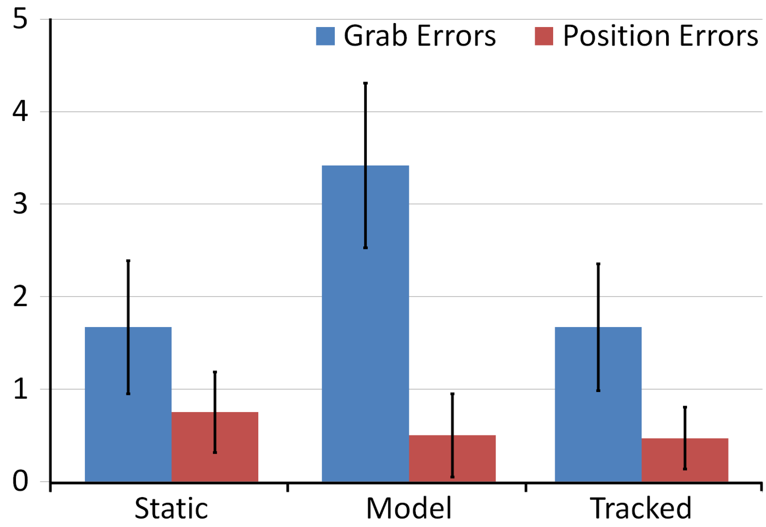


Figure 5.5: *Mean number of errors (with standard deviation) per session while performing grab and position instructions in each of the avatar gaze conditions.*

static gaze ($F(1,4) = 7.36$; $p < 0.01$) and tracked gaze ($F(1,4) = 6.84$; $p = 0.01$). A significant difference between the number of errors made during each of the five grab instructions, as an experimental task progressed, was also found ($F(2,4) = 4.91$; $p < 0.001$). No interaction effect was found between gaze condition and grab instruction.

Regarding position instruction errors, no main effect was found between avatar gaze condition ($F(2,4) = 0.64$; $P = 0.53$). However, a significant difference in the number of errors made during each of the five position instructions as an experimental task progressed was found ($F(2,4) = 5.03$; $p < 0.0001$). Similarly to grab instruction errors, no interaction effect was found between gaze condition and position instruction. The graph shown in Figure 5.5 illustrates the mean number, and standard deviation, of grab and position errors made by participants per experimental session in each of the three avatar gaze conditions. Values are presented in Table 5.4.

Time Taken

Two two-factor (3×5) repeated measures ANOVA were performed on the time taken per grab instruction and per position instruction, with the three avatar gaze conditions and the five grab or position instructions as input factors. Regarding grab instruction timing, a main effect of gaze condition ($F(2,4) = 3.67$; $p < 0.05$) was found. Post-hoc Tukey tests determined significant differences to lie between the gaze model condition and the other two classes of static gaze ($F(1,4) = 4.35$; $p < 0.05$) and tracked gaze ($F(1,4) = 4.5$; $p < 0.05$). A significant difference between the time taken for each of the five grab instructions, as an experimental task progressed, was also found ($F(2,4) = 5.0$; $p < 0.001$). No interaction effect was found between gaze condition and instruction.

Regarding position instruction timing, no main effect was found between avatar gaze condition ($F(2,3) = 0.91$; $p = 0.4$). However, a significant difference in the time taken to perform each of the five position instructions as an experimental task progressed was found ($F(2,3)=3.94;P<.0.01$). No interaction effect was found between gaze condition and position instruction. The graph shown in Figure 5.6

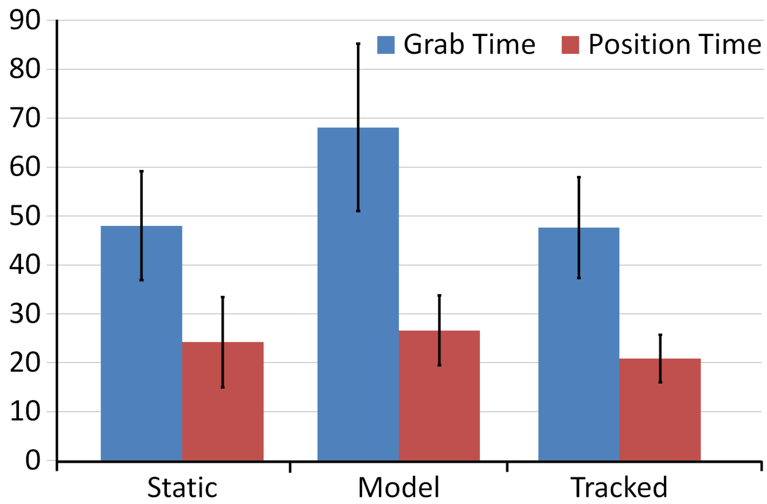


Figure 5.6: *Mean time taken in seconds (with standard deviation) per session to perform all grab and position instructions in each of the avatar gaze conditions.*

Table 5.3: *Mean time taken in seconds (with standard deviation) of each experimental session, grouped by session order.*

Order	1st	2nd	3rd
Total Time	370 (101.8)	304.3 (102)	207.9 (75.1)

illustrates the mean time taken in seconds, and standard deviation, for all grab and position instructions in an experimental task over the three avatar gaze conditions. Values are presented in Table 5.4.

A one-way ANOVA was calculated to investigate differences in total puzzle completion time over the three gaze conditions. However, no main effect was found ($F(2,33) = 0.41$; $p = 0.67$). Actual values are shown in Table 5.4, and are graphically illustrated in Figure 5.7. A subsequent one-way ANOVA was performed to investigate how completion time varied with participant experience over the three experimental sessions, using session order number (first, second, and third) as factors. A significant difference between session order numbers ($F(2,33) = 9.06$; $p < 0.001$) was found. The mean completion times in seconds, and standard deviation, for the first, second, and third experimental sessions are presented in Table 5.3, and illustrated, along with overall timing, in Figure 5.7.

5.3.2 Gaze Model Performance

Figure 5.8 aims to explicate the poor task performance observed during modelled gaze AMC. The graph compares components of participant behaviour: gaze direction (eye target, ET), the forward vector of head orientation (head target, HT), and direction of gaze generated by the model (model target, MT). Specifically, the equality of these elements is shown over all participants and sessions. For instance, “HT=ET” indicates that participants looked straight ahead (gaze direction was equal to head orientation) for a mean of 33.34% (with an 7.78% standard deviation) of the total time period of the experimental sessions. The low equalities shown for the model target comparisons (“ET=MT” and “HT=MT”) demonstrate the low likelihood of the model-generated gaze target being the same as a participant’s actual gaze direction (mean 9.2%, SD 2.8%), or their head orientation (mean 13.56%, SD 2.32%). While the primary input to the gaze model was FOV determined by head orientation, the gaze generated by the model was

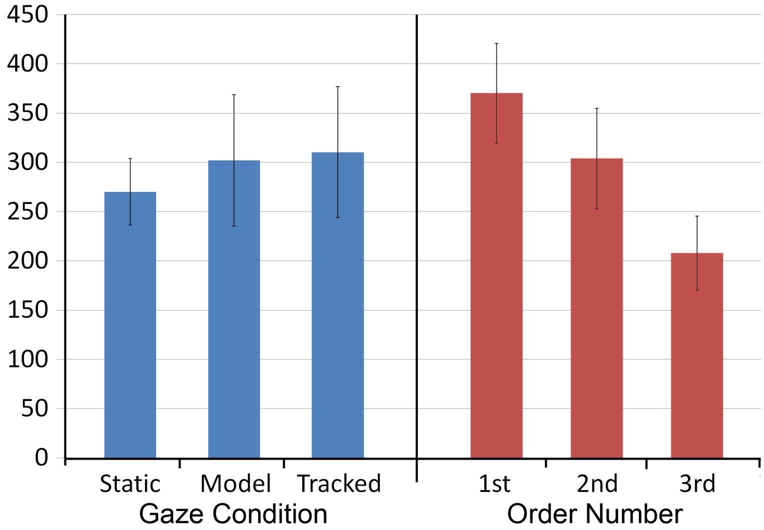


Figure 5.7: Left: *Mean total puzzle completion time in seconds (with standard deviation) in each of the avatar gaze conditions.* Right: *Mean total puzzle completion time, plotted by order of experimental session.*

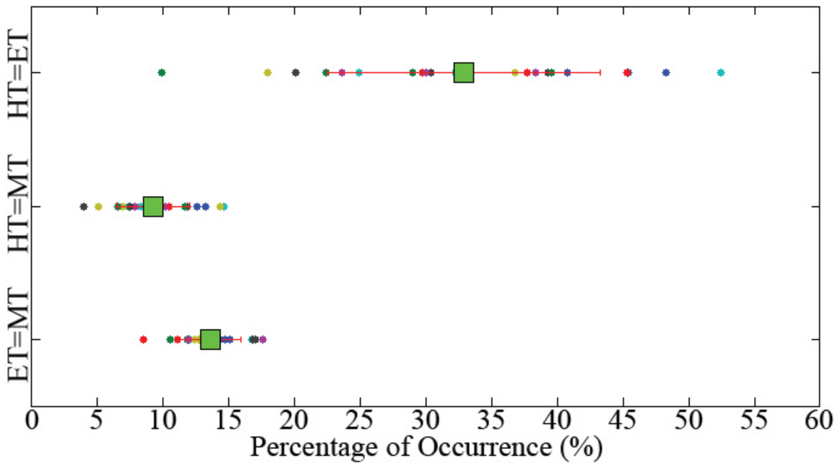


Figure 5.8: *Percentage equality of eye target (ET), head target (HT) and model target (MT) in all experimental sessions and participants. Small points denote individual participants. Square is mean. Range is standard deviation.*

generally different to both the actual gaze direction and also the head direction of the participant. Hence, the gaze model produced eye movements that were highly unfaithful to the users' actual gaze. Analysis of errors and timing demonstrated that this misleading gaze information was significantly detrimental to these metrics of task performance. Figure 5.8 also illustrates the wide variation of gaze behaviour over the twelve participants, who are each represented by a particular coloured point.

The frequency plots in Figure 5.9 show the combined frequencies of five gaze behaviours (proximity, saccade magnitude, saccade velocity, fixation duration and the eccentricity) from all twelve participants. A comparison of tracked and modelled gaze show a clear difference between the plots for each gaze parameter. The spread of the eccentricity for the model demonstrates the randomness of the targets chosen, as compared to the peakedness of actual gaze target eccentricity. The fixation duration plot also shows a peak in the gaze model's fixation durations at the 500 ms mark, which is in contrast to that of

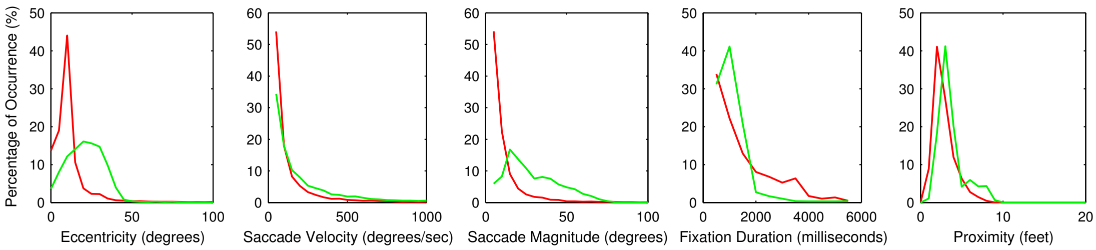


Figure 5.9: *Comparisons of gaze parameters (— tracked, — random).*

Table 5.4: *Mean task performance per session measured by* number of errors *and time in seconds, with standard deviation, for each condition.*

Condition	Static	Model	Tracked
Grab Errors	1.67 (1.44)	3.42 (1.78)	1.67 (1.37)
Position Errors	0.75 (0.87)	0.50 (0.90)	0.47 (0.67)
Grab Time	48.0 (22.3)	68.1 (34.2)	47.6 (20.6)
Position Time	24.2 (18.5)	26.6 (14.3)	20.9 (9.7)
Total Time	270.1 (67.2)	301.9 (133.4)	310.3 (132.8)

the tracked gaze.

Task Performance Summary

Regarding grab instructions, both tracked and static gaze AMC enabled participants to perform the experimental task both more accurately and in less time than modelled gaze AMC. However, no significant differences of these measures were found between the conditions of tracked and static gaze. Analysis of position instruction measures failed to expose any significant differences in number of errors or time taken between the three gaze conditions. Overall puzzle completion time was not significantly influenced by avatar gaze condition, but the order of the experimental sessions was significant, with participants requiring shorter periods of time to complete successive sessions. Table 5.4 presents an overview of the task performance analysis.

5.3.3 Subjective User Experience

Questionnaires after each of the three experimental sessions sought to elicit participants' judgement regarding user experience, self-performance ratings, and copresence. Repeated measures two-factor ANOVA, taking the three gaze conditions and the fourteen questions as factors, exposed no significant difference between conditions ($F(2,13) = 0.24$; $p = 0.79$). This finding is discussed in Section 5.4.

5.3.4 Interaction Analysis

The following analysis identifies specific behavioural practices and resources, including repair strategies, that participants were seen to employ during interaction. CA combined with gaze data forms the foundation of the analysis, and extracts describing participants' behaviour during both correct and incorrect responses to grab instructions are presented. Participants' use of movement and gaze as resources during the experimental task is then explored. Finally, post-hoc analysis of eye tracking data to explore the specific gaze practice of 'glancing' is presented. Glancing was used extensively by all participants as a repair strategy. This analysis is primarily concerned with interaction related to grab instructions,

```
CUBES - TRACKEDGAZE[02.b]          Participant: "Tim"

01   Rob: Tim
          assembled cubes

02   Rob: So     m- m:y    turn
          Rob    cubes     Rob

03   Tim: Okay
          Rob

                                   (Rob turns and looks at cube)
04   Rob: Can you pick up this cube where
          Rob

05        I'm looking at now
          Rob           designated cube

06   Tim: Is that this one (points to cube)
          cube          Rob

07   Rob: Yes.               (Tim grabs cube)
          Rob   cube
```

Figure 5.10: *CA transcript of tracked gaze AMC. Confederate, "Rob", issues a grab instruction to participant, "Tim", who responds by grabbing the correct cube.*

rather than position instructions. As suggested by the task performance analysis, responses to position instructions were not as revealing as those to grab instructions. This is further discussed in Section 5.4.

Responses to Grab Instructions

Participants responded to grab instructions either by correctly or incorrectly identifying the cube indicated by the confederate issuing the question. Figure 5.10 shows how an example case of correct and unproblematic cube identification and grabbing unfolds between a participant, "Tim", and the currently questioning participant, "Rob". At line *01*, Tim is gazing toward the puzzle base. Rob summons Tim's attention by calling his name and subsequently eliciting his gaze as indicated at the beginning of line *02*. Rob produces a pre-instruction, also at line *02*, which Tim acknowledges both vocally and with gaze at line *03*. At line *04*, Rob issues the instruction, during which Tim continues to gaze toward him. As he produces the deictic term "this", at line *04*, Rob turns and shifts his gaze to the designated cube. After a moment, and as Rob utters "now" at line *05*, Tim follows Rob's direction of gaze, shifting his own to focus on the correct cube. Tim points towards the cube at line *06*, and requests confirmation that his decision is correct. Tim performs this request verbally, with "Is that this one", followed by a return of gaze to Rob. At line *07*, following Rob's affirmative response, Tim focuses his gaze back on the cube, and grabs it.

In contrast, Figure 5.11 shows a case of problematic interaction during a grab instruction. In this extract, a participant, "Owen", firstly requests clarification as to which confederate is currently talking, and subsequently selects an incorrect cube. At line *01*, Rob issues an instruction as Owen gazes at various cubes in the VE and then at the bystanding confederate, "Paul". During the 3.2 seconds of

```
CUBES - MODELGAZE[01.b]         Participant: "Owen"

01   Rob:   Okay how about er you the pick cube I'm looking at now
            Gaze at various cubes                             Paul

02          (0.0)              (1.3)            (2.3)          (3.2)
            Paul               Rob              Paul           Rob

03   Rob:   Er: It's Rob here speaking
            Rob

04   Owen:  Yeah
     Paul:    Blue guy (points at Rob)
            Rob

05   Owen:  Pickup the one you're lookin' at: okay
            Rob                          cubes

06   Rob:   Yeah
            cubes

07          (           1.3)           (Owen grabs incorrect cube)
            Incorrect cube

08   Rob:   er No not that one
            Incorrect cube  Rob

09   Owen:  Oh Okay sorry                (Owen moves cube back)
            Rob     Incorrect cube

10   Owen:  Let me try again             (Owen moves close to Rob)
            Rob

11   Owen:  This one?                    (Owen grabs correct cube)
            Correct cube Rob

12   Rob:   Yeah
            Rob
```

Figure 5.11: *CA transcript of model gaze AMC highlighting problematic interaction and subsequent repair during a grab instruction. Confederates are "Rob" and "Paul". Participant is "Owen".*

silence that follows during line *02*, Owen alternates his gaze between the two confederates. Realising that Owen is unsure of which confederate just issued the question, Rob establishes that it was he, at line *03*. At line *04*, Owen vocally acknowledges this, as does Paul, who also points towards Rob. At line *05*, as Owen clarifies Rob's instruction, he shifts his gaze from Rob to a group of potential cubes. Rob confirms this action at line *06*, and after 1.3 seconds of deliberation, Owen grabs an incorrect cube at line *07*. Responding to this error, Rob states that the choice is incorrect, and Owen returns his gaze to Rob at line *08*. Owen apologises at line *09*, and returns the incorrect cube to its original position. At line *10*, Owen states that he will try again, and moves very close to Rob's avatar in order to examine its direction of gaze. At line *11*, Owen identifies and grabs another cube, and asks Rob for confirmation that his choice is correct. Rob confirms that Owen has grabbed the correct cube at line *12*, completing the grab instruction.

The above cases of successful and problematic interaction demonstrate how varying responses to instructions have the potential to significantly reduce task performance, in terms of both number of errors and time taken. The two cases also illustrate participants' strategic use of confederate avatars' gaze in order to inform their own action. This is particularly clear in the case of problematic interaction, in which, following an incorrect response to a grab instruction, the participant closely examines the instructing confederate's gaze in an attempt to repair the progress of interaction.

Movement as a Resource

When deciding which cube to grab, participants used confederates' proximity to cubes as a central cue. Correspondingly, confederates made extensive use of the ability to move freely within the VE, positioning themselves at an appropriate distance and angle (both vertically and horizontally) to the intended cube, thereby aiming to provide the participant with less ambiguous information. As the CA extract in Figure 5.11 illustrates at line *10*, participants often moved very close to the instructing confederate's avatar, in an attempt to more clearly determine their direction of gaze. This is one category of a number forms of locomotion that participants engaged in. In a small number of cases, participants were observed to take up an opposite position to the instructing confederate, so that potential cubes lay between the pair, in order to identify the target cube. However, a more common strategy adopted by participants was taking a position similar to that of the instructing avatar, by standing beside them, looking over their shoulder, or even standing "inside" them as allowed by the ICVE system. By adopting a common perspective, participants were able to share the confederate's view of the virtual scene. In this way, participants made use of the spatial characteristics of the ICVE system, employing movement strategically in an attempt to perform the task more effectively.

Gaze as a Resource

Gaze was employed as a general nonverbal resource during the multiparty interactions, and participants distributed their own gaze naturally between confederate avatars and cubes during periods of instruction, conversation, and silence. Participants were frequently observed to base their action on the direction they perceived the instructing confederate to be gazing. During both grab and position instructions, participants attempted to determine which cube or location was being indicated by focusing on, and following the instructing confederate's gaze. As exposed by task performance analysis, the gaze model condition in particular caused difficulties during interaction due to the unfaithful display of confederate gaze. The impact of this incorrect nonverbal information frequently surfaced during interaction with explicit participant vocalisation types including questions ("Are you looking at it now?"), complaints ("You're looking up! Not down.") or accounts ("It's a little difficult for me because your eyes keep moving about.").

These moments of problematic interaction, caused by confederate avatars displaying unfaithful gaze behaviour, highlight the operational importance of gaze as a communicative resource that is highly informative to subsequent action. Participants perceived the gaze of confederate avatars to reflect that of their embodied user, and thus, instinctively attempted to use the cue as a bidirectional resource, similarly to its role in collocated interaction.

Table 5.5: *Mean number (and standard deviation) of glances to and from the instructing avatar and target cubes per grab instruction.*

Condition	Static	Model	Tracked
# of Glances	2.42 (1.5)	3.58 (2.74)	2.25 (1.57)

Glances Per Grab Instruction

Following a grab instruction being issued be a confederate, a strategy adopted by all participants was to alternate gaze between the instructing confederate's eyes and the potential cubes. Over all twelve participants and experimental sessions, a total of 180 grab instructions were issued, and this strategy of glancing between confederates and cubes was employed in 176 (98.9%) of cases. By gathering information from confederates' gaze, participants were able to develop their judgement as to which cube was being indicated before taking action. The method was also employed extensively as a repair strategy following an incorrect grab. The number of glances per grab instruction was analysed with a repeated measures two-way ANOVA calculation, with the three gaze conditions and the five grab instructions as factors. A significant difference between conditions ($F(2,4) = 8.23$; $p < 0.0005$) and also between instructions ($F(2,4) = 4.54$; $p < 0.002$) was found. Similar to analysis of grab instruction errors and timings, post-hoc Tukey tests determined that the differences lay between the gaze model and the other two classes of static gaze ($F(1,4) = 8.81$; $p < 0.005$) and tracked gaze ($F(1,4) = 10.95$; $p < 0.001$). No interaction effect was found between condition and grab instruction.

Table 5.5 shows the mean number of glances performed per grab instruction under each gaze condition. These values are representative of a gaze control method's ability to clearly and efficiently communicate attentional information to an observer, with sufficient fidelity upon which to base critical action. Hence, when engaged in tracked and static gaze AMC, participants needed to performed fewer glances before responding to grab instructions than when engaged in modelled gaze AMC. This is likely to be an influential factor to the increased task performance (reducing the time taken and number of errors) observed in the tracked and static gaze conditions. Although the questionnaire did not expose significant differences of between gaze conditions, the user experience is also likely to be improved, with more efficient interactional flow, and reduced need to employ repair strategies. Although not statistically superior to static gaze, tracked gaze is identified as the highest-performing method of gaze control by this metric.

5.4 Discussion

Analysis aimed to measure quality of communication in terms of task performance, user experience, and interactional behaviour. In the categories of task performance and interaction analysis, tracked and static gaze AMC were seen to significantly outperform the gaze model condition. The poor performance of the model was explored in Section 5.3.2. This demonstrated how the simulated gaze behaviour was in contrast to the actual tracked gaze behaviour of participants, particularly in terms of eccentricity, saccade magnitude, and fixation duration. However, analysis of both task performance and interaction analysis failed to expose statistically significant differences between tracked and static gaze. Therefore, the initial hypothesis that tracked gaze avatars would enhance quality of communication during object-focused collaboration is not supported. Both intuitively, and when considering the enhanced throughput of nonverbal information delivered by tracked gaze, it is somewhat surprising that this was found to be the case. The following discussion explores influential factors toward this result, focusing particularly on the impact of gaze on the experimental interactions, and the technical properties of the ICVE setup. The lack of a significant result in subjective user experience is also discussed.

5.4.1 Gaze as a Determinant of Action

Participant population and experimental design are core determinants of data recorded during an applied experiment. When studying social interaction, mediated or otherwise, in which participants are able to act and communicate as desired, this impact is particularly prominent. The current experimental population consisted of inexperienced users of ICVE systems, who were positioned at the focal role of a three-party avatar-mediated puzzle scenario. Over the course of the three experimental interactions, participants were required to communicate both verbally and nonverbally with confederates, navigate the VE, make judgements, manipulate virtual objects, and conduct questionnaires. In this demanding situation, it is safe to state that many factors were influential to participants' action and perception during the AMC. For instance, as shown in Table 5.7, the order of the three experimental sessions was highly influential to overall puzzle completion time, which decreased with experience. The aim of this experiment was to assess the benefit of tracked gaze AMC in a multiparty task requiring physical movement and object manipulation. However, before specific benefits, if any, of tracked gaze can be explored, the fundamental question of whether avatar gaze itself was influential to the experimental interactions must be addressed.

Piloting and refinement of the experimental task design, detailed in Sections 5.2.2 and 5.2.5, aimed to establish a multiparty object-focused scenario which magnified the operational importance of avatar gaze during the collaborative interactions. The success of this aim, and of the influence of avatar gaze, may be confirmed by the significantly reduced quality of communication, measured by task performance and interaction analysis, observed when avatars' eye movement was controlled by the gaze model. Specifically, modelled gaze AMC resulted in: participants making more errors and taking more time when responding to confederates' instructions; higher degrees of ambiguity and confusion, as revealed by participants' verbal and nonverbal behaviour; and participants requiring to perform a higher number of glances between an instructing confederate's avatar and the potential cubes before responding to a grab instruction. Therefore, the poor performance of the modelled gaze AMC establishes eye movement as a practical nonverbal cue that is both trusted and acted upon when determining critical action during the object-focused task.

The three avatar gaze control methods may be positioned on a dimension of truthfulness to the actual gaze performed by embodied users. On this dimension, tracked gaze features a high degree of truthfulness, modelled gaze a low degree, and static gaze a neutral degree due to the absence of eye movement. Results of task performance and interaction analysis confirm that a low degree of truthfulness of avatar gaze, demonstrated by the gaze model, can be highly detrimental to participant action. However, data gathered between neutral and truthful gaze behaviour did not mirror this finding, and quality of communication supported by static gaze and tracked gaze was statistically identical. There are several explanations for this finding. The foremost reason is that avatars' head position and orientation, which was tracked and replicated regardless of gaze condition, provided participants with sufficient nonverbal information in order to perform the experimental task effectively. Human head orientation and gaze direction are closely coupled [ADRB91], and horizontal gaze shifts greater than 25° and vertical shifts greater than 10° typically produce combined eye and head movement to face a target [KB01]. Therefore, while tracked gaze does provide richer information of focus of attention, in the particular task, which featured relatively large ($0.3m^3$) target objects, this more accurate signalling was not practically required. In addition, the ability for users to move freely in the ICVE system allowed for strategic positioning and refinement of viewing perspective. Participants used this ability extensively as a tool to determine confederates' directional attention and signalling behaviour. This suggests that static gaze AMC, utilising just head and hand tracking, was able to support sufficiently rich NVC when engaged in this particular object-focused task. Finally, it should be noted that, while not statistically significant over static gaze, the superiority of tracked gaze is reflected in the results of task performance and interaction analysis: grab instruction time was less (grab instruction errors were equal); position instruction time and number of errors were less; and number of glances per grab instruction was less.

5.4.2 Technical Factors

During AMC in an ICVE system, the properties of the visual representation of fellow interactants' avatars is critical. These properties are defined by the technical characteristics of the ICVE system, both in terms of hardware and software. Regarding hardware, the brightness and resolution of the projection displays are central determinants, while for software, the fidelity of the geometric and articulated humanoid models, together with graphical rendering quality, are the major impact factors. The combination of these features form the core of the visual stimuli on which the mediated social interaction unfolds. Hence, a low-luminance projection system, or a particular contour to the shape of an avatar's head, are parameters informing how a user experiences the AMC. This experience is critical to the ability and manner in which users may perform specific action. Although no absolute measures were taken, the CAVE's projection system at UCL used by the experimental participants is relatively low-luminance compared to the system at UoR and UoS. As the central focus of this experiment was the investigation of varying methods of avatar gaze control, it should be noted that the visibility of the graphical representation of fellow avatars' eyes, which appeared life-size (a small number of physical pixels depending on rendering perspective) was not optimal. Thus, it is conceivable that the nature of the CAVE display system used by the participant during the experimental interactions was detrimental to observation and subsequent utilisation of the visually-subtle nonverbal channel of gaze.

Similarly to the prior conversation experiment, and as discussed in Section 4.4, the same basic avatar model was used in the current experiment to represent all three participants (albeit displaying differently coloured shirts). While the model does feature photographic texturing, its geometric fidelity is relatively basic in comparison to the detail of a human face. Also likely to be a significant factor to an observer's perception of the avatar was the lack of eyelid movement and blinking, causing the sclera of the eye to appear prominent and unnatural. These graphical characteristics should again be considered non-optimal and, recalling Garau's suggestion of there existing an interaction effect between visual and behavioural fidelity [Gar06], detrimental to the potential benefit of the faithful tracked gaze behaviour. Additionally, as exposed in the interaction analysis presented in Figure 5.11, the lack of avatar mouth movement led to difficulties regarding interaction and conversation management.

The overall implication of these findings promote AMC in ICVEs as a platform able to support high-quality object-focused interaction. This is reinforced by the subjective judgements of the experimental experience gathered by the questionnaire and interview data. Ratings both of ease of collaboration and of involvement in the task were rated highly, for all gaze conditions, on the 1..7 Likert scale. As a final indicator of the intuitive nature of telecommunication able to be supported by ICVEs, it should be noted that, despite the use of naïve participants and the complexity of the task, no critical failures of interaction occurred, and no outliers with regards to overall time taken were found.

5.5 Chapter Summary

The experiment reported in this chapter sought to investigate the impact of varying methods of avatar gaze control on quality of communication during object-focused multiparty AMC in ICVE systems. Through piloting, the experimental task, based on instructions issued by two confederates, and performed by a single participant, aimed to emphasise the operational importance of gaze during the collaborative interactions. The experiment compared three forms of AMC: tracked gaze, static gaze, and a simple gaze model. Data was collected from multiple sources, including eye tracking log files and videos, EyeCVE log files, gaze model log files, and subjective questionnaires. This data was analysed as metrics of task performance, subjective user experience, and interaction analysis. Analysis of task performance and interaction analysis revealed that tracked and static gaze AMC is able to support superior quality of communication than modelled gaze AMC. However, measures of subjective user experience did not show a significant difference between the gaze conditions. While the original hypothesis that tracked gaze AMC would support a higher quality of communication than both static and modelled gaze AMC is not fully supported, participants' action was significantly hindered in the modelled gaze condition, demonstrating the influence of misleading eye movement during object-focused AMC. Discussion focused on both experimental and technical characteristics of the studied AMC. Several factors, including display properties, avatar representation, naïvity of participants, and task design, were suggested as likely detractors from participants' ability and need to utilise, in practical terms, the richer nonverbal information supported by tracked gaze AMC.

The final telecommunication experiment presented in this thesis builds on lessons learned from Chapter 4's conversation experiment, and particularly, from the object-focused study documented in this chapter. Specifically, EyeCVE's avatar subsystem is improved, increasing the visual and behavioural fidelity of user embodiments, and a bespoke semi-immersive VR system, featuring a high luminance and high resolution projection display, is utilised to both enhance visual clarity and to ensure robust tracking of additional oculesic cues of blinks and pupil dilation.

Chapter 6

Experiment: Truth and Deception

This chapter presents two closely related experiments investigating truthful and deceptive interaction in AMC. Similarly to the experiments presented in the previous two chapters, eye tracking is used both interactively and analytically. However, while the previous experiments only investigated gaze, the experiments in this chapter also investigate blinking and pupil size, which are captured, represented in avatars, and analysed. Thus, interactively, avatars reproduced their embodied user's fuller oculesic behaviour of gaze, blinking, and pupil size during AMC, and analytically, all three cues are examined. To the author's knowledge, this is the first combined analysis of gaze, blinking and pupil dilation during social interaction in any field. The experiments are positioned in the social domain of interpersonal trust and deception, which presents a compelling array of issues by which to investigate interaction in visual telecommunication systems. Section 2.1.3 discussed how humans exhibit verbal and, critically, nonverbal behavioural cues correlated with lying and truth telling, and how oculesic cues of gaze, blinking, and pupil size are influenced accordingly. Detection of deception was also covered, demonstrating that people estimate the veracity of a communicative message by observing others, so media transmitting fewer nonverbal channels becomes preferable for the deceiver [CGB+04].

The first experiment (*E1*) in this chapter explores truthful and deceptive discourse between dyads in state-of-art AMC and VMC systems. The experimental task design was inspired by Walczyk et al.'s question-answer framework for manipulating cognitive load when lying and truth telling [WSC+05]. During E1's experimental interactions, a confederate issued questions to participants, who respond either truthfully or deceptively. Eye tracking was used to capture participants' oculesic behaviour of gaze, blinks, and pupil size during both AMC and VMC for post-experimental analysis. Following the interactions, a questionnaire collected data describing participants' psychological arousal and mood state.

The second experiment (*E2*) in this chapter investigates the impact of bestowing avatars with the oculesic cues of gaze, blinks, and pupil dilation, which are reproduced in real-time using eye trackers worn by embodied users. E2 follows on from E1, and invited a different set of participants to view audiovisual replays of E1's experimental interactions. E2's participants assessed E1's participants' degrees of veracity and engagement, together with confidence levels relating to the two judgements. These ratings were performed over three stimuli conditions: avatars exhibiting oculesics, avatars featuring no oculesics, and audio-only replays.

6.1 Experimental Aims and Expectations

The overall aim of the two experiments presented in this chapter was to investigate both behaviour (E1) and observation (E2) of truthful and deceptive discourse in AMC. The primary question E1 sought to answer is whether users' behaviour and response during social interaction in AMC and VMC is similar. More specifically, when engaged in audio-visual telecommunication, if the visual component of the communication is depicted by the virtual, graphical, stimuli of AMC, will users exhibit nonverbal behaviour and psychological response that is similar to face-to-face interaction simulated by VMC? This aim appears similar to that of the experiment presented in Chapter 4, which investigated behaviour in a multiparty conversational scenario in tracked gaze AMC and gaze aware VMC. However, that experiment was exploratory in nature, and rigorous analysis of oculesics was hindered by the ability of the MobileEye eye tracker. In contrast, E1 focuses on a specific social scenario (truth and deception) that is critical to practical usage of telecommunication systems, and considers a full range of oculesic behaviour coupled with rich interaction analysis as presented in Section 3.2. The comparison of AMC to the dominant form of visual telecommunication, VMC, aimed to provide a benchmark by which to measure users' behaviour and response during interaction mediated by virtual embodiments.

E2 aimed to investigate avatar fidelity, and how the increased nonverbal richness afforded by oculesic reproduction of user behaviour may influence how judgements of AMC may be formed. Specifically, E2 sought to investigate whether observers of AMC are able to detect truth and deception more accurately when an avatar's oculesic behaviour replicates the embodied user's.

The experimental hypotheses (noted *Hn*) were as follows:

- *E1H1:* During AMC and VMC, participants will exhibit similar patterns of the oculesic cues of gaze, blink rate, and pupil dilation. However, psychological arousal, measured by questionnaire data, will be greater following VMC.

- *E1H2:* When communicating **deceptive** messages in both mediums: participants' proportion of gaze directed at their partner will contrast to that measured when truth telling; participants' blink rate will decrease, followed by compensatory blinks after speech has ended; participants' pupils will dilate to a larger size than when truth telling.

- *E2H1:* When assessing the veracity of E1's participants, judgements will be more accurate and more confident when observing avatars featuring reproduction of oculesic behaviour than judgements of avatars displaying no oculesic expression, or audio-only stimuli. Similarly, higher ratings of engagement, and confidence in this rating, will be elicited when observing avatars featuring oculesic behaviour.

6.2 Experiment 1: Interactions

E1 compared the behaviour of participants when engaged in truthful and deceptive discourse in dyadic conversational interaction performed in AMC and VMC. The experiment employed a between-groups design with regards to mediation type, but a within-groups repeated measures design with regards to veracity of response. Hence, each single participant performed either AMC or VMC, but both told the truth and lied over the course of the experimental interaction.

Located at UCL, the experimental interactions took place between a participant using the WALL system outlined in Section 6.2.3, and a confederate using the CAVE. The experimental task design was inspired by Walczyk et al.'s question-answer framework for manipulating cognitive load [WSC+05]. Walczyk et al.'s study aimed to correlate the time it took an individual to respond to questions with the veracity of the individual. The experimental interaction consisted of five stages of questions that were issued by a confederate and answered by a participant. Each stage consisted of both open and closed (yes/no) questions regarding a single topic. Topics included general personal information, recent personal events, and remote personal events. For a given stage, participants were instructed either to respond with lies to all questions or to answer truthfully to all questions.

Borrowing Walczyk et al.'s framework, E1 presents a total of six stages each containing between 9–12 unique questions. The topics of the stages included two general personal information, two recent personal events, and two remote personal events. Thus, participants lied to three stages (one general, one recent, one remote), and responded truthfully to the remaining three. Unlike Walczyk et al., the current experiment did not aim to investigate behavioural differences between stage types relating to memory recall and cognitive load. Rather, the design is adopted as an effective framework for eliciting thoughtful responses to questions that may feasibly be answered truthfully and deceptively. For additional materials, see Appendix E.

6.2.1 Independent Variables

The independent variables were mediation type and veracity. Participants performed the experimental interactions either in VMC or in AMC. All participants answered the same six stages of questions, three of which they were instructed to respond to truthfully, and three deceptively. Stage order and veracity condition were resequenced over participants, so all stages in both mediation types were responded to both with lies and truths.

Mediation Type:

1. *AMC:* Participant and confederate were represented by avatars. Each user's natural head and hand movement, together with oculesic behaviours of gaze, blinks, and pupil size, were tracked and displayed by their avatar in real-time. Audio communication allowed the dyads to converse.

2. *VMC:* Two-way gaze aware video-conferencing enabled interactants to communicate both visually and aurally. The large-format and HD displays presented views of the top-half of interactants' bodies, allowing signalling and observation of gaze, head movement, and arm gesture.

Veracity:

1. *Truth:* Participants responded truthfully to all questions issued by the confederate in a particular stage.

2. *Deception:* Participants responded with lies to all questions issued by the confederate in a particular stage.

6.2.2 Piloting

Section 5.2.2 emphasised the importance of iterative prototyping before finalising an experimental design. Similarly, the current experiment evolved over several stages. Initially, the design considered AMC only, and was inspired by Leal and Vrij's [LV08] work investigating blinking during and after lying, and Lubow and Fein's [LF96] work investigating pupillary size response to the "visual guilty knowledge" test. The original experimental protocol featured two conversational stages and one intermediary object-focused action, in which the participant performed tasks independently and out of the confederate's sight. The first conversational stage involved the confederate providing training and information about the VR system. The critical object-focused period followed, during which participants either obeyed or disobeyed the confederate's prior orders based on instructions provided by the experimenter. Following this action, the second conversational stage began, which was also the critical period of data collection. During this stage, the participant was asked to recall the training information provided in the first conversational stage, and also their actions during the independent object-focused task, which was performed either truthfully or deceptively depending on their condition of obedience. Additionally, the visual guiltily knowledge test was performed, with the confederate showing various objects to the participant: some of which the participant would not have seen had they obeyed instructions during the object-focused stage. During this period of data collection, participants gaze, blinks, and pupil dilation was monitored for later analysis.

This experimental design, while incorporating physical movement in object-focused and conversational interaction, proved problematic in terms of data collection. Regarding pupil size, the low ambient luminance of the CAVE caused participants' pupils to remain in a dilated state, leading to limited variation of pupil size response during the interactions. Also, blink detection was not robust, as the CAVE's shutter glass frame forced the eye tracker's viewing angle to be off-centre, thus leading to a permanently low pupil aspect ratio. The elements of free movement and object-focused action also precluded the possibility of comparison between AMC and VMC. Following the more exploratory nature of Chapter 4's conversational experiment, further investigation into comparing AMC and VMC was deemed necessary for a full investigation into the issues raised when performing interaction mediated by virtual embodiments versus live video representations. Finally, the original design appeared to be too complex to investigate differences in truth telling and deception in AMC, and many participants expressed confusion during the experiment regarding their correct action. Therefore, E1's final focused interpersonal experimental task was designed, piloted, and refined.

6.2.3 Apparatus

The experimental AMC and VMC was performed between two immersive display systems at UCL: the CAVE and the WALL. The two labs were located physically in adjacent rooms, allowing for collocated communication between the main experimenter and the confederate. For reasons relating to blink and pupil size tracking outlined in Section 6.2.2, participants used the WALL, and the confederate used the CAVE. ViewPoint eye trackers were used in both AMC and VMC to capture gaze, blinks, and pupil size data.

Figure 6.1: *EyeCVE supporting AMC between users of the WALL (left) and CAVE (right) immersive projection systems.*

EyeCVE Setup

EyeCVE supported the AMC. Chapters 4 and 5 demonstrated the use of eye tracking to drive avatar gaze. The current study introduces blinking and pupil size as two additional tracked oculesic cues that are represented by the advanced avatars. Implementation details and screen captures are presented in Sections 3.1.3 and 3.1.4. In the CAVE, EyeCVE operated as described in Section 5.2.3, albeit a more mature version. The centralised EyeCVE server, maintaining the state of the virtual world and distributing changes to both clients, ran on the same machine driving the EyeCVE cluster displays. In the WALL, EyeCVE's VE was projected on a single full HD (1920×1080 pixels) display wall with perspective-correct monoscopic rendering. Head and hand tracking was achieved using a Polhemus Fastrak [Pol10], with sensors attached to a cap worn by participants as illustrated in Figure 6.5. Audio communication between users of the CAVE and the WALL was established externally to EyeCVE using Google Talk [Goo10b]. The confederate wore a wireless microphone, while a desk-mounted microphone detected participant vocalisation, aiming to reduce the number of devices required to be worn by the participant. Audio was output through speakers located in each lab. Figure 6.1 shows users engaged in AMC between both display systems.

The VE consisted of a meeting room populated with a rectangular table and two chairs. The absence of additional objects intended to eliminate the influence of distractions on participants' gaze behaviour. The VE is presented in Figure 6.2. Post-experimental analysis of gaze intended to investigate the proportion of time participants' gaze was directed *at* the confederate, and the time spent looking *away* from the confederate. To aid this classification, a virtual grid, that was not rendered by EyeCVE, was positioned behind the confederate avatar as shown in Figure 6.3. Section 3.2.3 presented a sample interaction analysis performed on data captured in the current experiment. The grid aided classification of at/away gaze, together with direction of away-gaze according to logical reference by column and row.

Video Conferencing System

VMC was hosted between the CAVE's front-wall and WALL displays. Bidirectional gaze awareness was achieved by aligning cameras and displays of remote users within Chen's threshold [Che02]. Video

Figure 6.2: *The VE in which the experimental AMC took place. The VE depicted a simple meeting room, with two chairs separated by a table. Before interaction, the virtual furniture was aligned to their real counterparts as illustrated in Figure 6.1.*

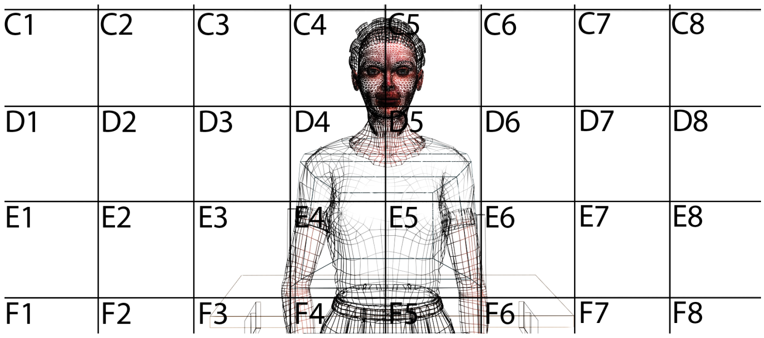


Figure 6.3: *Grid (not rendered by EyeCVE) positioned behind confederate avatar to aid classification of participants'* at *and* away *gaze.*

was streamed at full HD resolution in the WALL, and 1280×720 pixels in the CAVE using direct HDMI links between the camera located in one system and the projector display in the other. Users appeared life-size on both displays, and a plain black background behind the confederate intended to eliminate the influence of distractions on participants' gaze behaviour. Audio communication was supported by Google Talk [Goo10b]. Figure 6.4 shows VMC as it appeared for the participant using the WALL.

Eye Tracking and Data Collection

Similar to Chapter 5's object-focused experiment, eye tracking was achieved using the head-mounted ViewPoint EyeTracker from Arrington Research. Eye trackers were mounted on two frames as shown in Figure 6.5 for use in either the WALL or CAVE. An additional reason for participants using the WALL was so that no stereo glasses, that may affect pupil size due to dark lenses, had to be worn. Hence, the eye tracker was mounted on an empty frame. In the CAVE, an eye tracker was mounted identically to as described in Section 5.2.3. Both frames featured a camera which recorded the scene from the wearer's perspective at a 150° FOV. The wearer's foveal fixation point was overlaid on the scene video, allowing for post-session analysis of action once synchronized with the separate audio stream. As documented in Section 3.2, the eye tracker's logging facility also served as the primary method for experimental data collection, with the tracked behaviour of users being streamed to the process, and output to a log file at 60 Hz. Data included gaze (2D X/Y coordinate, 3D hit-point in the VE, and fixation and saccade timings), blink signals, pupil size, speech signals, head and hand tracking, and markup data input by the experimenter.

Figure 6.4: *Users engaged in VMC.*

Figure 6.5: *Eye tracker mounted on the WALL's lens-free frame (left), and on the CAVE's CrystalEyes 3 [Rea10] shutter glasses (right).*

6.2.4 Population

A total of 24 participants with normal or corrected vision and no previous experience of telepresence systems were recruited from UCL's campus through an advertising poster campaign. Participants were paid £10 to perform the study, which lasted for around one hour. Ages ranged from 19–62, with European, Asian and African origins were represented. While the effects of gender were not measured, both male and female participants took part, with 12 of each sex. Half of the participants performed the experimental interactions in AMC, and the remaining half in VMC. The two groups were balanced according to personality characteristics using the *NEO Five-Factor Inventory* questionnaire [CJM08], which was performed during the technical setup prior to the experimental interactions.

The experimental interactions consisted of six stages of questions, to which participants were instructed to respond to with lies or with truths. Stages of questions (but not questions within a stage) and instructed response veracity were resequenced over participants to negate inter-experimental effects. The same male confederate questioned all participants in all AMC and VMC sessions, and was blind to participants' current state of veracity. This aimed to minimise differences in confederate behaviour due to knowledge of how participants may be answering. Participants were aware of this, which in-

Table 6.1: *Question stage and veracity condition presented to participants for each mediation type. Stages were divided into* General 1 *(G1),* General 2 *(G2),* Recent 1 *(Rc1),* Recent 2 *(Rc2),* Remote 1 *(Rm1), and* Remote 2 *(Rm2). Note that stages are resequenced by type (General, Recent, Remote), but not within stage class (i.e. G1 always precedes G2). Participants were instructed to tell the truth (T) or lie (L) during specific stages as shown.*

Participant	Question Stage and Veracity											
	1st	*T/L*	2nd	*T/L*	3rd	*T/L*	4th	*T/L*	5th	*T/L*	6st	*T/L*
1	G1	*T*	G2	*L*	Rc1	*T*	Rc2	*L*	Rm1	*T*	Rm2	L
2	G1	*L*	G2	*T*	Rc1	*L*	Rc2	*T*	Rm1	*L*	Rm2	T
3	G1	*T*	G2	*L*	Rm1	*T*	Rm2	*L*	Rc1	*T*	Rc2	L
4	G1	*L*	G2	*T*	Rm1	*L*	Rm2	*T*	Rc1	*L*	Rc2	T
5	Rc1	*T*	Rc2	*L*	G1	*T*	G2	*L*	Rm1	*T*	Rm2	L
6	Rc1	*L*	Rc2	*T*	G1	*L*	G2	*T*	Rm1	*L*	Rm2	L
7	Rc1	*T*	Rc2	*L*	Rm1	*T*	Rm2	*L*	G1	*T*	G2	L
8	Rc1	*L*	Rc2	*T*	Rm1	*L*	Rm2	*T*	G1	*L*	G2	T
9	Rm1	*T*	Rm2	*L*	G1	*T*	G2	*L*	Rc1	*T*	Rc2	L
10	Rm1	*L*	Rm2	*T*	G1	*L*	G2	*T*	Rc1	*L*	Rc2	T
11	Rm1	*T*	Rm2	*L*	Rc1	*T*	Rc2	*L*	G1	*T*	G2	L
12	Rm1	*L*	Rm2	*T*	Rc1	*L*	Rc2	*T*	G1	*L*	G2	T

tended to motivate them to lie plausibly and also to negate any inter-experimental effects in this regard. Participants were not informed of the significance of their oculesic behaviour. Finally, as described in Section 6.2.5, setup procedures preceding the AMC and VMC were almost identical, aiming to provide a consistent experience to all participants prior to the experimental interaction. The condition sequences for all participants are presented in Table 6.1.

6.2.5 Procedure

The experimenter at UCL was responsible for organising and running the experiments as well as local technical setup. The confederate using the CAVE was familiar with the technical architecture of the system, and was able to calibrate his body tracking devices independently. The experimental procedure consisted of seven major phases: briefing upon participant arrival; *NEO Five-Factor Inventory* questionnaire; technical setup; eye tracker calibration; experimental interaction; *Profile of Mood States* (POMS) questionnaire [MLD92]; and finally a post-experimental interview.

Phase 1: Briefing

Upon arrival at the lab, participants were welcomed in a reception area, and were given an information sheet and ethical consent form. The information sheet set the scene for the forthcoming experiment involving a question-answer scenario, performed in AMC or VMC, with defined periods of truthful and deceptive response. Participants were left in private to fill out the forms, and were informed that they may ask any questions not directly relating to the experimental procedure or technical setup.

Phase 2: NEO Five-Factor Inventory

Following initial briefing, the participant was issued with the NEO Five-Factor Inventory questionnaire [CJM08]. The questionnaire is designed to measure an individual's psychological personality using a 60-item measure of the five factor model [Dig90]. The five-factor model rates extroversion, agreeableness, conscientiousness, neuroticism, and openness to experience. The questionnaire took around ten minutes for the participant to complete, and aimed to provide sufficient time for the experimenter to perform the necessary technical setup, and also in order to balance AMC and VMC groups according to personality score. The questionnaire responses were input using Google Documents [Goo10a], and automatically uploaded and processed, and an overall score determined. Participants were then assigned either to AMC or VMC, while keeping the cumulative scores for each mediation type similar between AMC and VMC as the 24 sessions progressed.

Phase 3: Technical Setup

While the participant completed the NEO Five-Factor Inventory, EyeCVE was set up between the CAVE and the WALL. Setup of AMC involved running an EyeCVE client for display in the WALL, and one for display in the CAVE, networked via an EyeCVE server. The clients loaded the shared VE shown in Figure 6.2, together with the remote user's avatar embodiment (male model for confederate, and male or female model for participant). The technical setup procedures preceding both AMC and VMC sessions were almost identical despite differences in the mediation type. The key difference was that, following setup of EyeCVE in the VMC condition, the WALL's projector input was switched to the video stream originating from the camera located in the CAVE and facing the confederate. Hence, EyeCVE operated in the 'background' during the VMC sessions for the following reasons: collection of body tracking data was managed by EyeCVE, and required filtering and processing before being sent for log file output (see Section 3.2.2); and to enable the interactions, visualized as AMC, to be captured in order to act as stimuli in E2. Audio communication was hosted by Google Talk [Goo10b].

Phase 4: Calibration

The participant was seated, and fitted with the head, hand, and eye tracking devices. Gaze, pupil aspect ratio for blinks, and pupil size range were calibrated. A further calibration step mapping 2D gaze to the correct avatar eye rotation in the VE was performed in EyeCVE. The experimenter checked the accuracy of the calibration and made adjustments if required. The confederate calibrated himself in the CAVE similarly. Microphones were positioned to ensure clear verbal communication for the forthcoming experimental interaction. Following setup, the experimenter left the WALL lab to observe the interactions and provide markup data for the logging process.

Phase 5: Experimental Interaction

The experimental interactions began with the confederate and participant greeting each other's avatar or video representation. The confederate ensured that the participant was clear with regards to their instructions, and proceeded to issue the first of the six stages of questions. Over the stages, the participant answered with lies or truths as per initial instructions, which were also provided on a sheet of paper in the WALL lab. Following the end of the participant's response to each question, the confederate paused for approximately two seconds before issuing the next question. Following the sixth stage of questions, the experimental interaction concluded, and the confederate bid farewell to the participant.

Phase 6: POMS Questionnaire

Immediately following the interaction, participants completed the POMS questionnaire. The questionnaire measures an individual's current mood state over six psychological traits: tension, anger, depression, fatigue, confusion, and vigour [MLD92].

Phase 7: Interview

Finally, the experimenter conducted an informal interview with the participant. Discussion revolved around the participant's general experience of the AMC or VMC, how they felt when lying and when telling the truth, and whether they followed instructions of veracity exactly.

6.2.6 Analysis

The majority of this analysis is based on eye tracking log files recorded from participants. Data collected during AMC and VMC were treated separately, and were divided according to stage. Stages were then grouped according to veracity of participant response, resulting in four data classes: AMC/truth, AMC/lies, VMC/truth, VMC/lies. Repeated measures two-way ANOVA was generally employed when determining statistical differences between the classes, with post-hoc Tukey tests performed where appropriate.

Gaze

Participants' gaze was categorized as *at* (looking towards the confederate's avatar or video representation) and *away* (looking elsewhere in the visual field). The grid positioned behind the confederate as illustrated in Figure 6.3 aided in this categorization. Table 6.2 and Figure 6.6 show the mean percentage and standard deviation of time that participants gazed at the confederate in each of the four data classes, together with combined truth and lie conditions for both AMC and VMC. The proportion of gaze directed at the confederate is greater during AMC than VMC in both conditions of veracity. Gaze at the confederate is also greater when telling lies than when telling truths in both AMC and VMC. The consistent trends in gaze behaviour between mediation types and veracity supports the validity of the experimental method and data collection. However, differences were not great enough to expose a main effect of gaze behaviour between AMC and VMC, or between truth telling and lying, when examined on this holistic scale. This result is to be expected, as analysis on this scale considered gaze behaviour over entire stages, which had a mean duration of 1.5 minutes.

Gaze behaviour during critical moments of question response was then explicitly targeted by analysing gaze during lying and truth telling when answering two individual open questions that were selected at random: "Tell me more about the last book you read" appearing in the *General 1* stage, and "What did you do last Saturday evening?" from the *Recent 2* stage. The period of time, starting when the

Table 6.2: *Mean percentage (and standard deviation) of gaze directed* at *the confederate in AMC and VMC during truth and lie stages, and combined.*

Mediation	*Truth*	*Deception*	*Combined*
AMC	80.8% (32.2)	85.6% (22.5)	83.6% (27.7)
VMC	75.9% (23.3)	83.7% (18.0)	79.7% (20.5)

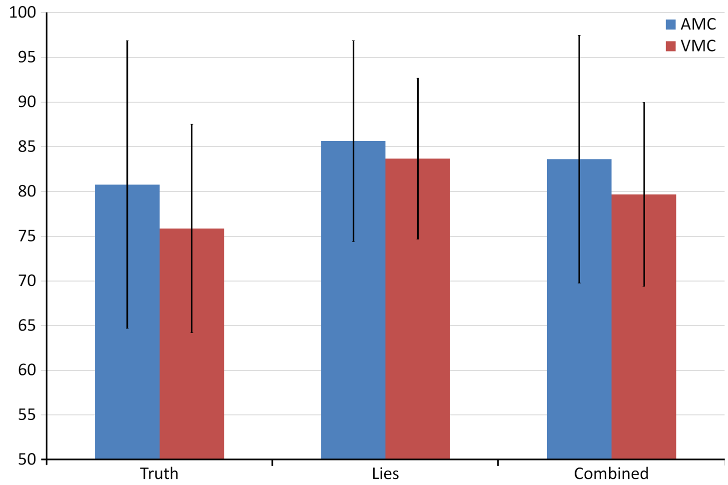


Figure 6.6: *Mean percentage and standard deviation of gaze directed* at *the confederate in AMC and VMC during truth and lie stages, and combined.*

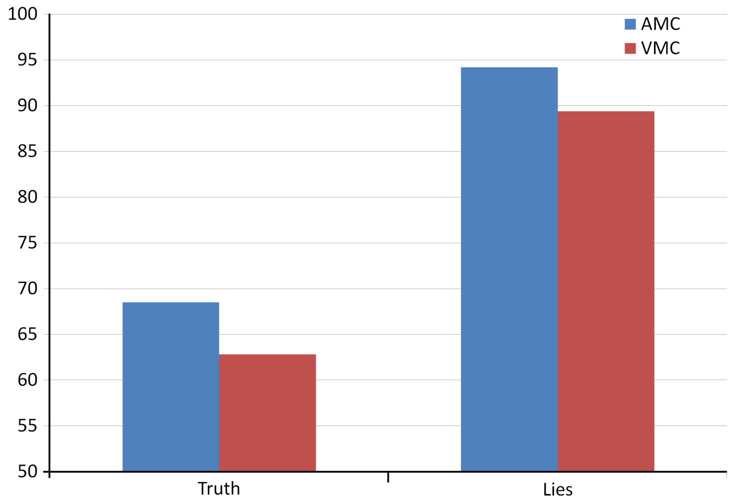


Figure 6.7: *Mean percentage of gaze directed* at *the confederate when answering two randomly-selected questions in AMC and VMC when telling truths and lies. Questions were "Tell me more about the last book you read" (from General 1 stage), and "What did you do last Saturday evening?" (from Recent 2 stage).*

confederate reached the semantic point of the question (i.e. following utterance of "book" and "evening" respectively), and ending two seconds following the end of participants' response speech was analysed. Considering both questions, mean results from all AMC participants revealed that, when telling the truth, gaze was directed at the confederate 68.5% of the time, but when lying, this figure rose to 94.2%. This difference is highly significant (p<.01). VMC participants demonstrated similar behaviour when answering the two questions, directing gaze at the confederate 62.8% of the time when responding truthfully, and 89.4% of the time when lying. Again, a main effect was found between conditions of veracity (p<.01). Finally, no significant difference in gaze behaviour was found between classes of AMC/truth and VMC/truth, or between classes of AMC/lies and VMC/lies. Figure 6.7 shows mean percentage of gaze directed at the confederate when answering the two questions truthfully and deceptively in AMC and VMC.

Table 6.3: *Mean number (and standard deviation) of blinks performed per question in AMC and VMC when answering truthfully, deceptively, and combined.*

Mediation	*Truth*	*Deception*	*Combined*
AMC	5.80 (4.36)	5.72 (4.30)	5.76 (4.33)
VMC	4.78 (4.42)	5.23 (4.58)	4.78 (4.25)

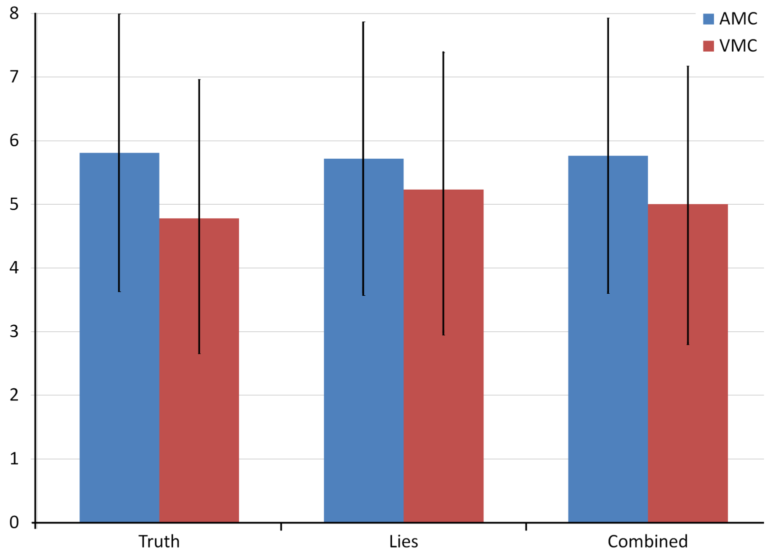


Figure 6.8: *Mean number and standard deviation of blinks performed per question in AMC and VMC when answering truthfully, deceptively, and combined.*

Blinking

Table 6.3 and Figure 6.8 show the mean number and standard deviation of blinks that participants performed per question in each of the four data classes, together with combined truth and lie conditions for both AMC and VMC. The values indicate that participants engaged in AMC blinked more regularly than those in VMC, but no significant differences between mediation type or veracity condition were found. Post-hoc tests focusing on specific questions also failed to expose reliable trends. The high standard deviation in blink rate demonstrates the influence of the unpredictable and idiosyncratic human element of the interactions, resulting in unfruitful attempts to generalise blink patterns over multiple participants.

Pupil Dilation

Table 6.4 and Figure 6.9 show participants' mean and standard deviation of pupil size in each of the four data classes, together with combined truth and lie conditions for both AMC and VMC. This measurement is normalized to 1: 0 representing participants natural state of constriction given the ambient luminance level when using the WALL; and 1 representing extreme dilation adjusted to darkness during calibration. Hence, when emotionally aroused or under cognitive load, pupil size is expected to be greater than the relaxed state of 0, but is not expected to approach extreme dilation of 1. Table 6.4 illustrates that mean pupil size is larger when lying than when telling the truth during both AMC and VMC, and that participants' pupils are more dilated during VMC than AMC. However, ANOVA calculations performed at this macro scale of stages failed to expose a main effect in pupil size between mediation type or veracity.

Table 6.4: *Mean pupil size (and standard deviation) in AMC and VMC during truth and lie stages, and combined. Values are normalized to 1: 0 represents natural pupil size given the ambient luminance level, and 1 represents extreme dilation in darkness.*

Mediation	*Truth*	*Deception*	*Combined*
AMC	0.17 (0.09)	0.21 (0.11)	0.19 (0.10)
VMC	0.21 (0.13)	0.26 (0.14)	0.23 (0.13)

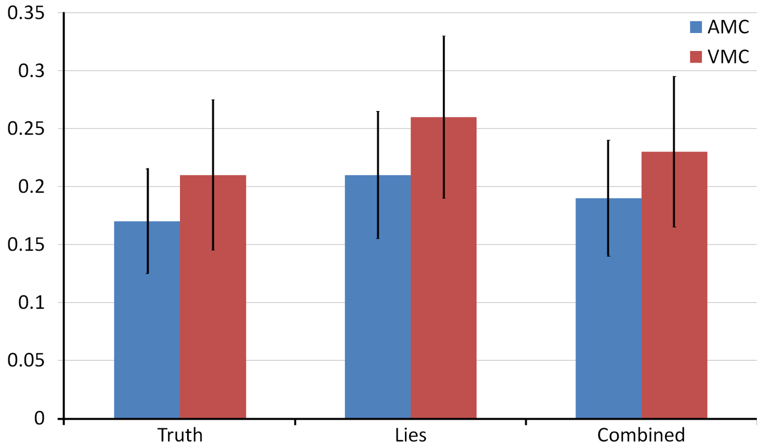


Figure 6.9: *Mean pupil size and standard deviation in AMC and VMC during truth and lie stages, and combined.*

Similarly to gaze analysis, participant pupil size during the critical moments of interaction were analysed by measuring the *change* in pupil size during response to two individual open questions that were selected at random: “What’s your first name?” appearing in the *General 2* stage, and “What’s the name of the school you attended?” from the *Remote 1* stage. This analysis was performed by computing the difference between each participant’s largest and smallest pupil size during the period starting two seconds prior to each question’s semantic point (i.e. “name” and “school” respectively) and ending two seconds after the end of participants’ vocal response. Considering both questions, the mean change in the pupil size of participants engaged in AMC when telling the truth was 0.07, but when lying, this change increased dramatically to 0.32, indicating increased dilation, and exposing a highly significant difference ($p<.01$). VMC participants’ mean pupil size change demonstrated similar behaviour, with a change of 0.11 for truth tellers and a change of 0.38 for liars, again exposing a main effect between response veracity ($p<.01$). Finally, no significant difference in pupil size change was found between classes of AMC/truth and VMC/truth, or between classes of AMC/lies and VMC/lies.

Profile of Mood States

The experimental interactions were designed to be moderately stressful for participants, who were required to lie or tell the truth over a series of semi-personal questions issued by a stranger (the confederate) in AMC or VMC. The POMS questionnaire, completed immediately following the interactions aimed to capture participants’ current psychological mood state. Table 6.5 and Figure 6.11 show the mean and standard deviation of the self-reported scores elicited by the questionnaire in AMC and VMC. Results indicate higher levels of arousal following VMC, with all six mood-factors of tension, anger, depression,

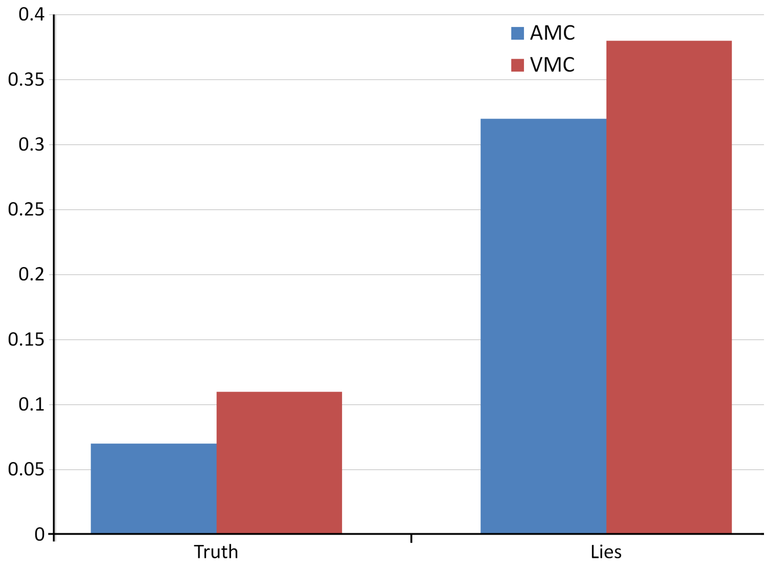


Figure 6.10: *Mean change in pupil size when answering two randomly-selected questions in AMC and VMC when telling truths and lies. Questions were "What's your first name?" (from General 2 stage), and "'What's the name of the school you attended?" (from Remote 1 stage).*

Table 6.5: *Mean (and standard deviation) mood-factor scores elicited by the POMS questionnaire following AMC and VMC.*

Mood State	*AMC*	*VMC*
Tense/Anxious	1.56 (0.81)	2.22 (1.22)
Angry/Resentful	1.22 (0.53)	1.50 (1.00)
Depressed	1.26 (0.65)	1.69 (1.14)
Fatigued/Tired	1.58 (0.74)	2.22 (1.17)
Confused	1.63 (0.90)	2.00 (1.04)
Vigorous	2.96 (0.97)	3.44 (1.18)

fatigue, confusion, and vigour (a positive state) eliciting higher scores than AMC. Repeated measures two-way ANOVA, taking mediation condition and the six mood-factors as factors uncovered a main effect (p<.01) between AMC and VMC, and post-hoc Tukey tests revealed that significant differences existed between all mood-factors (p<.05) excluding anger.

6.2.7 Discussion

Analysis approached the experimental interactions from two perspectives: comparison of AMC and VMC measured by eye tracking and psychological affect, and variance in behaviour between truthful and deceptive discourse between the two mediation types of AMC and VMC measured by eye tracking. Both experimental hypotheses, E1H1 and E1H2, are supported by the results. Regarding E1H1, participants engaged in AMC exhibited comparable oculesic behaviour to those engaged in VMC. As discussed below, both gaze behaviour and pupil size response tentatively suggest that participants engaged in VMC experienced a higher degree of psychological arousal than those who performed AMC. This conjecture is supported by the striking difference between affective states uncovered by the POMS questionnaire following the experimental interactions. Regarding E1H2, gaze behaviour was found to differ between states of veracity, with deception eliciting greater proportions of gaze directed at the confederate in

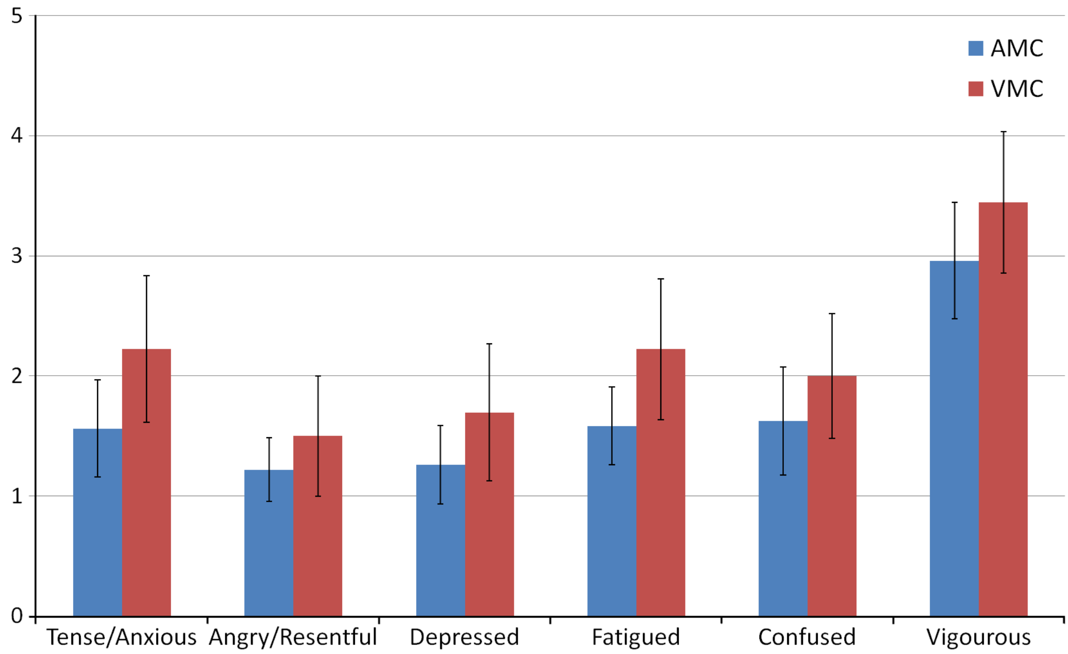


Figure 6.11: *Mean and standard deviation mood-factor scores elicited by the POMS questionnaire following AMC and VMC.*

both AMC and VMC. Participant pupil dilation was also greater during deception in both AMC and VMC. The following section discusses the results and experimental observations, focusing similarly on mediation type and state of veracity.

AMC and VMC

Comparing AMC and VMC, participants' oculesic cues of gaze, blinking, and pupil size were found to be similar in both forms of visual telecommunication. Analysis categorised participants' gaze distribution as that directed at the confederate, and away from the confederate. When considering the macro scale of question stages, AMC participants gazed at the confederate an average of 83.6% of the time. In VMC, this proportion was found to be slightly, but insignificantly, less at 79.7%. This relationship held true also when considering the micro scale of individual questions. Due to the structured nature of the experimental interactions, these values cannot be directly compared to classic studies of dyadic conversation (for instance, Argyle and Ingham suggested that people direct gaze at their conversational partners for 61% of the time [AI72]). However, it is informative to discuss factors influential to the current experiment's inflated proportion of gaze directed at the conversational partner with reference to such studies. Firstly, the experimental interactions were highly engaging for participants, who were required to respond either truthfully or deceptively over several stages of questions. This required a high level of concentration on the instructing confederate's verbal and nonverbal information, which was likely to elicit a high proportion of directed gaze. In addition, the immersive nature of the visual telecommunication systems investigated in this experiment, and indeed over the three telecommunication experiments presented in this thesis, implies that participants' FOV was consumed by the visual component of the mediation. As described in Sections 6.2.3 and 6.2.3, both AMC and VMC setups displayed the confederate avatar or video against a featureless background, thus eliminating visual distractions. This presented the confederate as the single salient object within the visual scene, and may have contributed to the increased proportion of gaze directed at the confederate compared to Argyle and Ingham [AI72].

While differences in gaze behaviour between AMC and VMC were not statistically significant, it is worthy to discuss the observation that AMC participants gazed at the virtual embodiment of the confederate more than VMC participants gazed at the live video of the confederate, during both macro and micro scales of analysis, and also when both lying and telling the truth. Two reasons for this are proposed. Firstly, for the experiment's inexperienced participants, the medium of interacting with another human who is embodied by a life-size avatar is likely to draw high levels of gaze. However, this novelty factor is likely to diminish after a short period of time, similarly to as discussed in Section 4.4.3. Hence, a more prominent reason is likely to be participants' knowledge that while their interactional partner can hear their voice and see various bodily movements in graphical form, their actual visual appearance is hidden and replaced by a digital puppet. In contrast, VMC participants were aware that a live video stream displaying their actual likeness at high resolution was being observed by the confederate. Additionally, the confederate was confronted with a live video of the participant as opposed to a generic digital avatar. Therefore, the intensity of the interpersonal interaction, and possible social awkwardness due to close-range eye contact with a stranger [AC76], were arguably heightened in the VMC sessions, thereby altering gaze behaviour. This argument is reinforced by the results of the POMS questionnaire shown in Section 6.2.6, which revealed that significantly higher degrees of both positive and negative psychological arousal were found in participants following VMC than following AMC.

Similarly to gaze behaviour, no main effect of mediation type was found when analysing pupil size on the macro scale of stage, or micro scale of individual questions. However, results reported in Section 6.2.6 indicated that, on both scales, VMC participants' pupils were slightly more dilated than the pupils of participants engaged in AMC. Pupil dilation indicates that an individual is in a state of either positive of negative emotional arousal [PS03]. Hence, the increased pupil dilation observed in VMC participants is consistent with results of both gaze analysis and the POMS questionnaire. Discussion of possible reasons for this have been presented above, and to which, pupil dilation results add further evidence. Finally, blink rate was similar between mediation types, with participants blinking slightly, but insignificantly, more frequently when engaged in AMC than when engaged in VMC. This indicates that blinking behaviour was not influenced by the visual stimuli of the varying telecommunication mediums.

Truth and Deception

Analysis of oculesics between conditions of truthful and deceptive response revealed significant differences in both gaze behaviour and pupil size. Initial analysis explored entire question stages, showing that lying participants in both mediation types gazed at the confederate for a consistently, but insignificantly, higher proportion of time than when telling the truth. However, a main effect of gaze behaviour was not exposed until the micro scale of individual questions was considered. This demonstrates that, during critical sequences of interaction, participants' state of veracity is highly influential to the gaze behaviour they exhibit. Previous studies in the deception literature have observed different gaze behaviours when lying and truth telling. For instance, Knapp et al. [KHD74] found that liars avoided directing gaze toward their conversational partner, resulting in significantly altered gaze behaviour than truth tellers. Conversely, a study by Burns and Kintz [BK76] found that liars employed a strategy of increasing the proportion of gaze directed at their conversational partner, which again resulted in a significant difference when compared to the gaze distribution of truth tellers. Explicating such contrasting findings, the detail of the specific experimental design and procedure provides the most logical rationale, and this was reflected by hypothesis E1H1, which predicted a contrast, rather than specifying a particular polarity. Results of the current experiment concurred with the latter finding, and participants were observed to direct significantly more gaze at the confederate when lying to them than when telling them the truth, as shown in Figure 6.7.

Pupil size was also affected by veracity, with participants exhibiting greater pupil dilation when telling lies than when telling truths. Again, analysis of key interactional segments revealed significantly higher levels of arousal, measured by increased pupil size, when engaged in deception compared to when engaged in truthful discourse. Blink rate again failed to reveal any consistent behavioural trends, suggesting that blinking behaviour was not impacted by state of veracity. Finally, strengthening the argument that participants behaved similarly in AMC and VMC, no significant differences in oculesics were found when comparing the two forms of visual telecommunication in truthful and deceptive scenarios: when telling the truth in AMC and VMC, similar oculesic patterns were found; and likewise, when lying in AMC and VMC, oculesic cues were similar.

6.3 Experiment 2: Detecting Deception

E2 aimed to assess the impact of avatars' oculesic behaviour on the extent to which observers are able to detect truthful and deceptive messages in AMC. E2 followed on from E1, inviting a different set of participants to view replays of the interactions performed in E1. Participants assessed the interactions in terms of veracity, rated on a scale of truthful or deceptive behaviour, and engagement, indicating how interested in the interaction the embodied users appear to be. Associated confidence levels for judgements of veracity and engagement were also collected.

6.3.1 Technical Preparation

Audiovisual captures of E1's interactions provided the experimental stimuli. During E1, two EyeCVE 'spectator' clients captured the unfolding AMC, regardless of whether the mediation, as presented to the participant and confederate, was being performed via video or avatar. Phase 3 of Section 6.2.5 explained how, for purposes of logging and keeping the pre-experimental procedure identical between both forms of telecommunication, EyeCVE operated in the background during the VMC condition. The intentions of E2 were also a motivation for this, doubling the amount of AMC footage for use as stimuli. The first spectator client processed participants' eye tracking data, thus capturing the AMC with full reproduction of oculesic cues. The second spectator client ignored eye tracking data, thereby capturing the AMC with no oculesic expression. In both cases, avatars' head and hand movement was driven by tracking, and mouth movement was synchronised with vocal input. Both spectator clients observed the

Figure 6.12: *E2's clip viewer interface.*

participant avatar from a perspective approximating the confederate's viewpoint during the experimental interactions. Beepa's Fraps [Pty10] was used to capture video at a resolution of 1200×720 pixels and 50 FPS and audio streams of both confederate and participant talk on the two spectator clients. The videos were then divided by question stages, resulting in a total of 132 clips over the 22 participants. A viewer application illustrated in Figure 6.12 was developed using Adobe Flash [Sys10a]. In terms of quality and perspective, the resulting audiovisual stimuli aimed to approximate that which was originally observed by the confederate during the AMC. Thus, E2 intended to gather many observers' perceptions of stimuli typically observed during conversational AMC, with the aim of collecting approximate judgements that may have formed had the observer been taking part in the actual AMC.

6.3.2 Experimental Design

A total of 27 participants (13 male) with normal or corrected vision and no knowledge of E1 were paid £5 to perform the experiment. Participants' ages ranged from 18–55, and European, Asian, and African origins were represented. Participants were divided into three groups (A, B, and C) of nine, with four or five males in each group. The clips representing E1's AMC were displayed in three conditions: video and audio with avatars exhibiting oculesic behaviour driven by eye tracking (*ET*); video and audio with avatars featuring no eye oculesic behaviour (*-ET*); and the audio-only component of the interactions (*AO*). Each group viewed a total of 24 clips, with an average duration of 1.5 minutes each. All 27 participants independently assessed each clip on a 1..7 Likert scale in terms of: veracity (1 = always lying, 7 = always telling the truth); engagement (1 = not at all engaged, 7 = completely engaged); and confidence corresponding to metrics of veracity and engagement (1 = very unsure, 7 = very sure). Each clip presented one question stage from E1 in which a participant responded by always lying or always telling the truth. Therefore, E2's Likert scale deliberately intended to mislead raters to believe that veracity varied over each clip. This was to preclude cases in which clearly true or false responses to individual questions could be used to decide the veracity rating of an entire stage, and also to encode

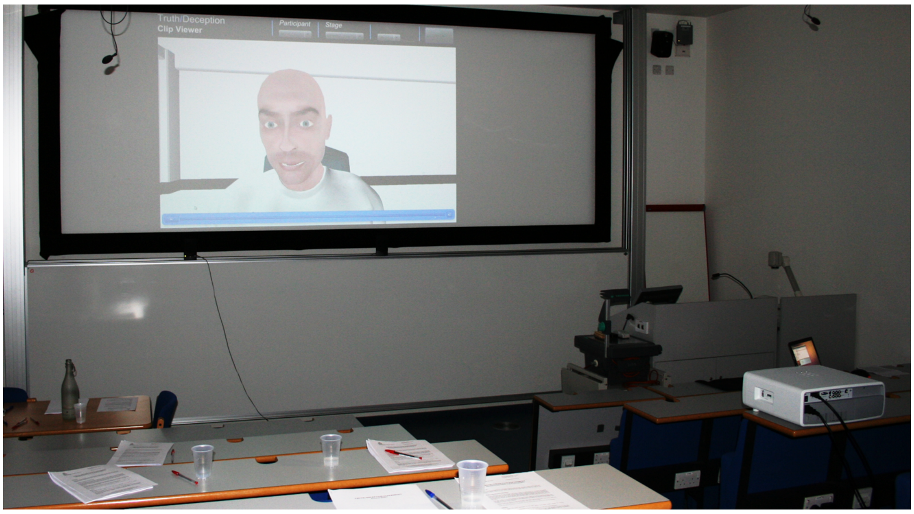


Figure 6.13: *Photograph of E2's setup, conducted in a lecture theatre at UCL.*

judgements in a richer manner than binary classification.

Each group rated the same clips in different conditions. For instance, if Group A viewed a particular clip in condition ET, then Group B would view the same clip in condition -ET or AO, and Group C in the remaining condition. Clip and condition orders were resequenced between the three groups so that all question stages and both response veracities were observed by each group. The experiment was performed in a lecture theatre at UCL, with the clips projected in their native resolution of 1200×720 pixels, and physically around 3×2 m, as shown in Figure 6.13. Prior to analysis, the Likert scores were encoded using a linear progression scale. Regarding a truthful clip, a participant who judged veracity with a rating of '1' would score 0% accuracy, '2' would equate to 16.7%, and so on until '7' resulting in 100% accuracy. Naturally, this scale was reversed for ratings of deceptive clips.

6.3.3 Procedure

Three experimental sessions were performed, with nine participants taking part in each. The experimental procedure consisted of three major phases: technical setup; briefing upon arrival of all participants; and the experimental rating task. The technical setup involved the experimenter initialising the projection and audio systems in the lecture theatre. Loud speakers were positioned to the sides of the projection screen. Following arrival of all participants, instruction sheets and ethics forms were provided. An answer booklet was also given to each participant in which to enter their responses. The Likert scale used for response was clarified, and any questions were answered. The experimental task began with the experimenter playing the first of the sequence of 24 clips, selected using the clip viewer shown in Figure 6.12. During the audio-only condition, the projector was simply turned off while the verbal component of the interactions were heard through the loud speakers. Following half of the 24 clips, the experimenter issued a five minute break before completing the second half of clips. Upon completion of all clips, participants were free to discuss the experiment before leaving. For additional materials, see Appendix E.

Table 6.6: *Percentages of veracity accuracy and confidence, and engagement and confidence between clip conditions when judging truth and lie question stages. ET: oculesic avatars, -ET non-oculesic avatars, AO: audio-only.*

Veracity	Metric	ET	-ET	AO
Truth	Veracity Accuracy	88%	70%	68%
	Confidence	78%	69%	72%
	Engagement	76%	58%	70%
	Confidence	85%	79%	82%
Deception	Veracity Accuracy	48%	39%	34%
	Confidence	74%	69%	73%
	Engagement	78%	54%	72%
	Confidence	83%	76%	80%

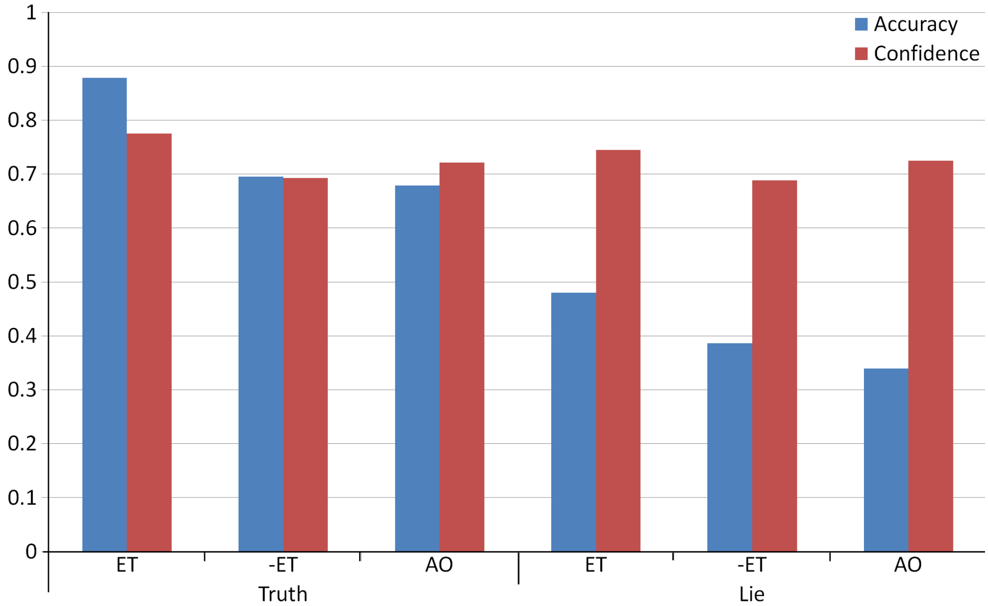


Figure 6.14: *Percentages of* veracity accuracy *and confidence between clip conditions when judging truth and lie question stages. ET: oculesic avatars, -ET non-oculesic avatars, AO: audio-only.*

6.3.4 Results

Table 6.6 shows all participants' mean scores of veracity accuracy, engagement, and confidence between the three experimental conditions, normalised to 1 and converted to percentages. Figures 6.14 and 6.16 clarify the data to display veracity accuracy and engagement individually. ET enables more accurate and confident assessment of the veracity of users engaged in both truthful and deceptive AMC than -ET or AO. Hence, an avatar displaying faithful replication of their embodied user's oculesic behaviour is able to inform more accurate judgement of their embodied user's state of truth telling and deception than an avatar featuring no oculesic behaviour (-ET), or audio-only stimuli (AO). The large difference in accuracy between judgements of truths and lies exposes the influence of the veracity effect [LPM99], with participants generally biasing responses toward truth. Additionally, ET elicited higher ratings of perceived engagement in the interaction than -ET and AO, and raters were again more confident in this decision.

Figure 6.15 illustrates where the main effects lie between the three stimuli conditions and for each

TRUTHS

Accuracy

ET	-ET	AO
88%	70%	68%

Engagement

ET	AO	-ET
76%	72%	58%

Confidence

ET	-ET	AO
78%	69%	72%

Confidence

ET	-ET	AO
85%	79%	82%

LIES

Accuracy

ET	-ET	AO
48%	39%	34%

Engagement

ET	AO	-ET
78%	72%	54%

Confidence

ET	-ET	AO
74%	69%	73%

Confidence

ET	-ET	AO
83%	76%	80%

Figure 6.15: *Significances of post-hoc Tukey tests for all questions and truth/lie data sets. Conditions jointly underlined are statistically similar. ET: oculesic avatars, -ET non-oculesic avatars, AO: audio-only.*

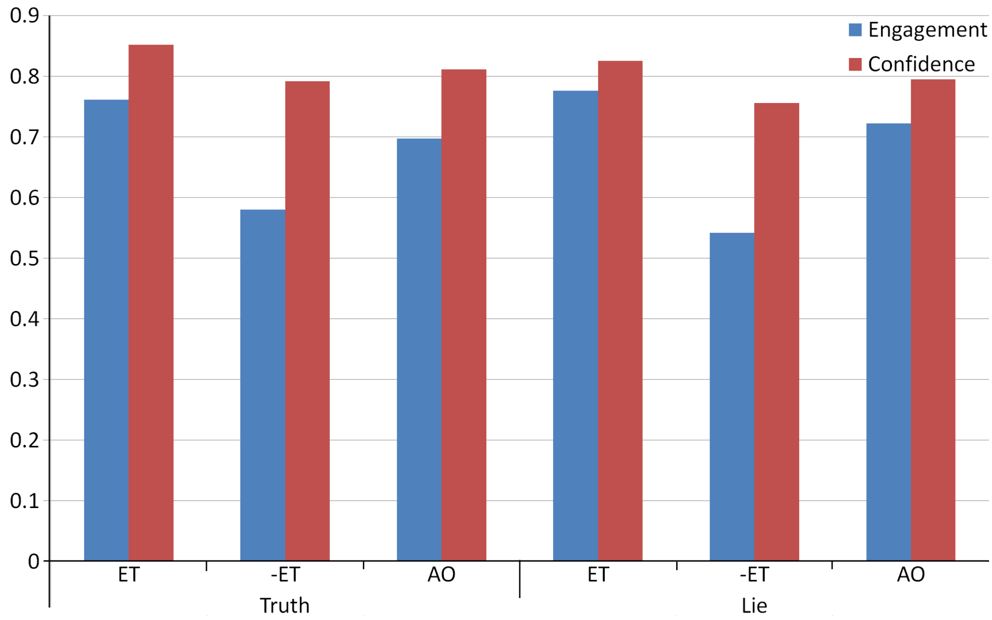


Figure 6.16: *Percentages of* engagement *and confidence between clip conditions when judging truth and lie question stages. ET: oculesic avatars, -ET non-oculesic avatars, AO: audio-only.*

response when rating both truths and lies. Conditions that are underlined by the same line may be considered statistically identical given a threshold of (p<.05). For instance, the line joining -ET and AO in the *Truths/Accuracy* block indicates that no significant difference was found between these two conditions when rating veracity of AMC in which the observed embodied user was telling the truth. However, as indicated by ET, which stands alone in the *Truths/Accuracy* block, a significant, and positive, difference in accuracy was found between ET and both -ET and AO when estimating the veracity of truth tellers. This finding of accuracy veracity is also observed for ratings of embodied users who are being deceptive, as indicated in the *Lies/Accuracy* block of Figure 6.15. Significantly more accurate judgements of the veracity of lying individuals are enabled when they are represented by oculesic avatars rather than non-oculesic avatars or the audio-only component of conversation. With regards to engagement, -ET performs significantly poorer than ET and AO when observing both truthful and deceptive AMC. ET elicits greatest confidence in all judgements, but this difference is only significant when judging truthful discourse.

6.3.5 Discussion

The stimuli presented during this experiment aimed to approximate that which was observed by the confederate during E1's experimental AMC, albeit in a non-immersive display. The motivation behind gathering a large number of subjective ratings of avatars operating with and without oculesics was to assess the extent to which faithful reproduction of oculesic cues is seen to enrich the nonverbal information transferred during AMC. These findings have implications for social interaction between users engaged in AMC, particularly relating to issues of interpersonal trust, including how the expressive characteristics of an avatar may influence observers' perceptions of the trustworthiness of interactional partners. The inclusion of an audio-only condition intended to provide a base class by which to compare the impact of the varying graphical stimuli.

Results of the subjective ratings support hypothesis E2H1: avatars featuring oculesic behaviour driven by eye tracking enable more accurate estimation of the veracity of an embodied user compared with assessment of avatars featuring no oculesics or audio-only stimuli. The addition of oculesic behaviour was also seen to raise the confidence of veracity judgements, but this was only significant when judging truthful discourse. With regards to engagement, oculesic avatars were rated similarly to audio-only stimuli, and significantly higher than avatars with no oculesic expression. The following discussion firstly addresses differences found between treatments, before focusing on the overall ability of observers to distinguish between truth and deception during AMC.

Judgements of veracity were more accurate when observing oculesic avatars, when their embodied users were both lying and telling the truth, than the other conditions of non-oculesic avatars and audio-only stimuli. As covered in Section 2.1.3, humans exhibit verbal and, critically, nonverbal behavioural cues correlated with lying and truth telling [DLM+03]. Correspondingly, people detect deception by observing others, so media transmitting fewer nonverbal channels becomes preferable for the deceiver [CGB+04]. The added nonverbal information that is captured and transmitted by oculesic avatars, resulting in more accurate estimation of veracity, demonstrates this phenomenon. Confidence in judgements of veracity was also seen to increase when avatars demonstrated oculesic cues. This is likely to be due to the added consistency between the aural and visual components of avatars' verbal and nonverbal behaviour. Differences between the non-oculesic avatars and audio-only stimuli were not significant when judging truths and lies. In this case, while the avatars did demonstrate head and mouth movement, the relative paucity of their expressive capacity resulted in ratings of veracity more similar to when judging the audio-only component of the interaction. However, it should be noted that, when judging both truths and lies, ratings of the non-oculesic avatars were slightly, but insignificantly, more accurate than the audio-only ratings. With regards to engagement, oculesic avatars and audio-only stimuli were rated significantly higher than non-oculesic avatars. This finding suggests that, when presented with a relatively inexpressive visual representation of an individual, ratings of liveliness are diminished in comparison with assessment of both solely aural information, and a depiction of the individual that demonstrates richer nonverbal expression.

Results revealed higher accuracy when judging truths than lies, with a mean accuracy of 75.3% over the three treatments for the former, and 40.3% for the latter. Two reasons for this disparity are suggested. Firstly, the veracity effect [LPM99], referring to peoples' natural tendency to judge more messages as truths than lies, is a factor often observed in experiments studying deception detection [AAD99]. Secondly, E1's experimental framework allowed participants to respond with deception in a manner open to interpretation. This may be further explained by post-experimental interviews following E1, revealing that many "lies" told by participants were in fact half-truths: deceptive statements that include some element of truth [Jam07]. For instance, when asked about what their activities "last" Saturday, post-experimental interviews revealed that some participants responded with information regarding some other Saturday. While such answers are still classified as lies in response to the specific question, they often resulted in credible responses, delivered with confidence, and perceptually inconspicuous in terms of cues to deception. Other questions, particularly those which followed on from one another, were more difficult to respond to with half-truths. For instance, in E1's General 2 phase, participants were asked *"Where was the last foreign country you visited?"*, which was then followed-up with the confederate asking them to *"Tell [him] some more about the trip"*. These issues demonstrate the complex nature of studying social interaction, with particular regard to scenarios involving deception and its detection. It should be noted that reported accuracies of deception detection are seen to vary greatly between studies and experimental design. Hence, a discussion of these results in relation to other studies, while interesting, is more suitable to a meta-analysis of deception detection studies as offered by Bond and DePaulo [BJD06], rather than specific experimental details as presented in this section.

6.4 Chapter Summary

The two experiments reported on in this chapter aimed to investigate a number of factors related to user behaviour, social presence, and media richness during truthful and deceptive discourse in visual telecommunication systems. E1 demonstrated that systematic differences in users' gaze and pupil size, correlated with telling truths and lies, are observed in both AMC and VMC. This suggests that users behave similarly when engaged in interpersonal communication when faced with video or avatar representations of interactional partners. However, the POMS questionnaire revealed psychological arousal to be greater following VMC than following AMC. E1's primary source of data collection and analysis was multimodal log files, which collated multiple streams from body tracking devices worn my participants. This process is detailed in Section 3.2. The NEO Five-Factor Inventory pre-experimental questionnaire was used to balance participant groups between the two conditions of AMC and VMC, and the POMS post-experimental questionnaire was used to gather subjective response.

E1's interactions were recorded and used as stimuli for E2, which involved another set of participants observing and rating clips of AMC. Results from E2 demonstrated the interactive benefit of eye tracking to bestow avatars with representations of their embodied users' oculesic cues of gaze, blinks, and pupil size, enabling observers to more accurately detect truth and deception. Thus, all experimental hypotheses were supported, but variations of blinking behaviour were not uncovered. Analysis of gaze and pupil dilation emphasised the importance of isolating critical periods of interaction due to the temporally-transient nature of a potential characteristic response following a given stimuli.

This chapter completes the telecommunication experiments presented in this thesis. The following chapter presents three further experiments which, similarly to E2, aim to gather subjective ratings of avatars and their oculesic behaviour over varying control methods and fidelities.

Chapter 7

Models of Oculesic Behaviour and Experiments

This chapter presents simulations of the oculesic behaviours of eyelid kinematics and gaze, together with associated experiments. The work does not investigate AMC directly, but rather assesses characteristics of avatars that may be influential to AMC. This work is motivated by the advance of real-time computer graphics, which enable the use of highly realistic virtual humanoids in ICVE systems. Thus, varying levels of representational and behavioural fidelity, likely to be influential to how virtual humanoids are perceived during AMC are investigated. The work appearing in this chapter was performed in parallel to the telecommunication experiments presented in Chapters 4–6. Sections 7.1 and 7.2 are concerned with eyelid animation, and hence are relevant to EyeCVE's advanced avatars which were utilised in Chapter 6's truth and deception experiments. As explained in Section 6.2.3, participants engaged in truthful and deceptive AMC were embodied by avatars featuring eyelid kinematics driven by the models presented in Section 7.2 of this chapter. Hence, chronologically, the eyelid models were developed and investigated before the truth and deception experiment. Section 7.3 is concerned with gaze simulation, and so is relevant to all types of virtual humanoids, regardless of facial animation capability. The gaze model was developed in response to the poor performance of the basic random gaze model presented and investigated in Chapter 5, and analysis of eye tracking gaze data captured during that object-focused experiment forms the basis of the model.

This chapter is divided into three sections. The first section details an experiment investigating the impact of varying the fidelity of an autonomous agent's oculesic behaviour on observers' ability to identify direction of gaze, and on subjective perceptions of realism. The second section documents two models generating the components of human eyelid movement. Simulations of blinking and lid saccades are presented algorithmically before they are validated against motion captured movements. An associated experiment, assessing the impact of the models on observers' perception of realism and believability of an avatar is then presented. Finally, the third section presents a gaze model developed using data collected from Chapter 5's object-focused experiment. The model is first presented algorithmically, followed by an experiment investigating how observers' subjective ratings of realism and believability are influenced by an avatar's method of gaze control.

7.1 Oculesic Behaviour Experiment

The following experiment explores the influence of varying oculesic characteristics of a virtual humanoid agent by measuring how accurately observers are able to identify the agent's direction of gaze, together with subjective assessments of realism and believability. Two elements of oculesic behaviour were studied: eyelid animation and vergence. Based on preliminary observation tests, these oculesic characteristics were expected to have a significant impact on both accuracy of determining gaze direction and perceived realism of the agent. Eyelid animation denotes that the agent's eyelids performed blinks and lid saccades. Models generating physiologically accurate animation of these two major components of human eyelid motion are presented in Section 7.2, while this experiment does not define dynamics. Vergence indicates that the agent's eyes rotate inwards to focus on a precise point in space, rather than the eyes gazing parallel to a focal point with the same rotation value for both eyes.

The two oculesic characteristics formed the experiment's independent variables, and were investigated over the following four stimuli conditions: no eyelid animation, no vergence; no eyelid animation, vergence; eyelid animation, no vergence; and eyelid animation and vergence. The main hypotheses were that, by more closely mimicking natural human behaviour, the inclusion of eyelid animation and vergence would lead to higher accuracy of gaze direction identification, and also to elicit higher subjective ratings of realism.

7.1.1 Technical Preparation

A virtual humanoid agent acted as the main experimental stimulus. The agent was developed using Smith Micro Poser 7 [Sof10b] and Autodesk 3DS Max 8 [Aut10]. The agent's eyes were comprised of separate geometry to the head, allowing for eye rotation. Eyelid animation was implemented using blend shapes, which linearly interpolate between target geometry models, each specifying an 'extreme' pose of facial expression. Controllers attached to the geometry of the eyes, allowing fine control of rotation were also implemented. The agent setup, in 3DS Max 8, is shown in Figure 7.1.

A simple VE was also designed in 3DS Max, consisting of an analogue clock face positioned directly in front of the agent's head at a distance of 0.5 m. Three virtual cameras were positioned directly in front of the agent and symmetrically to its front-left and front-right. The cameras intended to provide

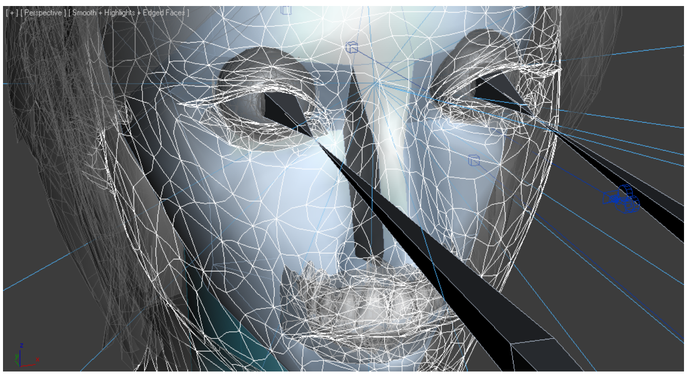


Figure 7.1: *Agent development in 3DS Max 8 [Aut10]. Image shows close view of agent's eye area. Note controllers extruding from the eyes to allow fine control of eye rotation.*

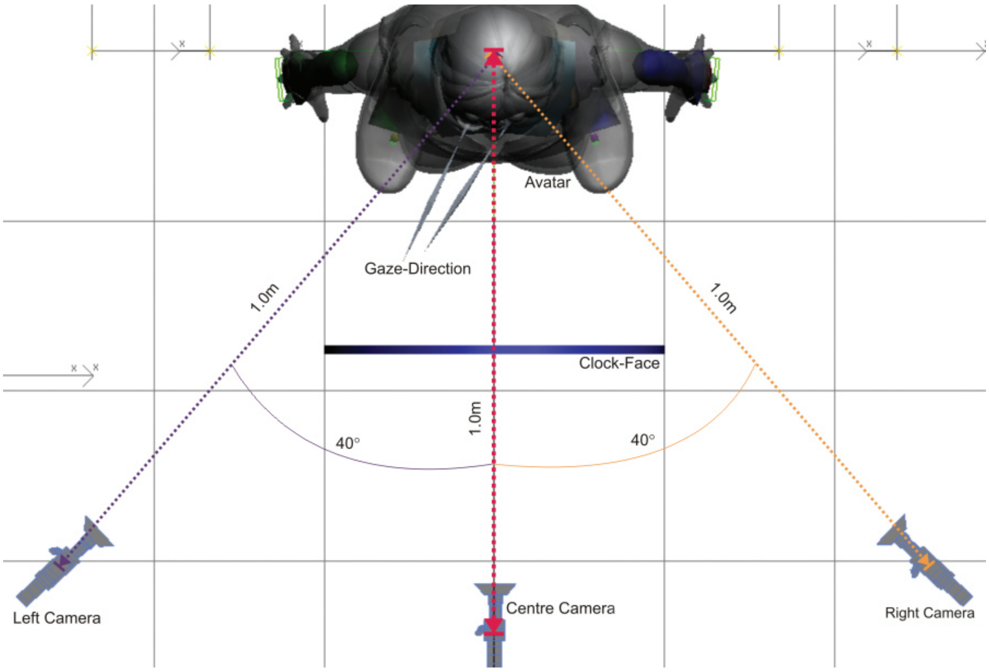


Figure 7.2: *Experiment setup from above. The virtual cameras are positioned at a distance of 1.0 m to the agent. This separation between bodies is in the far-phase of personal distance [Hal68], and aims to both promote affinity between participant and agent and maintain visual clarity for the task. Side-cameras were offset to +/-40° from the 0° central camera, remaining consistent with common conversational situations.*

viewpoints common to dyadic and multiparty AMC scenarios as demonstrated in the telecommunication experiments in Chapters 4–6. The experimental scene is illustrated in Figure 7.2.

Twelve video clips (one for each number around the clock face) were generated in each of the four combinations of eyelid animation and vergence, resulting 48 clips rendered from each camera angle, and 144 over the three cameras and twelve gaze positions. The duration of each clip was eight seconds, and the clips followed the sequence illustrated in Figure 7.3: the agent looks straight ahead; looks towards a specific number on the clock face for two seconds; returns gaze to the straight-ahead position for one second; returns gaze to the same number on the clock face for another two seconds; and finally returns gaze to the initial straight-ahead position. Each shift of gaze was performed in a single saccade. Hence, the stimuli presented an agent gazing at targets, which varied gaze direction horizontally, vertically, and to a lesser extent in depth. In order to isolate results to describe the impact of the varying oculesic behaviours, the eyes (and eyelids in the two eyelid animation conditions) were the sole animated elements of the agent. An experimental application illustrated in Figure 7.4 was developed using Adobe Flash [Sys10a].

7.1.2 Experimental Design

A within-subjects, repeated measures design was conducted over the four conditions to determine the impact of eyelid animation and vergence on participants' ability to identify the agent's gaze direction and on how the agent is perceived in terms of realism. At total of 48 (24 male) participants with normal or corrected vision were recruited from UCL's campus through an advertising poster campaign and paid

Figure 7.3: *Sequence of agent behaviour performed in each clip from left to right: agent begins looking straight ahead, gazes at a specific number on the clock face, returns gaze to the straight-ahead position, gazes at the number again, and finally returns gaze to the straight-ahead position.*

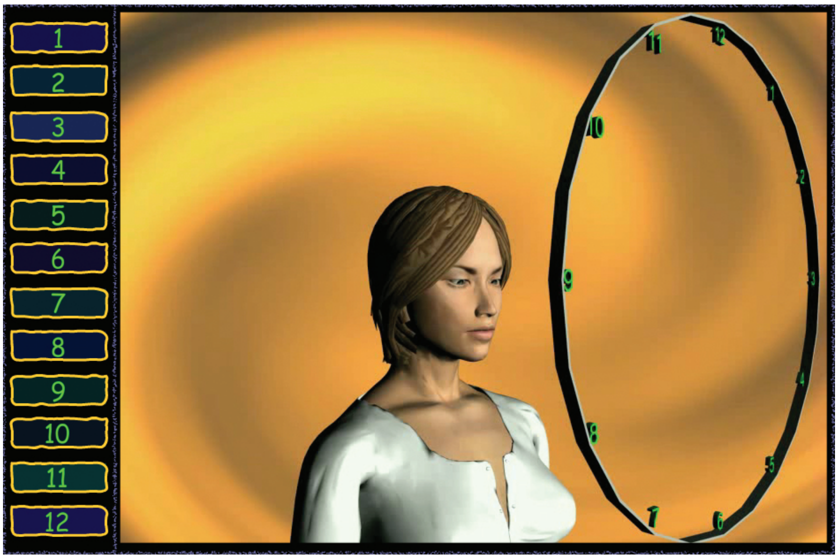


Figure 7.4: *Experimental interface for gaze direction identification task. Agent is being viewed from the left camera, and exhibiting* eyelid animation, no vergence *behaviour. In order to eliminate the use of markers, such as the mouse pointer, during POR identification, an input panel to the left of the main window was used rather than clicking on the clock face itself.*

£3 to perform the experiment, which lasted for approximately 20 minutes. Ages ranged from 18–30, and European, Asian, and African origins were represented. Participants viewed all 144 clips in blocks of 36. Each block consisted of clips featuring the agent performing in one of the four conditions, looking at the twelve numbers around the clock face from the three angles. The order of presentation of the conditions were resequenced over participants to eliminate inter-experimental effects. Two metrics were collected: estimation of gaze direction for each clip; and overall ratings of realism for each block of 36 clips, corresponding to a single experimental condition. Input was integrated into the experimental viewer, with input of gaze direction estimation performed using a side-panel of buttons as shown in Figure 7.4, and input of subjective ratings performed using a questionnaire following the last clip of each condition as shown in Figure 7.5. The subjective ratings were collected on a 1..7 Likert scale. Questions related to perception of the realism and authenticity of the agent's behaviour and appearance, together with self-performance ratings of the gaze direction identification task. For additional materials, see Appendix F.

7.1.3 Procedure

The experiment was performed on a 42" HD LCD screen which displayed the agent's head approximately life-size. Participants were seated 1.0 m from the screen at the agent's eye-level. The experi-

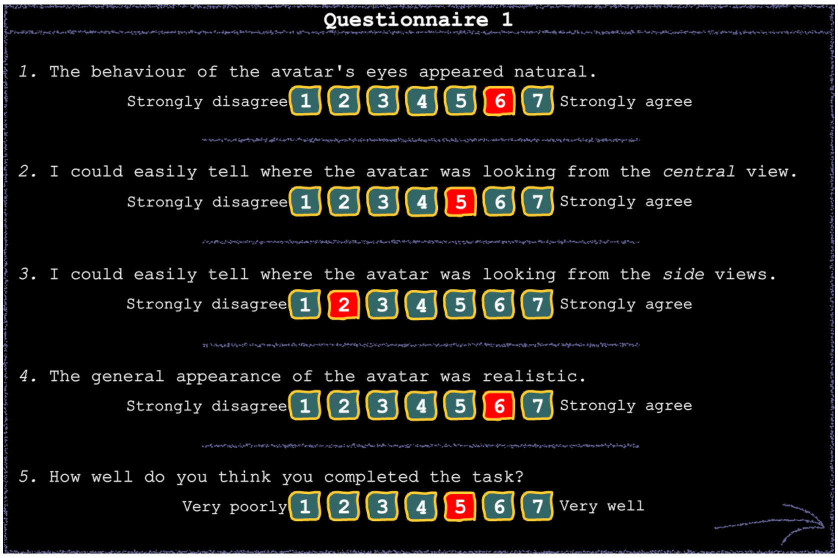


Figure 7.5: *Experimental interface when performing subjective questionnaire following a block of 36 gaze direction clips in which the agent exhibited a single oculesic condition. Ratings were input on a 1..7 Likert scale.*

mental procedure consisted of three major phases: briefing; experimental task; and post-experimental interview. Following arrival of the participant, instruction sheets and ethics forms were provided. The Likert scale used to gather subjective response was clarified, and any questions were answered. The experimental task began when the experimenter left the lab, and the participant initiated the first clip using the viewer application. The clips advanced automatically following the participant's input of estimated gaze direction, with the viewing camera alternating in a repeating sequence (centre, left, right), and the agent looking at a different number. Following the 36 clips, the application presented the questionnaire for the participant to complete. The application then proceeded with the next set of clips and questionnaire. This process repeated for all four conditions. In summary, for each of the four varying conditions of the agent's oculesic behaviour, participants' ability to identify direction of gaze, and subjective responses regarding realism were recorded.

7.1.4 Results

Results relating to gaze direction identification and subjective ratings were analysed independently.

Gaze Direction Identification

A repeated measures two-way ANOVA was performed on the gaze direction identification data, with the four oculesic conditions and the twelve gaze angles (numbers on clock face) as factors. This overall measure combined data from all three cameras, and indicated a main effect of both oculesic behaviour ($F(3,11)=4.19; P<0.01$), and gaze angle ($F(3,11)=11.26$; $P<.001$). Post-hoc Tukey tests determined that the *eyelid animation and vergence* condition performed significantly poorer than the classes of: *eyelid animation, no vergence* ($F(1,11)=11.73; P<.001$) and *no eyelid animation, no vergence* ($F(1,11)=4.03$; $P<.005$). An interaction effect was found between oculesic condition and gaze angle during the latter comparison ($F(1,11)=2.02; P<.05$). The column entitled 'All Cameras' in Table 7.1 shows this overall decrease in accuracy of identification of gaze direction as the behavioural fidelity of the agent increases.

Table 7.1: *Mean (and standard deviation) gaze direction identification accuracy for conditions and cameras.*

Condition	All Cameras	Centre Camera	Side Cameras
No Eyelid Animation, No Vergence	80.7% (0.72)	89.2% (0.53)	76.5% (0.79)
No Eyelid Animation, Vergence	79.4% (0.73)	92.2% (0.43)	73.0% (0.83)
Eyelid Animation, No Vergence	78.9% (0.58)	93.8% (0.24)	71.4% (0.67)
Eyelid Animation, Vergence	73.0% (0.64)	89.8% (0.54)	64.7% (0.67)

The most basic oculesic behaviour is seen to perform best, but the difference is only significant compared to the *eyelid animation and vergence condition*. This corresponds to the above ANOVA calculations, suggesting that, when analysed on this macro scale considering ratings from all three cameras, the inclusion, and in particular the combination, of the two oculesic cues of eyelid animation and vergence are detrimental to observers' ability to accurately judge gaze direction, thereby opposing part of the stated hypothesis.

The influence of camera angle on the accuracy of gaze direction identification was then explored. As shown in Table 7.1, the centre camera formed one group, and the two symmetrical side cameras were evaluated together. Starting with the centre camera, a repeated measures two-way ANOVA was performed with the four oculesic conditions and the twelve gaze angles as factors. A main effect was found between the oculesic conditions ($F(3,11)=3.26$; $P<.05$), and an interaction effect was found between factors ($F(3,11)=1.73$; $P<.01$). Post-hoc Tukey tests exposed that the differences lay between the *eyelid animation, no vergence* condition and two other classes: *eyelid animation and vergence* ($F(1,11)=7.64$; $P<.01$) and *no eyelid animation, no vergence* ($F(1,11)=6.99$; $P<.01$). The same tests were then performed on the data for the two side cameras, again finding a significant difference between oculesic conditions ($F(3,11)=4.02$; $P<.01$). Similar to the overall camera analysis, differences were found between the class of *eyelid animation and vergence* and the two other classes of *eyelid animation, no vergence* ($F(1,11)=6.92$; $P<.01$) and *no eyelid animation, no vergence* ($F(1,11)=11.83$; $P<.001$). An interaction effect was found during the latter comparison ($F(1,11)=1.98$; $P<.05$). Table 7.1 highlights consistently high accuracy from the centre camera over the four conditions, and lower, more variable accuracy, when judging from the side cameras.

The above analysis shows that accuracy of gaze direction identification is influenced by the agent's actual gaze direction, in combination with the observing camera angle. Therefore, analysis between symmetrical gaze angle 'pairs' was performed in order to isolate scenarios and expose specific combinations of gaze direction and camera angle with high impact. The experimental design of a clock face with one central and two symmetrical side cameras granted identical states for opposite gaze angles according to the side camera considered. For instance, a one o'clock gaze viewed from the right camera is equivalent to an 11 o'clock gaze from the left camera. Accordingly, 2 left = 10 right, 3 left = 9 right, 4 left = 8 right, 5 left = 7 right. Gaze angles 6 and 12 were treated independently, as the pair is not equivalent. These logical pairings of symmetrical gaze angles allowed direction of observed gaze to be classified more generally as *toward* and *away*: terms more meaningful to interaction in AMC as demonstrated during gaze analysis in the telecommunication experiments documented in Chapters 4–6. In this

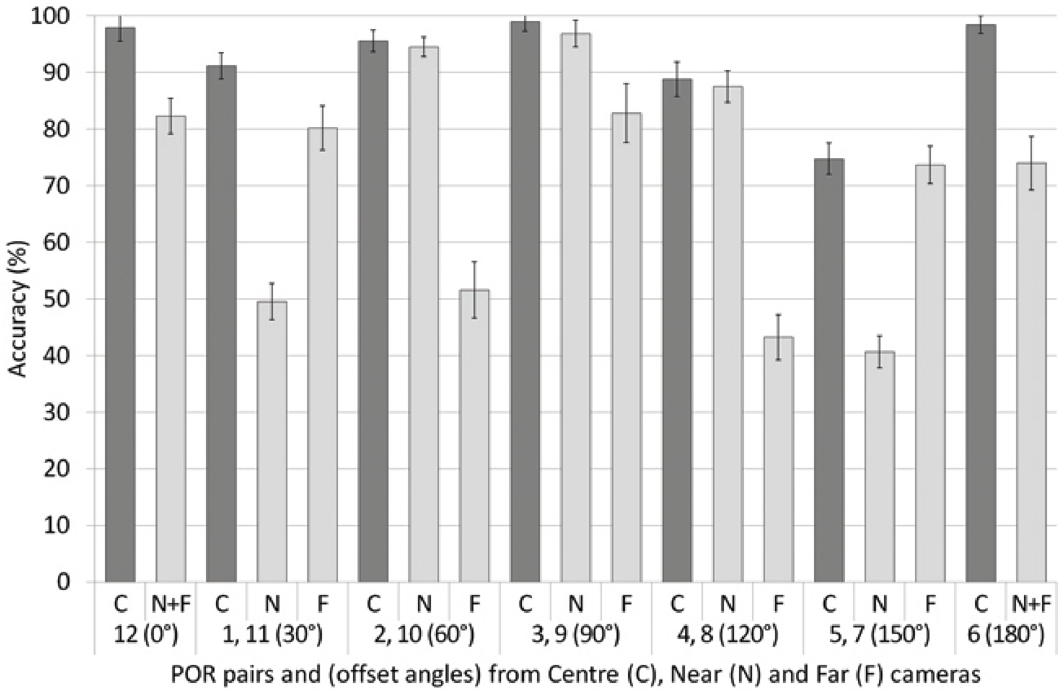


Figure 7.6: *Overall combined condition gaze identification accuracy from centre (C), toward (T), and away (A) cameras for symmetrical gaze angle pairs. Note that the 0°/180° vertical pair (6 and 12) remain separate, as the agent eye representation differs greatly.*

context, toward gaze indicated that the agent was looking in the hemisphere in which the observer lies, and away gaze indicated that direction of gaze was in the opposite direction of the perspective of the observer. Prior to this analysis, the validity of the pairings was confirmed by performing repeated measures two-way ANOVA on the data for the two gaze angles within each pair, and no significant differences were found. Overall accuracy levels for gaze angle pairs are shown in Figure 7.6 for centre, toward, and away perspectives. Certain combinations of gaze angle and camera angle show a significant and negative impact. In particular, 30° offset from vertical (pairs 1/11, 5/7) when observers are positioned in the toward-hemisphere, and the 60° offset from vertical (pairs 2/10, 4/8) when observers are positioned in the away-hemisphere.

Subjective Rating

The questionnaire sought to elicit judgements regarding the agent's realism and participants' ratings of self-performance during the gaze direction identification task for each condition. Question 1 (*"The behaviour of the avatar's eyes appeared natural"*) and Question 4 (*"The general appearance of the avatar was realistic"*) directly addressed perception of realism. A critical requirement of the experimental design was to isolate animation to the eyes and surrounding areas. Thus any variation in reported levels of perceived realism could be related solely to changes in oculesic representation. A one-way ANOVA evaluation of combined data from Questions 1 and 4 exposed a highly significant difference between conditions ($P<.001$). Post-hoc Tukey tests exposed differences to lie between the two classes featuring eyelid animation, and the two that did not, ($P<.001$). Table 7.2 presents the Likert scale responses, showing how the inclusion of eyelid animation is able to enhance perceived realism, while vergence was

Table 7.2: *Mean response and (standard deviation) for Questions 1, 4 and 5 on the 1..7 (negative..positive) Likert scale questionnaire.*

Condition	Question 1	Question 4	Question 5
No eyelid animation, no vergence	3.75 (1.48)	4.29 (1.32)	4.87 (1.24)
No eyelid animation, vergence	3.52 (1.53)	4.35 (1.38)	4.53 (1.36)
Eyelid animation, no vergence	5.69 (1.06)	5.98 (1.13)	5.42 (0.82)
Eyelid animation, vergence	5.33 (1.17)	5.85 (1.14)	5.16 (1.10)

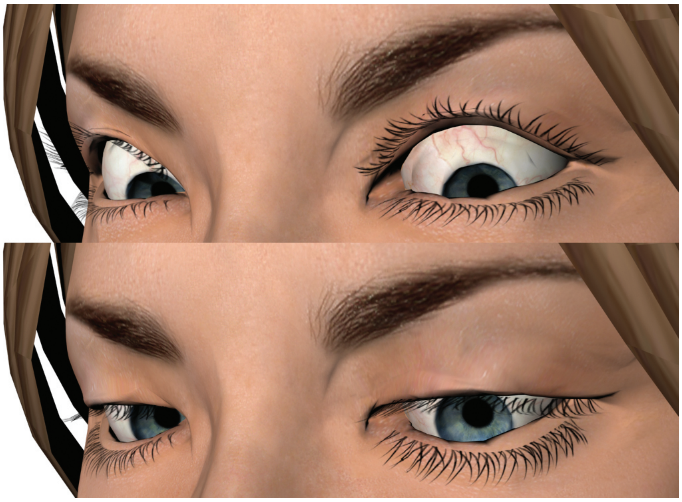


Figure 7.7: *Detailed screen captures of the agent's eyes when looking at number 5 (150° offset from vertical gaze angle 12) from the toward (right) camera. Top: no eyelid animation, no vergence. Bottom: eyelid animation and vergence.*

not a determining factor. Responses to Question 4 also demonstrate the influence of eyelid animation on the agent's holistic realism. There was no significant difference found between the two classes featuring eyelid animation (*P*=.14), adding to the evidence that vergence was not a factor affecting the agent's realism. Hence, the subjective component of the original hypothesis is supported by findings related to the impact of eyelid animation, but not supported by findings related to vergence.

One-way ANOVA evaluations of responses to Questions 2 and 3 (*"I could easily tell where the avatar was looking from the centre view / side views"*) did not expose significant differences between conditions despite significant differences in accuracy found during analysis of gaze direction identification data. However, a significant difference was found between conditions for Question 5 (*"How well do you think you completed the identification task?"*) ($P<.05$), with eyelid animation conditions earning superior ratings. This suggests that confidence in identifying the agent's direction of gaze was raised by the enhanced realism afforded by the addition of eyelid animation, despite actual gaze identification accuracy not always corresponding to this perception. An example rendering of the agent's eyes in the *eyelid animation, vergence* condition and the *no eyelid animation, no vergence* condition is presented in Figure 7.7.

7.1.5 Discussion

The experiment presented in this section explored how varying conditions of an autonomous agent's oculesic behaviour is seen to influence both accuracy of gaze direction identification and perceived realism. Results from the gaze direction identification task showed that accuracy varied depending on three factors: the agent's condition of oculesic behaviour; the agent's current gaze direction; and the angle from which the agent was being observed. In particular, accuracy of gaze direction identification was hindered when the agent exhibited eyelid animation and was viewed from side cameras. A particular example is demonstrated by the gaze angle pair 5/7 (30° vertical offset) shown in Figure 7.6. In this case, accuracy was seen to drop to 40.6% when viewed from the 'toward' camera. Conversely, when this same gaze angle pair was viewed from both centre and far cameras, accuracy increased to over 70%. Observations of human physiology elucidate this phenomenon: when operating with eyelid animation, the agent's eyelids are relatively closed in order to remain approximately aligned with the top of the cornea as the eyes depress to a low gaze angle. Thus, the agent's eyes are partially obscured depending on the angle of observation, but particularly in the toward hemisphere. Hence, when viewing from this angle, the visible area of the sclera is reduced, making the horizontal rotation of the agent's eyes more difficult to distinguish. The higher accuracy of gaze direction identification when viewing from both centre and away cameras reflects the fact that the horizontal offset of the agent's eyes was more clearly distinguishable. Finally, it is noteworthy to state that, in the example of the 5/7 pair, 99.2% of the total of 58.9% incorrect responses identified the agent's gaze direction as the vertical number 6. This observation applied to other situations where the agent's direction of gaze was unclear, and participants appeared to choose 'safe' or 'default' gaze directions of exactly vertical (numbers 6 and 12) or horizontal (numbers 3 and 9).

Judgement of the agent's realism, both as a whole and when considering just the eyes, was significantly enhanced by eyelid animation, while vergence was not a significant factor. The primary reason for the low-impact of vergence was likely to be the limited variation in depth of gaze targets due to the experiment's clock face design, together with static viewpoints restricting observation of the depth cue. Participants' self-performance assessments of the gaze direction identification task were significantly higher when judging eyelid animation conditions despite an overall performance reduction during the task. This implies that the increased visual and behavioural realism of the agent fostered higher levels of confidence when judging gaze direction, contradictory to actual identification accuracy performance.

Chronologically, this experiment was performed first all those detailed in this thesis. The experimental results informed development of EyeCVE's avatar subsystem, and also have implications on the more general design of virtual humanoids. The dynamics of lid saccades and blinking, which were approximated in this experiment, are subsequently formalised as parametric models, aiming to generate physiologically accurate eyelid kinematics. These models are presented and investigated in the following section of this chapter.

7.2 Eyelid Kinematics Models and Experiments

Section 7.1 demonstrated that animation of a virtual humanoid's eyelids plays an important part both in clearly conveying direction of gaze and in improving visual and behavioural realism. However, the dynamics of the agent's eyelid animation, featuring blinks and lid saccades, were not formalised. In the wider computer graphics and VE literature, eyelid kinematics for virtual humanoids is still a relatively unexplored area.

The human eyelids have two major motion components: lid saccades that follow the saccadic motion of the eyes, and blinking. Derived from literature in ophthalmology and psychology, the work presented in this section documents parametric models for both motion types, and emphasises their dynamic temporal behaviour. Firstly, the dynamics of lid saccades and blinks are parametrised in order to act as inputs to the models. The models are then presented algorithmically. Experimental validation follows, comparing animation generated by the models with the alternative methods of expensive motion captured data and simple linearly interpolated animation. Finally, an experiment investigating the general impact of the models on an avatar's overall realism is presented. The models are demonstrated to be suitable both for offline and real-time application: the former implementation provides stimuli during model validation, while the latter is implemented in EyeCVE during the general realism experiment.

7.2.1 Parametrisation

The equations presented in tables 7.3 and 7.4 are taken from empirical studies of eyelid movements in normal subjects presented by Evinger et al. [EMS91]. The metrics are based on data recorded from a comprehensive study measuring upper eyelid movements in normal human subjects. It should be noted that, while the lower eyelids also exhibit motion during blinks and lid saccades, their dynamics have never been measured without large intersubject variability [MeCM+05]. Compared to the upper eyelid, lower eyelid movements are relatively minor, with amplitudes of just a quarter of a corresponding upper eyelid movement [HD82]. They are also entirely passive, resulting from the transmission of forces of other muscles such as the levator, which contracts or relaxes to initiate upper eyelid movement [MeCM+05]. Consequently, the models do not consider lower eyelid movement explicitly. However, the virtual humanoids appearing in this section were developed so that slight lower eyelid movement is coupled to the movements of the upper eyelids, thereby exhibiting motion in the order of one quarter to that performed by the upper eyelid.

Regarding blinks, and promoting the suitability of modelling eyelid kinematics parametrically, Evinger et al. note the small amount of intersubject variability across a wide range of blink amplitudes and conditions [EMS91]. Thus, whether a blink originates from reflex, voluntary, or spontaneous cause, over 95% of more than 400 blinks obtained from nine subjects fell within a narrow window of velocities that the following equations describe. Likewise, when defining the characteristics of lid saccades, a similar relationship to blink dynamics between amplitude and maximum velocity also exists [EMS91]. Common to lid saccades and blinks are the up- and down- phases of movement. In the case of blinks, the down-phase refers to the initial downward movement of the eyelids (Equations 1 and 3), and the up-phase refers to the subsequent upward movement (Equations 2 and 4) to return the lids to their resting position based upon vertical eye rotation. In contrast, lid saccades exhibit only one phase of movement, which depends on the vertical direction of the eye saccade: an upward shift in eye gaze initiates a corresponding upward motion of the eyelids defined with up-phase properties (Equations 5 and 7), while a downward gaze shift initiates movements exhibiting down-phase dynamics (Equations 6 and 8).

Table 7.3: *Equations acting as input to the* ***lid saccade*** *model.*

E#	*Equation*
E1	$D=33.2+5.9A-0.069A^2$
E2	$D=98.9+3.6A-0.042A^2$
E3	$V=45.31A^{0.599}$
E4	$V=13.3A-14.82$

Table 7.4: *Equations acting as input to the* ***blink*** *model.*

E#	*Equation*
E5	$D=36.3+1.4A-0.016A^2$
E6	$D=87.9+4.3A-0.047A^2$
E7	$V=29.2A-35.9$
E8	$V=13.5A-5.87$

Lid Saccade Model Parameters

E1 defines the duration (*D*) of the down-phase of a lid saccade, given an amplitude of lid movement as measured in degrees (*A*). *E2* defines the duration (*D*) of the up-phase of a lid saccade, given an amplitude of lid movement (*A*). *E3* defines the maximum velocity *V* for the down-phase of a lid saccade, given an amplitude of lid movement (*A*). *E4* defines the maximum velocity *V* for the up-phase of a lid saccade, given an amplitude of lid movement (*A*).

Blink Model Parameters

E5 defines the duration (*D*) of the down-phase of a blink, given an amplitude of lid movement (*A*). *E6* defines the duration (*D*) of the up-phase of a blink, given an amplitude of lid movement (*A*). *E7* defines the maximum velocity *V* for the down-phase of a blink, given an amplitude of lid movement (*A*). *E8* defines the maximum velocity *V* for the up-phase of a blink, given an amplitude of lid movement (*A*).

7.2.2 Eyelid Saccade Model

Figure 7.8 presents a state diagram detailing the mechanics of the lid saccade model:

1. The initial input to the lid saccade model is gaze data. This will typically be generated by a gaze model or eye tracker. Given a gaze motion, the vertical amplitude shift is calculated by comparing the current vertical rotation of the eye with the new input angle.

2. This positive or negative (indicating an upward or downward shift) angle is then thresholded, possibly along with a signal sent from the source of gaze input to determine if the gaze motion is indeed a saccade or a minor or smooth pursuit movement.

3. Before the output from this classification is utilised, the final eyelid position is calculated by taking into account an overshoot probability function, and critically, the properties of the virtual character itself under control: this final position may depend on bone rotations if a skinning approach has been implemented, or blend shape weightings if linear geometry blending is the method of facial animation.

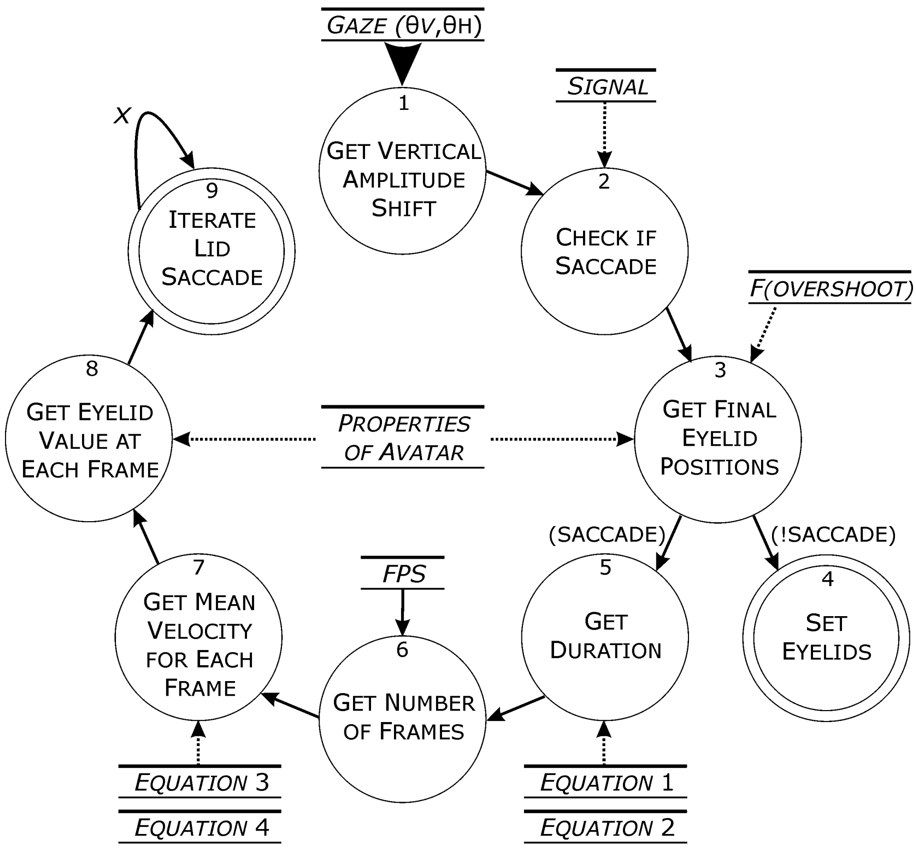


Figure 7.8: *State diagram detailing the lid saccade model.*

4. The controlling thread then splits depending on the classification of the input gaze motion: absence of a saccade classifies the movement as minor (for instance associated with a microsaccade or smooth pursuit) and sets the eyelids in their final position in a single frame before quitting. Execution continues if a saccade is detected.

5. If a saccade has been detected, vertical amplitude shift acts as input to Equation 1 (in the case of a downward saccade) or Equation 2 (for an upward shift) in order to determine the duration of the lid saccade.

6. The number of frames is then calculated, based on graphical frame rate, using the equation $N=DF$, where N is number of frames, D is duration of lid saccade, and F is FPS.

7. Either Equation 3 (downward saccade) or Equation 4 (upward saccade) is then used to calculate the mean velocity of each frame of the lid saccade.

8. The eyelid values for the particular virtual character are then determined for each frame using a variation of *distance=speed×time*: $Af=Vf((D/N)Nf)$, where Af is the amplitude for a given frame, Vf is mean velocity of the frame, D is the total duration of the lid saccade, N is number of frames and Nf is current frame being calculated.

9. Finally, for each of the calculated frames, the character's eyelid is set appropriately, animating the lid saccade.

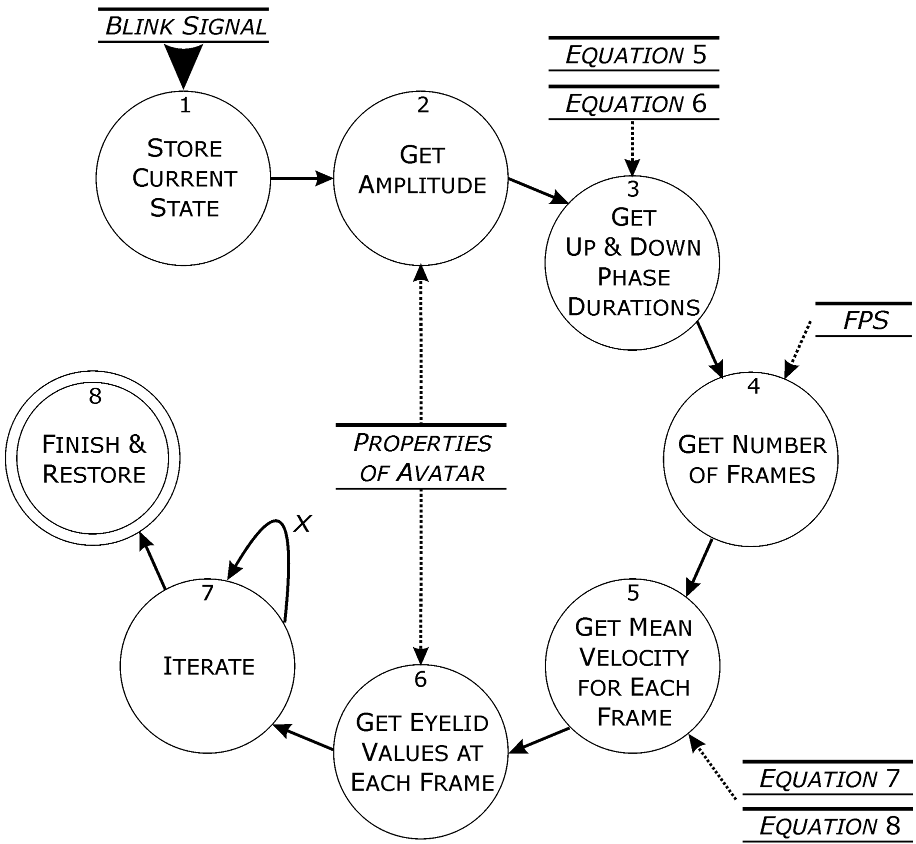


Figure 7.9: *State diagram detailing the blink model.*

7.2.3 Blink Model

Figure 7.9 presents the state diagram detailing the mechanics of the blink model:

1. Upon receiving a blink signal (for instance, from an eye tracker or gaze model), the current position of the character's eyelids are stored in order for the model to return to that position following completion of the blink.

2. Based upon the current state and the 'eyelid fully closed' state of the specific virtual character's eyelids, the angular amplitude of the blink motion is calculated.

3. Amplitude then acts as input to Equation 5 when calculating the down-phase of the blink and Equation 6 for the up-phase of the blink in order to determine the duration of the two blink phases.

4. The number of frames is then calculated, based on graphical frame rate, using the equation $N=DF$, where N is number of frames, D is duration of the blink, and F is FPS.

5. Equation 7 (down-phase) and Equation 8 (up-phase) are then used to calculate the mean velocity of each frame of the blink motion for both phases.

6. The eyelid values for the specific character are then calculated for each frame using the equation $Af=Vf((D/N)Nf)$, where Af is the amplitude for a given frame, Vf is mean velocity of the frame, D is the total duration of the down or up phase of the blink, N is number of frames and Nf is current frame being calculated.

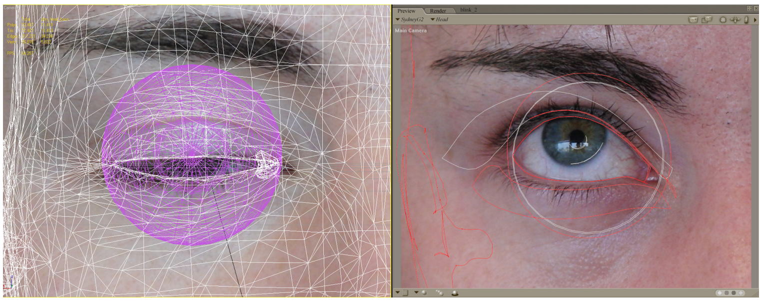


Figure 7.10: *Motion capture encoding onto two different avatars using 3DS Max 2010 [Aut10] (left) and Poser 7 [Sof10b] (right).*

7. The calculated frames are then iterated, animating the blink.

8. Finally, virtual character's initial eyelid state is restored, and control returns to the lid saccade model if it is operating.

7.2.4 Model Validation Experiment

The following experiment aimed to validate the models by comparing model-generated animation with two alternative and common methods of eyelid animation: motion capture and a linear interpolation. As covered in Section 2.3.7, motion capture can replicate an individual's behaviour with high-fidelity, and thus can be considered physiologically accurate. Linearly interpolated animation implies that a particular motion progresses over a defined number of frames, and that the change in motion is consistent between frames. Therefore, when applied to eyelid animation, linear interpolation does not consider the temporal dynamics of the actual human behaviour being simulated.

Participants rated the three methods of eyelid animation in terms of realism and similarity to a source video of an individual performing a lid saccade or blink. The similarity to source ratings served to verify that the encoded motion captured animations were valid. This source video also provided the data from which the motion captured animations were generated. The primary hypothesis was that model-generated animation would be judged similarly, in terms of realism, to that encoded from motion captured data, and significantly more realistic than linearly interpolated motion. In terms of similarity to the source video, the motion captured animations were expected to be rated most highly, followed by the model-generated animations. For additional materials, see Appendix F.

Technical Preparation

A series of blinks and lid saccades were captured from a normal human subject using a Casio® EXILIM Pro EX-F1 digital camera capable of taking 60 still photographs, at 1920×1080 pixels, per second. During each second of burst-shooting, images of the subject's eyes were captured at 16.7 ms intervals, thus capturing the current state of the eyelids during a blink or saccade at a consistent temporal rate. Sequences of photographs depicting blinks and lid saccades acted as the basis to encode the motion captured eyelid animations in two humanoid avatars. Images of this process, using 3DS Max 2010 [Aut10] and Poser 7 [Sof10b], are illustrated in Figure 7.10.

Three blinks and three lid saccades were encoded and rendered to video at a resolution of 800×600

Figure 7.11: *Experimental interfaces for eyelid model validation.* Left: *Participants ranked the three animation classes in terms of realism.* Right: *Participants ranked the three animation classes in terms of similarity to the source video, positioned in the top-left of the interface.*

pixels at 60 FPS. Likewise, the source image sequences of the real human subject performing the motions were cropped, as illustrated in Figure 7.10, to the same resolution and encoded to video also at 60 FPS. For each of the three blink clips, the vertical angle of the subject's eyes acted as the input parameter to the blink model in order to generate the simulated animations. Similarly, for each of the three lid saccade clips, the vertical shift in eye angle performed by the subject acted as the parametric input to the lid saccade model. The resulting six model-generated animations were rendered to video again at a resolution of 800×600 pixels and 60 FPS. Finally, simple linear animation was also generated for each blink and lid saccade motion using the same starting parameters of vertical angle and gaze shift respectively. Each linear motion was defined to match the number of frames present in the motion captured versions of the animation, but the change in eyelid position was consist between each frame. The linear animations were rendered to video at the same resolution and FPS as the other classes of animation. In summary, a series of videos of blinks and lid saccades were generated which were animated using three different methods: motion capture, the models as described in the Sections 7.2.2 and 7.2.3, and simple linear interpolation.

Following generation of the videos, an experimental interface was developed as shown in Figure 7.11. The interface featured two basic display layouts: one of which was used to record responses of realism, and the other to record responses of similarity to source. The realism display, shown in the left image of Figure 7.11, featured three synchronised videos, one representing each animation condition. The display when judging similarity to source, shown in the right image of Figure 7.11, featured four videos, one from each animation condition, together with the original human video which remained in the top-left of the interface. For each of the three blinks and lid saccade motions, a total of six experimental interface displays were generated, each with different positions of the three animation conditions. This was to negate the influence of placement on the experimental responses, and hence, each motion was viewed and ranked six times.

Experimental Design

The three conditions of eyelid animation were assessed using a within-subject, repeated measures ranking design. Ten members of the Computer Graphics and Virtual Environments group at UCL were

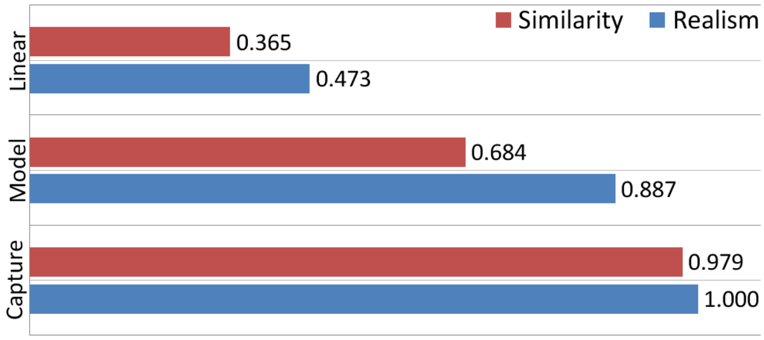


Figure 7.12: *Results of* realism *and* similarity to source *rankings of each animation type normalised to 1.*

recruited to perform the study which lasted for approximately ten minutes. All participants were experienced in 3D animation and visualisation. Participants ranked the animations over two stages. The first stage addressed the perceived realism of the eyelid motion, requiring participants to rank animations in order of what they perceived to be the most human-like movement. The second stage of evaluation required the same participants to rank the videos in terms of similarity to the source video of a human performing a blink or lid saccade motion, positioned in the top-left of the interface. Each of the six motions were viewed and ranked six times, so all three animation conditions appeared in all positions on the interface. The similarity to source task was performed after the realism task in order to avoid confounding the results of the latter (i.e. rating realism based on what was remembered from similarity to the source). Participants performed the ranking task on the same standard 19" display in the lab, with identical environmental luminance conditions.

Results

The validity of the two models, and also of the motion capture encoding, is strongly supported by the results. Figure 7.12 shows combined blink and lid saccade results from the realism and similarity rankings (normalised to 1).

When ranking similarity to the source video, the motion captured animations scored 97.9% of the maximum potential score aggregated from all ten participants. Repeated measures two-way ANOVA calculations revealed high consistency between participants (p=.94), and also that participants were consistent in their own ratings (p=.47). In addition to these rankings, several versions of the animations were presented to various other members of the Computer Graphics and Virtual Environments group during the encoding process for comment and modification. Thus, the faithfulness of the motion capture encoding to the original human movements can be affirmed with some confidence. In comparison, and as expected, the model-generated animations were ranked significantly less similar to the source video than the motion captured versions, at 68.4% similarity to the source. The linear animations were ranked lowest, at 36.5% similarity to the source. Repeated measures two-way ANOVA calculations revealed a statistically significant difference between the three animation classes ($p<.01$), and post-hoc Tukey tests revealed these differences to lie between all pairs ($p<.01$).

In terms of realism, the animation generated using the models received 88.7% of the total votes

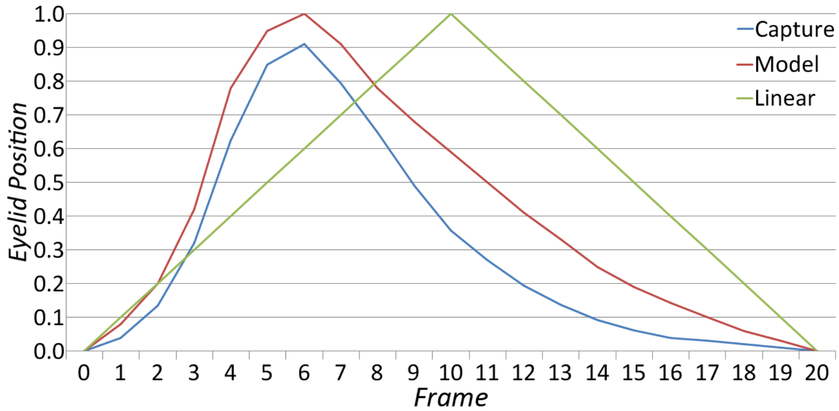


Figure 7.13: *Average blink profile of each animation method. Y=1 indicates full eyelid closure.*

scored by the motion captured animations, while the linear animation score was low, at 47% of the motion capture class. Repeated measures two-way ANOVA calculations revealed a highly significant difference between the realism of the animation types ($p<.01$). Post-hoc Tukey tests showed that the significant differences lay between the linear animation and both other animation classes of model ($p<.01$) and motion capture ($p<.01$). There was no significant difference found between the model-generated animations and those encoded from the motion capture data, promoting the ability of the models to generate realistic motion. Again, a high consistency between participants ($p=.89$) was found, strengthening the subjective votes.

Finally, figure 7.13 presents the dynamics of each animation class during blink motion (averaged from the three captured blinks). The y-axis represents the upper eyelid position, with a value of 1 being completely closed, while the x-axis represents time in frames. It is important to note that, while the graph suggests potential for optimisation of the model, it is not critical for the model's profile to exactly match the motion capture curve. The intention of the models is to present a generalised simulation of the kinematic behaviour of the eyelids, whereas Figure 7.13 (and this analysis) is based on just three blinks captured from a single subject. Hence, while the parametric input to the models may be modified to achieve a more precise fit in this particular instance, the aim of this analysis is to provide general validation of the models compared to alternative methods of animation.

7.2.5 General Impact Experiment

The following experiment aims to evaluate the impact of the two eyelid models on the general realism of an avatar. The models are assessed independently, and also when operating in conjunction, resulting in four comparison conditions: *no eyelid animation* (N); *lid saccade model only* (LS); *blink model only* (B); *lid saccade and blink models* (LSB). The main hypothesis was that, due to the inclusion of both major components of generic human eyelid behaviour, observers would judge an avatar operating in the combined lid saccade and blink (LSB) condition as more realistic than the other behaviour conditions. Regarding the comparison of the independent operation of the lid saccade model (LS) and the blink model (B), it was expected that they would both bestow an avatar with similar degrees of realism, as each model generated a component of human-like behaviour. The realism of avatars exhibiting no eyelid movement (N) was expected to be considered significantly lower than all other conditions.

Technical Preparation

The experimental stimuli comprised videos of an EyeCVE user's embodied avatar performing in two VEs. The capture session was performed in the CAVE lab at UCL. The subject was fitted with head and hand tracking and gaze was driven by the ViewPoint eye tracker. Performance was captured in two VEs: the object-focused puzzle scene investigated in Chapter 5, and a virtual living room furnished with various items. The subject was an experienced user of immersive VR systems, and no training was required prior to capture. During the capture session, the subject navigated and observed their surrounding environment naturally, and in the puzzle task, manipulated objects to solve the puzzle. The behaviour of the subject's embodied avatar was driven by the tracking devices, and included head, hand, and gaze motion. The eyelid models acted similarly to the sample implementation documented in Section 3.1.4, taking gaze as recorded by the eye tracker as input to the lid saccade model. In order to maintain consistent behaviour for evaluative purposes, blinks were generated at a consistent rate of 3.5 seconds (17 per minute), rather than detected and generated by the eye tracker. Bentivoglio et al. define 17 blinks per minute as the average blink rate for a person at rest, which is less than the average rate when engaged in conversation (26 per minute) and more than the average rate when reading (4.5 per minute) [BBC+97].

During the subject's performance, videos of the avatar was captured using 'spectator' clients, similar to as described in Section 6.3.1. The viewpoints of the spectator clients were set to pan and rotate around the avatar's face as the subject performed in the VEs. Fraps [Pty10] was used to record the action at a resolution of 592×384 pixels and 60 FPS in all eyelid animation conditions. Hence, four videos of the avatar, each featuring one of the eyelid animation conditions were captured in both VEs. The resulting videos were then divided into a series of shorter, 15-second clips, which formed the experimental evaluation stimuli. 12 clips in each condition were created. An experimental interface, shown in Figure 7.14, was developed using Adobe Flash [Sys10a].

Experimental Design

A balanced design paired comparison test was conducted to collect ratings of realism of the four eyelid behaviour conditions. The participants were 30 volunteers, who were offered a prize draw incentive to take part in the experiment. The experiment was performed online, and each participant used their own computer, and completed the experiment in their own time. As stated, conditions were N (no eyelid movement), LS (lid saccade model), B (blink model), and LSB (lid saccade and blink model). This resulted in six unique comparison pairs, and 12 when performed in the two VEs (puzzle and observation). The experimental interface shown in Figure 7.14 positions two video clips on the display. Hence, in order to negate any effect of vertical placement, the 12 comparison pairs were doubled to 24, with swapped placement and different video clips of the same condition presented. Participants were instructed to evaluate each of the 24 pairs by choosing the 'perceptually better' one by answering three questions focusing on different aspects of the realism of the avatar's behaviour. As shown on the interface presented in Figure 7.14, the questions were as follows:

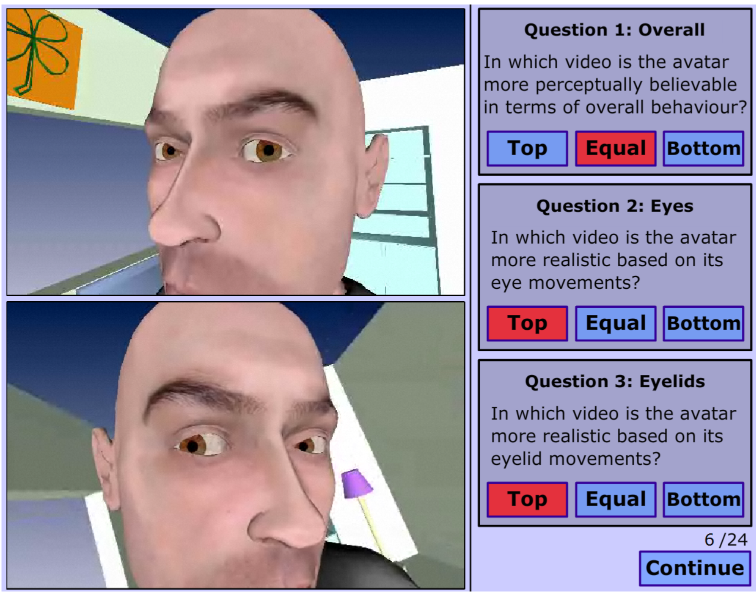


Figure 7.14: *Experimental interface for eyelid model general impact experiment. Top and bottom videos presented the avatar performing in different eyelid conditions. The virtual camera panned and rotated closely around the avatar's face, providing a varying but clear perspective.*

- *Question 1: Overall* - In which video is the avatar more perceptually believable in terms of overall behaviour?
- *Question 2: Eyes* - In which video is the avatar more realistic based on its eye movements?
- *Question 3: Eyelids* - In which video is the avatar more realistic based on its eyelid movements?

Question 3 explicitly concerns eyelid animation, while Questions 1 and 2 pertain to overall realism and eye realism respectively. A prediction informed by the results the experiment presented in Section 7.1 was that the improved fidelity of the avatar's eyelid animation would also lead to enhanced ratings of overall realism, and also of the eyes. For each question, the participant used the experimental interface to select *Top* if they judged the upper avatar's behaviour favourably, *Bottom* if they preferred the lower avatar, and *Equal* if they considered both to be similar. For additional materials, see Appendix F.

Analysis

Table 7.5 shows the computed preference matrix based on the voting results of the 30 participants. The number in each cell denotes the frequency-count of a specific method chosen as the perceptually better one for each of the three questions. 1 point was given for a clear *Top* or *Bottom* choice, while 0.5 was given to each condition in the case of an *Equal* decision. For instance, "89" in the first cell of the final row indicates that LSB was selected as the perceptually better condition, in terms of eyelid realism answered in Question 3, a total of 89 times when compared with N. Correspondingly, the fourth cell on the third row indicates that N was chosen a total of 31 times when compared with LSB for the same question.

Prior to performing vote comparisons for the four experimental conditions, two statistical tests were performed. Kendall and Smith's coefficient of consistency [KS40] for each participant in order to determine whether there was any intransitive vote, and an overall agreement test determining whether the

Table 7.5: *Computed preference matrix of all data for both VEs. One point was given for a clear preference, while 0.5 was given to each condition in the case of an equal rating. Q# refers to Question 1, 2, or 3.*

	Q#	N	LS	B	LSB	Total
N	1	-	48	38	38	124
	2	-	54.5	47.5	41	143
	3	-	37	33.5	31	101.5
LS	1	72	-	66	52	190
	2	65.5	-	59.5	49	174
	3	83	-	53.5	45	181.5
B	1	82	54	-	39	175
	2	72.5	60.5	-	51	184
	3	86.5	66.5	-	40.5	193.5
LSB	1	82	68	81	-	231
	2	79	71	69	-	219
	3	89	75	79.5	-	243.5

Table 7.6: *Comparisons of consistency (ζ) and agreement (μ) test statistics. Chi square (χ^2) and related **p** values given 6df and condition ranks for each question.*

Q#	ζ	μ	χ^2	*p*, 6 *d.f.*	1st	2nd	3rd	4th
1	0.02	0.07	55.10	<.01	LSB	LS	B	N
2	0.05	0.02	19.07	<.01	LSB	B	LS	N
3	0.21	0.11	84.73	<.01	LSB	B	LS	N

participants voted for all pairs similarly. Table 7.6 shows the averaged coefficient of consistency (ζ) for all participants, coefficient of agreement (μ) and corresponding chi square (χ^2) values with significance calculations (p) given the six degrees of freedom.

The significance values indicate that a statistically high variation was found between the conditions, but fails to identify where these significances lie. Figure 7.15 presents results from multiple comparison score tests between the six conditional pairs for the three combined questions. Conditions are ranked from lowest to highest (left to right) as established in Table 7.6. Any conditions that are underlined by the same line may be considered statistically identical given a threshold of (p=.05). In this case, conditions LS and B are underlined by the same line, echoing their similar ratings presented in Tables 7.5 and 7.6. Hence, there was a significant difference found between all other paired conditions.

Figure 7.16 presents a radial plot graph, pairing all conditions in order to highlight their relative differences. The graph is normalised to 1. The three questions are presented independently and also as a combined plot. The superiority of the combined lid saccade and blink model (LSB) condition is apparent, as well as the insignificant differences between between the independently-operating lid saccade (LS) and blink (B) models. No eyelid animation (N) is seen to consistently perform poorest.

7.2.6 Discussion

Previous work animating the eyelids of virtual humanoids has failed to formalise the define the kinematic behaviour of lid saccades and blinks, typically solving the problem through use of facial motion capture or crude approximation. The current section presented parametric models of lid saccades and blinks based on physiological data. The models are suitable for implementation in both real-time and

N	LS	B	LSB
368.5	545.5	552.3	693.5

Figure 7.15: *Eyelid models general impact experiment multiple comparison scores. Conditions underlined by the same line are statistically similar.*

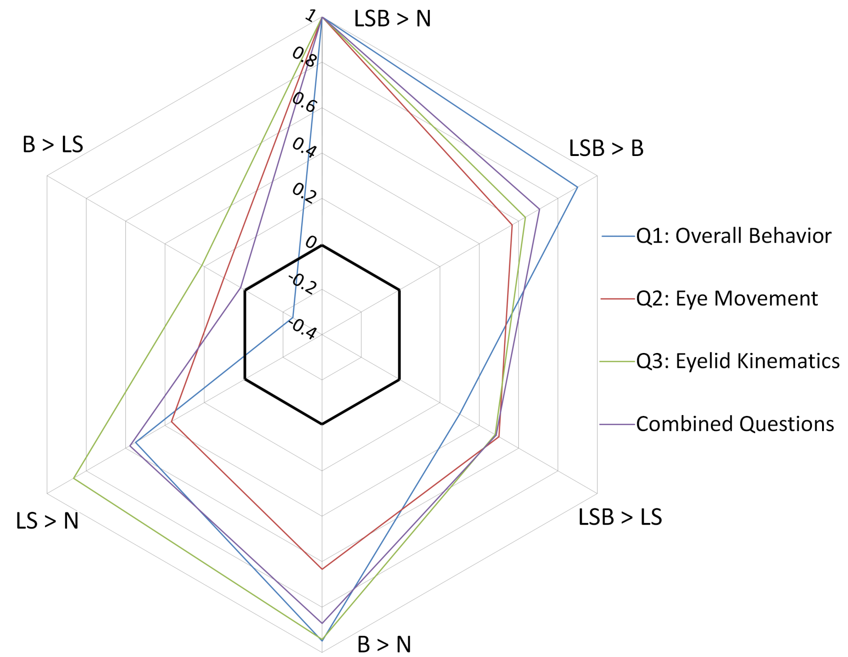


Figure 7.16: *Eyelid model general impact experiment radial plot comparing the six pairs of conditions. Each question and combined questions with relative ratings are normalised to 1. Note that the bold line around the axis indicates 0.*

offline animation systems. Animation generated by the models was experimentally validated against motion captured data and linear interpolation. Results supported the original hypothesis, indicating that the models performed statistically similar to motion captured animation in terms of realism. The validity of the motion captured animation was confirmed by the significantly higher ratings of similarity to the original source videos from which the motion capture data was encoded. In comparison, linearly interpolated animation performed poorly in both measures.

A subsequent experiment was then performed, assessing the general impact of the models on the overall realism of an avatar. The results supported the experimental hypothesis, indicating that the combined operation of the lid saccade and blink models can enhance the perceived realism of an avatar significantly above that attained when using either model independently. Also, independent operation of either model was rated as significantly more realistic than no eyelid animation. Comparing the independent operation of the two models, the avatar operating with the blink model only was rated slightly, but insignificantly, higher than the same avatar operating with the lid saccade model only. This difference is likely to be due to observers' expectancy for humans to blink regularly, and that a lack of blinking is perhaps more conspicuous than an absence of lid saccades, which, depending on gaze behaviour, may be subtle. The final section of this chapter presents a saliency-based approach to gaze modelling.

7.3 Saliency-Based Gaze Model

When immersed in a VE, salient stimuli such as the changing location and orientation of objects and other avatars attract the visual attention of users. This provides a method of driving the gaze behaviour of avatars, and is explored in this section. Content appearing in a VE has intrinsic saliency in terms of characteristics such as proximity, eccentricity, orientation and velocity. Proximity relates to the Euclidean distance between the position of an embodied user's viewpoint and an object, allocating more attention to near objects than those further away. Eccentricity is based on the angular distance of objects from the centrer of the a user's FOV (the head-centric vector), thus allocating more attention to object closer to the centre of the FOV than those in the periphery. Velocity is based on an object's speed across a user's visual field, allocating more attention to fast-moving objects than to slow-moving or still objects. Orientation is based on objects' rotation behaviour, with more attention being allocated to objects with high rotation speed. Extrinsic saliency may also be bestowed on an object by another user's interest in that object, which may be signalled by gazing or interacting with it.

This section presents a model for automatic gaze generation based on the saliency of objects that lie within an embodied user's FOV when performing in a VE. As documented in Section 2.3.6, previous algorithms generating gaze for use with virtual characters have typically been based upon observations in the social science literature and on eye tracking data collected during performance of the specific application scenario intended for the gaze model [PLBB02]. It may be argued that such gaze models suffer from three drawbacks: assumptions are made about gaze patterns that relate to certain social signals and cognitive states with limited understanding of how they fit into a temporal dimension; detecting cognitive state adds another layer of input which is not readily available in VE systems; and the focus of these gaze models has been on general realism of gaze, while target relevancy has been largely ignored. In contrast, the approach to gaze modelling presented in this section takes a user's head position and orientation as input to gaze calculation, and distributes gaze between targets based on their saliency. The model is based on data captured from the object-focused experiment presented in Chapter 5, but aims to generate gaze that is not specific to a particular social or task-based scenario. Firstly, this section presents the gaze model algorithmically. An experiment follows, comparing animation generated by the model with alternative methods of gaze control: tracked gaze, the random model used in Chapter 5's object-focused experiment, and static gaze, featuring still centred eyes.

7.3.1 Gaze Model

The saliency model is designed to adapt to complex interaction unfolding within a virtual scene. It considers both the varying behaviour of the controlling user, and properties of objects within the scene. Saliency computation is based on the functions generated in the analysis of the random gaze model presented in Section 5.3.2. The distribution of saliency values and objects of interest are often spatially biased towards the centre of an observer's viewpoint [MK01]. In order to decrease the high probability of centre bias in the saliency model, the random model takes control of gaze 25% of the time, while the saliency model computes the target object 75% of the times. To implement this, a uniformly distributed random integer between 0 and 3 is generated and the random model takes control whenever the number

Algorithm 2 *saliency model* computes target object (o_x, o_y, o_z)

Require: scene database of objects $\{O_1, O_2, O_3, \ldots O_n\}$
Require: FOV, $fov = 70°$ (i.e. eccentricity, $\theta \leq 35°$)
1: **for** each frame **do**
2: include phantom object for head-centric vector in scene database
3: compute elapsed time since previous frame Δt
4: seed random to 1 second {determines saliency state and fixation duration}
5: compute avatar's eye location in world coordinates (e_x, e_y, e_z)
6: **for** each object in the scene database **do**
7: determine object's location and orientation in world coordinates
8: compute eccentricity, θ {equation 7.2}
9: compute vertical angle, θ_v
10: **if** $(\theta < 35°)$ and $(-25° < \theta_v < 25°)$ **then**
11: compute saliency scores for:
12: - change in orientation, Δq {equation 7.4}
13: - object's velocity v {equation 7.3}
14: - proximity p {equation 7.1}
15: - eccentricity θ {equation 7.2}
16: compute total saliency score and store in list A
17: add to list of objects within field of view, list B
18: **end if**
19: **end for**
20: **if** $(\Delta tc < 300ms)$ and (current target object is still within FOV) **then**
21: return current target object's location
22: **end if**
23: **if** saliency state = true **then**
24: determine object with highest saliency from list A
25: **else** {considers less salient objects (in the periphery)}
26: pick randomly from objects within field of view from list B
27: **end if**
28: aim avatar's eyes at centre of selected target object
29: **if** previous target $\neq$ selected target **then**
30: compute elapsed time since last target changed, Δtc
31: compute eyeball interpolation {Wquation 7.6}
32: **end if**
33: **end for**

3 is generated (i.e. when saliency state equals false as described on line 23–27 of Algorithm 2. The probability that the saliency algorithm is used is thus given by: $P(salience) = 3/4$.

Spatial and Temporal Distribution of Fixations

The main input to the model is the scene database, which stores all the objects within the scene. Algorithm 2 determines the target object by examining the intrinsic saliences of objects within the current FOV, defined as (35° from the head-centric vector).

Intrinsic Saliency Criteria

The saliency gaze model generates gaze target based on four intrinsic saliency of the objects:

1. Given the user's eye, $E = (e_x, e_y, e_z)$, and the object, $O_i = (o_x, o_y, o_z)$, the *proximity, p* is computed from the Euclidean distance between the two 3D points as:

$$p = \sqrt{(e_x - o_x)^2 + (e_y - o_y)^2 + (e_z - o_z)^2}, \tag{7.1}$$

2. *Eccentricity*, θ, defined as the magnitude of the dot product is computed as:

$$\theta = \arccos\left(\frac{u \cdot v}{|u||v|}\right), \tag{7.2}$$

where $u = (u_x, u_y, u_z)$ is the head-centric vector and $v = (v_x, v_y, v_z)$ is the direction vector of the eye to the object, $(e_x, e_y, e_z) - (o_x, o_y, o_z)$.

3. *Velocity*, v, is defined as the rate of change of the object's location and is computed as:

$$v = \frac{\Delta O_i}{\Delta t}, \tag{7.3}$$

where ΔO_i is the Euclidean distance between an object's location at time t_1 and its location at time t_2, and Δt is the time interval of the frame duration. The normalised saliency score, S_v of the object's velocity is given by $v/20$ (i.e. a reasonable maximum speed of six metres per second).

4. *Orientation*, Δq, defined as the change in object's angular position over time and is computed as:

$$\Delta q = 2 \arccos({q_1}^{-1}.q_2) \tag{7.4}$$

where quaternions q_1 and q_2 represent two orientations at time t_1 and t_2 respectively. The normalised saliency score, $S_{\Delta q}$ of the object's orientation is given by $\Delta q/180$ (i.e. a reasonable maximum change in orientation of 180°).

Saliency Scoring and Fixation Duration

The saliency of each object within the field of view is computed from a summation of the normalised saliency scores and is used to guide attention.

$$S_O = S_\theta + S_p + S_v + S_{\Delta q}, \tag{7.5}$$

The object with the highest combined saliency score is determined as the target. The computation of these scores relies on appropriate normalisation and summation steps in a competitive manner to determine most likely target object to be allocated fixations. The fixation duration is limited to 300 ms as long as the target object remains within the FOV. This duration is informed by Henderson's average fixation duration during scene viewing [HH99].

7.3.2 Eyeball Dynamics

The eyeball is interpolated over six frames by fitting to an exponential velocity curve as presented in Lee et al. [PLBB02], and subsequently, Vinayagamoorthy et al. [VGSS04]:

$$y = 14e^{[-\pi/4(x-3)^2]}, \tag{7.6}$$

where $x = frame\{1,2,3,4,5,6\}$. The eye is moved to the intermediate positions within each frame to produce smooth movement during saccadic gaze shifts.

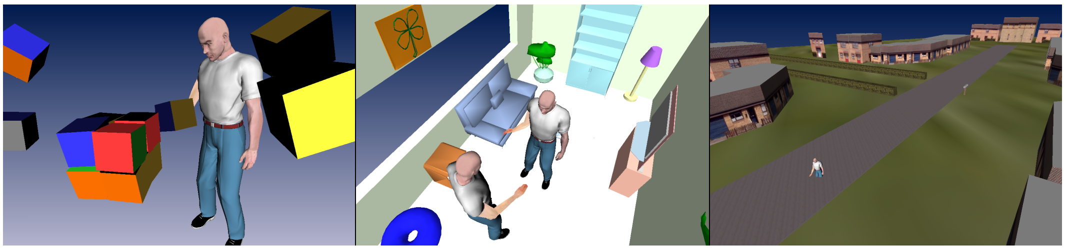


Figure 7.17: *VEs used for the gaze model experiment: object-focused puzzle (left), conversation with another user (center), and navigating through a large town environment.*

7.3.3 Experiment

The following experiment aims to compare the gaze model with three other methods of gaze control. The experimental method was identical to that employed in Section 7.2.5, and assessed perceived realism of an avatar across the following four gaze control conditions: no gaze movement (N), random gaze model (R), saliency gaze model (S), and tracked gaze (T). The main hypothesis was that observers would judge an avatar operating with tracked gaze as more realistic than all other gaze conditions. However, the saliency gaze model was expected to approach ratings of tracked gaze. The other classes of random and static gaze were expected to perform poorest.

Technical Preparation

The technical preparation followed the procedure adopted in the eyelid model experiment presented in Section 7.2.5. The performance of an expert user in EyeCVE was captured in UCL's CAVE system. In order to assess the ability of the saliency model to generate natural and meaningful gaze, it was important to measure performance in varying VE scenarios. Performance of a subject in three situations relevant to AMC were captured: the object-focused puzzle scene investigated in Chapter 5, a two-party conversation taking place in a virtual living room furnished with various items, and navigating through a large-scale town scene. The VEs are shown in Figure 7.17. During the object-focused puzzle scenario and the navigation scene, the system operated in single-user mode, while the subject conversed with another expert user located in the WALL display at UCL during the conversational scenario. The subject performed each scene under the four gaze control conditions, and engaged in the puzzle, conversation, or navigation in a relaxed and natural manner. The behaviour of the avatar embodiment was driven by head and hand tracking devices. Gaze was controlled by the ViewPoint eye tracker in the tracked gaze condition, was simulated by the random model or saliency model in the two model conditions, and remained static in the final condition. In the conversational VE, mouth-movement of both avatars was animated by a speech detector as detailed in Section 3.1.4. Finally, in all conditions, the eyelid models presented in the previous section in this chapter were used to generate blinks (17 per minute) and lid saccades based on gaze movement.

EyeCVE spectator clients were used to capture the subject's performance in the four gaze conditions from a close-up rotating viewpoint around the avatar's face. Fraps [Pty10] was used to record the action at a resolution of 592×384 pixels and at 60 FPS. Hence, four videos of the avatar, each featuring one of the gaze behaviour conditions, were captured in the three VEs. The resulting videos were then divided

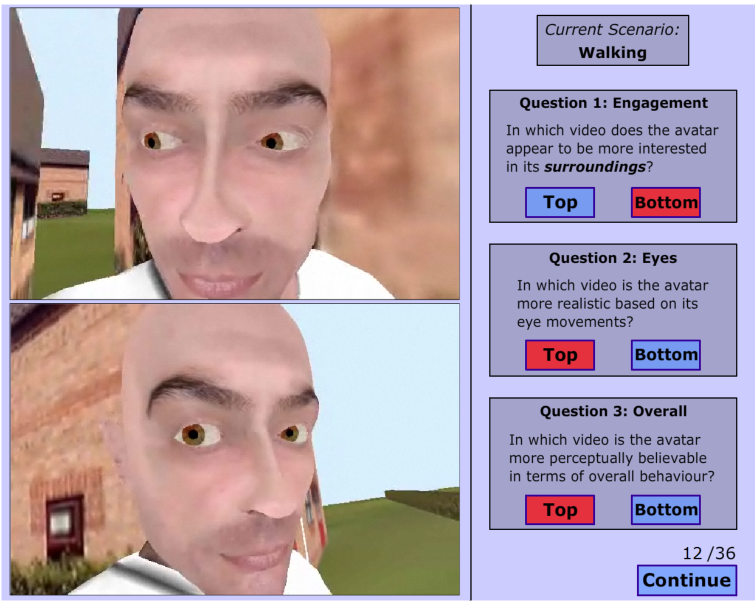


Figure 7.18: *Gaze model experimental interface. Top and bottom videos presented the avatar performing in different gaze conditions. The virtual camera panned and rotated closely around the avatar's face, providing a varying but clear perspective.*

into a series of shorter, 15-second clips, which would form the experimental evaluation stimuli. 18 clips in each condition were created. The experimental interface developed for the eyelid model was adapted for the current experiment, and is shown in Figure 7.18.

Experimental Design

Similarly to the eyelid model experiment in Section sec:eyelidgenexp, a balanced design paired comparison test was conducted to collect ratings of realism of the four gaze behaviour conditions. The participants were 70 volunteers, who were offered a prize draw incentive to take part in the experiment. The experiment was performed online, and each participant used their own computer, and completed the experiment in their own time. As stated, conditions were N (non-moving static eyes), R (random gaze model), S (saliency gaze model), and T (tracked gaze). This resulted in six unique comparison pairs, and 18 over the three VEs. To negate any effect of vertical placement, the 18 comparison pairs were doubled to 36 with swapped placement and different video clips of the same condition. Participants were instructed to evaluate each of the 36 pairs by choosing the 'perceptually better' one by answering three questions focusing on different aspects of the realism of the avatar's behaviour. As shown on the interface, the questions were as follows:

- *Question 1: Engagement* - In which video does the avatar appear to be more (engaged in the puzzle / engaged in the conversation / interested in its surroundings)?
- *Question 2: Eyes* - In which video is the avatar more realistic based on its eye movements?
- *Question 3: Overall* - In which video is the avatar more perceptually believable in terms of overall behaviour?

The three questions were designed accordingly to extract information regarding how involved in the particular scenario (object-focused puzzle, conversation, or navigation) the avatar appeared to be, the natural quality of eye motion, and the overall realism of the avatar. It was important to design the experiment to generate data for analysis which maintained the context of the three VE scenarios. Therefore, while always eliciting metrics of engagement, Question 1's phrasing varied slightly between three versions according to the VE scenario currently under inspection. This framing of Question 1 also served to contextualise the following two questions, which focused on the realism of the eye motion, and the overall perception of the avatar's believability. For each question, the participant used the experimental interface to select *Top* if they judged the upper avatar's behaviour favourably, or *Bottom* if they preferred the lower avatar. For additional materials, see Appendix F.

7.3.4 Analysis

A preference matrices for each VE scenario are shown in Tables 7.7(a), relating to the puzzle scene, 7.7(b) for conversation scene, and 7.7(c) for navigation scene. The number in each cell denotes the selection frequency of a specific method when answering one of the three questions, with 1 point given for a *Top* or *Bottom* choice. For instance, in Table 7.7(a), '110' in the first cell of the final row indicates that condition T was selected as the perceptually better condition a total of 110 times when compared with condition N and when answering Question 3. Correspondingly, the fourth cell on the first row indicates that condition N was selected as the perceptually better condition a total of 30 times when comparing the same conditions and when answering the same question. Note that the totals for each such corresponding condition pair sum to 140 (70 participants and two ratings with reversed vertical position).

Prior to performing vote comparisons for the four experimental conditions, Kendall and Smith's coefficient of consistency [KS40] for each participant was performed, together with an overall agreement test determining whether the participants voted for all pairs similarly. Table 7.8 shows the averaged coefficient of consistency (ζ) for all participants, coefficient of agreement (μ), and corresponding chi square (χ^2) values with significance values (p) given the six degrees of freedom.

A main effect ($p<0.001$) was found between gaze conditions. However, this does not determine between which conditions the significances lie. Figure 7.19 presents results from multiple comparison score tests between the six conditional pairs for the three combined questions. Conditions are ranked from lowest to highest (left to right) as established in Table 7.8. Any two conditions that are underlined by the same line may be considered statistically identical given a threshold of (p=.05). For instance, the line connecting S and T in Q3 of the conversation scene indicates that there were no significant differences in terms of overall believability between the saliency model and tracked gaze.

Figure 7.20 presents a radial plot graph, pairing all conditions in order to highlight their relative differences. The graph is normalised to 1. The three questions are presented independently and also as a combined plot.

Gaze data comparing tracked gaze, the saliency gaze model, and the random model is plotted in Figure 7.21, enabling correlation between conditions. The plots show the frequencies of five gaze param-

Table 7.7: *Computed preference matrix for:*

(a) *object-focused puzzle* VE.

	Q#	N	R	S	T	Total
N	1	-	53	43	35	131
	2	-	44	48	31	123
	3	-	58	51	30	139
R	1	87	-	58	59	204
	2	96	-	30	39	165
	3	82	-	35	34	151
S	1	97	82	-	67	246
	2	92	110	-	54	256
	3	89	105	-	56	250
T	1	105	81	73	-	259
	2	109	101	86	-	296
	3	110	106	84	-	300

(b) *conversation* VE.

	Q#	N	R	S	T	Total
N	1	-	63	57	45	165
	2	-	58	46	32	136
	3	-	59	47	39	145
R	1	82	-	47	66	195
	2	87	-	46	58	191
	3	86	-	44	55	185
S	1	88	98	-	75	261
	2	99	99	-	63	261
	3	99	101	-	67	267
T	1	99	79	70	-	248
	2	112	87	83	-	282
	3	105	90	79	-	274

(c) *navigation* VE.

	Q#	N	R	S	T	Total
N	1	-	24	21	19	64
	2	-	41	39	23	103
	3	-	39	36	26	101
R	1	116	-	54	75	245
	2	99	-	60	29	188
	3	101	-	57	34	192
S	1	119	86	-	84	289
	2	101	80	-	29	210
	3	104	83	-	30	217
T	1	121	65	56	-	242
	2	117	111	111	-	339
	3	114	106	110	-	330

eters (proximity, saccade magnitude, saccade velocity, fixation duration and the eccentricity) generated in each gaze condition during the data capture. Generally, the saliency model produces plots that are reasonably correlated with those of tracked gaze, while plots for the random model differed greatly to tracked gaze.

7.3.5 Discussion

The saliency model presented in the current section takes a novel approach to gaze computation by estimating interactions between dynamic objects, rather than attempting to develop a model for specific environments and tasks. While this approach does not consider explicit interactional states (such as speaking and listening) it has the virtue of a straightforward implementation that can be applied to most VE systems comprising a scene database. The experimental results support the hypothesis, which stated that avatars operating with tracked gaze would be rated as superior those exhibiting to other conditions of gaze control. Additionally, the saliency gaze model was expected to generate gaze that would be rated significantly better than both static and random model gaze. Average rankings across scenes and questions indicate the superiority of tracked gaze, followed next by the saliency model, then the random model, and lastly by static gaze. These results are supportive of the saliency model's approach to scene analysis towards realistic gaze generation.

Overall, the plots shown in Figure 7.21 demonstrate reasonable correlation between the actual

Table 7.8: *Comparisons of consistency (ζ) and agreement (μ) test statistics. Chi square (χ^2) and related* ***p*** *values given 6df. Ranking of conditions for each question and scene also shown.*

S#,Q#	ζ	μ	χ^2	*p*, 6 *d.f.*	1st	2nd	3rd	4th
1,1	0.380	0.079	71.9	<0.001	T	S	R	N
1,2	0.393	0.181	157.1	<0.001	T	S	R	N
1,3	0.384	0.158	137.8	<0.001	T	S	R	N
2,1	0.320	0.123	108.8	<0.001	S	T	R	N
2,2	0.386	0.186	161.4	<0.001	T	S	R	N
2,3	0.393	0.174	151.3	<0.001	T	S	R	N
3,1	0.493	0.253	217.0	<0.001	S	R	T	N
3,2	0.520	0.249	213.5	<0.001	T	S	R	N
3,3	0.459	0.237	203.4	<0.001	T	S	R	N

Puzzle Scene

Q1				*Q2*				*Q3*			
N	R	S	T	N	R	S	T	N	R	S	T
131	204	246	259	123	165	256	296	139	151	250	300

Conversation Scene

Q1				*Q2*				*Q3*			
N	R	T	S	N	R	S	T	N	R	S	T
165	195	248	261	136	191	261	282	145	185	267	274

Navigation Scene

Q1				*Q2*				*Q3*			
N	T	R	S	N	R	S	T	N	R	S	T
64	242	245	289	103	188	210	339	101	192	217	330

Figure 7.19: *Multiple comparison score for all data. Any conditions whose scores are underlined are considered statistically similar.*

tracked gaze and that generated by the saliency model. In particular, properties of velocity and magnitude have a $\rho_{T,S}$ value ranging between 0.987 to 0.999. However, the plots for view proximity indicate that the random model generated gaze more similar to tracked gaze in this regard. The saliency model performed particularly poorly in the navigation scene, generating erratic and unrealistic gaze behaviour. This particular scene featured a large number objects dispersed over a wide area, indicating that the model would benefit from adapting input parameters based on pre-processing the properties of a given scene.

Finally, the author's position, as detailed in Section 2.4, arguing that behavioural simulation in AMC is problematic should be restated. An individual's gaze behaviour cannot be determined or predicted by any practical means, including by the individual themselves in the unfolding interaction. Therefore, gaze models (and other algorithms generating autonomous behaviour) should not aspire or claim to faithfully replicate an individual's behaviour. Rather, their purpose is to act both as general indicators of an embodied user's attention, and to bestow an avatar with general characteristics of liveliness; particularly when tracking may be unavailable.

7.4 Chapter Summary

This chapter presented three sections investigating the impact of varying oculesic behaviours of avatars and agents, together with associated experiments. The first section explored the influence that eyelid animation and vergence of an autonomous agent had on accuracy of gaze direction identification and

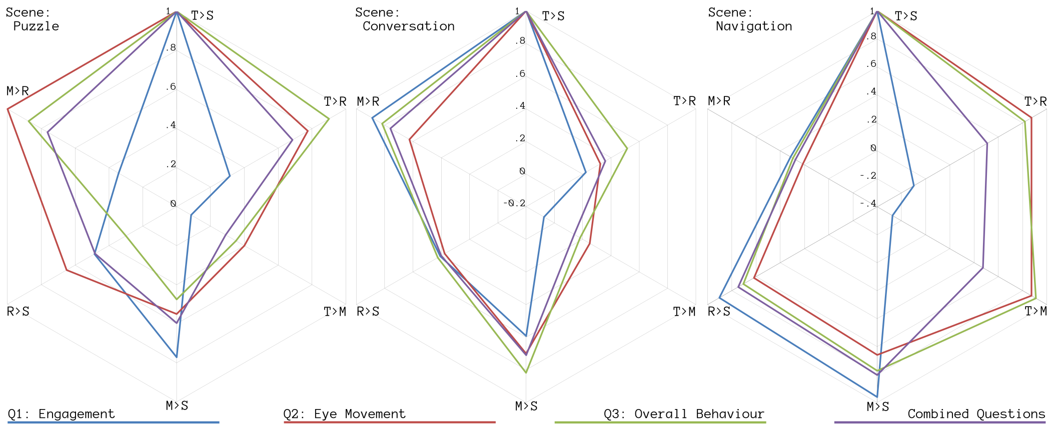


Figure 7.20: *Radial plot comparing the six pairs of conditions with relative ratings normalized to 1 for each question and combined questions. Note that the bold line around the axis indicates 0.*

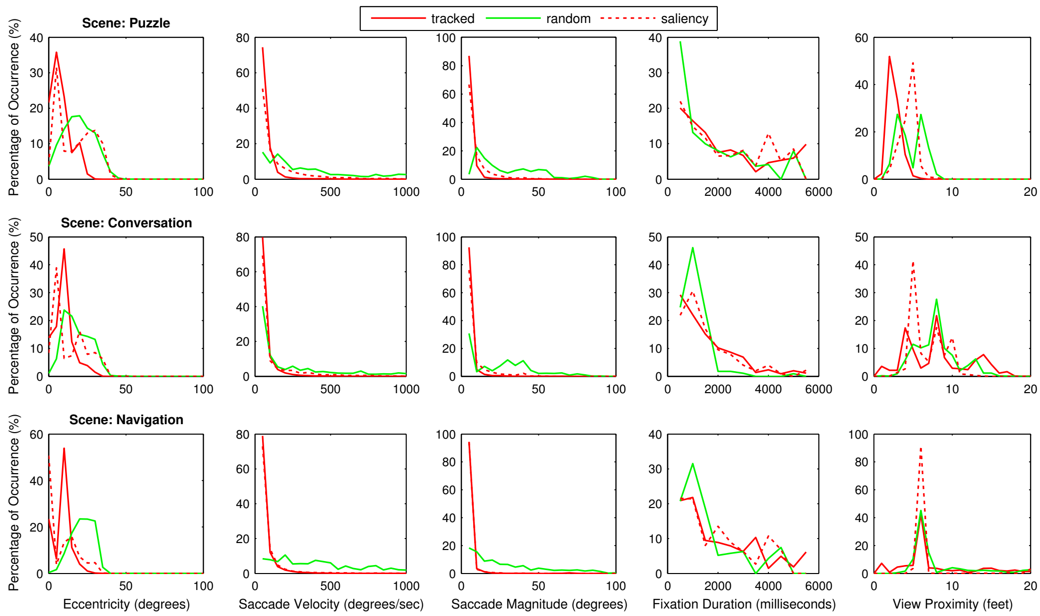


Figure 7.21: *Comparisons of gaze observed in tracked, saliency model, and random model conditions for each scene.*

perceived realism. Eyelid animation was found to significantly increase the realism of the agent, but in some situations, hindered accuracy of gaze direction identification. This was contradictory to participants' self-performance judgements, which were higher when assessing the agent operating with eyelid animation. Vergence was not found to be a significant factor influencing accuracy of gaze direction identification or realism in the particular experimental setup. However, it should be noted that the experiment featured an agent gazing at targets which varied only slightly in depth, and viewed was from static camera angles on a non-immersive display.

The second section formalised the eyelid animation featured in the first experiment, presenting two parametric models generating physiologically-accurate movement of the two major components of human eyelid kinematics. The lid saccade model generates realistic motion based on vertical shifts in a given gaze gaze signal, while the blink model animates realistic blink motion given a current eyelid position. The quality of the models was validated by comparing against motion captured data, and observers judged the realism of the simulated and captured animations as statistically identical. The general impact of the models on the overall realism of an avatar was then assessed. Data was recorded in EyeCVE, using the ViewPoint eye tracker to capture gaze. This gaze stream acted as input to the lid saccade model, which was interrupted as blink signals were generated. Participants judged animations of an avatar exhibiting varying eyelid animation conditions, with results promoting the combined use of the blink and lid saccade models to enhance the avatar's realism.

The third section focused on gaze modelling, approaching the problem of generating realistic and meaningful gaze behaviour by assigning saliency to objects within a user's current FOV and distributing gaze accordingly. The design and development of the model sought to determine characteristics of objects which attract attention in VEs, and how they could be weighted to drive an avatar's gaze. Properties of proximity, eccentricity, orientation, and velocity were used to calculate the saliency of each object in a user's FOV. A model of eyeball dynamics was also implemented. The ability of the saliency model to generate realistic gaze was then assessed in an experiment which compared tracked, random, static, and saliency model gaze. Experimental results and graphical analysis of gaze behaviour demonstrated a promising gaze model that is able to generate realistic and relevant gaze in varying tasks, and does not require prior knowledge of the VE's content or scenario. However, tracked gaze was ranked as the superior method of avatar gaze control.

Chapter 8

Conclusions

Telecommunication is increasingly being carried out in multi-user VEs, in which users are embodied by avatars. Millions of people work, play, and socialise for large amounts of time in online virtual worlds such as Linden Lab's Second Life [Res10b] and massively multiplayer online games such as Blizzard Entertainment's World of Warcraft [Ent10a]. World of Warcraft alone boasts over 12 million paying subscribers [Ent10b]. Currently, the vast majority of users enjoying this emerging form of visual remote interaction, are interfacing via consumer hardware, with standard displays and input devices. In comparison with the technical apparatus employed in the experimental work presented in this thesis, the experience offered by this mode of AMC is primitive: both in terms of degree of immersion, and capture and transmission of nonverbal behaviour. These two characteristics have been identified in the VE literature as critical requirements for the support of high-quality AMC, increasing presence [Sla09], and copresence [Gar03] respectively. However, along with the proliferation of AMC, such technology for home and commercial use is also undergoing rapid development, with a variety of motion tracking interfaces and stereoscopic displays on the horizon. For instance, Microsoft's Kinect (formerly Project Natal) [Cor10b] is a natural user interface device which promises to enable real-time skeletal mapping, gesture recognition, facial tracking, and voice recognition. Coupled with a high definition, large-format, stereoscopic display, it is certain that such technologies will enable sophisticated avatar communication systems to be built and deployed over the coming years.

The work in this thesis explored how AMC in state-of-art ICVE systems may be enhanced and understood by the use of eye tracking. As argued in Section 2.4, humans habitually use gaze practices which are consequential for the interactions in which they occur, yet which cannot be predicted or inferred from any other observable behavioural state, including talk. Thus, if AMC aspires to support such natural interactions, the possibility of simulating avatar gaze is precluded. Consequently, tracking and replication of oculesic cues was established as the option best suited to preserving the original communicative intent of the interactants. Chapter 3 documented development and evaluation of EyeCVE. The system uses eye tracking to drive avatar oculesics, including gaze, blinking, and pupil dilation, which are transmitted and displayed to fellow interactants in real-time.

Chapters 4–6 presented a series of telecommunication experiments. Each chapter focused on a common and specific collaborative scenario. Chapter 4 explored conversation, Chapter 5 investigated object-focused interaction, and Chapter 6 examined truth and deception. The overarching goal of this experimental work was to explore how eye tracking may be applied in AMC, both interactively to enhance quality of communication by enriching avatars' nonverbal expressiveness, and analytically to explore how users behave and respond in AMC, with particular regard to social presence in comparison with VMC. Finally, Chapter 7 presented models of oculesic behaviour, together with associated experiments. This work investigated behavioural and representational characteristics of virtual humanoids likely to influence AMC. Over the course of the experimental work presented in this thesis, eye tracking was used as the primary method of data collection. However, other methods of subjective and objective data collection were also employed, including head and hand tracking, speech tracking, performance metrics, questionnaires, and interviews.

This closing chapter summarises the work presented in this thesis. Firstly, the findings of each experiment are recounted, followed by the holistic conclusions, relating back to the research questions and contributions established in Chapter 1. Finally, potential directions for future work are established.

8.1 Conversation Experiment

Figure 8.1: *Revisiting the conversational experiment in AMC* (left and centre) *and VMC* (right).

The first telecommunication experiment, investigating triadic conversation in tracked gaze AMC and gaze aware VMC, was reported in Chapter 4. In the VMC system, gaze awareness was realised by careful alignment of video displays and camera positions, while AMC was supported by an early version of EyeCVE, driving avatar gaze. The goal of the experiment was to compare users' behaviour in the two visual communication mediums, with a particular focus on the operational importance of gaze. A semi-structured interview scenario was formulated, featuring two confederates interviewing a single participant. In both mediums, gaze data recorded from eye tracking was combined with CA as the primary means of evaluation and analysis of the interactions. To support this qualitative analysis, participants' gaze behaviour during the critical time periods in which confederates were issuing questions was also measured in both AMC and VMC.

Findings indicated that multiparty conversation in both AMC and VMC was able to be conducted, and that gaze behaviour was similar in both mediums. CA demonstrated that interaction in both systems was influenced by the technical characteristics and limitations of each medium. In particular, the fragmented workspace presented by VMC, and the relative paucity of avatar expression during AMC, presented problems with regards to turn-taking and conversation management. However, these issues were not critical, and gaze behaviour in both AMC and VMC followed established social norms observed in collocated interaction. This experiment was less formal, both in design and analysis, and more exploratory in terms of the general issues regarding gaze-enabled telecommunication systems than the experiments presented in the subsequent chapters. In particular, the lack of a condition in which other users' gaze information was made unavailable precluded the ability to explore the interactive benefit of eye tracking in AMC.

8.2 Object-Focused Experiment

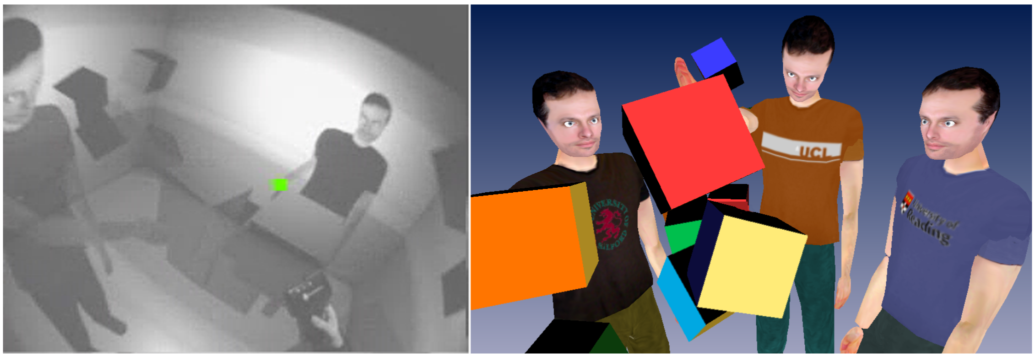


Figure 8.2: *Revisiting the object-focused experiment, investigating varying methods of avatar gaze control. View of the puzzle task from the eye tracker scene camera* (left)*, and corresponding virtual capture* (right).

The second telecommunication experiment, reported in Chapter 5, sought to investigate the impact of varying methods of avatar gaze control on quality of communication during object-focused multiparty AMC in ICVE systems. An experimental puzzle scenario was devised to emphasise the operational importance of gaze during the collaborative interactions. The task was based on instructions, that were issued by two confederates, and performed by a single participant, towards the puzzle's eventual goal state. The experiment compared three forms of AMC: tracked gaze, static gaze, and a randomised gaze model. Data was collected from multiple sources, including eye tracker videos and log files, EyeCVE log files, gaze model log files, and subjective questionnaires. This data was analysed as metrics of task performance, subjective user experience, and interaction analysis.

The main finding of the experiment was that quality of communication, measured by task performance and interaction analysis, was reduced by avatars demonstrating misleading gaze behaviour. When compared to both tracked and static gaze AMC, modelled gaze AMC resulted in more mistakes and slower action during the object-focused collaboration. Measures of subjective user experience were generally high, and did not show a significant difference between gaze conditions, thus promoting AMC in ICVEs as a platform able to support high-quality spatial interaction.

8.3 Truth and Deception Experiments

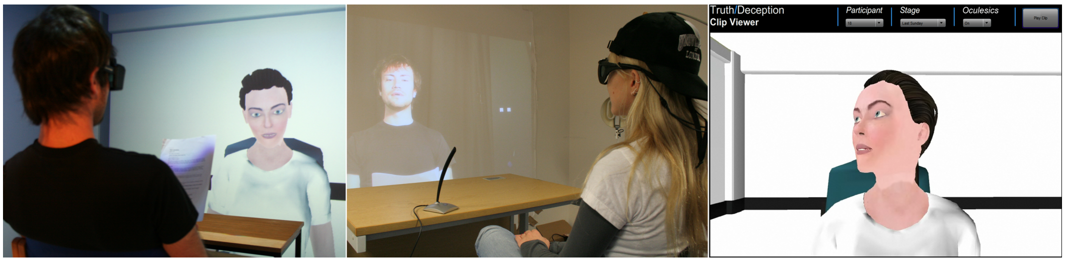


Figure 8.3: *Revisiting the truth and deception experiments. The first experiment investigated AMC* (left) *and VMC* (centre)*, and the second experiment measured accuracy of detection of deception* (right).

Chapter 6 presented the final telecommunication study, consisting of two closely related experiments investigating truthful and deceptive interaction in AMC and VMC. The experiments were positioned within the social domain of interpersonal trust and deception, which presents a compelling array of issues by which to investigate social interaction in visual telecommunication systems.

The first experiment in the chapter explored truthful and deceptive discourse between dyads in state-of-art AMC and VMC systems. The experimental task involved a question-answer framework in which a confederate issued questions to a participant, who responded either truthfully or deceptively. Alongside gaze, the oculesic cues of blinking and pupil size were also tracked, both for real-time representation during AMC, and for post-experimental analysis. Following the interactions, a questionnaire collected data describing participants' psychological arousal and mood state. The second experiment investigated the impact of bestowing avatars with their embodied users' oculesic cues of gaze, blinks, and pupil dilation. A different set of participants viewed audiovisual replays of the experimental interactions collected during the first experiment. Ratings of veracity and engagement were collected, together with confidence levels relating to the two judgements. These ratings were performed over three stimuli conditions: avatars exhibiting oculesics, avatars featuring no oculesics, and audio-only replays.

The main finding of the first experiment was that similar oculesic behaviour and response was demonstrated during both AMC and VMC, but that psychological arousal was greater in following the video-based interactions. This supports the finding of the conversation experiment that people respond similarly to embodied avatars as they do to live video, but also reveals that a higher degree of social presence may be fostered by VMC. The first experiment also revealed that lying and telling the truth elicited systematic differences in users' gaze and pupil size during both AMC and VMC. The temporally transient nature of oculesic cues following a given stimuli was also demonstrated, implying that isolation of specific interactional time periods was essential when performing analysis of eye tracking data. The second experiment found that avatars exhibiting oculesic behaviour, driven by eye tracking, increased the richness of NVC transmitted during AMC to the extent that observers were able to detect truth and deception more accurately. It was also found that users embodied by oculesic avatars were perceived to be more engaged in interaction than when embodied by less expressive non-oculesic avatars. Finally, the increased consistency between the nonverbal and verbal components of communication, exhibited by oculesic avatars, was found to increase confidence when rating of veracity and engagement.

8.4 Oculesic Models and Experiments

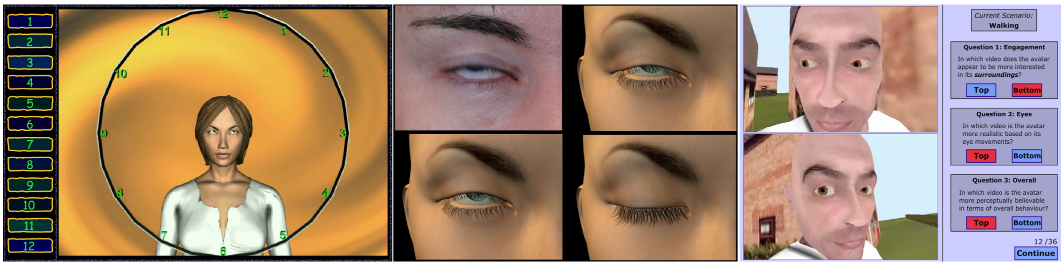


Figure 8.4: *Revisiting the oculesic models and associated experiments.* Left: *Experiment investigating varying characteristics of a virtual humanoid's oculesic behaviour on POR identification and realism.* Centre: *Kinematic models of human eyelid motion, with associated validation and general impact experiments.* Right: *Saliency-based gaze model, with associated experiment comparing to tracked gaze.*

Rather than directly studying telecommunication, the three sections presented in Chapter 7 sought to model and investigate oculesic characteristics of virtual humanoids. The first section detailed an experiment investigating the impact of varying fidelity of an autonomous agent's oculesic behaviour on observers' ability to identify its direction of gaze, and also on subjective perceptions of realism. The main finding was that eyelid animation is able to significantly increase the realism of an agent, but in some situations, may hinder accuracy of gaze direction identification. This negative impact on accuracy was contradictory to participants' self-performance judgements, which were higher when assessing the agent operating with eyelid animation.

The eyelid motion that appeared in the first experiment was simple, and not based on human physiology or motion captured data. This was subsequently addressed in the chapter's second section, which presented two parametric models generating physiologically-accurate movement of lid saccades and blinks, the two major components of human eyelid kinematics. The models were detailed algorithmically, before being validated against motion captured data, and finally assessed in terms of general impact on the realism of an avatar. The main finding of the validation experiment was that the model-generated animation was rated similar to motion captured data in terms of realism. The general impact experiment found that combined use of the two eyelid models was able to enhance the overall realism of an avatar.

The third and final section in Chapter 7 focused on gaze modelling. The approach taken aimed to generate realistic and meaningful gaze behaviour by computing the saliency of objects within an embodied user's FOV, and animating their avatar's eye movement accordingly. The parameters of proximity, eccentricity, orientation, and velocity acted as input to the saliency calculation, while an eyeball dynamics simulation was also implemented to control the saccadic motion of the eyes. An experiment collected subjective ratings of an avatar's realism when operating with varying methods of gaze control. The main finding of the experiment was that tracked gaze was rated as most realistic, followed by the saliency model, which was rated higher than both random gaze and static gaze.

8.5 Contributions

The overarching goal of the research was to investigate how AMC between users of ICVE systems can be both enhanced and understood through the use of eye tracking. Chapter 1 introduced three central research questions that are components of this goal:

1. *Can eye tracking be used to increase the nonverbal information transmitted during AMC, and does this improve quality of communication between users?*

2. *Measured by eye tracking, do users engaged in AMC behave and respond in a socially realistic manner compared to users engaged in VMC?*

3. *Can simulations of oculesic behaviour have a positive impact on observers' perceptions of virtual humanoids, and can such models be used to complement tracked behaviour in AMC?*

The first two questions are concerned with theories of media richness and social presence respectively. They address the central premise of whether eye tracking may be applied both interactively, to enhance the richness of AMC (Question 1), and analytically, to further understand peoples' social behaviour and response during AMC (Question 2). The final question (Question 3) is secondary to the focus of the overall research, and addresses how simulated behaviour may be used to complement a user's actual, tracked, movement in AMC. This thesis made both substantive and methodological contributions. The substantive contributions consist of empirical findings concerning the interactive and analytical application of eye tracking in AMC. The methodological contributions concern the theory of user embodiment in ICVEs, and technical approaches to achieve tracked oculesics in AMC.

8.5.1 Methodological Contributions

Presented in the following section, the substantive contributions made by the experimental research concern the interactive application of eye tracking to replicate users' oculesic cues in embodied avatars during real-time AMC. However, in order to facilitate the work from which these contributions are derived, development of a technical platform was required. This development work included an ICVE system supporting tracked oculesics, and a method of multimodal data collection.

Chapter 3 presented the technical details of EyeCVE, the ICVE system that was developed over the course of this research. EyeCVE enables AMC between users of various immersive and semi-immersive VR display technologies, such as CAVE and WALL systems. Other ICVE systems, such as DIVE [CH93] and MASSIVE [GB95], also support this application. However, the unique feature of EyeCVE is the integration of mobile eye tracking, which interfaces with the avatar subsystem to achieve tracked oculesic avatars. Chapter 3 detailed methods of processing the oculesic cues of gaze, blinks, and pupil size, and how to map them onto an embodied avatar for real-time oculesic animation during AMC.

A range of data collection and analysis techniques were employed over the course of the experimental work. Some methods were inspired by existing studies in the fields of social science (for instance CA), and VEs (for instance presence questionnaires). However, in a system such as EyeCVE, a user's performance is captured from multiple tracking sources, each monitoring a separate channel of natural communication. These channels include oculesics, body motion, and verbal utterances. Chapter 3 documented a solution to the problem of collating such multimodal behaviour, in addition to markup data critical in an experimental procedure, in a manner that preserves the temporal relationships between multiple, synchronous, input streams. The method was used in the object-focused and truth and deception experiments presented in Chapters 5 and 6.

Additional methodological contributions include the experimental task designs, which may be used and adapted for studies on interpersonal and object-focused AMC (Chapters 4–6), and experimental frameworks for studying perceptual aspects of virtual humanoids (Chapter 7).

8.5.2 Substantive Contributions

The substantive contributions of the experimental work, documented throughout Chapters 4–7, directly address the three central research questions posed at the beginning of the thesis. The first question asked whether eye tracking could be applied to enrich the nonverbal information transmitted during AMC, and consequently, does this improve quality of communication. Work aiming to address this question is concerned with the theory of media richness, which describes the ability of a medium to transmit and reproduce information about the individuals who are communicating, with particular focus on transmission and display of users' natural nonverbal behaviour [DL84]. The goal of AMC is to support high-quality nonverbal and verbal interpersonal communication between remote users. Hence, a crucial barrier towards achieving this goal is the relative paucity of avatar expressivity compared with live video. This is due to the lack of comprehensive capture (and subsequent transmission and display) of the gamut of human nonverbal cues, which limits the richness of AMC. A central hypothesis of this research is that, by increasing capture, transmission, and display of nonverbal cues, AMC may be enriched, and medium will be more able to support collaborative interactions that are more similar to those that take place in the real world. This view is shared by researchers in the field [Sch02, SRSH05, GSBS01]. Findings from the experiment on object-focused collaboration, presented in Chapter 5, and from the deception detection experiment, presented in Section 6.3 of Chapter 6, form the main contributions to this topic. Findings from the object-focused collaborative task did not position tracked gaze AMC as superior to static gaze AMC. In contrast, the detection deception experiment clearly demonstrated the practical implications of bestowing avatars with the tracked oculesic cues of gaze, blinks, and pupil dilation. In this case, the enhanced richness of the visual, nonverbal, component of communication enabled the veracity of embodied users to be identified more accurately in comparison with less expressive avatars.

These two contrasting findings may be reconciled through examination of the key differences between the communicative scenarios and experimental designs in question. In the object-focused collaboration, the primary role of gaze was to act as an indicator of attention. However, the addition of tracked gaze did not enhance the attentional information provided by tracked head orientation. The highly interpersonal scenario presented by deception detection experiment added the additional tracked oculesic cues of blinks and pupil dilation to the established cue of gaze. In this case, the primary role of tracked oculesics was not solely for attentional means, but also served to provide expressive insight into the psychological and cognitive state of embodied users. The roles of oculesic cues, in terms of expressing and observing deception, were established in Chapter 2 as salient and influential in the context of the experimental interactions. Additionally, and unlike head orientation in the object-focused scenario, no surrogates or approximations for these expressive channels of oculesics exist. This is true for both avatar-mediated and non-mediated discourse. Hence, this positions oculesics as critical and unique expressive cues, with operational relevance to specific scenarios such as detection of deception. Therefore, the first research question may be answered affirmatively, with a caveat stressing the importance of matching the technical capability of an ICVE system with the characteristics of an intended communicational scenario.

The second question asked whether users engaged in AMC behave and respond in a socially realistic manner, with particular reference to behaviour during VMC. Similarly to the preceding question, which is concerned with the theory of media richness, this question is related to another theory grounded in the domain of CSCW. The theory of social presence is defined as the degree of salience of a conversational partner in a one-to-one interaction [SWC76], and is also referred to as copresence in the VE literature [Gar03]. Findings from the conversational study, presented in Chapter 3, and from the truthful and deceptive interaction study, presented in Section 6.2 of Chapter 6, form the main contributions to this topic. Both experiments compared behaviour and response during interpersonal scenarios performed using AMC and VMC. The comparison condition of VMC acknowledges its status as the optimal form of visual telecommunication with regards to nonverbal exchange. Findings from the conversation experiment positioned the two mediums similarly in terms of degree of copresence, and in both forms of mediated communication, interactants were able to signal using their own gaze and observe others' gaze, similarly to collocated communication. These claims are given further support from the findings of the truth and deception interaction study, which, together with gaze, considered the additional oculesic cues of blinking and pupil dilation. The study found that users exhibited similar oculesic behaviours during interaction in both AMC and VMC. This is consistent with Reeves and Nass' theory of the medium as social actor, which hypothesises that people will tend to anthropomorphise media to an extent that they treat them as social entities. However, it was also found that psychological arousal was greater following VMC than following AMC, suggesting that live video increases the interpersonal experience of the mediated interaction. Hence, the second question may be answered affirmatively, but, similarly to the first question, the relationship between the telecommunication medium and its intended application should be carefully considered.

The third question asked whether simulations of oculesic behaviour could have a positive impact on observers' perceptions of avatars, and also, can such simulations be used to complement a user's tracked behaviour in AMC. Findings from the oculesic models and experiments presented in Chapter 7 form the main contributions to this topic, and support an affirmative answer to the question. Findings from the eyelid models' general impact experiment, presented in Section 7.2.5, indicated that the perceived realism of an avatar is increased by the combination of tracked gaze with simulated eyelid kinematics. The work presented in Section 7.3.3 found that avatars exhibiting gaze generated by the saliency model were rated as more realistic than those exhibiting random or static gaze. Hence, in the absence of eye tracking apparatus, the gaze model could be coupled with head tracking, thereby bestowing avatars with relevant gaze behaviour. This is also related to another contribution of the thesis, which reviewed the problematic use of behavioural modelling in AMC (Chapter 2). Focusing on gaze, the review argued that, due to the unpredictable, interpersonal, and idiosyncratic nature of human social interaction, it is not feasible to simulate a user's nonverbal behaviour without distorting the semantic content of the original action. Hence, for real-time AMC, tracking and replication must be implemented wherever possible. However, the review did not focused on use of modelling and tracking in combination. The gaze and eyelid models present an interesting combination for such hybrid form of avatar control, which is primarily driven by tracking, but includes elements of automated control. For instance, the initial input to the blink model may be a user's actual blink, detected by eye tracking. However, the resulting animation of this behaviour is not tracked, and rather is generated by the blink model's algorithm. Similarly, the main input to the gaze model is a user's tracked head movement, which is used to calculate current FOV and subsequently to calculate the saliency of objects lying within, to which gaze is distributed accordingly.

8.6 Directions for Future Work

The mediated communication captured during the three telecommunication studies were guided by experimental frameworks designed to engage participants in a particular social and collaborative scenario. In order to control data collection, and focus on the aims of each experiment, it was critical to impose a degree of control on the unfolding interactions. Thus, while participants were free to act as they wished within each particular framework, it may be argued that the captured AMC is removed somewhat from how such systems may be used for daily work and leisure. This criticism may be levelled at much work in the field of VEs, and relatively few studies (an exception being [SSH+03]) have investigated AMC over long time periods and with expert users. Such investigation of tracked oculesic AMC between friends and strangers, and between naïve and experienced users, is a potentially revealing avenue of research. Throughout the experimental work, collaboration around the object-focused puzzle, detailed in Chapter 5, arguably placed the highest demands on participants, in terms of task difficulty and cognitive load. However, this study also revealed the significant impact of learning effects, which resulted in increased performance with experience. Hence, further benefit of an AMC system featuring tracked oculesics may be revealed through more organic collaboration between expert users.

The framework imposed during the truthful and deceptive interactions detailed in Chapter 6 was also likely to have implications on the captured communication. The experimental task was designed to manipulate participants' cognitive load and psychological arousal by asking them to tell truths and lies during a structured question-answer scenario. While this structure enabled analysis of explicit interactional states, it did, however, present an artificial social scenario in which participants' oculesic behaviours were measured when they were being *told* to lie as opposed to *choosing* to be deceptive. Hence, the measured cues of oculesics may conceivably differ in more natural and unstructured interaction common to how CSCW systems are used in daily life. The scenario of truth and deception suggests a compelling application of tracking an individual's natural action. The study demonstrated that people exhibit systematic differences in oculesic behaviour between truth telling and lying. This finding may be more generally applicable to classifying an individual's mental state. For example, alongside veracity, information regarding current cognitive load or emotional arousal may also be inferred directly from eye tracking data. This may find diverse applications in the real-world, such as for lie detection, competency tests, and medical use.

The eyelid kinematics models were designed to generate generic motion based on normal human physiology. However, due to their parametric implementation, several opportunities are presented for alternative expressive forms of motion. For instance, simply by reducing the velocity of the down-phase of a blink, observers are likely to perceive a virtual humanoid as fatigued [SBS94]. Results from the saliency gaze model were promising, and while the author does not consider behavioural modelling as an alternative to tracking in AMC, the development of effective and reusable gaze models present a compelling array of problems and opportunities. The gaze model may be extended to more rigorously analyse objects within a user's FOV in order to enable allocation of gaze to surfaces of objects instead of simply their centre, as is currently implemented. To achieve this, further investigation into how attention is allocated during object scrutiny would be required. Secondly, extending the gaze model to also drive an agent's head direction would be an interesting challenge. Indeed, automated full-body animation may also be considered, including pointing and locomotion. Finally, the allotment of attention, based on the saliency of objects within, may be a useful resource for VE designers regarding construction of scenes.

To further enhance AMC's ability to support rich interpersonal telecommunication, additional channels of natural nonverbal expression must be tracked and reproduced in real-time. To this end, facial expression and accurate lip synchronisation should be considered as high priorities. However, alongside this gradual progression towards incorporating the gamut of human nonverbal behaviour, the operational benefit of each should be a central consideration. It is possible that capture and display of some behavioural cues may provide only minor benefits to communication, or, due to additional worn devices, may cause discomfort to users. A particular consideration should be the coupling of an intended communication with a telecommunication medium's characteristics of social presence and media richness. AMC may be particularly suited to applications in which a high degree of social presence and media richness is desirable, in combination with a preference for anonymity and a less affecting level of psychological arousal. This may include communication interfaces for virtual tutoring, people with social anxiety, or online socialising. However, it should be acknowledged that critical and highly interpersonal telecommunication is likely to always be more suited to the faithful representation of fellow interactants as provided by VMC. In this sense, mediated telecommunication systems can be regarded as filters for behavioural cues, losing the fidelity of some, while making others more salient. This is particularly the case in AMC, as each expressive channel of nonverbal communication can be tightly controlled; from complete neglect to faithful reproduction. This ability for may be exploited to suit the characteristics of a collaborative task or intentions of an individual, thereby having far-reaching implications on how future telecommunication mediated by avatars will unfold.

This thesis has aimed to investigate the use of eye tracking to both enhance and understand AMC. Research covered multiple collaborative scenarios, comparisons with state-of-art VMC, methods for representing the oculesic cues of gaze, blinks, and pupil dilation in AMC, models of gaze and eyelid movement, networked VE system design and development, data collection methods, and analytical techniques. The findings suggest that eye tracking is able to enhance AMC towards a richer medium for interpersonal telecommunication in which interactants are able to transmit and recognise subtle nonverbal signals relating to underlying communicative intent. Also, the degree of social presence experienced in AMC was found to be less acute than that fostered by VMC, but that users' social behaviour in AMC was similar to that demonstrated in VMC. The importance of matching the properties of a communication with those of the communication medium was stressed. Future work will build on these findings by exploring natural interaction scenarios, which both objectively and subjectively assess the use of AMC as a practical telecommunications medium suitable for mainstream adoption.

Appendices

Appendix A

Publications

The following publications, all appearing in peer-reviewed international conferences and journals, are presented in chronological order according to date of publication. Where appropriate, the section in this thesis corresponding to the work presented in the publication is referenced.

IEEE Virtual Reality, 2008

William Steptoe and Anthony Steed

High-Fidelity Avatar Eye-Representation.

Features extracts of work presented in Chapter 7.

```
@inproceedings{DBLP:conf/vr/SteptoeS08,
    title = {High-Fidelity Avatar Eye-Representation},
    author = {William Steptoe and Anthony Steed},
    publisher = {IEEE},
    booktitle = {VR},
    year = {2008},
    pages = {111-114},
    ee = {http://dx.doi.org/10.1109/VR.2008.4480759}
}
```

Presence, 2008

John Rae and Estefania Guimaraes and William Steptoe

Simulation versus Reproduction for Avatar Eye-Gaze in Immersive Collaborative Virtual Environments.

Features extracts of work presented in Chapter 2.

```
@inproceedings{DBLP:conf/presence/RaeGS08,
    title={Simulation versus Reproduction for Avatar Eye-Gaze in Immersive Collaborative Virtual
    Environments},
    author={John Rae and Estefania Guimaraes and William Steptoe},
    publisher = {IEEE},
    booktitle = {Presence},
    year={2008}
    pages = {85-94}
}
```

ACM Computer Supported Cooperative Work, 2008

William Steptoe and Robin Wolff and Alessio Murgia and Estefania Guimaraes and John Rae and Paul Sharkey and David Roberts and Anthony Steed

Eye-Tracking for Avatar Eye-Gaze and Interactional Analysis in Immersive Collaborative Virtual Environments.

Features extracts of work presented in Chapter 4.

```
@inproceedings{DBLP:conf/cscw/SteptoeWMGRSRS08,
    author = {William Steptoe and Robin Wolff and Alessio Murgia and Estefania Guimaraes and
        John Rae and Paul Sharkey and David Roberts and Anthony Steed},
    title = {Eye-tracking for avatar eye-gaze and interactional analysis in immersive collaborative
        virtual environments},
    publisher = {ACM},
    booktitle = {CSCW},
    year = {2008},
    pages = {197-200},
    ee = {http://doi.acm.org/10.1145/1460563.1460593}
}
```

IEEE/ACM Distributed Simulation and Real Time Applications, 2008

Robin Wolff and David Roberts and Alessio Murgia and Norman Murray and John Rae and William Steptoe and Anthony Steed and Paul Sharkey

Communicating Eye Gaze across a Distance without Rooting Participants to the Spot.

Features extracts of work presented in Chapter 3.

```
@inproceedings{DBLP:conf/dsrt/WolffRMMRSSS08,
    title = {Communicating Eye Gaze across a Distance without Rooting Participants to the Spot},
    author = {Robin Wolff and David Roberts and Alessio Murgia and Norman Murray and
    John Rae and William Steptoe and Anthony Steed and Paul Sharkey},
    publisher = {IEEE Computer Society},
    booktitle = {DS-RT},
    year = {2008},
    pages = {111-118},
    ee = {http://dx.doi.org/10.1109/DS-RT.2008.28}
}
```

IEEE/ACM Distributed Simulation and Real Time Applications, 2008

Alessio Murgia and Robin Wolff and William Steptoe and Paul Sharkey and David Roberts and Estefania Guimaraes and Anthony Steed and John Rae

A Tool for Replay and Analysis of Gaze-Enhanced Multiparty Sessions Captured in Immersive Collaborative Environments.

Features extracts of work presented in Chapter 3.

```
@inproceedings{DBLP:conf/dsrt/MurgiaWSSRGSR08,
    title = {A Tool for Replay and Analysis of Gaze-Enhanced Multiparty Sessions Captured in
```

Immersive Collaborative Environments},
author = {Alessio Murgia and Robin Wolff and William Steptoe and Paul Sharkey and
David Roberts and Estefania Guimaraes and Anthony Steed and John Rae},
publisher = {IEEE Computer Society},
booktitle = {DS-RT},
year = {2008},
pages = {252-258},
ee = {http://dx.doi.org/10.1109/DS-RT.2008.25}
}

IEEE Virtual Reality, 2009

William Steptoe and Oyewole Oyekoya and Alessio Murgia and Robin Wolff and John Rae and Estefania Guimaraes and David Roberts and Anthony Steed

Eye Tracking for Avatar Eye Gaze Control During Object-Focused Multiparty Interaction in Immersive Collaborative Virtual Environments.

Features extracts of work presented in Chapter 5.

@inproceedings{DBLP:conf/vr/SteptoeOMWRGRS09,
title = {Eye Tracking for Avatar Eye Gaze Control During Object-Focused Multiparty
Interaction in Immersive Collaborative Virtual Environments},
author = {William Steptoe and Oyewole Oyekoya and Alessio Murgia and Robin Wolff and
John Rae and Estefania Guimaraes and David Roberts and Anthony Steed},
publisher = {IEEE},
booktitle = {VR},
year = {2009},
pages = {83-90},
ee = {http://dx.doi.org/10.1109/VR.2009.4811003}
}

IEEE Virtual Reality, 2009

David Roberts and Robin Wolff and John Rae and Anthony Steed and Rob Aspin and Moira McIntyre and Adriana Pena and Oyewole Oyekoya and William Steptoe

Communicating Eye-gaze Across a Distance: Comparing an Eye-gaze enabled Immersive Collaborative Virtual Environment, Aligned Video Conferencing, and Being Together.

@inproceedings{DBLP:conf/vr/RobertsWRSAMPOS09,
author = {David Roberts and Robin Wolff and John Rae and Anthony Steed and Rob Aspin
and Moira McIntyre and Adriana Pena and Oyewole Oyekoya and William Steptoe},
title = {Communicating Eye-gaze Across a Distance: Comparing an Eye-gaze enabled Immersive
Collaborative Virtual Environment, Aligned Video Conferencing, and Being Together},
publisher = {IEEE},
booktitle = {VR},
year = {2009},
pages = {135-142},
ee = {http://dx.doi.org/10.1109/VR.2009.4811013}
}

ACM Virtual Reality Software and Technology, 2009

Oyewole Oyekoya and William Steptoe and Anthony Steed

A Saliency-based method of Simulating Visual Attention in Virtual Scenes.

Features extracts of work presented in Chapter 7.

@inproceedings{DBLP:conf/vrst/OyekoyaSS09,
    author = {Oyewole Oyekoya and William Steptoe and Anthony Steed},
    title = {A saliency-based method of simulating visual attention in virtual scenes},
    publisher = {ACM},
    booktitle = {VRST},
    year = {2009},
    pages = {199-206},
    ee = {http://doi.acm.org/10.1145/1643928.1643973}
}

ACM Human Factors in Computing Systems (CHI), 2010

William Steptoe and Anthony Steed and Aitor Rovira and John Rae

Lie Tracking: Social Presence, Truth and Deception in Avatar-Mediated Telecommunication.

Features extracts of work presented in Chapter 6.

@inproceedings{DBLP:conf/chi/SteptoeSRR10,
    title = {Lie Tracking: Social Presence, Truth and Deception in Avatar-Mediated Telecommunication},
    author = {William Steptoe and Anthony Steed and Aitor Rovira and John Rae},
    publisher = {ACM},
    booktitle = {CHI},
    year = {2010},
    pages = {1039–1048},
    ee = {http://doi.acm.org/10.1145/1753326.1753481}
}

Computer Animation and Virtual Worlds, Volume 21 Issue 3–4, May 2010

William Steptoe and Oyewole Oyekoya and Anthony Steed

Eyelid Kinematics for Virtual Characters.

Features extracts of work presented in Chapter 7. Presented at Computer Animation and Social Agents (CASA) 2010.

@inproceedings{DBLP:conf/chi/SteptoeSRR10,
    title = {Eyelid Kinematics for Virtual Characters},
    author = {William Steptoe and Oyewole Oyekoya and Anthony Steed},
    publisher = {John Wiley & Sons},
    booktitle = {Computer Animation and Virtual Worlds},
    year = {2010},
    pages = {1546–4261},
    ee = {http://dx.doi.org/10.1002/cav.354}
}

Appendix B

List of Acronyms

The following acronyms appear in this thesis:

2D	Two Dimensional
3D	Three Dimensional
ADCM	Activation-Decision-Construction Model
AG	Access Grid
AMC	Avatar-Mediated Communication
ANOVA	Analysis Of Variance
ASL	Applied Science Laboratories
CA	Conversation Analysis
CAVE	Cave Automatic Virtual Environment
CSCW	Computer Supported Cooperative Work
DOF	Degree-of-Freedom
DVCR	Digital Video Camera Recorder
FPS	Frames-per-Second
GUI	Graphical User Interface
HD	High Definition ("Full HD" indicates 1920×1080 pixels)
HITL	Human-in-the-Loop
HMD	Head-Mounted Display
ICVE	Immersive Collaborative Virtual Environment
IK	Inverse Kinematics
LED	Light Emitting Diode
MTU	Maximum Transmission Unit
NVC	Nonverbal Communication
PI	Place Illusion
POMS	Profile of Mood States (questionnaire)
POR	Point-of-Regard
SC	Sensorimotor Contingencies
UCL	University College London

UDP	User Datagram Protocol
UoS	University of Salford
UoR	University of Reading
VE	Virtual Environment
VMC	Video-Mediated Communication
VoIP	Voice over Internet Protocol
VR	Virtual Reality
VRPN	Virtual Reality Peripheral Network

Appendix C

Materials for the Conversation Experiment

C.1 Information Sheet for Participants

Department of Computer Science
Mel Slater
Professor of Virtual Environments

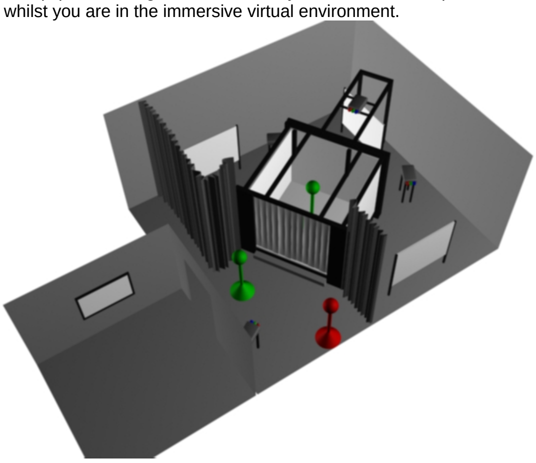
University College London
Gower Street
London WC1E 6BT UK

+44 (0)20 7679 3709
+44 (0)20 7387 1397 Fax
M.Slater@cs.ucl.ac.uk
http://www.cs.ucl.ac.uk/staff/M.Slater

INFORMATION SHEET FOR PARTICIPANTS

Thank you for participating in our study. This is one of a long series of studies into understanding the responses of people within virtual environments. This study has been approved by *University College London's Committee on the Ethics of Non-NHS Human Research*. Please read through this information sheet and feel free to ask any questions. The experimenters will answer any general questions; however the specific aspects regarding this study cannot be discussed with you until the end of the session. The whole study will take about *one hour*.

You will be using the CAVE™-like system called the ReaCTor. See figure below. The ReaCTor is a VR system made up of 3 walls measuring roughly 3m x 3m x 3m. You will wear VR glasses and a tracker. The virtual reality viewing equipment can be worn over eyeglasses. You may be asked to take off your shoes in order to protect the virtual reality equipment. In addition to the tracking equipment used to navigate the system, you will also be fitted with physiological equipment designed to measure your heart rate, respiration and galvanic skin responses whilst you are in the immersive virtual environment.



In this particular study you will be going into a party by yourself and there are a few other people there. Some of them may talk to you, and you can talk with them if you wish. We the experimenters will not be there with you.

PLEASE TURN OVER

<u>*Please ask any questions*</u> that come to mind. Read and sign the **Consent Form**.

Information that we collect will never be reported in a way that specific individuals can be identified. Information will be reported in a statistical and aggregated manner, and any verbal comments that you make, if written about in subsequent papers, will be presented anonymously.

IMPORTANT

When people use virtual reality systems, some people sometimes experience some degree of nausea. If at any time you wish to stop taking part in the study due to this or any other reason, please just say so and we will stop.

There has been some research, which suggests that people using head-mounted displays might experience some disturbances in vision afterwards. No long term studies are known to us, but the studies which have been carried out do testing after about 30 minutes, and find the effect is still sometimes there.

There have been various reported side effects of using virtual reality equipment, such as 'flashbacks'.

With any type of video equipment there is a possibility that an epileptic episode may be generated. This, for example, has been reported for computer video games.

PROCEDURES

- You will be asked to read, understand and sign a **Consent Form**. If you sign it the study will continue with your participation. **Note that you can withdraw at any time without giving any reasons.**
- You will be asked to complete a number of questions on paper, so that we can try to understand your responses during the study.
- You will be fitted with sensors to measure your heart rate, respiration and galvanic skin responses.
- You may be asked to remove your shoes and switch off mobile phones before using the VR equipment.
- You will have a training period standing in the CAVE for us to collect your physical data as a baseline. You will then go into the environment as mentioned above and stay there for a few minutes during which you will be videotaped.
- After the visit to the environment you will complete a questionnaire about your experience, and a questionnaire which is similar to the one you did before coming.
- Finally there will be a small discussion with the experimenters about your experiences. During this time, you might be audio or video taped.
- **Thank you** for your participation. Please do not discuss this study with others for about **three months**, since the study is continuing.
- Any other questions?

William Steptoe
Department of Computer Science
University College London
Gower Street
London WC1E 6BT

+44 (0) 20 7679 3672 *Tel*
+44 (0) 20 7387 1397 *Fax*

w.steptoe@cs.ucl.ac.uk
http://www.cs.ucl.ac.uk/staff/w.steptoe

C.2 Consent Form for Participants

University College London
Gower Street
London WC1E 6BT UK

Department of Computer Science
William Steptoe

W.Steptoe@cs.ucl.ac.uk

ID

PROJECT: EYE CATCHING
Investigators: William Steptoe, Anthony Steed

To be completed by volunteers:
We would like you to read the following questions carefully.

Have you read the information sheet about this study?	YES/NO
Have you had an opportunity to ask questions and discuss this study?	YES/NO
Have you received satisfactory answers to all your questions?	YES/NO
Have you received enough information about this study?	YES/NO
Which investigator have you spoken to about this study?	…………
Do you understand that you are free to withdraw from this study?	
• At any time	YES/NO
• Without giving a reason for withdrawing	YES/NO

Do you understand and accept the risks associated with the use of virtual reality equipment? YES/NO

Do you agree to take part in this study?	YES/NO
Do you agree to be audio taped?	YES/NO
Do you agree to have your head and eye motion tracked ?	YES/NO
Do you agree for the above motion to be mapped to a virtual character along with your audio recording for other confidential experimental participants to view?	YES/NO

I certify that I do not have epilepsy.

I certify that I will not be driving a car, motorcycle, bicycle, or use other types of complex machinery that could be a danger to myself or others, within 3 hours after the termination of the study.

Signed…………………………………Date……………………………...

Name in block letters.……………………………………………………...

Investigator.…………………………………………………………..

In case you have any enquiries regarding this study in the future, please contact:

William Steptoe
Department of Computer Science
University College London
Gower Street
London WC1E 6BT

+44 (0) 20 7679 7215 *Tel*
+44 (0) 20 7387 1397 *Fax*

W.Steptoe@cs.ucl.ac.uk

C.3 Personal Information Form

ID ______________________

Your Age	
How fluent is your English?	Basic ☐ Proficient ☐ Fluent ☐
Occupational status *(If other, please specify and also your area of interest)*	☐ Undergraduate Student ☐ Masters Student ☐ PhD Student ☐ Research Assistant/Fellow ☐ Staff - systems, technical ☐ Faculty ☐ Administrative Staff ☐ Other
Please state your level of computer literacy on a scale of (1...7)	(novice) 1 2 3 4 5 6 7 (expert)
Please rate your level of experience with computer *programming*	(novice) 1 2 3 4 5 6 7 (expert)
Have you ever experienced 'virtual reality' before?	(no experience) 1 2 3 4 5 6 7 (extensive experience)
How many times did you play video games (at home, work, school, or arcades) in the last year?	Never ☐ 1 - 5 ☐ 6 - 10 ☐ 11 - 15 ☐ 16 - 20 ☐ 21 - 25 ☐ > 25 ☐
How many *hours per week* do you spend playing video games?	0 ☐ < 1 ☐ 1 - 3 ☐ 3 - 5 ☐ 5 - 7 ☐ 7 - 9 ☐ > 9 ☐

C.4 Extended Conversation Analysis

ANALYSIS

In order to examine participants' interactional behaviour we have used conversation analysis, which is geared to identifying systematic practices used by people in social interaction. The present analysis is limited to reporting on aspects of the interviewees' gaze practices, although it will be necessary to comment briefly on some of the interviewers' behaviour. The transcripts presented below to support our observations show the talk of a given speaker (and/or moments of silence) in the numbered lines and a line below the numbered line shows where the interviewee was gazing at that moment. All the participants were given pseudonyms in transcripts. The interviewers were called: Robert (Rob), Peter (Pet) and Paul (Pau) and the interviewees were given different names. In order to show where the speakers were gazing at we used the following initials: R (Rob); P (Peter or Paul, for Access Grid and CAVE respectively) – always in capitals for Access Grid and in capital letters for the CAVE when showing gaze directed at centre and/or face of the avatar, but in lower caps for their hands and arms; C (centre); A (away from speakers and from the centre); T (table, in the CAVE).

1) Supporting Eye-Gaze Practices

The two environments support certain eye-gaze practices that are observed in co-present interaction. As in co-present interactions (Kendon, 1973; Stivers and Rossano, date), in the two studied environments participants tend to (a) look at a speaker (when being asked a question); (b) they look away when starting answering the question but (c) bring gaze back to the questioner as the turn construction units - out of which answers are built - reach completion.

1.1 <u>Access Grid</u>
<u>#1. AG#04-Interviewee View</u>

```
05   Wal:    (huh)[(huh)
06   Rob:         [Can you tell us briefly about your
     (W): a→ (R)_________________________________________

07           professional academic background °please°.
     (W): a→(R)_________________________________________..

08   Wal:    Er:m >sure< em I'm a PhDv
     (W): b→ (C)__________________________

09           student at (names institution)(.)
     (W): c→___.(R)____________________________.
```

1.2 <u>CAVE</u>
<u>#2. CAVE#06-Interviewee View</u>

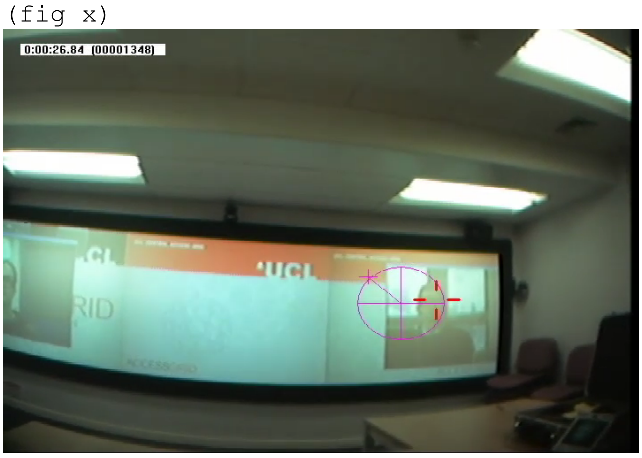
```
04   Rob:     Okay thanks er a what kind of per son:
          a→ R_______________________________________

05             (.) would you refuse to work with
          a→ R_____________________________  P___
```

```
06              (.)
          b→ C__
07   Ric:       Er:: (.) someone stubborn (.) that doesn't
          b→ C____________ R________ P____________________
                           ↑
                           c
08              listen to others
          c→ P_____________ R_
09              (2.0)
                R___
```

One feature of participant's eye-gaze is that in Access Grid when they gaze at an interviewer their gaze is directed at the interviewer's face (see fig. x). In the CAVE, some participants, at the start of the interview, did not gaze at the avatar's face region (figure y) but started doing so as the interview developed (figure w, z), this provides evidence that they learned that the avatars eye gaze was relevant.

(fig x)



(AG#03 – Interviewee View : "I'm a computer scientist")

(fig y)


(fig w)


(fig z)

2) Management of speaker transition

The interaction, an interview with clear roles of interviewers – who ask scripted questions in a pre-allocated order – and interviewee – who answers questions posed by the interviewers – ran smoothly in the two environments. The question-answer format had clear constraints regarding the selection of next speaker and the kind of talk that was relevant after a question. Speakers transition was only a matter of concern after a question was answered by the interviewee and the questioners had to self-select to (1) produce a receipt of the answer (as having answered the question), or (2) produce a next question.

2.1 Access Grid

In the video-conferencing environment, in those moments in which turn allocation had to be locally (and collaboratively) managed by the participants, eye-gaze was an important element in turn-taking. The data extract presented below shows one of those instances of speakers transition and the role of eye-gaze in turn taking as line 05 shows the prior questioner (Rob) bringing his gaze to next questioner (Pet) and this being followed by the participant's gaze, which is then brought to the next questioner, who subsequently (line 06) starts asking his next question.

#3. AG#03-Interviewee View

```
01   Cha:   uh: .hh and I am working as a: uh: research
     (C):    ..(C)__________.(R)________.(A)___.(R)______

02          fellow here at (Place)¿=.h Uh in computer
     (C):   __________________________________________

03          science¿.hhh Primarily character animation¿
     (C):   __________________________________________

04           (0.8)
     (C):   _____

             (Rob brings gaze to Pet and Charles follows)
              /
05          (0.3)
     (C):   __..

06   Pet:   But do::: do you c- consider yourself
     (C):   (P)___________________________________

07          successful.
     (C):   __________
```

The spatial fragmentation of this environment created, in some instances, some difficulties for this local management of turn-taking. Fragment (4), presented below, shows how the participant witnesses the unfolding of speaker transition in a long period of silence (4.2 seconds, line 09). The participant uses gaze to monitor who is speaking next: he gazes at ‘P’ as the next questioner, but in the absence any talk from Peter he moves his gaze to the centre, back to Pete and then to Rob, where he sees Rob shifting his gaze towards Pete and then gazes back again at Pete and sustains this gaze as Pete asks his question.

#4. AG#03-Interviewee View

```
01   Cha:   But maybe sort of I wouldn’t .hh uh::
     (C):   _____________________________________

02          youknow it’s not so good sort of .h
     (C):   ___________________________________

03          working with somebody who’s: t- too
     (C):   .(P)_________.(A)______________..(R)

04          rigid in their way of doing things that
     (C):   _____________________________________

05          won’t listen and I suppose that’s sort
     (C):   _________________.(A)__________________

06          of something that is difficult¿=but .hh
     (C):   ___.(P)_____.(A)___.(P)________.(A)_____

07          (oi) is something I could be accused of
     (C):   __________________.(R)__________________

08          sometimes so: .hh youkno:w huh huh
     (C):   _______________________________..(P)
```

```
                                 (Rob gazes at Pet)
                                          /
09    →     (- - - - - - - - - - - - - - - - - - - - - - -)
    (C):    ______..(C)_..(P)__..(C)__..(R)_...(P)_____

10  Pet:    And my last question is uh:m what would
    (C):    ______________________________________________
```

2.2 CAVE

Some limitations of the CAVE environment, such as the lack of lip movement from the avatars and its non spatialised sound created further challenges to participants in terms of 'recognizing' which avatar was speaking. In order to make clear to the participants who was speaking, the interviewers adopted a system of naming and pointing to a next speaker to mark the turn allocation of next speakers in their interviewing role. This limitation made turn-taking on the CAVE a bit artificial in comparison to ordinary conversation (Sacks, Schegloff and Jefferson, 1974).

Extract 5 (below) shows, similarly to the extract presented above, that after completion of his own turn, the participant gazes at the interviewers at the end of his turn (line 02) and then uses gaze (line 03) to monitor who is going to speak next. Differently from the AG environment and due to the limitations described above, the transition of speakers is aided by pointing (lines 04-05) and figures (a-b).

#5. CAVE#03-Interviewee View

```
01  Sim:    Uh:m: (.) before coming to (name of place).=
    (S):    __________________________.(P)________________

02          =To:: run the (Name) la:b.
    (S):    ________..(R)___.(R)_____.

03          ( -  -  - - )
    (S):    (P)_.(r).(R)

                (R extends arm to P)
                         /
04  Rob:    O:ka:y (0.2) So: to yo:u¿
    (S):    ________________.(r)_______

              (P extends arm too)
               /
05          (.)
    (S):    (p)

06  Pau:    Oka:y=uh:: do you consider yourself up till
    (S):    _____.(P)___________________________________

07          no::w: (.) successful¿=And why:.
    (S):    __________________________.(p)_______
```


(fig a)


(fig b)

This communicational feature, although somewhat artificial in comparison to co-present speaker transitions, makes evident another important feature of the CAVE environment: that its shared space makes it possible for gesture to be used as a communicational resource.

3) Gaze Distribution

Another feature observed in those intreractions was that in the Access Grid, participants often looked at the questioner when answering a posed question, wheres in the CAVE it was more common for participants lo gaze at the two interviewers when responding to a question, which can be seen as greater feeling of multi-party conversation in the shared space of the CAVE in comparison with the Access Grid.

#6. AG#03-Interviewee View

```
15  Rob:    Okay can you tell us u:h a bit (0.2) about
     (C):   (R)________________________________________

16          your professional academic background please.
     (C):   .(R)______________.___________________..____..

17  Cha:    Uh::m: we:ll I am a computer sci:entis:t.
     (C):   (C)_______________.(R)____________________

18          Uh:: I've got a .hhh m- degree in computer
     (C):   _.(C)____________________.(R)______________

19          science and a PhD in computer graphics and
```

```
     (C):      ________________________________________
01             animation¿ uh: .hh and I am working as a:
     (C):      ____________..(C)__________.(R)________.(A)

02             uh: research fellow here at (Place)¿=.h Uh in
     (C):      ___.(R)__________________________________

03             computer science¿.hhh Primarily character
     (C):      _________________________________________

04             animation¿
     (C):      __________
```

#7. CAVE#06-Interviewee View

```
01   Rob:       So I'm gonna ask the first question (.)
                P___________ R__________________________

02              Okay (.)
                R_______

03              Can you tell us about your
                R_________________________

04              academic and professional background
                R___________________________________

05              °please.°
                P_______

06   Ric:       Err Academic er I did a bachelors in
                P ..C___________________ R___________

07              computer science.
                R________________

08              (.)and I'm doing a PhD in bio
                C_____ R______________________

09              informatics: now. (.) at the moment
                R_____P______ R_________________ P__

10              °h professional background
                P______________ R__ P_____

11              pwpwt! nothing.
                P_.R______________
```

C.5 Extract of Mobile Eye Log File

Avi TimeSt:	Frame	Spot x	Spot y	Pupil x	Pupil y	Pupil r	Scene x	Scene y
--------	321807	398.36	406.14	-2000	-2000	-2000	-55.95	-298.32
--------	321808	409.21	398.33	-2000	-2000	-2000	-14.09	-276.26
--------	321809	417.36	393.67	-2000	-2000	-2000	16.42	-264.22
--------	321810	404.34	394.12	-2000	-2000	-2000	-26.89	-257.09
--------	321811	411	389	-2000	-2000	-2000	-0.95	-242.32
--------	321812	414.77	385.64	-2000	-2000	-2000	14.11	-232.21
--------	321813	417.23	383.77	-2000	-2000	-2000	23.67	-226.85
--------	321814	419.5	381.91	-2000	-2000	-2000	32.6	-221.38
--------	321815	421	379.25	-2000	-2000	-2000	39.61	-212.38
--------	321816	425.83	381.33	419.76	423.66	73.72	319.29	-37.07
--------	321817	424.82	382.86	-2000	-2000	-2000	314.77	-42.12
--------	321818	-2000	-2000	-2000	-2000	-2000	-2000	-2000
--------	321819	-2000	-2000	-2000	-2000	-2000	-2000	-2000
--------	321820	-2000	-2000	-2000	-2000	-2000	-2000	-2000
--------	321821	-2000	-2000	-2000	-2000	-2000	-2000	-2000
--------	321822	-2000	-2000	-2000	-2000	-2000	-2000	-2000
--------	321823	-2000	-2000	-2000	-2000	-2000	-2000	-2000
--------	321824	398	442.6	-2000	-2000	-2000	179.93	-248.96
--------	321825	400.26	428.36	-2000	-2000	-2000	198.43	-196.84
--------	321826	403.09	417.18	-2000	-2000	-2000	216.44	-156.65
--------	321827	407.13	406.75	381.94	480.16	66.69	343.98	-7.01
--------	321828	409.13	401.21	383.61	471.57	68.01	343.03	0.52
--------	321829	397.54	408.93	365.68	487.21	64.47	304.72	-25.89
--------	321830	401.79	403	-2000	-2000	-2000	323.36	-6.49
--------	321831	404.77	396.23	-2000	-2000	-2000	338.41	16.91
--------	321832	407.08	389.63	-2000	-2000	-2000	351.12	40.05
--------	321833	409	384.23	389.62	445.98	70.65	368.72	36.58
--------	321834	-2000	-2000	-2000	-2000	-2000	-2000	-2000
--------	321835	384.17	395.33	-2000	-2000	-2000	278.17	11.73
--------	321836	379.11	384.94	337.3	450.7	71.22	277.62	33.46
--------	321837	384.21	377	341.42	435.73	70.68	289.93	55.49
--------	321838	389.71	370.79	-2000	-2000	-2000	312.86	74.94
--------	321839	389	366.56	-2000	-2000	-2000	313.79	91.28
--------	321840	355.42	366.83	289.58	424.97	73.17	195.48	89
--------	321841	360.87	360.33	299.8	414.08	75.49	217.45	92.11
--------	321842	366	357.92	306.49	407.51	73.66	229.41	101.13

Appendix D

Materials for the Object-Focused Experiment

D.1 Consent Form for Participants

UCL DEPARTMENT OF COMPUTER SCIENCE

Project: Eyecatching 2007-2008

Investigators: Anthony Steed, William Steptoe, Wole Oyekoya (UCL)
David Roberts, Norman Murray, Robin Wolff (University of Salford)
Paul Sharkey, Alessio Murgia (University of Reading)
John Rae, Paul Dickerson, Estefania Guimaraes (Roehampton University)

To be completed by volunteers:
We would like you to read the following questions carefully.

Have you read the information sheet about this study? YES/NO
Have you had an opportunity to ask questions and discuss this study? YES/NO
Have you received satisfactory answers to all your questions? YES/NO
Have you received enough information about this study? YES/NO

Which investigator have you spoken to about this study? …………

Do you understand that you are free to withdraw from this study?

- At any time YES/NO
- Without giving a reason for withdrawing YES/NO

You received an information sheet which outline the potential risks of using virtual reality equipment, including (1) the risk of driving a car, motorcycle, bicycle and/or using complex machinery; (2) the risk of getting flashbacks and/or nausea from taking part in the experiment; and (3) the potential discomfort of VR glasses.

Do you understand and accept the risks associated with the use of virtual reality equipment? YES/NO
Do you agree to take part in this study? YES/NO
Do you agree to the data being stored and used in the researcher's ongoing research? YES/NO
Do you agree to the data being stored indefinitely? YES/NO
Do you agree to be video taped? YES/NO
Do you agree to be audio taped? YES/NO
Do you agree to let us make transcripts of your interactions and present them in printed publications? YES/NO
Do you agree to let us make transcripts of your interactions and present them in presentations? YES/NO
Do you agree to let us use video frames or photo stills of you or your avatar in printed publications? YES/NO
Do you agree to let us use the video or photo stills of you or your avatar in presentations? YES/NO
Do you agree to have your body movements tracked while in the virtual reality system? YES/NO
Do you agree to have your eye movements tracked while in the virtual reality system? YES/NO
Do you agree to let us use video recordings showing what your eye is tracking in presentations? YES/NO
Do you agree to let us use video recordings showing what your eye is tracking in printed publications? YES/NO
Do you agree to the audio, video and other data being shared with academic collaborators on the project at University of Salford, Roehampton University and University of Reading? YES/NO

Do you agree to completing a personality questionnaire? YES/NO
I certify that I do not have epilepsy.

I certify that I will not be driving a car, motorcycle, bicycle, or use other types of complex machinery that could be a danger to myself or others, within 3 hours after the termination of the study.

Signed……………………………………Date……………………………...

Name in block letters………………………………………………………...

Investigator…………………………………………………………………..

In case you have any enquiries regarding this study in the future, please contact:

Anthony Steed
Department of Computer Science
University College London
Gower Street
London WC1E 6BT
+44 (0) 20 7679 4435 *Tel*
Email: A.Steed@cs.ucl.ac.uk

Prof. Paul Sharkey
Department of Cybernetics
University of Reading
Reading RG6 6AY
Ph: (+44) 118 378 6704
Email: p.m.sharkey@reading.ac.uk

David Roberts
The Centre for Virtual Environments,
Business House,
University Road,
University of Salford,
Salford, Greater Manchester,
M5 4WT
+44 161 29 52926
Email:D.J.Roberts@salford.ac.uk

Dr John Rae
Roehampton University
Whitelands College
Holybourne Avenue,
London
SW15 4JD
Email: j.rae@roehampton.ac.uk
020 8392 3612

D.2 Ethics Form for Participants

Roehampton University

ETHICS BOARD

RESEARCH PARTICIPANT CONSENT FORM

Title and brief description of Research Project:

Eye catching: Supporting tele-communicational Eye-gaze in Collaborative Virtual Environments – copresent tasks with tracked eye-gaze

Thank-you for your considering taking in this research project. We are collecting experimental video film data of people undertaking collaborative activities. Some of these involve the manipulation of physical objects (e.g. assembling flat pack furniture or repairing an object that appears to be broken), others involve thinking towards a collaborative solution (e.g. undertaking a negotiation task), and others might just involve some kind of talk (e.g. an interview). The research is interested in how we engage in social interaction – from studying these interactions it is hoped that we can get a better understanding regarding the use of talk, gesture, gaze and such in interaction. This is one of a series of studies that aims to contribute our understanding of how people interact in “virtual environments”.

Please read through this information sheet and feel free to ask any questions. The investigators will answer any general questions, however the specific aspects regarding this study cannot be discussed with you until the end of the session.

In particular, the experiment you are taking part involves moving cubes around to complete a puzzle under the direction of two other people. Once you have completed filming you will be invited to delete or identify for deletion any parts of the film that you wish (you can delete the entire film if you would like to). The film will be analyzed to explore features of the interaction so as to provide a better understanding about how people interact. Thus we might look at how gestures are used, how we co-ordinate gaze and talk, how movement is integrated into activity sequences or how intonation shapes are used and responded to – anything which helps deepen our understanding about how we interact with one another.

You will be asked to wear eye-tracking glasses. The wearing of such glasses can cause discomfort to some participants, especially if used for a prolonged time. In this study you will use eye-tracking glasses for periods of no more than about one hour, but you should let us know if you get any discomfort from using them, and you can stop the experiment at any time if you experience some discomfort.

- You will be asked to read, understand and sign a **Consent Form**. If you sign it, the study will continue with your participation. **Note that you can withdraw at any time without giving any reasons.**
- You will be asked to complete a number of questions online and/or on paper, so that we can try to understand your responses during the study.
- After the task, we may ask you complete a questionnaire about your experience.
- Finally there will be a short discussion with the experimenters about your experiences. During this time, you might be audio or video taped.

Please note, you can withdraw from the study or delete any part of the film which you appear at any moment. If you have any remaining questions concerning the research that you have been involved in please contact either Dr Paul Dickerson, Dr John Rae or Estefania Guimaraes using the contact details below.

Name and status of Investigators:

Dr Paul Dickerson (p.dickerson@roehampton.ac.uk 020 8392 3613)

Dr John Rae (j.rae@roehampton.ac.uk 020 8392 3612)

Estefania Guimaraes E.Guimaraes@roehampton.ac.uk (020 8392 3000 ext 4515)

Please read the consent form and feel free to decline entirely or only agree to those aspects that you are completely happy with.

PART 1. CONSENT FOR YOU TO BE FILMED FOR RESEARCH PURPOSES.

The study involves a video recording being made of you. Some of the data recorded may involve interactions with the researcher(s). You can leave the scene of the recording or request that the filming stops at any stage.

You may withdraw your consent at any time before, during or after the filming. You can request the deletion of any part of the film in which you appear.

We would like to make transcripts of the talk and actions that happen in the video and store them for research purposes. They are likely to be of research value long into the future.

Are you happy to be filmed as part of this project?

yes / no (if "yes" please sign) ________________________

Are you happy to be fitted with eye-tracking equipment in order to record your gaze responses? Are you happy to have your eye movements tracked while taking part in the study?

yes / no (if "yes" please sign) ________________________

Consent Statement:

I agree to take part in this research, and am aware that I am free to withdraw at any point. I understand that the information I provide will be treated in confidence by the investigator although I understand that if a risk of serious harm arises that the researchers may need to take appropriate action. I understand that my identity will be protected in the publication of any findings unless I unless I specifically choose for my the recording, or stills from it, to be seen by other people in part 2 or part 3 below.

yes / no (if "yes" please sign) ________________________

Name: ________________________

Please note: if you have a concern about any aspect of your participation, please raise this with the investigator, or with the Dean of School who is

Name: Mr Michael Barham
Contact Address: School of Human and Life Sciences; Roehampton University, Whitelands College, Holybourne Avenue, LONDON, SW15 4JD
Direct Phone No: 020 8392 3620 Email: m.barham@roehampton.ac.uk

PART 2. CONSENT FOR FILM AND SOUNDTRACK TO BE USED FOR RESEARCH PURPOSES.

Transcripts of your talk and actions

1. Do you give permission for written transcripts of your talk and actions to be to be presented at conference & workshop presentations?

yes / no (if "yes" please sign) ___________________

2. Do you give permission for written transcripts of you talk and actions appear in academic and professional publications (such as journals, including electronic and paper versions)?

yes / no (if "yes" please sign) ___________________

Photo stills of you

3. Do you give permission for still photographs of you to be presented in conference & workshop presentations?
yes / no (if "yes" please sign) ___________________

4. Do you give permission for still photographs of you to be presented in research and educational publications?
yes / no (if "yes" please sign) ___________________

Audio fragments

5. Do you give permission for audio fragments involving you to be presented in conference & workshop presentations?
yes / no (if "yes" please sign) ___________________

6. Do you give permission for audio fragments involving you to be presented in research and educational publications?
yes / no (if "yes" please sign) ___________________

Video fragments

7. Do you give permission for video fragments involving you to be presented in conference & workshop presentations?
yes / no (if "yes" please sign) ___________________

8. Do you give permission for video fragments involving you to be presented in research and educational publications?
yes / no (if "yes" please sign) ___________________

9. Do you give permission for video recordings showing what your eye is tracking to be presented in conference & workshop presentations?
yes / no (if "yes" please sign) ____________________

10. Do you give permission for video recordings showing what your eye is tracking to be presented in research and educational publications?
yes / no (if "yes" please sign) ____________________

PART 3. DATA TO BE SHARED WITH OTHER RESEARCHERS

1 .Do you give permission for transcripts involving you to be shared with other researchers?

yes / no (if "yes" please sign) ____________________

2. Do you give permission for audio recordings involving you to be shared with other researchers?

yes / no (if "yes" please sign) ____________________

3. Do you give permission for photo stills involving you to be shared with other researchers?

yes / no (if "yes" please sign) ____________________

4. Do you give permission for video recordings involving you to be shared with other researchers?

yes / no (if "yes" please sign) ____________________

5. Do you give permission for video recordings showing what your eye is tracking to be shared with other researchers?

yes / no (if "yes" please sign) ____________________

6. Do you give permission for your data to be used by computer scientists in the design of computer programmes that simulate general, non-personalised human behaviour?

yes / no (if "yes" please sign) ____________________

PART 4. OTHER QUESTIONNAIRES

In addition to taking part in a video-recorded task, we would also be interest in
You completing a personality questionnaire before the study and then completing a questionnaire about your experience after the interaction. We are also interested in discussing your experience in a short interview, which we would like to audio or video record.

Date: ______________________________

D.3 Extended Conversation Analysis

IA3 [02a.Static] Participant "Martin"

```
1     R:    >°can you can you°< pick up
            [R                ][In front of R]

2           the cube I am looking at
            [R                        ]

3     M:    err:::: (.) which errmm (.)
            [S              ][Cube near S]

4           is that this one (- - - - - - - -)
            [R               ][cube near S][R ]

5     S:    err no (.) er no
            [S ][R ] [other cubes]

6     M:    where are you looking
            [scanning cubes][S  ]

7           (.) oh across the room
            [at cubes across the room]

8           (- - - - - - - - -)
            [at cubes across room]

9     M:    err:::: this one
            [Scans cubes][fixes on one cube]

10          (- - - -)
            [moves to R]

11    S:    err (.) no
            [R        ]

12          (- - - - -)
            [Walks behind R's avatar]
              [S - back of R's avatar]

13    M:    hang on (- - - - - - - - -)
            [Back of R] [S      ]

14          Oh (- - - -)
            [at specific cube]

15          this one up here
            [Cube]

16          (- - - - - - - - - - 1.0 -)
            [Cube                     ]

17    S:    yeah should be (the other)one
            [Cube                    ]

18          yeh the one yeah that's it
            [Cube
                   [moves cube       ]

19          >got it<
            [Cube  ]
            [moving cube]
```

In extract IA3 M makes three attempts to identify the correct cube (lines 4, 9 and 15). In attempt one (line 4) both Stephen and Robert have been gazed at prior to the participant, Martin, producing a candidate response regarding the correct box. In the second attempt Robert and Stephen are looked at (albeit briefly) before the boxes are scanned and a second box identified. Following these two failed attempts Martin now makes a somewhat more dramatic move – he moves towards and behind Robert’s avatar and looks from that visual perspective towards the cubes at the other side of the CAVE. After gazing briefly at Stephen, Martin gazes at a box and subsequently manipulates the correctly identified box as required.

For the current analysis it is worth noting three things about extract IA3. First, the participant (Martin) appears to take into account where others are gazing. That is, guesses are produced after gazing at the facial region of co-present others who are involved in selecting the box to be manipulated. Second, in IA3 *both* co-present others (Robert and Stephen) were gazed at prior to Martin making a box choice. Third, and perhaps most important Martin walks behind Robert’s avatar and gazes at the boxes from the (body) position of Robert’s avatar (lines 12 and 13). In this way the participant has made use of not just one source but several available sources of information regarding the identity of the designated cube. In the course of this the participant has made full use of the avatars’ spatial position and perspectives in seeking to identify the correct box.

INTERACTION ANALYSIS

```
++++++++++++++++++++++++++++++++++++++++++++
```

Figure {IA} [02.b.Tracked] Participant: “Tim”

```
01  S:   [(Timothy)       ]
         [ assembled cubes]

02  S:   [So [(.) m- m:y ] [turn
         [S  [ S, cubes, ] [S

03  T:   [Okay=
         [S

                           [turns and looks at cube
04  S:   =[Can you pick up [this cube where
          [S

05       [I’m looking at [now
         [S              [designated cube

06  T:   [Is that this [one      ((puts cursor on cube))
         [cube         [S

07  S:   [Yes.                   ((T moves cube))
         [cube
+++++++++++++++++++++++++++++++++++++++++++++++
```

```
++++++++++++++++++++++++++++++++++++++++++
```

Figure {IA2} [01.b.Model] Participant: “Owen”

```
01  R:   [Okay how about er you the pick cube I’m looking][at now]
         [  Gaze at source cubes and assembled cubes     ][  S   ]

02       ([- - -] - -[ - - - - ]- 1 - -[ - - - ]- - [- - - 2])
          [  S  ]    [R        ]       [  S    ]    [  R    ]

         ((looks from one avatar to the other))

03  R:   [Er: It’s Robert here speaking]
         [            R                ]

04  O:   [Yeah      ]
    S:   [Blue guy  ]
         [   R      ]

05  O:   [Pickup the one you’re] lookin’ [at okay     ]
         [   R                 ]         [source cubes]

06  R:   Yeah

07       (---------1)           ((moves a cube))

08  R:   er No not that one

09  O:   Oh Okay sorry          ((moving cube back))
```

```
10        (---------1)           ((replaces cube))

11  O:    [Let me try again] ((moves v. close to R)
          [     R         ]

12  O:    This one?

13        (------)               ((picks and moves cube))

14  R:    Yeah
+++++++++++++++++++++++++++++++++++++++
```

This section aims to complement the quantitative behavioural analysis through identifying specific practices and resources that the confederates and participants employ. Figure {ia1} shows an example of how a case of a correct and unproblematic cube identification unfolds interactionally. Here Stephen summons the participant's attention (line 1), eliciting his gaze (in line 2), and produces a pre-instruction (line 2) which receives a go-head for the participant (line3). When Stephen gives his instruction (line 4) Tim is gazing at him, he turns and looks at the designated cube as he produces the deictic term “this” (line4). Timothy follows Stephen's gaze and looks straight at the correct cube and places the cursor on it (line 6). At the same time he asks for confirmation, briefly returning his gaze to Stephen at turn completion. Stephen gives confirmation and Timothy starts to move the correct cube. By contrast, Figure {ia2} shows an example of a case where the participant selects the wrong cube, in addition it provides an example of repair-work to establish which avatar is providing instructions. Robert issues an instruction (line 1) but during the 2 second silence that follows, Owen looks from Robert to Stephen and back. Robert establishes that he is speaking (line 3) and reissues his instruction. As he does this, Owen is at first looking at him and then looks to the source cubes (line 5). Owen picks and moves a cube (line 7) but gets feedback from Robert that his choice is incorrect (line 8). So although Owen has been gazing at Robert throughout lines 3 -5 he was unable to identity the cube that Robert was looking at. Owen moves close to Robert to examine his gaze in line 11 gets the correct cube on a second attempt.

Based on the examination of such details the following observations can be made about the way in which the participants and confederates interact.

Movement as a resource for designation. Proximity to cubes seems to be very helpful in determining the cube being looked at correctly. The confederates made extensive use of the possibility of being able to move and get very near to the cubes being looked at or positioned themselves in front of the cubes being designated.

Participant's use of movement. In fig ia2 we noted that the participant moves close to Roberts face. This is one category of a number forms of locomotion that participants engage in. In a small number of cases participants look at the avatar's face and then take up an *opposite* position behind the cubes 3b, 5a (tracked), 5 b (static) -- * 5 b (1^{st} cube) in order to see what the avatar is gazing at. However a more common strategy was taking a position similar to the one of the avatar – standing beside them (or even inside them), presumably to see the avatar's point of view. In one case (2a, static) the participant first tries to figure out where Robert is looking at, but as he fails twice he goes behind Robert's avatar. <might offer transcript of this>

Reports of Problems with following avatar's gaze Difficulties in the identification of the cube being looked at frequently came to the surface of the interaction with participants vocalizations consists such as *questions* (e.g. "Are you looking at it now?" - 1b); *complaints* ("You are looking up! Not down." – 10a) or *accounts* ("It is a little difficult for me because your eyes are moving" – 6b). There were very much more common in interactions involved modelled gaze.

Overall, participants frequently – but not always – sought information about the cube being designated by gazing at the avatar's face. Confederates, frequently made use of the possibility of locomotion with the cave as a resource for designating a cube. On the occasions when participants moved, they engaged in an unexpected behaviour: moving alongside the avatar in order to examine the scene from the avatar's perspective.

Appendix E

Materials for the Truth and Deception Experiments

E.1 Consent Form for Participants (Experiment 1)

University College London
Gower Street
London WC1E 6BT UK

Department of Computer Science
William Steptoe

W.Steptoe@cs.ucl.ac.uk

ID

PROJECT: EYE CATCHING
Investigators: William Steptoe, Anthony Steed

To be completed by volunteers:
We would like you to read the following questions carefully.

Have you read the information sheet about this study? YES/NO
Have you had an opportunity to ask questions and discuss this study? YES/NO
Have you received satisfactory answers to all your questions? YES/NO
Have you received enough information about this study? YES/NO
Which investigator have you spoken to about this study? ………..
Do you understand that you are free to withdraw from this study?

- At any time YES/NO
- Without giving a reason for withdrawing YES/NO

Do you understand and accept the risks associated with the use of virtual reality equipment? YES/NO
Do you agree to take part in this study? YES/NO
Do you agree to be audio taped? YES/NO
Do you agree to have your head and eye motion tracked ? YES/NO
Do you agree for the above motion to be mapped to a virtual character along with your audio recording for other confidential experimental participants to view? YES/NO

I certify that I do not have epilepsy.

I certify that I will not be driving a car, motorcycle, bicycle, or use other types of complex machinery that could be a danger to myself or others, within 3 hours after the termination of the study.

Signed…………………………………Date……………………………...

Name in block letters.……………………………………………………...

Investigator.……………………………………………………………..

In case you have any enquiries regarding this study in the future, please contact:

William Steptoe
Department of Computer Science
University College London
Gower Street
London WC1E 6BT

+44 (0) 20 7679 7215 *Tel*
+44 (0) 20 7387 1397 *Fax*

W.Steptoe@cs.ucl.ac.uk

E.2 Information Sheet for Participants (Experiment 1)

University College London
Gower Street
London WC1E 6BT UK

Department of Computer Science
William Steptoe

W.Steptoe@cs.ucl.ac.uk

TRUTH/DECEPTION EXPERIMENT
INFORMATION SHEET FOR PARTICIPANTS

General

Thank you for participating in our study. This is one of a long series of studies into understanding the responses of people within virtual environments. Please read through this information sheet and feel free to ask any questions. The experimenters will answer any general questions, however the specific aspects regarding this study cannot be discussed with you until the end of the session.

Experiment Instructions

In this particular study you will be taking part in a two-way interaction with another person (confederate). During the interaction, the confederate will ask you a series of questions. The questions are divided into six groups, with around ten questions in each group, and therefore around 60 in total. Some questions can be answered with a "yes/no" response, while others are more open-ended.

You will be given an additional sheet which tells you how you must answer for each of the six question groups. For each group, you will be instructed either to answer **truthfully** to ALL questions in the group, or to **lie plausibly** when answering ALL questions in the group. The confederate will be **unaware** of your truth or lie instructions, and so will not know if you are telling the truth or lying, and you must not tell him.

The total question period will last approximately 15-20 minutes.

Communication System

The interaction will be performed using **one** of the following three methods:

1. Virtual reality (VR) system.
   If you are selected for the VR system, the confederate will appear on your display as a virtual character or 'avatar'. Likewise, the confederate will see you as an avatar. The avatars' behavior is mapped to your real movements, so for instance, as you nod, your avatar will nod, and as you look down, so will your avatar.

2. Video-conferencing system.
   If you are selected for the video-conferencing system, the confederate will appear as a live video stream, and also you will appear to the confederate as a live video stream.

3. Face-to-face.
   Finally, if you are selected for the face-to-face scenario, you will simply be seated opposite the confederate in the same room.

Eye Tracking

Before the experiment begins, you will be fitted with an eye tracker. The eye tracker will monitor the movements of your eyes. You will be seated during the experiment. You will be free to move your head as you would do in a normal conversation with someone. The eye tracker is designed to be as light and comfortable as possible to enable you to act as you would if it wasn't there!

Ethics

This study has been approved by the UCL ethics committee under project ID number 0432/001.

Please ask any questions that come to mind. Read and sign the **Consent Form**.

Information that we collect will never be reported in a way that specific individuals can be identified. Information will be reported in a statistical and aggregated way, and any verbal comments that you make, if written about in subsequent papers, will be presented anonymously.

When people use virtual reality systems, some people sometimes experience some degree of nausea. If at any time you wish to stop taking part in the study due to this or any other reason, please just say so and we will stop.

With any type of video equipment there is a possibility that an epileptic episode may be generated. This, for example, has been reported for computer video games.

PROCEDURES

- You will be asked to read, understand and sign a **Consent Form**. If you sign it the study will continue with your participation. **Note that you can withdraw at any time without giving any reasons.**
- You will be asked to complete an online questionnaire so that we can try to understand your responses during the study.
- You are asked to switch off mobile phones before the experiment.
- After the experiment you will complete an online questionnaire about your experience.
- Finally there will be a small discussion with the experimenters about your experiences. During this time, you may be audio taped.

- **Thank you** for your participation. Please do not discuss this study with others for about **three months**, since the study is continuing.
- Any other questions?

The whole study will take about one hour.

In case you have any enquiries regarding this study in the future, please contact:

William Steptoe
Department of Computer Science
University College London
Gower Street
London WC1E 6BT

+44 (0) 20 7679 7215 *Tel*
+44 (0) 20 7387 1397 *Fax*

W.Steptoe@cs.ucl.ac.uk

E.3 Technical Procedure (Experiment 1)

Machine	IP Address	Reference	Process	Path	Notes
Clustermaster	**128.16.13.169**	CM_1	trackd.exe -file EyeCVE.conf	C:\Program Files\trackd\bin	Should be
		CM_2	WorldServer.exe	C:\eyecatching\EyeCVE_svn\bin\Debug	lie_room.
		CM_3	aitor.bat	C:\eyecatching\EyeCVE_svn\bin\Debug	
		CM_4	gtalk		Should be
		CM_5	detector.bat	C:\eyecatching\EyeCVE_svn\bin\Debug	Enter Aito
		CM_6	Microphone		Turn on r
		CM_7	Call Fleming		Call truth
Slave1	**N/A**	SL1_1	eyecve1.bat	C:\eyecatching\eyecve_lie\Debug	
Slave2	**N/A**	SL2_1	spec.bat	C:\eyecatching\eyecve_lie\Debug	Repositio
		SL2_2	Fraps		F9, 50fps
Slave3	**N/A**	SL3_1	spec.bat	C:\eyecatching\eyecve_lie\Debug	Repositio
		SL3_2	Fraps - 50fps, audio		F9, 50fps
Slave4	**N/A**	SL4_1	eyecve4.bat	C:\eyecatching\eyecve_lie\Debug	
Eyecatcher	**128.16.14.69**	EC_1	Connect eye tracker		Yellow, R
		EC_2	viewpoint.exe	C:\Viewpoint-60	resize cal
		EC_3	mobileeyeserver.exe	C:\Viewpoint-60	Select VI
		EC_4	detector.bat	C:\Viewpoint-60	Enter Par
		EC_5	VNC server		Should be
		EC_6	Record data		Ctrl-U
Fleming	**128.16.3.58**	FL_1	detector.bat	E:\eyecatching\EyeCVE\bin\Debug	Enter Par
		FL_2	ViewPointClient.exe	E:\eyecatching\EyeCVE\bin\Debug	connect t
		FL_3	gtalk		User: trut
		FL_4	judy.bat (female participant)	E:\eyecatching\EyeCVE\bin\Debug	Repositio
		FL_5	don.bat (male participant)	E:\eyecatching\EyeCVE\bin\Debug	Repositio
		FL_6	VNC viewer to Eyecatcher		Passworc
		FL_7	Microphone		Turn on r
		FL_8	VPRN_Server.exe	Desktop	Turn on F
		FL_9	Calibrate in Viewpoint		Via VNC
		FL_10	Calibrate in EyeCVE		
		FL_11	Pick up call		gtalk

		Final Order	
e running already		CM_1	Set up CM
cfg in ..\data		CM_2	
		CM_3	
e running already. User: truthexp1, Pass: truthexp		CM_4	End
or in Name, uncheck 'Tracked Blinking and Pupil Dilation'		SL1_1	Set up slaves
eceiver and transmitter, check line-in levels		SL4_1	
exp2 from gtalk		SL2_1	
		SL3_1	End
		EC_1	Set up EC
		EC_2	
n to face participant, move cursor		EC_3	
, audio		EC_5	End
		FL_3	Set up FL
n to face participant, move cursor		FL_2	
, audio		FL_8	
		FL_4 or FL_5	
		FL_6	End
		CM_6	Set up Audio
ed, White, power		FL_7	
libration plane, calibrate, remove calibration plane, check blink threshold, check video settings		CM_7	
EWPOINT and unselect log		FL_11	
ticipant in name, perform min and max detection, o, n, m during interaction		CM_5	
e running already		FL_1	
		Participant Needed!	
		FL_9	Calibrate eyetracker
ticipant in name, used only for speech input, 180 threshold?		FL_10	
o 128.16.13.69 (eyecatcher)		EC_4	End
hexp2, Pass truthexp		Final Check	Audio, tracking, blink, VPX video
n		SL2_2	Record Data
n		SL3_2	
d: wsteptoe		EC_6	End
eceiver and transmitter, check line-in levels			
Fastrak			

E.4 NEO Five-Factor Inventory Questionnaire (Experiment 1)

NEO Five-Factor Inventory

Instructions:

Please read the intructions carefully before beginning. This questionnaire contains 60 statements. Read each statement carefully. For each statement choose the option with the response that best represents your opinion.

Choose "Strongly Disagree" if you *stronly disagree* or the statement is definitely false.

Choose "Disagree" if you *stronly disagree* or the statement is mostly false.

Choose "Neutral" if you are *neutral* on the statement, you cannot decide, or the statement is about equally true and false.

Choose "Agree" if you *agree* or the statement is mostly true.

Choose "Strongly Agree" if you *strongly agree* or the statement is definitely true.

Your Given ID number:	
1	**I am not a worrier.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
2	**I like to have a lot of people around me.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
3	**I don't like to waste my time daydreaming.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
4	**I try to be courteous to everyone I meet.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
5	**I keep my belongings clean and neat.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐

6	**I often feel inferior to others.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
7	**I laugh easily.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
8	**Once I find the right way to do something, I stick to it.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
9	**I often get into arguments with my family and co-workers.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
10	**I'm pretty good about pacing myself so as to get things done on time.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
11	**When I'm under a great deal of stress, sometimes I feel like I'm going to pieces.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
12	**I don't consider myself especially "light-hearted".** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
13	**I am intrigued by the patterns I find in art and nature.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
14	**Some people think I'm selfish and egotistical.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
15	**I am not a very methodical person.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐

16	**I rarely feel lonely or blue.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
17	**I really enjoy talking to people.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
18	**I believe letting students hear controversial speakers can only confuse and mislead them.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
19	**I would rather cooperate with others than compete with them.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
20	**I try to perform all the tasks assigned to me conscientiously.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
21	**I often feel tense and jittery.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
22	**I like to be where the action is.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
23	**Poetry has little or no effect on me.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
24	**I tend to be cynical and sceptical of other's intentions.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
25	**I have a clear set of goals and work toward them in an orderly fashion.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐

26	**Sometimes I feel completely worthless.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
27	**I usually prefer to do things alone.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
28	**I often try new and foreign foods.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
29	**I believe that most people will take advantage of you if you let them.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
30	**I waste a lot of time before settling down to work.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
31	**I rarely feel fearful or anxious.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
32	**I often feel as if I'm bursting with energy.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
33	**I seldom notice the moods or feelings that different environments produce.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
34	**Most people I know like me.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
35	**I work hard to accomplish my goals.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
36	**I often get angry at the way people treat me.**

	Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
37	**I am a cheerful, high-spirited person.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
38	**I believe we should look to our religious authorities for decisions on moral issues.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
39	**Some people think of me as cold and calculating.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
40	**When I make a commitment, I can always be counted on to follow through.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
41	**Too often, when things go wrong, I get discouraged and feel like giving up.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
42	**I am not a cheerful optimist.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
43	**Sometimes when I am reading poetry or looking at a work of art, I feel a chill or wave of excitement.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
44	**I'm hard-headed and tough-minded in my attitude.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
45	**Sometimes I'm not as dependable or reliable as I should be.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐

46	**I am seldom sad or depressed.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
47	**My life is fast-paced.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
48	**I have little interest in speculating on the nature of the universe or the human condition.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
49	**I generally try to be thoughtful and considerate.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
50	**I am a productive person who always gets the job done.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
51	**I often feel helpless and want someone else to solve my problems.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
52	**I am very active person.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
53	**I have a lot of intellectual curiosity.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
54	**If I don’t like people, I let them know it.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
55	**I never seem to be able to get organised.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐

56	**At times I have been so ashamed I just wanted to hide.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
57	**I would rather go my own way than be a leader of others.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
58	**I often enjoy playing with theories or abstract ideas.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
59	**If necessary, I am willing to manipulate people to get what I want.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐
60	**I strive for excellence in everything I do.** Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree ☐

SUBMIT

E.5 POMS Questionnaire (Experiment 1)

Profile of mood states

Below is a list of words that describe feelings people have. Please read each one carefully, then circle the number which best describes the extent to which you have this feeling **now**.

		Not at all	A little	Moderately	Quite a lot	Extremely
1.	Tense	0	1	2	3	4
2.	Worn out	0	1	2	3	4
3.	Angry	0	1	2	3	4
4.	Lively	0	1	2	3	4
5.	Confused	0	1	2	3	4
6.	Shaky	0	1	2	3	4
7.	Sad	0	1	2	3	4
8.	Grouchy	0	1	2	3	4
9.	Active	0	1	2	3	4
10.	On edge	0	1	2	3	4
11.	Annoyed	0	1	2	3	4
12.	Energetic	0	1	2	3	4
13.	Hopeless	0	1	2	3	4
14.	Relaxed	0	1	2	3	4
15.	Resentful	0	1	2	3	4
16.	Unworthy	0	1	2	3	4
17.	Uneasy	0	1	2	3	4
18.	Can't concentrate	0	1	2	3	4
19.	Fatigued	0	1	2	3	4
20.	Listless	0	1	2	3	4
21.	Nervous	0	1	2	3	4
22.	Lonely	0	1	2	3	4
23.	Muddled	0	1	2	3	4
24.	Furious	0	1	2	3	4

		Not at all	A little	Moderately	Quite a lot	Extremely
25.	Cheerful	0	1	2	3	4
26.	Exhausted	0	1	2	3	4
27.	Gloomy	0	1	2	3	4
28.	Sluggish	0	1	2	3	4
29.	Weary	0	1	2	3	4
30.	Bewildered	0	1	2	3	4
31.	Alert	0	1	2	3	4
32.	Bitter	0	1	2	3	4
33.	Efficient	0	1	2	3	4
34.	Forgetful	0	1	2	3	4
35.	Guilty	0	1	2	3	4
36.	Vigorous	0	1	2	3	4

E.6 Experiment 1 Questions (Experiment 1)

General Questions 1

You are now to be asked the first of two sets of general information questions. Be sure to answer questions the way you are instructed on the information sheet. Please let me know when you are ready.

1. What is your first name?
2. What is your mother's first name?
3. *Are you a member of any societies?*
4. *Do you smoke?*
5. *Have you ever been arrested?*
6. What is the most important quality you look for in a mate?
7. *Have you ever cheated on an exam?*
8. What was the title of the last book you read?
9. Tell me more about the book.

OK, that's the end of this stage of questions. Please look at your instructions for the next stage of questions and let me know when you are ready.

General Questions 2

You are now to be asked the second of two sets of general information questions. Be sure to answer questions the way you are instructed on the information sheet. Please let me know when you are ready.

1. What is your age?
2. *Do you have any birthmarks?*
3. What year were you born?
4. *Do you have siblings?*
    5. How old is the oldest child in your family?
6. *Do you have any pets?*
7. What is your natural eye colour?
8. *Have you ever visited Spain?*
9. Where was the last foreign country you visited?
10. Tell me some more about the trip.

OK, that's the end of this stage of questions. Please look at your instructions for the next stage of questions and let me know when you are ready.

Last Sunday

You are now to be asked a set of questions about the previous Sunday. Think about what you did this past **Sunday**. Take as much time as you need, and let me know when you feel ready to answer questions about this. Be sure to answer questions the way you are instructed on the information sheet.

1. About what time did you awaken?
2. *Did you go to church on Sunday?*
    3. About what time did the service begin?
    4. What was the sermon about?
5. What did you eat for your main meal?
6. Who was the cook?
7. Did you do any work or study?
    8. How many hours did you do?
9. What did you do in the evening?
10. What time did you go to bed?

OK, that's the end of this stage of questions. Please look at your instructions for the next stage of questions and let me know when you are ready.

Last Saturday

You are now to be asked a set of questions about the previous Saturday evening. Think about what you did this past **Saturday evening**. Take as much time as you need, and let me know when you feel ready to answer questions about this. Be sure to answer questions the way you are instructed on the information sheet.

1. Around 9 PM Saturday night, where were you?
2. What were you doing?
3. *Were you with others?*
    4. How many people were you with?
    5. *Were you with someone you are attracted to romantically?*
6. *Were you drinking alcohol?*
7. *At the time, were you drunk?*
8. *Did you enjoy what you were doing?*
9. What was the last thing you ate that evening?
10. What time did you go to bed?

OK, that's the end of this stage of questions. Please look at your instructions for the next stage of questions and let me know when you are ready.

High-School Senior Year

You are now to be asked a set of questions about your final year of high-school. Think back to your final year in high school. Think about who your teachers were, who your friends were, and so on. Think also about how good a student you were. Take as much time as you need, and let me know when you feel ready to answer questions about this. Be sure to answer questions the way you are instructed on the information sheet.

1. What's the name of the school you attended?
2. Which subjects were you studying?
3. What was the name of your <subject> teacher?
4. Approximately how old was this teacher?
5. *As a senior, did you have a best friend?*
    6. Was your best friend a male?
    7. What was your best friend's first name?
8. *Were you well-behaved as a student?*
9. *Were you disappointed with your final grades?*
10. *During your senior year, did you fail any classes?*
    11. Which classes did you fail?
12. *Do you still keep in contact with anyone from school?*

OK, that's the end of this stage of questions. Please look at your instructions for the next stage of questions and let me know when you are ready.

University/Employment History

You are now to be asked a set of questions about your university and employment history. Think about your employment history or university history. Also, if you are currently employed or studying, think about the main position you have. Let me know when you feel ready to answer questions about these. Be sure to answer questions the way you were told to outside and answer as quickly as you can.

1. Are you currently employed or are you studying?

    2. What do you do?

    3. When did this position start?

4. Have you ever lied on a job application?

    5. What did you lie about?

6. *Do you consider yourself a good <employee/student>?*

7. What job have you had that you enjoyed the most?

8. When you have finished studying, what do you want to do professionally?

9. Why does this attract you?

10. *Do you think your degree major relates to your professional interests?*

OK, that's the end of this stage of questions. Please look at your instructions for the next stage of questions and let me know when you are ready.

E.7 Veracity Orders (Experiment 1)

TRUTH / DECEPTION EXPERIMENT

You will be asked a total of six stages of questions ordered as shown in the table below. Each stage is around 10 questions.

When answering each stage of questions, you must either **tell the truth** OR **lie as convincingly as possible** as instructed below.

Question Stage	You must:
General 1	Tell the Truth
General 2	Lie
Last Sunday	Tell the Truth
Last Saturday Evening	Lie
High-School	Tell the Truth
Employment/University History	Lie

TRUTH / DECEPTION EXPERIMENT

You will be asked a total of six stages of questions ordered as shown in the table below. Each stage is around 10 questions.

When answering each stage of questions, you must either **tell the truth** OR **lie as convincingly as possible** as instructed below.

Question Stage	You must:
General 1	Lie
General 2	Tell the Truth
High-School	Lie
Employment/University History	Tell the Truth
Last Sunday	Lie
Last Saturday Evening	Tell the Truth

TRUTH / DECEPTION EXPERIMENT

You will be asked a total of six stages of questions ordered as shown in the table below. Each stage is around 10 questions.

When answering each stage of questions, you must either **tell the truth** OR **lie as convincingly as possible** as instructed below.

Question Stage	You must:
Last Sunday	Tell the Truth
Last Saturday Evening	Lie
General 1	Tell the Truth
General 2	Lie
High-School	Tell the Truth
Employment/University History	Lie

TRUTH / DECEPTION EXPERIMENT

You will be asked a total of six stages of questions ordered as shown in the table below. Each stage is around 10 questions.

When answering each stage of questions, you must either **tell the truth** OR **lie as convincingly as possible** as instructed below.

Question Stage	You must:
Last Sunday	Lie
Last Saturday Evening	Tell the Truth
High-School	Lie
Employment/University History	Tell the Truth
General 1	Lie
General 2	Tell the Truth

TRUTH / DECEPTION EXPERIMENT

You will be asked a total of six stages of questions ordered as shown in the table below. Each stage is around 10 questions.

When answering each stage of questions, you must either **tell the truth** OR **lie as convincingly as possible** as instructed below.

Question Stage	You must:
High-School	Tell the Truth
Employment/University History	Lie
General 1	Tell the Truth
General 2	Lie
Last Sunday	Tell the Truth
Last Saturday Evening	Lie

TRUTH / DECEPTION EXPERIMENT

You will be asked a total of six stages of questions ordered as shown in the table below. Each stage is around 10 questions.

When answering each stage of questions, you must either **tell the truth** OR **lie as convincingly as possible** as instructed below.

Question Stage	You must:
High-School	Lie
Employment/University History	Tell the Truth
Last Sunday	Lie
Last Saturday Evening	Tell the Truth
General 1	Lie
General 2	Tell the Truth

E.8 Experimenter's Notes Sheet (Experiment 1)

1 | Participant Notes

Calibration:

Blinking:

Pupil Dilation:

Notes:

Post:

E.9 Extract of Log File (Experiment 1)

Stage	StageName	T/F	Time	X-Gaze	Y-Gaze	ROI	Head	Hand	Pupil Size	Blinks	Talk	Hit Point
4	General 2	T	317.6061	1.6363	0.3127	44	0.0228	0.0228	0.634433	BLINK		don_body-FACES100
4	General 2	T	317.6227	1.697	0.1638	44	0.0225	0.0225	0.634433			don_body-FACES100
4	General 2	T	317.6394	1.7238	0.0656	44	0.0326	0.0212	0.634433			don_body-FACES100
4	General 2	T	317.6561	1.7238	0.0656	44	0.0706	0.0244	0.634433		TALK	don_body-FACES100
4	General 2	T	317.6727	1.4403	0.6962	44	0.0581	0.035	0.634433			don_body-FACES100
4	General 2	T	317.6894	1.4293	0.7335	44	0.056	0.035	0.634433			don_body-FACES100
4	General 2	T	317.7061	1.6444	0.3136	46	0.0512	0.0325	0.634433			don_body-FACES100
4	General 2	T	317.7227	1.6444	0.3136	46	0.0871	0.0451	0.634433			don_body-FACES100
4	General 2	T	317.7394	1.6444	0.3136	46	0.0853	0.0407	0.634433		TALK	don_body-FACES100
4	General 2	T	317.756	0.8523	1.4074	46	0.0237	0.0229	0.634433			don_body-FACES100
4	General 2	T	317.7727	0.8523	1.4074	46	0.0811	0.0397	0.634433			don_body-FACES100
4	General 2	T	317.7893	0.8523	1.4074	46	0.1632	0.0745	0.634433			don_body-FACES100
4	General 2	T	317.806	0.8523	1.4074	44	0.1773	0.0708	0.634433			don_body-FACES100
4	General 2	T	317.8227	0.8523	1.4074	44	0.1071	0.0574	0.634433			don_body-FACES100
4	General 2	T	317.8394	0.8523	1.4074	44	0.1347	0.0798	0.634433			don_body-FACES100
4	General 2	T	317.856	0.8523	1.4074	44	0.0851	0.0444	0.634433		TALK	don_body-FACES100
4	General 2	T	317.8727	0.8523	1.4074	44	0.1508	0.0667	0.634433			don_body-FACES100
4	General 2	T	317.8893	0.7497	0.1719	44	0.1723	0.1076	0.634433			don_body-FACES100
4	General 2	T	317.906	0.7522	0.1362	46	0.1671	0.131	0.724425		TALK	don_body-FACES100
4	General 2	T	317.9226	0.7428	0.1706	46	0.1683	0.1218	0.724425			don_body-FACES100
4	General 2	T	317.9393	0.742	0.1793	46	0.1689	0.1237	0.758975			don_body-FACES100
4	General 2	T	317.9559	0.739	0.1876	46	0.1686	0.1245	0.758975			don_body-FACES100
4	General 2	T	317.9726	0.7371	0.2027	46	0.1689	0.1226	0.758975			don_body-FACES100
4	General 2	T	317.9893	0.7397	0.2049	25	0.1682	0.1233	0.758975			chair02-FACES
4	General 2	T	318.0068	0.7359	0.2176	25	0.1687	0.1254	0.758975			chair02-FACES
4	General 2	T	318.0226	0.7359	0.2176	25	0.1632	0.0745	0.758975			chair02-FACES
4	General 2	T	318.0393	0.7408	0.2378	25	0.1667	0.1291	0.758975	BLINK		chair02-FACES
4	General 2	T	318.0559	0.736	0.2475	46	0.1699	0.1304	0.737407			don_body-FACES100
4	General 2	T	318.0726	0.7326	0.2549	46	0.1684	0.1304	0.7154			don_body-FACES100
4	General 2	T	318.0892	0.7332	0.2551	46	0.1689	0.1304	0.7154			don_body-FACES100
4	General 2	T	318.1059	0.7324	0.2618	21	0.1688	0.1324	0.7154			B6-FACES
4	General 2	T	318.1225	0.727	0.2749	21	0.1694	0.1327	0.7154			B6-FACES
4	General 2	T	318.1396	0.7278	0.2827	21	0.1693	0.1355	0.704777			B6-FACES

E.10 Consent Form for Participants (Experiment 2)

University College London
Gower Street
London WC1E 6BT UK

Department of Computer Science
William Steptoe

W.Steptoe@cs.ucl.ac.uk

ID

PROJECT: EYE CATCHING
Investigators: William Steptoe, Anthony Steed

To be completed by volunteers:
We would like you to read the following questions carefully.

Have you read the information sheet about this study?	YES/NO
Have you had an opportunity to ask questions and discuss this study?	YES/NO
Have you received satisfactory answers to all your questions?	YES/NO
Have you received enough information about this study?	YES/NO
Which investigator have you spoken to about this study?	…………
Do you understand that you are free to withdraw from this study?	
• At any time	YES/NO
• Without giving a reason for withdrawing	YES/NO
Do you agree to take part in this study?	YES/NO

I certify that I do not have epilepsy.

Signed…………………………………Date……………………………...

Name in block letters.…………………………………………………...

Investigator.……………………………………………………………..

In case you have any enquiries regarding this study in the future, please contact:

William Steptoe
Department of Computer Science
University College London
Gower Street
London WC1E 6BT

+44 (0) 20 7679 7215 *Tel*
+44 (0) 20 7387 1397 *Fax*

W.Steptoe@cs.ucl.ac.uk

E.11 Information Sheet for Participants (Experiment 2)

University College London
Gower Street
London WC1E 6BT UK

Department of Computer Science
William Steptoe

W.Steptoe@cs.ucl.ac.uk

TRUTH/DECEPTION EXPERIMENT 2
INFORMATION SHEET FOR PARTICIPANTS

General

Thank you for participating in our study. Please read through this information sheet and feel free to ask any questions. The experimenters will answer any general questions, however the specific aspects regarding this study cannot be discussed with you until the end of the session.

Experiment Instructions

In this particular study you will be viewing a series of video-clips of computer-generated characters (avatars). In each video, the avatar will be seen to answer a series of questions, and the avatar will move and talk. These videos were captured during a previous experiment in which people were put in an 'interview' scenario and asked a number of questions. Certain movement and speech of these people was recorded and 'mapped' onto an avatar. So for example, if one of the people shakes their head and says "no", the avatar will be seen to do the same.

Your task today is to make judgments for each of the clips. You will be asked whether you think the person is telling the truth or lying, and how confident you are about this rating. You will also be asked how interested in the interaction you think the person is and again how confident you are in your answer.

Toady you will be shown a total of 24 clips. 16 of them will feature video and audio, and on 8 of them, the video will be turned off and you will only hear the audio.

Each clip lasts for around 1.5 minutes, so the experiment will last for around 40 minutes.

Ethics

This study has been approved by the UCL ethics committee under project ID number 0432/001.

Please ask any questions that come to mind. Read and sign the **Consent Form**.

Information that we collect will never be reported in a way that specific individuals can be identified. Information will be reported in a statistical and aggregated way, and any verbal comments that you make, if written about in subsequent papers, will be presented anonymously.

PROCEDURES

- You will be asked to read, understand and sign a **Consent Form**. If you sign it the study will continue with your participation. **Note that you can withdraw at any time without giving any reasons.**
- You are asked to switch off mobile phones before the experiment.
- **Thank you** for your participation. Please do not discuss this study with others for about **three months**, since the study is continuing.
- Any other questions?

In case you have any enquiries regarding this study in the future, please contact:

William Steptoe
Department of Computer Science
University College London
Gower Street
London WC1E 6BT

+44 (0) 20 7679 7215 *Tel*
+44 (0) 20 7387 1397 *Fax*

W.Steptoe@cs.ucl.ac.uk

E.12 Participant Response Sheet (Experiment 2)

1/24

Do you think the person is *telling the truth* or *lying*?

Always Lying	1	2	3	4	5	6	7	**Always Telling the Truth**

How *confident* are you in your answer above?

Very Unsure	1	2	3	4	5	6	7	**Very Sure**

How *engaged* in the interaction does the person appear to be?

Not at all Engaged	1	2	3	4	5	6	7	**Completely Engaged**

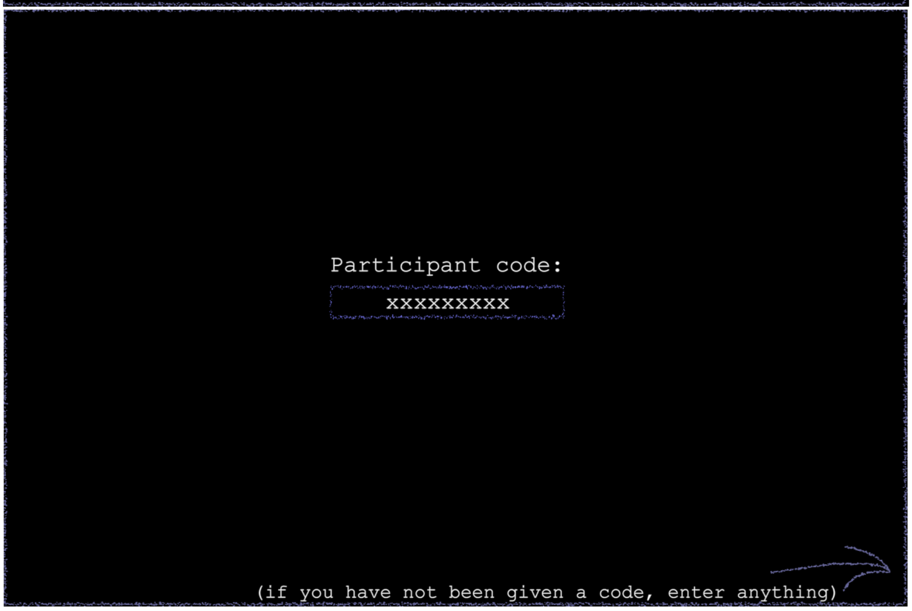
How *confident* are you in your answer above?

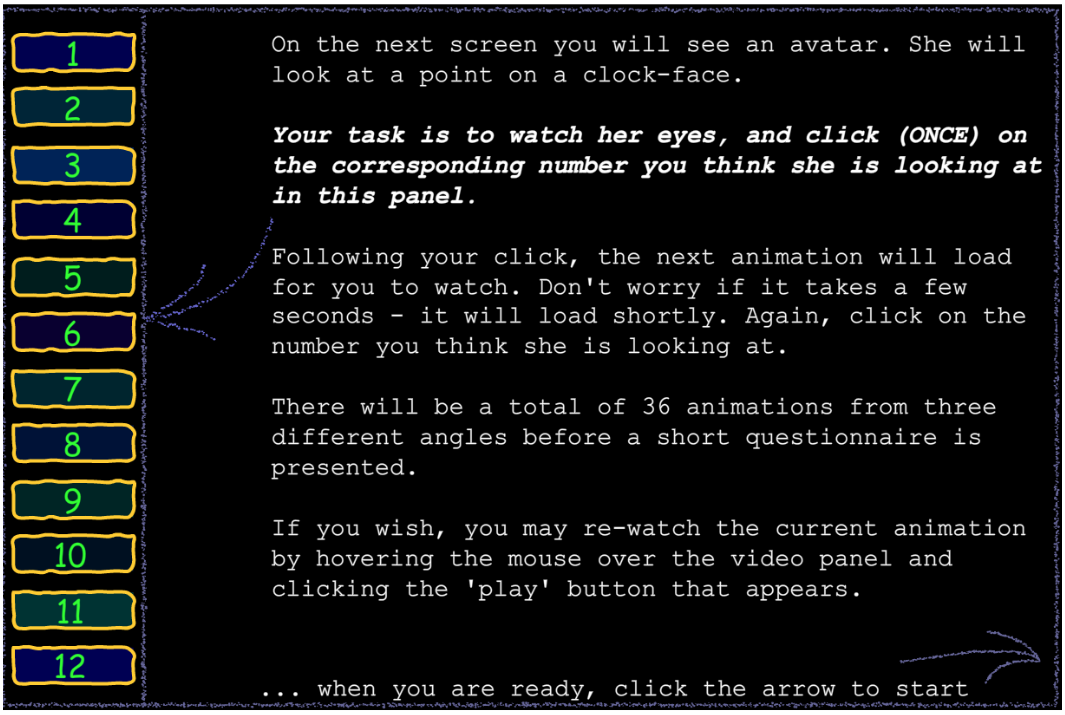
Very Unsure	1	2	3	4	5	6	7	**Very Sure**

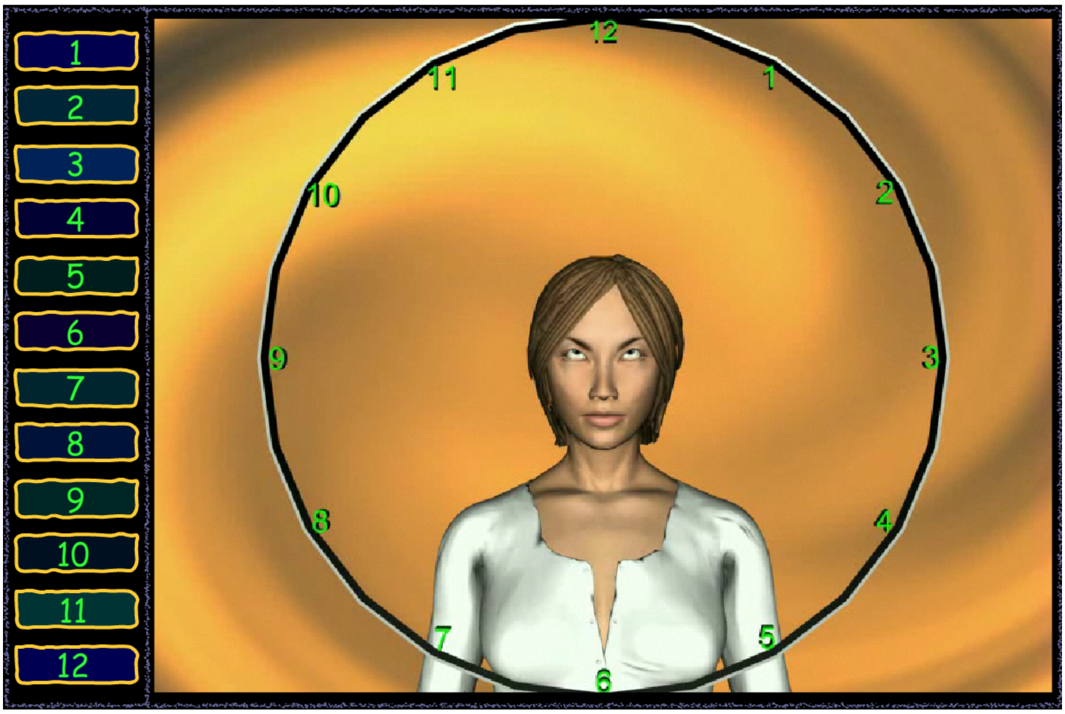
Appendix F

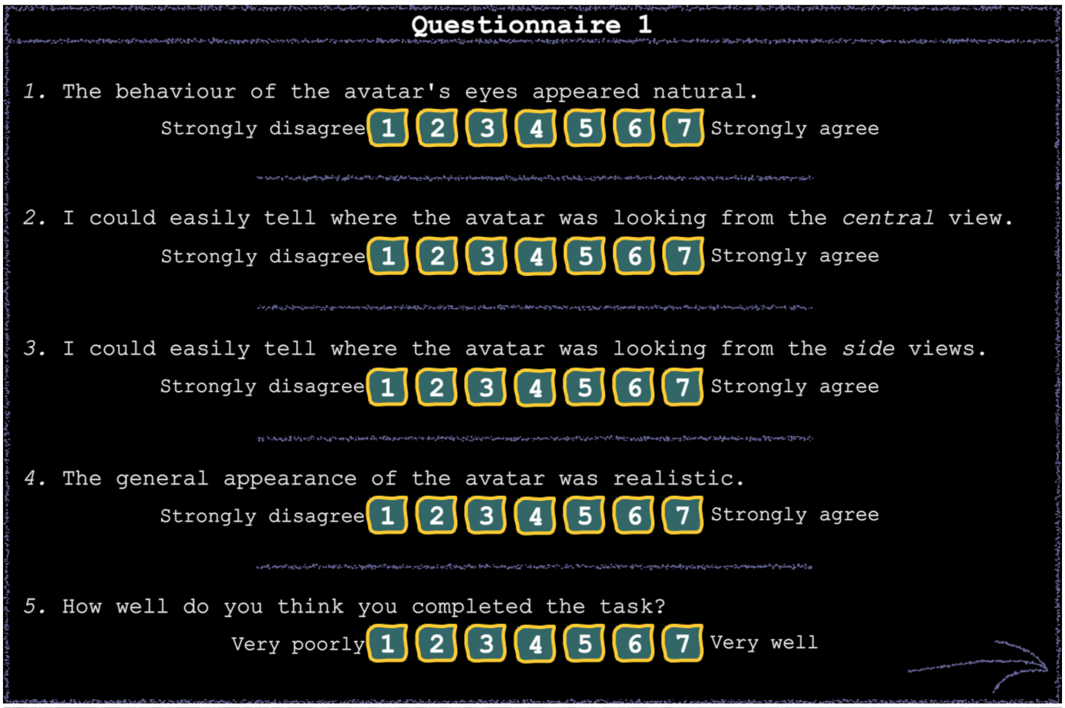
Materials for the Oculesic Models Experiments

F.1 Images from the Oculesic Behaviour Experiment




1
2
3
4
5
6
7
8
9
10
11
12
On the next screen you will see an avatar. She will look at a point on a clock-face.
Your task is to watch her eyes, and click (ONCE) on the corresponding number you think she is looking at in this panel.
Following your click, the next animation will load for you to watch. Don't worry if it takes a few seconds - it will load shortly. Again, click on the number you think she is looking at.
There will be a total of 36 animations from three different angles before a short questionnaire is presented.
If you wish, you may re-watch the current animation by hovering the mouse over the video panel and clicking the 'play' button that appears.
... when you are ready, click the arrow to start


1
2
3
4
5
6
7
8
9
10
11
12
12
1
2
3
4
5
6
7
8
9
10
11

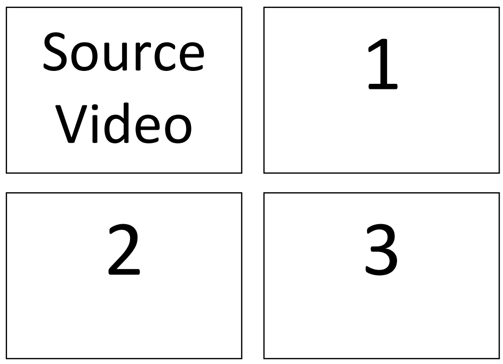

Questionnaire 1
1. The behaviour of the avatar's eyes appeared natural.
Strongly disagree 1 2 3 4 5 6 7 Strongly agree
2. I could easily tell where the avatar was looking from the central view.
Strongly disagree 1 2 3 4 5 6 7 Strongly agree
3. I could easily tell where the avatar was looking from the side views.
Strongly disagree 1 2 3 4 5 6 7 Strongly agree
4. The general appearance of the avatar was realistic.
Strongly disagree 1 2 3 4 5 6 7 Strongly agree
5. How well do you think you completed the task?
Very poorly 1 2 3 4 5 6 7 Very well


UCL
Thank you! You have completed all sections of the experiment.
Your responses have been submitted for analysis. If you have any
questions or comments, please email me by clicking here.
William Steptoe - w.steptoe@cs.ucl.ac.uk

F.2 Participant Response Sheet for Eyelid Models Validation Experiment

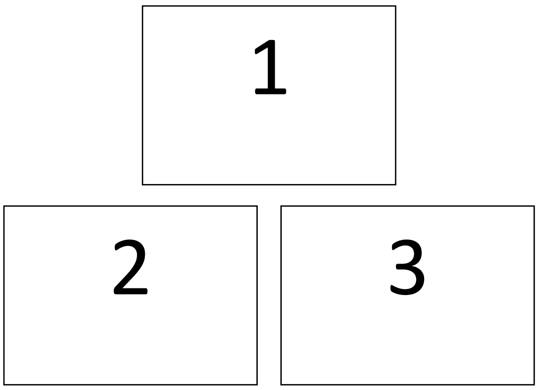
Most Similar to Source Video

Clip	*1st*	*2nd*	*3rd*
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			

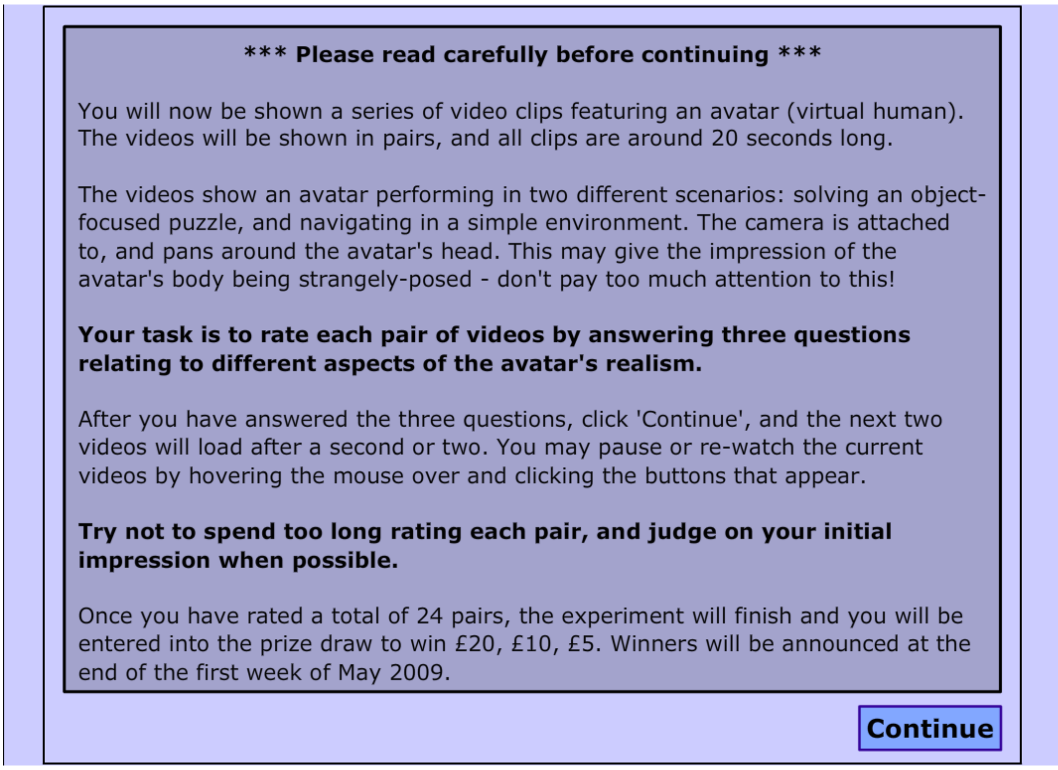


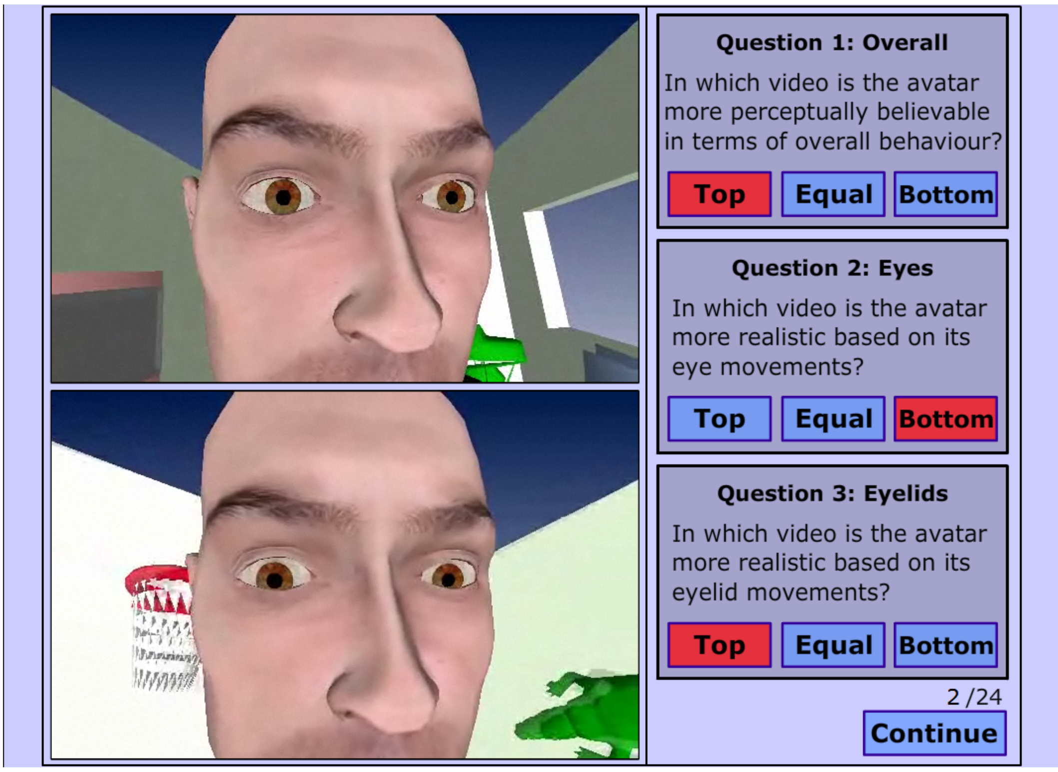
Most Realistic

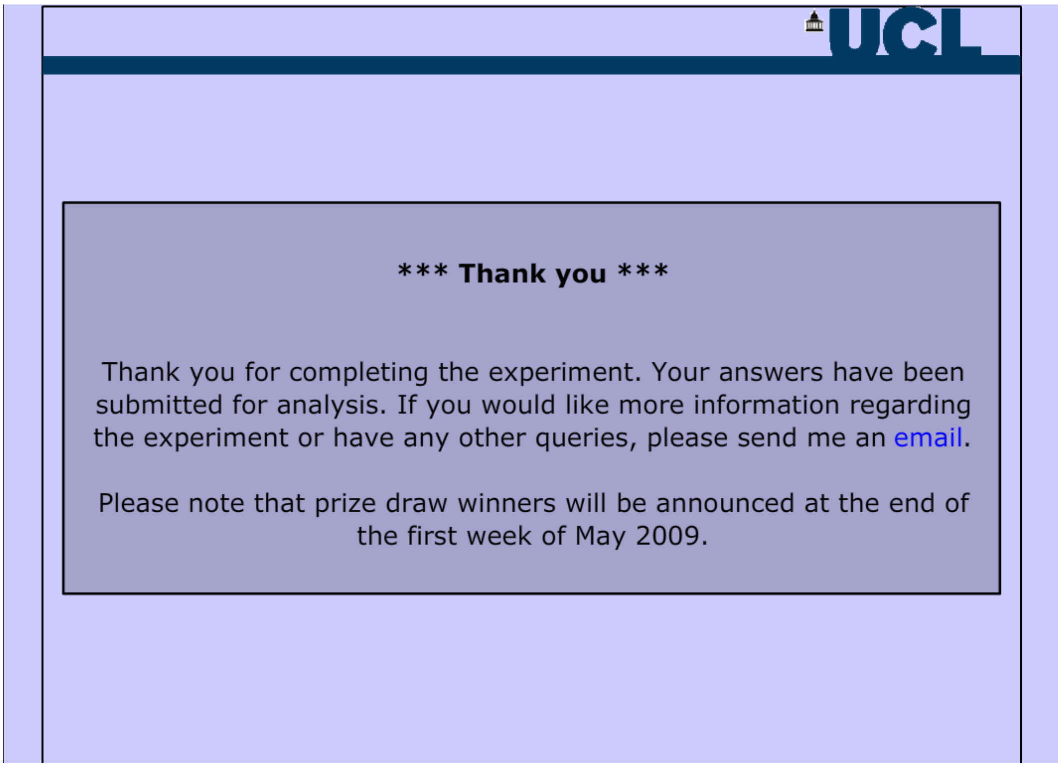
Clip	***1st***	***2nd***	***3rd***
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			



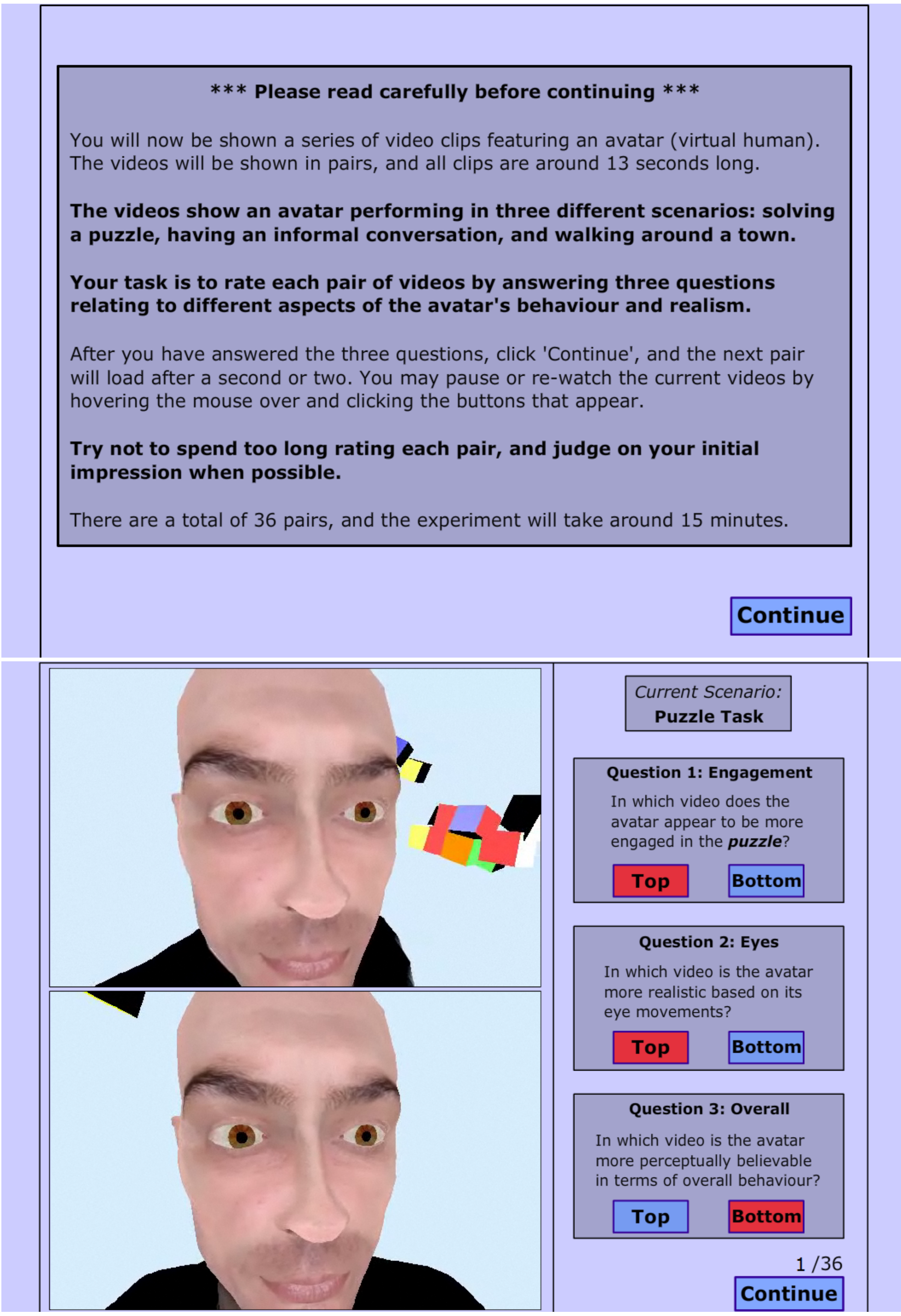
F.3 Images from the Eyelid Models General Impact Experiment

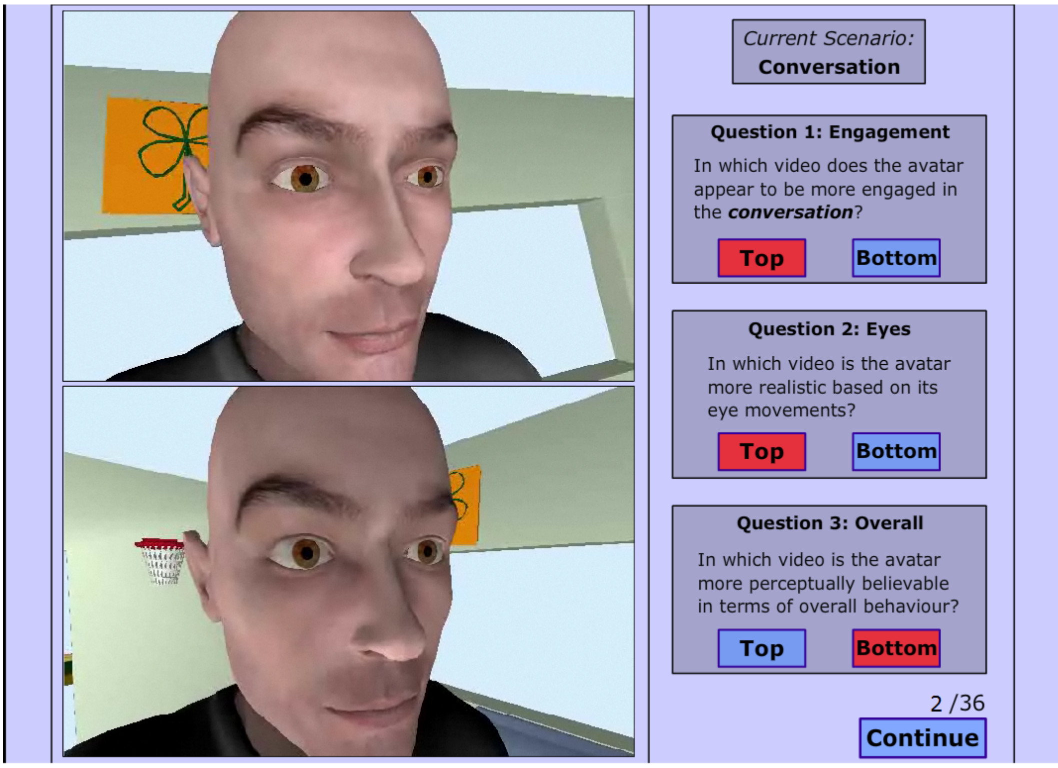


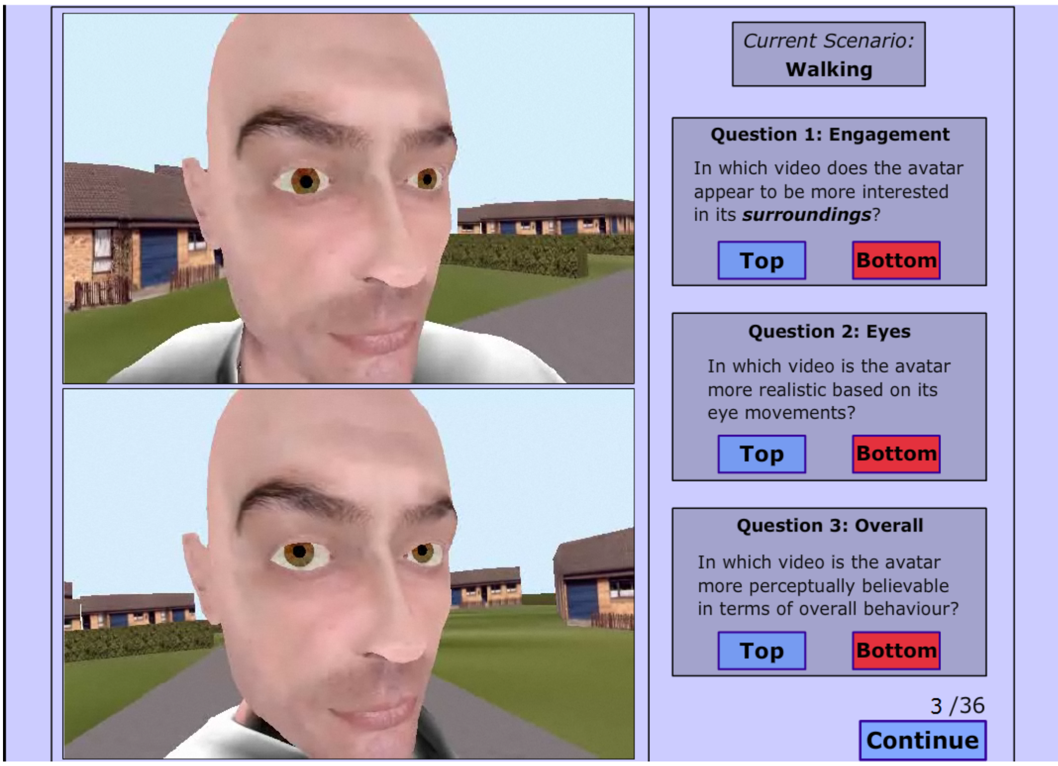

Question 1: Overall
In which video is the avatar more perceptually believable in terms of overall behaviour?
Top
Equal
Bottom
Question 2: Eyes
In which video is the avatar more realistic based on its eye movements?
Top
Equal
Bottom
Question 3: Eyelids
In which video is the avatar more realistic based on its eyelid movements?
Top
Equal
Bottom
2 /24
Continue

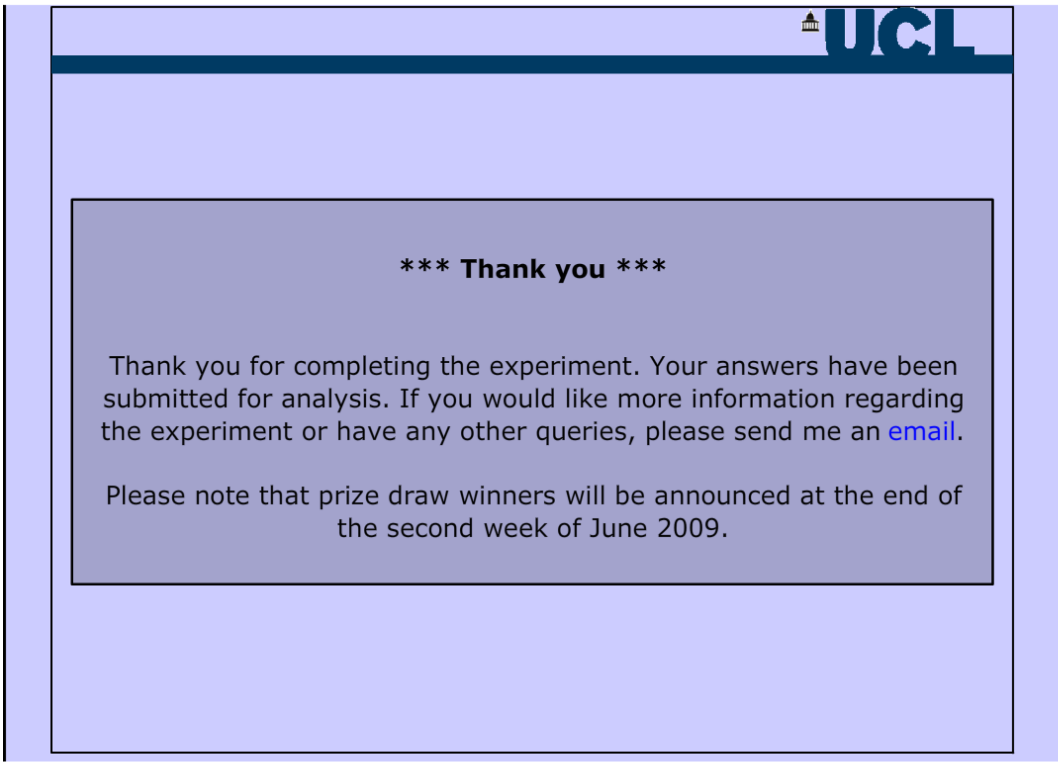

UCL
*** Thank you ***
Thank you for completing the experiment. Your answers have been submitted for analysis. If you would like more information regarding the experiment or have any other queries, please send me an email.
Please note that prize draw winners will be announced at the end of the first week of May 2009.

F.4 Images from the Gaze Model Experiment




Current Scenario:
Conversation
Question 1: Engagement
In which video does the avatar appear to be more engaged in the conversation?
Top
Bottom
Question 2: Eyes
In which video is the avatar more realistic based on its eye movements?
Top
Bottom
Question 3: Overall
In which video is the avatar more perceptually believable in terms of overall behaviour?
Top
Bottom
2 /36
Continue


Current Scenario:
Walking
Question 1: Engagement
In which video does the avatar appear to be more interested in its surroundings?
Top
Bottom
Question 2: Eyes
In which video is the avatar more realistic based on its eye movements?
Top
Bottom
Question 3: Overall
In which video is the avatar more perceptually believable in terms of overall behaviour?
Top
Bottom
3 /36
Continue


UCL
*** Thank you ***
Thank you for completing the experiment. Your answers have been submitted for analysis. If you would like more information regarding the experiment or have any other queries, please send me an email.
Please note that prize draw winners will be announced at the end of the second week of June 2009.

Bibliography

[AAD99] D.E. Anderson, M.E. Ansfield, and B.M. DePaulo. Loves best habit: Deception in the context of relationships. *The social context of nonverbal behavior*, pages 372–409, 1999.

[AC76] M. Argyle and M. Cook. *Gaze and Mutual Gaze*. Cambridge University Press Cambridge, 1976.

[AD65] M. Argyle and J. Dean. Eye contact, distance and affiliation. *Sociometry*, 28(3):289–304, 1965.

[ADRB91] C. Andre-Deshays, M. Revel, and A. Berthoz. Eye-head coupling in humans. *Experimental Brain Research*, 84(2):359–366, 1991.

[AF67] M. Argyle and B.M. Foss. *The psychology of interpersonal behaviour*. Penguin Books Middlesex. England, 1967.

[AG77] M. Argyle and J. Graham. The Central Europe Experiment-looking at persons and looking at things. *Journal of Environmental Psychology and Nonverbal Behaviour*, 1(1977):6–16, 1977.

[AH84] J.M. Atkinson and J. Heritage. *Structures of Social Action: Studies in Conversation Analysis*. Cambridge University Press, 1984.

[AI72] M. Argyle and R. Ingham. Gaze, mutual gaze and proximity. *Semiotica*, 6(1):32–50, 1972.

[AIAM73] M. Argyle, R. Ingham, F. Alkema, and M. McCallin. The different functions of gaze. *Semiotica*, 7(1):19–32, 1973.

[AL87] S.R. Acker and S.R. Levitt. Designing Videoconference Facilities for Improved Eye Contact. *J. Broad. & Elec. Media*, 31:181, 1987.

[ALC68] M. Argyle, M. Lalljee, and M. Cook. The effects of visibility on interaction in a dyad. *Human Relations*, 21(1):3–17, 1968.

[Arg88] M. Argyle. *Bodily communication*. Taylor & Francis, 1988.

[AS96] A.N.B. AS. Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? *Advances in experimental social psychology*, page 389, 1996.

[AT79] M. Argyle and P. Trower. *Person to Person: Ways of Communicating*. HarperCollins Publishers, 1979.

[Aut10] Autodesk. 3DS Max, http://usa.autodesk.com. Software, 2010.

[BB01] R. Bradley and R.T. Bradley. The origin and organization of socioaffective dialogue. *International Journal of Public Administration*, 24(7):799–842, 2001.

[BB04] J.N. Bailenson and J. Blascovich. Avatars. Encyclopedia of Human-Computer Interaction, 2004.

[BB09] M. Banf and V. Blanz. Example-Based Rendering of Eye Movements. In *Computer Graphics Forum*, volume 28, pages 659–666. Blackwell Publishing Ltd, 2009.

[BBBL01] J.N. Bailenson, J. Blascovich, A.C. Beall, and J.M. Loomis. Equilibrium Theory Revisited: Mutual Gaze and Personal Space in Virtual Environments. *Presence: Teleoperators & Virtual Environments*, 10(6):583–598, 2001.

[BBC+97] A.R. Bentivoglio, S.B. Bressman, E. Cassetta, D. Carretta, P. Tonali, and A. Albanese. Analysis of blink rate patterns in normal subjects. *Movement Disorders*, 12(6), 1997.

[BBF+95] S. Benford, J. Bowers, L.E. Fahlén, C. Greenhalgh, and D. Snowdon. User embodiment in collaborative virtual environments. *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 242–249, 1995.

[BBF+97] S. Benford, J. Bowers, L.E. Fahlén, C. Greenhalgh, and D. Snowdon. Embodiments, avatars, clones and agents for multi-user, multi-sensory virtual worlds. *Multimedia Systems*, 5(2):93–104, 1997.

[BBH+90] C. Blanchard, S. Burgess, Y. Harvill, J. Lanier, A. Lasko, M. Oberman, and M. Teitel. Reality built for two: a virtual reality tool. In *Proceedings of the 1990 symposium on Interactive 3D graphics*, page 36. ACM, 1990.

[BBL+03] A.C. Beall, J. Bailenson, J. Loomis, J. Blascovich, and C. Rex. Non-zero-sum gaze in immersive virtual environments. In *Proceedings of HCI International*. Citeseer, 2003.

[BBL+05] J.N. Bailenson, A.C. Beall, J. Loomis, J. Blascovich, and M. Turk. Transformed social interaction, augmented gaze, and social influence in immersive virtual environments. *Human Communication Research*, 31(4):511, 2005.

[BBP01] G. Breton, C. Bouville, and D. Pele. FaceEngine a 3D facial animation engine for real time applications. In *Proceedings of the sixth international conference on 3D Web technology*, pages 15–22. ACM New York, NY, USA, 2001.

[BBRG96] S. Benford, C. Brown, G. Reynard, and C. Greenhalgh. Shared spaces: transportation, artificiality, and spatiality. In *Proceedings of the 1996 ACM conference on Computer supported cooperative work*, pages 77–86. ACM New York, NY, USA, 1996.

[BBW96] J.K. Burgoon, D.B. Buller, and W.G. Woodall. *Nonverbal communication: The unspoken dialogue*. McGraw-Hill New York, 1996.

[BC95] S. Baron-Cohen. *Mindblindness: An Essay on Autism and Theory of Mind*. MIT Press, 1995.

[BCWJ97] S. Baron-Cohen, S. Wheelwright, and T. Jolliffe. Is There a" Language of the Eyes"? Evidence from Normal Adults, and Adults with Autism or Asperger Syndrome. *Visual Cognition*, 4(3):311–331, 1997.

[BF88] W. Becker and AF Fuchs. Lid-eye coordination during vertical gaze changes in man and monkey. *Journal of neurophysiology*, 60(4):1227–1252, 1988.

[BFI95] W. Barfield and T.A. Furness III. *Virtual environments and advanced interface design*. Oxford University Press, Inc. New York, NY, USA, 1995.

[BH97] D.A. Bowman and L.F. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics*. ACM New York, NY, USA, 1997.

[BHG93] N.I. Badler, M. Hollick, and J. Granieri. Real-time control of a virtual human using minimal sensors. *Presence*, 2(1):82–86, 1993.

[Bir60] R.L. Birdwhistell. Kinesics and communication. *Exploration in communication. Boston: Beacon Hill*, pages 54–64, 1960.

[BJD06] C.F. Bond Jr and B.M. DePaulo. Accuracy of deception judgments. *Personality and Social Psychology Review*, 10(3):214, 2006.

[BK76] J.A. Burns and BL Kintz. Eye contact while lying during an interview. *Bulletin of the Psychonomic Society*, 7(1):87–89, 1976.

[BL88] K.R. Boff and J.E. Lincoln. User's Guide Engineering Data Compendium Human Perception and Performance, 1988.

[Bla02] J. Blascovich. Social influence within immersive virtual environments. *The social life of avatars: Presence and interaction in shared virtual environments*, pages 127–145, 2002.

[BLB+02] J. Blascovich, J. Loomis, A.C. Beall, K.R. Swinth, C.L. Hoyt, and J.N. Bailenson. Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, 13(2):103–124, 2002.

[BM90] B. Buxton and T. Moran. EuroPARC's integrated interactive intermedia facility (IIIF): early experiences. In *Proceedings of the IFIP WG 8.4 confernece on Multi-user interfaces and applications table of contents*, pages 11–34. Elsevier North-Holland, Inc. Amsterdam, The Netherlands, The Netherlands, 1990.

[BN08] D. Bitouk and S.K. Nayar. Creating a Speech Enabled Avatar from a Single Photograph. In *Proceedings of the IEEE Virtual Reality (VR) Conference*, pages 107–110, 2008.

[BO69] N. Bernick and M. Oberlander. Effect of verbalization and two different modes of experiencing on pupil size. *Perception and Psychophysics*, 3:327–329, 1969.

[Bod97] M. Boden. Douglas Hofstadter and the Fluid Analogies Research Group, Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought. *Minds and Machines*, 7(3):460–464, 1997.

[Bor04] T. Borchers. Interpersonal Communication. *Moorehead State University*, 2004.

[Bra71] P.T. Brady. Effects of transmission delay on conversational behavior on echo-free telephone circuits. *Bell System Technical Journal*, 50(1):115–134, 1971.

[Bri89] E. Britannica. *Encyclopaedia Britannica*. Chicago: Encyclopædia Britannica, 1989.

[Bru80] T.J. Bruneau. Chronemics and the Verbal-Nonverbal Interface1. *The Relationship of verbal and nonverbal communication*, page 101, 1980.

[BTG03] L.M. Berry, M. Taflinger, and L. Grebe. *Employee selection*. Thomson/Wadsworth, Belmont, CA, 2003.

[Cap85] J.N. Cappella. Controlling the floor in conversation. *Multichannel integrations of nonverbal behavior*, pages 69–103, 1985.

[Car88] R.H.S. Carpenter. Movement of the eyes. *Volume 2nd ed. London Pion Limited, London*, 1988.

[CB01] J.S. Casanueva and E.H. Blake. The effects of avatars on co-presence in a collaborative virtual environment. In *Annual Conference of the South African Institute of Computer Scientists and Information Technologists (SAICSIT2001). Pretoria, South Africa*. Citeseer, 2001.

[CCD00] A. Colburn, M.F. Cohen, and S. Drucker. The Role of Eye Gaze in Avatar Mediated Conversational Interfaces. *Microsoft Research Report*, 81:2000, 2000.

[CDO+00a] L. Childers, T. Disz, R. Olson, M.E. Papka, R. Stevens, and T. Udeshi. Access Grid: Immersive Group-to-Group Collaborative Visualization. *Proceedings of the 4th International Immersive Projection Technology Workshop*, 2000.

[CDO+00b] L. Childers, T. Disz, R. Olson, M.E. Papka, R. Stevens, and T. Udeshi. Access grid: Immersive group-to-group collaborative visualization. In *Proceedings of the 4th International Immersive Projection Technology Workshop*. Citeseer, 2000.

[CEH75] R.A. Crone and Y. Everhard-Halm. Optically induced eye torsion. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 195(4):231–239, 1975.

[CET+01] T.F. Cootes, G.J. Edwards, C.J. Taylor, et al. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[CGB+04] J.R. Carlson, J.F. George, J.K. Burgoon, M. Adkins, and C.H. White. Deception in computer-mediated communication. *Group Decision and Negotiation*, 13(1):5–28, 2004.

[CH93] C. Carlsson and O. Hagsand. DIVE- a platform for multi-user virtual environments. *Computers & Graphics(Pergamon)*, 17(6):663–669, 1993.

[Che02] M. Chen. Leveraging the asymmetric sensitivity of eye contact for videoconference. *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*, pages 49–56, 2002.

[CJM08] P.T. Costa Jr and R.R. McCrae. Revised NEO Personality Inventory (NEOPI-R). *The Sage Handbook of Personality Theory and Assessment, Volume 2: Personality Measurement and Testing*, page 179, 2008.

[CNSD93] C. Cruz-Neira, D.J. Sandin, and T.A. DeFanti. Surround-screen projection-based virtual reality: the design and implementation of the CAVE. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, page 142. ACM, 1993.

[Cor10a] Ascension Technology Corporation. Ascension MotionStar, http://www.ascension-tech.com/realtime/RTMotionSTARTethered.php. Hardware, 2010.

[Cor10b] Microsoft Corporation. Project Natal, http://www.xbox.com/en-US/live/projectnatal/. Consumer electronics device, 2010.

[Cos95] J.P. Costella. A Beginner's Guide to the Human Field of View. *School of Physics, The University of Melbourne, November*, 1995.

[Cra71] M. Cranach. The role of orienting behavior in human interaction. *Behavior and Environment: the Use of Space by Animals and Men. Plenum Press, New York*, pages 217–237, 1971.

[CWS99] T.K. Capin, J. Wiley, and I. Sons. *Avatars in Networked Virtual Environments*. John Wiley & Sons, 1999.

[Dar04] C. Darwin. *The Expression of the Emotions in Man and Animals*. J. Murray, 1904.

[DeP80] B. DePaulo. Humans as lie detectors. *Journal of Communication*, 30(2):129–39, 1980.

[DeP92] B.M. DePaulo. Nonverbal behavior and self-presentation. *Psychological Bulletin*, 111(2):203–243, 1992.

[DGHP01] D.P. Dionisio, E. Granholm, W.A. Hillix, and W.F. Perrine. Differentiation of deception using pupillary responses as an index of cognitive processing. *Psychophysiology*, 38(02):205–211, 2001.

[Dig90] J.M. Digman. Personality structure: Emergence of the five-factor model. *Annual review of psychology*, 41(1):417–440, 1990.

[DK98] B.M. DePaulo and D.A. Kashy. Everyday lies in close and casual relationships. *Journal of Personality and Social Psychology*, 74(1):63–79, 1998.

[DKU98] J.V. Draper, D.B. Kaber, and J.M. Usher. Telepresence. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 40(3):354–375, 1998.

[DL84] R.L. Daft and R.H. Lengel. Information richness: A new approach to managerial information processing and organization design. *Research in organizational behavior*, 6:191–233, 1984.

[DLM+03] B.M. DePaulo, J.J. Lindsay, B.E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper. Cues to deception. *Psychological Bulletin*, 129(1):74–118, 2003.

[DLN05] Z. Deng, J.P. Lewis, and U. Neumann. Automated Eye Motion Using Texture Synthesis. *IEEE Computer Graphics and Applications*, pages 24–30, 2005.

[Duc86] S. Duck. *Human relationships: An introduction to social psychology*. Sage, 1986.

[Duc03] A.T. Duchowski. *Eye Tracking Methodology: Theory and Practice*. Springer, 2003.

[Dun74] S. Duncan. Some signals and rules for taking speaking turns in conversations. *Nonverbal Communication: Readings with Commentary*, 1974.

[DV99] A.R. Dennis and J.S. Valacich. Rethinking media richness: Towards a theory of media synchronicity. 1(1), 1999.

[EF69] P. Ekman and W.V. Friesen. *The repertoire of nonverbal behavior*. Mouton de Gruyter, 1969.

[EF71] P. Ekman and W.V. Friesen. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2):124–129, 1971.

[EF75] P. Ekman and W.V. Friesen. *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*. Prentice-Hall, 1975.

[EGR91] C.A. Ellis, S.J. Gibbs, and G. Rein. Groupware: some issues and experiences. *Communications of the ACM*, 34(1):39–58, 1991.

[EGS65] R.V. Exline, D. Gray, and D. Schuette. Visual behavior in a dyad as affected by interview content and sex of respondent. *Journal of Personality and Social Psychology*, 1(3):201–209, 1965.

[EI87] R.J. Edelmann and S. Iwawaki. Self-reported expression and consequences of embarrassment in the United Kingdom and Japan. *Psychologia*, 30(4):205–216, 1987.

[Ekm92] P. Ekman. An argument for basic emotions. *Cognition & Emotion*, 6(3-4):169–200, 1992.

[Ell81] C.J. Ellis. The pupillary light reflex in normal subjects. *British Medical Journal*, 65(11):754, 1981.

[EMS91] C. Evinger, K.A. Manning, and P.A. Sibony. Eyelid movements. Mechanisms and normal data. *Investigative ophthalmology & visual science*, 32(2):387–400, 1991.

[Ent10a] Blizzard Entertainment. World of Warcraft, http://blizzard.com. Software, 2010.

[Ent10b] Blizzard Entertainment. World of Warcraft subscriber base reaches 12 million worldwide, http://eu.blizzard.com/en/press/081223.html. Software, 2010.

[EPM08] I.M. Ekman, A.W. Poikola, and M.K. Makarainen. Invisible eni: using gaze and pupil size to control a game. In *CHI'08 extended abstracts on Human factors in computing systems*, pages 3135–3140. ACM, 2008.

[ETHG70] R.V. Exline, J. Thibaut, C.B. Hickey, and P. Gumpert. Visual interaction in relation to Machiavellianism and an unethical act. *Studies in machiavellianism*, pages 53–75, 1970.

[EW65] R.V. Exline and L.C. Winters. Effects of cognitive difficulty and cognitive style upon eye to eye contact in interviews. In *annual meeting of the Eastern Psychological Association, Philadelphia, PA*, 1965.

[Exl63] R.V. Exline. Explorations in the process of person perception: Visual interaction in relation to competition, sex, and need for affiliation. *Journal of Personality*, 31(1):1–20, 1963.

[FAM+00] J. Freeman, S.E. Avons, R. Meddis, D.E. Pearson, and W. IJsselsteijn. Using behavioral realism to estimate presence: A study of the utility of postural responses to motion stimuli. *Presence: Teleoperators & Virtual Environments*, 9(2):149–164, 2000.

[FAPI99] J. Freeman, S.E. Avons, D.E. Pearson, and W.A. IJsselsteijn. Effects of sensory information and prior experience on direct subjective ratings of presence. *Presence: Teleoperators & Virtual Environments*, 8(1):1–13, 1999.

[FB09] J. Fox and J.N. Bailenson. Virtual self-modeling: The effects of vicarious reinforcement and identification on exercise behaviors. *Media Psychology*, 12(1):1–25, 2009.

[FOM+02] A. Fukayama, T. Ohno, N. Mukawa, M. Sawaki, and N. Hagita. Messages embedded in gaze of interface agents—impression management with agent's gaze. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*, pages 41–48. ACM New York, NY, USA, 2002.

[FSS07] D. Friedman, A. Steed, and M. Slater. Spatial social behavior in second life. *Lecture Notes in Computer Science*, 4722:252, 2007.

[FUS97] C. Ferrán-Urdaneta and J. Storck. Truth or deception: the impact of videoconferencing for job interviews. In *Proceedings of the eighteenth international conference on Information systems*, pages 183–196. Association for Information Systems Atlanta, GA, USA, 1997.

[FW99] J.M. Findlay and R. Walker. A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences*, 22(04):661–674, 1999.

[Gar03] M. Garau. *The Impact of Avatar Fidelity on Social Interaction in Virtual Environments*. PhD thesis, University College London, 2003.

[Gar06] M. Garau. Selective fidelity: Investigating priorities for the creation of expressive avatars. *Avatars at work and play: Collaboration and interaction in shared virtual environments*, pages 17–38, 2006.

[GB95] C. Greenhalgh and S. Benford. MASSIVE: a collaborative virtual environment for teleconferencing. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 2(3):239–261, 1995.

[GB04] M. Gillies and D. Ballin. Integrating Autonomous Behavior and User Control for Believable Agents. In *International Conference on Autonomous Agents: Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-*, volume 1, pages 336–343, 2004.

[GB06] E. Gu and N.I. Badler. Visual Attention and Eye Gaze During Multiparty Conversations with Distractions. *Intelligent Virtual Agents*, 2006.

[GG86] M.H. Goodwin and C. Goodwin. Gesture and coparticipation in the activity of searching for a word. *Semiotica*, 62(1/2):51–75, 1986.

[GH90] C. Goodwin and J. Heritage. Conversation analysis. *Annual review of anthropology*, 19(1):283–307, 1990.

[GL86] Lucasfilm Games and Quantum Link. Habitat. Software, 1986.

[Goo00] C. Goodwin. Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32(10):1489–1522, 2000.

[Goo10a] Google. Google Documents, http://www.google.com/documents. Software, 2010.

[Goo10b] Google. Google Talk, http://www.google.com/talk. Software, 2010.

[Gor51] G. Gordon. Observations upon the movements of the eyelids. *British Medical Journal*, 35(6):339, 1951.

[Gra18] H. Gray. *Anatomy of the human body*. Lea & Febiger, 1918.

[GRA+02] J. Gratch, J. Rickel, E. André, J. Cassell, E. Petajan, and N. Badler. Creating interactive virtual humans: Some assembly required. *IEEE intelligent systems*, pages 54–63, 2002.

[Gre88] I. Greif. *Computer-Supported Cooperative Work: A Book of Readings*. Morgan Kaufmann, 1988.

[Gro10] Oxford Metrics Group. VICON, http://www.vicon.com. Hardware, 2010.

[GSBS01] M. Garau, M. Slater, S. Bee, and M.A. Sasse. The impact of eye gaze on communication using humanoid avatars. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 309–316. ACM Press New York, NY, USA, 2001.

[GSV+03] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M.A. Sasse. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. *Proceedings of the conference on Human factors in computing systems*, pages 529–536, 2003.

[GW60] E.J. Gibson and R.D. Walk. The visual cliff. *Scientific American*, 202(4):64–71, 1960.

[GWN+03] M. Gross, S. Wurmlin, M. Naef, E. Lamboray, C. Spagno, A. Kunz, E. Koller-Meier, T. Svoboda, L. Van Gool, and S. Lang. blue-c: a spatially immersive display and 3D video portal for telepresence. *ACM Transactions on Graphics*, 22(3):819–827, 2003.

[Hal36] A.J. Hall. Some observations on the acts of closing and opening the eyes. *British Medical Journal*, 20(5):257, 1936.

[Hal45] A.J. Hall. The origin and purposes of blinking. *British Medical Journal*, 29(9):445, 1945.

[Hal63] E.T. Hall. A System for the Notation of Proxemic Behavior. *American Anthropologist*, 65(5):1003–1026, 1963.

[Hal68] E.T. Hall. Proxemics. *Current Anthropology*, 9(2/3):83, 1968.

[HD82] M.J. Hawes and R.K. Dortzbach. The microscopic anatomy of the lower eyelid retractors. *Archives of Ophthalmology*, 100(8):1313, 1982.

[HEHR+08] K. Hagni, K. Eng, M.C. Hepp-Reymond, L. Holper, B. Keisker, E. Siekierka, and D.C. Kiper. Observing Virtual Arms that You Imagine Are Yours Increases the Galvanic Skin Response to an Unexpected Threat. *PLoS One*, 3(8), 2008.

[Hei76] I. Heilveil. Deception and pupil size. *Journal of Clinical Psychology*, 32(3):675–76, 1976.

[Her70] J. Heron. The phenomenology of social encounter: The gaze. *Philosophy and Phenomenological Research*, 31(2):243–264, 1970.

[HFH+98] J. Hindmarsh, M. Fraser, C. Heath, S. Benford, and C. Greenhalgh. Fragmented interaction: establishing mutual orientation in virtual environments. *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pages 217–226, 1998.

[HFH+00] J. Hindmarsh, M. Fraser, C. Heath, S. Benford, and C. Greenhalgh. Object-focused interaction in collaborative virtual environments. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(4):477–509, 2000.

[HGSV04] M. Hartwig, P.A. Granhag, L.A. Stromwall, and A. Vrij. Police officers' lie detection accuracy: Interrogating freely versus observing video. *Police Quarterly*, 7(4):429, 2004.

[HH99] J.M. Henderson and A. Hollingworth. High-level scene perception. *Annual Review of Psychology*, 50(1):243–271, 1999.

[Hin72] R.A. Hinde. *Non-verbal communication*. Cambridge Univ Pr, 1972.

[Hof95] D.R. Hofstadter. *Fluid concepts and creative analogies*. Basic Books New York, 1995.

[HP60] E.H. Hess and J.M. Polt. Pupil size as related to interest value of visual stimuli. *Science*, 132(3423):349–350, 1960.

[HR92] L. Hettinger and G. Riccio. Virtually induced motion sickness in virtual environments. *Presence: Teleoperators and Virtual Environments*, 1(3):306–310, 1992.

[HRBC06] J. Hauber, H. Regenbrecht, M. Billinghurst, and A. Cockburn. Spatiality in videoconferencing: trade-offs between efficiency and social presence. *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 413–422, 2006.

[HSS+05] I. Heldal, A. Steed, M. Spante, R. Schroeder, S. Bengtsson, and M. Partanen. Successes and Failures in Co-Present Situations. *Presence: Teleoperators & Virtual Environments*, 14(5):563–579, 2005.

[HT72] M.K. Holland and G. Tarlow. Blinking and mental load. *Psychological reports*, 31(1):119, 1972.

[HTS66] C.S. Harris, R.I. Thackray, and R.W. Shoenberger. Blink rate as a function of induced muscular tension and manifest anxiety. *Perceptual and Motor Skills*, 22(1):155–160, 1966.

[HWM78] R.G. Harper, A.N. Wiens, and J.D. Matarazzo. *Nonverbal communication: The state of the art*. John Wiley & Sons Inc, 1978.

[IDP03] L. Itti, N. Dhavale, and F. Pighin. Realistic avatar eye and head animation using a neurobiological model of visual attention, 2003.

[IK92] H. Ishii and M. Kobayashi. ClearBoard: a seamless medium for shared drawing and conversation with eye contact. *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 525–532, 1992.

[IKA94] H. Ishii, M. Kobayashi, and K. Arita. Iterative design of seamless collaboration media. *Communications of the ACM*, 37(8):83–97, 1994.

[IKG93] H. Ishii, M. Kobayashi, and J. Grudin. Integration of interpersonal space and shared workspace: ClearBoard design and experiments. *ACM Transactions on Information Systems (TOIS)*, 11(4):349–375, 1993.

[IKN98] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11):1254–1259, 1998.

[Ins01] B.E. Insko. *Passive haptics significantly enhances virtual environments*. PhD thesis, Citeseer, 2001.

[Int10] InterSense. IS900, http://www.intersense.com/IS-900. Hardware, 2010.

[IT94] E.A. Isaacs and J.C. Tang. What video can and cannot do for collaboration: A case study. *Multimedia Systems*, 2(2):63–73, 1994.

[Ive27] H.E. Ives. Television. *Bell System Technical Journal*, 6:551–559, 1927.

[Jam07] W. James. Pragmatism's conception of truth. *The Journal of Philosophy, Psychology and Scientific Methods*, 4(6):141–155, 1907.

[JH95] B. Jordan and A. Henderson. Interaction analysis: Foundations and practice. *The Journal of the learning sciences*, 4(1):39–103, 1995.

[Joh88] R. Johansen. *GroupWare: Computer Support for Business Teams*. The Free Press New York, NY, USA, 1988.

[KB66] D. Kahneman and J. Beatty. Pupil diameter and load on memory. *Science*, 154(3756):1583–1585, 1966.

[KB01] S.C. Khullar and N.I. Badler. Where to Look? Automating Attending Behaviors of Virtual Human Characters. *Autonomous Agents and Multi-Agent Systems*, 4(1):9–23, 2001.

[KBBB00] B. Keysar, D.J. Barr, J.A. Balin, and J.S. Brauner. Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, pages 32–38, 2000.

[KC69] A. Kendon and M. Cook. The consistency of gaze patterns in social interaction. *British Journal of Psychology*, 60(4):481–494, 1969.

[Ken67] A. Kendon. Some functions of gaze-direction in social interaction. *Acta Psychol (Amst)*, 26(1):22–63, 1967.

[KF73] A. Kendon and A. Ferber. A description of some human greetings. *Comparative ecology and behaviour of primates*, pages 591–668, 1973.

[KG64] D.W. Kennard and G.H. Glaser. An analysis of eyelid movements. *The Journal of Nervous and Mental Disease*, 139, 1964.

[KG08] M. Kipp and P. Gebhard. IGaze: Studying Reactive Gaze Behavior in Semi-immersive Human-Avatar Interactions. *Intelligent Virtual Agents: 8th International Conference, IVA 2008, Tokyo, Japan, September 1-3, 2008, Proceedings*, page 191, 2008.

[KHD74] M.L. Knapp, R.P. Hart, and H.S. Dennis. An exploration of deception as a communication construct. *Human Communication Research*, 1(1):15–29, 1974.

[Kid75] R.F. Kidd. Pupil size, eye contact, and instrumental aggression. *Perceptual and motor skills*, 41(2):538, 1975.

[Kid05] M. Kidwell. Gaze as Social Control: How Very Young Children Differentiate" The Look" From a" Mere Look" by Their Adult Caregivers. *Research on Language and Social Interaction*, 38(4):417–449, 2005.

[KJC+09] A. Kotranza, K. Johnsen, J. Cendan, B. Miller, D.S. Lind, and B. Lok. Virtual multi-tools for hand and tool-based interaction with life-size virtual human agents. In *Proceedings of the 2009 IEEE Symposium on 3D User Interfaces-Volume 00*, pages 23–30. IEEE Computer Society Washington, DC, USA, 2009.

[KKD+05] A. Kenny, H. Koesling, D. Delaney, S. McLoone, and T. Ward. A preliminary investigation into eye gaze data in a first person shooter game. In *Proceedings of the 19th European Conference on Modelling and Simulation*, pages 733–740. Citeseer, 2005.

[KR05] Y.M. Kalman and S. Rafaeli. Email chronemics: Unobtrusive profiling of response times. In *Proceedings of the 38th International Conference on System Sciences, HICSS*, volume 38, page 108, 2005.

[Kra04] R.J. Krauzlis. Recasting the Smooth Pursuit Eye Movement System. *Journal of Neurophysiology*, 91(2):591–603, 2004.

[KS40] M.G. Kendall and B.B. Smith. On the method of paired comparisons. *Biometrika*, 31(3-4):324–345, 1940.

[KSJ00] E.R. Kandel, J.H. Schwartz, and T.M. Jessell. *Principles of Neural Science*. Appleton & Lange, 2000.

[KU85] C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Human neurobiology*, 4(4):219–227, 1985.

[Lab10] Applied Science Laboratories. Applied Science Laboratories, http://www.asleyetracking.com. Eye tracker manufacturer, 2010.

[Lan03] J.P. Lantolf. Intrapersonal communication and internalization in the second language classroom. *Vygotsky's educational theory in cultural context*, pages 349–370, 2003.

[Lee07] M.R. Lee. Solanaceae IV: Atropa belladonna, deadly nightshade. *J.R. College Physicians Edinburgh*, 37:77–84, 2007.

[Ler03] G.H. Lerner. Selecting next speaker: The context-sensitive operation of a context-free organization. *Language in Society*, 32(02):177–201, 2003.

[LF42] O. Lowenstein and E.D. Friedman. Pupillographic Studies: I. Present State of Pupillography; Its Method and Diagnostic Significance. *Archives of Ophthalmology*, 27(5):969, 1942.

[LF96] R.E. Lubow and O. Fein. Pupillary size in response to a visual guilty knowledge test: New technique for the detection of deception. *Journal of Experimental Psychology: Applied*, 2(2):164–177, 1996.

[LPM99] T.R. Levine, H.S. Park, and S.A. McCornack. Accuracy in detecting truths and lies: Documenting the veracity effect. *Communication Monographs*, 66(2):125–144, 1999.

[LV08] S. Leal and A. Vrij. Blinking During and After Lying. *Journal of Nonverbal Behavior*, 32(4):187–194, 2008.

[LZ99] R.J. Leigh and D.S. Zee. *The neurology of eye movements*. Oxford University Press, Oxford, 3 edition, 1999.

[Mas96] J.T. Masterson. Nonverbal communication in text based virtual realities. Master's thesis, University of Montana, 1996.

[Mat74] E. Matin. Saccadic suppression: a review and an analysis. *Psychol Bull*, 81(12):899–917, 1974.

[MD09] X. Ma and Z. Deng. Natural Eye Motion Synthesis by Modeling Gaze-Head Coupling. In *Proceedings of the 2009 IEEE Virtual Reality Conference*, pages 143–150. IEEE Computer Society Washington, DC, USA, 2009.

[MeCM+05] J.M.C. Malbouisson, A.A.V. e Cruz, A. Messias, L.V.O. Leite, and G.D. Rios. Upper and lower eyelid saccades describe a harmonic oscillator function. *Investigative ophthalmology & visual science*, 46(3):857–862, 2005.

[Meh07] A. Mehrabian. Communication without words. *Communication Theory*, page 193, 2007.

[Mei62] R.L. Meier. *A communications theory of urban growth*. MIT Press Cambridge, Mass., 1962.

[MG02] A.F. Monk and C. Gale. A Look Is Worth a Thousand Words: Full Gaze Awareness in Video-Mediated Conversation. *Discourse Processes*, 33(3):257–278, 2002.

[Mil63] S. Milgram. Behavioral Study of obedience. *Journal of Abnormal and Social Psychology*, 67(4):371–378, 1963.

[Min80] M. Minsky. Telepresence. *Omni*, 2(9):45–52, 1980.

[MIWBJ02] M. Meehan, B. Insko, M. Whitton, and F.P. Brooks Jr. Physiological measures of presence in stressful virtual environments. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 645–652. ACM New York, NY, USA, 2002.

[MK01] D. Melcher and E. Kowler. Visual scene memory and the guidance of saccadic eye movements. *Vision Research*, 41(25-26):3597–3611, 2001.

[MLD92] D.M. McNair, M. Lorr, and LF Droppleman. Profile of mood states. *San Diego, California*, 53:6, 1992.

[Mof01] J. Moffitt. Ogg VorbisOpen, Free AudioSet Your Media Free. *Linux journal*, 2001(81es):9, 2001.

[Mol69] J.P. Molnar. Picturephone service-a new way of communicating. *Bell Laboratories Record*, 47(5):134–135, 1969.

[MR03] M. Meehan and S. Razzaque. Effect of latency on presence in stressful virtual environments. *Proceedings of the IEEE*, pages 141–148, 2003.

[MRS$^+$07] N. Murray, D. Roberts, A. Steed, P. Sharkey, P. Dickerson, and J. Rae. An assessment of eye-gaze potential within immersive virtual environments. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, 3(4):8, 2007.

[NB03] K.L. Nowak and F. Biocca. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 12(5):481–494, 2003.

[NM00] C. Nass and Y. Moon. Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1):81–103, 2000.

[NMK09] H. Nakanishi, Y. Murakami, and K. Kato. Movable cameras enhance social telepresence in media spaces. In *Proceedings of the 27th international conference on Human factors in computing systems*, pages 433–442. ACM New York, NY, USA, 2009.

[NR04] S. Noël and J.M. Robert. Empirical Study on Collaborative Writing: What Do Co-authors Do, Use, and Like? *Computer Supported Cooperative Work (CSCW)*, 13(1):63–89, 2004.

[OA02] M. O'Hagan and D. Ashworth. *Translation-mediated communication in a digital world: facing the challenges of globalization and localization*. Multilingual Matters Limited, 2002.

[OAS+02] N. Osawa, K. Asai, M. Suzuki, Y.Y. Sugimoto, and F. Saito. An immersive programming system: Ougi. In *The 12th International Conference on Artificial Reality and Telexistence (ICAT 2002)*, pages 36–43, 2002.

[OMIM94] K.I. Okada, F. Maeda, Y. Ichikawaa, and Y. Matsushita. Multiparty videoconferencing at virtual social distance: MAJIC design. *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pages 385–393, 1994.

[oO10] Atlas of Ophthalmology. Atlas of Ophthalmology, http://www.atlasophthalmology.com. Website, 2010.

[Pat90] M.L. Patterson. Functions of non-verbal behavior in social interaction. *Handbook of language and social psychology*, pages 101–120, 1990.

[PHG+04] B. Pan, H.A. Hembrooke, G.K. Gay, L.A. Granka, M.K. Feusner, and J.K. Newman. The determinants of web page viewing behavior: an eye-tracking study. In *Proceedings of Eye tracking Research & Applications*, pages 147–154, New York, NY, USA, 2004. ACM.

[PK27] E. Ponder and W.P. Kennedy. On the act of blinking. *Experimental Physiology*, 18(2):89, 1927.

[PLBB02] S. Park Lee, J.B. Badler, , and N.I. Badler. Eyes alive. *ACM Transactions on Graphics (TOG)*, 21(3):637–644, 2002.

[PLN02] D. Parkhurst, K. Law, and E. Niebur. Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42(1):107–123, 2002.

[PM04] J. Pavlik and S. McIntosh. *Converging media: An introduction to mass communication*. Allyn & Bacon, 2004.

[PO03] C. Peters and C. O'Sullivan. Attention-driven eye gaze and blinking for virtual humans. In *International Conference on Computer Graphics and Interactive Techniques*, pages 1–1. ACM New York, NY, USA, 2003.

[POB09] V.F. Pamplona, M.M. Oliveira, and G.V.G. Baranoski. Photorealistic models for pupil light reflex and iridal pattern deformation. *ACM Transactions on Graphics (TOG)*, 28(4):106, 2009.

[Pol10] Polhemus. Fastrak, http://www.polhemus.com. Hardware, 2010.

[PPB+05] C. Peters, C. Pelachaud, E. Bevacqua, M. Mancini, and I. Poggi. A Model of Attention and Interest Using Gaze Behavior. *Lecture Notes In Computer Science*, 3661:229, 2005.

[PS03] T. Partala and V. Surakka. Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies*, 59(1-2):185–198, 2003.

[PS07] X. Pan and M. Slater. A preliminary study of shy males interacting with a virtual female. In *Presence: The 10th Annual International Workshop on Presence*, 2007.

[PSB02] D.P. Pertaub, M. Slater, and C. Barker. An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators & Virtual Environments*, 11(1):68–78, 2002.

[PSS+01] S. Penny, J. Smith, P. Sengers, A. Bernhardt, and J. Schulte. Traces: Embodied Immersive Interaction with Semi-Autonomous Avatars. *Convergence*, 7(2):47, 2001.

[Pty10] Beepa Pty. Fraps, http://www.fraps.com. Software, 2010.

[PW96] F.I. Parke and K. Waters. *Computer facial animation*. A.K. Peters, Ltd. Natick, MA, USA, 1996.

[Rae01] J. Rae. Organizing Participation in Interaction: Doing Participation Framework. *Research on Language and Social Interaction*, 34(2):253–278, 2001.

[Ray98a] K. Rayner. Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin*, 124:372–422, 1998.

[Ray98b] K. Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 85:618–660, 1998.

[Rea10] RealD. CrystalEyes 3, http://www.reald.com/Content/CE3.aspx. Hardware, 2010.

[Res10a] Arrington Research. Arrington Research, http://www.arringtonresearch.com. Eye tracker manufacturer, 2010.

[Res10b] Linden Research. Second Life, http://lindenlab.com/. Software, 2010.

[RM95] V. Richmond and J. McCroskey. *Immediacy. Nonverbal Behavior in Interpersonal Relations*. Boston, Allyn & Bacon, 1995.

[RMM+05] D. Roberts, D. Marshall, S. McLoone, D. Delaney, and R. Aspin. Exploring the use of local inconsistency measures as thresholds for dead reckoning update packet generation. In *Distributed Simulation and Real-Time Applications*, pages 95–202, Montreal, Canada, 2005.

[RN96] B. Reeves and C. Nass. *The media equation: how people treat computers, television, and new media like real people and places*. Cambridge University Press New York, NY, USA, 1996.

[Ros05] F. Rossano. On sustaining vs. withdrawing gaze in face-to-face interaction. In *91st annual convention of the National Communication Association, Boston, MA, November*, 2005.

[Rub74] Z. Rubin. Measurement of romantic love. *International Journal of Group Tensions*, 4:71–81, 1974.

[Rut84] D.R. Rutter. *Looking and seeing: The role of visual communication in social interaction*. John Wiley & Sons Inc, 1984.

[RVB02] D. Reiners, G. Voß, and J. Behr. OpenSG: basic concepts. In *Proceedings of the OpenSG Symposium*. Citeseer, 2002.

[RWOS03] D. Roberts, R. Wolff, O. Otto, and A. Steed. Constructing a Gazebo: Supporting Teamwork in a Tightly Coupled, Distributed Task in Virtual Reality. *Presence: Teleoperators & Virtual Environments*, 12(6):644–657, 2003.

[Sac95] H. Sacks. *Lectures on Conversation*. Blackwell, 1995.

[SB92] K. Schmidt and L. Bannon. Taking CSCW seriously. *Computer Supported Cooperative Work (CSCW)*, 1(1):7–40, 1992.

[SBA92] A. Sellen, B. Buxton, and J. Arnott. Using spatial cues to improve videoconferencing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 651–652. ACM New York, NY, USA, 1992.

[SBS94] J.A. Stern, D. Boyer, and D. Schroeder. Blink rate: a possible measure of fatigue. *Human factors*, 36(2):285–297, 1994.

[Sch02] R. Schroeder. Social interaction in virtual environments: Key issues, common themes, and a framework for research. *The social life of avatars: Presence and interaction in shared virtual environments*, pages 1–18, 2002.

[Sch06] R. Schroeder. Being there together and the future of connected presence. *Presence: Teleoperators and Virtual Environments*, 15(4):438–454, 2006.

[Sch07] E.A. Schegloff. *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Cambridge University Press, 2007.

[SFH$^+$01] S.A. Spence, T.F.D. Farrow, A.E. Herford, I.D. Wilkinson, Y. Zheng, and P.W.R. Woodruff. Behavioural and functional anatomical correlates of deception in humans. *Neuroreport*, 12(13):2849, 2001.

[SHS+00] M. Slater, J. Howell, A. Steed, DP Pertaub, and M. Garau. Acting in virtual reality. *Proceedings of the third international conference on Collaborative virtual environments*, pages 103–110, 2000.

[SKMY09] M. Slater, P. Khanna, J. Mortensen, and I. Yu. Visual Realism Enhances Realistic Response in an Immersive Virtual Environment. *IEEE Computer Graphics and Applications*, 29(3):76–84, 2009.

[Sky10] Skype. Skype, http://www.skype.com. Software, 2010.

[Sla04] M. Slater. How colorful was your day? Why questionnaires cannot assess presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 13(4):484–493, 2004.

[Sla09] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549, 2009.

[SMIRM07] H. Skaf-Molli, C.L. Ignat, C. Rahhal, and P. Molli. New work modes for collaborative writing. In *International Conference on Enterprise Information Systems and Web Technologies (EISWT-07)*, 2007.

[Sof10a] Jenkins Software. RakNet, http://www.jenkinssoftware.com. Software, 2010.

[Sof10b] Smith Micro Software. Poser 7, http://poser.smithmicro.com/poser.html. Software, 2010.

[Spe10] OpenAL Specification. Reference, v. 1.0, 2000, Loki Software. Software, 2010.

[SRSH05] A. Steed, D. Roberts, R. Schroeder, and I. Heldal. Interaction between Users of Immersion Projection Technology Systems. *HCI International 2005, the 11th International Conference on Human Computer Interaction*, pages 22–27, 2005.

[SSA+01] R. Schroeder, A. Steed, A.S. Axelsson, I. Heldal, Å. Abelin, J. Wideström, A. Nilsson, and M. Slater. Collaborating in networked immersive spaces: as good as being there together? *Computers & Graphics*, 25(5):781–788, 2001.

[SSH+03] A. Steed, M. Spante, I. Heldal, A.S. Axelsson, and R. Schroeder. Strangers and friends in caves: an exploratory study of collaboration in networked IPT systems for extended periods of time. *Proceedings of the 2003 symposium on Interactive 3D graphics*, pages 51–54, 2003.

[SSJ74] H. Sacks, E.A. Schegloff, and G. Jefferson. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, 50(4):696–735, 1974.

[SSS+99] A. Steed, M. Slater, A. Sadagic, A. Bullock, and J. Tromp. Leadership and collaboration in shared virtual environments. In *Proceedings of the IEEE Virtual Reality*, page 112. IEEE Computer Society Washington, DC, USA, 1999.

[SSUS00] M. Slater, A. Sadagic, M. Usoh, and R. Schroeder. Small-group behavior in a virtual and real environment: A comparative study. *Presence: Teleoperators & Virtual Environments*, 9(1):37–51, 2000.

[Ste96] A. Steed. *Defining Interaction within Virtual Environments*. PhD thesis, PhD thesis, Department of Computer Science, Queen Mary and Westfield College, University of London, London, 1996.

[Str96] J. Streek. How to do things with things: Objects and symbolization. *Human Studies*, 19:365–384, 1996.

[SU93] M. Slater and M. Usoh. Presence in immersive virtual environments. In *1993 IEEE Virtual Reality Annual International Symposium, 1993.*, pages 90–96, 1993.

[SUS94] M. Slater, M. Usoh, and A. Steed. Depth of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 3(2):130–144, 1994.

[Sut68] I.E. Sutherland. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, pages 757–764. ACM, 1968.

[SVS05] M.V. Sanchez-Vives and M. Slater. From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6(4):332–339, 2005.

[SW67] J.W. Stass and F.N. Willis. Eye contact, pupil dilation, and personal preference. *Psychonomic science*, 7(10):375–376, 1967.

[SW97] M. Slater and S. Wilbur. A Framework for Immersive Virtual Environments(FIVE)-Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 6(6):603–616, 1997.

[SWC76] J. Short, E. Williams, and B. Christie. *The social psychology of telecommunications*. John Wiley & Sons, 1976.

[Sys10a] Adobe Systems. Adobe Flash, http://www.adobe.com/products/flash. Software, 2010.

[Sys10b] Cisco Systems. Cisco TelePresence, http://www.cisco.com. Hardware, 2010.

[TBS+98] J. Tromp, A. Bullock, A. Steed, A. Sadagic, M. Slater, and E. Frecon. Small group behaviour experiments in the Coven project. *Computer Graphics and Applications, IEEE*, 18(6):53–63, 1998.

[Tha99] D. Thalmann. The Role of Virtual Humans in Virtual Environment Technology and Interfaces. In *Proceedings of Joint EC-NSF Advanced Research Workshop, Bonas, France*, 1999.

[Tid95] J.B.W.L.C. Tidwell. Nonverbal cues in computer-mediated communication, and the effect of chronemics on relational communication. *Journal of Organizational Computing*, 5(4):355–378, 1995.

[Toa74] F.M. Toates. Vergence eye movements. *Documenta Ophthalmologica*, 37(1):153–214, 1974.

[TOY+08] K. Takashima, Y. Omori, Y. Yoshimoto, Y. Itoh, Y. Kitamura, and F. Kishino. Effects of avatar's blinking animation on person impressions. In *Proceedings of graphics interface 2008*, pages 169–176. Canadian Information Processing Society Toronto, Ont., Canada, Canada, 2008.

[TRH+01] I.I. Taylor, M. Russell, T.C. Hudson, A. Seeger, H. Weber, J. Juliano, and A.T. Helser. VRPN: a device-independent, network-transparent VR peripheral system. In *Proceedings of the ACM symposium on Virtual reality software and technology*, page 61. ACM, 2001.

[TTO05] K. Tateno, M. Takemura, and Y. Ohta. Enhanced eyes for better gaze-awareness in collaborative mixed reality. In *Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 100–103. IEEE Computer Society Washington, DC, USA, 2005.

[UAW+99] M. Usoh, K. Arthur, M.C. Whitton, R. Bastos, A. Steed, M. Slater, and F.P. Brooks Jr. Walking $>$ walking-in-place $>$ flying, in virtual environments. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 359–364. ACM Press/Addison-Wesley Publishing Co. New York, NY, USA, 1999.

[UCAS00] M. Usoh, E. Catena, S. Arman, and M. Slater. Using presence questionnaires in reality. *Presence: Teleoperators & Virtual Environments*, 9(5):497–503, 2000.

[US68] W.R. Uttal and P. Smith. Recognition of alphabetic characters during voluntary eye movements. *Perception and Psychophysics*, 3:257–264, 1968.

[VBR+03] F. VanderWerf, P. Brassinga, D. Reits, M. Aramideh, and B. Ongerboer de Visser. Eyelid movements: behavioral studies of blinking in humans under different stimulus conditions. *Journal of neurophysiology*, 89(5):2784–2796, 2003.

[VC98] H.H. Vilhjálmsson and J. Cassell. BodyChat: autonomous communicative behaviors in avatars. *Proceedings of the second international conference on Autonomous agents*, pages 269–276, 1998.

[Ver99] R. Vertegaal. The GAZE groupware system: mediating joint attention in multiparty communication and collaboration. *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, pages 294–301, 1999.

[VGRG81] J.A. Van Gisbergen, D.A. Robinson, and S. Gielen. A quantitative analysis of generation of saccadic eye movements by burst neurons. *Journal of Neurophysiology*, 45(3):417–442, 1981.

[VGSS04] V. Vinayagamoorthy, M. Garau, A. Steed, and M. Slater. An Eye Gaze Model for Dyadic Interaction in an Immersive Virtual Environment: Practice and Experience. *Computer Graphics Forum*, 23(1):1–11, 2004.

[Vin06] V. Vinayagamoorthy. *User Responses to Virtual Humans in Immersive Virtual Environments*. PhD thesis, PhD thesis, University College London, University of London, London, UK, 2006, 2006.

[Vri94] A. Vrij. The impact of information and setting on detection of deception by police detectives. *Journal of Nonverbal Behavior*, 18(2):117–136, 1994.

[VSS05] V. Vinayagamoorthy, A. Steed, and M. Slater. Building Characters: Lessons Drawn from Virtual Environments. *Proceedings of Toward Social Mechanisms of Android Science: A CogSci 2005 Workshop*, pages 119–126, 2005.

[VSvdVN01] R. Vertegaal, R. Slagter, G. van der Veer, and A. Nijholt. Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 301–308. ACM Press New York, NY, USA, 2001.

[VWSC03] R. Vertegaal, I. Weevers, C. Sohn, and C. Cheung. GAZE-2: conveying eye contact in group video conferencing using eye-controlled camera direction. *Proceedings of the conference on Human factors in computing systems*, pages 521–528, 2003.

[WAS+00] J. Wideström, A.S. Axelsson, R. Schroeder, A. Nilsson, I. Heldal, and Å. Abelin. The collaborative cube puzzle: a comparison of virtual and real environments. In *Proceedings of the third international conference on Collaborative virtual environments*, pages 165–171. ACM New York, NY, USA, 2000.

[WBJ+67] P. Watzlawick, J.H. Beavin, D.D. Jackson, et al. Pragmatics of human communication: A study of interactional patterns, pathologies, and paradoxes. 1967.

[Wei66] J. Weizenbaum. ELIZA-a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, 1966.

[WGT57] R.D. Walk, E.J. Gibson, and T.J. Tighe. Behavior of light-and dark-reared rats on a visual cliff. *Science*, 126(3263):80–81, 1957.

[Wil77] E. Williams. Experimental comparisons of face-to-face and mediated communication: A review. *Psychological Bulletin*, 84(5):963–976, 1977.

[Wil90] L. Williams. Performance-driven facial animation. In *Proceedings of the 17th annual conference on Computer graphics and interactive techniques*, page 235. ACM New York, NY, USA, 1990.

[Wor10] WorldViz. Live Characters, http://www.worldviz.com/products/livecharacters. Software, 2010.

[WPKD02] J.B. Walther, M.R. Parks, M.L. Knapp, and J.A. Daly. Handbook of interpersonal communication, 2002.

[WSC+05] J.J. Walczyk, J.P. Schwartz, R. Clifton, B. Adams, M. Wei, and P. Zha. Lying person-to-person about life events: A cognitive framework for lie detection. *Personnel Psychology*, 58(1):141–170, 2005.

[WSC06] J.T. Wang, M. Spezio, and C.F. Camerer. Pinocchios pupil: Using eyetracking and pupil dilation to understand truth-telling and deception in games. *forthcoming in American Economic Review*, 2006.

[WVdBS+95] R.J. Wouters, W.A. Van den Bosch, T. Stijnen, A.C. Bubberman, H. Collewijn, and H.G. Lemij. Conjugacy of eyelid movements in vertical eye saccades. *Investigative ophthalmology & visual science*, 36(13):2686–2694, 1995.

[WWEP94] B. Winn, D. Whitaker, D.B. Elliott, and N.J. Phillips. Factors affecting light-adapted pupil size in normal human subjects. *Investigative ophthalmology & visual science*, 35(3):1132, 1994.

[Yar67] A.L. Yarbus. *Eye movements and vision*. Plenum press, 1967.

[YB07] N. Yee and J. Bailenson. The Proteus effect: the effect of transformed self-representation on behavior. *Human Communication Research*, 33(3):271, 2007.

[YB09] N. Yee and J.N. Bailenson. The Difference Between Being and Seeing: The Relative Contribution of Self-Perception and Priming to Behavioral Changes via Digital Self-Representation. *Media Psychology*, 12(2):195–209, 2009.

[YBU+07] N. Yee, J.N. Bailenson, M. Urbanek, F. Chang, and D. Merget. The Unbearable Likeness of Being Digital: The Persistence of Nonverbal Social Norms in Online Virtual Environments. *CyberPsychology & Behavior*, 10(1):115–121, 2007.

[ZDR81] M. Zuckerman, B.M. DePaulo, and R. Rosenthal. Verbal and nonverbal communication of deception. *Advances in experimental social psychology*, 14:1–59, 1981.

[Zha03] S. Zhao. Toward a taxonomy of copresence. *Presence: Teleoperators & Virtual Environments*, 12(5):445–455, 2003.

[ZKD81] M. Zuckerman, R. Koestner, and R. Driver. Beliefs about cues associated with deception. *Journal of Nonverbal Behavior*, 6(2):105–114, 1981.