PRESENCE 2008

Proceedings of the 11th Annual International Workshop on Presence
Padova, 16-18 October 2008

Printed by
CLEUP Cooperativa Libraria Universitaria Padova
Padova 2008

Edited by Anna Spagnolli, Luciano Gamberini

ISBN: 978-88-6129-287-1

Simulation versus Reproduction for Avatar Eye-Gaze in Immersive Collaborative Virtual Environments

John Rae[2], Estefania Guimaraes[1], William Steptoe[2]

[1]School of Human and Life Sciences Department of Computer Science
[2]Roehampton University, London University College London
{J.Rae@Roehampton.ac.uk}

Abstract

Using video technology in telecommunication systems poses spatial orientation problems because participants have a window onto each other's space, rather than sharing one space. These problems can be overcome using immersive virtual reality in which remote participants are represented as avatars. Because of the challenges involved in capturing eye-gaze reliably and problems of loss in transmission, the simulation of eye-gaze behaviour has been attempted. We review the work on eye-gaze models for avatars and autonomous agents. These models infer eye-gaze from interactional states (e.g. speaking or listening). However, through reviewing detailed analyses of social interaction we identify eye-gaze practices that cannot be inferred from participants' talk. We discuss the extent to which such practices constrain the prospects for the simulation of participants' eye-gaze in telecommunication systems, and discuss some future work for better designing autonomous agents and representing avatars.

Keywords--- **Gaze, Avatars, Agents, Telecommunication, Immersive Collaborative Virtual Environments, Simulation, Mediation**

1. Introduction

When people interact face-to-face, the surrounding environment is a contiguous space in which participants are able to use a full range of non-verbal communicational resources: they can move their eyes and head to look at others, change facial expression, gesture, posture, and move as desired. Despite development of highly sophisticated Computer Supported Cooperative Work (CSCW) systems, there is no substitute for such co-located interaction which supports rich non-verbal communication while allowing free movement in a perceptually unfragmented workspace.

Traditional video-mediated communication (VMC) systems provide remote participants with synchronous video and audio channels, and have been found to improve ability to show understanding, forecast responses, give non-verbal information, enhance verbal descriptions, manage pauses and express attitudes [1]. However, VMC compresses the representation of 3D space, constraining rich cues available in co-located collaboration such as depth, resolution and field of view, and thereby limiting awareness and the ability to point at and manipulate objects [2].

In terms of presenting a spatially-unified shared environment in which multiple remote participants are able to interact and move freely, Immersive Collaborative Virtual Environments (ICVEs) have been shown to be an effective medium for both communication and entertainment. ICVEs connect remote or co-located users of immersive display systems (such as the CAVE™) within a spatial, social and informational context, with the aim of supporting high-quality interaction [3]. Co-presence is the extent to which the computer becomes transparent and there is a sense of being present with other people in the virtual environment, and there is a direct working with the other people [4]. ICVEs often combine high degrees of presence and co-presence, because the sense of being in another place and of being there with other people reinforce each other [5]. Consequently, user embodiment is a fundamental issue when designing CVEs [6], and this is typically maintained using an avatar - a graphical representation of a human. In this paper we follow Bailenson and Blascovich's (2004) [7] definition and use the term 'avatar' to refer to a virtual human used to represent a participant to others in a shared virtual environment and 'agent' to refer to a virtual human with pre-scripted behaviours (see also Vinayagamoorthy, Steed and Slater 2005)[8].

Virtual humans are capable of eliciting appropriate responses from people, and it has been observed that unwritten social norms such as proxemics and unease from close-range mutual eye-contact with unknown partners occur in CVEs as they do in real-life. Bailenson and colleagues tested Argyle and Dean's (1965) equilibrium theory's specification of an inverse relationship between mutual gaze and interpersonal distance [9]. They found that all participants maintained more space around virtual humans than they did around similarly sized and shaped but nonhuman-like objects. Correspondingly, people's perception of their own virtual representation has been found to have a significant and instantaneous impact on user behaviour, with more attractive avatars prompting an increase in self-disclosure, and relatively taller avatars raising confidence during negotiation tasks [10].

Avatars exhibiting higher levels of visual and behavioural fidelity can potentially communicate the subtleties of human nonverbal communication more successfully, thereby enhancing the perceived authenticity of the interaction [8, 11]. A consistent interaction effect between visual and behavioural realism has been found [12], 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 the converse is true for lower fidelities. Therefore, it is the combination of the extent to which a person feels that they are in the presence of real person whose actions are represented by an avatar (termed social presence) and the degree to which an avatar behaves as people do in the real world (termed behavioural realism) which determines the level of social influence it can have on people [13]. Although social presence and behavioural realism cannot be separated during avatar interaction [14], they pose significantly different design and development challenges.

The behavioural realism of avatar gaze is particularly important. Representing gaze information has long been recognised as a requirement for natural communication through visual remote collaboration and conferencing systems [15]. This is a logical extension of Argyle's conviction that gaze is of central importance in social behaviour and non-verbal communication where it is used as a bidirectional channel monitoring initiation, maintenance and termination of messages [16]. (It is necessary to distinguish eye-gaze from *head-gaze* or (the focus of attention inferred from head-orientation); in this paper we use *gaze* to refer to eye-gaze.)

One strategy for driving avatar gaze behaviour is through tracking participants' gaze as introduced by Steptoe, Wolff, Murgia, Guimaraes, Rae, Sharkey, Roberts and Steed [19] (in press) and Wolff, Roberts, Murgia, Murray, Rae, Steptoe, Steed and Sharkey [17] (in press) which reproduces participants' gaze on their avatars by using mobile eye-trackers, thus allowing free-movement. There are however, considerable challenges in tracking, transmitting and representing gaze behaviour. Consequently, from an engineering point of view there are considerable benefits to, and a need for, at least some degree of avatar's gaze behaviour to be simulated. Work on the development of avatars and autonomous agents shows that there is scope for simulating participants' gaze behaviour. Simulation at the local site would circumvent the need for the capture and transmission of gaze information. In addition, even if remote participants' gaze is tracked, an eye-gaze simulation model is likely to be needed to cope with errors or failures of the eye-tracking system.

This paper explores the behavioural realism of avatar gaze. We firstly present methods of avatar control and present a critical review of gaze models, which have become the predominant method for the control of avatar gaze. These models infer eye-gaze from interactional states (e.g. speaking or listening). Such modelling strategies assume that a person's gaze behaviour can be inferred from other things that they are doing. Whether or not this assumption is generally true is an empirical matter. We review detailed analyses of social interaction and we identify certain eye-gaze practices that cannot be inferred from participants' talk.

We discuss the implications of this for the use of simulated gaze in telecommunication contexts and we identify directions for future research.

2. Avatar Control

Although the application domains of ICVEs are diverse, the fundamental scenarios which they simulate can be generalised as pure conversation (analogous to a face-to-face meeting or videoconference) [20] and object-focused tasks (analogous to manipulating and discussing artefacts or documents) [21]. Therefore, in each case and as in face-to-face scenarios, it would be a particular hindrance to communicate with other people while lacking the rich feedback and signalling abilities granted by natural (as opposed to inferred) non-verbal communication, particularly in emotionally charged or decisive situations.

In ICVEs, non-verbal communication is mediated through each participant's avatar, and there is ongoing research towards being able to represent participants' natural full-body movement, posture and facial expression in real-time to support interaction to the level we expect from co-located situations. There are two predominant methods of avatar control – *tracking* and *simulation*. These methods are often used in combination to control different aspects of an avatar.

2.1. Tracking (Reproduction)

Tracking methods aim to capture movements made by a human participant and to reproduce them on an avatar. Users of immersive display systems such as the CAVE wear a pair of liquid crystal shutter glasses to resolve the stereoscopic imagery. A head-tracker positioned on the shutter glasses together with a hand-held tracked input device for 3D interaction within the virtual environment allows the system to determine the location and orientation of the user's head and hand in real-time [22] (Leigh, DeFanti, Johnson, Brown, Sandin 1997). Hence, we are able to animate avatars from the head and hand tracking data, thus capturing some high-level non-verbal communication channels such as pointing, interpersonal space and head-orientation. Even these minimal cues have been shown to significantly contribute toward perception of other's visual attention [14].

Head-orientation is also a useful indicator of where a person is looking, and it has been suggested that (during conversation) gaze generally corresponds to head-orientation 90% of the time [23]. However, Murray and Roberts [24] recently determined that augmenting avatar head-orientation with gaze (replayed from pre-recorded eye-tracking data) is of vital importance for observers to be able to correctly identify where an avatar is looking for object selection in ICVEs [16].

Most ICVEs currently fail to track participant gaze; an issue which we addressed by developing *EyeCVE* [17, 18, 19] which uses mobile head-mounted eye-trackers to drive avatar gaze.

2.2. Modelling (Simulation)

Unlike tracking methods, modelling systems animate one behavioural aspect of an avatar based on inferences about *other* behaviours of a human participant. Models simulating one or more aspects of human nonverbal communication have been developed and investigated with the aim of presenting more believable avatars, and also agents, for real-time interaction in ICVEs, commercial videogame middleware, and pre-rendered character animation in computer-generated movies. The development of gaze and attention models has been a research focus; and some of the most significant work is summarised in the next section.

3. Gaze Models for agents and avatars

Gaze models have been developed toward simulation of naturalistic gaze behaviour for avatars and agents. The operation of the majority of gaze models cited here takes a state-based goal-directed approach to gaze, focusing on major gaze changes such as those for creating conversational turn taking. Values for behavioural properties such as fixation duration, angular velocity and saccade magnitude act as input parameters for a broader analytical model. These timings implement statistical generalisations about human gaze behaviour often derived from social science literature. Timings have been shown to be able to change an avatar's mental state, such as excited or sleepy [25].

3.1. Statistical gaze models

The Eyes Alive [26] model was based on both empirical studies of saccades and statistical models of eye-tracking data from a single dyadic (two-person) conversation. The speaker was allowed to move her head freely while the video was recorded. Eye trajectory kinematics were extracted from the eye-tracking session, which was further segmented and classified into two modes: *talking* and *listening*. Thus, the model took into account the dynamic characteristics of eye-movement, including saccade magnitude, direction, duration, velocity, and inter-saccadic interval. The authors used an autonomous virtual agent head (not full-body and not user-controlled) using three types of gaze control: stationary, random, and model-based. On a non-immersive display, subjects were asked to give feedback relating to the perceived naturalness of the character's gaze. The results showed 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 character.

Garau, Slater, Bee, and Sasse [27] presented a parametric gaze model which took timings from research on face-to-face dyadic conversations by Argyle and Cook (1976), Argyle and Ingham (1972) and Kendon and Cook (1969). Similarly to the Eyes Alive model, animations were based on "while speaking" and "while listening" modes. For the talking mode, mean duration of gaze was 1.8 seconds for "at partner" gaze, and 2.1 seconds for "away" gaze, with an average frequency of 14 "at partner" glances per minute. For the "while listening" mode, mean duration of gaze was 2.5 seconds for "at partner" gaze and 1.6 seconds for "away" gaze, with an average frequency of 17 "at partner" glances per minute. For "at partner" gaze, the avatar's eyes focused directly ahead. The values for vertical and horizontal angles of "away" gaze were chosen randomly from a uniformly distributed range of 0 to 15 degrees.

A more comprehensive user-study than presented by the Eyes Alive project was performed. The experiment was designed to investigate the impact of gaze on the perceived quality of communication by comparing the effects of random eye-movement and the gaze model. Pairs of participants were asked to conduct a role-playing conversational task over a non-immersive video-tunnel link on which the avatar's head was displayed with gaze based on the model and random movement. It was found that having an avatar whose gaze behaviour was directly related to the conversation consistently and significantly improved the quality of communication compared to random gaze. This is consistent with the Eyes Alive study, and also supports the assertion introduced in [28] that, in order for avatars to meaningfully contribute to communication, their animation needs to reflect some aspect of the interaction that is taking place.

Vinayagamoorthy, Garau, Steed and Slater [20] adapted the Eyes Alive model to include the timing data found in Garau, Slater, Bee, and Sasse's model [27]. Consequently, theoretical information from social psychology studies was augmented with the spatiotemporal eye trajectories derived from eye-tracking data. Extending the user-studies associated with the prior two models, experiments were performed using an immersive CAVE™ setting with full-body representation of the participants. Again, the *talking* and *listening* states were divided into the sub-states of "at partner" and "away". These were determined by head-orientation as derived from the head-tracker. Therefore, the virtual character presented in this study should be classified as a semi-autonomous avatar or cyborg. In fact, this is true for all virtual characters partially controlled by behaviour models.

Participants were represented by low or high visual-realism avatars exhibiting random or model-based gaze behaviours and animated using position and orientation of head and hand-trackers. Similarly to Garau, Slater, Bee, and Sasse, the impact of the gaze model on perceived quality of communication was investigated, but the varying visual fidelity of the avatar also allowed insight into the level of correlation between social presence and behavioural realism and consequently, social-influence of the avatar for communication.

Results were consistent with the previous findings, and the gaze model outperformed the random gaze behaviour overall. However, regarding fidelity, in the case of the visually low-fidelity avatar, the more realistic gaze model behaviour did not improve the perceived quality of communication, whereas for the higher-fidelity avatar representation, the gaze model increased effectiveness. This supports the theory of consistency between behavioural and visual realism as more recently addressed in [11]. It was also noted that experimental participants stood facing each other and maintained personal space as also observed by Yee [10] in online virtual worlds such as Second Life (Linden Research).

3.2. Agent attention models

In addition to the gaze models that have been developed for avatars, simulation work that has been developed for agents illustrate further ways of modelling gaze. Level of engagement is largely controlled by amount of gaze and gaze models have been implemented as modules in broader simulations of human attention. Gu and Badler [23] aimed to provide agents with human-like responses to environmental stimuli by modelling aspects of human vision, memory and attention. The gaze model implements low-level motor control (smooth pursuits and saccades) and high-level gaze patterns which also consider multi-party turn-taking (inspired by Miller, 1999), engagement, cognitive workload and distractions and all their interrelations. The model views cognitive resources of agents as a finite resource. As an agent is assigned more demanding tasks or conversational situations, their mental workload increases and more attention is devoted, consequently increasing the likelihood of missing an unexpected event or environmental distraction.

Peters [29] introduced a model with agents capable of visually perceiving another's interest level based on gaze, head and body direction and locomotion. Subsequently, the observing agent makes the decision to continue speaking or take another action. This model also takes into account occurrences in the external environment for both the speaker and the listener and based upon gaze direction, to adjust engagement levels accordingly.

Although both presented in rather abstract and theoretical manners, these models highlight some types of complexities of human behaviour that are beginning to be tackled by avatar/agent models towards realistic human-like behaviour. However, as we will discuss in the following section, there are several conflicting interests and goals running between the presented models.

3.3. Limitation of current gaze models

Gaze models have been shown to be capable of producing simulated gaze that achieves certain levels of perceived realism and task-relevant interactional cues. Associated user-studies investigated whether gaze models are more effective than static or random eye-movement. In all such experiments covered and also as seen in Deng, Lewis and Neumann [25], Colburn, Cohen and Drucker [30], and Fukayama, Ohno, Mukawa, Sawaki, and Hagita [31], the models are judged to be more authentic and natural than random or stationary gaze.

However, there are a number of limitations to these models. The empirical or measured data on which these models rest has overwhelmingly come from dyadic 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. While some models do exist to support richer, multiparty interaction between agents, it remains to be seen if they can be adapted for use in real-time participant-controlled avatars in ICVEs for use in telecommunication.

A general issue which may be perceived as a conflict for the support of using simulated gaze in ICVEs in a telecommunication context is that aspects of non-verbal communication are likely to be highly contextual across different moments in interaction and also to vary across participants. Therefore, generalised models, while they have been shown to be superior to random gaze (and indeed can successfully indicate current mood or cognitive load), will fail to support the communicational nuances that gaze allows.

We consider *simulation* versus *reproduction* (i.e. *models* versus *tracking*) of avatar gaze as an issue of behavioural fidelity. When defining the fidelity of an avatar's gaze behaviour, a distinction can be made between what we term *attentional* and *communicational* gaze [13] These are informed by Vertegaal's requirements for video and virtual conferencing systems [32]. *Attentional* gaze is able to infer only focus of attention (i.e. a head-orientation metaphor), and does not support true awareness of other's gaze thus not allowing mutual gaze to be established. This is due to for instance, a fragmented workspace or narrow field of view as seen in non-immersive CVEs and general many videoconferencing systems (excepting the MAJIC [33] and similar systems). The higher fidelity *communicational* gaze locates the attentional properties in a perceptually unified space and preserves participant gaze, thus supporting fuller gaze awareness and mutual gaze.

So, it is the nature of the display system combined with the fidelity of the conserved human gaze that defines a system's capacity for using gaze as a communicational resource during mediated interaction. We classify avatar gaze-models (no matter the display system) as supporting attentional gaze, because even though an avatar might be perceived by other participants to be producing realistic gaze behaviour, these models do not preserve the actual gaze behaviour of the signaller. In contrast, a tracked-gaze system that reproduces actual gaze as presented by *EyeCVE* would be classified as supporting communicative gaze, because the participants are able to use their gaze as a communicational resource in a spatially-consistent shared environment. However, it is ultimately the intended application scenario that will determine the required fidelity of gaze-preservation, which may or may not be considered critical.

A general issue for all these models is, of course, that they infer gaze behaviour from other interactional events. Therefore, a particular problem for such models is posed by interactional practices that are accomplished through gaze, rather than other resources, such as talk and gestures. In the following section we review research into such practices. We draw on findings emerging from conversation analysis, which aims to provide detailed technical analyses of the resources out of which participants build social interaction.

4. Gaze practices in social interaction

Underlying the idea of better modelling for avatars and agents is the idea of making their behaviour life-like and 'realistic'; consequently, simulation requires knowledge of human practices such that they can be modelled. In this section, drawing on conversation analytical research, we show that in social interaction humans sometimes use gaze practices which are consequential for the interactions in which they occur yet which cannot be predicted and/or inferred from talk. Such practices are important because they cannot be predicted and/or inferred from interactional states (such as speaking and listening), which many models are based upon. There are several practices accomplished with the deployment of gaze that pose challenges to simulation that we now address in turn.

4.1. Speaker selection in multi-party interaction

One important use of gaze that poses serious difficulties for simulation is its deployment for speaker selection. In a multiparty setting (i.e. three or more persons) a speaker may sometime address all other persons but sometimes they 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 address term, e.g. the addressee's name. However, gaze appears to be an important and widely used practice. Commonly, a speaker will gaze at the party whom they are addressing (in particular bringing their gaze to them as their turn reaches its end). Often their talk will include the pronoun "you" - showing that someone is being addressed - but whom that someone is shown by their gaze [34].

The fragment presented below (Figure 1) shows an example of this phenomenon. It shows how there is no vocal clue about who is being addressed by 'you' and that participants' monitor one another's gaze to check who is being *shown* to be selected as next speaker by gaze. Four friends are having dinner when one participant – Vivian – asks a question (line 12). We can note (lines 13-15) that the participants are not clear about who is being addressed by the verbal aspect of the question, so Nancy and Michael look up and at Vivian (see annotation Nv and Mv on line 13 – note that Vs means Vivian gazes at Shane, Nv means Nancy gazes at Vivian, and Mv means Michael gazes at Vivian) during the production of her talk, to check who is being addressed. As she gazes at Shane during the production of the question both Nancy and Michael keep on eating, and only when Shane looks up (line 15) and finishes chewing his food, he produces a response at line 16.

In this case, 'you' shows that the question is being addressed to a single participant, thus selecting that participant to speak next. However, all of Vivian's coparticipants could take it that they were possibly being addressed by 'you'. However, when Nancy and Michael look up (at a point when a response may soon be due), they can see from Vivian's gaze direction that the question was visibly directed to Shane, and neither responds in the 1.2 seconds before Shane speaks at line 16.

```
[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 3 Extract from Lerner (2003, p.183)

This human social practice has two important upshots for the work on telecommunication in virtual digital environments. Firstly: this shows the importance of gaze - and the need to capture it and represent such meaningful form of communication for enhanced presence. Secondly: this shows an important limitation of simulation. Given that there is nothing in the speaker's talk that a co-participant can use to see who is being addressed, they have to monitor the speaker's gaze. Consequently, there is also nothing in the speaker's talk that a simulator could use to know where the speaker is gazing.

4.2. Promoting sequence expansion

Gaze may be used in a particular sequential position, namely following a response (such as an answer) to an initiating action (such as a question) in order to pursue further talk. For example, Rossano [35] has shown that following an answer to a question, a questioner may gaze at the answerer showing that they do not take the answer that has been provided to be a sufficient answer for the question-answer 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 participants can show themselves not to be treating a sequence as being over, so the sequence is often extended as their gaze behaviour pursues further talk from an interlocutor. Moreover participants 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 gaze, showing their attention to whether possible sequence completion has been treated as actual completion or not (see Schegloff, [36], p. 116, on Rossano's unpublished work). This careful monitoring of participants' gaze in interaction is only understood in terms of talk as the place in which a sequence reaches possible completion, but what is treated as *actual* completion by participants (that is a satisfactory second pair part to a base first pair part) is only decided locally by participants and is accomplished by gaze.

4.3. Characterising a suspension of talk

The examples presented above show how participants can engage in 'meaningful' gaze practices that cannot be understood only by elements of the talk they are associated 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 in a conversation. The practices examined in this section involve interruptions to talk with different gaze practices that are deployed to achieve very different actions.

```
          Caregiver looks up at children (0.7)
                |   looks down at book
                |         |  looks up at children
                |         |         |
             .. X______ , x        .X
1 CG:  I wish I could ho[p like that
2 H :                   [Aihhhhhhhhhhhhhhhh!
```

Figure 2 Extract from Kidwell (2005, p. 426)

Kidwell [37] shows how very young children understand the import of different looks produced by their care givers. She distinguishes between 'mere looks' and 'the look' which can be used for sanctioning a child. She shows that an interruption to talk associated with fixed and relatively long gaze at a child produces sanctioning of some kind of bad behaviour/activity and is often successful in stopping the sanctioned activity [37]. For example, in Figure 2, a caregiver looks up as while she reads from a story book. During this 'mere look', one child pulls the shirt of the other child, Heather, who cries out as the caregiver's gaze returns to the book. 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 whether they are being looked at and *how* they are looked at.

Not every interruption to talk is, however, a sanctioning activity. Goodwin and Goodwin [38] showed that when engaging in word searches, and interrupting the progressivity of their talk, participants look away from the recipients of the talk. Such interruption of the talk can be understood in those cases as being performed as the course of activity of finding the right term, name or word to, then, progress with the activity that was in course. Other interruptions and re-starts while the speaker looks at their interlocutor can be used when speakers seek to attract gaze toward them [35] It can be seen from those cases that similar vocal practices, such as silences and the interruption of talk, can be associated to different actions. As Kidwell [33] puts it "the import of a gaze is a locally contingent, interactionally achieved matter." (p. 419). Another relevant import from the examples shown earlier is the fact that some activities are, in effect, implemented with the use of gaze rather than simply relying on a participant's talk (e.g.: speaker selection, to pursue an action, to sanction a bad behaviour). Another example of an activity which cannot be inferred by the talk of the participants and is achieved by non-vocal resources can be seen in interactions that involve objects.

4.4. Projecting involvement with a different activity

As shown by Rae (2001) [40], participants can project an involvement with some other activity and/or object (e.g. a telephone and the action of ringing someone) via non-vocal resources. The extract presented below (Figure 3) makes evident that, although a participant does not make any reference to phoning someone, he has non-vocal resources for implying 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 3 Extract from Rae (2001, pp. 263-264)

Two particular body movements are especially relevant here. First, well before the speaker picks up the receiver, he turns his gaze toward the phone. This redirection of gaze involves a head turn of about 90 degrees and a subsequent swivelling of his chair toward the phone. These non-vocal actions, together with his talk, clearly project involvement in phoning. As Rae (2001) [40] suggests, 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. 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 and it can be seen, however, that the specific use of head orientation and gaze direction in connection with the talk of the participant are meaningful and interactionally relevant here.

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 to understand that a head turn towards the phone and a subsequent repositioning of his seating arrangement towards the phone mean 'looking at the phone' and project the making of a phone call. Participants, and analysts, can and often do infer what/who is being looked at by their interlocutor's head orientation. It should be pointed out that research in human interaction using conversation analysis, the methodological approach to data analysis considered here, focuses on naturally occurring, spontaneous social interaction and uses video-recording to capture it. As such (with limited exceptions, e.g. Steptoe, Wolff, Murgia, Guimaraes, Rae, Sharkey, Roberts and Steed [19] it does not use eye-trackers in the analysis of gaze in social interaction, and consequently the focus of participants' gaze sometimes must be inferred from head-orientation.

The human practices associated with gaze reviewed here have shown the importance of gaze and head orientation in human interaction and, consequently, the need to capture them and represent them in telecommunication systems. Gaze is "part of the interactional machinery by which participants sustain and regulate their conjoined activities" [37, p. 420]. Consequently, there are profound problems in attempting to simulate the use of gaze when it, rather than talk, is used to accomplish certain interactional practices (e.g.: speaker selection, to pursue an action, to sanction a bad behaviour, to project an involvement with a different activity).

5. Discussion

5.1. The Use of ICVEs as interface systems for telecommunication

The representation of participants' gaze behaviour poses particular challenges in using ICVEs for telecommunication. Whilst certain aspects of human behaviour, (e.g. talk) can easily be captured, transmitted and replicated, gaze is technically challenging, particularly in CAVE-like displays [17, 18. 19]. Clearly, in case of technical failures and other shortcomings there are practical benefits to at least some degree of avatar's eye gaze behaviour being simulated. Simulation at each local site might appear to be an attractive strategy as it would circumvent the need for the capture and transmission of gaze information. In addition, when remote participants' gaze has been tracked, transmitted and represented, smoothing or filtering have been needed [17] due to inherent limitations with data input (i.e. the constraints of the eye tracking sensors) and losses in data transmission. Nevertheless, with the exception perhaps of certain degrees of low-level filtering, simulation requires knowledge of human gaze practices.

Gaze models for simulating gaze seek to display gaze behaviour based on other things that the participant is doing (e.g. their state of talk). In this way, interactional states are used to determine gaze behaviour. In the case of statistical models, this is through a statistical model of behavioural norms. However, based on conversational analytic research, we have shown that there are certain classes of gaze behaviour where interactional work is done by gazing itself. Our review identified four cases: selecting next speaker in multiparty interactions, sanctioning, showing that a sequence has not reached completion, and proposing a course of action. These behaviours cannot be simulated from co-occurring behaviours.

5.2. Limitations of the data presented

Previously, we noted the distinction between head orientation and gaze (or head-gaze and gaze); when a person's gaze displays what they are doing it is possible that head orientation may carry the relevant information. Indeed, we noted that sometimes when carrying out interactional analysis, as a result of limitations due to the setting of video recording equipment, participants gaze is, in fact, inferred from head orientation. It is evident however that this is not always the case that gaze and that the availability of co-participant's gaze is sometimes interactionally relevant.

Whilst it appears to be the speaker's gaze that is the heart of this issue, there is still the possibility that head orientation might be enough of a clue for a participant to infer a co-participant's gaze direction. However this seems unlikely: as documented by Lerner [34], although gaze is an explicit form of addressing, the success of this practice depends upon the separate gazing practices of co-participants. So, the addressed participant has to recognize that she/he is being addressed and other participants have to understand they are *not* being addressed. Lerner documented the practice of *gazing around* a participant who is an unintended but possible (and proximate) addressee, as one of the evidences for such matter. So, sometimes a speaker addressing someone beyond a proximate participant can produce a gazing pose that markedly produces *gazing around* this proximate participant (when they are gazing at the speaker), even when this participant is not obscuring their view. This shows that gaze placement is carefully used and monitored by participants. Moreover, in ICVEs, gaze has been shown to be of vital importance for the correct identification of what a participant is looking at [24].

5.3. Speaker selection through gaze in multiparty interaction

The case of speaker selection is particularly interesting because this has received attention within the VMC literature [32] and within the conversation analysis literature [34]. In co-located interaction, the success of speaker selection through gaze depends upon careful attention by participants as to where their co-participants' gaze is. As such, it seems likely that mutual gaze is relevant at this juncture because it enables a speaker who is selecting a next-speaker through gaze to see that this party has seen their gaze (it also puts that 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 we are 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 co-present 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 identifying who has been selected. So, it is not enough to know when gaze is directed at oneself, it is also crucial for participants to see who else the speaker is gazing at.

Whilst speaker selection is clearly important, it should be remembered though that there are other reasons for mediating gaze accurately, e.g. so that participants can see what others are gazing at. Sometimes addressing is done not be gazing at the addressee but by gazing a certain objects. (E.g. in when saying something like "Can you pass that" in a co-located multi-party setting where gaze shows what "that" refers to and "you" can then be inferred by proximity to that object. Incidentally, in such cases – but not all – it seems very plausible that hand gesture would be implicated).

5.4. Avatar Realism and Fidelity to Remote Participant

One aim of work on avatar development is the achievement of realism such that participants feel that they are co-present with the persons represented by the avatars. A part of this project has been the emulation, or simulation of human behaviour. The potential to model human gaze presents itself as a solution to various challenges that arise in using ICVEs as interfaces in telecommunication systems. However, work on avatar realism is partially orthogonal to accurate mediation. A particular problem is that inferred gaze may be rated as "realistic" yet fail to replicate crucial gaze behaviour that a participant engages in. That is, a gaze model could receive high levels of ratings for perceived authenticity yet not actually be faithful to the behaviour of the remote participant. 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 telecommunications, there is actually a human participant whose actual gaze behaviour could be quite different from that which a model of them would display. Indeed, in some contexts ethical issues might arise if a participant does not know how their behaviour is being represented.

It should be noted that if an avatar that has modelled behaviours it is no longer an actual avatar (in the sense of a representation of a human): it is part avatar and part agent. Such hybridization is perhaps evident in Vinayagamoorthy, Garau, Steed and Slater's [20] closing 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." (emphasis supplied)

Although we have emphasised here some problems of using models to represent an actual human in virtual environments, we are aware of the fact that the studies on modelled agents and avatars have their value as empirical demonstrations of how certain aspects of avatar behaviours are perceived. We are also aware of the potential of *virtual characters* (in the sense we noted above) in educational and entertainment applications to systems where participants either choose, or have allocated to them, interactional styles that differ from their actual behaviour.

6. Future research

The issues reviewed point to the need for future work on the representation of avatar gaze for telecommunication to proceed along five lines.

- Further empirical interactional analysis of co-located (or co-present) human interaction will enable us to develop a better understanding of the practices that occur in social interaction.There are two particular matters to address here. Firstly, research is needed in order to examine the structures and practices that naturally occur. For this, we recommend using conversation analysis, which is geared to examining such phenomena in interaction. As mentioned previously, conversation analysis uses video-recordings of social interaction, these may fail to show participants' gaze in the detail that is necessary to distinguish gaze from head-orientation. Indeed participants may use head orientation and ocular orientation for specific interaction purposes as is suggested by the familiar phenomenon of looking "out of the corner" of one's eye. Further research in this area could draw on suitably positioned high definition cameras and eye-tracking equipment. Secondly, in order to establish the distribution of the phenomena that we have drawn attention to, such an analysis would require content analysis. Thirdly, further quantitative analysis of eye-tracked data is necessary.

- Further empirical evaluation of how social interaction occurs when it is mediated through different systems is needed. In particular, we need to know more about the problems, if any, that may occur through gaze behaviour not being supported either because of technical limitations or through the use of models that do not represent certain behaviours. In addition, we need to understand the resources that participants are able to develop to address such problems if they occur.

- In addition to examining behaviours in different environments, further empirical analysis is needed of participants' experience, in particular the level of social presence that they feel.

- It is necessary to develop more advanced gaze models. Whether or not it is desirable or necessary for telecommunication systems, there is scope for the development of enhanced gaze models that go beyond the use of states of talk. Such research is perhaps more relevant for the development of autonomous agents than for telecommunication. Models covered from Gu and Badler [23] and point the way forward for gaze models. The latter proposed a gaze model with associated parameters to enable avatars to convey different impressions to users, rather than simply low-level eye-movement properties. This is a step towards implementing the assertion that the management of one's own impression to influence the behaviours of others plays a pivotal role in human communication, and is an essential function to enable avatars to adapt to the multiplicity of social communication scenarios.

- Finally, there a need for more work on developing ways of capturing, transmitting and representing actual gaze behaviour. Unless the interactional phenomena that we have focused on here are generally rare or any problems engendered by them not being supporting are readily fixed by participants, more work is needed on capturing, transmitting and representing participants' gaze behaviour. Each of these three phases requires development. In particular, there are ergonomic issues concerning the usability and comfort of data-capture equipment and there are aesthetic issues concerning the appearance of avatars: in addition to general avatar design issues, there is the need for unobtrusive gaze capturing devices that are easy to use.

Conclusions

ICVEs offer a medium for enabling telecommunication in which, in addition to real-time audio communication, remote participants can see each other's nonverbal behaviour in a shared space. There are a number of reasons why the simulation of eye-gaze is an attractive strategy. Such a strategy aims to enable participants to interact with realistic avatars without the need for the problems for the difficulties in using eye-trackers to capture gaze behaviour. We have shown, however, that there are some moments in interaction when it is not possible to infer gaze behaviour from states of talk. Further empirical work is needed to establish the payoffs between the costs of mediating participants' gaze on the one hand and not doing so on the other.

Acknowledgements

This work is supported by the UK EPSRC award EP/E007570/1 Eye Catching: Supporting tele-communicational eye-gaze in Collaborative Virtual Environments (Dr John Rae and Dr Paul Dickerson, School of Human and Life Sciences, Roehampton University, London), Project Partners SGI, VISUAL ACUITY LIMITED, Electrosonic Ltd and Avanti Communications Limited. This is a collaborative project also supported by the following EPSRC awards EP/E007406/1 (Professor David Roberts and Dr Norman Murray, Informatics Research Institute, University of Salford); EP/E010032/1 (Dr A Steed, Computer Science, University College London) and EP/E008380/1 (Professor Paul Sharkey, School of Systems Engineering, University of Reading) and involving input from Dr Robin Wolff (Salford); Dr Alessio Murgia (Reading); Dr Vinoba Vinayagamoorthy and Dr Wole Oyekoya (UCL). Particular thanks to Anthony Steed for critical comments and suggestions on the manuscript. We are also grateful to comments from three anonymous reviewers.

References

[1] E. Isaacs, J. Tang. What video can and cannot do for collaboration. *Multimedia Systems,* 2, 63–73. 1994.

[2] J. Hauber, H. Regenbrecht, M. Billinghurst, A. Cockburn. Spatiality in videoconferencing. In: Proceedings of the 2006 20th anniversary conference on Computer Supported Cooperative Work, 413–422. November. 2006.

[3] D. Roberts, R. Wolff, O. Otto, A. Steed. Constructing a Gazebo: supporting teamwork in virtual reality. *Presence: Teleoperators and Virtual Environments*, 12, 644–657. 2003.

[4] M. Slater, J. Howell, A. Steed, D. Pertaub, M. Garau. Acting in virtual reality. In: *Proceedings of the third international conference on Collaborative virtual environments*, 103–110. September. 2000.

[5] R. Schroeder. Social Interaction in Virtual Environments: Key Issues, Common Themes, and a Framework for Research. In: R. Schroeder (Ed.) *The Social Life of Avatars: Presence and Interaction in Shared Virtual Environments*. New York: Springer-Verlag. pp. 1-18. 2002.

[6] S. Benford, J. Bowers, L. Fahl´en, C. Greenhalgh, D. Snowdon. Embodiments, avatars, clones and agents for multi-user, multi-sensory virtual worlds. *Multimedia Systems,* 5, 93–104. 1997.

[7] J. N. Bailenson, J. Blascovich. Avatars. In: W. S. Bainbridge (Ed.) *Encyclopedia of Human-Computer Interaction*. Great Barrington, MA: Berkshire Publishing Group. pp. 64-68. 2004.

[8] V. Vinayagamoorthy, A. Steed, M. Slater. Building Characters: Lessons Drawn from Virtual Environments. In: *Proceedings of Toward Social Mechanisms of Android Science: A CogSci 2005 Workshop*, 119–126. July. 2005.

[9] J. Bailenson, J. Blascovich, A. Beall, J. Loomis. Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators and Virtual Environments,* 10, 583–598. 2001.

[10] N. Yee, J. Bailenson. The Proteus Effect: Self Transformations in Virtual Reality. *Human Communication Research*, 33, 271-290. 2007.

[11] M. Garau. Selective fidelity: Investigating priorities for the creation of expressive avatars. In R. Schroeder, A.S. Axelsson (Eds.) *Avatars at work and play: Collaboration and interaction in shared virtual environments.* New York: Springer-Verlag. pp. 17–38. 2006.

[12] M. Garau. The Impact of Avatar Fidelity on Social Interaction in Virtual Environments. PhD thesis, University College London, 2003.

[13] J. Blascovich, J. Loomis, A. Beall, K. Swinth, C. Hoyt, J. Bailenson. Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry,* 13, 103–124. 2002.

[14] W. Steptoe, A. Steed. High-Fidelity Avatar Eye-Representation. *IEEE Virtual Reality Conference*, 111-114. March. 2008.

[15] S. Acker, S. Levitt. Designing Videoconference Facilities for Improved Eye Contact. *Journal of Broadcasting and Electronic Media,* 31, 181-91. 1987.

[16] M. Argyle, M. Cook. *Gaze and Mutual Gaze.* Cambridge: Cambridge University Press. 1976.

[17] R. Wolff, D. Roberts, A. Murgia, N. Murray, J. Rae, W. Steptoe, A. Steed, P.Sharkey. Communicating Eye Gaze across a Distance without Rooting Participants to the Spot. In

Proceedings of the 12th IEEE International Symposium on Distributed Simulation and Real Time Applications (DS-RT) [in press].2008.

[18] A. Murgia, R. Wolff, W. Steptoe, P. Sharkey, D. Roberts, E. Guimaraes, A. Steed, J.Rae A Tool For Replay And Analysis of Gaze-Enhanced Multiparty Sessions Captured in Immersive Collaborative Environments. In: *Proceedings of the 12th IEEE International Symposium on Distributed Simulation and Real Time Applications (DS-RT)* .[in press]. 2008.

[19] W. Steptoe, R. Wolff, A. Murgia, E. Guimaraes, J. Rae, P. Sharkey, D. Roberts, A. Steed . Eye-Tracking for Avatar Eye-Gaze and Interactional Analysis in Immersive Collaborative Virtual Environments. Paper presented to *CSCW 2008: Computer Supported Cooperative Work*. November. 2008.

[20] V. Vinayagamoorthy, M. Garau, A. Steed, M. Slater. An Eye Gaze Model for Dyadic Interaction in an Immersive Virtual Environment: Practice and Experience. *Computer Graphics Forum*, 23, 1–11. 2004.

[21] R. Schroeder, A. Steed, A.S. Axelsson, I. Heldal, A.Abelin, J. Widestrom, A. Nilsson, M. Slater. Collaborating in networked immersive spaces: as good as being there together? *Computers & Graphics Journal*, 25, 781-788. 2001.

[22] J. Leigh, T. DeFanti, A. Johnson, M. Brown, D. Sandin, D. Global tele-immersion: Better than being there. In *Proceedings of the Seventh International Conference on Artificial Reality and Tele-existence*, 10-17. December. 1997.

[23] E. Gu, N. Badler. Visual Attention and Eye Gaze During Multiparty Conversations with Distractions. In: *Proceedings of the International Conference on IVA 6*, 193–204. August. 2006.

[24] N. Murray, D. Roberts. Comparison of head gaze and head and eye gaze within an immersive environment. In: *Proceedings of the the 10th IEEE International Symposium on Distributed Simulation and Real Time Applications*, 70-76. October. 2006.

[25] Z. Deng, J. P. Lewis, U. Neumann. Automated Eye Motion Using Texture Synthesis. *IEEE Computer Graphics and Applications*, 25, 24-30. 2005.

[26] S. Lee, J. Badler, N. Badler. Eyes alive. *ACM Transactions on Graphics*, 21, 637–644, 2002.

[27] M. Garau, M. Slater, S. Bee, 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*, 309-316. March-April. 2001.

[28] H. H. Vilhjalmsson, J. Cassell. BodyChat: autonomous communicative behaviors in avatars. In: *Proceedings of the second international conference on Autonomous agents*, 269-276. July. 1998.

[29] C. Peters. Direction of attention perception for conversation initiation in virtual environments. In: *Proceedings of the International Working Conference on Intelligent Virtual Agents*, 215-228. September. 2005.

[30] A. Colburn, M. F. Cohen, S. Drucker. The Role of Eye Gaze in Avatar Mediated Conversational Interfaces. Microsoft Research Report. MSR-TR-2000-81. 2000.

[31] A. Fukayama, T Ohno, N. Mukawa, M. Sawaki, 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*, 41-48. July. 2002.

[32] R. Vertegaal. The GAZE groupware system. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, 294–301. May. 1999.

[33] K. Okada, F. Maeda, Y. Ichikawaa, Y. Matsushita. Multiparty videoconferencing at virtual social distance MAJIC design. In: *Proceedings of the ACM conference on CSCW*, 385–393. October. 1994.

[34] G. H. Lerner. Selecting next speaker: The context-sensitive operation of a context-free organization. *Language in Society*, 32, 177-201. 2003.

[35] F. Rossano. When it's over is it really over? On the effects of sustained gaze vs. gaze withdrawal at sequence possible completion. Paper presented to the *International Pragmatics Association*, Riva del Garda, Italy, 2005.

[36] E. A. Schegloff. *Sequence Organization in Interaction: A Primer in Conversation Analyis*, *1*. Cambridge: Cambridge University Press. 2007.

[37] 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, 417-449. 2005.

[38] M. H. Goodwin, C. Goodwin. Gesture and Coparticipation in the Activity of Searching for a Word. *Semiotica*, 62, 51-75. 1986.

[39] C. Goodwin. Notes on Story Structure and the Organization of Participation. In: M. Atkinson, J. Heritage (Eds.) *Structures of Social Action*. Cambridge: Cambridge University Press. pp. 225-246. 1984.

[40] J. Rae. Organizing Participation in Interaction: Doing Participation Framework. *Research on Language and Social Interaction*, 34, 253–278. 2001.`