GAZE-2: Conveying Eye Contact in Group Video Conferencing Using Eye-Controlled Camera Direction

Roel Vertegaal, Ivo Weevers, Changuk Sohn and Chris Cheung
Human Media Lab
Queen's University
Kingston, ON K7L3 N6, Canada
{roel,weevers,sohn,cheung}@cs.queensu.ca

ABSTRACT

GAZE-2 is a novel group video conferencing system that uses eye-controlled camera direction to ensure parallax-free transmission of eye contact. To convey eye contact, GAZE-2 employs a video tunnel that allows placement of cameras behind participant images on the screen. To avoid parallax, GAZE-2 automatically directs the cameras in this video tunnel using an eye tracker, selecting a single camera closest to where the user is looking for broadcast. Images of users are displayed in a virtual meeting room, and rotated towards the participant each user looks at. This way, eye contact can be conveyed to any number of users with only a single video stream per user. We empirically evaluated whether eye contact perception is affected by automated camera direction, which causes angular shifts in the transmitted images. Findings suggest camera shifts do not affect eye contact perception, and are not considered highly distractive.

KEYWORDS: Attentive User Interfaces, Multiparty Video Conferencing, Gaze, Eye Contact, Eye Tracking.

INTRODUCTION

With the recent resurge of interest in video conferencing as a means to conduct business meetings without travel comes a renewed interest in the usability of this technology to support group conversations. One potential usability problem is that video conferencing does not necessarily support the regulation of conversational turn taking any better than telephony-based systems [28]. In multiparty conversations, when the current speaker falls silent, it is not obvious who will be the next speaker. Previous research suggests that the looking behavior of conversational partners, or more specifically, their eye contact with each other, plays a critical role in determining who is to be the next speaker in group conversations [15, 28]. Most group video conferencing systems do not convey eye contact correctly, and thus inhibit a smooth group turn taking process [25]. There are a number of reasons for this. Firstly, most systems employ only a single video camera, mounted on top of the screen. This causes a visual parallax between the location of the camera lens and the on-screen representation of participants. According to Chen [4], to avoid the appearance of eye gaze being lowered or otherwise skewed, the video camera needs to be located to within 1 degree horizontally and 5 degrees vertically from the on-screen representation of the eyes of each conversational partner. With multiple participants represented on the screen, correct camera placement can therefore not be guaranteed. The solution to this problem is to use multiple cameras, one for each participant [3]. However, the use of multiple cameras provides no guarantee that eye contact is conveyed. Any on-screen head movement by participants greater than 1 degree of visual angle will re-introduce a parallax [28]. As such, previous systems supporting eye contact have required their participants to sit still, with on-screen video images placed at a fixed and predetermined location [21]. The second problem with the use of multiple cameras is that the number of video connections rises almost quadratically with the number of participants. In such systems, eye contact is conveyed by sending each participant images taken from the perspective of their eyes on the screens of other users. When there are four participants, each participant therefore requires three cameras, one for each other participant, resulting in 12 unique video streams. With six participants, this number is 30. These unique video streams cannot be compressed into a single stream [11]. Given the network-intensive nature of video conferencing this may result in severely reduced picture quality, possibly forfeiting the purpose of the video link altogether.

GAZE-2 [26] provides a novel hybrid between single camera and multiple camera approaches. It employs a large array of cameras, but selects only one camera from this array for transmission at any time. Using a low-cost eye tracker, the system measures which participant each user looks at. It selects the corresponding camera for multicast to all participants, thus guaranteeing a parallax-free image with eye contact. Video images of participants in GAZE-2 are represented as 2D planes suspended in a 3D virtual meeting room. To avoid all participants having continuous eye contact with each other, the system automatically rotates each video image to face the participant the respective user looks at, as measured by

CHI 2003, April 5–10, 2003, Ft. Lauderdale, Florida, USA.
Copyright 2003 ACM 1-58113-630-7/03/0004…$5.00.

her eye tracker. This way, eye contact can be accurately conveyed to any number of users, rendered at arbitrary on-screen locations, with only a single stream of video per user. Whenever the automated camera director selects a new camera for transmission, which is every time a user looks at a new participant, this results in an angular shift in the video image. We empirically evaluated whether users of our automated camera direction system would correctly perceive eye contact during these shifts. Findings suggest angular shifts do not affect the perception of eye contact. Moreover, participants of our study did not consider the shifts to be highly distractive. However, we do recommend that cameras be placed within 8 degrees (or about 10 cm at arms length) of each other. We will first provide an overview of previous work, including a discussion of systems that mediate gaze direction. We will then discuss the design of the camera director hardware, and of our 3D communication environment. Finally, we will discuss the empirical study and design recommendations.

PREVIOUS EMPIRICAL WORK

There is strong empirical evidence justifying the value of conveying eye contact in group communication systems. In this section, we will discuss several of the experiments that demonstrate the sensitivity of participants to eye contact, as well as the role of eye contact in regulating multiparty conversation.

User Sensitivity to Eye Contact

People are capable of determining whom others are looking at with great accuracy. Von Cranach and Ellgring [30] reported that observers, located 1.5 m away and at right angles to the axis between two interactors, correctly identified more than 60% of the fixations by one interactor at the nose bridge of the other interactor as being inside the facial region. Given the extreme angle, they found observers relied mostly on head position. According to Argyle and Cook [1], when the observer is the other interactor, the accuracy is much greater as observers can rely on eye positional information. Participants, at a distance of 2 m from another person facing them, have been reported to judge 84% of fixations by that person at their nose bridge correctly as 'looking directly at me' [14]. Jaspars et al. [16] suggest that from a distance of about 1 m, people are able to discriminate the gaze position of someone facing them with an accuracy of approximately 1 cm in their facial plane (which relates to .6 degrees). The accuracy of eye contact perception may, however, be altered by the use of a video link. In [4], Chen challenged the idea that sensitivity to eye contact perception in video images is symmetric. He asked actors to look at targets projected around a video camera lens mounted behind a projection screen. Observers then looked at the recorded images to judge whether they experienced eye contact with the looker. He found a tolerance to vertical, downward, parallax of about 5 degrees. This suggests that when a video conferencing camera is mounted on top of the screen, with the image of a participant positioned just below it, eye contact should easily be maintained. However, according to Chen's data, in order to achieve 90% accuracy of ratings, parallax in any other direction should be between 0 and 1 degrees of visual angle [5]. This means that when multiple participants are represented, multiple video cameras are required. It also means that participant windows cannot be positioned arbitrarily, and participants should not move their heads within their video frames. Depending on the distance of observers and the type of camera lens, any on-screen horizontal movement of their eyes greater than about 1 cm may be sufficient to break eye contact.

Eye Contact in Multiparty Mediated Conversation

Most studies of eye contact during conversations focused on two-person communication [1]. However, multiparty conversational structure is much more complicated than its dyadic equivalent. As soon as a third speaker is introduced, the next turn is no longer guaranteed to be a specific non-speaker. This poses problems for the regulation of turn taking. When a speaker yields the floor in a multiparty situation, the question arises to *whom* she yields the floor. It has long been presumed that eye contact provides critical information in resolving this process [3, 22]. Isaacs and Tang [15] performed a usability study of a group of five participants using a desktop video conferencing system. They found that during video conferencing, users addressed each other by name and started explicitly requesting individuals to take turns. In face-to-face interaction, they found people used their eye gaze to indicate whom they were addressing or to suggest a next speaker. Similarly, O'Connaill et al. [19] found that in video conferencing more formal techniques were used to achieve speaker switching than in face-to-face interaction. They too attribute this to the absence of gaze-related speaker-switching cues. Sellen [23] was one of the first to formally investigate the effects of eye contact on the turn taking process in four-person video conferencing. Unfortunately, she found no effects because the video conferencing system she implemented did not accurately convey eye contact [23]. Vertegaal et al. [25] studied effects of eye contact on turn taking in three-person video mediated conversations. They found that when eye contact was not conveyed, participants took about 25% fewer turns. Without eye contact, 88% of the participants indicated they had trouble perceiving whom their partners were talking to. Two explanations were suggested for these findings:

1) Eye contact is used to convey whom one speaks or listens to. According to Vertegaal et al. [29], in four-person conversations, there is about a 77% chance that the person being addressed is the person being looked at by the speaker, and about an 88% chance that the person being listened to is the person being looked at. As such, when users cannot observe eye contact, they cannot accurately estimate whether they are being addressed or expected to speak, posing problems in floor management.

2) Eye contact is used to regulate intimacy and arousal in conversations. Argyle and Dean [2] suggested there is an optimal level of intimacy for different communication situations, and that eye contact is an important factor in maintaining this equilibrium. Other factors which affect the equilibrium include physical proximity and intimacy of topic. As such, when users cannot observe eye gaze, they feel less inclined to take the floor.

More recently, Vertegaal et al. [27] compared speaking behavior between two conditions: (1) in which participants experienced eye contact synchronized with turn taking, and (2) in which participants experienced random eye contact. The amount of eye contact in this study was measured as a co-variate. Results showed participants were 22% more likely to speak when looking behavior was synchronized with conversational attention (i.e., whom they were speaking or listening to at the time). However, covariance analysis showed these results were due to minor differences that occurred in the *amount* of eye contact between conditions. As much as 49% of the variance in speaking behavior was explained by the variance in the amount of eye contact. According to Duncan [10], the very reason why turn taking exists is to optimize the participants' attention for a single speaker. Vertegaal et al. suggest the above two explanations may be more intricately related than previously thought, and that the amount of eye contact is used to regulate and optimize the attention of listeners and speakers in multiparty conversation [27]. As such, video mediated systems should support eye contact in order to allow smooth negotiation of participant attention.

PREVIOUS SYSTEMS

Over the years, a number of solutions have been developed to convey eye contact during multiparty video conferencing. These systems can be classified as belonging to one of three categories:

1) *Multi-camera Perspective Systems.* These systems employ multiple cameras to convey unique perspectives from the point of view of each user's virtual on-screen eyes. Often, a video tunnel is used to allow co-location of cameras and images.
2) *Attention-Tracking Avatar Systems.* In these systems, eye contact between users is measured with, for example, an eye tracker and conveyed separately from the images that represent the users.
3) *Eye Contact Synthesis Systems.* In these systems, two or more cameras are used to obtain an image of each user. Images are morphed between cameras to provide a synthesized appearance of eye contact.

Multi-camera Perspective Systems

Hydra [22] is a multi-camera perspective system that simulated a four-way round-table meeting. To convey eye contact, boxes containing a camera, a small monitor and speaker were positioned on a table in places that would otherwise be held by each remote participant. The cameras in Hydra were located below the monitor, rather than above it. This means the parallax was likely too large to convey eye contact properly [23]. Consequently, Sellen did not observe any effects of the use of her system during empirical evaluation [23]. In the MAJIC system [20], cameras were placed behind the eyes of users projected on a semi-transparent life-size screen. Although MAJIC exhibited a greater tolerance to parallax due to the large distance of users to the camera units, when users would move significantly, the cameras would need to be repositioned to achieve a parallax-free image. Both Hydra and MAJIC suffer from problems with regards to network performance. Since, at each site, a camera is required for each other participant, network bandwidth use does not scale linearly with the number of users. In fact, all multi-camera perspective systems suffer this problem.

Using Video Tunnels to Position Cameras

To allow parallax-free images, most other systems in this category employ the use of video tunnel technology originally developed by Rosenthal [21]. Using a half-silvered mirror placed at a 45 degree angle to the screen, video tunnels allow placement of camera units behind a projection of a virtual screen in front of the user (see Figure 1). This allows the camera to be located on the same axis as the eyes of a participant represented on the screen. However, video tunnels do not negate afore mentioned problems of multi-camera perspective systems. As Chen's results demonstrate [4], when the eyes of a person represented on the screen shift relative to the camera unit, for example, because that person moves his chair, eye contact is easily lost. To solve this problem, most video tunnels require their users to put their heads inside the video tunnel, so that they remain on-axis with the camera unit. However, we believe it is problematic to require users to restrain head movement in this way.

Attention-Tracking Avatar Systems

A relatively new approach to the conveyance of eye gaze in group communication is the use of avatar-based systems in combination with sensing technology. The original GAZE Groupware System [28] prevented visual parallax by presenting each participant with animated snapshots of other participants, taken before the meeting while looking into the camera lens. These snapshots were suspended in a 3D virtual meeting room that presented each participant with their own unique point of view. Using an eye tracker, GAZE measured which person each participant looked at during a meeting. It then automatically oriented each participant's image towards that person. Although GAZE conveyed eye contact in a manner that preserved its effect on multiparty conversation, it did not allow the use of real-time video images. The iCon System by Taylor and Rowe [24] simulated eye contact between users by superimposing video images onto 3D head models displayed in a shared virtual meeting room. Using computer vision, the iCon

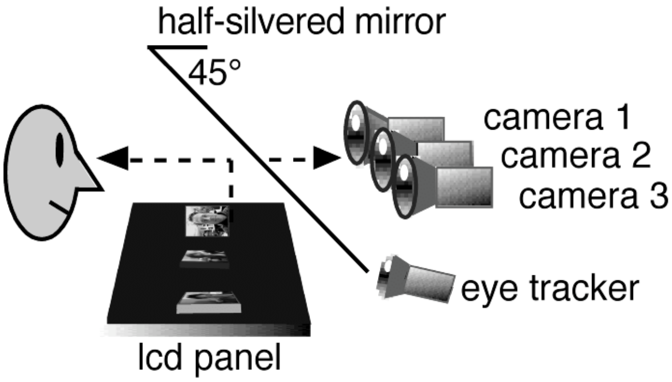

Figure 1. Attentive Video Tunnel. To prevent parallax, an eye tracker mounted in a video tunnel selects the camera the user looks at for broadcast.

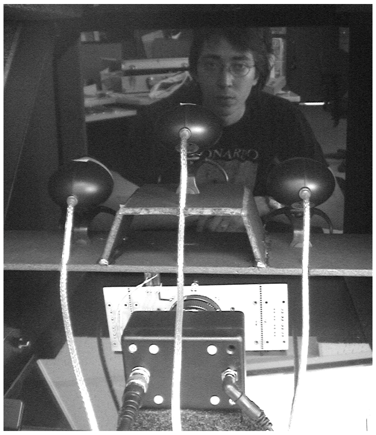

Figure 2. Back side of the Attentive Video Tunnel, with cover removed. Inside, 3 video cameras are mounted behind a half-silvered mirror. The eye tracker is visible below the cameras.

system estimated the gaze orientation of the head by tracking three markers on each user's audio headset. A main benefit of this system is that it allows the rendering of perspective views that show the side of the head. However, camera placement remains an issue since the system employs live video texturing to animate facial features of the head models. To avoid this problem altogether, Colburn et al. [7] developed a multiparty conferencing system that employs realistic 3D avatars of users. Their system is very advanced in that it uses images from a regular video camera to produce realistic 3D head models of each user. A clear benefit of this technology is that there is no parallax problem as no video is captured to render eye contact. However, one of the drawbacks of avatar-mediated conferencing systems is their current lack of realism as compared to video conferencing. It is very difficult to create and animate realistic models of human beings in a conversation [12]. To achieve realism, such systems require elaborate sensing technology that measure or predict the location of facial features, including eye movements, throughout a conference.

Eye Contact Synthesis Systems

A very recent development is that of systems that morph video images taken from different angles in order to produce an accurate representation of gaze direction [13, 31, 17]. Although this is a promising approach, the main problem with these systems has been the associated computational complexity. Yang and Zhang [31] describe a system that finds and warps the eyes in the user's video image to correct visual parallax. However, authors report it is difficult to achieve the correct warp equation for all desired angles, given any initial angle of the user's eyes in a video image [17].

GAZE-2 DESIGN RATIONALE

GAZE-2 provides a novel hybrid between single camera and multiple camera approaches. We will first describe the design of an eye-controlled camera director that automatically selects the camera with the least parallax for transmission to other users. We then describe how images of participants are rotated towards the person they look at in a 3D virtual meeting room.

Designing An Eye-Controlled Camera Director

From the above discussion, it becomes apparent that one of the improvements to be made to Rosenthal's video tunnel is to allow for flexible positioning of participant images on the screen. By placing an array of cameras to cover the display area inside a video tunnel, an automated camera director could ensure that participants would always receive the image from the closest camera.

Figure 1 shows the principle of our Attentive Video Tunnel design using 3 cameras. Cameras are placed behind a half-silvered mirror placed at a 45 degree angle inside a dark box. Below the mirror, a 17" LCD screen is mounted that is reflected by the mirror into the eyes of the user. Three cameras, placed to cover the horizontal range of the monitor, look through the mirror to capture a frontal image of the user's face (see Figure 2). Below the three cameras, we mounted a low-cost eye tracker of our own design, based on IBM's Pupilcam [18]. The eye tracker measures where the user looks on the screen, and uses this information to change the location of the cursor on the LCD screen. Our automated camera director automatically selects the camera closest to where the user looks for multicasting [11] to all other participants.

Figure 3 shows the three different images generated during this switching process. Each picture shows the image from a different camera inside the video tunnel, from left to right, while the user is looking at that camera. Note that as our system switches to a different camera angle, the background of the image shifts. However, when participants shift to a new camera, the eyes of the user are recaptured without any visible parallax. In the next section, we will explain how we designed the GAZE-2 virtual meeting room to communicate looking behavior by rotating parallax-free frontal images of users towards the person they look at.

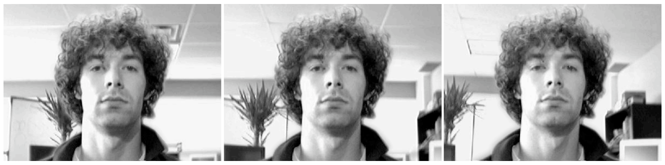

Figure 3. Parallax-free images generated by cameras in the Attentive Video Tunnel. Images from the left, middle and right cameras were taken while the user looked into the lens of the respective camera.

Designing an Attentive Virtual Meeting Room

In GAZE-2, the frontal, parallax-free image of each user is projected on a 2D plane suspended in a 3D virtual meeting room at locations that would otherwise be held in a round-table meeting. Figure 4 shows a four-way conference in progress. The eye tracker inside the video tunnel measures not only which camera a user is looking at, but also which participant that user has eye contact with. This information is sent across the network to all meeting rooms. In each meeting room, each user's video image is automatically rotated in 3D towards the participant he looks at. As users look at different participants, their video tunnel switches to a new camera, portraying a video image from a different camera angle. The rotation of the video window of a user provides a surprisingly powerful affordance of head orientation by the corresponding user. Firstly, like head orientation, the projected surface of a face shrinks with rotation. Secondly, since interlocutors typically establish eye contact with the person they listen or speak to, the limited resolution of peripheral vision strengthens the illusion of head orientation by unattended individuals. When the array of video cameras inside the video tunnel is sufficiently large to cover all angles, this system can convey eye contact accurately to any number of users, rendered at any on-screen location, with only a single stream of video per user.

Networks of Attention

Network bandwidth requirements are a key aspect of the usability of video conferencing systems. In Internet-based video conferencing systems, unavailability of network resources leads to poor image quality through high compression rates, low frame rates, or decreased resolution. Multicasting alleviates network problems by allowing a single video stream to be sent to all users simultaneously, occupying only a single unit of network bandwidth. When GAZE-2 is used by 4 users in multicast mode, only 4 video streams need to be sent over the network. However, when multicasting is unavailable, the use of a single video feed yields little or no bandwidth gains, as video streams need to be sent separately to each participant. To increase network efficiency in such cases, our system allows the employment of attentive video compression [8]. During any multiparty video conferencing session a considerable amount of network bandwidth is wasted. This is because each user is capable of looking only at a single person at a time. The human retina is capable of high-resolution vision only within an area of approximately 2 degrees from the visual axis of the eye, a region called the *fovea* [9]. As such, users that are displayed in peripheral vision can be rendered with reduced resolution. The eye tracker inside the video tunnel allows us to measure exactly which video streams are currently observed, and which are not. The system only requests high-resolution video images from those users that are currently being fixated upon by other users. As exemplified by Figure 4, this is not noticeable to the human eye. In Figure 4, only the left image is sent in hi-res. Try focusing on that image, with your head at a distance of 10 inches. When you now try observing the other users without moving your eye, you notice the limited resolution of your peripheral vision.

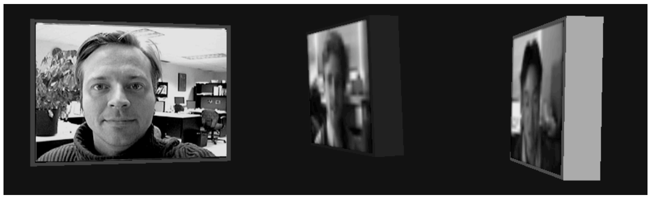

Figure 4. GAZE-2 session with 4 users. Everyone is currently looking at the left person, who's image is broadcast in a higher resolution. Frontal images are rotated to convey gaze direction.

An Artificial Cocktail Party Effect

GAZE-2 employs a similar technique for the attentive compression of audio streams. During multiparty conversations in which there is more than one speaker, the human auditory system is capable of attenuating the voices of unattended individuals. This phenomenon, known as the Cocktail Party Effect [6], allows us to focus our attentive resources on a single speaker, reducing the distraction by background voices. Since people tend to look at the individual they are listening to [29], our system can predict the occurrence of situations in which participants are listening to different speakers (known as side conversations). When users are looking at each other for sustained periods of time, audio from other users that are not part of the side conversation (i.e., that are not looking at a person in the sub-group), is low-pass filtered to about 75% of the original volume. To ensure that this artificial cocktail party effect does not negatively impact the conversational turn taking process, the audio is faded back to its original level as soon as a user looks at another user in a different side conversation. Although we are still in the process of evaluating this feature, we believe that the artificial Cocktail Party Effect may ease the management of side conversations during a conference. Furthermore, since low-pass filtering is implemented by reducing the sampling rate of unattended audio streams, it allows us to further reduce network bandwidth requirements of the system. GAZE-2 provides a dynamic quality of service for video and audio streams that manages network resources to match the attentive resources of the participants.

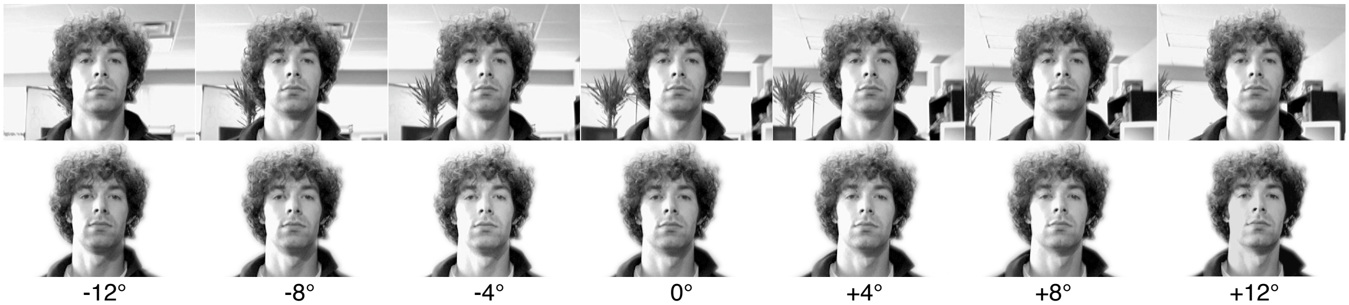

Figure 5. *Stimuli images with and without background generated by cameras positioned at different angles. Images were taken while looking into the respective camera lens.*

EVALUATION

GAZE-2 was designed to support the attentive structure of interlocutors during mediated conversations. As such, the eye controlled camera director is the most critical aspect of this system. We designed an empirical study to evaluate whether the automated camera director is capable of supporting eye contact during camera switches. We also wanted to establish whether angular shifts, particularly in the background of an image, produced any distraction effects. We predicted that distraction would depend on (1) the angular distance between cameras within the video tunnel; (2) the presence of background objects and (3) the size of the image. The angular distance between cameras affects the size of the shift between images. The presence of background objects also affects the size of this shift, since objects located further away from the camera shift more. Finally, we believed image size might affect the perception of background shifts. With large images, the face of the onlooker is more likely to occupy the subject's fovea, with background shifts rendered in peripheral vision.

Experiment Design

Ten volunteers participated in the study. We used a within-subjects design, in which each participant was presented with 32 sequences of images. Each image sequence consisted of two pictures taken from two different camera angles, that were looped at a 1 second interval.

To constitute our first factor, we produced 7 images of a male onlooker, each taken from a different camera angle while looking into the camera lens. Figure 5 shows the different images of the onlooker, taken with 4 degree increments from the center location. Since the onlooker was seated 70 cm away from the camera, each 4 degree shift corresponded to a horizontal shift in camera location of about 5 cm. To produce a +12° shift, the camera was shifted 15 cm to the left of the onlooker, and oriented at a 12 degree angle towards the onlooker. To produce a –12° shift, the camera was shifted 15 cm to the right of the onlooker, and oriented at a 12 degree angle towards the onlooker. We constructed image pairs by combining images from cameras at different angular distances: 2 images pairs with 4 degree shifts; 2 pairs with 8 degree shifts; 2 pairs with 12 degree shifts; 1 pair with a 16 degree shift and 1 pair with a 24 degree shift. Care was taken that shifts were always balanced around the center location, such that the average camera location in all image sequences was at 0 degrees. To constitute our second factor, we duplicated the 8 image pairs and removed the background from all image sequences in the second set. The resulting sequence is shown in the *second* row of Figure 5. To allow for fair comparison between effects of background and foreground shifts, the image of the onlooker in the sequence with backgrounds was made the same for all angles. This sequence is shown in the *first* row of Figure 5. To constitute our third factor, we duplicated and enlarged the resulting 16 image pairs, to constitute a total number of 32 image pairs. The first set of 16 was presented with a size of 10 cm x 7 cm. The second set of 16 was presented with a size of 28 cm x 20 cm. During the experiment, all subjects were seated at a 70 cm distance from the presentation screen. This means that in the large image set, the subject's fovea was completely occupied by the face of the onlooker when looking at the eyes of that onlooker. Images were presented on a Powerbook G4 laptop at full display brightness in a controlled lighting environment. Image sequence presentation order was fully counterbalanced between participants for all factors.

Results

After each image pair, participants were asked to score 5 questionnaire items. The first three measured agreement with a statement using a 7-point Likert scale (from *strongly disagree* to *strongly agree*):

1) This person kept looking straight at me;
2) The images looked the same;
3) This sequence of images was distracting.

The other two questions asked participants to rank on a 7-point scale (from *not noticeable* to *very large*) their perception of the presence of the following:

4) Any color or brightness change between images;
5) Any movement between images.

Effects of Camera Shift Angle

Table 1 shows the results for each factor in our experiment, averaged over 160 trials. Although dependent variables were of an ordinal level of measurement, means and standard deviations are displayed for clarity. All results were tested for significance using non-parametric two-tailed paired tests evaluated at □=0.05. Figure 6 shows the

Condition	Eye Contact	Similarity	Distraction	Color/Brightness	Movement
Large Shift (> 8°)	6.7 *(1.0)*	**3.1*** *(1.7)*	**3.7***** *(1.7)*	**3.1***** *(1.9)*	**4.2***** *(1.6)*
Small Shift (< 8°)	6.7 *(0.9)*	**3.6*** *(2.0)*	**2.9***** *(1.4)*	**2.4***** *(1.5)*	**3.6***** *(1.5)*
Background Shift	**7.0***** *(0.2)*	**2.9***** *(1.7)*	**3.8***** *(1.6)*	**1.7***** *(1.0)*	**4.7***** *(1.4)*
Foreground Shift	**6.4***** *(1.3)*	**3.8***** *(2.0)*	**2.9***** *(1.5)*	**3.7***** *(1.8)*	**3.0***** *(1.4)*
Large Image	6.7 *(1.0)*	**3.2*** *(1.8)*	3.4 *(1.6)*	2.7 *(1.7)*	3.9 *(1.5)*
Small Image	6.7 *(0.9)*	**3.5*** *(2.0)*	3.3 *(1.7)*	2.8 *(1.8)*	3.8 *(1.7)*

Table 1. *Mean scores and std. dev. for 3 experimental factors. Bold pairs indicate significant differences between conditions.*

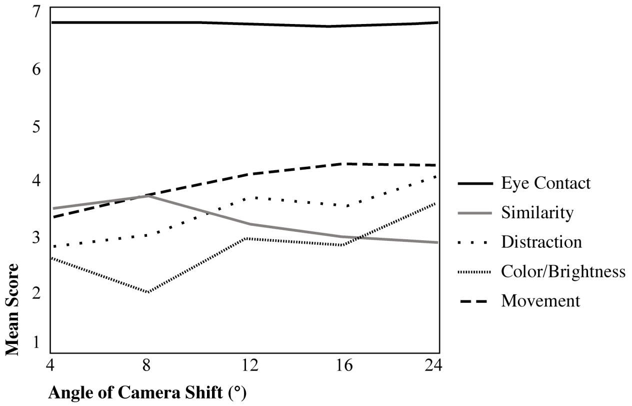

Figure 6. *Mean score for questionnaire items as a function of camera shift angle.*

mean score for each questionnaire item according to the degree of angular shift present in image pairs. The figure shows that most ratings are negatively affected beyond a camera angle of 8 degrees. We used this break point to evaluate the effect of camera angle on our ratings, comparing scores for shifts in camera angle less than 8 degrees with scores for camera angles larger than 8 degrees. The only rating not affected by angular shift was the degree of eye contact, which was consistently scored high at 6.7 (Mann-Whitney Z=-.49, p=0.62)). This is because the looker was instructed to always look into the lens of the camera. As expected, image sequences with large shifts were rated as less similar (3.1) than those with small shifts (3.6), although differences were not large (Mann-Whitney Z=-2.33, p=0.02). Large shifts were ranked more distractive (3.7) than small shifts (2.9) (Mann-Whitney Z=-4.16, p<0.001). Although the degree of distraction is not high in either case, this does suggests shifts are best kept within 8 degrees. Large shifts produced a noticeable change in color or brightness in the image (3.1) as compared to small shifts (2.4) (Mann-Whitney Z=-3.33, p=0.001). This is because differences in camera angle affected the lighting conditions of the onlooker's face. As expected, large shifts were ranked as producing more noticeable movement (4.2) than small shifts (3.6) (Mann-Whitney Z=-3.42, p=0.001).

Effects of Background vs. Foreground Shift

As expected, the presence of a background shift had a significant effect on all ratings. Images with background shifts (7.0) produced higher eye contact rankings than images with only foreground shifts (6.4) (Mann-Whitney (Z=-4.91) p<0.001)). This is because the face was the same for all image pairs with backgrounds, causing no shift in eye contact between images. Images with background shifts were ranked as less similar (2.9) than images with foreground shifts (3.8) (Mann-Whitney Z=-5.77, p<0.001). Background shifts were also ranked more distractive (3.8) than foreground shifts (2.9) (Mann-Whitney Z=-5.48, p<0.001). Although the degree of distraction is again not high in either case, this suggests background removal may be beneficial. However, background shifts hardly produced a noticeable change in color or brightness in the image (1.7) as compared to foreground shifts (3.7) (Mann-Whitney Z=-9.20, p<0.001). The face was more affected by changes in color or brightness than the background, because it was located closer to the light source and camera. As expected, background shifts were ranked as producing more noticeable movement (4.7) than foreground shifts (3.0) (Mann-Whitney Z=-8.23, p<0.001).

Effects of Image Size

Image size produced no significant effect on any of the questionnaire items, except for similarity rankings. Large images were ranked less similar (3.2) than small images (3.5) (Mann-Whitney Z=-2.15, p<0.03). This may be due to the presence of JPEG compression artifacts in the images, which were more noticeable in the large image condition.

DISCUSSION

The animations of camera shifts in this experiment very much represented a worst-case scenario. During multiparty conversations the eyes of an interlocutor tend to shift with speaking turns, which typically take much longer to complete than one second [29]. Results demonstrate that as long as cameras are aligned with images of participants on the screen, eye contact should be maintained during camera switches. The effect of camera shift is most distractive when backgrounds are visible and when the camera shifts are larger than 8 degrees of visual angle. However, the degree of distraction is not high in either case.

We implemented our findings in the design of our attentive video tunnel by reducing the distance between cameras to 10 cm. This decreases the chance of shifts larger than 8 degrees during the rapid iterative looking behavior that occurs when addressing a group [29]. We also implemented a simple background removal algorithm that further mitigates any negative effects of camera shifts.

CONCLUSIONS

We presented GAZE-2, a novel video conferencing system that uses eye-controlled camera direction to ensure parallax-free transmission of eye contact in mediated group conversations. In GAZE-2, multiple cameras are placed in a video tunnel behind on-screen images of participants. Using an eye tracker, the system automatically selects the camera a user looks at for multicast to all participants. This ensures a single parallax-free video stream per user, and scalability of the system. Video images of participants are rendered in a 3D virtual meeting room, where they are automatically oriented to convey whom each user looks at. We evaluated whether eye contact perception is retained during automated camera switching, which causes angular shifts in the transmitted image. Findings suggest such angular shifts do not affect eye contact perception, and are not considered highly distractive.

REFERENCES

1. Argyle, M. and Cook, M. *Gaze and Mutual Gaze*. London: Cambridge University Press, 1976.
2. Argyle, M. and Dean, J. Eye-contact, Distance and Affiliation. *Sociometry* 28, 1965, pp. 289-304.
3. Buxton, W.S., Sellen, A.J., and Sheasby, M.C. Interfaces for Multiparty Videoconferences. In Finn, K.E., Sellen, A.J., and Wilbur, S.B. (Ed.), *Video-Mediated Communication.* Mahwah: Laurence Erlbaum Associates, 1997, pp. 385-400.
4. Chen, M. Leveraging the Asymmetric Sensitivity of Eye Contact for Videoconference. In *Proceedings of CHI 2002.* Minneapolis: ACM Press, 2002, pp. 49-56.
5. Chen, M. *Personal Communication*, 2002.
6. Cherry, C. Some Experiments on the Reception of Speech with One and with Two Ears. *Journal of the Acoustic Society of America* 25, 1953, pp. 975-979.
7. Colburn, A., Cohen, M.F., and Drucker, S.M. *The Role of Eye Gaze in Avatar Mediated Conversational Interfaces*. Microsoft Research Report 2000-81, 2000.
8. Duchowski, A. & McCormick, B. Gaze-Contingent Video Resolution Degradation. In *Human Vision and Electronic Imaging III*. SPIE, 1998.
9. Duchowski, A. *Eye Tracking Methodology: Theory & Practice.* Berlin: Springer-Verlag, 2003, (in press).
10. Duncan, S. Some Signals and Rules for Taking Speaking Turns in Conversations. *Journal of Personality and Social Psychology* 23, 1972.
11. Ericksson, H. MBONE: The Multicast Backbone. *Communications of ACM* 37(8), 1994, pp. 54-60.
12. Garau, M., Slater, M., Bee, S., and Sasse, M. The Impact of Eye Gaze on Communication Using Humanoid Avatars. In *Proceedings of CHI 2001.* Seattle: ACM Press, 2001, pp. 309-316.
13. Gemmell, J., and Zhu, D. Implementing Gaze-Corrected Video Conferencing. In *Proceedings of CIIT 2002.* IASTED, 2002.
14. Gibson, J.J. and Pick, A.D. Perception of Another Person's Looking Behavior. *American Journal of Psychology* 76, 1963, pp. 386-394.
15. Isaacs, E. and Tang, J. What Video Can and Can't Do for Collaboration. In *Proceedings of ACM Multimedia '93.* Anaheim: ACM Press, 1993, pp. 199-206.
16. Jaspars J. et al. Het Observeren van Oogcontact. *Nederlands Tijdschrift voor de Psychologie* 28, 1973.
17. Jerald, J. and Daily, M. Eye Gaze Correction for Videoconferencing. In *Proceedings of ETRA 2002*. New Orleans: ACM Press, 2002, pp. 77-81.
18. Morimoto, C. et al. Pupil Detection and Tracking Using Multiple Light Sources. *Image and Vision Computing* 18, 2000.
19. O'Connaill, B., Whittaker, S., and Wilbur, S. Conversations Over Video Conferences: An Evaluation of the Spoken Aspects of Video-Mediated Communication. *Human Computer Interaction* 8, 1993.
20. Okada, K. et al. Multiparty Videoconferencing at Virtual Social Distance: MAJIC Design. In *Proceedings of CSCW'94.* Chapel Hill: ACM Press, 1994, pp. 385-393.
21. Rosenthal, A.H. *Two-way Television Communication Unit.* US Pat. 2,420,198, 1947.
22. Sellen, A., Buxton, B., and Arnott, J. Using Spatial Cues to Improve Desktop Videoconferencing. In *Proceedings of CHI'92.* Monterey: ACM Press, 1992.
23. Sellen, A.J. Remote Conversations: The Effects of Mediating Talk with Technology. *Human Computer Interaction* 10(4), 1995.
24. Taylor, M. and Rowe, S.. Gaze Communication Using Semantically Consistent Spaces. In *Proceedings of CHI 2000*. The Hague: ACM Press, 2000, pp. 400-407.
25. Vertegaal, R., Van der Veer, G. and Vons, H. Effects of Gaze on Multiparty Mediated Communication. In *Proceedings of Graphics Interface 2000*, 2000, pp.95-102.
26. Vertegaal, R., Weevers, I., and Sohn, C. GAZE-2: An Attentive Video Conferencing System. In Extended Abstracts of CHI 2002. Minneapolis: ACM Press, 2002.
27. Vertegaal, R. and Ding, Y. Explaining Effects of Eye Gaze on Mediated Group Conversations: Amount or Synchronization? In *Proceedings of CSCW 2002*. ACM Press, 2002, pp. 41-48.
28. Vertegaal, R. The GAZE Groupware System: Mediating Joint Attention in Multiparty Communication and Collaboration. In *Proceedings of CHI'99.* Pittsburg: ACM Press, 1999, pp. 294-301.
29. Vertegaal, R., Slagter, R., Van der Veer, G., and Nijholt, A. Eye Gaze Patterns in Conversations: There is More to Conversational Agents Than Meets the Eyes. In *Proceedings of CHI 2001.* Seattle: ACM, 2001, pp. 301-308.
30. Von Cranach, M. and Ellgring, J.H. The Perception of Looking Behaviour. In Von Cranach, M. and Vine, I. (Ed.), *Social Communication and Movement.* London: Academic Press, 1973.
31. Yang, R., Zhang, Z. Eye Gaze Correction with Stereovision for Video-Teleconferencing. *Microsoft Research Technical Report 2001-19*, 2001.