MultiView: Spatially Faithful Group Video Conferencing

David Nguyen
nguyendt@eecs.berkeley.edu

John Canny
jfc@cs.berkeley.edu

Department of Electrical Engineering and Computer Science
University of California, Berkeley
Berkeley, CA 94720-1776

ABSTRACT

MultiView is a new video conferencing system that supports collaboration between remote *groups* of people. MultiView accomplishes this by being *spatially faithful*. As a result, MultiView preserves a myriad of nonverbal cues, including gaze and gesture, in a way that should improve communication. Previous systems fail to support many of these cues because a single camera perspective warps spatial characteristics in group-to-group meetings. In this paper, we present a formal definition of spatial faithfulness. We then apply a metaphor-based design methodology to help us specify and evaluate MultiView's support of spatial faithfulness. We then present results from a low-level user study to measure MultiView's effectiveness at conveying gaze and gesture perception. MultiView is the first practical solution to spatially faithful group-to-group conferencing, one of the most common applications of video conferencing.

Author Keywords

Video Conferencing, Spatial Faithfulness, Gaze, Eye Contact, Deixis

ACM Classification Keywords

H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces—Computer-Supported Cooperative Work.

INTRODUCTION

The goal of any computer-mediated communication system is to enable people to communicate in ways that allow them to effectively accomplish the task at hand. However, most systems do a poor job of preserving non-verbal, spatial, and turn-taking cues that have been shown to be important for group activities [2]. In spite of prior work in video conferencing, MultiView (Figure 1) is the first practical system to support these cues by preserving what we will define as *spatial faithfulness* for the important case of group-to-group meetings, arguably the most common application of video conferencing.

CHI 2005, April 2–7, 2005, Portland, Oregon, USA.

Figure 1. This photograph shows a MultiView site used in our experimental setup.

Spatial faithfulness is the system's ability to preserve spatial relationships between people and objects. Typical video conferencing systems distort these relationships. For example, consider two groups of people using a standard video conferencing system. Because this system uses only one camera at each site, all viewers at the other site see the same view – in effect, they share the same set of eyes. A byproduct of this phenomenon is what is known as the *Mona Lisa Effect* – either everyone or no one feels the remote person is making eye contact with them. MultiView aims to preserve lost spatial information such as this and restore the many cues used in communication, particularly gaze and gesture information. MultiView accomplishes this by providing unique and correct perspectives to each participant by capturing each perspective using one of many cameras and simultaneously projecting each of them onto a directional screen that controls who sees which image.

In addition to being spatially faithful, MultiView has other attractive features. Using available off-the-shelf components allows MultiView to maintain a low cost. Initiating a MultiView meeting is very easy since little setup is required before each meeting after the initial installation of the system. The design of the system affords correct viewing for a finite number of viewing positions at a conference table.

In outline, we begin by defining spatial faithfulness. We then present the metaphor for MultiView and detail its implementation. We then give an overview of the affordances of the system. Finally, we present the results of a user study that measures the perception of nonverbal cues through MultiView – specifically gaze and gesture.

SPATIAL FAITHFULNESS

In this section, we introduce a vocabulary to facilitate a discussion of the capabilities of MultiView. We begin with a discussion of gaze awareness then use it to help define spatial faithfulness.

Defining Gaze Awareness

In analyzing video conferencing systems, it is helpful to characterize the different types of *gaze information* that such systems can support. The literature uses the following definitions widely. Following Monk and Gale [10]:

Mutual Gaze Awareness – knowing whether someone is looking at you. Often times known as "eye contact."

Partial Gaze Awareness – knowing in which direction someone is looking (up, down, left, or right).

Full Gaze Awareness – knowing the current object of someone else's visual attention.

There is a slight ambiguity in the definitions above. For instance, Chen [3] discovered that viewers are less sensitive to an image of an interlocutor looking slightly *below* their eyes than in other directions in perceiving eye contact. So "knowledge of the other's gaze" is subject to ambiguity due to the perception of the viewer.

For practical reasons, most video conferencing systems rely on a camera displaced relative to the image of the remote participant, which leads to an immediate misalignment and loss of spatial faithfulness. A few notable exceptions are described in prior work. Dourish et al. observed that with the initial use of this type of setup, users at first obliged the remote user by looking into the camera, but then re-adapted to looking at their interlocutor's face as their understanding of the visual cues evolved [5].

The above issues demonstrate that a better understanding of the effects of the sensation of eye contact versus the knowledge of eye contact is required. Using the immense size of prior work that try to mitigate the parallax created by a displaced camera in video conference systems design as well as work that show the existence of specialized brain functions for gaze detection [13], we take the stance that it is the sensation that is important. Furthermore, non-verbal communication can function beyond any knowledge of it actually occurring – much non-verbal communication is neither consciously regulated nor consciously received, though its effects are certainly observable [6]. Returning to our definition problem, the above considerations lead to the following re-framing of spatial faithfulness.

Defining Spatial Faithfulness

In this section, we define spatial faithfulness. Our definition emphasizes the perception of nonverbal cues as opposed to knowledge of the intended cues. We use gaze awareness as a starting point in defining spatial faithfulness, but generalize it to include other spatial cues. First, we introduce a simple abstract model.

A Simple Abstract Model

Our model consists of the following objects which act upon attention:

Attention Source – a person who provides attention to the attention target. The method of attention can manifest itself in many different ways including, but not limited to, visual, gestural, positional, directional, etc.

Attention Target – an object (could be a person or anything else) that receives attention from the source.

Observer – the person charged with understanding the presented information about attention – its source, its target, and any attached meaning.

Two common terms used in the gaze research community are *observer* and *looker*. Observer is used in the same way as it is used here, but looker is a special case of an attention source where the type of attention is limited specifically to gaze information. Similarly, we can define a *pointer*, which would be an attention source who uses gesture cues.

Spatial Faithfulness

The definitions below are general terms that can be applied to different types of attention, such as gaze or pointing.

Mutual Spatial Faithfulness – a system is said to be mutually spatially faithful if, when the observer or some part of the observer is the object of interest, (a) it appears to the observer that, when that object is the attention target, it actually is the attention target, (b) it appears to the observer that, when that object is not the attention target, the object actually is not the attention target, and (c) that this is simultaneously true for each participant involved in the meeting.

Partial Spatial Faithfulness – a system is said to be partially spatially faithful if it provides a one-to-one mapping between the apparent direction (up, down, left, or right) of the attention target as seen by the observer and the actual direction of the attention target.

Full Spatial Faithfulness – a system is said to be fully spatially faithful if it provides a one-to-one mapping between the apparent attention target and the actual attention target, whether the target is a person or an object.

The notion of *simultaneity* is important in characterizing video conferencing systems. Consider a dyadic system of two people, X and Y. A system supports mutual gaze awareness if when X makes eye contact with Y it appears to Y that X is indeed making eye contact. *At the same time*, it must also appear to X that Y is making eye contact when that is

the case. Simultaneity can apply to meetings of more than two members.

Group Use of Spatial Information

Gaze has a critical role in group communication. Its functions include turn-taking, eliciting and suppressing communication, monitoring, conveying cognitive activity, and expressing involvement [8]. By removing or distorting gaze perception, we risk adversely affecting the processes of communication that depend on these functions. For instance, Vertegaal et al. [20] found that participants took 25% fewer turns when eye contact was not conveyed in a three-person meeting. However, an arbitrarily added video channel will not necessarily result in better communication. Connell et al. found that audio alone may be, in fact, preferable in routine business communication [4]. Bos et al. measured the effects of four different mediated channels – face-to-face, text, audio only, and video and audio – on trust building [2]. They found that adding video did not significantly contribute to trust building over audio-only channels in people who have not met face-to-face. Furthermore, Short et al. [17] notes that a video channel may actually further disrupt some communication processes. For instance, the lack of mutual eye contact can lead one participant to feel like she is making eye contact with a remote participant when the other does not. Argyle et al. [1] found that such asymmetries lead to noticeable increases in pause length and interruptions.

Another important cue that heavily depends on spatial information is gesture. Collocated groups in an office environment often point and gesture toward spaces where ideas were formulated and discussed as if that particular space is a marker of knowledge. Groups of people may also use gesture to measure and regulate understanding. Standard video conferencing systems often distort or destroy these gesture cues. In particular, group-to-group systems with one camera per site will necessarily distort gesture for the same reasons that they distort gaze (Mona-Lisa effect).

PRIOR WORK

Hydra [16] supports multi-party conferencing by providing a camera/display surrogate that occupies the space that would otherwise be occupied by a single remote participant. Because of the scale and setup of a Hydra site, there is still a noticeable discrepancy between the camera and the image of the eyes, resulting in the same lack of support for mutual gaze awareness that standard desktop setups have. Hydra does add an element of mutual spatial faithfulness in that it appears to an observer that she is being looked at when she is indeed the attention target and not being looked at when she is not the attention target in group meetings.

GAZE-2 [21] is another system developed to support gaze awareness in group video conferencing. GAZE-2 uses an eye tracking system that selects from an array of cameras the one the participant is looking directly at to capture a frontal facial view. This view is presented to the remote user that the participant is looking at, so that these two experience realistic eye contact. However, views of other participants are synthesized by rotating the *planar frontal* views of those other participants. Because of the Mona Lisa effect, even significant rotations of frontal views will still be perceived as frontal ones, while a side view of those participants is what is desired. To mitigate this, GAZE-2 uses extreme rotations (70 degrees or more) of these other views, and attaches them to a 3D box to create a spatial perception that overwhelms the perception of the face itself. This distortion is not spatially faithful, and there is no attempt to preserve gesture or relations with objects in the space.

MAJIC [11] produces a parallax-free image by placing cameras behind the image of the eyes using a semi-transparent screen. MAJIC supports mutual, partial, and full spatial faithfulness since the images are free of parallax, so long as there is only one participant at each site since they employ single view displays.

An extreme approach to preserving spatiality is to use a mobile robotic avatar or PRoP (Personal Roving Presence) as a proxy for a single remote user [12]. PRoPs suffer from a Mona-Lisa effect at both ends, but are not intended for group-to-group interaction. At the robot end, they mitigate the effect by using the robot's body and camera as a gaze cue (rather like GAZE-2's virtual monitors). When multiple users operate PRoPs in a shared physical space, full spatial faithfulness is preserved.

All the above systems claim to support multi-site meetings. A striking limitation on all these systems, however, is that they only work correctly and provide their claimed affordances when used with *one participant per site*. This will be a problem with any system based on viewer-independent displays. In real physical space, different users *do not share* the same view of others. MultiView provides a practical solution to this problem, using a custom view-dependent display.

A DESIGN FOR SPATIAL FAITHFULNESS

We start with a "virtual conference room" which contains a large conference table as per Figure 2. Two groups of people sit on opposite sides of the table. The spatiality of the room is visually coherent – all members on one side of the table see the entirety of the other side as if the glass pane is not there. This allows visual communication to occur naturally since it supports all the visual cues that are typically present in face-to-face meetings: stereo vision, unique perspectives depending on position, life size, high resolution, appropriate brightness, etc. This, in turn, supports nonverbal cues including gaze and gesture. This environment is mutually, partially, and fully spatially faithful.

Implementation

Our goal is to realize the spirit of the metaphor with groups in two different locations. Figure 3 is a diagram of a two site implementation of MultiView with three participants at each site. The display screen lies in the plane of the remote participants. The display is designed in such a way that when multiple projectors project onto it at once, each image will only be seen by a person who is in the viewing zone for that projector. In Figure 3, person 'L' will only see the image

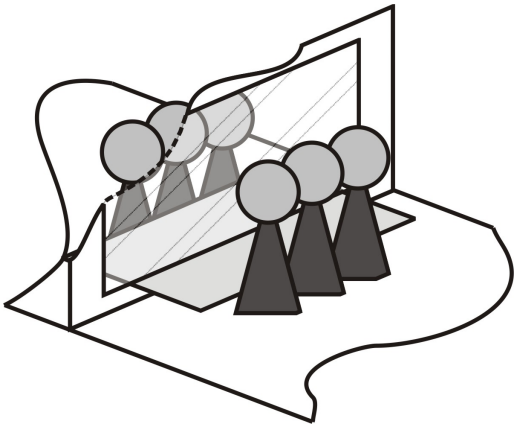

Figure 2. An illustration of the metaphor used by MultiView. A conference table with two groups on people on either side.

produced by projector 'L', person 'C' will only see the image produced by projector 'C', and person 'R' will only see the image produced by projector 'R'. They can all view their respective images simultaneously. The critical feature of the design for full spatial faithfulness is that the cameras must be at the exact position of the remote participants' virtual images. The screens can actually be moved forward or backward of this plane and the images scaled appropriately. The cameras then accurately capture what a remote participant would see if they were physically at the location of their virtual image. The design of the MultiView screen is discussed in a later section. We used BenQ PB2120 projectors with resolutions of 800x600 pixels.

The simplest realization of Multiview is to place the cameras on top of the viewing screen. In that case, the projected images should be life-sized and the virtual images should be in the plane of the screen (i.e. no lateral disparity between images from different projectors). Each camera is placed directly above the image of the remote person and is centered on the middle of the viewing area (or the middle participant if there is one). Each camera is connected to the corresponding projector at the other site.

The sites do not necessarily have to have the same number of camera/projector pairs. The top site in figure 3 is illustrated to support up to five viewing zones and output video streams while the bottom site only supports three of each. In addition, no special configuration is needed if fewer than the supported number of participants are present – the seats are simply left empty as denoted by the dotted cameras/projectors.

This setup introduces the parallax issues seen in desktop video-conferencing systems, since the positions of the cameras are above the position of the eyes of the image. However, because of the scale of the system, we can leverage Chen's findings that show an asymmetry in a person's sensitivity to eye contact [3]. He found that people would still perceive eye contact if the eyes are less than 5° below the camera. Because of the scale of the system and the distance

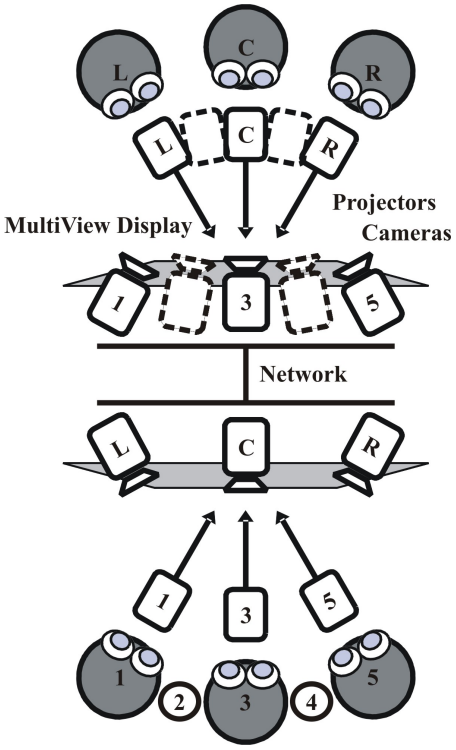

Figure 3. A diagram of MultiView.

the viewers sit from the screen, the parallax is still small enough to provide the sensation of eye contact. In our setup, the average angle was about 3°.

A problem we ran into during early configuration was determining the height of the screen. Our first attempt put the bottom of the screen at the level of the tabletops so that group members could look straight ahead. Since the cameras were on top of the screens, the camera's aim was excessively downward and produced a "bird's eye" view and a large disparity between the camera and the image of the remote viewer's eyes. A better approach was to fix the cameras at a height slightly above eye level and allow the screen to hang below. We didnt need the lower part of the screen, which showed that this type of setup prefers a wider than normal screen aspect ratio (more like 2:1 than 4:3).

Designing the MultiView Directional Screen

The MultiView screen's main function is to display the image produced by a projector only to a person in a very specific viewing zone. Conventional screens will diffuse an image so that it is visible from a wide range of angles and only support a single large viewing zone. MultiView's screen carefully controls diffusion and produces relatively narrow viewing zones above, below, and slightly to the side of a light source. The viewing zones are roughly vertical "pie slices" centered on the middle of the screen. Therefore, a person looking over the top of a projector sees only the image from that projector. This is simultaneously true for all projectors.

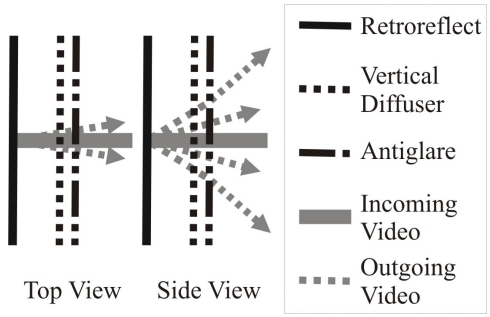

Figure 4. The multiple layers of the MultiView screen.

The MultiView display uses multiple layers to create its viewing zones. A diagram of layers is shown in Figure 4. The back-most layer is a retroreflective cloth. An ideal retroreflective material bounces all of the light back to its source ($\theta_r = \theta_i$). This differs from an ideal mirror where the light bounces along the reflective path ($\theta_r = -\theta_i$). Additionally, materials can exhibit properties of a Lambertian surface that, ideally, diffuses light in all directions equally. A practical retroreflective material exhibits all three properties – given a source of light, some of the light bounces back to the source, some of the light gets reflected along the reflective path, and light gets diffused by a small angle along both the retroreflected and reflected paths. The 3M 8910 fabric was used for two reasons: 1) it had a strong retroreflective characteristic, and 2) because of its exposed lens design, it has minimal reflective properties and good diffusive properties to reduce glare effects.

The next layer is a one-dimensional diffuser which extends the viewing zone for one projector to a vertical "slice". Without it, the image would only be visible directly on the projection axis. This is problematic because if you were in front of the projector, you would block the projected image, and if you were behind it, the projector would block your view. In our implementation, we used a lenticular sheet as the diffuser[1]. A spacing of 1/4" or more between retroreflector and lenticular sheet is recommended, otherwise the diffusion effects of the lenticular will be undone by the retroreflector (outgoing and incoming rays will be close relative to lenticular spacing). It is possible to reduce this spacing if needed by using a lenticular sheet with finer pitch e.g. 80 LPI or greater.

The last layer is an anti-glare layer. The high gloss finish of the lenticular sheets produced a very distracting glare along the path of reflection. As a result, we applied an anti-glare film produced by DuPont (HEA2000 Gloss 110). To apply the film, the smooth side of the lenticular sheet must face the viewer and projectors.

[1]Note: Lenticular sheets are often used for directional displays or for multiple image merging or separation and have been used in this way in previous spatial displays. This often confuses readers trying to understand MultiView. In our application we are *not* using the lenticular sheet as a lenticular imager, but simply as a directional diffuser. Any other diffuser could be used, but others are currently much more expensive.

Qty	Item	Cost/Unit	Total Cost
1	Retroreflective Sheet	$50.00	$50.00
1	Lenticular Sheet	$50.00	$50.00
1	Anti-Glare Layer	$600.00	$600.00
3	Camera + Lenses	$100.00	$300.00
3	Projectors	$900.00	$2700.00
		Total	$3700.00

Table 1. Cost for a three person MultiView site.

The result is a screen that is capable of showing multiple unique views to different viewing zones in space. With proper alignment, those particular views can simulate the perspective of actually being there. Figure 5 shows photos taken from positions 1 and 3 (see Figure 3) of the same display and of the same people asked to look at position 5.

Cost

One of the benefits of MultiView is its relatively low cost and potentially high gain. The cost to build a single three-person site is shown in Table 1. The cost presented does not include hardware to transport the video from one site to another or other miscellaneous building hardware. Clearly, the projectors account for most of the cost. Projectors, like computers, have a history of decreasing cost and increasing picture quality. Recently, projectors fell below $1,000 and continue to decrease in cost. In addition, they are becoming smaller and consuming less power, which, as we will see, present some very interesting scenarios.

Setup

Each projector must be positioned correctly to present the view of a remote camera; however, the alignment step is straightforward. Each camera is set permanently at a certain view angle when it is attached to the viewing screen. For the screen+cameras at site A, assume a list of view angles is saved in a file at site A. This configuration should never need to be changed, as long as enough cameras are used to support the largest anticipated conference. To determine the correct projector placement at site B for a conference with site A, the site A camera file is first downloaded at site B. Then video from the *site B* center camera is fed back to any display (including the projector being set up) at site B. On this display, red vertical lines are rendered that show the angles of all possible *remote* camera views (using the site A camera data file), and these lines are superimposed on the local (site B) view of table and projectors. The projector can be moved left or right until it aligns with one of the red lines in the local view. Once it does, the site B projector is switched to the video feed from the corresponding site A camera, and it will faithfully reproduce the view from that angle. This setup process takes only a few seconds, which is important if the system is to be used with varying numbers of participants or with a stowable display screen.

A Three Site Implementation

The current implementation of MultiView supports group-to-group, two site meetings. However, it is possible to extend MultiView to support more than two sites. Figure 6 il-

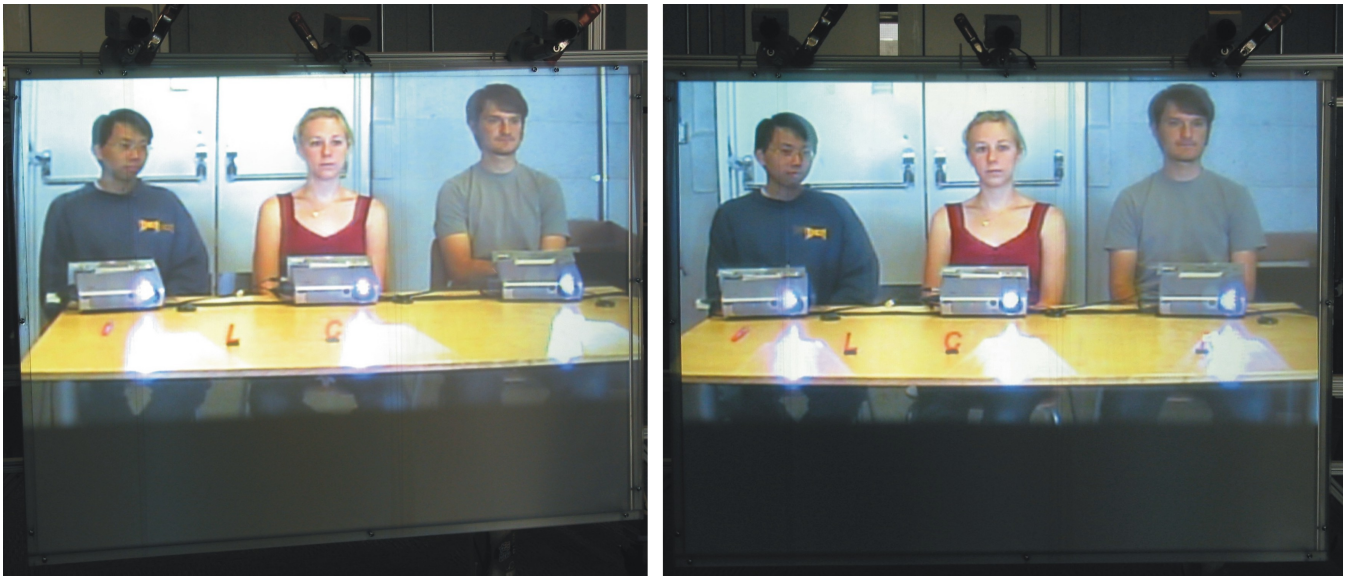

Figure 5. Two different photos of the same display and scene from two different perspectives. The left photo was taken from position 1, the right photo was taken from position 3. Everyone in this photo was looking at position 5. See Figure 3 for positions.

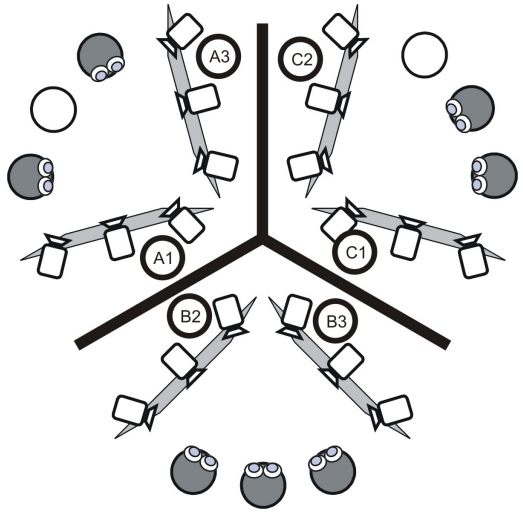

Figure 6. A three site setup of MultiView. Each site can support up to three participants but sites A and C are not fully populated.

lustrates a three site setup supporting multiple people. In the three site configuration, the A1 cameras are projected onto screen B2, the A3 cameras on screen C2, the B2 cameras on screen A1, the B3 cameras on screen C1, C1 cameras on screen B3, and the C2 cameras on screen A3. With a wide enough throw and some image shape correction, one projector could be used to project images to both screens at a site. This preserves mutual, partial, and full spatial awareness across all sites – a person at site A would be able to determine the attention target of a person at site B even if the attention target was at site C. Notice that in this illustration, not all the seats are filled.

AFFORDANCES OF MULTIVIEW

We list some of the affordances of the MultiView system that are relevant to video conferencing systems.

Multi-Modal Cues: As with face-to-face, MultiView can support multiple types of cues concurrently. During calibration, a person setting up the system was able to look at someone at the remote site and point in a direction to say tacitly, "Hey you, go that way." He was able to use two non-verbal deictic cues – gaze to identify the person and hand gestures to identify the direction he wanted them to move – at the same time. No verbal communication is required.

Life-Size Images: Reeves and Nass have shown that the size of a display can affect the levels of cognitive arousal and we wished to preserve this effect [14]. Many common systems use typical computer monitors to display the video stream and, oftentimes, the image itself is only a fraction of the screen. GAZE-2 [21] uses the entirety of the monitor's real estate, but the actual images of people are quite small. The rest of the monitor space is required for recreating a sense of spatial relations among the participants. Hydra [16] uses small LCD panels as a display. In MultiView, the entirety of the display is used.

Wide Field of View: The view that each group member receives is a single, coherent, wide view. This allows them to use any object or person as an attention target. This differs from most previous video conferencing systems that favor head and shoulders perspectives. In ClearBoard [7], remote participants share an electronic white board. This supports full gaze awareness of graphics on the whiteboard, but not for other objects in the space.

High Resolution Video Streams: The resolution of MultiView is limited by the capacity of the cameras and projectors used. Several current multiple-view display systems use a single display and filter method [15] or a lenticular separation method [9] to produce different views. These methods divide the resolution of a display among multiple views so that each view has only N/K pixels, where N is the pixels of the full display and K is the number of views. MultiView supports K full-resolution views. MultiView uses CCTV cameras capable of capturing 420 lines of resolution and projectors capable of projecting 800x600 pixels. Therefore, the cameras used in our current implementation set the image quality limit.

EVALUATION

Our evaluation involved a user study to 1) demonstrate its ability to naturally represent gaze and gesture information to the viewer, 2) characterize the accuracy of our implementation, and 3) get feedback to guide future possible user studies. The primary goal of experiment 1 is to demonstrate MultiView's support of *partial and full spatial awareness with respect to gaze* for all participants simultaneously in the meeting. The primary goal of experiment 2 is like that of experiment 1, except we test with respect to *gesture*. The primary goal of experiment 3 is to test MultiView's support for *mutual spatial awareness with respect to gaze* (or mutual gaze awareness) for all pairs simultaneously.

Participants

Seven groups of three and one group of two were used for testing. Overall, 23 participants took part in our user study. They were recruited from the undergraduate and graduate student population at University of California, Berkeley. Each participant was paid $10 upon completion of the experiment. In addition to the participants, a set of researchers were recruited from our lab to provide the visual stimuli in our experiments. There were six researchers used in sets of three. The makeup of the researcher group for each session was determined by availability.

Experimental Setup

In all three of our experiments, we used the MultiView setup shown in Figures 1 and 3. Everyone sat approximately 12' from the screen. Because of available materials, we used a less-than life-sized (48"x36") viewing screen. The screen image was scaled by 2/3 to fit the image of all three participants on the screen. This scaling puts the virtual participants a distance *behind* the plane of the screen making the total effective distance to the remote participants was 18'. Since we kept the cameras on top of the display screen our setup had a slight distortion from full spatial faithfulness. A fully accurate setup would have required either life-size images with the camera on top of the screen, or cameras set behind the screen corresponding to the location of the actual participants. We did the experiment with this caveat. Later we discuss other methods of overcoming it.

Each participant was about 25" from his or her neighbor. On the screen, each person was about 16" apart. At one site, three researchers – designated as L, C, and R for left, center, and right – were asked to provide the visual stimulus. These positions were marked with standing acrylic letters. At the other site, small acrylic numbers – 1, 2, 3, 4, and 5 – marked five positions on the conference table. Each position designated an attention target for our experiment. There was about 8" of separation between each attention target on the remote screen. Participants in the study sat behind 1, 3, or 5. At the end of each study, comments were solicited to provide insight into the results and feedback about design improvements.

Viewing Position	μ	σ
1	0.70	0.65
3	0.63	0.67
5	0.60	0.70
Combined	0.64	0.68

Table 2. The mean (μ) and standard deviation (σ) of error by gaze direction perception by viewing position.

Experiment 1

Task

In experiment 1, each researcher was instructed to look at one of the 5 positions. The positions were randomly generated prior to each session of the experiment and provided to each researcher on a sheet of paper. If the position happened to have a participant in it (positions 1, 3, and 5), they were instructed to look into the image of the person's eyes on the screen. If the position was in between two participants (positions 2 and 4), they were asked to look at that position, but at the average eye level of the participants. The participants were then asked to record which position each researcher appeared to be looking at on a multiple choice answer sheet. They were carefully instructed to avoid trying to determine which target they felt like the researcher *actually* was looking at, but to instead concentrate on which target the image of the researcher *appeared* to be looking at. This process was repeated 10 times.

Results

The results of experiment 1 are presented in different ways that are relevant to the discussion that follows. The primary measurement in our results is the error in perceiving the attention target. We define error of any given stimulus i (ϵ_i) to be the difference between what the observer perceived to be the attention target of the image (t_{pi}) and the actual attention target of the researcher producing the gaze stimulus (t_{ai}):

$$\epsilon_i = |t_{pi} - t_{ai}|$$

Table 2 presents the mean and standard deviation of error by the observer's viewing position. For instance, the mean error for observers sitting at position 1 is 0.70. An analysis of variance shows that viewing position had no significant effect on mean error, $F(2, 687) = 1.48, p = 0.23$. This is to be expected, in fact, it is a validation of the Mona Lisa principle – the principle implies that perceived view is not affected by viewer angle relative to a screen.

Table 3 presents the mean and standard deviation of error by the target of the gaze stimuli. For instance, the mean

Gaze Target	μ	σ
1	0.28	0.63
2	0.79	0.67
3	0.68	0.71
4	0.73	0.62
5	0.43	0.65

Table 3. The mean error (μ) and standard deviation (σ) in perceived gaze direction for each set of stimuli directed at each target in experiment 1.

Stimulus Target

Perceived Target	1	2	3	4	5
1	78.5	24.2	10.6	1.0	0.0
2	18.5	33.9	18.8	6.2	0.9
3	0.0	30.3	46.3	24.9	5.6
4	3.1	10.3	20.6	35.8	29.0
5	0.0	1.2	3.8	32.1	64.5

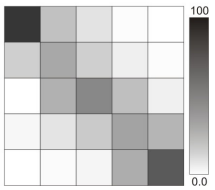

Figure 7. The confusion matrix for experiment 1. Each column represents the actual target of the gaze stimulus and each row represents the target as perceived by the participants. The confusion matrix is represented textually on the left and graphically on the right.

error of responses to all stimuli targeted at position 2 is 0.79. The Tukey HSD procedure showed significant differences in any pairing between stimuli whose target was 2, 3, or 4 and stimuli whose target was 1 or 5. There is no significant difference for any other pairing.

Figure 7 presents the results in the form of a confusion matrix. Each column represents the actual target of the gaze stimulus and each row represents the target as perceived by the participant given the gaze stimulus. For example, for all gaze stimuli directed at position 3 (column 3), 10.6% of the responses perceived that the gazer was looking at position 1, 18.8% at position 2, 46.3% at position 3, 20.6% at position 4, and 3.8% at position 5.

Discussion

Referring back to Figure 3, we consider the seventh trial of our third session. Researcher L is instructed to look at target 1, Researcher C at target 1, and Researcher R at target 5. All the participants, mindful of being asked to record where they think the *image* of the researcher is looking, respond correctly for each researcher. If this trial were reproduced using a standard single view setup, with the camera positioned at the center of the screen (correlating to position 3), then the observer sitting at position 1 would feel as though Researchers L and C were looking to her left (beyond available targets) and Researcher R at position 3. An observer at position 5 would also have these sorts of distortions. The only one with the correct perspective would be the observer at position 3 since the position of the remote camera correlates to that person's perspective.

Viewing Position	μ	σ
1	0.55	0.61
3	0.53	0.60
5	0.65	0.67
Combined	0.58	0.63

Table 4. The mean (μ) and standard deviation (σ) of error by gesture direction perception by viewing position.

Gaze Target	μ	σ
1	0.23	0.46
2	0.65	0.55
3	0.59	0.61
4	0.76	0.69
5	0.55	0.71

Table 5. The mean error (μ) and standard deviation (σ) in perceived gesture direction for each set of stimuli directed at each target.

The position of the observer had no significant effect on the mean error. Observers were often able to respond to a stimulus in a matter of a second. The mean error in determining the direction of a person's gaze was 0.64. The rather low accuracy is probably due to the large distance between the two sets of participants, discussed later.

The two end positions, 1 and 5, enjoyed a significantly lower mean error than the interior positions, 2-4. From the comments gathered during the experiment, it seems that this is due to a self-calibration phenomenon resulting from the setup of the experiment. The participants were aware that the target set consisted of only five positions, and quickly learned what the images looked like when looking at the end positions. Comments like "I thought the last one was a 5, but it wasn't because this time she's looking even more to the right," were common.

Experiment 2

Task

Experiment 2 is similar to experiment 1, except that instead of gazing at each of the positions, the researchers were asked to point in the direction of the position. This process was repeated ten times.

Results

The results found in experiment 2 were very similar to those found in experiment 1. They are summarized in Tables 4 and 5 and Figure 8 without further discussion.

Experiment 3

Task

In experiment 3, participants and researchers were paired off. The researchers were asked to gaze at points on the screen relative to their participant partner's eyes. They were asked to look at one of the following: above the camera, at the camera, at the participant's eyes, below the eyes, slightly past the right of the eyes, or slightly past the left of the eyes. The targets were randomly generated before each session

Perceived Target \ Stimulus Target	1	2	3	4	5
1	78.9	35.7	4.1	2.7	0.0
2	19.3	38.4	23.4	6.8	1.3
3	1.8	22.2	47.4	29.9	9.0
4	0.0	3.8	22.8	35.4	33.3
5	0.0	0.0	2.3	25.2	56.4

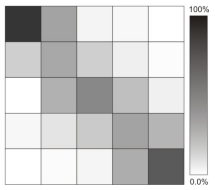

Figure 8. The confusion matrix for experiment 2. Each column represents the actual target of the gesture stimulus and each row represents the target as perceived by the participants. The confusion matrix is represented textually on the left and graphically on the right.

Gaze Direction	Total	Yes	No	Rate
Above Cam	100	54	46	54.0%
At Cam	132	91	41	68.9%
At Eyes	127	81	46	63.8%
Below Eyes	136	76	60	55.9%
Left of Eyes	123	74	49	60.2%
Right of Eyes	72	37	35	51.4%

Table 6. The responses of the participants based on the direction of gaze in experiment 3.

of the experiment. Each participant was asked, "Do you feel as though the researcher is looking directly into your eyes?" After 10 trials, participants and researchers switched partners. This process was repeated until all pairs were exhausted.

Results

A summary of the results from this experiment are given in Table 6. The first column ("Gaze Direction") describes the direction of the gaze. The second column ("Total") is the total number of stimuli presented in that direction. The third column ("Yes") is the number of times a participant replied positively as to whether or not they felt the researcher was looking directly into their eyes. The fourth column ("No") is the number of times a participant replied negatively to that same question. The fifth column ("%Rate") is the rate at which the participants answered positively.

Discussion

This experiment was designed to provide more precise characterization of MultiView's support for mutual gaze awareness. Our expectation was that participants would answer "yes" near 100% of the time when gaze was directed at the camera. However, we see that the rate for this case was actually at 68.9%. In addition, there is little difference between the rates of perceived eye contact between each gaze direction. When asked for comments at the end of the experiment, it was repeatedly mentioned that it was difficult to make out the exact position of the pupil.

However, the participants also mentioned that they had a strong sensation of eye contact during impromptu conversations with researchers between experiments. They felt like the entire *context* of the conversation, combined with the visual information, provided a strong sensation of eye contact even with the limited ability to determine pupil position.

This highlights a separation between the ability to determine the position of a pupil and the sense of eye contact. In [13], Perrett describes the existence of a *direction-of-attention detector* (DAD), which is a specialized brain function used to determine the attention target. His theory suggests that, though the eyes are the primary source of information, the DAD can come to depend more on other cues such as head orientation and body position when the eyes are viewed from a distance or otherwise imperceptible, as is the case with MultiView. The task we presented to our participants required them to judge pupil direction, but the differences between the images of two different gaze points were apparently imperceptible.

FUTURE WORK

Design Lessons Learned: From the results of experiment 3, it is clear that the image quality could be improved in order to help gaze estimation accuracy. Three improvements can be implemented straightforwardly. First, since the current cameras are limiting image quality, higher quality cameras can be used without significantly adding to overall cost. Secondly, the screen size should be larger to eliminate the need for scaling, and to preserve spatial faithfulness with the cameras placed on top of the screen. Thirdly, we can reduce the distance the participants sit from the screen, which was 12' in the study. This was set by the throw distance of the projectors we used. Taking into account the image scaling, the participants were sitting at an equivalent distance of about 18'. This is a very large "virtual conference table", and it is perhaps not surprising that participants had some difficulty determining remote participants' gaze direction. Fortunately, inexpensive short-throw projectors are available, and some feature a 16:9 image form which is a better fit to our application. Current low cost projectors can produce a full-width image (72"Wx45"H) at a throw distance of only 8'. Combined with an appropriate screen, this would allow intimate meetings with an effective participant separations of only 8'.

We expect much more accurate gaze estimation at a virtual distance of 8'. First, remote participants see the local participants more closely, and the angular changes in their gaze will be two times larger. These magnified changes will be rendered on a screen that is two times larger in visual angle to the *local* participants. These effects are multiplicative in terms of the viewer's retinal perception of gaze displacement (4x), which should give much better gaze estimation.

Higher Level User Tests: In the previous experiments, we presented a low level, perception-based user study. The participants were simply asked if they perceived some visual phenomenon provided by a spatially faithful system. As experiment 3 demonstrates, perception of the stimuli we measured provides only a hint of the sensations preserved by nonverbal cues through MultiView. In future user testing, we would like to determine whether or not a spatially faithful system like MultiView affects the way people work together

on a variety of tasks. For instance, Bos et al. [2] measured differences in a trust building exercise using a variant of the prisoner's dilemma.

Personalized views: Three evolving technologies will make MultiView match its metaphor even better. The first is the development of micro-projectors [19]. This new breed of projectors are predicted to scale down to the size of matchbooks and use a fraction of the energy required by current projectors. Though they are predicted to produce lower light levels, they are an ideal match for MultiView because the high gain (directional) screens concentrate the brightness back to the viewer. The second is the development of algorithms for *synthetic* video views that interpolate a set of fixed cameras [18]. The third are a set of tracking technologies. With these technologies, every person could have a microprojector embedded into their laptop. They can all walk into a conference room and sit wherever they wish. The tracking system would automatically figure out their position and synthesize, *exactly*, the appropriate view for that observer, even if the observer decides to move around.

CONCLUSION

We developed MultiView in order to give remote groups of people the advantages of meeting face-to-face without the disadvantages of traveling. We approached this goal by designing a system that concentrates on the broader goal of spatial faithfulness versus just eye contact alone. In this paper, we defined spatial faithfulness and concentrated specifically on its gaze and gesture aspects. We then proposed a spatially faithful metaphor of a large conference table. Based on this metaphor, we presented the design of MultiView, a multiparty video conferencing system capable of supporting multiple people at each site. Evaluating MultiView consisted of 1) analyzing metaphor matches and mismatches, and 2) performing a low-level user study that demonstrates MultiView's support for mutual, partial, and full spatial faithfulness.

REFERENCES

1. Argyle, M., Lalljee, M., and Cook, M. The effects of visibility on interaction in a dyad. *Human Relations, 21*, (1968), 3–17.
2. Bos, N., Olson, J., Gergle, D., Olson, G., and Wright, Z. Effects of four computer-mediated communications channels on trust development. *Proc. CHI 2002*, ACM Press (2002), 135-140.
3. Chen, M. Leveraging the asymmetric sensitivity of eye contact for videoconference. *Proc. CHI 2002*, ACM Press (2002), 49-56.
4. Connell, J. and Mendelsohn, J. and Robins, R. and Canny, J. Dont hang up on the phone, yet! *ACM GROUP (Conf. on Group support)*, Sept 2001, 117-124.
5. Dourish, P., Adler, A., Bellotti, V., and Henderson, A. Your place or mine? learning from long-term use of audio-video communication. *Computer-Supported Cooperative Work 5*, 1 (1996), 33-62.
6. Ekman, P. Telling Lies: Clues to Deceit in the Marketplace, Marriage, and Politics, *Third edition*, W.W. Norton, 2002
7. Ishii, H., and Kobayashi, M. Clearboard: a seamless medium for shared drawing and conversation with eye contact. *Proc. CHI 1992*, ACM Press (1992), 525-532.
8. Kendon, A. Some functions of gaze-direction in social interaction. *Acta Psychologica 26* (1967), 22-63.
9. Lipton, L., and Feldman, M. A new autostereoscopic display techology: The synthagram. http://www.stereographics.com/.
10. Monk, A., and Gale, C. A look is worth a thousand words: Full gaze awareness in video-mediated conversation. *Discourse Processes 33*, 3 (2002), 257-278.
11. Okada, K., Maeda, F., Ichikawaa, Y., and Matsushita, Y. Multiparty videoconferencing at virtual social distance: MAJIC design. *Proc. CSCW 1994*, ACM Press (1994), 385–393.
12. Paulos, E. and Canny, J. Prop: Personal roving presence. *ACM SIGCHI*, 1998. Los Angeles, pp 296-303.
13. Perrett, D. I. Organization and Functions of Cells Responsive to Faces in the Temporal Cortex. *Phil. Trans.: Biological Sciences 335*, 1274 (1992), 23–30.
14. Reeves, B., and Nass, C. *The media equation: how people treat computers, television, and new media like real people and places.* Cambridge University Press, 1998.
15. Schmidt, A., and Grasnick, A. MultiViewpoint autostereoscopic dispays from 4d-vision gmbh. *Stereoscopic Displays and Virtual Reality Systems IX 4660*, 1 (2002), 212-221.
16. Sellen, A., Buxton, B., and Arnott, J. Using spatial cues to improve videoconferencing. *Proc. CHI 1992*, ACM Press (1992), 651-652.
17. Short, J., Williams, E., and Christie, B. *The social psychology of telecommunications*. Wiley & Sons, London, 1967.
18. Slabaugh, G. G., Culbertson, W. B., Malzbender, T., Stevens, M. R., and Schafer, R. W. Methods for volumetric reconstruction of visual scenes. *Int. J. Comput. Vision 57*, 3 (2004), 179-199.
19. Upstream Engineering, Inc. http://www.upstream.fi/.
20. Vertegaal, R., Van der Veer, G. and Vons, H. Effects of Gaze on Multiparty Mediated Communication. *Proc. Graphics Interface 2000*, (2000), 95–102.
21. Vertegaal, R., Weevers, I., Sohn, C., and Cheung, C. Gaze-2: conveying eye contact in group video conferencing using eye-controlled camera direction. *Proc. CHI 2003*, ACM Press (2003), 521-528.