MultiView: Spatially Faithful Group Video Conferencing

David Nguyen

May 23, 2005

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 – such as 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. 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. As part of the MultiView design, we introduce a new design for a new multiple-viewpoint display. 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 for spatially faithful group-to-group conferencing, one of the most common applications of video conferencing.

Contents

Figure 1: This photograph shows a MultiView site which supports up to three participants and can provide three perspectives. Notice MultiView consists of multiple cameras, multiple projectors, and a specially designed screen. Two setups like this were used in our evaluation.

1 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 the non-verbal, spatial, and turn-taking cues that many important group tasks depend on [2]. In spite of much 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.

Spatial faithfulness is a 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 the remote viewers at a 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 like the remote person is making eye contact with them (see Figure 2 for a demonstration). 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 what 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 or joining one that's currently in session (hot-joining) is very easy since little setup is required after the initial installation of the system. The design of MultiView 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 MultiView. Finally, we present the results of a user study that measures the perception of nonverbal cues through MultiView – specifically gaze and gesture.

2 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 extend it to define spatial faithfulness.

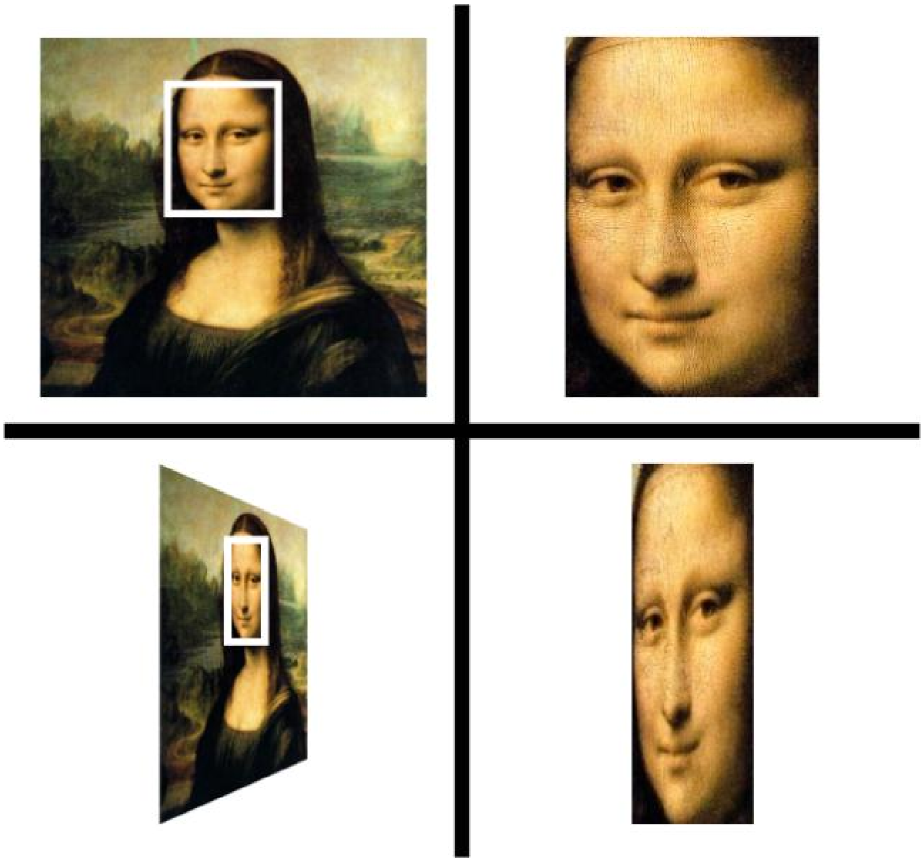


Figure 2: A demonstration of the Mona Lisa Effect. On the top left corner is a frontal image of the Mona Lisa. On the top right is a closeup of the face from this perspective. Notice how Mona Lisa seems to be looking directly at you. On the bottom left corner is an image of the Mona Lisa rotated as if it were viewed from 50° to your left. On the bottom right is a closeup of her face from this perspective. From the closeup, there is still a strong perception that Mona Lisa is still looking directly at you because all the gaze direction stimuli are maintained. A tempting alternative interpretation may be to say that Mona Lisa is looking forward, and because of the rotation, she must be looking in the direction of the rotation. However, humans are good at interpreting projective transformations, thus people tend to correct the warped image, and the sensation that Mona Lisa is still looking at you dominates [23].

2.1 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 [11]:

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 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 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]. Once this level of understanding is achieved, the interlocutors can accurately judge whether their partner is engaging in eye contact or not. Can this system now be classified as supporting mutual gaze awareness? or is the actual sensation of eye contact required?

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 tries 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 [14], 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]. Next, we consider the definitions of gaze awareness and the the above considerations to define spatial faithfulness.

2.2 Defining Spatial Faithfulness

In this section, we define spatial faithfulness. Our definition emphasizes the sensation 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.

2.2.1 A Simple Abstract Model

Our simple model is based on attention and 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 thing) 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 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 the gestural cue of pointing.

2.2.2 Spatial Faithfulness

The definitions below are general terms that can be applied to all types of attention cues, including gaze or gesture.

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.

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, then 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 if that is the case. Simultaneity can apply to meetings of more than two members.

A system can be spatially faithful with respect to a certain type of attention. Most common are systems that explicitly support some level of spatial faithfulness for only gaze and not gesture. This is true for GAZE-2 [21].

2.3 Group Use of Spatial Information

2.3.1 Gaze

Gaze has a critical role in group communication. According to Kendon [9], its functions include turn-taking, eliciting and suppressing communication, monitoring, conveying cognitive activity, and expressing involvement. By removing or distorting gaze perception, we risk adversely affecting the processes of communication that depend on these functions. For instance, Vertegaal et al. [21] 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. [4] found that audio alone may be, in fact, preferable in routine business communication. Bos et al. [2] measured the effects of four different mediated channels – face-to-face, text, audio only, and video and audio – on trust building. They found that adding video did not significantly contribute to trust building when compared to audio-only channels in people who have not met face-to-face.

Furthermore, Short et al. [18] notes that a video channel may actually disrupt some communication processes when compared to audio only channels. 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, leading to an asymmetry in the understanding of the shared context. Argyle et al. [1] found that such asymmetries lead to noticeable increases in pause length and interruptions.

2.3.2 Gesture

Another important, but much less studied, cue that heavily depends on spatial information is gesture. Collocated groups in an office environment often point and gesture toward spaces where ideas were previously 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, standard video conferencing system which use single camera perspectives will necessarily distort gesture in group-to-group meetings for the same reasons that they distort gaze (Mona-Lisa effect)

3 Prior Work

Hydra [17] (Figure 3) supports multi-party conferencing by providing a camera/display surrogate for each remote participant in the meeting. This surrogate occupies the space that would otherwise be occupied by the corresponding 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 [22] (Figure 4) 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. The other participants in the group

Figure 3: Hydra

meeting see this *frontal planar* image in a 3D space rotated toward the image of the person being looked at. 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 blurs the image, 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 [12] (Figure 5) 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

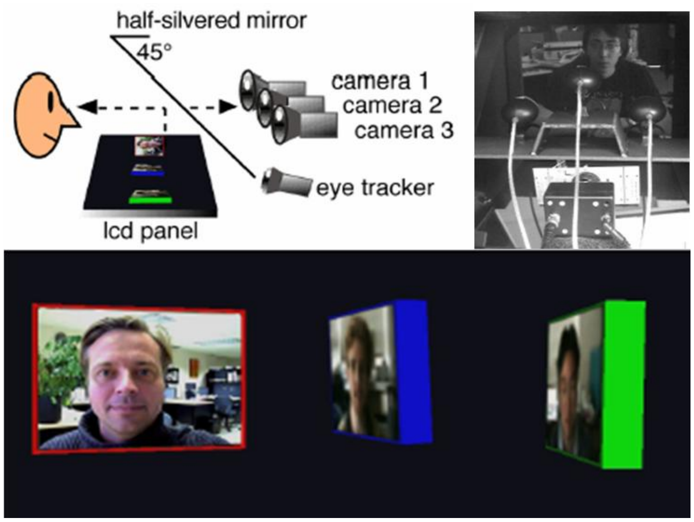


Figure 4: GAZE-2

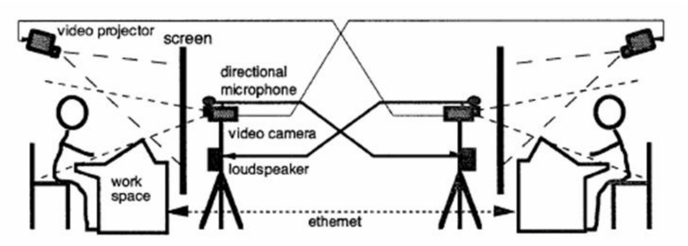


Figure 5: System architecture of MAJIC.

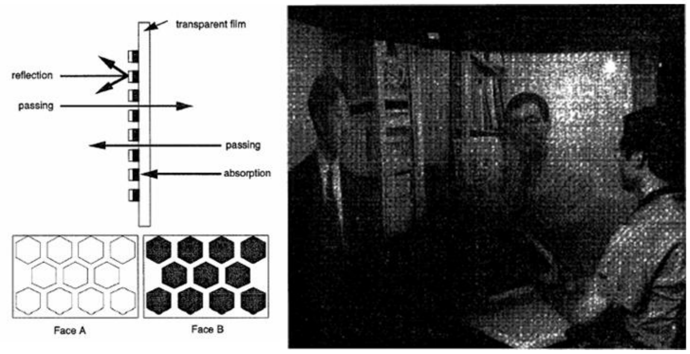


Figure 5: MAJIC

Figure 6: PRoP

user [13] (Figure 6). PRoPs suffer from the 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 with others. MultiView provides a practical solution to this problem, using a custom view-dependent display.

4 A Spatially Faithful Design

4.1 The MultiView Metaphor

We start with a "virtual conference room" which contains a large conference table as per Figure 7. 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.

4.2 Collapsed Viewer and Parallax Effects

Now consider a standard video conferencing system. There is one camera, and everyone views the image from this single camera. By placing a camera in the remote site, a *virtual viewing position* is created. Anyone viewing the video stream from this remote camera will take on the perspective of this virtual viewing position. When multiple people view the same video stream, everyone perceives the meeting from the same virtual viewing position regardless of their actual local position. This is what we call the *Collapsed Viewer Effect*. Because they all share the same virtual viewing position, it will be impossible for any participant

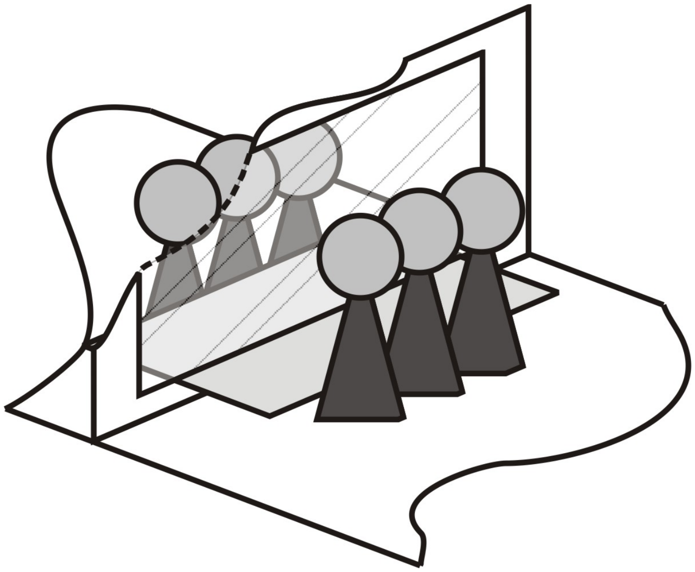


Figure 7: An illustration of the metaphor used by MultiView. A conference table with two groups of people on either side. The two groups of people are meeting through a large transparent window.

in a group to distinguish among them who any nonverbal cues from the remote site are directed at.

Additionally, every local participant is represented by an image at the remote site creating a *virtual presence position.* The virtual presence position is where all the remote participants will look when they intend to look directly at the corresponding participant. However, in standard video conferencing systems, the virtual viewing position from where the perspective is taken and the virtual presence position where a participant is represented as an image will most likely not match resulting in a *parallax effect.* As a result, all cues will undergo a geometric transformation defined by the difference in location between the two positions. This geometric transformation can be difficult to interpret by the viewer and may squelch any sensations caused by nonverbal cues.

To solve the collapsed viewer effect and the parallax effect, MultiView creates multiple *virtual positions* at the remote site which correspond to multiple specific positions at the local site. For each virtual position, a unique view is provided to the appropriate participant. Additionally, the virtual representation of that participant is projected to this same virtual position. Thus, a virtual position is created when the virtual viewing position and the virtual presence position for a particular participant at a particular local position are merged. Since the virtual viewing position and the virtual presence position are the same, there will be no parallax effect. By creating multiple virtual positions, the collapsed viewer effect is also eliminated. By carefully controlling the geometric relationships and creating a one-to-one mapping between the virtual positions and the local positions, MultiView can simulate a meeting around a common table.

An illustration of our approach is shown in Figure 8. On the top is a face-to-face meeting between two groups consisting of members [1, 3, and 5] and [R, C, and L] (these numbers and letters are used to maintain consistency with the description of the experimental setup that comes later). When the two groups are separated from each other, MultiView can effectively recreate the meeting setup shown in the top figure according to the illustration in the bottom figure. Participants 1, 3, and 5 have corresponding virtual representations in the remote site labeled 1', 3', and 5'. The virtual representations are located at virtual positions which correspond to the actual positions of the participants. The same goes for participants R, C, and L. Because MultiView carefully maintains a one-to-one relationship between the virtual and actual positions, it preserves many nonverbal cues.

4.3 Implementation

The MultiView solution is diagrammed in Figure 9. This solution supports two sites, and each site supports up to three participants. In order to provide multiple perspectives, MultiView uses multiple cameras, each providing a unique perspective to the appropriate remote participant.

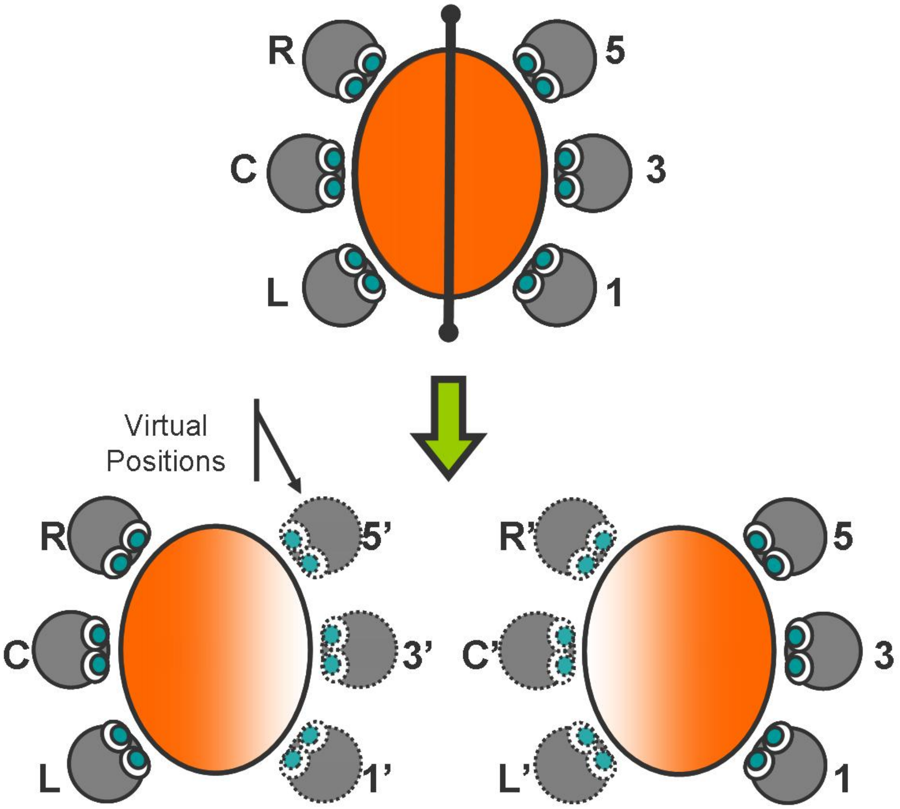


Figure 8: An illustration of how MultiView solves the *Collapsed Viewer Effect* and the *Parallax Effects*. The top image is a face-to-face meeting between two groups (R, C, and L) and (1, 3, and 5). MultiView creates multiple virtual positions from where the perspective is taken and the image of the participant is projected. Additionally, there is a one-to-one correspondence between the remote virtual positions and the local actual positions.

Term	Definition
Virtual Viewing Position	The position at the remote site from which a local participant's perspective is captured.
Virtual Presence Position	The position at the remote site where the image of the local participant is displayed.
Virtual Position	A position where the virtual viewing position and virtual present position coincide.
Collapsed Viewer Effect	The effect where multiple local participants share the same virtual viewing position. This phenomenon commonly occurs in standard video conferencing systems since everyone views a video stream from the same camera.
Parallax Effect	The effect where the perceived actions of remote participants does not correspond to the actual actions of the remote participants as a result of the virtual viewing position and virtual presence position not matching.

Table 1: A list of key concepts from section 4.2

Because we have multiple perspectives, MultiView provides a multiple viewpoint display of our own design so that each participant in a group meeting can see their own unique video stream while looking at the same display as all the other participants.

The front projected 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 (directly behind the projector). The design of the MultiView display is discussed in Section 4.3.1. Considering Figure 9, person 'L' will only see the image 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. At each site, there are multiple cameras, each providing a unique view for the remote site. Camera 'L' is projected only by projector 'L' on the other site, camera 'C' for projector 'C', and camera 'R' for projector 'R'. By transitivity, person 'L' will only see the video stream produced by

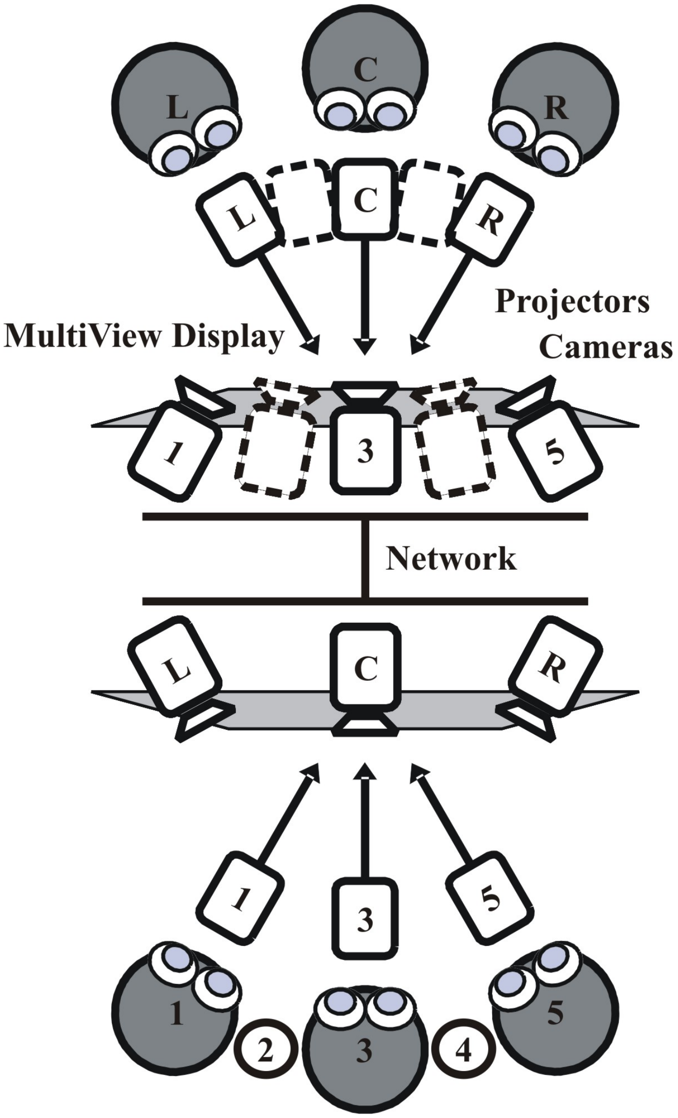


Figure 9: A diagram of the current MultiView Setup with two sites. Each site can support up to three participants. In our experiments, researchers sat at positions L, C, and R (for left, center, and right, respectively). Experiment participants sat at position 1, 3, and 5. Positions 2 and 4 were targets between participants 1 and 3, and 3 and 5, respectively. The dashed projectors and cameras are used to illustrate that MultiView sites do not necessarily have to be symmetric in the number of cameras and projectors. The upper site can support up to five viewers and send up to five video streams while the bottom site can only support three of each. The extra cameras and projectors are simply left unconnected.

camera ‘L’, person ‘C’ will see through camera ‘C’, and person ‘R’ will see through camera ‘R’. The same is true for participants 1, 3, and 5.

By placing each of the multiple cameras at the exact position of the remote participant’s virtual image, MultiView effectively creates a virtual position by merging the virtual viewing position with the virtual presence position. This simultaneously solves both the collapsed viewer effect and the parallax effects. To further maintain geometric relationships, the effective distance of the screen from the participants at one site has to be the same as the distance of the participants from the screen at the other site. The physical distance can can be adjusted if the image is scaled appropriately.

The simplest realization of camera placement is to put the cameras on top of the viewing screen right above the image of the person who will be seeing through that camera. In this way, there is no horizontal parallax. This setup does introduce the vertical parallax issues seen in standard desktop video conferencing systems (Figure 10), since the position of the cameras are above the position of the eyes of the image. However, we can leverage Chen’s findings that show an asymmetry in a person’s sensitivity to eye contact [3]. He found that people will perceive eye contact as long as the angle between the image of the eyes and the camera is less than 5°. Because of the scale of MultiView and the distance 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 is about 3°.

The resulting MultiView system is shown in Figure 1. Figure 11 shows photos taken from directly behind the projectors at positions 1 and 3 (see Figure 9 for positions) of the same screen and of the same people. All the participants in that photo were asked to look at position 5. Notice that for each participant, there is a camera directly above their image. This provides each remote participant with their own unique perspective.

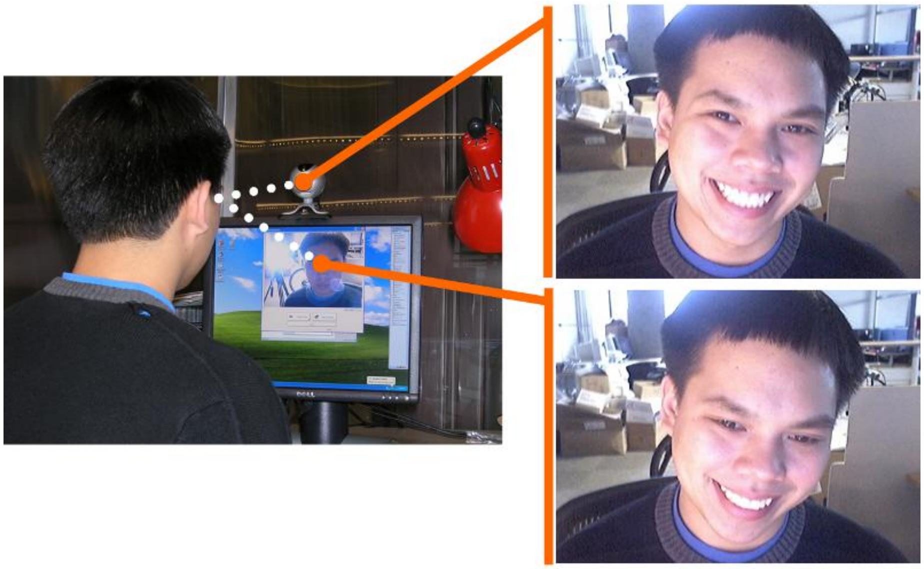


Figure 10: A demonstration of gaze parallax in standard desktop systems. When the local participant looks at the image of the remote participant in the eyes, the remote participant sees an image which suggests they are being *looked down upon*. The local participant can simulate direct eye contact for the remote participant by looking directly into the camera, but now the local participant is forced to look at the camera.

Figure 11: Two different photos of the same screen 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 9 for positions).

The sites do not necessarily have to have the same number of camera/projector pairs. The top site in Figure 9 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.

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. Additionally, the participants had to look up to look at the image of the remote participants. A better approach was to fix the cameras at a height slightly above eye level and allow the screen to hang below. We didn't 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).

4.3.1 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.

The MultiView screen uses multiple layers to create its viewing zones. A diagram of layers is shown in Figure 12. The back-most layer is a retroreflective cloth. An ideal retroreflective material bounces all of the

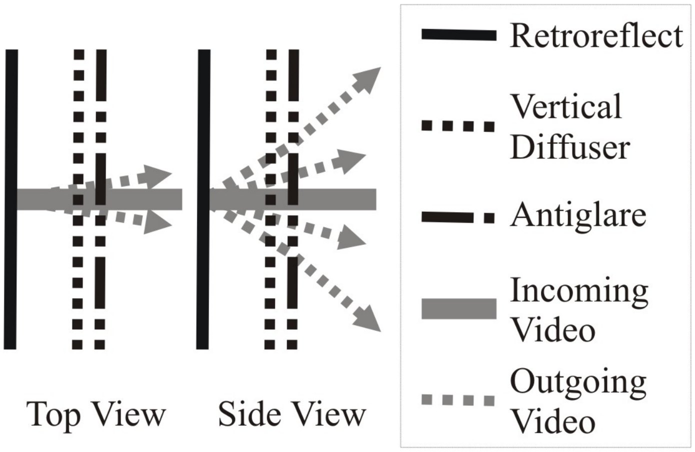


Figure 12: Slice views showing the multiple layers of the MultiView screen. The backmost layer is is a retroreflective sheet. The center layer is a vertical diffuser. The frontmost layer is an antiglare layer. The “Top View” shows a small amount of diffusion in the left and right directions. The “Side View” shows a large amount of diffusion in the up and down direction.

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. There are many retroreflective materials available in the market, but we chose to use the 3M 8910 fabric 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 specular effects.

The next layer is a one-dimensional diffuser which extends the viewing zone vertically creating a “slice”. Without it, the image would only be visible directly on the projection axis. This is problematic because if a person were in front of the projector, she would block the projected image, and if she were behind it, the projector would block her view. In our implementation, we used a lenticular sheet as the diffuser[1]. A spacing of 1/4" or more between retroreflect and lenticular sheet is recommended, otherwise the diffusion effects of the lenticular will be undone by the retroreflect (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 antiglare 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 antiglare film produced by DuPont (HEA2000 Gloss 110). The film has a pressure sensitive adhesive (PSA) so we used a method similar to that used in applying window tints to the the smooth side of the lenticular sheet.

The result is a display that is capable of showing multiple unique views to different viewing zones in space. We measured the diffusion properties of the MultiView screen by placing a projector at -10° from the normal and measuring the brightness of the reflected image at varying degrees. The diffusion profile is shown in Figure 13. As can be seen from the figure, the brightness is greatest at the angle of the projector at 35 lumens. The brightness quickly drops off and settles to about 5 lumens creating a narrow viewing zone of a few degrees. So, when we add a second and third projector at 0° and 10°, for a person viewing from -10°, the image from the projector in front of her would be about 7 times as bright as the image from the other projectors.

---

[1] Note: Lenticular sheets are often used in directional displays for multiple image 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.

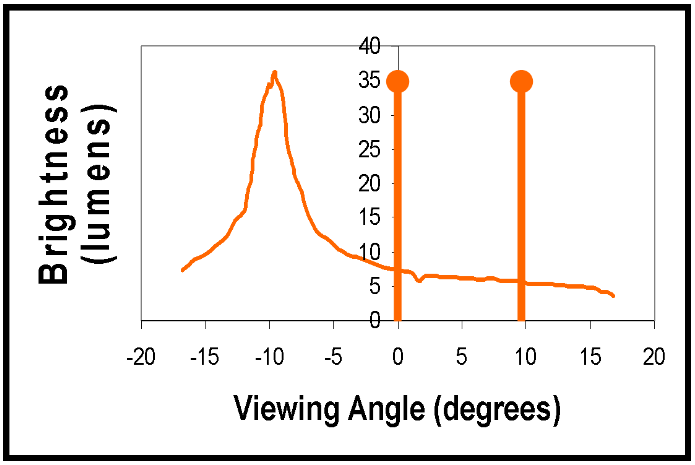


Figure 13: The diffusion profile of the MultiView screen. Light was measured at the same height of the projectors along multiple viewing angles. The projector was positions at −10° from the normal (coinciding with the peak of this curve). The two additional stems at 0° and +10° represent additional hypothetical projectors.

4.4 Cost

One of the benefits of MultiView is its relatively low cost and potentially high gain. The fixed cost is building the MultiView screen, costing around $700 (Table 2). The variable costs include the cameras and the projectors (Table 3). One camera ($100) is needed for each remote person, and one projector ($1000) is needed for each local viewer. The total cost of one MultiView site as it was implemented in our setup (3 cameras, 3 projectors) was about $4000. A key advantage for MultiView is that the variable costs increase linearly with the number of participants. Clearly, the projectors account for most of the cost. Fortunately, projectors, like computers, have a history of decreasing cost and increasing picture quality. Recently, projectors fell below $1,000 and continue to decrease. In addition, they are becoming smaller and consuming less power, which, as we will see, present some very interesting scenarios.

Item	Cost
Retroreflective Sheet	$50.00
Lenticular Sheet	$50.00
Antiglare Layer	$600.00
Total	$700.00

Table 2: Fixed cost for building a MultiView site. These are the materials needed to build the screen.

Item	Cost/Unit
Camera + Lenses	$100.00
Projectors	$1000.00
Total	$1100.00

Table 3: Variable costs for building a MultiView site. One camera is needed for each remote participant supported. One projector is needed for each local viewer supported.

4.5 Setup

Each projector must be positioned correctly to present the view of a remote camera; however, the alignment step can be easy and straightforward. In our current implementation, sitting behind a projector automatically configures you for a meeting by virtue of the fact that both sites were set up with fixed projector and camera positions.

We can relax the constraint of fixed projectors and cameras with little added cost in initiating a meeting. Each camera is set permanently at a certain position and view angle when it is attached to the screen. For the screen+cameras at site A, assume this camera information 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 and participant placement at site B for a conference with site A, the site A camera file is first downloaded. Then *local video* from the site B center camera is fed back to the projector whose position is being configured so that the participant will see himself and the local scene. Superimposed on the video will be red vertical lines that show all the local positions that correspond to the remote virtual position using the camera data file from site A. The participant must then move the projector and himself left or right until he aligns with one of the red lines in the local view. Once he does, the site B projector is switched to the video feed from the corresponding site A camera he selects, and it will faithfully reproduce the view from that angle. This setup process should take only a few seconds, which is important if MultiView is to be used with varying numbers of participants or with a stowable screen.

4.6 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 14 illustrates a three site setup supporting multiple participants at each site. In the three site configuration, the A1 cameras are projected onto screen B2, the A3 cameras onto screen C2, the B2 cameras onto screen A1, the B3 cameras onto screen C1, C1 cameras onto screen B3, and the C2 cameras onto 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.

5 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 meetings, MultiView can support multiple types of cues concurrently. During calibration, a person setting up MultiView was able to look at someone at the remote site and

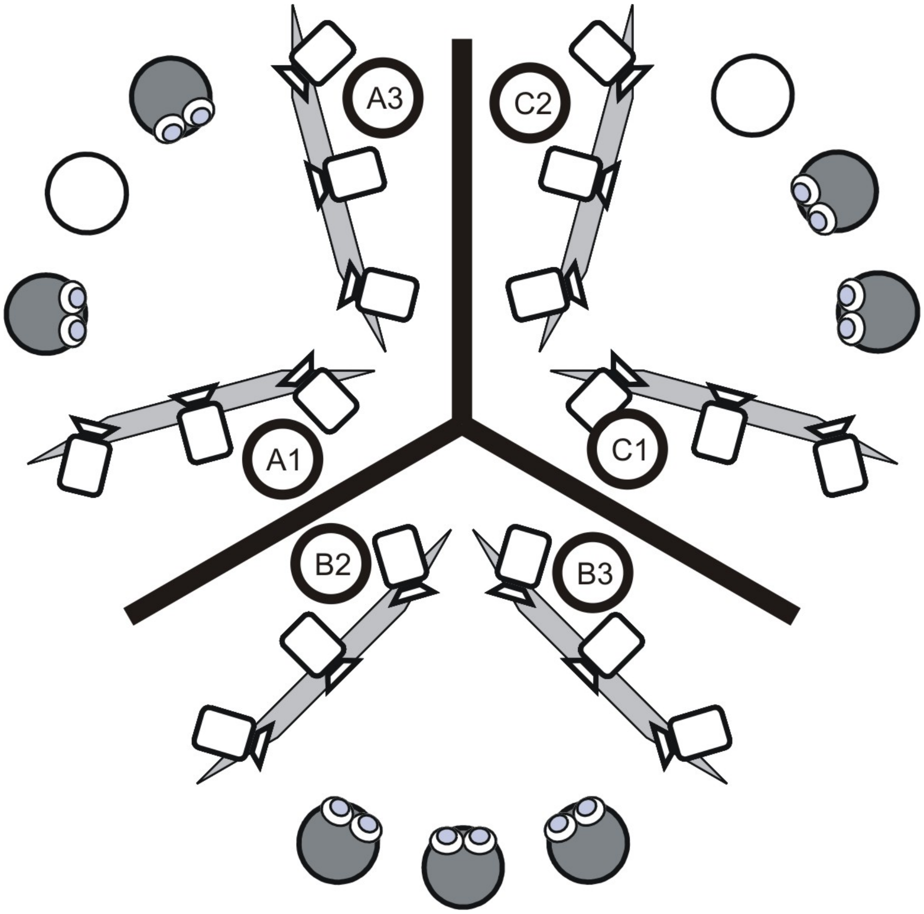


Figure 14: A three site setup of MultiView. Each site can support up to three participants but sites A and C are not fully populated.

point in a direction to say tacitly, “Hey you, go that way.” He was able to use two nonverbal 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 was 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 [15]. 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 [22] 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 [17] uses small LCD panels as a display. In the MultiView display, the entirety of the 36”x48” screen 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 [8], remote participants share an electronic white board. This supports full gaze awareness of graphics on the white board, 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 [16] or a lenticular separation method [10] 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 number of pixels for 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.

6 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 are testing with respect to *gesture*. The primary goal of experiment 3 is to test MultiView's support for *mutual spatial awareness with respect to gaze* (i.e. mutual gaze awareness) for all pairs simultaneously.

6.1 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 was a pool of six researchers used in sets of three. The makeup of the researcher group for each session was determined by availability.

6.2 Experimental Setup

In all three of our experiments, we used the MultiView setup shown in Figures 1 and 9. Everyone sat approximately 12' from the screen. Because of available materials and the size of the scene we wanted to capture, we used a less-than life-sized (36”x48”) screen. The 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’.

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 stimuli. 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 screen. Participants in the study sat behind 1, 3, or 5.

At the end of the experiments, the participants were asked the following question in order to help us interpret results, provide insight into the way MultiView was used, determine possible design improvements, and guide future work:

> “Please use the space below for any comments you have on our new system. This may include, but is not limited to, details about the system, reactions to how you felt about using the system, any perceived differences between using MultiView and face-to-face meetings, perceived differences between MultiView and other video conference systems you’ve used, etc.”

6.3 Experiment 1: Group Gaze

6.3.1 Task

In experiment 1, each researcher was instructed to look at one of the five 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 participant's eyes on the screen. If the position was in between two participants (positions 2 and 4), they were asked to look toward that position 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 ten times.

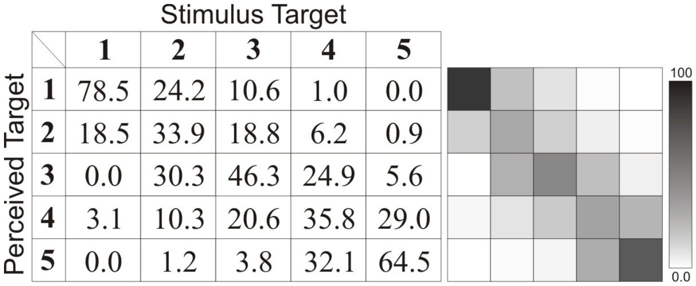
6.3.2 Results

The results of experiment 1 are presented in different ways that are relevant to the discussion that follows. Figure 15 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. For the condition of gaze, 90% of the responses were at most one target off.

Another measure takes a closer look at 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 4 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 was 0.70. An analysis of variance showed that viewing



	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 15: 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.

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

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 5 presents the mean and standard deviation of error by the target of the gaze stimuli. For instance, the mean error of responses to all stimuli targeted at position 2 was 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 was no significant difference for any other pairing.

6.3.3 Discussion

Referring back to Figure 9, we consider the seventh trial of our third session. Researcher L is instructed to look at target 1, Researcher C at

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

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 similar 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.

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.

In much of the established literature on gaze, acuity is often measured in degrees. Given the above geometry, we can calculate the change in angle between any two adjacent attention targets as:

$$\Delta\alpha = \arctan\left(\frac{20cm}{300cm}\right) = 3.82^\circ$$

If we multiply this value by the mean error, an extremely rough estimate of sensitivity in degree measure can be produced: $0.64 \cdot 3.82^\circ = 2.45^\circ$. This value is roughly on par with previous empirical values for gaze direction

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

acuity [3, 7]. This experiment does not have the precision or setup required to accurately measure human acuity and was not intended to do so, but we felt the coincidence compelling.

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 was 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.

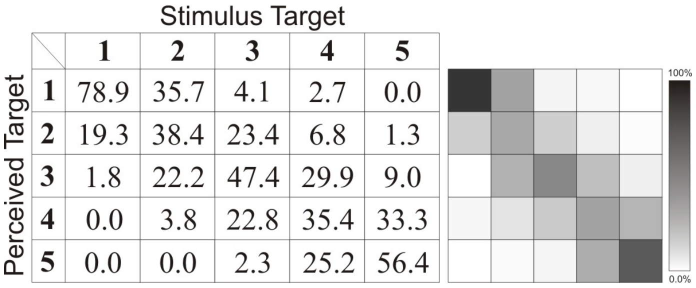
6.4 Experiment 2: Group Gesture

6.4.1 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.

6.4.2 Results

The results found in experiment 2 were very similar to those found in experiment 1. They are summarized in Figure 16, Tables 6 and 7 without further discussion. For the condition of gesture, 94% of the responses were at most one target off.



	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 16: 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.

Gesture 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 7: The mean error (μ) and standard deviation (σ) in perceived gesture direction for each set of stimuli directed at each target.

6.5 Experiment 3: Mutual Gaze

6.5.1 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 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.

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 8: The responses of the participants based on the direction of gaze in experiment 3.

6.5.2 Results

A summary of the results from this experiment are given in Table 8. 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.

6.5.3 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 sensation of eye contact. In [14], 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.

7 Future Work

7.1 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 physical distance the participants sit from the screen, which was 12’ in the study. This distance was set by the throw distance of the projectors we used. Combining the effects of the physical distance from the screen and the image scaling, the participants were sitting at an effective distance of about 18’. This was a very large “virtual conference table”, and it is

Figure 17: The new MultiView site. Improvements from the previous version include a 48”x96” screen, and short throw projectors reducing the effective viewing distance from 18’ to 8’.

perhaps not surprising that participants had some difficulty determining remote participants’ gaze direction.

With these improvements in mind, we have completed construction of a new MultiView system (Figure 17) which incorporates the changes mentioned above. Specifically, we increased the size of the screen from 36”x48” to 48”x96”. Additionally, we are using short throw projectors (Hitachi CP-X275W) which are capable of projecting an 80” wide image from just 8’. The result is that we reduce the effective viewing distance from 18’ in the prior system to 8’ in the new system.

We expect much more accurate gaze estimation at an effective 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.

7.2 Higher Level User Tests

In the experiments discussed above, we performed low level, perception-based user studies. 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.

7.3 Longitudinal Studies

In addition to laboratory testing, we plan on setting up a MultiView system between our lab and another lab for use in long term collaboration. As Dourish et al. has shown in [5], the behavior of using a computer-medicated communication system can change with long term use. We are interested in seeing how the behavior of participants using our system changes over time.

7.4 Personalized views

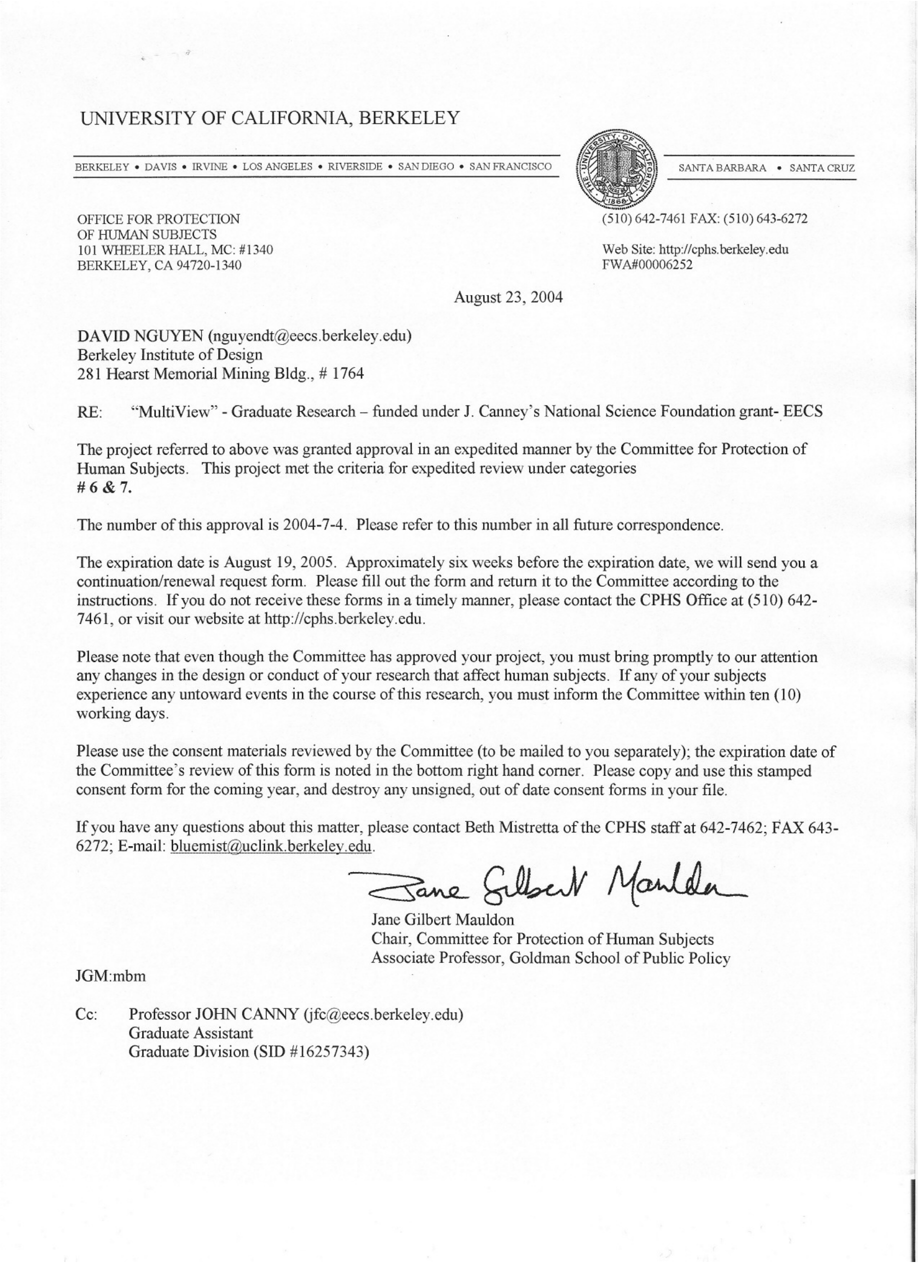
Three evolving technologies will make MultiView match its metaphor even better. The first is the development of micro-projectors [20]. This new breed of projectors are predicted to scale down to the size of matchbooks and use a fraction of the energy required by current projectors. Though they are predicted to produce lower light levels, they are an ideal match for MultiView because the high gain (directional) display concentrate the brightness back to the viewer. The second is the development of algorithms for *synthetic* video views that interpolate from a set of fixed cameras [19]. The third are a set of tracking technologies. With these technologies, every person could have a micro-projector 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.

8 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 multi-party 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.

A Human Subjects

A.1 CPHS Approval Letter



UNIVERSITY OF CALIFORNIA, BERKELEY

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO   SANTA BARBARA • SANTA CRUZ

OFFICE FOR PROTECTION
OF HUMAN SUBJECTS
101 WHEELER HALL, MC: #1340
BERKELEY, CA 94720-1340

(510) 642-7461 FAX: (510) 643-6272

Web Site: http://cphs.berkeley.edu
FWA#00006252

August 23, 2004

DAVID NGUYEN (nguyendt@eecs.berkeley.edu)
Berkeley Institute of Design
281 Hearst Memorial Mining Bldg., # 1764

RE: "MultiView" - Graduate Research – funded under J. Canney's National Science Foundation grant- EECS

The project referred to above was granted approval in an expedited manner by the Committee for Protection of Human Subjects. This project met the criteria for expedited review under categories
 # 6 & 7.

The number of this approval is 2004-7-4. Please refer to this number in all future correspondence.

The expiration date is August 19, 2005. Approximately six weeks before the expiration date, we will send you a continuation/renewal request form. Please fill out the form and return it to the Committee according to the instructions. If you do not receive these forms in a timely manner, please contact the CPHS Office at (510) 642-7461, or visit our website at http://cphs.berkeley.edu.

Please note that even though the Committee has approved your project, you must bring promptly to our attention any changes in the design or conduct of your research that affect human subjects. If any of your subjects experience any untoward events in the course of this research, you must inform the Committee within ten (10) working days.

Please use the consent materials reviewed by the Committee (to be mailed to you separately); the expiration date of the Committee's review of this form is noted in the bottom right hand corner. Please copy and use this stamped consent form for the coming year, and destroy any unsigned, out of date consent forms in your file.

If you have any questions about this matter, please contact Beth Mistretta of the CPHS staff at 642-7462; FAX 643-6272; E-mail: bluemist@uclink.berkeley.edu.

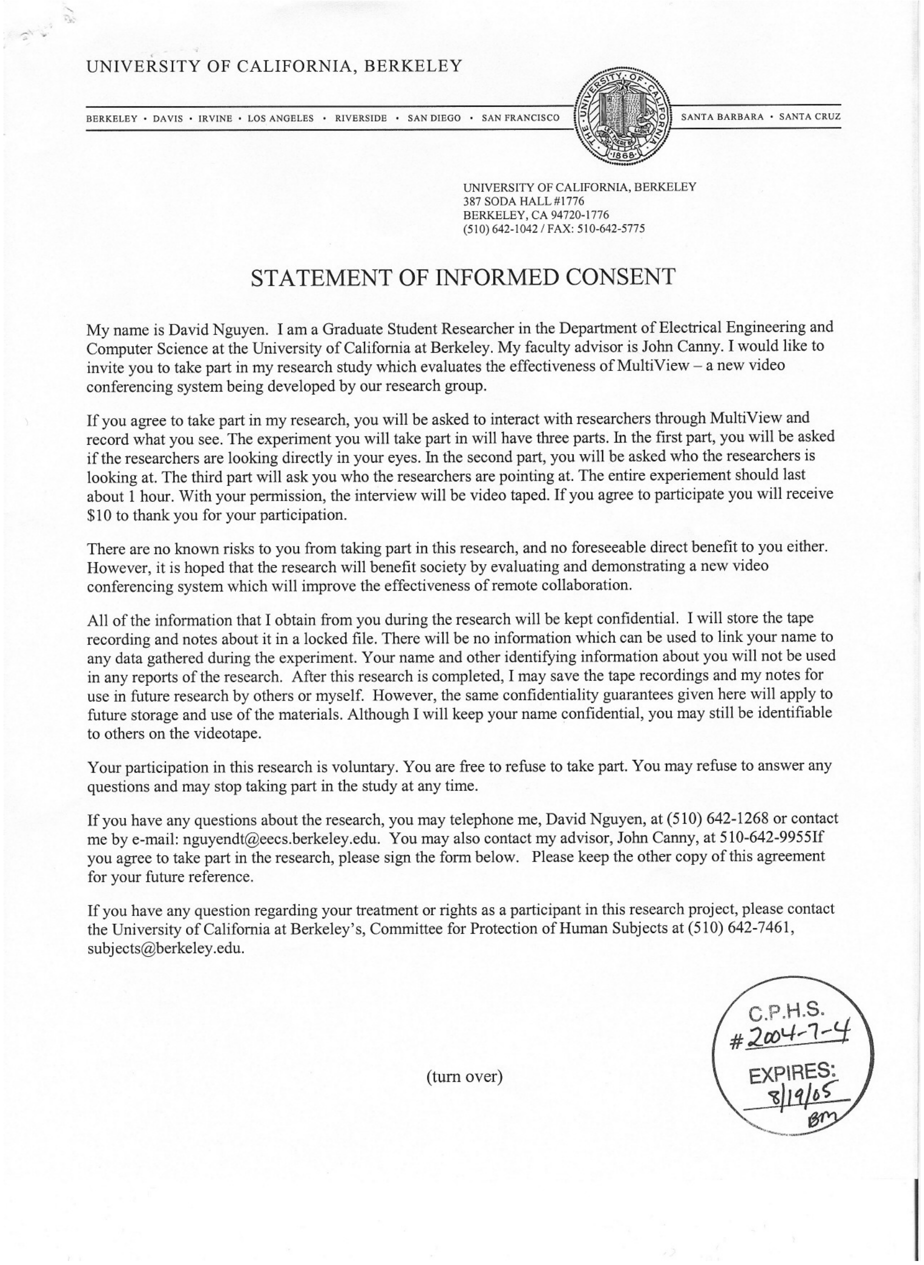
Jane Gilbert Mauldon

Jane Gilbert Mauldon
Chair, Committee for Protection of Human Subjects
Associate Professor, Goldman School of Public Policy

JGM:mbm

Cc: Professor JOHN CANNY (jfc@eecs.berkeley.edu)
Graduate Assistant
Graduate Division (SID #16257343)

A.2 Statement of Informed Consent



UNIVERSITY OF CALIFORNIA, BERKELEY

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO 1868 SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, BERKELEY
387 SODA HALL #1776
BERKELEY, CA 94720-1776
(510) 642-1042 / FAX: 510-642-5775

STATEMENT OF INFORMED CONSENT

My name is David Nguyen. I am a Graduate Student Researcher in the Department of Electrical Engineering and Computer Science at the University of California at Berkeley. My faculty advisor is John Canny. I would like to invite you to take part in my research study which evaluates the effectiveness of MultiView – a new video conferencing system being developed by our research group.

If you agree to take part in my research, you will be asked to interact with researchers through MultiView and record what you see. The experiment you will take part in will have three parts. In the first part, you will be asked if the researchers are looking directly in your eyes. In the second part, you will be asked who the researchers is looking at. The third part will ask you who the researchers are pointing at. The entire experiement should last about 1 hour. With your permission, the interview will be video taped. If you agree to participate you will receive $10 to thank you for your participation.

There are no known risks to you from taking part in this research, and no foreseeable direct benefit to you either. However, it is hoped that the research will benefit society by evaluating and demonstrating a new video conferencing system which will improve the effectiveness of remote collaboration.

All of the information that I obtain from you during the research will be kept confidential. I will store the tape recording and notes about it in a locked file. There will be no information which can be used to link your name to any data gathered during the experiment. Your name and other identifying information about you will not be used in any reports of the research. After this research is completed, I may save the tape recordings and my notes for use in future research by others or myself. However, the same confidentiality guarantees given here will apply to future storage and use of the materials. Although I will keep your name confidential, you may still be identifiable to others on the videotape.

Your participation in this research is voluntary. You are free to refuse to take part. You may refuse to answer any questions and may stop taking part in the study at any time.

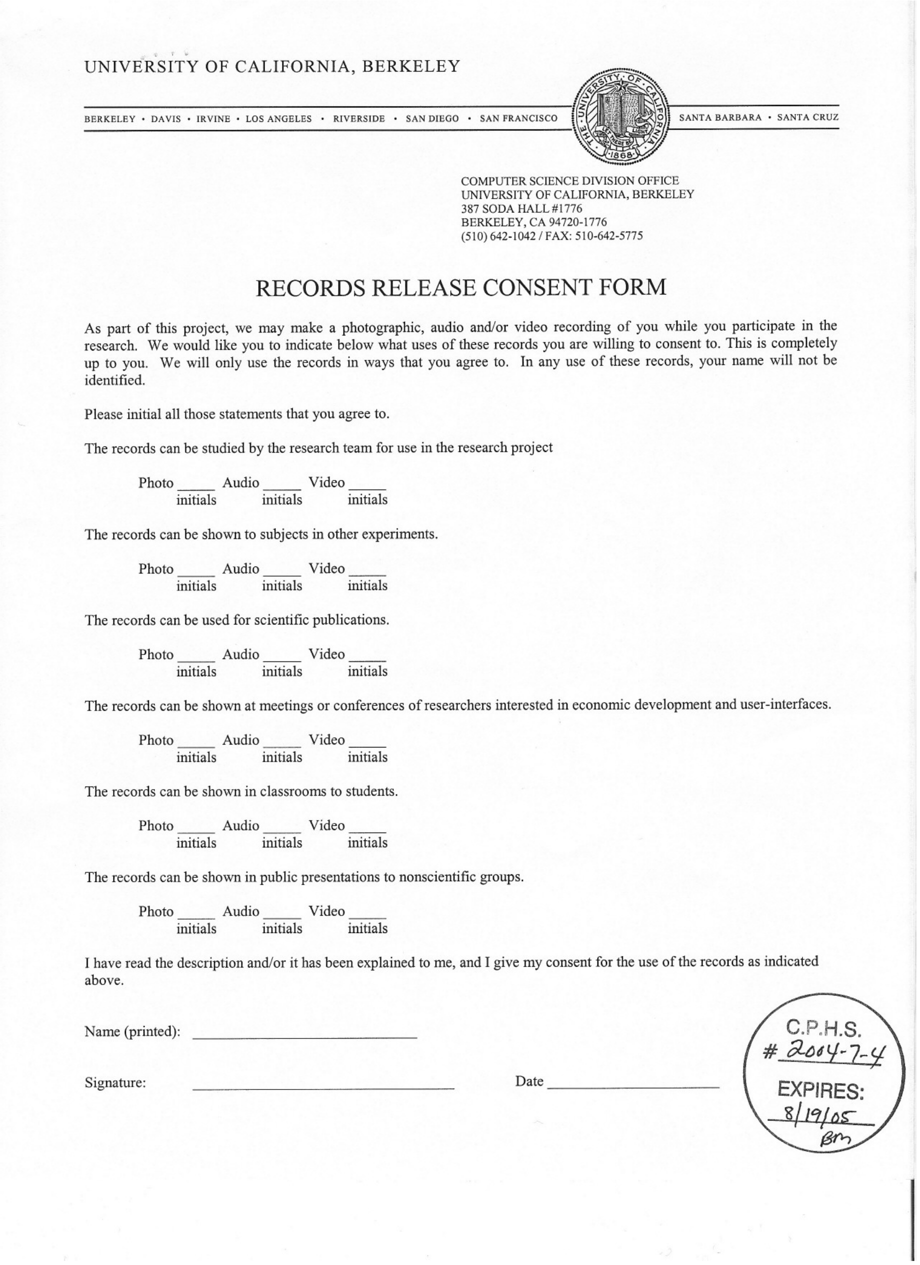
If you have any questions about the research, you may telephone me, David Nguyen, at (510) 642-1268 or contact me by e-mail: nguyendt@eecs.berkeley.edu. You may also contact my advisor, John Canny, at 510-642-9955If you agree to take part in the research, please sign the form below. Please keep the other copy of this agreement for your future reference.

If you have any question regarding your treatment or rights as a participant in this research project, please contact the University of California at Berkeley's, Committee for Protection of Human Subjects at (510) 642-7461, subjects@berkeley.edu.

C.P.H.S.
#2004-7-4
EXPIRES:
8/19/05
BM

(turn over)

A.3 Record Release Form



UNIVERSITY OF CALIFORNIA, BERKELEY

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO 1868 SANTA BARBARA • SANTA CRUZ

COMPUTER SCIENCE DIVISION OFFICE
UNIVERSITY OF CALIFORNIA, BERKELEY
387 SODA HALL #1776
BERKELEY, CA 94720-1776
(510) 642-1042 / FAX: 510-642-5775

RECORDS RELEASE CONSENT FORM

As part of this project, we may make a photographic, audio and/or video recording of you while you participate in the research. We would like you to indicate below what uses of these records you are willing to consent to. This is completely up to you. We will only use the records in ways that you agree to. In any use of these records, your name will not be identified.

Please initial all those statements that you agree to.

The records can be studied by the research team for use in the research project

Photo _____ Audio _____ Video _____
initials initials initials

The records can be shown to subjects in other experiments.

Photo _____ Audio _____ Video _____
initials initials initials

The records can be used for scientific publications.

Photo _____ Audio _____ Video _____
initials initials initials

The records can be shown at meetings or conferences of researchers interested in economic development and user-interfaces.

Photo _____ Audio _____ Video _____
initials initials initials

The records can be shown in classrooms to students.

Photo _____ Audio _____ Video _____
initials initials initials

The records can be shown in public presentations to nonscientific groups.

Photo _____ Audio _____ Video _____
initials initials initials

I have read the description and/or it has been explained to me, and I give my consent for the use of the records as indicated above.

Name (printed): ____________________________

Signature: ________________________________ Date ______________________

C.P.H.S.
2004-7-4
EXPIRES:
8/19/05
BM

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., Mendelsohn, J., Robins, R., 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] Gibson, J. and Pick, A. Perception of another persons looking behavior. *American Journal of Psychology* (1963), 386-394.

[8] 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.

[9] Kendon, A. Some functions of gaze-direction in social interaction. *Acta Psychologica 26* (1967), 22-63.

[10] Lipton, L. and Feldman, M. A new autostereoscopic display techology: The synthagram. http://www.stereographics.com/.

[11] 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.

[12] 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.

[13] Paulos, E. and Canny, J. Prop: Personal roving presence. *ACM SIGCHI*, 1998. Los Angeles, pp 296-303.

[14] Perrett, D. I. Organization and Functions of Cells Responsive to Faces in the Temporal Cortex. *Phil. Trans.: Biological Sciences 335*, 1274 (1992), 23–30.

[15] 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.

[16] Schmidt, A. and Grasnick, A. Multiviewpoint autostereoscopic dispays from 4d-vision gmbh. *Stereoscopic Displays and Virtual Reality Systems IX 4660*, 1 (2002), 212-221.

[17] Sellen, A., Buxton, B., and Arnott, J. Using spatial cues to improve videoconferencing. *Proc. CHI 1992*, ACM Press (1992), 651-652.

[18] Short, J., Williams, E., and Christie, B. *The social psychology of telecommunications*. Wiley & Sons, London, 1967.

[19] 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.

[20] Upstream Engineering, Inc. http://www.upstream.fi/.

[21] Vertegaal, R., Van der Veer, G. and Vons, H. Effects of Gaze on Multiparty Mediated Communication. *Proc. Graphics Interface 2000*, (2000), 95–102.

[22] 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.

[23] Zitnick, C. L., Gemmell, J., and Toyama, K. Manipulation of Video Eye Gaze and Head Orientation for Video Teleconferencing. *Microsoft Research Technical Report*, MSR-TR-99-46, June 1999.