Lie Tracking: Social Presence, Truth and Deception in Avatar-Mediated Telecommunication

William Steptoe*[1], Anthony Steed[1], Aitor Rovira[1], John Rae[2]
[1]Department of Computer Science, University College London, UK
[2]School of Human and Life Sciences, Roehampton University, UK
*W.Steptoe@cs.ucl.ac.uk

ABSTRACT
The success of visual telecommunication systems depends on their ability to transmit and display users' natural nonverbal behavior. While video-mediated communication (VMC) is the most widely used form of interpersonal remote interaction, avatar-mediated communication (AMC) in shared virtual environments is increasingly common. This paper presents two experiments investigating eye tracking in AMC. The first experiment compares the degree of social presence experienced in AMC and VMC during truthful and deceptive discourse. Eye tracking data (gaze, blinking, and pupil size) demonstrates that oculesic behavior is similar in both mediation types, and uncovers systematic differences between truth telling and lying. Subjective measures show users' psychological arousal to be greater in VMC than AMC. The second experiment demonstrates that observers of AMC can more accurately detect truth and deception when viewing avatars with added oculesic behavior driven by eye tracking. We discuss implications for the design of future visual telecommunication media interfaces.

Author Keywords
Avatar- and Video-Mediated Communication, Eye Tracking, Social Presence, Virtual Environments, Trust, Deception

ACM Classification Keywords
H.4.3 Communications Applications: Computer conferencing, teleconferencing, and videoconferencing; I.3.7 Three-Dimensional Graphics and Realism: Virtual Reality

General Terms
Experimentation, Human Factors, Theory

INTRODUCTION
Visual telecommunication systems enable remote users to interact synchronously, by means of natural linguistic and nonverbal channels of communication. Video-mediated communication (VMC) provides a rich mode of interpersonal remote interaction, capturing and displaying users' behavior including facial expression and, potentially, gaze direction. Recently, avatar-mediated communication (AMC), in which users are embodied by graphical humanoids within a shared virtual environment (VE), has rapidly increased in prevalence and popularity as an emerging form of visual remote interaction [17, 31]. However, compared with a live human face on video, the current paucity of avatar expression generally fails to capture and display many nonverbal cues, significantly hindering the richness of AMC [20].

Through two experiments, this paper investigates a number of factors related to user behavior, social presence, and media richness during truthful and deceptive discourse in visual telecommunication systems. Primarily we focus on AMC, but we also investigate VMC. A central theme of the work is the investigation of eye tracking for interactive and analytical uses. Interactively, eye tracking is used to drive avatars' oculesic behavior of gaze, blinking, and pupil size during AMC. Analytically, eye tracking is used to examine user behavior during both AMC and VMC. We frame our work in the social domain of interpersonal trust and deception, which we consider to present a compelling array of issues by which to investigate social interaction in visual telecommunication systems. Humans exhibit verbal and, critically, nonverbal behavioral cues correlated with lying and truth telling [15]. Correspondingly, people detect deception by observing others, so media transmitting fewer nonverbal channels becomes preferable for the deceiver [7].

Our first experiment (E1) explores truthful and deceptive discourse between dyads (*N=22* pairs) in state-of-art VMC and AMC systems. During the experimental interactions, a confederate issued questions to participants, who responded either with truths or with lies. Eye tracking was used to capture participants' oculesic behavior of gaze, blinks, and pupil size during AMC and VMC, for post-experimental analysis. Following the interactions, a questionnaire collected data describing participants' psychological arousal and mood state, which is also indicative of the degree of social presence engendered by the experience. The primary question E1 seeks to answer is whether users' behavior and response is similar in both AMC and VMC. 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 behavior and psychological response that is similar to the face-to-face interaction presented by VMC? In terms of accessibility and quality of nonverbal exchange, VMC is the dominant form of visual telecommunication, and provides a benchmark by which to measure AMC.

CHI 2010, April 10-15, 2010, Atlanta, Georgia, USA.

Our second experiment (E2) investigates the impact of bestowing avatars with faithful reproduction of their embodied users' tracked oculesic behaviors of gaze, blinks, and pupil size. E2 follows on from E1, inviting a different set of participants (*N=27*) to view replays of the interactions performed in E1. Participants assess the interactions in terms of veracity, rated by an estimate of truthful or deceptive behavior; engagement level, indicating how interested in the interaction the embodied user appears to be; and also confidence levels relating to the two judgments. These ratings were performed over three stimuli conditions: avatars exhibiting oculesics, avatars featuring no oculesics, and audio-only. Hence, E2 investigates avatar fidelity, and how the increased nonverbal bandwidth afforded by oculesic reproduction of user behavior may influence AMC. Therefore, the major question this experiment seeks to answer is whether observers of AMC are able to detect truth and deception more accurately when an avatar's oculesic behavior replicates the embodied user's.

The main contributions of this paper are as follows. Firstly, analysis of eye tracking data captured in E1 demonstrates systematic differences in users' gaze and pupil size, correlated with telling truths and lies in both AMC and VMC. To the authors' knowledge, this is the first combined analysis of gaze, blinking and pupil dilation in any field. Secondly, E1 also shows that users behave similarly in both forms of mediated communication, but that psychological arousal is greater in the video-based interactions. Thirdly, E1 demonstrates the interactive use of eye tracking to reconstruct a user's gaze, blinking, eyelid movement, and pupil dilation in a virtual humanoid avatar for real-time telecommunication. Finally, E2 shows that observers of AMC are able to more accurately detect truth and deception when avatars replicate the oculesic behavior of their embodied user.

SOCIAL PRESENCE AND MEDIA RICHNESS

Introduced by Short, Williams, and Christie [30], the theory of social presence explains the extent to which social cues are conveyed by different forms of mediated communication. The authors define social presence as the degree of salience of a conversational partner in a one-to-one interaction. Audio and text media fails to convey the visual cues present in collocated interaction including gaze, facial expression, gesture, and proximity. Thus, a medium's capability to transmit and display users' natural nonverbal behavior defines its potential to support social presence amongst its users. In the VE literature, the terms 'social presence' and 'copresence' are often used interchangeably, but both definitions aim to capture the sense of being in the company of another person during the course of mediated interaction. Measurement of social presence in mediated communication has been performed using questionnaires, task performance, and behavioral response [19].

A closely-related theory in HCI, which owes its origins to social presence, is Daft and Lengel's media richness theory [10]. Media richness theory describes the ability of a medium to transmit and reproduce information about the individuals who are communicating, and 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 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 [11]. Positioning these theories in the context of our two experiments, social presence describes the sense of collocation with a remote user during mediated interaction, and is measured by users' behavior and response during the interactions performed in E1. Media richness describes the ability of a medium to communicate information, and is assessed over varying levels of avatar fidelity in E2.

AVATAR- AND VIDEO-MEDIATED COMMUNICATION

VMC provides a rich mode of remote interaction between dyads and small groups. In comparison with audio conferencing, VMC has been found to improve users' ability to show understanding, forecast response, enhance verbal description, express attitudes, and manage pauses [22]. Several technological tiers of VMC systems exist, and here we are interested in state-of-art solutions which support high definition video and audio streams, display remote interactants at life-size, and support 'gaze awareness'. Gaze is of central importance in social behavior and nonverbal communication [1], and correspondingly, gaze awareness is a key requirement for effective VMC that has been shown to improve both task performance and sense of social presence [23]. Gaze awareness implies that users are able to determine where others are looking using natural lines of sight, and also allows perception of mutual gaze (eye contact). Technically, gaze awareness may be achieved through physical alignment of cameras and displays to enable natural lines of sight operating within Chen's offset threshold of 1° horizontal 5° vertical [8]. This narrow threshold indicates that the support for communicative gaze in VMC is not robust, and even minor physical movement of a user may introduce parallax between camera position and video display resulting in loss of gaze awareness. VMC presents a compressed representation of 3D space, constraining the rich spatial cues common to collocated interaction such as depth, resolution, and field of view. In the regard to spatiality, VMC has proven to be more similar to audio conferencing than to unmediated interaction [37].

Some of the spatial limitations of VMC can be overcome by immersive collaborative virtual environments (ICVEs), which connect remote users of immersive projection technology (IPT) systems, such as the CAVE™, within a spatial, social and informational context, with the aim of supporting high-quality interaction [29]. Several classes of collaborative VE systems exist, which may be categorized by the level of *immersion* they support. Immersion is defined by the technology a system is comprised of, which determines the degree to which a user's sensory input channels to all modalities are stimulated by the VE interface [16]. This work investigates state-of-art AMC supported by an ICVE system operating between networked IPT displays featuring perspective-correct graphical rendering via head tracking, life-size display of remote users' avatars, and body and eye tracking. User embodiment is a fundamental issue to ICVEs, and is typically maintained using an avatar: a graphical representation of a human [5]. Avatars generally exhibit humanoid form which grants a direct relationship between the natural movement of a user, and the corresponding animation of their avatar, which is generally tracked with minimal bodily tracking sensors attached to the head, trunk, and limbs [3]. This control metaphor becomes critical in multi-user VEs, as users' embodiments act directly as communication mediators, providing information regarding position, identification, focus of attention, gesture, and action [33]. While ICVEs are able to support spatial collaboration between remote users, the minimal tracking generally employed, and subsequent low levels of expressivity and fidelity demonstrated by avatars, is currently the critical constraint toward supporting high-quality interpersonal AMC. Our ICVE system, *EyeCVE*, integrates head-mounted eye tracking to faithfully replicate users' oculesic behaviors of gaze, blinks, eyelid movement, and pupil size, which are simulated and displayed by their virtual embodiments at remote sites.

DECEPTION, DETECTION, AND OCULESICS

Deception, or lying, is a deliberate attempt to mislead others. Deception is defined as an act intended to foster in another person a belief or understanding which the deceiver considers false [39]. Lying is a fact of everyday life, and it is estimated that people tell an average of one or two lies a day [14]. People's motivation for lying varies, 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 [15]. 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 [15]. Most lies are unremarkable, but betrayals of intimacy or trust, or those told in critical situations, can have serious implications.

In their meta-analysis collating results of previous studies, DePaulo et al. present a comprehensive catalog of the extent to which 158 nonverbal and verbal cues occur during deceptive communication [15]. 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 are influenced by a subject'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 [13]. Cognitive load, and emotional response to both positive and negative stimuli is correlated with pupil dilation [28]. A person's pupils dilate when they communicate deceptive messages, and the amount of dilation has been shown to increase linearly with deceptions of greater magnitude [36]. Gaze behavior is not classified so clearly, and is highly dependent on the social scenario, and idiosyncrasies, and culture. Indeed, studies have shown liars to gaze at their conversational partner both for less [24] and more [6] time than truth tellers. Finally, blink rate is influenced by cognitive load, becoming less frequent as cognition increases, and followed by a compensatory effect in the form of increased blinking after demand drops [25].

In many situations, people question the credibility or accuracy of statements that another person makes. The 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 [12]. An explanation for this poor performance is that 'detectives' seem to focus on the wrong behaviours when trying to distinguish lies from truths [40]. It has been shown that providing people with knowledge about cues to deception can improve detection accuracy [34]. 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 [26]. In 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 [7]. For instance, VMC allows rich exchange of nonverbal cues including facial expression, oculesics, and upper-body language, and studies have shown that peoples' ability to detect deception is statistically identical when assessing subjects via video or in reality [21].

HYPOTHESES

E1 investigates truthful and deceptive discourse in AMC and VMC, and E2 studies detection of deception in AMC. Drawing on reviewed work in the HCI, VE, and social science literature, we make the following hypotheses preceding our two experiments:

- *E1H1:* During AMC and VMC, participants will exhibit similar oculesic behaviors of gaze, blink rate, and pupil size. 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 performing AMC, judgments will be more accurate and confident when observing avatars featuring faithful reproduction of oculesic behavior than judgments of avatars with 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 behavior.

Figure 1. EyeCVE supporting AMC between users of the WALL (left) and ReaCToR (right) immersive projection systems.

EXPERIMENT 1: INTERACTIONS

Technical Setup

ICVE System

Our ICVE system, EyeCVE [38], supported AMC between users of two networked IPT displays that were physically located in adjacent rooms. The ReaCToR is UCL's immersive CAVE™-like system featuring perspective-correct stereoscopy and four surrounding display walls projected at 1024×768 pixels. The second display, referred to as the WALL hereafter, consisted of a single full HD (1080p) projection display wall with perspective-correct mono rendering. In both displays systems, tracking captured a user's head, hand, and oculesic behavior, and EyeCVE used these data streams to drive the behavior of an avatar embodiment, which appeared life-size and animated at 60 frames-per-second on the partner's display. Audio communication between users of the ReaCToR and the WALL was established externally to EyeCVE with Google™Talk. Figure 1 shows EyeCVE running in both display systems.

Our previous work demonstrated the use of eye tracking to drive avatar gaze [32]. The current study introduces blinking and pupil size as two additional behaviors which are similarly tracked and represented. Figures 2 and 3 show screen-captures in EyeCVE of avatar pupil dilation and blinking driven by eye tracking. Reproduction of pupil size was achieved by determining the current user's natural pupil size range by placing them in a high-luminance environment (to trigger extreme constriction), followed by a low-luminance environment (triggering extreme dilation) during eye tracker calibration. The eye tracker then monitored and streamed current pupil size at a rate of 60Hz (every 16ms) to the local EyeCVE client, which resolved the value relative to the established range, and distributed the changes to the remote client. Upon reception of updated pupil size data, EyeCVE's animation system blended between the two eye models shown in Figure 2 to faithfully represent a user's current pupil size in their avatar embodiment. Blinks were detected by monitoring the user's pupil aspect ratio with the eye tracker: as the eyelid comes down during a blink, the elliptical fit to the pupil becomes increasingly flat before it disappears, and this characteristic change in aspect ratio of the ellipse may be used to detect blinks. While the threshold required minor adjustment for each user, a mean aspect ratio of 0.6 (1.0 being a perfect circle) provided robust classification of blink signals. Eyelid kinematics

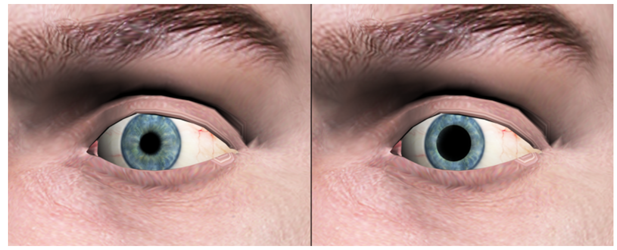

Figure 2. Screen-captures in EyeCVE of extreme constriction (left) and extreme dilation (right) of male avatar's pupils.

Figure 3. Screen-captures in EyeCVE of female avatar blinking.

during a blink were animated using a model which generated physiologically-accurate human blink motion. Lid saccades (vertical movements of the eyelids which accompany vertical changes in gaze [18]) were also simulated by a model.

Video Conferencing System

VMC was hosted between the ReaCToR's front-wall and WALL displays. Gaze awareness was achieved by aligning cameras and displays of remote users within Chen's threshold [8]. Video was streamed at 1080i resolution in the WALL and 720p in the ReaCToR 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 audio communication was supported by Google™Talk. Figure 4 shows VMC viewed in the WALL.

Eye Tracking and Data Collection

Eye tracking was achieved using the head-mounted ViewPoint EyeTracker from Arrington Research shown in Figure 5. A camera mounted on the eye-glass frame recorded the scene from the wearer's perspective at 150° field-of-view (close to human limitation ∼180°). The wearer's foveal fixation point is overlaid on the scene video, allowing for post-session analysis of action once synchronized with the separate audio stream. The eye tracker's logging facility also served as the primary method for experimental data collection, with the tracked behavior of users being streamed to the process and output to a log file at 60Hz. Data included gaze (2D X/Y coordinate, 3D hit-point in EyeCVE, together with fixation and saccade timings), blink signals, pupil size, speech signals, head and hand tracking, and markup data input by the experimenter.

Experimental Design

Our first experiment compared users' behaviors 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 truth telling and deception. A

Figure 4. Users interacting via the video conferencing system.

Figure 5. Eye tracker mounted on the WALL's lens-free frame (left), and on the ReaCToR's CrystalEyes® 3 shutter glasses (right).

total of 22 participants (11 male) with normal or corrected vision and no previous telepresence experience were paid £10 to perform the experiment. Ages ranged from 19-62, and European, Asian and African origins were represented. Upon arrival, participants completed the NEO personality inventory [9], and were then assigned either to the AMC or the VMC condition depending on social score and gender, with the intention of balancing the groups. The participants read the experimental instructions while the experimenters performed technical setup.

Our experimental task design is inspired by Walczyk et al.'s question-answer framework for manipulating cognitive load when lying and truth telling [35]. That study consisted of five stages of questions that were issued by a confederate and answered by a participant. Each stage consisted of open and closed (yes/no) questions regarding a single topic including general personal information, recent personal events, and remote personal events. During each stage, participants were instructed to either 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 (two general, two recent, two remote) each containing around ten unique questions. Thus, participants lied during three stages (one general, one recent, one remote), and told the truth in the remaining three. See [35] for the question list. Unlike Walczyk et al., we do not aim to investigate behavioral differences between stages. Rather, we consider it an effective framework for eliciting thoughtful response. Stages of questions, but not questions within a stage, were resequenced over participants and mediation type.

The same male confederate questioned all participants in all AMC and VMC sessions, and was blind to participants' current state of veracity. Participants were aware of this, which intended to motivate them to lie plausibly and also to negate any inter-experimental effects in this regard. The low ambient luminance of the ReaCToR caused users' pupils to remain in a dilated state, leading to limited variation of pupil size response. Hence, participants were positioned in the WALL, and the confederate in the ReaCToR. The technical setup procedure for both AMC and VMC sessions was similar despite the differences in mediation. Setup of AMC involved running EyeCVE clients for display on the WALL and ReaCToR, networked via an EyeCVE server. The clients loaded the shared 'meeting room' VE shown in Figure 1, together with the partner's avatar embodiment (male model for confederate, and male or female model for participant). 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 3D avatar eye rotation was performed in EyeCVE. The confederate calibrated himself in the ReaCToR similarly. Finally, microphones were positioned to ensure clear verbal communication for the experimental interaction. Preparation of VMC adhered to an identical procedure, but projector input was switched to the video stream following setup completion. Hence, EyeCVE operated in the 'background' during the VMC sessions for the following reasons: collection of bodily tracking data was managed by EyeCVE, and required filtering and processing before being sent for log file output; and to enable the interactions, visualized as AMC, to be captured in order to act as stimuli for E2.

Following setup, the experimenter left the WALL lab to observe the interactions and provide markup data for the logging process. The experimental interactions began with the confederate and participant greeting each other in the AMC or VMC. 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 instruction. Following the sixth stage, the experimental interaction concluded, and the confederate bid farewell to the participant. Immediately following the interaction, participants completed the 'Profile of Mood States' (POMS) [27] questionnaire measuring current mood state over six psychological traits: tension, anger, depression, fatigue, confusion, and vigor. Finally, the experimenter conducted an informal interview with the participant.

Results

The majority of this analysis is based on log files recorded by the eye tracker worn by 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 analysis of variance (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 representation) and *away* (looking elsewhere in the visual field). Table 1 shows 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. It is greater when telling lies than when telling truths in both AMC and VMC. The consistent trends in gaze behavior between conditions of mediation and veracity supports the validity of our experimental method and data collection. However, these differences were not great enough to expose a main effect between veracity condition or mediation condition when gaze behavior was assessed on this holistic scale. This result is to be expected, as analysis on this scale considers gaze behavior over entire stages (mean duration 1.5 minutes), and so gaze behavior during critical moments of question response is not explicitly targeted. Hence, we analyzed 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. We considered the period of time from when the confederate reached the semantic point of the question (i.e. following utterance of "book" and "evening") until two seconds following the end of participants' response speech. 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%; a highly significant difference ($p<.01$). VMC participants demonstrated similar behaviour, 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 veracity conditions ($p<.01$). Finally, no significant difference in gaze behavior was found between classes of AMC/truth and VMC/truth, or between classes of AMC/lies and VMC/lies.

Table 1. Mean (and standard deviation) percentage 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)

Blinking

Table 2 shows the mean number and standard deviation of blinks 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 condition 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 generalize blink patterns over multiple participants.

Table 2. 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)

Pupil Dilation

Table 3 shows 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 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. Table 3 illustrates that mean pupil size is larger when lying than when telling the truth in both mediation types, and that participants' pupils are more dilated during VMC than AMC. However, ANOVA calculations performed at this macro 'stage' scale failed to expose a main effect in pupil size between conditions of mediation or veracity. Hence, similarly to gaze analysis, we focused on critical moments of interaction 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 *General 2* stage, and "What's the name of the school you attended?" from *Remote 1*. 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") and finishing two seconds after the end of participants' vocal response. Considering both questions, mean change in AMC participants' pupil size 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 behavior, with a 0.11 for truth tellers and 0.38 for liars, again exposing a main effect ($p<.01$) between veracity condition. 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.

Table 3. 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)

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 confederate via AMC or VMC. The POMS questionnaire, completed immediately following the interactions aimed to capture participants' current psychological state. Table 4 shows the mean and standard deviation of 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, fatigue, confusion, and vigor (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 significant differences to lie between all mood-factors ($p<.05$) excluding anger.

Table 4. 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)

Discussion

Our analysis approached the experimental interactions from two perspectives: comparison of AMC and VMC measured by eye tracking and psychological affect, and variance in behavior between truthful and deceptive discourse in both mediation types measured by eye tracking. Comparing AMC and VMC, participants' oculesic behavior of gaze, blinking, and pupil size was found to be similar in both forms of visual telecommunication. Gaze distribution was similar between mediation conditions, with proportion of gaze at the confederate comparable to classic studies of dyadic conversation in the social science literature [2]. Participants gazed slightly, but insignificantly, more at the confederate when the visual component of communication was depicted by a virtual embodiment than when faced with video, and this held true when considering both macro scale of question stage, and micro scale of individual questions. Blink rate was similar between mediation types, with participants blinking slightly, but insignificantly, more frequently when engaged in AMC than when performing VMC. Pupil size was also comparable when analyzed on both macro and micro scales, and no main effect between AMC and VMC was found. However, results of the subjective self-ratings elicited by the POMS questionnaire indicated significantly higher arousal of both positive and negative moods following VMC than AMC. In summary, our results support hypothesis *E1H1*: participants exhibited comparable oculesic behavior during interaction in both mediation types. Pupil size was consistently, but insignificantly, larger during VMC than AMC, tentatively suggesting that participants engaged in VMC experienced a higher degree of arousal than those in AMC. This conjecture is supported by the striking difference between affective states uncovered by the POMS questionnaire following the experimental interactions.

With regards to variance in oculesic data between conditions of truthful and deceptive response, analysis of both gaze and pupil size revealed significant differences. When considering the macro scale of stages, lying participants in both mediation types gazed at the confederate for a higher proportion of time than truthful tellers. However, a main effect of gaze behavior was not exposed until the micro scale of individual questions was considered. This demonstrated that, during critical sequences of interaction, a participants' state of veracity is highly influential to their gaze behavior. Pupil size was also affected by states of truth and deception, with increased pupil dilation observed when lying than when truth telling. Again, analysis of key interactional periods revealed significantly higher levels of arousal, measured by increased pupil size, when engaging in deceptive discourse compared with truthful response. Finally, strengthening the argument that participants behaved similarly in AMC and VMC, differences in both truthful behavior and deceptive behavior between both forms of visual telecommunication were insignificant. In summary, with the exclusion of blink behavior, our results support *E1H2*. Gaze behavior was found to differ between states of veracity, with deceptive discourse eliciting greater proportions of gaze directed at the confederate in both AMC and VMC. Pupil size was also greater during deception in both AMC and VMC.

EXPERIMENT 2: DETECTING DECEPTION

Technical Preparation

Our second experiment aimed to assess the impact of avatars' oculesic behavior on the extent to which observers are able to detect truthful and deceptive messages in AMC. Video and audio captures of E1's interactions provided the experimental stimuli. During E1, two EyeCVE 'spectator' clients captured the unfolding AMC regardless of whether the actual mediation was being performed via video or avatar. 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 synchronized with vocal input. Both spectator clients observed the participant avatar from a perspective approximating the confederate's viewpoint during the experimental interactions. Fraps®(Beepa®) was used to capture video at 1200×720 pixels at 50 frames-per-second and audio streams of both confederate and participant talk on each spectator client. The videos were then divided by question stage, resulting in a total of 132 clips over the 22 participants. A viewer application illustrated in Figure 6 was developed using Adobe®Flash®. In terms of quality and perspective, the resulting audio-visual stimuli aimed to approximate that which was originally observed by the confederate during the AMC. Thus, E2 intended to gather many observers' perceptions of the stimuli that may approximate judgments made by the single confederate during the original AMC.

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. Ages ranged from 18-55, and European, Asian, and African origins were represented. The participants were divided into three groups (A, B, and C) of nine, with

Figure 6. E2's clip viewer interface.

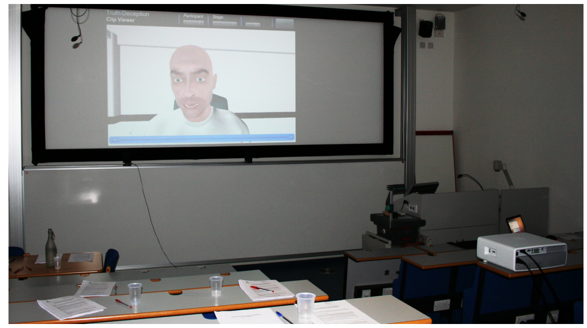

Figure 7. Photograph of E2's setup, conducted in a lecture theater.

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 behavior driven by eye tracking (*ET*); video and audio with avatars featuring no eye oculesic behavior (*-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 featured a 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, and also to encode judgments in a richer manner than binary classification.

Each group rated the same clips in different conditions, which cycled as clips advanced. 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 theater, with the clips projected in their native resolution of 1200×720 pixels, and physically around 3×2 meters, as shown in Figure 7. 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.

Results

Table 5 shows all participants' mean scores of accuracy, engagement, and confidence between the three experimental conditions, normalized to 1 and converted to percentages. The table shows that ET enables more accurate and confident assessment of the veracity of users engaged in both truthful and deceptive AMC than -ET or AO. Hence, avatars displaying faithful replication of embodied users' oculesic behavior are able to inform more accurate judgment of the users' state of truth telling and deception than observation of avatars featuring no oculesic behavior (-ET), or audio-only stimuli (AO). The large difference in accuracy between judgments of truths and lies exposes the influence of the veracity effect [26], 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.

Table 5. Percentages of 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	Accuracy	88%	70%	68%
	Confidence	78%	69%	72%
	Engagement	76%	58%	70%
	Confidence	85%	79%	82%
Deception	Accuracy	48%	39%	34%
	Confidence	74%	69%	73%
	Engagement	78%	54%	72%
	Confidence	83%	76%	80%

Figure 8 illustrates where the main effects lie between the three conditions and for each score 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 lies between these two conditions, but that a significant difference does exist between both conditions and ET. This relationship also applies for accuracy of lie detection. 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 judgments, but this difference is only significant when judging truthful discourse.

Discussion

Our results support hypothesis *E2H1*: avatars featuring oculesic behavior 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. Oculesic behavior raises raters' confidence levels, but only significantly when judging truthful interaction. With regards to engagement, eye tracked avatars were

TRUTHS

Accuracy

ET	-ET	AO
88%	70%	68%

Confidence

ET	-ET	AO
78%	69%	72%

Engagement

ET	AO	-ET
76%	72%	58%

Confidence

ET	-ET	AO
85%	79%	82%

LIES

Accuracy

ET	-ET	AO
48%	39%	34%

Confidence

ET	-ET	AO
74%	69%	73%

Engagement

ET	AO	-ET
78%	72%	54%

Confidence

ET	-ET	AO
83%	76%	80%

Figure 8. 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.

rated similarly to audio-only stimuli, and significantly higher than avatars with no oculesic expression. Participants were more accurate when judging truths than lies, demonstrating the veracity effect. The influence of this truth bias 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. For instance, when asked about what they did "last" Saturday, many participants responded with information regarding some other Saturday. Finally, it should be noted that reported accuracy of deception detection varies between studies and experimental design, so a discussion of our results in relation to others is problematic.

CONCLUSIONS AND FUTURE WORK

E1 demonstrated that users' oculesic behavior is similar during AMC and VMC, but that users' psychological affect is greater when faced with video. E2 showed that avatars featuring oculesic behavior driven by their embodied users are able to transmit nonverbal cues critical to informing accurate perception of truth and deception in AMC. Overall findings present implications for the design of future visual telecommunication media interfaces, with particular regard to matching an intended application with desirable characteristics of social presence and media richness. We can hypothesize from E1 that the degree of social presence experienced in AMC is less acute than that fostered by VMC, but that users' behavior in AMC is no less socially 'real' than that demonstrated in VMC. Also, we can hypothesize from E2 that AMC featuring oculesic avatars is able to provide a rich medium for interpersonal telecommunication in which users are able to transmit and recognize subtle nonverbal signals relating to underlying communicative intent. Hence, we suggest communication 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, could conceivably benefit from avatar mediation. This may include communication interfaces for virtual tutoring, people with social anxiety, or online socializing. Additionally, we suggest that critical or highly interpersonal telecommunication is more suited to the faithful representation of fellow interactants provided by video. In this sense, mediated telecommunication systems can be regarded as filters for behavioral 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. Hence, while there may be ethical issues to consider, this ability for transformed social interaction [4] may be exploited to suit the characteristics of a collaborative task, thereby potentially shaping the unfolding communication.

The mediated interactions captured in E1 were designed to manipulate participants' cognitive load and 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 behaviors were measured when they were being *told* to lie as opposed to *choosing* to be deceptive. In future work, we aim to investigate natural and unstructured interaction that is common to how computer mediated communication systems are used in daily life. Investigation of AMC between friends and strangers, and between naïve and experienced users is another potentially revealing avenue of research. 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. We consider facial expression and accurate lip synchronization as high priorities. Finally, given the variety of motion tracking interfaces and 3D displays for both home and commercial use currently in development, we are confident that more sophisticated avatar communication systems will be built and deployed over the coming years.

REFERENCES

1. M. Argyle and M. Cook. *Gaze and Mutual Gaze*. Cambridge University Press Cambridge, 1976.
2. M. Argyle and R. Ingham. Gaze, mutual gaze and proximity. *Semiotica*, 6(1):32–50, 1972.
3. N. Badler, M. Hollick, and J. Granieri. Real-time control of a virtual human using minimal sensors. *Presence*, 2(1):82–86, 1993.
4. J. Bailenson, A. Beall, J. Loomis, J. Blascovich, and M. Turk. Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments. *Presence: Teleoperators & Virtual Environments*, 13(4):428–441, 2004.
5. S. Benford, J. Bowers, L. Fahlén, C. Greenhalgh, and D. Snowdon. User embodiment in collaborative virtual environments. *Proc. SIGCHI conference on Human factors in computing systems*, pages 242–249, 1995.
6. J. Burns and B. Kintz. Eye contact while lying during an interview. *Bulletin of the Psychonomic Society*, 7(1):87–89, 1976.
7. J. Carlson, J. George, J. Burgoon, M. Adkins, and C. White. Deception in computer-mediated communication. *Group Decision and Negotiation*, 13(1):5–28, 2004.
8. 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.

9. P. Costa Jr and R. McCrae. Revised NEO Personality Inventory (NEOPI-R). *The Sage handbook of personality theory and assessment: Personality measurement and testing*, page 179, 2008.
10. R. Daft and R. Lengel. Information richness: A new approach to managerial behavior and organizational design. *Research in organizational behavior*, 6:191–233, 1984.
11. A. Dennis and J. Valacich. Rethinking media richness: Towards a theory of media synchronicity. In *Proc. 32nd Hawaii International Conference on System Sciences*, volume 1, pages 1–10. Citeseer, 1999.
12. B. DePaulo. Humans as lie detectors. *Journal of Communication*, 30(2):129–39, 1980.
13. B. DePaulo. Nonverbal behavior and self-presentation. *Psychological Bulletin*, 111(2):203–243, 1992.
14. B. DePaulo and D. Kashy. Everyday lies in close and casual relationships. *Journal of Personality and Social Psychology*, 74(1):63–79, 1998.
15. B. DePaulo, J. Lindsay, B. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper. Cues to deception. *Psychological Bulletin*, 129(1):74–118, 2003.
16. J. Draper, D. Kaber, and J. Usher. Telepresence. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 40(3):354–375, 1998.
17. N. Ducheneaut, N. Yee, E. Nickell, and R. Moore. Alone together?: exploring the social dynamics of massively multiplayer online games. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, page 416. ACM, 2006.
18. C. Evinger, K. Manning, and P. Sibony. Eyelid movements. Mechanisms and normal data. *Investigative ophthalmology & visual science*, 32(2):387–400, 1991.
19. M. Garau. *The Impact of Avatar Fidelity on Social Interaction in Virtual Environments*. PhD thesis, University College London, 2003.
20. M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M. Sasse. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. *Human factors in computing systems*, pages 529–536, 2003.
21. M. Hartwig, P. Granhag, L. Stromwall, and A. Vrij. Police officers' lie detection accuracy: Interrogating freely versus observing video. *Police Quarterly*, 7(4):429, 2004.
22. E. Isaacs and J. Tang. What video can and cannot do for collaboration: A case study. *Multimedia Systems*, 2(2):63–73, 1994.
23. 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.
24. M. Knapp, R. Hart, and H. Dennis. An exploration of deception as a communication construct. *Human Communication Research*, 1(1):15–29, 1974.
25. S. Leal and A. Vrij. Blinking During and After Lying. *Journal of Nonverbal Behavior*, 32(4):187–194, 2008.
26. T. Levine, H. Park, and S. McCornack. Accuracy in detecting truths and lies: Documenting the veracity effect. *Communication Monographs*, 66(2):125–144, 1999.
27. D. McNair, M. Lorr, and L. Droppleman. Profile of mood states. *San Diego, California*, 53:6, 1992.
28. 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.
29. 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.
30. J. Short, E. Williams, and B. Christie. *The social psychology of telecommunications*. John Wiley & Sons, 1976.
31. W. Steptoe, O. Oyekoya, A. Murgia, R. Wolff, J. Rae, E. Guimaraes, D. Roberts, and A. Steed. Eye Tracking for Avatar Eye Gaze Control During Object-Focused Multiparty Interaction in Immersive Collaborative Virtual Environments. In *Proc. 2009 IEEE Virtual Reality Conference*, pages 83–90. IEEE Computer Society Washington, DC, USA, 2009.
32. W. Steptoe, R. Wolff, A. Murgia, E. Guimaraes, J. Rae, P. Sharkey, D. Roberts, and A. Steed. Eye-tracking for avatar eye-gaze and interactional analysis in immersive collaborative virtual environments. In *Proc. ACM 2008 conference on Computer supported cooperative work*, pages 197–200. ACM New York, NY, USA, 2008.
33. D. Thalmann. The Role of Virtual Humans in Virtual Environment Technology and Interfaces. In *Proc. of Joint EC-NSF Advanced Research Workshop, Bonas, France*. ACM New York, NY, USA, 1999.
34. A. Vrij. The impact of information and setting on detection of deception by police detectives. *Journal of Nonverbal Behavior*, 18(2):117–136, 1994.
35. J. Walczyk, J. 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.
36. J. Wang, M. Spezio, and C. Camerer. Pinocchios pupil: Using eyetracking and pupil dilation to understand truth-telling and deception in games. *forthcoming in American Economic Review*, 2006.
37. E. Williams. Experimental comparisons of face-to-face and mediated communication: A review. *Psychological Bulletin*, 84(5):963–976, 1977.
38. R. Wolff, D. Roberts, A. Murgia, N. Murray, J. Rae, W. Steptoe, A. Steed, and P. Sharkey. Communicating Eye Gaze across a Distance without Rooting Participants to the Spot. In *Proc. IEEE/ACM Symposium on Distributed Simulation and Real-Time Applications-Volume 00*, pages 111–118. IEEE Computer Society Washington, DC, USA, 2008.
39. M. Zuckerman, B. DePaulo, and R. Rosenthal. Verbal and nonverbal communication of deception. *Advances in experimental social psychology*, 14:1–59, 1981.
40. M. Zuckerman, R. Koestner, and R. Driver. Beliefs about cues associated with deception. *Journal of Nonverbal Behavior*, 6(2):105–114, 1981.