Real faces and robot faces: The effects of representation on computer-mediated communication

Emma L. Clayes\textsuperscript{a,}, *, Anne H. Anderson\textsuperscript{b}

\textsuperscript{a}University of the Highlands and Islands Millennium Institute, Perth College, Scotland, UK
\textsuperscript{b}College of Art & Design, Architecture, Engineering and Physical Sciences, Dundee University, Scotland, UK

Received 9 June 2004; received in revised form 10 October 2006; accepted 10 October 2006
Communicated by E. Churchill
Available online 10 January 2007

Abstract

While there is much research regarding audio, video and text based communication, there has been little work concerning how users communicate via avatars—that is graphical embodiments of remote users. The aim of this study was to explore the effects of different forms of representation, by examining how users communicate via high quality video images and basic graphical representations in different communicative contexts. Communication analysis revealed that video images facilitate turn-taking, although they are not necessarily perceived very differently from basic avatars in terms of questionnaire responses. Using eye-tracking techniques, we found that while participants generally gaze more often at video images, this is dependent on the communicative context and is not necessarily an advantage in a problem-solving situation. This study has demonstrated the value of employing various measures and tasks in order to evaluate computer-mediated interactions. The results have implications for the use of video and graphical representations in computer mediated communication and suggest that the benefits of video must be considered in relation to the user’s requirements (e.g. communication process versus outcome) and context in which the system is to be applied (e.g. problem-solving or social).

Keywords: Video-mediated communication; Avatars; Eye-tracking

1. Introduction

Despite recent technological advances in the use of graphical representations, there have been few attempts to evaluate these newer forms of representation in comparison to traditional video-conferencing links for example. This study will attempt to provide a detailed evaluation of video and graphical representations by employing eye-tracking techniques alongside measures of communication and questionnaires. In this way, we can investigate the effects of representation on attention and communication, and provide an opportunity for participants to record their subjective experiences. In addition, unlike many studies in this area, a number of different experimental tasks will be included in order to provide a variety of different communicative contexts.

1.1. Representations of remote participants

Although there have been many studies comparing face-to-face communication and video mediated communication (see Finn et al., 1997) there have been very few studies comparing different types of representations displayed on screen. Sellen (1995) compared 3 different types of video conferencing systems with audio and face-to-face communication. Measures of communication revealed differences between video mediated and face-to-face communication, regardless of the type of video representation. The results suggest that that the type of visual information available is not important but rather whether the conversation is mediated by technology. However, questionnaires and interview data did reveal differences in preference for the three video-conferencing systems and this would seem to indicate that the type of visual representation is important, at least in terms of subjective reports.
Parise et al. (1996) employed four different types of representations to investigate how co-operation with a computer agent is affected by the agent’s pictorial realism, human-likeness and likeability. Participants played a social dilemma game with a talking computer agent that resembled a person, a dog, a cartoon dog, or with a real human confederate interacting through a video link. They found that participants co-operated highly with the human computer agent and the confederate. This suggests that the type of representation presented on screen does affect co-operation as both the ‘human’ computer agent and the dog agents had the same capabilities in terms of simulated speech. Interestingly, the human computer agent was rated similarly to a real confederate interacting through a video link and this has implications for the use of video in mediated communication.

Lantz (2001) compared small groups communicating face to face, using a chat medium and a collaborative virtual environment (CVE). The chat environment was accessed via one of the participants’ home page on the Internet. This enabled conversation via written text. The CVE meetings took place on an existing environment on Active Worlds™. This had a chat-like facility (written text) and a 3D-environment with avatars representing the participants. The avatars were basic graphics that were often referred to by the users as ‘cones’. Questionnaires were used to measure subjective ratings concerning efficiency, communication process, problems with the technology, enjoyment and competence development. CVE meetings were rated as more efficient overall than face to face meetings and both chat and CVE conditions were rated as more efficient for task-oriented work.

An interesting finding from this study was that the avatars seemed to facilitate turn taking as the participants found it easier to begin meetings and communicate compared to the chat condition despite the fact that both shared the same text facility. This seems to suggest that the shared environment in the CVE condition and the graphical presence of others on screen provides additional benefits for remote users in terms of turn taking and efficiency for task oriented work. However, the results were based on subjective responses only and compared text-based communication to normal speech in face-to-face communication. The study does suggest that representations do affect mediated communication and the benefit of having graphical representations is most marked for task-oriented work.

The avatars used in the Lantz (2001) study were basic graphical representations that were often referred to by the users as ‘cones.’ Advances in computer graphics have led to more focus on animated avatars and interface agents which can exhibit life-like behaviours such as speech, gaze, blinking and other body movements. The basic assumption underlying these developments is that users will find it easier to communicate with animated interface agents (or with others via animated avatars), as this more closely resembles human behaviour (Takeuchi and Naito, 1995; Elliot et al., 1997; Cassell et al., 2000). However, opponents point out that animated agents may harm communication as the user will be lead to expect more of the computer than is actually possible (Norman, 1994; Walker et al., 1994; Wilson, 1997).

A number of studies have investigated the value of adding gaze direction to graphical representations. Hindmarsh et al. (1998) found problems with humanoid style avatars as participants assumed their vision to resemble human characteristics when in fact their field of view was restricted by the technology. Problems occurred due to fragmented views of embodiments in relation to shared objects, and participants had to compensate with verbal accounts of their actions. They suggest that difficulties in understanding others perspectives could be addressed by providing more explicit representations of actions than those provided by pseudo humanoid embodiments and that support for navigation and co-participation should be oriented to the activities of others rather than the individual’s current actions.

Garau et al. (2001) found favourable reports for informed gaze avatars compared to random gaze avatars. In the informed gaze condition, the eye animations were driven by conversational turn-taking (timings were based on previous research on face-to face dyadic conversations e.g. Argyle and Cook, 1976). In the random gaze condition, the avatar’s head and eye animations were unrelated to conversational flow. By comparing the subjective responses of dyads in four conditions of communication (video, audio and two avatar conditions) they found that participants responded most favourably to video followed closely by the informed gaze avatar condition. The results revealed that responses for the random gaze avatars and audio only condition did not differ. They suggest that an avatar whose gaze behaviour is related to the conversation is more beneficial than one that merely exhibits liveliness. These findings suggest that while random gaze avatars may not improve on video, the informed gaze avatars are rated more favourably for remote communication than audio only. One factor to consider when comparing simple graphical avatars with those that exhibit more lifelike features is the accuracy of simulated behaviour. The informed gaze avatars in the Garau et al. (2001) study were successful as their gaze was informed by the conversation and was not merely generated for cosmetic purposes as in the Hindmarsh et al. study described above. Vertegaal and Ding (2002) suggest that synchronised gaze models are preferable when designing CVEs. However, they also found that task performance is most affected by the synchronisation of gaze and in some situations random gaze models may suffice.

Several studies have attempted to empirically test the usefulness of animated agents with conflicting results. Sproull et al. (1996) found that users responded more favourably to a virtual careers counsellor when presented with textual output only compared to when a face accompanied the text. In contrast to these findings, Koda
and Maes (1996) reported higher likeability ratings for a virtual poker player visualised by a face than for one without a face. In a review of the empirical research on animated agents, Dehn and Mulken (2000) suggest there is no evidence for a general advantage of an interface with an animated agent to one without. In terms of users’ attitudes there have been positive findings and in general users rate a system with an animated agent as more entertaining than one without an animated agent. Dehn and Mulken (2000) propose that studies should conform to the same methodological standards, and take into account other factors such as task and type of agent.

Although human-human interaction with graphical representations of users in a virtual space is certainly more complex than human-computer interaction via interface agents, there is probably more need for this type of technology to support collaborative work and social play. In order to inform the design and development of avatars, studies must attempt to examine the nature of social interactions using different representations. Bowers et al. (1996) used simple 3D box-polygons to represent users in the MASSIVE CVE for an internationally distributed, real-time, multi-party meeting. From analyses of social interaction they found problems with turn-taking and participation. It was also found that participants move the embodiments to face whoever they wish to interact with. This was unnecessary and suggests that the simple graphical shapes do have a social interactional role and are not merely providing a view of the virtual world for the users.

There is much debate therefore, concerning whether development of this type of technology should be concerned with the humanisation of interfaces and the development of anthropomorphic avatars or whether basic representations may be adequate in some cases. It has also been suggested that humanoid agents and avatars may be confusing and lead the user to expect more from the virtual world than is actually possible. The difficulty with this type of research is that studies differ with regard to the experimental approach, including the type of animation used, the kind of comparison made (e.g. video and animation or two different types of animation), the dependent variables and the type of task. Many studies have relied on subjective responses and basic key questions remain unanswered, such as how the form of representations may have an effect on group communication, and how individuals distribute their visual attention to representations on screen.

1.2. Visual information and communication

The absence of certain types of visual information in mediated communication has been used to explain observed differences between face-to-face and computer mediated communication. Daft and Lengel (1986) introduced the notion of ‘media richness’, which refers to the amount of information available to participants in any communication medium. They suggest that communication media differ in ‘richness’ with rich media such as face-to-face communication providing both visual and verbal information. Rich media are thought to be able to support the rapid exchange of multiple types of information (e.g. in face-to-face communication both visual and verbal types of feedback are transmitted immediately). However ‘leaner’ media such as email provide fewer types of information and feedback is much slower.

Rutter (1987) introduced the concept of social cuelessness to describe the subjective differences felt when communicating with different media. He suggests that when we communicate we quickly form an impression of psychological closeness or distance based on the availability of social cues. ‘According to our cuelessness model, the central difference between media is the extent to which they encourage psychological distance’ (Rutter 1987, p. 99). Usually, Rutter suggests, ‘when social cues are denied us... we feel distant psychologically... Cuelessness leads to psychological distance, psychological distance leads to task-oriented and depersonalised content, and task-oriented, depersonalised content leads in turn to a deliberate, unspontaneous style and particular types of outcome [less likelihood of compromise]’ (Rutter 1987, p. 74). Interestingly, Rutter’s theory predicts particular outcomes which can be tested, namely the prediction that ‘cuelessness’ leads to more task-oriented discussion with less likelihood of compromise. It is less clear how one accurately measures the psychological distance between different media although subjective reports from users of these media may provide insights.

CVEs that utilise avatar representations may appear to be richer than text or audio only communication but the visual information they provide is essentially different from video images. For example, the static avatar representations employed in this study do not provide information about non-verbal behaviours. Visual cues are thought to be important in terms of establishing social presence, that is the salience of the participants interacting and the relationship between them. Short et al. (1976) suggest that communication media that provide a wide range of visual cues engender a strong sense of social presence among users. They found that users ranked various media in decreasing order of social presence thus: face-to-face, video links, the telephone, monaural audio and the business letter. They argued that various tasks might require more or less social presence and that with simple tasks involving ‘information transmission’ or problem-solving, which require little social presence, the medium used matters less.

Social presence is a subjective quality and is concerned with the users’ perception of the media and their interlocutors. One would expect users communicating via video to have a stronger sense of social presence in comparison to those communicating via avatars. Social presence is a concept similar to ‘social cuelessness’ and in common with Rutter (1987), Short et al. (1976) predict task differences in relation to the importance of social presence,
and this issue will be explored in the various task situations employed in this study.

A number of studies have confirmed the importance of visual information in communication. Boyle et al. (1994) found that the visibility of one's conversational partner improves information transfer and the management of turn taking. When speakers had access to visual cues, participants expended less effort in terms of dialogue to achieve common ground. The task used in the study was a problem-solving task and the same level of task performance was found in both conditions of communication. The benefit of seeing one's conversational partner was reflected in terms of shorter dialogues. This may be because visual signals appear to compensate for the underspecification of verbal information and are used to infer additional shared information. Turn taking was also managed more efficiently due to availability cues from non-verbal information. The experiment highlights the importance of the visibility of interlocutors in face-to-face communication and how it affects the communication process in terms of shorter, more efficient dialogues. However, it also suggests that the same level of task performance can be achieved in the absence of visual information about one's conversational partners.

Other studies have investigated the effects of different types of information in mediated communication. These have demonstrated that good quality audio is important for remote communication (O'Connell et al., 1993) and indeed should take precedence over video if there are limits on bandwidth (Anderson et al., 1997). Another major theme in mediated communication research is the importance of shared workspaces over views of remote participants (Gaver et al., 1993; Anderson et al., 2000; Fussell et al., 2000). Therefore, there are many different types of visual information to consider including views of remote participants and views of shared data. By focusing on the visual representation of participants in mediated communication, one can attempt to determine what types of visual information are most valuable for remote users. Also, by including different types of task information, one can compare the relative value of representations and task information in different communicative contexts as theorists suggest that the importance of visual behaviours varies with the type of communicative task (Short et al., 1976; Rutter, 1987).

1.3. Tasks and methodology

Although there are many studies investigating computer mediated communication there are often conflicting results found due to the fact that many different factors are manipulated at the same time. Many researchers are now calling for a more unified approach and for studies to employ the same methodological standards and to take into account the various extraneous factors in studies of remote communication systems (Monk et al., 1996; Anderson et al., 1997; Dehn & Mulken, 2000). This study will use three-party groups, as many studies have focused on dyads or single users (especially when evaluating graphical representations). As mentioned previously, studies also differ widely in terms of the type of comparison made and by focusing on different types of visual representation (and using the same audio system) this study will attempt to minimise the various extraneous factors involved in evaluating different types of communication systems. The avatars in this study are static, robotic heads that differ slightly in terms of gender appearance (e.g. the female head is smaller with slighter features). There are a huge variety of graphical representations and many studies have attempted to evaluate the usefulness of animated agents. It is important to minimise the influence of additional factors when comparing the effects of different representations. By selecting a basic, static avatar with no additional features this should provide a more accurate comparison of these very different types of representation without having to account for the behaviours of more animated agents. For example, it would be very difficult to investigate the usefulness of video representations in providing non-verbal information if the avatars were animated, anthropomorphic agents. In addition, a number of studies have observed the use of simple graphical representations (e.g. Bowers et al., 1996; Lantz, 2001) and it will be informative to explore further by employing the various measures used in this study.

By concentrating on the type of visual information available to remote participants, one can utilise techniques such as eye-tracking to obtain a measure of information processing at the interface. While eye-tracking has been used extensively in the field of psychology (see Rayner, 1998), it is only recently that the value of this measure has been exploited in the area of human-computer interaction (Mullin et al., 2001). Eye-tracking provides an objective, reliable measure of attention at the interface and can determine how often different types of information are used on screen, and therefore what types of information are most valuable for remote users. By using this measure alongside more traditional measures such as communication analysis and questionnaires, we will hopefully provide a more detailed picture of mediated communication than previous studies in this area.

It is suggested that the type of communicative context may affect the importance of social behaviours (Short et al., 1976) and may in fact lead to particular outcomes depending on the type of medium used (Rutter, 1987). For example, it may be less important to view one's conversational partner in a problem-solving task while the absence of non-verbal behaviours may inhibit communication and lead to less compromise in a social task. A number of studies have suggested that video links can be useful for social tasks (Isaacs and Tang, 1993; Olson et al., 1995) whereas for problem solving tasks the addition of video does not seem to provide many advantages over audio
Studies concerned with avatars and interface agents have suggested that for social tasks computer agents are rated similarly to video links (Parise et al., 1996) while Lantz (2001) found that CVE business meetings were rated as being more efficient than face to face and text meetings.

The tasks in this study were designed to explore participants’ gaze and communicative behaviour during different contexts. By manipulating the type of information displayed on screen it is possible to explore how video and avatar representations are used in comparison to task artefacts. Therefore, this study will include four experimental tasks including a problem-solving task, business meeting scenario (incorporating organisational status differences) and two social exchange tasks that allow participants to communicate without the demands of a particular task (in order to explore how the different representations are used when no other information is displayed on screen). In this way we can explore mediated group communication in a variety of different situations unlike previous studies, which are often limited to one type of scenario.

Other benefits of employing this design include the fact that participants will have time to become familiar with the technology (the four tasks will take place over two hours) and communicating with others in various task situations. The extended time period in this study also allows participants to become more familiar and comfortable with one another. Experimental situations often involve very limited exposure to the laboratory environment and it is particularly important when evaluating the effects of new technologies that participants are allowed time to become familiar with their surroundings, thereby precluding any novelty effects.

1.4. Hypotheses

This study will address the following issues: Does the extra visual information provided by video affect group communication? For example, one may expect there to be more difficulties in managing turn taking in the avatar condition as previous studies have found that visual signals (e.g. posture, facial expressions, gaze) are important for co-ordinating conversation (Boyle et al., 1994). Another question relates to how often the representations are actually used by the participants in terms of how often they gaze at the representations on screen. Do video images attract more gaze than avatar representations? In addition to analyses of communication and gaze behaviour, questionnaires will be employed as many studies have found subjective differences between representations using this measure (Sellen, 1995; Parise et al., 1996; Lantz, 2001).

The four tasks were designed to explore specific issues concerning computer-mediated communication and are described in more detail under the heading for each task in the method section below. The main hypotheses regarding the different forms of representation are as follows:

H1. Participants will gaze more often at video images of their remote interlocutors in comparison to avatar icons, as live faces provide more information than static, grey faces.

H2. Conversations will be more difficult to manage in the avatar condition as reflected in a higher number of interruptions compared to the video condition. This is expected due to the lack of visual cues for turn taking in the avatar condition.

H3. Participants will rate the video condition more favourably in subjective ratings of the technology. This is predicted as the video condition provides a wider range of visual cues (for example, gaze, facial expressions and gestures) compared to the avatar condition.

2. Method

Groups of three participants collaborated to complete four different tasks using one of two types of remote communication systems. One person in each trio was eye tracked during each task and the data was stored for off-line analysis. The conversations were audio recorded on tape for analysis. All participants used the same high quality audio system and in one condition live video images of the remote participants were presented on screen. In the avatar condition, the video images were replaced with an audio conferencing application. BT Conference Call Presence™ is an advanced phone conferencing application that allows users to communicate by creating a virtual meeting room with icons representing all the attendees on screen. It also allows real-time sharing of documents, spreadsheets and other applications. Many other features such as chair control are included but for this experiment the basic graphic of the meeting room was used. The avatar icons were grey static faces (slightly different for males and females), and the participant’s name was displayed at the top of the avatar icon. Fig. 1 is an example of what participants in the video condition saw during task 2 and Fig. 2 is an example of what participants in the avatar condition saw on screen during task 4 (the task information remained the same in both conditions, while in tasks 3 and 4 only the representations were displayed on screen). The video images were sized to approximate the area of the virtual meeting room (see Section 2.3).

2.1. Design

The study is a 2 × 4 mixed design. The first factor is type of representation (video images/avatar icons) and is between groups. Twelve groups of three took part in the video condition and twelve groups of three were exposed to the avatar condition. The second factor is task and is within groups as all participants completed 4 different tasks in the same order for one session lasting approximately two hours. This was done in order to provide participants with enough time to become familiar with the
systems, and the tasks were run in the same order to enable comparisons across conditions. The first and third tasks are similar in terms of allowing the participants to chat freely without the demands of a particular task so comparisons can be made.

2.2. Participants

Seventy-two participants in total took part in this study. The three participants did not know each other previously to ensure that all participants in each condition began communicating on the same level (i.e. with no prior knowledge or familiarity with the others in the group). The participants were recruited from the student population at Glasgow University and varied in age between eighteen and forty years old. All of the participants had previous experience of using computers but most were unfamiliar with remote communication systems and none had used the Presence audio-conferencing system. The participants were required to have normal vision for eye tracking purposes and be native English speakers. Each participant was paid eight pounds for taking part in the study.

2.3. Set-up (for each task)

The participants were placed in three different rooms and communicated via high quality video and, full duplex analogue audio connections. The audio equipment included two maplin six channel stereo mixing desks, one spirit twelve channel desk and one mackie twelve channel desk. All of the participants wore headsets (pro-luxe headsets with microphones attached). The interactions
were recorded on tape using a denon tape deck. The participants were seated in front of a computer with a 19 inch monitor and the screen resolution was 1024 \times 768 pixels.

The a vatar window was 120 \times 170 mm in size and the video images were both 100 \times 90 mm. In taking into account both video images, the areas of representation were roughly the same (avatar: 120 \times 170 mm, video: 100 \times 180 mm). Although the actual avatar icons were much smaller (15 mm \times 10 mm) than the video images, the amount of gaze directed at the avatar window was used as the dependent measure on gaze analyses. This was done in order to ensure a close comparison between how often participants gazed at the virtual meeting room (including the table which the avatar icons surround) and the two video images. In addition, the actual virtual meeting room provides a context for the avatar icons to be viewed in and is therefore as much a part of the representation as the icons themselves. The representations were always presented in the top left of the screen. In tasks two and four, task information was presented on the right hand side of the screen. This was the same size in both conditions and tasks (165 \times 205 mm).

The avatars displayed on screen were grey robotic heads that displayed the participants’ gender and name. The BT Conference Call Presence system was on loan to the department although this system is also available to download from the Internet for users with a PC and separate Internet and telephone lines. In the video condition, three JVC camcorders were used with videum capture software (version 2.9.2q) and the video images were refreshed at over 24 times per second.

The eye-tracked subject sat approximately 70 cm from a 19 inch computer screen. For the participants being eye tracked, the remote eye-tracking device (RED) was positioned just below the computer monitor and in the video condition the JVC camcorder was positioned on the top of the monitor. For the two participants not being eye tracked in the video condition, the camcorders were positioned to the side of the monitor on a tripod stand. The non-invasive eye-tracker was manufactured by SMI (sensori motor instruments) and the software used was Iview V.31. Calibrations were conducted at the beginning of each session and task.

2.4. Procedure

Participants were taken to separate rooms and given a consent form and a full set of instructions for each task. The experimenter explained that there were instructions to be read before the beginning of each task, and a questionnaire to complete at the end of each task. The experimenter left each participant to read the instructions and complete the consent form before returning to make sure that they were ready for the first task. The experimenter was able to communicate with all participants through the audio link and indicated to them when to begin each task. At the end of each task the experimenter checked on each participant to make sure they had completed the questionnaire and were reading the instructions for the next task. These details remained the same for each of the four tasks.

2.5. Experimental tasks

The tasks were designed to explore participants’ gaze and communicative behaviour during different contexts. In tasks 1 and 3 the only information displayed on screen are the representations of the interlocutors (video images/avatars) and thus there is no focus of attention for the actual task other than the representations. This allows us to compare how the different representations are utilised when there is no other information presented on screen. In task 2 a picture of a map is displayed on screen, which is central to the task and therefore provides total focus for the participants although the representations are also displayed alongside the map. In task 4 a picture of four mobile phones is displayed which is relevant to the task but not essential in order to complete the task. This provides a mixed focus of attention for the participants and enables us to explore how the different representations are utilised in a context when there is other relevant information on screen. While the main aim of this study is to explore gaze and communicative behaviour, the manipulation of tasks according to focus of attention provides an adequate range of communicative contexts in which to explore patterns of communication and the content of each is described in more detail below.

2.5.1. Task 1: no focus (social exchange)

The participants were asked to find out general information about their communicative partners and ‘chat’ to each other for approximately ten minutes. The only information presented on screen was the avatar window or video images representing the other participants. This provides an opportunity to examine how often participants gaze at the virtual meeting room without the inclusion of a task feature on screen. Although participants may gaze at the representations only because there is no other information presented on screen, it is equally true that they may ignore the representations if there is other information to look at on screen.

2.5.2. Task 2: total focus (map task)

The map task (Brown et al., 1983) is a collaborative problem-solving task and originally involved two participants. In this study the three-person version adopted by Anderson et al. (1999) will be employed. In this task two participants (referred to as instruction givers) have to instruct one other person (the instruction follower) on how to draw a route on their map. The instruction follower uses a mouse to draw the route on screen. All participants have the same map although individual landmarks may differ. The eye-tracked subject was always one of the instruction
The experimenter ensured that both the participants and instructions for the roles were very detailed and the four mobile phones, to be sold in their retail outlets. The participant’s task was to persuade the confederates in a somewhat formal manner in order to simulate the setting of a business meeting and ensure the status differences were apparent. The participant was introduced to the confederates in a somewhat formal manner and was asked to play the role of a sales representative and was required to indicate their response be marking a vertical line on a scale of 0–100. The main prediction regarding subjective responses was that the video condition would be rated more favourably compared to the avatar condition on all of the questions and especially the four questions concerned with the social presence of the medium (e.g. how sociable/warm/sensitive/personal did you find this medium of communication?). This was predicted as the video condition provides visual information about the remote participants (i.e. appearance, posture, gaze and other non-verbal behaviours) and this should engender a stronger sense of social presence among users (Short, et al., 1976).

2.5.3. Task 3: no focus (social exchange)

The participants were instructed to discuss their experience of the session so far. The task is comparable to the first task as the participants are merely being asked to chat and in this way we can compare communication and gaze behaviour over time as by the third task the users have had time to become familiar with the different representations on screen. This is important, as many studies allow only brief exposure to the technology and by comparing behaviour over the first and third tasks we can determine whether time does influence behaviour. As in the first task, only the representations were displayed on screen.

2.5.4. Task 4: mixed focus (business meeting)

This task is a simulation of a business meeting as this type of scenario might be more realistic for contexts in which mediated communication systems are used. The task roles were designed to explore the effects of status on multi-mediating groups as previous research in this area has produced mixed findings (e.g. Silver et al., 1994; France et al., 2001).

The business meeting task involves one participant and two confederates. The confederates were role-playing people of different occupational status. One was to act as the manager of Mobilephone warehouse and the other was an employee from Mobilephone warehouse. The confederates were the same people throughout the study and were postgraduates in the department. The participant was asked to play the role of a sales representative and was introduced to the confederates in a somewhat formal manner in order to simulate the setting of a business meeting and ensure the status differences were apparent. The participant’s task was to persuade the confederates that their product should be selected from a shortlist of four mobile phones, to be sold in their retail outlets. The instructions for the roles were very detailed and the experimenter ensured that both the participants and the confederates were prepared for the task. While this task is somewhat similar to the negotiation task employed by Rutter (1987), the main purpose of this task was to incorporate a realistic scenario into the investigation.

The visual information displayed on screen was a picture of four mobile phones and this graphic was the same size as the map used in Task 2 (165 x 205 mm) and was presented on the right of the screen. We can explore how often representations are looked at when there is other information presented on screen, while the information presented is not essential to complete the task as the map was in Task 2. In comparison to the no focus of attention in tasks 1 and 3, and the total focus of attention in task 2, this task enables us to compare how different representations are used when there is a mixed focus of attention. The fact that the task is a simulation of a business meeting provides realism and while the confederates mean that the status manipulation is artificial, it allows us to tentatively explore whether technology may have an equalising effect on interactions of this type. For example, the avatar representations used in this study are static and convey no aspects of non-verbal behaviour that may play an important role in determining communication in status-differentiated meetings.

2.6. Questionnaires

After completing each task all participants were given a short questionnaire to record their subjective impressions of the technology. The participants were asked to rate the medium of communication on twelve questions concerning their perceptions of the technology. The participants were required to indicate their response be marking a vertical line on a scale of 0–100. The main prediction regarding subjective responses was that the video condition would be rated more favourably compared to the avatar condition on all of the questions and especially the four questions concerned with the social presence of the medium (e.g. how sociable/warm/sensitive/personal did you find this medium of communication?). This was predicted as the video condition provides visual information about the remote participants (i.e. appearance, posture, gaze and other non-verbal behaviours) and this should engender a stronger sense of social presence among users (Short, et al., 1976).

3. Analysis of eye-tracking data

For each task the area representing the participants on screen was defined for percentage analysis on the Iview program (the two video images and the avatar window). For tasks 2 and 4 the task feature (the map or the visual graphic of the mobile phones) was also defined. The data sets were checked by defining the screen as one area and obtaining the total percentage of gaze directed on screen including fixations of 20 msec and above (as the horizontal and vertical data points are sampled every 20 msec). The criterion for accepting a data set has been arbitrarily set at 50% (see Mullin et al., 2001). If less than
50% of the participants’ gaze behaviour is captured on screen then the data set is rejected. However, in this study the criteria was lowered to 40% for the social exchange tasks as the participants tended to spend a lot of the time looking away from the screen. This was probably due to the fact that there was little information on screen, only the representations of the other participants.

Although 58% of participants’ data files were discarded in the social exchanges tasks, this is comparable to other studies that have reported losses of even 62% (Schnipke and Todd, 2000). One must also take into account the nature of the tasks (socially demanding, especially the first task which requires participants to converse with complete strangers) and the length of the session (approximately 2 h long, which is obviously tiring considering the number of tasks the participants had to complete in this time).

The data files were then modified by removing any missing data points (when the participant looked away from the screen no data points were recorded) as these will contribute to the overall percentage scores. After the missing data points have been removed the percentage of time on screen should be almost 100% although there is a small amount lost due to blinks and horizontal or vertical data points recorded off screen (e.g. when the participant glanced to the side or just below/above the screen data points were still recorded). The number of files in each condition (from a maximum number of 12 for the first 3 tasks and 24 for the mobile phone task) and average percentage of gaze captured on screen is summarised for each task in Table 1.

4. Results

4.1. Gaze behaviour

For all tasks a 2 × 2 mixed analysis of variance on the amount of gaze directed at defined areas (as a percentage of the time spent looking at the screen) was conducted to examine the effects of relevance of information as a repeated measure (2 levels: representation of the other participants and the rest of the screen/task feature) and representation as a between subjects variable (2 levels: video and avatar representations). Analyses have been conducted on all fixations of 20 msec and above.

4.1.1. Task 1. No focus (social exchange): gaze behaviour

A significant effect of relevance of information was found \( F(1, 8) = 46.04, p < .05 \) as participants in both conditions gazed more often at the representations of the other participants compared to the rest of the screen. There was no effect of representation \( F < 1 \) and no interaction between relevance of information and representation \( F(1, 8) = 1.85, p > .05 \).

It was predicted that video images would attract more gaze as live faces provide more information compared to static, grey icons in the avatar condition (H1). The hypothesis however, is only partially supported. Participants do gaze more often at the representations of other people in the group compared to the rest of the screen, and more often at representations in the video condition (see Table 1). However, there was no statistically significant interaction between relevance of information and representation. Participants gaze at avatar icons representing the other people in the group almost as often as participants gaze at video images. This occurs despite the fact that these icons provide no information in terms of physical appearance, facial expressions and other types of non-verbal behaviour.

4.1.2. Task 2. Total focus (map task): gaze behaviour

An interaction was predicted as it was expected that users would gaze more often at the map compared to the representations of the other participants in both conditions, and more often at the video images compared to the avatar icons. A significant effect of relevance of information was found \( F(1, 16) = 369.22, p < .05 \) as participants in both conditions gazed more often at the map compared to the representations of the other participants. There was no effect of representation \( F(1, 16) = 1.04, p > .05 \) but there was a significant interaction between relevance of information and representation \( F(1, 16) = 83.22, p < .05 \). All of the simple main effects were significant \( ps < .05 \).

The means can be found in Table 1 and these reveal that participants in the avatar condition looked almost exclusively at the map and spent only 2% of their fixations at the

<table>
<thead>
<tr>
<th>Task</th>
<th>( N )</th>
<th>Condition</th>
<th>Avatar/video representations</th>
<th>Rest of screen/task feature</th>
<th>Total % on screen</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:Social exchange</td>
<td>5</td>
<td>Video</td>
<td>82.6</td>
<td>8.4</td>
<td>91.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Avatar</td>
<td>69.2</td>
<td>19.8</td>
<td>89.0</td>
</tr>
<tr>
<td>2:Map task</td>
<td>9</td>
<td>Video</td>
<td>29.7</td>
<td>61.4</td>
<td>91.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Avatar</td>
<td>2.0</td>
<td>91.2</td>
<td>93.2</td>
</tr>
<tr>
<td>3:Social exchange</td>
<td>5</td>
<td>Video</td>
<td>83.6</td>
<td>13.2</td>
<td>96.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Avatar</td>
<td>67.8</td>
<td>17.0</td>
<td>84.8</td>
</tr>
<tr>
<td>4:Business meeting</td>
<td>14</td>
<td>Video</td>
<td>56.6</td>
<td>33.9</td>
<td>90.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Avatar</td>
<td>15.3</td>
<td>64.8</td>
<td>80.1</td>
</tr>
</tbody>
</table>

Table 1
Mean percentages of gaze directed on screen to representations and task feature/rest of screen for each task by condition.
avatar icons on screen. While participants in the video condition also gazed more often at the map they did direct almost 30% of their fixations to the video images presented on screen. This interaction indicates that although participants do gaze more often at the task feature in problem solving tasks, there appears to be some benefit of providing a video link over static forms of representation as participants direct almost a third of their attention to the videos on screen.

4.1.3. Task 3. No focus (social exchange): gaze behaviour

As in the first task, the only information presented on screen was the representations of the other participants. A significant effect of relevance of information was found (F(1, 8) = 45.32, p < .05) as participants in both conditions gazed more often at the representations of the other participants compared to the rest of the screen. There was no effect of representation (F(1, 8) = 2.15, p > .05) and no interaction between relevance of information and representation (F < 1.25).

The patterns of gaze found were remarkably similar to the first social exchange task. Although participants did gaze slightly more often at the video images compared to the avatar icons (see Table 1) this difference was not significant. Again, this occurs despite the fact that the avatar icons provide no information in terms of physical appearance, facial expressions and other types of non-verbal behaviour.

4.1.4. Gaze behaviour over time

It was expected that users’ gaze behaviour might differ in the third task compared to the first task, as the participants will have had time to become familiar with the different communication media. For example, they may gaze more often at the video images or less often at the avatar icons. However, a 3 factor mixed ANOVA comparing gaze behaviour across the two tasks (with task as a within subjects factor, relevance of information as a within subjects factor and representation as a between subjects factor) revealed a significant effect of relevance of information only (F(1, 8) = 89.71, p < .05) as participants gazed more often at the representations compared to the rest of the screen in both tasks. There were no statistically significant effects of representation (F(1, 8) = 1.94, p > .05), task (F < 1) or any interactions between the three factors (all Fs < 3.01). This could be explained in terms of the task and the fact that there was no other information to look at and therefore the representations always attract the highest percentage of gaze regardless of length of exposure.

4.1.5. Task 4. Mixed focus (business meeting): gaze behaviour

The business meeting task was included to explore how often representations are looked at when there is other task information on screen, relevant to the task but not essential for task completion as the map was in task 2. A significant interaction was found F(1, 26) = 34.21, p < .05 as partici-

pats gazed more often at the shared visual graphic (main effect of relevance of information F(1, 26) = 4.71, p < .05) compared to the representations of the other participants, but looked more often at the shared data in the avatar condition. There was also a significant effect of representation F(1, 26) = 18.96, p < .05. The interaction indicates that video images attract more attention compared to avatar icons, and that shared visual data relevant to the task are gazed at more often than avatar representations.

This interaction can be compared to gaze behaviour in the second task, the map task. An interaction was found in both instances but the means reveal differences related to the two tasks. In the map task for example, participants in the avatar condition directed only 2% of fixations to the avatar icons whereas in the business meeting task they direct 15% of fixations to the avatar icons on screen. In the video condition, participants direct 30% of fixations to the video images in the map task and 57% to video images in the business meeting task. The patterns of gaze found in the avatar condition are somewhat surprising, as there is a marked increase in the percentage of gaze directed to the avatar icons in the business meeting task, despite the fact that these icons display little personal information and that there is other information to look at on screen.

4.2. Communication analysis

All twenty-four three-person discussions for each task were audio recorded. Due to recording problems unknown at the time, six of the conversations were lost. In order to maintain the same number in each condition for each task, two other conversations were discarded. These interactions were transcribed (including interruptions) and checked. For each task, the transcriptions of sixteen three-person interactions were measured in terms of conversation structure—the total number of words, turns, and interruptions by role. In addition, the number of interruptions as a percentage of the total number of turns was also calculated in order to compare interruptions across the two conditions while accounting for variations in the number of turns. Further analyses carried out for the second and fourth tasks are described in more detail below.

The main hypothesis regarding communication in this study was that participants in the avatar condition would experience more difficulty in co-ordinating turns of speech, due to the lack of information about non-verbal behaviours (H2). This should be reflected in an increased number of interruptions made by participants in comparison to those in the video condition. For each measure, number of interruptions and percentage of interruptions, a 2 × 2 mixed design ANOVA was carried out with representation as a between subjects variable (two levels: video and avatar) and participant/task role as a repeated measure (three levels: representing each participant in the group).
4.2.1. Task 1. No focus (social exchange): communication

It was predicted that due to the novelty of the representations on screen and lack of visual cues that are known to be important for turn taking, participants would interrupt one another more often in the avatar condition. A significant effect of role was found ($F(2, 28) = 3.61, p < .05$) and pair wise comparisons indicated that person 1 in each condition (identified as the first person to speak) interrupted more often than the other two participants. There was no effect of representation ($F < 1.7$) and no interaction between role and representation ($F < 1$). The hypothesis is not supported although the means indicate a small numerical difference, with more interruptions in the avatar condition (average per person, 9.3) compared to the video condition (5.8) as illustrated in Table 2.

In order to standardise for conversations of different lengths, the percentage of interruptions was also calculated in terms of number of interruptions as a percentage of the total number of turns. A main effect of representation just missed significance ($p = .07$). There was no effect of role or an interaction between role and representation (all $Fs < 1$). This indicates that despite being the first task for all of the participants, those in the avatar condition seem to have more difficulty in co-ordinating turns and this is reflected in a trend towards more interruptions in the avatar condition.

4.2.2. Task 2. Total focus (map task): communication

The means can be found in Table 3. For number of interruptions, there were no significant effects of role, medium of communication and no interaction between the two factors (all $Fs < 2.17$).

The analysis of percentage of interruptions revealed a significant effect of media ($F(2, 28) = 4.29, p = .058$) as there were more interruptions in the avatar condition compared to the video condition. There was no effect of role and no interaction between role and media (all $Fs < 1.12$).

The hypothesis was confirmed as the total number of interruptions by each participant revealed a significant effect of representation ($F(1, 14) = 5.29, p < .05$) as participants in the avatar condition interrupted one another more often compared to participants in the video condition (see Table 4). There was no effect of role and no significant interaction between role and representation (both $Fs < 1$). The fact that this result was not significant for the first task also suggests that the benefits of video may be related to exposure.

The number of interruptions as a percentage of the total number of turns was analysed to account for differences in conversation lengths. A significant effect of representation was found ($F(1, 14) = 6.64, p < .05$) as there were more interruptions in the avatar condition compared to the video condition. There was no effect of role and no interaction between role and representation (both $Fs < 1.12$).

4.2.3. Task 3. No focus (social exchange): communication

The hypothesis was confirmed as the total number of interruptions by each participant revealed a significant effect of representation ($F(1, 14) = 5.29, p < .05$) as participants in the avatar condition interrupted one another more often compared to participants in the video condition. The conversations did not differ significantly across conditions in terms of total number of words and turns per person (all $ps > .05$). Therefore, while the presence of visual feedback in the video condition enables the participants to manage turn taking, the same information does not seem to result in more efficient communication compared to the avatar condition.

### Table 2

<table>
<thead>
<tr>
<th>Condition-Representation</th>
<th>Person 1 (%)</th>
<th>Person 2 (%)</th>
<th>Person 3 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Interruptions (avatar)</td>
<td>12.25 (9.91)</td>
<td>9.00 (8.34)</td>
<td>6.62 (3.7)</td>
</tr>
<tr>
<td>N Interruptions (video)</td>
<td>7.50 (6.04)</td>
<td>4.37 (4.30)</td>
<td>5.50 (3.50)</td>
</tr>
<tr>
<td>% Interruptions (avatar)</td>
<td>22.09 (11.07)</td>
<td>19.08 (12.53)</td>
<td>18.42 (8.52)</td>
</tr>
<tr>
<td>% Interruptions (video)</td>
<td>13.35 (7.4)</td>
<td>11.21 (7.27)</td>
<td>14.02 (6.53)</td>
</tr>
</tbody>
</table>

### Table 3

<table>
<thead>
<tr>
<th>Condition-Representation</th>
<th>Person 1 (%)</th>
<th>Person 2 (%)</th>
<th>Person 3 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Interruptions (avatar)</td>
<td>8.12 (6.10)</td>
<td>15.25 (16.67)</td>
<td>7.87 (4.01)</td>
</tr>
<tr>
<td>N Interruptions (video)</td>
<td>8.12 (7.21)</td>
<td>10.25 (4.59)</td>
<td>7.12 (5.19)</td>
</tr>
<tr>
<td>% Interruptions (avatar)</td>
<td>16.02 (11.58)</td>
<td>23.14 (12.81)</td>
<td>20.11 (8.29)</td>
</tr>
<tr>
<td>% Interruptions (video)</td>
<td>14.04 (7.44)</td>
<td>17.64 (8.23)</td>
<td>14.43 (5.75)</td>
</tr>
</tbody>
</table>

### Table 4

<table>
<thead>
<tr>
<th>Condition</th>
<th>Person 1 (%)</th>
<th>Person 2 (%)</th>
<th>Person 3 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N Interruptions (avatar)</td>
<td>8.75 (13.13)</td>
<td>7.00 (6.27)</td>
<td>7.62 (4.10)</td>
</tr>
<tr>
<td>N Interruptions (video)</td>
<td>2.12 (2.03)</td>
<td>1.75 (1.28)</td>
<td>1.87 (1.64)</td>
</tr>
<tr>
<td>% Interruptions (avatar)</td>
<td>20.90 (13.61)</td>
<td>19.01 (8.15)</td>
<td>23.93 (10.55)</td>
</tr>
<tr>
<td>% Interruptions (video)</td>
<td>8.89 (8.12)</td>
<td>10.72 (6.99)</td>
<td>12.74 (12.16)</td>
</tr>
</tbody>
</table>
the first task overall (average of 7.5 per person) compared to the second task (4.8). There were no effects of role or any interactions between the three factors (all Fs<2.53). There was however, a trend towards an effect of representation (F(1, 14)= 3.47, p = 0.08) and this reflects the fact that there were more interruptions in the avatar condition compared to the video condition. The percentage of interruptions was also compared in order to account for any variances in the total number of turns in each task and condition. There were no main effects of task, role or any interactions between the three factors (all Fs<2.05) but there was a significant effect of representation F(1, 14) = 5.80, p < .05. This revealed a significantly higher percentage of interruptions in the avatar condition.

While users experience more difficulty in managing their turns in the avatar condition overall, the benefits of video are clearly related to exposure as the percentage of interruptions is reduced in the second social exchange task compared to the first (refer to Tables 2 and 4). As participants become more familiar with the video images, they are able to use them to provide information about the other participants and to co-ordinate turn taking. This obviously does not happen in the avatar condition, as there is no visual information about the other participants’ non-verbal behaviours.

4.2.5. Task 4 mixed focus (business meeting): communication

The task was designed in order to explore whether avatars may have an equalising effect on status-differentiated interactions. It was predicted that there would be more participant-high status interactions compared to participant-low status interactions (i.e. exchanges of turns of speaking) in both conditions although this difference should be greater in the video condition due to extra status cues in terms of visual information about the participants. This was analysed by counting the number of interactions between the participant and high status confederate, and the participant and the low status confederate in both conditions. A 2 x 2 mixed ANOVA with status as a within subjects measure (2 levels; interactions with the high and low status confederate) and representation as a between subjects measure (2 levels: video and avatar) was conducted. A main effect of status was found as there were more participant-high status confederate interactions compared to participant-low status confederate interactions in both conditions F(1, 14) = 87.1, p<.05. There was no effect of representation and no interaction between representation and status (both Fs<1). While this result does not reflect the equalising effect of avatar icons, it can be explained in terms of the task roles as it is likely that the high status confederate would ask the participant more questions and therefore interact more often with the participant compared to the low status confederate. While the confederates in this study were following a script, it may be that in real-life business meetings the effects of status on communication patterns are so strong that they overcome any differences in communication media.

While the patterns of interactions indicate that the technology has no equalising effect on meetings between persons of different status, the interruption data suggested some interesting differences. A significant interaction (F(2, 28) = 7.2, p < .05) revealed that the high status confederate interrupted significantly more often in the video condition. The same result was found for the percentage of interruptions (ps<.05). While this result was not predicted one can attribute it to a number of factors. As it was only the high status confederate who interrupted more often in the video condition this may be related to the fact that differences in status will be more pronounced in the video condition due to extra visual information about the participants. Perhaps the confederate was more self-conscious in the video condition and interrupted more often in order to enhance their role as the high status person who would be most likely to dominate the discussion (as found in face to face discussions, Carletta et al., 1998). One must also acknowledge the problems involved in using confederates. While the set up did seem to be entirely convincing to the real participant, it is probable that the confederates became tired of reading from the same script and this obviously affects patterns of communication. Nevertheless, it would be interesting to see if this finding would be replicated in a real-life business meeting, incorporating persons of different occupational status.

4.3. Task 2: map task performance

The map task provides a quantitative measure of task performance. A t-test was used to compare route deviation scores in the two conditions. This was calculated as the total number of cm² that the instruction follower’s route deviated from the correct route on the map. There was no significant difference in route deviations in the two conditions t(18) = 0.29, p > .05. The mean scores (n = 12) were extremely close with participants in the avatar condition averaging 35.6 cm deviation and those in the video condition 38.1 cm. The maps were completed equally well in both conditions of representation. This replicates previous findings with this task (e.g. Boyle et al., 1994; Anderson et al., 1997) which found that task performance is unaffected by the medium of communication.

4.4. Questionnaire results

It was predicted that participants would rate the video condition more favourably in subjective ratings of the technology due to the extra visual information provided (e.g. gaze, gestures, posture) compared to the avatar condition (H3). All questions were posed using a scale of zero to one hundred and for each question zero represented a negative and one hundred represented a positive response (e.g. zero represented very cold/very unsociable/impersonal and one hundred represented very warm/very sociable and
very personal). The data can be described as interval data as the scale is continuous and this was done so that comparisons could be made between groups and within tasks using analysis of variance. Although all of the participants were required to complete a questionnaire after each task only 20 participants in total completed all of the 4 tasks and these responses were analysed.

Overall, the participants’ evaluations were very similar and positive in both conditions. Surprisingly, only one question revealed a significant main effect of representation: ‘How useful were the representations of the other participants on screen (video images/avatars)?’ 0 = not useful at all, 100 = very useful. As expected, the video images were rated as being far more useful (71.5) than the avatar icons (36.1), $F(1, 19) = 31.2, p < .05$. There was also a significant effect of task for this question ($F(1, 19) = 3.86, p < .05$) as scores for the map task (task 2) were significantly lower than those for task 4. There was no interaction between task and representation ($F = 2.04$). Although this is an important result, it does suggest that the overall usefulness of the representations as perceived by participants does not necessarily match with other ratings of the medium. Of the four questions regarding the more social aspects of the medium such as sociability, warmth and sensitivity, there were no main effects of representation (all $F_s < 1$). There were however, two interactions between task and representation for the questions regarding how social participants found the medium ($F(1, 19) = 2.76, p = .05$) and how sensitive they found the medium ($F(1, 19) = 3.68, p < .05$). These interactions revealed that responses in the video condition varied across tasks, decreasing for the map task and then increasing for the final two tasks. This may be expected, especially as the last task was the business meeting task and placed greater emphasis on social behaviours. However, the most surprising finding is that this interaction is not in any way due to any differences between the video and avatar conditions and in fact responses in the avatar condition were more consistent across the four different tasks as illustrated in Table 5.

<table>
<thead>
<tr>
<th>How sociable would you rate the medium of communication? 0 = very unsociable, 100 = very sociable</th>
<th>Task 1</th>
<th>Task 2</th>
<th>Task 3</th>
<th>Task 4</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avatar</td>
<td>69.2</td>
<td>61.9</td>
<td>69.0</td>
<td>64.1</td>
<td>66.1</td>
</tr>
<tr>
<td>Video</td>
<td>62.0</td>
<td>58.6</td>
<td>69.2</td>
<td>66.7</td>
<td>64.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How sensitive did you find this medium of communication? 0 = very insensitive, 100 = very sensitive</th>
<th>Task 1</th>
<th>Task 2</th>
<th>Task 3</th>
<th>Task 4</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avatar</td>
<td>62.0</td>
<td>59.6</td>
<td>55.7</td>
<td>60.4</td>
<td>59.4</td>
</tr>
<tr>
<td>Video</td>
<td>57.8</td>
<td>57.0</td>
<td>65.9</td>
<td>69.7</td>
<td>62.1</td>
</tr>
</tbody>
</table>

There were no differences between the two conditions for the three questions concerning the technology and how easy it was to use, to communicate information and to understand the other participants in the group. For all three questions there was only a significant effect of task (all $p_s < .05$) and this revealed that responses were lower for the map task compared to the other three tasks in both conditions of communication. As the map task was the only problem solving task this indicates that participants experienced more difficulty with this task and found it more difficult to communicate information and to understand the other participants in the group. This is confirmed by the fact that participants did rate the map task as most difficult in both conditions on the question concerning task difficulty (only significant effect of task, $p < .05$). These results are still interesting as they reveal that despite having no visual feedback in the avatar condition; participants did not rate the medium as being any more difficult to use and to communicate with other participants than the video condition.

For the two questions relating to the participants’ enthusiasm to use the system in the future there were only significant effects of task ($p_s < .05$). In both instances responses were significantly lower (and reflected less enthusiasm) for the map task compared to the other three tasks. Finally, the question ‘How do you rate your enthusiasm to use the system in the future?’ (0 = very negative, 100 = very positive), revealed no significant effects of task, representation or an interaction between the two (all $F_s < 2.5$). The responses were quite positive and in fact were numerically higher for the avatar conditions (average scores: 75.6 for the avatar condition and 72.2 for the video condition).

It is perhaps surprising that participants reacted so favourably to the simple representations in our avatar condition. It seems that the simple device of adding personal names and a gendered icon provides a similar experience to that provided by high quality video images. This result does not support theories emphasising social presence (Short et al., 1976), which suggest that media that provide a wide range of visual cues engender a strong sense of social presence among users. However, many of these results may be task dependent and reflect the participants’ reaction to the task and not the technology. For example, the map task was rated as being the most difficult task out of the four and participants also found it more difficult to communicate with others in the group during this task regardless of medium of communication. In terms of users’ subjective ratings, these results suggest that avatars are able to support remote communication as well as video images on a number of measures, including social aspects, and especially in the case of problem solving tasks.

5. Discussion

5.1. Discussion of the findings on behavioural data

The main prediction regarding video images would attract higher percentages of gaze compared to avatar icons, as live faces provide more
information than static grey faces. Although participants did gaze more at video images in the first and third tasks (no focus-social exchange), the difference was not statistically significant. In the second and fourth tasks, an interesting interaction between representation and relevance of information was found. Participants gazed more often at the shared data presented on screen in both of these tasks, but participants did gaze more often at the representations of the other participants in the video condition. The results from the four tasks tend to support the main prediction that video images are gazed at more often than avatar icons.

Two of the four tasks involved presenting shared information on screen and the results suggest that shared data is more important than views of remote participants and this supports previous research in this area (e.g. Anderson et al., 2000). Interestingly, the gaze patterns strongly reflected task differences. In the business meeting task, participants directed substantially more gaze to the avatar icons than they did in the map task. While the same pattern was found in the video condition, this result is interesting as it suggests that users do gaze at representations on screen, even when there is other relevant information on screen, and despite the fact that these icons present little communicative information. This suggests that the named icon provides a minimal cue that is able to support group interaction, and this therefore attracts more attention in a social task compared to a problem-solving task. This is important for research on different types of representations (Bowers et al., 1996; Dehn and Mulken, 2000) as the results suggest that even basic types of representations attract participants’ gaze and therefore seem to be utilised in some way during remote communication.

It was hypothesised that video images would attract more gaze as the communicative information provided is thought to be important for users in terms of managing turn-taking and improving informational transfer. If video images attract more gaze than avatar icons then this should be reflected in a reduced number of interruptions for users interacting with video. Generally, this was found although in the first task the difference was not statistically significant. Interestingly, the social exchange tasks revealed a small effect of exposure for video images as the number of interruptions was reduced in the third task compared to the first (no difference for the avatar condition). This finding relates to a study by Boyle et al. (1994) who investigated the effect of familiarity during the Map Task in face to face interactions. They found that familiar pairs said significantly more and looked at one another more often than unfamiliar pairs but interrupted one another significantly less often. Although this study observed no difference in gaze patterns, we did find that as participants became more familiar they interrupted one another less, noticeably in the video condition. This may indicate that the benefits of video, in terms of providing important communicative information, are utilised when users have had the opportunity to become familiar with the medium of communication.

The business meeting task incorporated status differences and it was expected that status differences would be less pronounced in the avatar condition due to the absence of visual status cues. Firstly, there was a main effect of status as conversations in both conditions consisted of more interactions between the participant and the high status confederate compared to the participant and the low status confederate. Secondly, there was an interaction found for the number of interruptions whereby the high status confederate interrupted more often in the video condition. These results are probably due to the task manipulation although they do indicate that the status differences worked equally well in both conditions of communication. Obviously it would be more beneficial to conduct a field study whereby the participants were interacting according to real-life status-differentiated roles. Nonetheless, it appears that status effects are robust, even when participants are represented by basic graphical representations, and this reveals another aspect of how participants interact via avatars. The results suggest that status effects are not wholly dependent on visual information and supports research by France et al. (2001). However, one could argue that avatars may be able to provide some equalising effect if visual cues in video mediated communication were reinforcing stereotypical cues. For example, in meetings where someone representing the higher status person is older, dressed formally and typically male, then perhaps avatar representations may offer some benefit in terms of equalisation compared to video images, particularly in a field study when the participants are not following a task script. It would be particularly interesting to see if real-life participants communicated differently according to how they were represented on screen, for example would the high status participant interrupt more often when represented by video instead of avatar as found in this study?

The four tasks were included to compare behaviour in a range of different contexts and in the map task it was possible to measure task performance. Previous research using this task would predict that task performance would be the same in both conditions but that more talk would be required to complete the task in the avatar condition (Boyle et al., 1994). While the same performance was found in both conditions, there was no difference in the number of words needed to complete the task. This suggests that users completing a problem-solving task perform equally well without visual information about their remote collaborators and without any cost in terms of extra dialogue. This result relates to a study by Matarazzo and Sellen (2000) which found that participants rated poor quality video conditions more highly than good quality video conditions and also completed tasks faster in the poor quality condition. The findings were explained in terms of a ‘distraction effect’ that suggests participants may be distracted by high quality video links of remote collaborators and this may affect task
performance. Our findings suggest that as basic representations enable participants to achieve the same level of task performance as video images, the amount of visual information required to complete a problem-solving task is minimal in terms of how the remote collaborators are represented. Fussell et al. (2003) suggest that people rarely look at faces during collaborative physical tasks and instead distribute their gaze evenly across other targets including task objects and actions.

5.2. Discussion of the findings on user’s attitudes

The objective measures of gaze behaviour and communication have revealed interesting findings although perhaps the most surprising results were reflected in the questionnaire responses. It was expected that participants would rate the video condition more highly and more favourably than the avatar condition, especially on several questions concerning social factors. However, there was only one significant effect of representation as participants rated the representations as being far more useful in the video condition. On all of the other questions, there was no main effect of representation and the responses were remarkably similar, and overall quite positive. While this result is somewhat surprising, it does indicate that subjective measures are useful in this type of study as they provide additional information to the objective measures employed. For example, while video images may be effective in terms of reducing the number of interruptions, they are not perceived to be more sociable, warm or sensitive compared to avatar icons. Also, while avatar icons may be rated similarly to video images on a range of questions, there was a large difference between how useful the participants perceived the representations to be. It would have been beneficial to interview the participants at the end of the session in order to try to understand these subjective preferences and perhaps this could be incorporated in any future studies of this kind. One factor that may have influenced these results is familiarity and the fact that participants were communicating with strangers. It would be interesting to see if the same ratings would be found if participants were communicating with strangers. A number of studies have suggested that video images can be distracting for task oriented work and ‘poorer’ forms of video are preferred over high-quality video (e.g. Takao, 1999; Matarazzo and Sellen, 2000) and there may a practical use for employing basic representations in mediated communication. There is also a body of research concerning gaze aversion during face-to-face communication, particularly during difficult discussions (e.g. Kendon, 1967; Beattie, 1981) and this may also be related to the distracting effect of employing ‘real faces’ as representations of remote participants. Avatar representations may hold more for remote communication than videos as they allow users to construct their own representation of their interlocutors without the distraction of a live video image.

This study compared basic avatar representations to high quality video images and there is much scope to explore the influence of more advanced types of graphical representations on remote communication. Garau et al. (2003) investigated avatar realism and eye gaze control in a shared virtual environment. Following on from their previous study (Garau et al., 2001) they found that inferred-gaze avatars were preferred to random gaze although this depended on appearance as responses to a low-realism avatar (match-stick appearance in comparison to photo realistic high-realism avatars) were adversely affected by inferred gaze. Interestingly, responses to a number of questions were unaffected by avatar realism and this relates to the questionnaire findings in this study.

This study has demonstrated the value of employing a number of measures to evaluate computer-mediated interactions. By using techniques such as eye tracking alongside more traditional measures of communication, this type of research could reveal more about attention and communication in realistic mediated communication settings. One could manipulate many different types of information in order to determine the optimum visual representations for remote users in different contexts (e.g. dress, gestures, size, and surrounding virtual environment). As mentioned previously, it would be highly beneficial to conduct studies of real-life interactions using different representations, and in a variety of communicative contexts (e.g. friends chatting, status-differentiated business meetings and problem-solving situations). While the scope of this study has been rather broad in exploring how the focus of attention impacts on gaze and communication, it would be interesting to take each different real-life situation and explore in detail the optimum representation and screen environment.

5.3. Future directions

The critical comparison to investigate is whether avatar representations on screen can provide extra benefits for audio only communication. For example, we found that while these icons display little personal information (name and gender only) they do attract more attention in a social task compared to a problem-solving task. This suggests that they may be used as social prompts for the participants, as they serve as a reminder of other participants presence, and display basic information that may be extremely useful when communicating with strangers for example. A number of studies have suggested that video images can be distracting for task oriented work and ‘poorer’ forms of video are preferred over high-quality video (e.g. Takao, 1999; Matarazzo and Sellen, 2000) and there may a practical use for employing basic representations in mediated communication. There is also a body of research concerning gaze aversion during face-to-face communication, particularly during difficult discussions (e.g. Kendon, 1967; Beattie, 1981) and this may also be related to the distracting effect of employing ‘real faces’ as representations of remote participants. Avatar representations may hold more for remote communication than videos as they allow users to construct their own representation of their interlocutors without the distraction of a live video image.

5.4. Implications for the design of remote communication systems

There are a number of implications for the design of remote communication systems. For example, the context
of the communication is important, as it appears to be more important to see one’s conversational partner in more social tasks compared to problem solving tasks. The requirements of the user must also be taken into consideration as if the outcome of a remote meeting is more important than the actual process of communication then one could recommend that avatars be employed instead of video images as they enable the same level of task performance in a problem solving situation. However, if the process of communication was a more crucial factor, then video images of remote participants would be recommended as the visual feedback they provide can facilitate turn taking. In terms of applying this type of technology to remote distributed groups, it can be concluded that the only major advantages of providing rich video representations of participants is that these attract more gaze and can facilitate turn taking. Avatar representations enable participants to achieve the same level of task performance and they are rated equally to video representations despite the fact that users may perceive videos to be more useful representations than avatar icons. The fact that this study compared video to one of the most basic graphical representations available makes these findings even more remarkable.

6. Conclusion

The aim of this study was to explore the effects of different types of representation on patterns of group communication, gaze behaviour and questionnaire responses. Communication results indicate that video images may be beneficial for remote users in terms of reducing the number of interruptions and this is probably due to the fact that videos attract more attention in comparison to avatar icons and provide useful visual cues for managing turn taking. However, this benefit is not related to task performance or subjective preferences as in both measures avatars function as well as video images. These results are quite inconsistent with theories that emphasise the importance of visual information in communication (Short et al., 1976; Daft and Lengel, 1986). The results do however support studies that suggest video images do not provide extra benefits for problem solving tasks (e.g. Williams et al., 1977; Matarazzo and Sellen, 2000). In addition, the results extend previous research on representations (Parise et al., 1996; Lantz, 2001) and reveal more about how participants interact with graphical representations of their remote collaborators.

In conclusion, it appears that low cost, low bandwidth forms of representation are able to provide remote interlocutors with a context for communication that is not very different to that obtained from paying for the cost, bandwidth and equipment for high-quality, live video images. As more people are attracted to communicating with others via computer it appears that future developments may revolve around graphical representations rather than the use of video. If this is the case then it is no doubt valuable to conduct more research into how these graphical representations affect both human-computer and human-human interactions.

This research was supported by a UK Economic and Social Research Council studentship (Award no: 500429937017) with co-sponsorship from BT plc. We thank Elizabeth Churchill and three anonymous reviewers for valuable comments on a previous version of this paper.

References


