Rogers, Bill; Apperley, Mark; Delos Reyes, Ervin; Masoodian, Masood

Wedge Video: Supporting Remote Participants in a Mixed-Mode Videoconference Meeting

Published in:
Interacting with Computers

DOI:
10.1093/iwc/iwad032

Published: 10/05/2023

Document Version
Publisher's PDF, also known as Version of record

Interacting with Computers, 2023, **00**, 1–15
https://doi.org/10.1093/iwc/iwad032
Advance access publication date 14 April 2023
Article

Wedge Video: Supporting Remote Participants in a Mixed-Mode Videoconference Meeting

Bill Rogers[1], Mark Apperley[1,*], Ervin Delos Reyes[1] and Masood Masoodian[2]

[1]School of Computing and Mathematical Sciences University of Waikato Private Bag 3105 Hamilton 3240 New Zealand
[2]Aalto University, Finland
*Corresponding author: m.apperley@waikato.ac.nz

Abstract

As global COVID-19 pandemic response has moved from full lockdowns and partial lockdowns in most parts of the world to a post-COVID era, an interesting new phenomenon that has emerged is the increased prevalence of hybrid meetings with a mixture of online and in-person attendees. The opportunity for remote participants to observe the responses and interactions of others in the meeting is generally accepted as being limited. An experimental prototype system, called Wedge Video, has been constructed as an attempt to improve the experience of remote participants in hybrid in-person/remote meetings. Wedge Video uses standard screen and camera equipment with existing video conferencing software (Zoom). An evaluation of the prototype system was conducted based on three simple games that each required players to interact rapidly and with some use of body language or gaze direction. Encouraging results led to the examination of the geometry of screen and camera placement in detail. A system that has a somewhat 'virtual reality' feeling to it has now been developed. The remote user is given a view of the in-person part of the meeting with participants at the same scale and location as they would be if the remote user were at the table themselves. Similarly, the local participants see the remote person in place at their table, at a realistic scale and with close to accurate gaze direction. A very preliminary evaluation of these concepts has been promising.

RESEARCH HIGHLIGHTS:

- Remote participants in hybrid meetings are disadvantaged in engagement and social interaction.
- Screen arrangement and camera placement at both meeting room and remote site can be modified to improve this.
- Providing dual 'wedge' screens for the remote participant's view of the meeting room and their 'proxy' presentation in the meeting room can be very effective.

Keywords: hybrid meeting, remote participants, COVID, videoconferencing

1. INTRODUCTION

As COVID-19 pandemic restrictions on social contact and requirements for remote working have changed over the past 3 years, varying from country to country and region to region, an interesting new phenomenon has been the increased prevalence of 'hybrid' meetings (Saatçi *et al.*, 2019) with a mixture of online and in-person attendees, using various conferencing tools to support in-group conversations and interactions. Although such meetings have long been a feature of the videoconferencing scene involving several colocated participants connecting to others remotely, in many cases, the remote person is considered to have a different status in the meeting. For example, sometimes the remote person is invited in to give a presentation or is specifically consulted about a topic, perhaps attending only part of the 'main' colocated meeting. The contribution of such remote participants to the meeting in these situations may in some cases work well, especially when their video is presented on a large screen to the colocated participants. What is more questionable, however, is what the remote participants can themselves take from such meetings because their opportunity to observe the responses and interactions of others in the meeting is generally considered as being limited or at least less than others. If the remote person has no special status in a hybrid meeting, then they often give and gain little from attending, being in effect 'second class' attendees and missing out on important social cues (e.g., through direct gaze), which can in turn lead to lack of interest and attention and generally distraction and boredom (Xu *et al.*, 2017). Despite such limitations, due to the new circumstances arising from remote working during, and post, the COVID-19 pandemic, it is becoming increasingly common for many 'ordinary' would-be full participants to have to attend hybrid meetings remotely.

Interest in studying the problems of remote participants in hybrid meetings arose from the authors' personal experiences. However, others have also reported on the topic (op den Akker *et al*, 2009). In a study of engagement in virtual meetings, participants in a hybrid meeting were interviewed (Kuzminykh and Rintel, 2020). Here are two of four quotes published under the heading 'Remote engagement is difficult':

> *'The external people just participate much less. It's much more difficult to be a part of the conversation. Even if the chair is mindful about asking them. It's very hard for them to indicate that they want to say something, or to jump into a conversation.'*

Received: October 27, 2021. **Revised:** January 30, 2023. **Accepted:** February 22, 2023

> *'If you are one of the few people calling in, it's very hard to be heard. Because people don't give you space.'*

There is a large range of software, both commercial and experimental, for supporting virtual meetings with two principal paradigms (Rogers *et al.*, 2021). One, the 'virtual world', uses computer game technology, allowing users, represented by avatars, to navigate a first-person view of an (often extensive) model world; Second Life (Linden Research, 2023) is the iconic example. The second is the modern 'videoconference', exemplified by Zoom (Zoom Video Communications Inc, 2023). Although the display format can be varied, the typical view is a screen divided into a grid of small live video images or 'talking heads'. Since the start of the pandemic, Zoom has seen a huge increase in use, and the name has become synonymous with 'videoconference' or 'remote meeting' (Bailenson, 2021).

Building on these paradigms, there are many mixed systems in which video is used in some way that combines with navigation of a virtual space; for example, using small video windows as avatars to position in a model world. Most recently, variations of the virtual world style, designed for use with immersive virtual reality displays, have been developed (McVeigh-Schultz and Isbister, 2021). These systems are all designed, however, for pure virtual meetings.

There is little reported work on systems to support hybrid virtual/colocated meetings. The videoconference paradigm (Zoom) can be, and is used, in hybrid situations. It works for 'predominantly remote' hybrid meetings. For example, in a gathering of mostly lone remote participants, two or more people can colocate in front of a single camera, participating together as though they were a 'single participant'. A difficulty with this scenario is that their joint video will be presented as a single image in the display grid, giving smaller images of each person in comparison with those of lone participants. The videoconference format is also commonly used for 'predominantly colocated' situations, where most participants are together in a meeting room but a few 'Zoom' in. This is the format that led to the quotations noted earlier.

Kuzminykh and Rintel (2020) point out that not all (remote) attendees have difficulty with disengagement. Those with a peripheral interest in a meeting reported that attending remotely allowed them to deliberately disengage without seeming impolite. They were able do other work whilst monitoring the meeting for items of interest.

The research described in this article focuses on the specific hybrid meeting situation where a single person is remotely joining a relatively small in-person meeting with a handful other attendees. Of particular interest is the case in which the remote person wishes to be a full and equal participant. Such a meeting is usually conducted with participants seated around a table, all able to see each other's upper bodies, faces, activities and interactions. The reason for choosing to focus on this type of meeting is that this is a situation in which everyone is likely to want to engage and that it is arguably one of the worst situations for a single remote participant. It was decided to have only a single remote participant for simplicity. In this scenario, typical meeting room arrangements place the remote attendee's video display on a wall (or a large moveable platform) at one end of the table and usually above normal eye contact level. The camera feeding the meeting to the remote attendee is also usually near the display, typically looking down on the meeting. It may or may not have a wide enough field of view (FOV) to capture all in-person attendees. There is often a poor (distant) view of people at the other end of the table. As such, it can be difficult for the remote attendee even to tell who is speaking, let alone observe nuances of personal interactions and non-verbal cues, thus leading to the remote participant feel excluded from the meeting (Grønbæk *et al.*, 2021).

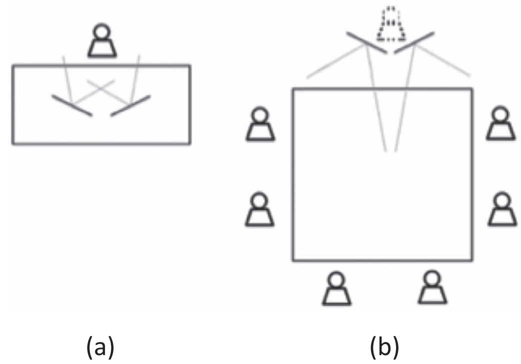

FIGURE 1. The Wedge Video configuration: (a) the remote participant's setup and (b) the meeting room configuration with the remote participant's proxy shown dashed at the top. Screen/camera combinations and viewing angles are also shown.

This article is divided into two main sections, each corresponding to different phases of the research. The work began with the Wedge Video concept. A prototype was built, then a method of evaluation devised and a usability trial conducted. Section 1 describes that process and the findings from the trial. The results were encouraging, suggesting that the system is a potential improvement to a standard videoconferencing setup, but they also suggested that the layout of cameras and monitors of the prototype could be refined for it to better support the interaction of the meeting participants. Section 2 of the article carefully analyses the identified issues relating to participants' views of a mixed-mode meeting and then proposes and investigates a more precise setup/layout. An initial pilot trial of this reconfigured prototype has also been conducted, leading to some promising findings.

2. INITIAL PROTOTYPE WEDGE VIDEO

The prototype Wedge Video setup is based on off-the-shelf computer hardware using standard videoconferencing software (Zoom). The physical configuration for a typical meeting is shown in Figure 1. The meeting uses two simultaneous bidirectional Zoom conference connections from the meeting room (with audio running on only one), with two cameras and two monitors at each end. The remote participant sits looking at a pair of standard 24-in. computer monitors each in landscape orientation. The monitors face the participant in a wedge-shape arrangement, and being moderately large, they occupy the part of the participant's FOV that, in an in-person meeting, would cover the meeting table and the other participants. In the meeting room, a similar pair of 24-in. monitors are placed, this time in portrait orientation, as an outward-facing wedge, oriented such that each monitor is most visible from one side of the meeting table. The meeting room monitors are placed slightly back from the table in a chair position. Each of the four monitors has a camera facing in the same direction as the monitor. Zoom conferences link the left remote monitor/camera to the left local monitor/camera and right to right. This gives the remote person a view as from the proxy position (shown dashed in Figure 1(b)) and the local people views of the remote person as though they were seated in the proxy position, and from the correct side.

Although the primary interest is enabling participation, it should be noted that the Wedge Video format may, at least partly, address some of the issues identified as likely causes of 'video fatigue'—the fact that people using videoconferencing during the pandemic found meetings more fatiguing than face-to-face meetings, despite total time in meetings slightly reduced over prepandemic experience (Standaert *et al.*, 2022). Bailenson (2021) argues that fatigue is caused by the following: 'eye gaze at a close distance'—being tightly face to face with others on a screen; 'cognitive load'—from trying to interpret limited body language; 'all-day mirror'—constantly seeing yourself on-screen leading to continuous self-evaluation and 'reduced mobility'—having to stay on camera. He also points out that it is rare for people to stare continuously at each other for long periods. Wedge Video presents all participants at roughly the same size and apparent distance as would occur in a real meeting. Potentially, it allows more natural body language, at least in part. In addition, because the screens provide no self-image, the user is not presented with an image of themselves to worry about. However, reduced mobility is still a problem because the remote person needs to stay at the focus of their camera. In a way, this configuration has characteristics of a virtual reality system in that it puts the remote user in a position in which their view is very like that of a person at the table. Sometimes others will be looking at them, but not always. Similarly, the remote person appears 'life-sized' in the proxy position, has two screens and can 'look around' the table. It is suggested that this should be a system that avoids much of the 'staring continuously at each other' that is characteristic of the grid-style videoconference.

2.1. Evaluation of the initial prototype

The primary aim of Wedge Video is to enable the remote participants of a hybrid videoconference to feel that they are part of the colocated meeting, to have access to as much of the necessary visual information as possible and to enable them to follow the personal interactions going on with, and between, all the attendees. Specifically, it was chosen to focus the evaluation on the ability of participants to be aware of 'who was looking at whom'. This could enable people to behave courteously by making sure that lines of view were open to others, to keep track of which attendees seem to be paying attention to speakers and to notice side interactions. A good solution would enable the remote attendee to have, and benefit from, the awareness of presence and location of the others and also keep colocated meeting participants aware of the remote person. Although this would fall short of the more exact interaction made possible by direct eye contact, it is important that such information is available in a form that makes it quick and 'easy' to use. Of course, it is probably not true that personal interactions of others are 'easy' to follow, but it is true that innate mental processing, refined by previous experience of similar situations, allows people to interpret what they see quickly and often without much conscious awareness of the processes. The aim of the Wedge Video system, therefore, is not so much to be 'easy' to use, but to leverage skills developed in the real world and to do so rapidly.

Consequently, in evaluating this experimental setup, it was not so much of interest to know whether the remote meeting participants could (slowly) work out who was looking at whom, but rather whether they could pick up this visual information easily enough to be able to make use of it in real time whilst mostly paying attention to verbal communication taking place at the meeting. However, although much is said about the importance of social awareness in remote meetings (Jackson, 2021), it is in fact rather difficult to measure such awareness. Depending on the meeting, situations in which awareness is important may not occur with high frequency.

To address this evaluation problem, three experiments were devised and conducted to test the real-time effectiveness of the Wedge Video meeting setup. These experiments were each based on simple games, one in each of the three experiments, in which the importance and speed of recognizing social interactions are greater than in most real meetings. The interactions in these games are not socially deep, but they do depend on an awareness of other people and their interactions.

2.2. Study tasks

The first game/experiment asks the question: 'Can a remote participant effectively keep up with, and participate in, a conversation involving body language in real time?' In particular, this task depends on sufficient support for turn-taking. In a game called 'Counting', the group of meeting participants are simply asked to count to 40 as quickly as they can as a group. Successive numbers must be spoken by different people, and everyone must participate. Turn-taking is the challenge because the participants are required to restart the count if two people speak at once. No instruction or time for planning is given, and participants may not say anything other than to contribute numbers to the count. This game requires that participants communicate by gaze observation and body language simply because no other means of communication is permitted.

The game can be better explained by looking at possible game sequences. After the game starts (in this experiment the researcher simply said 'start' to the group), someone must decide to say '1'. If two people make the decision and speak at the same time, then the count must restart. Assuming one person says '1', then another person (other than the starter) must then say '2'. It is only by looking at one another that people can guess for a good time to contribute. Sometimes, a group will themselves come up with the idea of speaking in order of seat placement. Sometimes the speaker might look at someone else to 'ask' them to be the successor. Groups differ in the solutions they adopt, but communication by 'looking' is always important. Having a remote participant adds the problem of audio lag, which must be accommodated. In this experiment, the lag was ~0.5 seconds, which may not seem much, but enough to effect the flow of normal human conversation. Groups therefore tended to experiment until they found a workable solution.

The evaluation was concerned only with finding out if it is possible to play this game using a test setup. There was no concern with the possibility that the game might run more slowly than in a wholly in-person setting, but rather, the aim was to provide evidence that the remote person could function as part of a group. The audio and video lag would certainly make the game slower, but that would not matter too much so long as the participants could find a way of playing the game that involved everyone and was acceptable to them.

The second game/experiment directly asks the question: 'Can participants tell who is looking at whom?' The game is called 'Winking Murder' ('Wink Murder,' 2023) and requires the players to watch each other. However, the layout of the group sitting around a table makes it impossible for a person to watch all other players at all times, so they must scan the group. If a person (A) is looking at someone else (B) and sees that person (B) wink 'at them', then they (A) 'die'—by closing their eyes and putting their head down on the table. The person who 'dies' must do so quietly and not do it in such a way as to deliberately draw attention to their 'killer'.

If a third person observes the wink interaction, then they call out the would-be 'killer', and that person must 'die' instead. The game continues until only one person is left 'alive'.

This game cannot be expected to work with the remote person in exactly the same way it would in a wholly in-person setting. The idea of winking 'at' someone in a real person-to-person setting will probably rely on eye contact to give a clear experience of being winked 'at'. If one of the participants is remote, the idea of being winked 'at' becomes less precise—it becomes the person looking in 'my' direction and winking—or if the winker is remote, reduces to looking in a direction close to 'toward me' and winking. Accordingly, this experiment set out to determine whether players felt that the information they had in the setup was close enough to 'reality' to allow them to play the game at all; in other words, whether they could gather enough 'looking' data quickly enough to make the game functional. It was not expected that the game would be exactly the same as it would be in real life, but to see if it could be played and if the participants would enjoy it, with the remote person actively taking part.

The final game/experiment more generally explores levels of participation in the group, with the aim of again trying to determine whether the remote person can fully participate in the activity. The original game is called 'Press Conference' ('Perfect Party Games,' n.d.). However in this evaluation, it was used only as the basis of a game, in that 'players' were not organized into teams or awarded any points. Instead, participants were asked to play iterations of the basic game format. In Press Conference, one person is the 'interviewee' and is assigned an identity (of a well-known person) without knowing that identity. All other players are 'reporters' and are told the identity of the interviewee. Although in the original game the reporters choose the identity, in this game, the identity was assigned at random from a preprepared list. The game format is for the reporters to ask questions that lead the interviewee to discover their identity. The name of the identity or their occupation (or close synonyms) cannot be used in questions. The interviewee is required to give 'correct' answers in the sense that they are consistent with answers they could sensibly give if they did know who they were. If the interviewee works out who they are without giving an answer inconsistent with their hidden role, they win. A bad answer or an incorrect guess causes loss. If fact, the 'game' can be played more for amusement and clever verbal interaction than for meeting the win condition. Making the other reporters laugh without actually causing the lose condition is a common goal.

In this experiment, the aim of using the game was to provide an engaging and fast-paced conversation, hopefully involving all participants including the remote person. The game would also avoid the artificiality of asking participants to 'conduct a meeting' with an agenda that would be of no significance to them. The interest here was mainly to see if the remote participant could take part in conversations sensibly.

2.3. Experimental setup

The three games/experiments were conducted using the monitor and camera layout described earlier (Figure 1). Large 24-in. video monitors (actually all-in-one computers) were used, with those for the remote participant being in landscape orientation and those in the proxy position turned to portrait. The experiment took advantage of the fact that Zoom can have a single image in full-screen mode in the case where there are only two end points in a conversation. The video camera's field-of-view was an issue, and a small hardware modification was made to the computers in the proxy position. Their built-in cameras would have been put

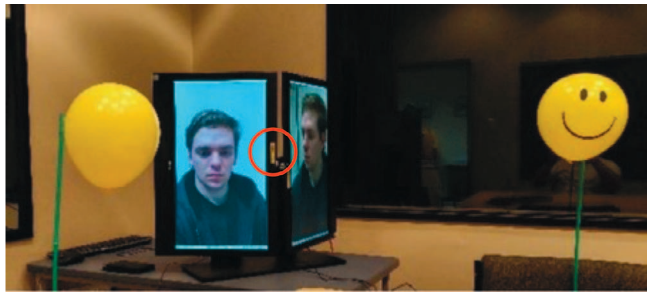

FIGURE 2. The physical setup in the meeting room showing the remote participant's proxy wedge image, with balloons used as dummy in-room participants. The red highlight circle shows the position of the modified cameras.

FIGURE 3. The remote participant's setup, with two screens showing the meeting room and balloons used as dummy participants.

into portrait orientation with a resulting FOV of ~40 degrees. This was insufficient to get the required panorama of the meeting. Accordingly, the cameras were taken out of their mounting and rotated into landscape orientation to give a nearly 70-degree FOV (actually measured at 66 degrees). The video management package OBS (Open Broadcast System) was used to modify the video coming from the remote cameras, both to rotate it and to crop the image to portrait view. Figure 2 shows the proxy setup with balloons standing in as meeting participants for visibility checks, and Figure 3 shows the remote participant's view. A variety of participant and proxy positions were trialled—for example, Figure 2 shows the proxy position at a corner of the main table—but the one shown in Figure 1 was settled on for this set of experiments. Preliminary visibility tests using balloons in place of most of the participants, as seen in Figures 2 and 3, were used to check that the camera orientations gave fields of view sufficient to show all participants at the table and to get the image of the remote participant to a sensible size.

The experiments had the proxy position in the meeting room forward of the position suggested in Figure 1, actually on the table itself, and had the proxy monitors angled at ~45 degrees. The proxy-position cameras were placed vertically in the middle and horizontally at the inside of the proxy monitors, as shown in Figure 2. This geometry element was dictated by the length of cables connecting the cameras. The cameras on the remote participant's monitors (Figure 3) were at the normal top centre positions and were not rotated or moved.

2.4. Findings from the initial prototype experiments

The experiments were repeated with four groups of participants, and each group played each of the three games 3–4 times. The participants were third year university computer science students, generally aged ~20–21 years and predominantly men. The

TABLE 1. Responses to the active engagement question for the three games

Game	*Meeting room participants*								*Remote participants*							
	Count	**1**	**2**	**3**	**4**	**5**	**Mean**	**SD**	**Count**	**1**	**2**	**3**	**4**	**5**	**Mean**	**SD**
Counting	11	0	0	0	2	9	4.82	0.40	4	0	0	1	2	1	4.00	0.82
Press Conference	11	0	0	1	3	7	4.55	0.69	4	0	0	1	1	2	4.25	0.96
Winking Murder	10	0	0	2	6	2	4.00	0.67	4	1	0	1	2	0	3.00	1.41

experiments were conducted as one of the activities in a human–computer interaction university course. Because it proved difficult to recruit enough student participants to fill all the places in each group, one or two of the researchers filled the empty spaces to ensure that each group comprised five participants. Generally, the participants did know one another before the experiment. As discussed further, the experiments showed that, in this environment, the participants were able to successfully play the 'Counting' and 'Press Conference' games, but did experience some problems with 'Winking Murder'.

To investigate whether the experience was satisfactory, each participant was asked to complete a questionnaire after each experiment session. For each of the three games, participants were asked to respond to the question: *'Did you feel like you were part of the group during the activities? Were you able to participate?'* on a 5-point Likert scale (anchors 1: Not active and 5: Very active). There were a total of 15 student participants and all completed this questionnaire. Eleven responses were from participants who took part in seated positions at the table and four from participants taking the remote role. The responses to this question for each game are summarized in Table 1.

For the 'Counting' game, the in-person participant responses averaged 4.82 and the remote responses averaged 4.00. Although the remote participants' average is lower, it still indicates that they felt they were able to take part in the game activity. Furthermore, the fact that the 'Counting' games were all completed is a strong evidence of effective participation.

For the 'Press Conference' game, the in-person participant responses averaged 4.55 and the remote responses averaged 4.25. Again, these ratings show a good level of participation and engagement.

For the 'Winking Murder' game on the other hand, the in-person participant responses averaged 4.00 and the remote responses averaged 3.0. This shows that for this game, the setup did not work as well for the remote participants as for the in-person ones. Furthermore, the less positive in-person results suggest concern about the difficulty of working with the remote person in this context.

A summary yes/no question was asked of in-person participants after all three games: *'Did you feel like the remote person was part of the group'*. Ten of the participants responded 'Yes' and just one 'No'.

More specific questions were also asked with respect to each game. For the 'Counting' game, participants were asked: *'In your opinion, how well did your group cooperate during this activity?'* with a 5-point Likert scale (anchors 1:Poor and 5:Great). These responses are summarized in Table 2. The meeting room participant responses averaged 3.91 and the remote responses averaged 4.75. It is interesting that in this case, the remote participants rated the cooperation higher. Perhaps the reason is that the setup worked well enough for the remote participants—who may have had low expectations—whereas the in-person participants were forced to cooperate with them, making their experience worse than might occur in a fully in-person situation.

Participants were also asked for a yes/no response to: *'Do you think you struggled with being involved in taking turns (audio lag, struggling with observation of the room, etc.)?'* Only one of the four remote participants answered yes to this question, in contrast to 5 of the 11 in-person participants. Again, this may indicate that dealing with the remote participants was seen as more difficult by the in-person ones.

After the 'Winking Murder' game, the participants were asked yes/no questions as to whether they struggled to make and observe others making eye contact. Three of the four remote participants reported struggling both with making and observing eye contact with in-person participants, whereas 5 of the 11 in-person participants reported struggling to make eye contact both with in-person and the remote participants. In addition, 7 of the 11 in-person participants reported struggling both to see in-person and the remote participants make eye contact with others. The results suggest that although this game is not easy to play in this experimental setup, it is interesting that in-person participants reported a similar experience in interacting with the remote person as with local participants. The results do show, however, that the remote participants had difficulty playing this game.

Finally, for the 'Press Conference' game, the number of successes was recorded—i.e., the number of people who successfully guessed their character without 'failing' an interview question. In all three instances of this game with the remote participant as the interviewee, the character was guessed successfully. In the 10 games with an in-person participant as the interviewee, half (five) of them were successful. Based on these results, it seems that this was a well-balanced game for this experimental setup because the game was playable and could include a remote participant effectively.

2.5. Summary of initial prototype findings

The physical configuration used in these experiments provides the remote participant with a complete and reasonably natural view of the meeting table. Similarly, it also provides those seated around the meeting table with an approximation to a natural view of the remote person, both in position and size, with careful adjustment of the video settings. The local people have a view of the remote person effectively 'sitting at the table'.

The three games used for the experiments demonstrate that the prototype setup is effective in various ways. The 'Counting' game shows that people can manage to use gaze and body language to communicate. The 'Press Conference' game shows that a remote participant can take part in a quick series of verbal interactions; in particular, they are able to manage turn-taking. The 'Winking Murder' shows that 'eye contact' was partly possible but did not work reliably enough to allow the game to run well.

Camera placement, even in one-on-one videoconferencing, usually means that there is little or no real sense of eye contact.

TABLE 2. Responses to the group cooperation question for the Counting game

Meeting room participants								*Remote participants*							
Count	1	2	3	4	5	Mean	SD	Count	1	2	3	4	5	Mean	SD
11	0	0	5	2	4	3.91	0.94	4	0	0	0	1	3	4.00	0.50

The prototype setup described here, with compromised camera positions chosen for practical implementation, cannot fully solve this problem either. However, the partial success of 'Winking Murder' suggests that the lesser goal—to see if the experimental setup could be refined to provide as much information as possible about 'who is looking at whom'—might be practical and could be refined from this initial ad hoc implementation.

3. MIXED-MODE MEETINGS; REFINING THE CONFIGURATION

This section turns to more carefully examining problems, options and geometry in the mixed-mode meeting. It examines the effects of the Wedge Video prototype geometry discussed in section 2 by developing the configuration in three stages.

Consider the situation of a single remote participant in an otherwise in-person meeting. The most common meeting room layout for this scenario is to have a rectangular table with participants sitting on each side and at one end. At the opposite end, there is typically a large monitor screen that can be used to display the video link to the remote participant when there is one. A camera, usually mounted on or near the monitor, can relay a view of the meeting to the remote person—along with a microphone and speakers providing the audio connection. In this type of setup, if there is need for a shared document it normally replaces the video from the remote person, although picture-in-picture can be used to keep a small live video. Anyone who has participated in such a meeting is likely to have observed some strengths and weaknesses of this arrangement.

Figure 4 shows a fairly extreme, although nevertheless real, example of this scenario involving two remote and 16 in-person attendees. Firstly, it should be noted that even the experience of the in-person participants is not ideal. Linear placement of people down the sides of a fairly narrow (1.8-m estimated) table makes it difficult for them to see most of the others along their own side of the table. However, the people at the end are visible to all and there is good line of sight between those on opposite sides.

To enable rapid exploration of configurations, a 3D first-person graphics model was built of a setting, with participants, cameras and monitors. This is a simulator in the sense that cameras are implemented as virtual cameras in the assigned positions, and it is the output of those cameras that is displayed on the monitor models. Cameras were implemented with a 70-degree horizontal FOV, which is a typical field width for Web cameras. Each simulated person can be run through a simple animation that moves their heads and arms, allowing observation of whether or not small movements would compromise views. The model uses 24-in. (diagonal), 1920 × 1080 monitors and cameras. The modelled table height is 75 cm. Size varies, but most of the simulated images that follow show a 2 × 2 m table. The simulated person used is a 186-cm tall man, with eye level of 170 cm standing and 123 cm seated. This is taller than average (the average human male height is 174 cm and female height is 161 cm) but was chosen to ensure there was enough space to accommodate most real people.

FIGURE 4. The remote participants' view of a large, mostly in-person, meeting using standard Zoom configurations.

Figure 5 shows the 3D simulation representation of the real arrangement shown in Figure 4 from the viewpoint of the fifth person on the left-hand side, looking toward the head of the table. Although this arrangement is less than ideal in terms of participants being able to see each other, adhering to common social conventions can improve visibility along the sides to some extent. For instance, as can be seen from Figure 5, slightly leaning forward as shown by second person (pink shirt) on the left side puts their face into view. If that person was speaking and leaning forward a little more, it is natural for those along the same side to lean back slightly to clear lines of sight for each other. Although such social conventions are commonplace, one could argue that it is the awareness of the physical presence of other people that makes them likely to be followed. These conventions, however, might not be followed for the benefit of the remote participant, or they may be affected by the fact that the image of the remote participant (in this case on the large screen) is not necessarily coincident with the camera providing the remote view. The situation in fact may be rather different, for instance, in a meeting with two uncooperative factions, when it is likely that the people of each faction would position themselves on opposite sides of the table, avoiding the need for direct cooperation with their opposition, but still achieving good visibility.

The experience of the remote people in this meeting is quite different to that of those present in person. In the example of Figure 4, the camera is mounted above table eye level of one of the remote participants (on the left), causing their video image to seem as if they are looking down. Furthermore, in the main meeting site, the remote participants are shown on a large screen—also positioned above eye level—requiring local participants to look up to see them. Figure 6 shows a 3D model representation of such a meeting with a single remote participant as viewed from the opposite end of the table in the meeting room. Similarly, the view from the remote person's perspective is shown in Figure 7.

The focus of this work is on making sure that all of those in the meeting, including remote participants, can see each other and can see who is looking at whom. This analysis can begin by

FIGURE 5. A simulated view from the perspective of the fifth person on the left-hand side of the meeting room table in Figure 4.

FIGURE 6. The simulated view from the head of the meeting room table, with the remote person in view on a large display, as is commonly done.

FIGURE 7. The simulated view for a single remote participant. The small red cube on the top of the monitor screen shows the position of the camera in this and subsequent figures.

looking at direct visibility. Tables 3 and 4 show direct visibility in pure in-person and pure online meetings, respectively. The only difference is that participants in online meetings can (usually) see themselves, potentially with a negative effect over time (Bailenson, 2021). Table 5 shows visibility for the mixed meeting,

TABLE 3. Visibility for in-person meetings

Can see	Self	Other
Self	✗	✓
Other	✓	✓

TABLE 4. Visibility for online meetings

Can see	Self	Other
Self	✓	✓
Other	✓	✓

TABLE 5. Visibility for mixed in-person/online meetings

View L = Local R = Remote		Self L	Self R	Other L	Other R
Self	L	✗	×	✓	✓
	R	×	✓	?	✓
Other	L	✓	✓	✓	✓
	R	✓	✓	✓	✓

with the red and green entries copied from Tables 3 and 4. The question mark for a remote person seeing others who are local reflects limited perspective and resolution and is discussed in the following paragraphs.

3.1. Perspective, camera angle and resolution

Human vision is capable of high resolution, at least near the centre of the fovea when a person fixates. A recent survey of issues in human visual acuity (Strasburger, 2020) quotes estimates for maximal visual acuity under good illumination as two-thirds minute of arc in young adults. As a rough approximation, taking the minimal acuity angle as equivalent to a pixel, this would give an image of a human face (taking the human face width as 15 cm) at 8-m distance (the distance from a participant at the screen position looking toward a person at the opposite end of the table) of width 100 pixels. This is consistent with the general observations that even when sitting at the opposite end of a long meeting room table, normal vision allows one to see a person at the far end in enough detail to read their facial expressions and be aware of their gaze direction. Interestingly, these observations also corroborate with reports from the Battle of Bunker Hill (Lockhart, 2011), in which American soldiers were told not to fire until they 'could see the whites their [British troops] eyes'. An eyewitness of that particular battle put that distance at 4 rods—or 80 m—suggesting a very high resolution in fixated vision.

In addition, the human vision system accounts for perspective in real-world settings, leading to the perception of even a long table as rectangular and the people at the far end as 'normal'-sized human beings when in a physical meeting. In a remote meeting setting, on the other hand, although the camera captures perspective in a video image from the same position as the remote participant's eye(s)—if they were in the physical meeting—they would still fail to view the resulting image on their monitor from a distance and direction that would allow their visual perception to correctly interpret the image and compensate for the perspective.

(a) (b)

FIGURE 8. Face images with widths of (a) 30 and (b) 100 pixels, respectively, showing the poor detail to be seen of participants some distance from the camera, as at the far end of the meeting table.

FIGURE 9. Magnified section of the far end of the table from the actual screenshot of Figure 4 revealing the poor representation of detail in the image itself.

In the simulation setup shown in Figure 7, the remote person is sitting close to their monitor, seeing an image with perspective appropriate for a greater distance. In such a scenario, the viewer's inability to properly use their perceptual 'perspective correction' exacerbates their sense of distance from people at the far end of the table; clearly, this would not happen in a real physical meeting.

In comparison, a camera with a 70-degree FOV (typical of cameras used for videoconferencing) and high-definition resolution (1920 × 1080 pixels) would see a face at the end of a long table (8 m) with width of 30 pixels. Figure 8 shows faces at pixel widths of 30 and 100 pixels. As expected, 100 pixels gives a detailed view, but the 30-pixel version lacks detail and is marginal for interpreting expression. In addition, a less-than-optimal choice of display scale and variations in assembly of the remote image on a monitor can significantly erode the viewing resolution. For example, the screenshot shown in Figure 4 is from the actual screen of a remote participant. This does not have the meeting room image presented at its full resolution and, as a result, loses almost all detail on the far-end faces, as can be seen in the zoomed-in detail section of Figure 9. The centre face in that image has a width of ~8 pixels, which is of course too little to show facial features, let alone expression. As such, in this particular example, the remote meeting participant would be unable to read facial expressions of the other meeting participants, contributing to them feeling themselves to be an 'outsider'.

3.2. How might the situation be improved?

What options are available for reorganizing meetings? This discussion has focused on the case of a single remote participant taking part in an otherwise in-person meeting and also been restricted to small meetings, mostly with ≤6 in-person participants. These conditions largely avoid problems with video resolution and make it possible (although still surprisingly difficult) to assemble groups of people for experiments. The scale is that of a tutorial discussion group or a subcommittee meeting.

What would be desirable to achieve? It is recognized that it is impossible for participants to achieve the kind of eye contact that would occur in an in-person conversation, even in a one-to-one conversation. In the simple one-to-one video conversation, it is obvious to each person that the other is looking at an image on a screen and not directly at the camera feeding the images. Experimental systems have achieved some successes with placement of small cameras directly on screen and with image manipulation of gaze direction. It should be noted, however, that this is difficult in the one-to-one situation, and equivalents for the meeting context have not been considered. Instead, the research has been directed toward a lesser goal: allowing participants to be able to see at whom each other participant is looking, with enough detail to be able to capture facial expressions. The reason for this choice is to try to keep the remote participant aware of, and participating in, the personal interactions of the meeting.

The goal of this research is then to enable one remote participant to take part in an otherwise small in-person meeting (up to 6 participants). By 'taking part', it is meant feeling like a full participant: being able talk and listen; being aware of who is talking; being aware of who is looking at whom; being able to see facial expressions and more generally to be able to do these things as an equal participant—not feeling left out and ideally not feeling special. Being aware of who is looking at whom helps give an understanding of the dynamics of the personal interaction and which is suggested is a key feature of being a full participant.

The meeting room seating arrangement for the following sections is as shown in Figure 10. There are six 'local' meeting room participants, numbered 1 to 6 and identified in the simulated images by the colours of their T-shirts, and the remote participant.

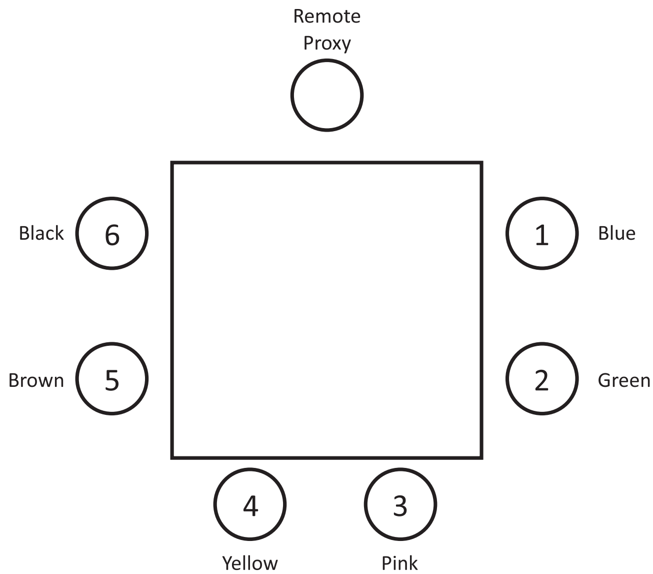

FIGURE 10. Seating arrangement for the following discussion. Meeting room participants are referred to by their index number (1–6) and the colours of their T-shirts, as indicated here.

3.3. The organization of the space

Having a remote participant on a 'cinema screen', as shown in Figure 6, sets them apart. Being looked up to makes a person

FIGURE 11. Improvement step 1: Moving the remote image to a natural position at the meeting table (C.F. Figure 6).

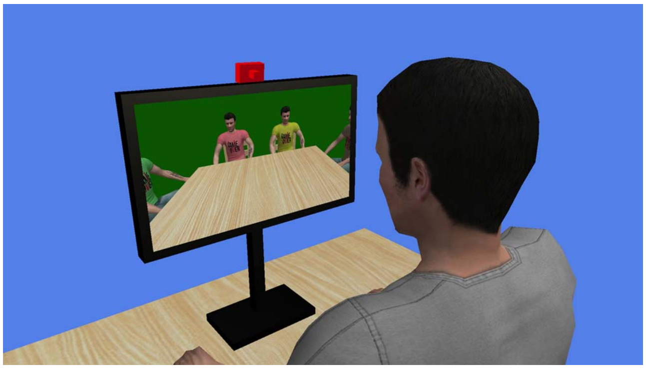

FIGURE 12. Improvement step 1: Remote view corresponding to the arrangement shown in Figure 11.

FIGURE 13. Improvement step 2: The remote participant's view, with two monitors set more closely to them. (The inset shows the dual proxy meeting room cameras).

special, usually superior, but in the meeting context where participants are focused downward on documents, ironically easier to ignore or forget. If a participant glances around the meeting to try to get a sense of consensus, for example, they may not look up at the screen.

3.4. Improvement step 1: Presenting the remote participant in the correct position at the table

The first improvement step is to use a smaller monitor for the remote person and move it to a natural seating position as shown in Figure 10. Note that the image is displayed on the monitor in full high-definition resolution. Although the screenshot captures only a few pixels from the remote monitor, a person at the opposite end of the table has an image of the remote person with a face width of ∼600 pixels, in fact higher resolution than resolvable by their vision at that range (they could resolve at most 200 pixels). The vertical aspect ratio on a 24-in. portrait-oriented monitor allows sufficient height for a life-size head image.

In Figure 11, the remote person is still looking down on the meeting (because their camera is still mounted at the top of their monitor—see Figure 12), although not as much as in the earlier meeting scene depicted in Figure 4 (real) and Figure 6 (simulated). This is a compromise between ideal camera angle and ideal remote viewer relationship to their monitor screen.

Local people can see each other and the remote person. They can see who is looking at whom and also see who is looking at the remote person. Although the view of the remote position for the in-person participants is good, this configuration gives a narrow angle of view for the remote participant (using a 70-degree horizontal FOV camera). Their view is shown in Figure 12.

3.5. Improvement step 2: Providing a better view of the meeting room for the remote participant

The remote view shown in Figure 12 is unsatisfactory in that the remote participant cannot see everyone else. This could be resolved by moving the camera generating the image, but at the cost of reducing image size, particularly of people at the other end of the table. It would also make the setting up more difficult. It is convenient in practice to have a camera mounted on the monitor, and most camera mounts are designed to sit on the top of monitors. The solution that has been adopted for this problem is to provide the remote person with two monitors set up to fill their horizontal FOV in a similar way to that they would experience in person at the meeting, as shown in Figure 13. Two cameras are mounted on the meeting room proxy monitor (shown in the inset in Figure 13) to provide the images for the two remote monitors.

The meeting room cameras are angled to provide non-overlapping images of their respective sides of the table; the remote monitors are angled to match the camera angles of their incoming signals. The monitors have also been moved closer to the remote person, who is sitting closely to the monitor stand, as can be seen in Figure 14. The proximity of the remote person to the monitors is chosen to correctly size the image in their FOV. Note that this leaves the person uncomfortably close to the screens, with little desk space on which to work. That issue could be alleviated by using larger monitors, moved back correspondingly. For the purposes of this discussion, the 24-in. simulated monitor size has been retained because that matches the equipment used in real-life experiments. The seating position is shown in Figure 14.

There are subtleties to the monitor placement shown in Figure 14 when it is examined more closely. Figure 15 was produced by generating outlines of participant heads in a first-person view generated by looking to the left from the proxy position in the meeting room. These white outlines are superimposed on the actual view that the remote participant has, looking to their left. It can be seen that the remote person is presented with an image in which the perspective is flattened—they see the furthest person as larger than they would be in real life. This is surprising at first glance. Because their monitors are turned to move the centre back, it might be expected to see the distant (centre) person smaller than in first-person view because of the additional perspective effect. In fact, the remote person has

FIGURE 14. The remote participant seated with the dual monitor configuration. The close proximity issue is discussed in the text.

FIGURE 15. The remote view of the meeting room with superimposed first-person view head outlines (the white lines) showing that perspective has been flattened or distorted.

FIGURE 16. Superimposed outlines of correct-perspective head sizes, with the remote monitor moved back 0.5 m from the remote participant.

been placed so close to the monitor that they are actually closer to the middle of the joint image than its edges, thus producing the observed and counterintuitive perspective effect. If the monitors are enlarged and placed further from the remote person, there is a sweet spot (back 0.5 m) at which the head sizes look correct. This is illustrated in Figure 16. Note though, that the spacing of the participants is not quite right, leaving the centre person a bit little to the right of where they should be, in contrast to Figure 15, in which the centre person is more left than they should be.

The other aspect of Figure 13 that is of interest is the appearance of the table. The far edge of the table appears as a straight line across the two monitors. Arranging the viewpoint so that

FIGURE 17. Meeting room with the blue participant (no. 1, Figure 10) looking toward the pink participant (no. 3).

this occurs is helpful in providing a realistic view for the remote participant. It happens when the relative heights of the remote person and their screens are such as to match camera and view perspectives. It is easily arranged by raising or lowering the remote monitors, but this does require adjustment to suit each individual participating as the remote person.

The adjustment required when practically setting up the system follows these steps:

- Pitch proxy cameras to centre local participants in vertical view.
- Angle proxy cameras so that their fields of view just do not overlap.
- Angle remote monitors to match cameras.
- Set remote person viewing distance to get correct FOV for in-person attendees.
- Adjust remote monitor heights to give straight back table edge.
- Adjust pitch and angle of the remote camera to centre remote person's image on the proxy monitor.

These adjustments could be a little tedious to get right, but near independence of the settings make it perfectly possible to get a good setup in a single pass.

What are the strengths and weaknesses of the step 2 improvements? The main achievement is that the remote participant has a realistic view of each in-person participant. The geometry is accurate enough for the remote person to be able to see who of the others is looking at whom, at least in the sense of seeing what direction their head is pointing. In Figure 17, the participant in the blue T-shirt is looking at the participant in pink. The first-person view from blue is shown in Figure 18, and the view of this action as seen by the remote participant is shown in Figure 19.

3.6. Improvement step 3: Improving the view for all

The next question is: 'what can local people see of the remote person?' Local meeting room people see the remote person in a position consistent with that person sitting at the table. This means, at least roughly, that the remote person is of the correct scale and at the correct height, and they are in a sensible place 'at the table'. If their audio is sourced from the proxy monitor, then their voice comes from the right place. This should help local people be aware of the remote person and be conscious of the effect on their view if they lean back or forward. There is however, only a single view of the remote person.

FIGURE 18. The first-person view as seen by the blue participant (no. 1, Figure 10) when looking at pink (no. 3) as in Figure 17.

FIGURE 19. The first-person view as seen by the remote participant, when blue (no. 1, Figure 10) is looking at pink (no. 3) as in Figure 17.

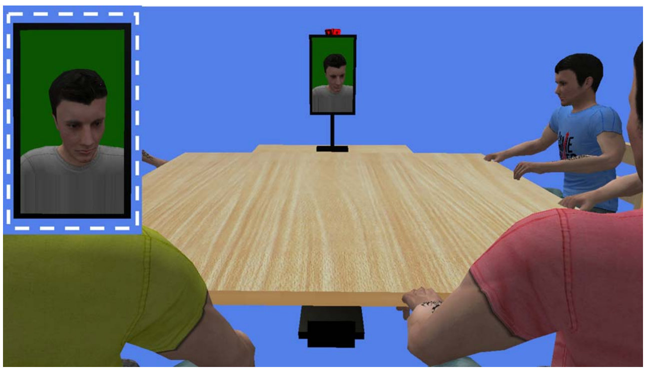

FIGURE 20. The remote participant looking at pink (no. 3) as seen in the meeting room. The inset shows a zoomed-in version of the proxy screen.

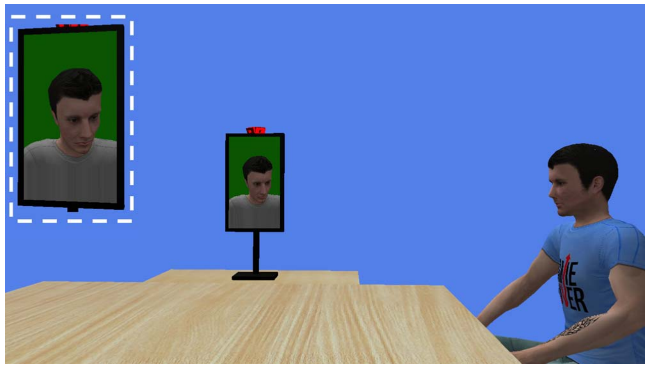

FIGURE 21. The remote person looking at pink (no. 3) as seen by pink. The inset shows a zoomed-in version of the proxy screen from pink's direction.

FIGURE 22. The remote person looking at pink (no. 3) as seen by blue (no. 1). The inset shows a zoomed-in version of the proxy screen from blue's direction.

Viewed from a position between the pink and yellow participants (no. 3 and no. 4 in Figure 10) at the head of the table (as in Figure 18), the image of the remote person can be accurately seen to be looking toward any of the local participants. Figure 20 shows the remote looking at pink (no. 3), as viewed from that head-of-table position.

Unfortunately, viewing a monitor from a position to one side of centre doesn't alter the direction in which a face on the screen appears to be looking. In fact, this is an everyday experience, albeit one that is not typically of concern. When a person is seated to the side, say, watching television, a news reader still seems to be facing them. Shifting the point of view to that of the pink person (no. 3) in Figure 21, or even to blue's viewpoint (no. 1) in Figure 22, always leaves the viewer with the impression that the remote person is looking to their right. As one moves around, the remote image gets narrower, but one compensates for that and doesn't give it much thought. Indeed, the model chosen in the simulation has quite broad cheeks, and their narrower version tends to look like a person with a narrow face. Does the degree of turning change with our viewing angle on the screen? In fact, no. If the face turns 90 degrees, we see it in profile. The narrowed image is in profile, so we see it as turned 90 degrees from any viewing angle. Similarly, turning 45 degrees alters our view of the face in a way we will interpret as a 45-degree turn in the narrowed view, and so on.

To summarize the discussion in relation to step 3, the system so far allows for the remote person being able to see others and see who is looking at whom from a place close (a little high) to a natural seating position. Local people can see the remote person, also in an appropriate position. Local people can see the direction that the remote person has turned, but the appearance is relative to their position, rather than absolute. The error in apparent identification of gaze direction is summarized in Table 6. The local (meeting room) participants in the meeting are numbered clockwise from the left of the proxy position as shown in Figure 10. The columns in Table 6 represent the local participant at which the remote person is directly looking; the rows represent the local participants. The numbers in the table are the differences between apparent view target and actual view target. Half values represent points between participants. View targets can be outside the range 1–6 if the remote person appears to be looking beyond the circle of targets. Clearly, the situation as shown in this table is not ideal.

TABLE 6. Meeting room view error for remote participant's gaze direction with a single proxy monitor in the meeting room. Numbered participant positions are as shown in Figure 10

Error in apparent view (average absolute, 2.5)		Remote person is looking at local person: 1	2	3	4	5	6
	1	−4.5	−3.5	−2.5	2.5	3.5	4.5
	2	−3.5	−2.5	−1.5	1.5	2.5	3.5
	3	−2.5	−1.5	−0.5	0.5	1.5	2.5
Local person	4	−2.5	−1.5	−0.5	0.5	1.5	2.5
	5	−3.5	−2.5	−1.5	1.5	2.5	3.5
	6	−4.5	−3.5	−2.5	2.5	3.5	4.5

TABLE 7. A summary of the ability to identify who is looking at whom with a single monitor representing the remote participant in the meeting room

Who can see who is looking at whom (average gaze error)		Another local	The remote person
A **local** can see that:	One **local** is looking at	✓	✓
	The **remote** person is looking at	2.5	
The **remote** person can see that:	One **local** is looking at	✓	✓

TABLE 8. With one laptop per meeting room attendee, ability to detect who is looking at whom

Who can see who is looking at whom		Another local	The remote person
A **local** can see that:	One **local** is looking at	✓	✗
	The **remote** person is looking at	Nearby only	
The **remote** person can see that:	One **local** is looking at	✓	✓

The question of who can actually see whom with the mixed-mode system developed so far in this section 2 discussion is summarized in Table 7. A tick in this table means that the 'looking-at' information can be derived just as it would be in a wholly in-person meeting. For example, a local can see that another local is looking at the remote person in just the same way they would if the remote person was actually sitting in the proxy position.

The challenge is to try to reduce the view error, highlighted by Tables 6 and 7, without losing the gains already made, ideally to get all the numbers in the view error table to zero. The simplest way to get the correct view for each local person is to provide separate images of the remote person for each. This could be done by using multiple cameras or by image manipulation of the image from one or a small number of cameras. Further image manipulation has not been considered because the goal of this system is to maximize the quality of personal interaction. At the current state of the art, it is conceivable that artificial or manipulated images could look quite realistic, but it doesn't seem likely that they could convey a real-time accuracy of expression. It is not unreasonable however, to consider one camera at the remote end for each local person and corresponding video feeds. This would require a number of cameras and bandwidth roughly the same as that used for a fully virtual meeting—currently widely used and considered to have an acceptable cost. The difficulty is finding a way to display those images that preserves the meeting layout. A possibility is to use a display in front of each meeting room participant—for example, each participant having a laptop computer in front of them. This would provide an accurate view of the remote person to each local person, clearly showing their view direction. However, the system would not allow a local participant to determine which other locals the remote might be looking at, because the way in which the view direction would be presented is relative to the observer (local person) in the sense of being pointed directly at them (good) or turned away left or right to a degree depending on the relative position of the target of their gaze (not so good). Such a relative orientation can be interpreted for movements along a line of people sitting side by side—looking slightly to the left means looking at the person sitting on the left; a bit further means the next person along. It is the way in which we normally interpret the gaze direction of a person sitting on the opposite side of a table when present in person. The angles don't work for people not on our side of the table. More importantly, there is no capability for seeing who is looking at the remote person. When a local looks at the remote person, they look straight ahead and direct their gaze a little downward. Table 8 summarizes the view situation, assuming that the camera solution of step 2 delivering images to two monitors is present, as well at the monitors for each local.

To preserve the capability for a local to see if another local is looking at the remote person, a solution is needed in which the image(s) of the remote person are (all) located in the proxy position. Interestingly, there is technology that provides this capability; autostereoscopic displays are capable of providing different images at different viewing angles ('Autostereoscopy,' 2023). Their primary application is to provide 3D content without the need for viewers to wear special eyewear, but they have also been recommended for videoconferencing applications; Lincoln *et al.* (2010) demonstrated a system that allowed for two separate images on a single monitor. Systems with multiple viewing angles exist, but repeat pairs of images across their horizontal view span. Autostereoscopic displays have been produced commercially but have not become sufficiently popular to be readily available or have low cost.

In the context of this work, of looking at options using readily available equipment, it was chosen to extend the practical system by a modest increase in video channels. Specifically, in step 3, the proxy monitor is replaced with a pair of monitors, angled to focus participants' attention on the monitor giving them the most relevant view. This won't give local users exact information on the remote person's view direction, but is intended to reduce the view error.

This solution is illustrated in Figures 23, 24 and 25. Two cameras are needed at the remote end to feed the (now) two proxy monitors. Figure 23 shows the remote person's setup, unchanged from step 2 except that cameras are now positioned over the centre person on each side: green on the left and brown on the right. As in step 2, the remote person's view enables them to see the view directions of locals, as shown in Figure 24. The

FIGURE 23. With step 3, the remote participant now has two cameras.

FIGURE 24. The remote view, with blue (no. 1) looking at remote and green (no. 2) at pink (no. 3).

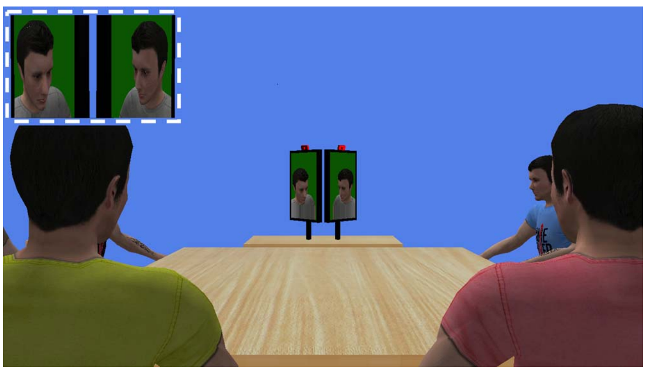

FIGURE 25. With step 3, the meeting room now has two screens in the proxy position of the remote participant. The inset shows zoomed-in versions of these images as seen from the head of the table.

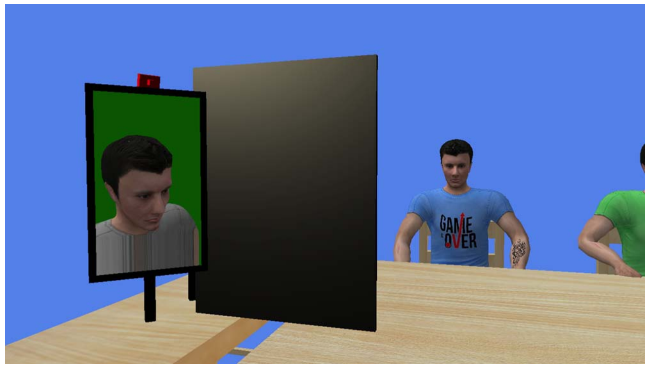

FIGURE 26. A baffle placed between the two proxy screens to block the view of the 'wrong' monitor, in this case as seen from seat 6 (black).

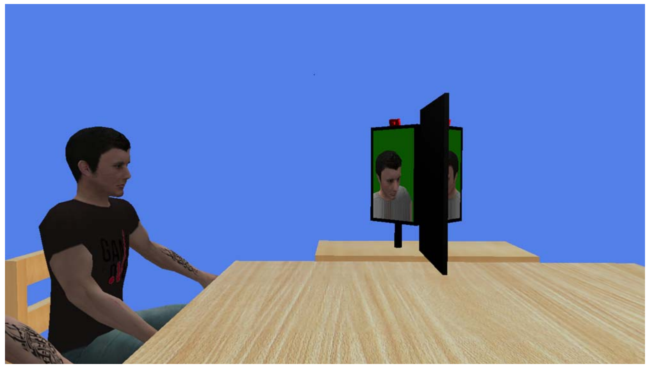

FIGURE 27. The baffle as seen from the position of seat 4 (yellow).

TABLE 9. Meeting room view error for remote participant's gaze direction with a dual proxy monitor in meeting room, as in Figure 25 25. Absolute average error is 0.7. Numbered participant positions are as shown in Figure 10

Local person	Remote person is looking at local person					
	1	2	3	4	5	6
1	−1	−1	−1	−1	−1	−1
2	0	0	0	0	0	0
3	1	1	1	1	1	1
4	−1	−1	−1	−1	−1	−1
5	0	0	0	0	0	0
6	1	1	1	1	1	1

corresponding local meeting room view from the end of the table is shown in Figure 25. Note the two different views of the remote person. The idea here is that each local participant should look at the monitor on their side of the table. Placement of the monitors still puts the image of the remote person in an appropriate proxy position. The angle of the monitors mean that locals 1 (blue) and 6 (black) can see only the appropriate monitor (refer back to Figure 10 for participant locations). Locals 2 (green) and 5 (brown) can't see very much of the wrong monitor, but locals 3 and 4 (pink and yellow) can see both monitors, the correct one being only slightly better aligned for their gaze. Participants' gaze can be directed to the appropriate monitor by placing a baffle, as shown from seat 6 (black)'s point of view in Figure 26 and in Figure 27 from 4 (yellow)'s perspective. Placement of the baffle is a compromise between having an obstruction between the front two seats and a block for the end seats. The position chosen does not completely block the wrong monitor from the end seats, but it does make the right one the more obvious to watch.

If the remote person looks directly at the camera on their left, then local no. 2 (green) will correctly be the target of their gaze. Of course, locals no. 1 and no. 3 (blue and pink) will also see remote looking at them; no. 4, 5 and 6 (yellow, brown and black) will however see the remote person looking somewhere on the opposite side of the table. The result is not perfect, but is a considerable improvement on the situation with a single proxy monitor. A person seated to the right of a centre person will see one gaze to the left of where it should be; a person to the left will see one gaze to the right. These errors are presented in Table 9.

The final result for this step 3 improvement, the implementation of the dual proxy monitor, is summarized in Table 10. The

TABLE 10. The overall 'who is looking at whom' summary for the dual proxy monitor implementation

Who can see who is looking at whom (average gaze error)		**Another local**	**The remote person**
A **local** can see that:	One **local** is looking at	✓	✓
	The **remote** person is looking at	**0.66**	
The **remote** person can see that:	One **local** is looking at	✓	✓

gaze error is never worse than 1 for a meeting with six people, with an average positional error of 0.66. If the remote person seems to be looking at a person, two-thirds of the time they may in fact be looking at that person's neighbour. The remote person is spatially part of the meeting and has a close to normal view of what is going on. Only ordinary conferencing equipment is in use and only twice the video bandwidth of a picture-on-the-wall conference is required. The visibility data are easily tested in simulation. Whether or not the system provides a satisfactory conferencing setup, especially for the remote participant, must be judged experimentally.

3.7. Practical trial

Following this analysis, a new version of prototype setup has been constructed, matching as closely to the ideal geometry as possible. A preliminary evaluation has been conducted, with two groups of participants running through sets of the three games. Selection of participants was not ideal because of lockdown conditions. Essentially only researchers and friends took part. The results were similar to those of the first trial for Counting and Press Conference. Winking Murder worked better in that games were played successfully, implying that estimates of who was looking at whom were improved. This evaluation is continuing.

4. CONCLUSIONS AND FURTHER WORK

This article has explored issues associated with hybrid videoconference meetings, where many of the participants are colocated, whereas some are remote. Recognizing the issues for a remote participant—those of a sense of relative isolation and lack of direct engagement with the group through conventional social tools such as eye contact—section 2 of the article began by describing the evaluation of an experimental setup specifically attempting to address and assess some of these issues for a single remote participant. This approach, using standard videoconferencing displays and cameras, aimed at (i) improving the remote participant's view of the others in the meeting and (ii) improving the perception of the remote participant's presence by the others in the meeting. This was achieved using pairs of displays and cameras arranged in a wedge formation. The initial experiments, which used simple games to explore engagement and interaction, showed promise in the approach. However, they also highlighted the need for improving the perception of who was looking at whom, both for the remote and group participants, through more careful experimentation with positioning and angles of the cameras and displays used in the setup.

Section 3 of the article then approaches these issues by analysing the layout and organization of the meeting room itself and the setup of a remote participant's workstation. This focuses as much as possible on conventional videoconference displays and cameras and explores the geometry in an attempt to provide as realistic a view as possible for the remote participant of the meeting room and for the meeting room participants of the remote person to reinforce effective presence. From section 2, the challenge of both types of participant being able to determine 'who is looking at whom' was the highlighted task. The analysis has shown that although a perfect solution may not be possible, significant improvements can be achieved by careful consideration of this geometry and layout, particularly for the size of the group considered here (six colocated participants, one remote). Subsequent experimental evaluation has supported these concepts.

It is suggested that even as shown here in section 3, without further development, the proposed setup, which does involve remote participants having two displays, provides sufficient improvement in engagement for the additional cost to be fully justified. This is particularly true, given the current trends toward increased remote working, even postpandemic. Further research on the configuration, both in terms of hardware and software, is continuing.

References

op den Akker, R., Hofs, D., Hondorp, H., op den Akker, H., Zwiers, J., & Nijholt, A. (2009). Supporting engagement and floor control in hybrid meetings. In: Esposito, A., Vích, R. (eds) *Cross-Modal Analysis of Speech, Gestures, Gaze and Facial Expressions*. Lecture Notes in Computer Science, vol 5641. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-03320-9_26

Autostereoscopy (2023). *Wikipedia*. Retrieved January 30, 2023 from https://en.wikipedia.org/w/index.php?title=Autostereoscopy&oldid=1135401644

Bailenson, J. N. (2021) Nonverbal overload: a theoretical argument for the causes of Zoom fatigue. *Technol Mind Behav*, **2**. Retrieved from. https://doi.org/10.1037/tmb0000030.

Grønbæk, J. E., Saatçi, B., Griggio, C. F. and Klokmose, C. N. (2021) MirrorBlender: supporting hybrid meetings with a malleable videoconferencing system. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21). Association for Computing Machinery, New York, NY, USA, Article*, **451**, 1–13. https://doi.org/10.1145/3411764.3445698.

Jackson, S. P. (2021) Three bodies: problems for video-conferencing. *Phenomenol Mind*, **20**, 42–50. http://journals.openedition.org/phenomenology/406.

Kuzminykh, A. and Rintel, S. (2020) Low engagement as a deliberate practice of remote participants in video meetings. In *CHI Conference on Human Factors in Computing Systems*, pp. 1–9. ACM, Honolulu.

Lincoln, P., Nashel, A., Ilie, A., Towles, H., Welsh, G. and Fuchs, H. (2010) *Multi-View Lenticular Display for Group Teleconferencing*. In The 2nd *International ICST Conference on Immersive Telecommunications*. Berkeley.

Linden Research. (2023). *Second Life*. Retrieved January 30, 2023 from https://secondlife.com/

McVeigh-Schultz, J. and Isbister, K. (2021) A "beyond being there" for VR meetings: envisioning the future of remote work. *Hum Comput Interact*, **37**(5), 433–453.

Wink Murder (2023). *Wikipedia*. Retrieved January 30, 2023 from https://en.wikipedia.org/w/index.php?title=Wink_murder&oldid=1123549999

Perfect Party Games (n.d.). Retrieved January 30, 2023 from https://www.perfect-partygames.com/press-conference.html

Rogers, B., Apperley, M. and Masoodian, M. (2021) BubbleVideo: Supporting Small Group Interactions in Online Conferences. *International Conference on Human-Computer Interaction*, C. Ardito *et al* (Eds). vol. 12933, pp. 67–75. Springer, Bari.

Saatçi, B., Rädle, R., Rintel, S., O'Hara, K., Nylandsted Klokmose, C. (2019). Hybrid meetings in the modern workplace: stories of success and failure. In: Nakanishi, H., Egi, H., Chounta, IA., Takada, H., Ichimura, S., Hoppe, U. (eds) *Collaboration Technologies and Social Computing. CRIWG+CollabTech 2019*. Lecture Notes in Computer Science, vol 11677. Springer, Cham. https://doi.org/10.1007/978-3-030-28011-6_4

Shockley, K. M., Gabriel, A. S., Robertson, D., Rosen, C. C., Chawla, N., Ganster, M. L. and Ezerins, M. E. (2021) The fatiguing effects of camera use in virtual meetings: a within-person field experiment. *J. Appl. Psychol.*, **106**, 1137–1155.

Standaert, W., Muylle, S. and Basu, A. (2022) Business meetings in a postpandemic world: when and how to meet virtually. *Business Horizons*, **65**, 267–275.

Strasburger, H. (2020) Seven myths on crowding and peripheral vision. *i-Perception*, **11**, 204166952091305–204166952091346.

Xu, B., Ellis, J. and Erickson, T. (2017). Attention from afar: simulating the gazes of remote participants in hybrid meetings. In: *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17)*. Association for Computing Machinery, New York, NY, USA, 101–113. https://doi.org/10.1145/3064663.3064720

Zoom Video Communications Inc. (2023). *Zoom*. Retrieved January 30, 2023 from https://zoom.us/