Seconds Matter: Improving Distributed Coordination by Tracking and Visualizing Display Trajectories

Mike Fraser, Michael R. McCarthy, Muneeb Shaukat and Phillip Smith
Department of Computer Science, University of Bristol, Merchant Venturers Building, Woodland Road, Bristol, BS81UB, UK.
{fraser, mccarthy, muneeb}@cs.bris.ac.uk
+44 (0)117 9545144

ABSTRACT
Pauses in distributed groupware activity can indicate anything from technical latency through infrastructure failure to a participant’s thoughtful contemplation. Unraveling these ambiguities highlights mismatches between unseen off-screen activities and on-screen cursor behaviors. In this paper we suggest that groupware systems have typically been poor at representing off-screen activities, and introduce the concept of display trajectories to bridge the sensor gap between the display and its surrounding space. We consider requirements for display trajectories using the distributed social scientific analysis of video data as an example domain. Drawing on these requirements, we prototype a freeform whiteboard pen tracking and visualization technique around displays using ultrasound. We describe an experiment which inspects the impact of display trajectories on remote response efficiency. Our findings show that visualization of the display trajectory improves participants’ ability to coordinate their actions by one second per interaction turn, reducing latency in organizing turn taking by a ‘standard maximum’ conversation pause.

Author Keywords
Display Trajectory, Off-Screen Tracking, Pen-based Interaction, Groupware.

ACM Classification Keywords
H5.2. [User Interfaces]: Input devices and strategies; J.4. [Computer Applications]: Social and Behavioral Sciences.

INTRODUCTION
Co-present situations are often regarded as optimal for high-quality real-time collaboration, because of the subtlety of mutual awareness available between people as they work together, often with a variety of media and materials. Such situations offer the ability to annotate, animate, share, juxtapose and mutually refer to information both over and on screens and documents. Workplace studies highlight the ability of people to rely on others’ capacity to see, understand and respond to these behaviors [20]. Pointing at, pointing towards, imitating, showing and exaggerating features of interest for others are complex practices, but all draw on the assumption that they are intelligible to whomever might take the next interaction turn. Significantly, if there is a pause in proceedings the cause of no next turn can be identified – an unexpected interruption, a moment of careful thought and so on, are all noticeable phenomena in co-present situations.

As new forms of distributed system emerge, such moments become less intelligible because the awareness relationships between places, people and materials of relevance are ‘fragmented’ [8]. In this paper, we outline our approach to sustaining distributed coordination in the face of off-screen events by providing an additional information channel between users that we have called the display trajectory. This channel works by tracking the user’s movements around a display and then showing them to remote participants. Our intention is to provide distributed users with enough information about remote movements that they can predict to some extent when remote off-screen users are about to interact on-screen. Our hope is that this will allow them to better implicitly organize, interleave and coordinate their actions. In the following sections, we describe the domain of distributed social scientific video analysis that motivated our technique and review related work.

ANALYSIS DISPLAY TRAJECTORIES
In the MiMeG project [7], we are investigating how to support social scientists in distributed collaborative analysis of video data. Clearly in an ideal world data analysis can be conducted co-presently, but personal constraints along with research funding increasingly directed to distributed teams means that typically researchers need to share and inspect research data across geographical distances at the same time. Simultaneously, digital video is becoming an invaluable tool for social and cognitive scientists to capture and analyse a wide range of social action and interactions. As a result,
video-based research is increasingly undertaken by research teams distributed across continental and worldwide institutes, yet there are few tools explicitly designed to support distributed real-time analysis of video materials.

As a result, we have been developing groupware to support skilled analytic work with digital video data. To facilitate this work, we have undertaken studies of existing analytic practice to develop requirements for the design of the tools, including detailed studies of video-based research in both social scientific and professional communities. These have included reflections on co-present ‘data sessions’, or physical gatherings to share perspectives on video data [7]. We argue that collaborative analytic work in qualitative social sciences requires significant skill. Researchers encounter new challenges in reconciling perspectives on action, perspectives which are key to developing a joint understanding of the phenomenon of interest. Viewing data on its own fragments the relationship between remote analyst and remote data making communicating and interpreting insights difficult.

![Figure 1. Projected MiMeG interface](image)

Our first prototype (Figure 1) attempted to meet some needs of conducting co-located video data sessions remotely. We began by designing a groupware system which played video data at multiple sites simultaneously, included support for real time audio, and an electronic whiteboard marker system [23]. Researchers at each site could annotate the video data for each other during discussions. By providing the ability to draw on the screen, we hoped analysts would be gesturally explicit about their perspectives, drawing a particular event or area of analytic interest to focus. We encountered a number of issues with this approach which are detailed, along with our initial study of co-present analysis, elsewhere [7]. In this paper, we are particularly concerned with just how difficult remote users found a pen-to-screen approach. Whilst co-located researchers could see the analyst prepare to produce a pen stroke in front of the screen, researchers at remote sites were only made aware of the stroke at the time it was being produced. This was often too late to determine the annotation’s contextual relevance, because users were finding out where and when the annotations were being produced simultaneously with trying to understand their relationship with the video data, all this during which the data itself was changing in playback. Seeing a particular subject required being ready to notice when and where it would be highlighted in order to simultaneously understand what motivated its selection. In other words, video annotated during playback places critical demands on approach and timing for designing and understanding annotations. Typically, this display trajectory is lost on remote participants, with the result that the substance of the annotation is missed completely at the appropriate time, or else the cross-site interaction slows, simplifies and repeats while researchers have to talk others explicitly through their activities, to the frustration of both remote and co-located participants.

Some groupware solutions have drawn on rich representations of the remote space to show user-display relationships, such as video views. However, video is notoriously difficult to interrelate coherently with shared digital data. For example, one growing form of real-time collaboration software based on video is the Access Grid [2], which is becoming a widespread tool for conducting remote meetings, presentations and so on. However, if we consider the use of Access Grid Nodes (AGNs) to support expert collaborative analysis of video data between remote sites, problems start to emerge. We might transmit both video of the analyst(s) and views of the data under scrutiny to remote participants, but without more explicit representation of analysts’ mutual orientation to one another and to the data, the ability to share perspectives is undermined.

As a result, previous work has attempted to capture and preserve the relationship between distributed users, their materials and displays. Support for groups of users working with and around digital artefacts has been studied in a variety of remote communication types, from video media spaces [8] through collaborative virtual environments [14] to remote embodiment through robot proxies [13]. In these cases, preserving the relationship between action and object across sites is critical to supporting the work of users. This issue is highlighted by Tang and Minneman [34], who demonstrate how video shadows of off-screen distributed pairs of users can support collaboration in a groupware drawing tool. However, they also report that accurate pointing can be difficult using this technique due to blurring of the body at a distance, as well as local group occlusions and confusions due to multi-user blurring. Whilst we do not plan to use video or shadows to indicate off-screen behaviours for these and related reasons, the qualitative improvements observed are an important motivating factor.

Some researchers have highlighted difficulties particular to groups using groupware. Tang et al. [31] use the terms display disparity and presence disparity to describe coordination difficulties in mixed presence (i.e. group-to-
Related work by Kirk and Stanton Fraser on mixed ecologies [18] describes how collaborative physical tasks can be supported by the transmission of bird’s eye video views of participants’ hands over views of physical artifacts. Clearly, whilst best practice in working with physical objects may vary from that relating to digital data, the point that freeform interaction techniques are valuable for complex and rich forms of task performance remains for our case. Tang et al. [33] also project bird’s eye views but extract arm movements using image processing techniques and map them over a groupware application. These techniques are particularly designed for group-to-group applications, although they do not explicitly illustrate distance from the screen and therefore the pace of approach to the display is hard to gauge, an issue which is particularly critical in our case.

As well as these groupware designs, there are single-user interface techniques designed to support co-present collaboration which we might draw upon for enhancing distributed coordination. Large displays have often been used to provide mutual information and peripheral awareness [21], or to share information among a group [16, 24] often using Single Display Groupware (SDG) techniques [30]. Indeed, the properties of traditional shared surfaces such as table tops and whiteboards have often been used to motivate the design of shared display systems (e.g. [26]). All these techniques capitalize on untracked display trajectories for supporting co-present interaction.

Also important are techniques which enable freeform gestures and annotations between displays (e.g. [28, 15]), although these techniques focus on matching on-screen behavior across multiple displays, rather than explicitly determining off-screen interaction. Li et al. [19] also describe techniques for switching between ink and gesture modes, although they rely on on-screen data for performing both of these, rather than separating them into on-screen and off-screen behaviors. Tano et al. [35] suggest combining off-screen 3D gestures with on-screen 2D sketches, and apply these techniques to single-user applications such as creative design.

Relevant for our interest in off-screen activity is the Hover Widget technique [9] which uses an alphabet of pen movements above the screen in the tracking state to perform a variety of additional functions available without altering the pen’s focus. This work found significant time reductions enabled by selecting new functions without moving the focus of attention. Following work with remote pointing devices for large displays [1], Parker et al. [25] also show off-screen tracking efficiency gains for pointing at targets with a magnetically-tracked stylus. We are also interested in an efficiency gain using off-screen tracking, although to enhance reactions in collaborative situations, rather than actions in single-user applications.

Drawing on these techniques and our requirements led us to explore the significant benefits to be gained by tracking and visualizing display trajectory data. We require visualizations that support the richness of freeform skilled work as with the original on-screen electronic whiteboard markers, but also ensure predictability of the on-screen annotations so that the implicit position and timing of a new contribution or turn at the display can be judged. To enable both richness and predictability, we need to consider how to capture and represent the position and pace of off-screen movements. Our review of previous work also shows it would enhance our design to scale to large numbers of co-located displays and annotation devices. In the following section we describe our prototype solution to these requirements, and then describe an experiment which demonstrates how effective our solution is at supporting coordination.

TRACKING AND VISUALIZING TRAJECTORIES

We have created a prototype system which can approximately convey the trajectory of electronic marker pens with respect to projected displays. Our solution requires two components: tracking the device’s location, and visualizing that information on remote displays. In the following section we describe how the pens are tracked, and we return to visualizing trajectories in the next section.

Tracking Trajectories

Our original electronic whiteboard system triangulates a marker pen in two dimensions. The marker pen emits ultrasound and infrared signals when the nib is pressed against the display surface and these are used to calculate location to a high degree of precision. To create a system capable of tracking the pen around the display, we have used a recently developed system for three-dimensional ultrasonic positioning [22, 27]. Six transmitters send ultrasonic pulses which are received and processed on a receiver attached to a Gumstix embedded computer [10]. The Gumstix uses a time differential to calculate the receiver’s position with respect to the known transmitter locations. The hardware cost of this system compares favourably with the existing electronic whiteboard marker system, at around $150 for the transmitters which surround the screen, and $225 for the embedded computer, receiver and battery, all attached to each marker pen.
We calibrated the tracking system with all six measured transmitter positions to achieve good overall coverage and moderate effects of intermittently obscured transmitters. Testing showed that good coverage of the region in front of the display was reached by angling four transmitters at corners inwards towards the centre of the display raised by 45º, and two vertically down from the ceiling (Figure 2). The transmitters have an active cone of 140º allowing a large area to be covered. Multiple pens can be tracked with no reconfiguration of the transmitters or hardware.

Figure 3 illustrates our original electronic whiteboard marker pen enhanced with additional components to support the ultrasound tracking system. A battery powers the Gumstix for approximately eight hours of constant use. The Gumstix provides a 200MHz Intel XScale processor, running Linux, which receives data from the attached ultrasound receiver, calculates position, and then sends on data over Bluetooth to a machine running the MiMeG groupware. This data can then be streamed by the application to create an appropriate visualization on displays at remote groupware clients.

During our initial testing of the system, we stabilized the tracking and investigated the interoperation of our tracking with the marker pen’s own system.

Stabilizing Tracking
At this prototype stage we have not directly integrated hardware/firmware for the two tracking systems. In any case, they are designed with different goals and accuracies. The pen’s own system enables input to be more accurate when the pen is in contact with the screen, and the receiver relies on infrared signals to determine latency of the ultrasound signal and provide high levels of accuracy in two dimensions (approx 40 dots per centimeter). On the other hand, our additional components work purely on ultrasound signals (reducing transmitter/receiver hardware costs) and only provide a minimum accuracy of 10cm \( (p<0.05) \) [22]. Error in reported position consists mostly of high frequency environmental noise whilst the pen’s true position consists mostly of lower frequency information, so we implemented an FIR low-pass filter to stabilize and improve the tracking accuracy. The filter proved effective with imperceptible latency and very little noticeable noise present in the final signal, with an approximate accuracy of 3cm. The smoothing effect of the filter (blue dots) on a typical set of raw data (red dots) from the tracking system is shown in Figure 4 (coordinates in meters).

Interoperating Tracking
When the marker pen nib is pressed on the display screen, it emits its own 2D ultrasound signals. The 3D positioning system also receives these signals as they operate over the same frequencies. This causes the 3D receiver to lose the correct position of the pen. However, the whiteboard receiver is not reciprocally affected by the transmitters of the 3D positioning system, due to the difference in intensity of the pulses transmitted. This means it is possible to support on-screen tracking using the 2D system when the nib is activated, and off-screen tracking using the 3D system without interference when the nib is deactivated. We wanted to exploit this feature of the tracking interoperation, so we explored the behavior of the interference further. Interference is largely independent of the pen’s position in relation to the 3D transmitters, and of the pen’s current motion. The pen’s position is reacquired in under a second of the interference finishing provided the transmitters are not occluded. We configured our 3D tracking against erroneous readings by forcing the tracking software to enter a state where it tries to
reacquire the pen’s position to a very high degree of certainty when it detects very large changes in position readings. Once the pen stops emitting the 2D signal and the interference passes, valid pulses are received allowing tracking to resume and the position is reacquired correctly. Configuring the 3D tracking system to self-repair independently of the 2D system’s activation means we can support more than one display with the same 3D tracking hardware, although this would depend on the interference not propagating to all 3D transmitters, thresholds for which we have not yet explored. Thus far we have written software to calculate and transmit the position of any number of independent pens via their Bluetooth aerials to use in front of any number of distributed displays.

**Alternative Tracking Methods**

We have been working with a semi-opaque solid back projected screen because it works well with the whiteboard markers at relatively low cost. Clearly our approach could draw on any 2D/3D tracking combination that interoperates successfully. For example, we could use touch-sensitive smart boards or touch screens as the 2D displays, and we could use magnetic tracking systems such as Polhemus to obtain 3D data. Indeed this approach would be more simple because there would be no interference between the on-screen and off-screen sensors. Equally, incorporating our 3D ultrasound receiver into a capped Bluetooth-enabled pen (such as the Nokia SU-27W) could allow us to use it on a touch screen. So, our approach deliberately represents the most difficult case with regard to tracking signal integration. However, this approach also has the benefit that the tracking systems would be the most cheap, coherent and effective to combine for potential further hardware integration.

Our transmission signal is independent of the number of receivers whereas many touch screens and electronic whiteboard systems are not, so unlike many existing tracking systems for displays our approach is scalable to large numbers of co-located users each holding their own input device. However, our approach is still constrained by the whiteboard system to one input device in contact with the screen at any one time, and interference effects across larger multi-display distances remain unexplored. Nonetheless the system could be used with multiple pens and also across multiple displays enhancing existing screen-to-screen techniques [15, 28, 29]. This could be achieved either by deploying multiple sets of transmitters and thresholding signal strengths to select transmitter groups. It would also be cheaper but have greater opportunity for transmitter occlusion to use a single six transmitter system for a large space, and calibrate the application for multiple displays in known locations within that space.

**Trajectory Visualizations**

As well as developing hardware and software for tracking pens’ locations around displays, we have considered how to visualize display trajectories at remote clients. Based on our investigations in our example domain of distributed video analysis, we wanted display trajectory visualizations to have the following properties:

1) Show off-screen position accurately
2) Obscure as little of the video data as possible
3) Display the pace of off-screen positional changes

After initial testing, we decided that these objectives were conflicting in some cases – more trajectory information typically required obscuring more of the screen. We particularly wanted to illustrate pace (i.e. rate of changes in distance from the display) well as this would show a user’s transition from standing back and observing to using the pen on the screen and directly taking a turn. We settled on three visualizations which balanced these goals in different ways (Figure 5).

![Figure 5. From left to right: Crosshair (V1), Orb (V2) and Streamer (V3) visualizations](image)

The first visualization is a crosshair spread over the entire display area with the opacity of both axes scaled according to distance away from the board (x and y map to their axes, z maps to opacity). Crosshair was primarily designed to fulfill objective 1, i.e. to show current screen position accurately. It is relatively unobtrusive, but does not display trajectory pace well as the change in fade is not highly visible and no other changes in distance from the display are shown over time.

The next visualization is an orb filled with a gradient color. Both the opacity and radius are scaled according to the distance away from the board, becoming smaller and less opaque the closer the pen is to the display (x and y map to centre position, z maps to opacity and radius). Orb was primarily designed to fulfill objective 2, i.e. to obscure as little of the data being inspected as possible, particularly when at high opacity a good distance from the display. It is reasonable at illustrating position and displays some pace information, with changes in distance from the screen visible in two ways – both radius and fade of the orb.

The final visualization uses streamer trails that leave a history of circles of varying radius representing the distance away from the board and an increasing opacity along its length fading from most recent to the least recent position (x and y map to position, z maps to radius, time maps to fade). Streamer was primarily designed to fulfill objective 3, i.e. to show the pace and timing of movement well. It does not illustrate position well because multiple positional representations overlap, and it obscures more of the screen than the other two approaches.

We expected any of these visualizations would improve the predictability of the display trajectory’s position and pace. We also anticipated that as a result of these trade-offs
visualizations would have benefits over one other. For assisting in judging the position of the trajectory, we expected Crosshair to be better than Orb and Orb to be better than Streamer. On the other hand, for assisting in judging the pace of the trajectory, we expected Streamer to be better than Orb and Orb to be better than Crosshair.

EXPERIMENT

Aims and Rationale
We conducted an experiment to compare the impact of our three visualizations. We have proposed that display trajectories may enhance the richness and prediction aspects of coordination. For example, richness of interaction might be enhanced by tracked visualizations adding to the complexity and detail with which users can gesture or produce subtle nonverbal cues for others. In future work we plan to explore support for interaction richness, and we return to that idea in the discussion section. However, in this experiment we only investigated support for predictability, specifically whether the ability to judge trajectory pace and position (i.e. when and where another’s turn ends) is enhanced. We were therefore interested in any benefits to response accuracy and response time that each visualization might provide. We have numbered visualization conditions V0 (no visualization), V1, V2 and V3 (as in Figure 5). As per our expectations, we thought that any visualization would improve response accuracy (predicting position) and response time (predicting pace) compared to no display trajectory. We also expected that they would have an order from best to worst for response accuracy (V1, V2, V3) and for response time (V3, V2, V1).

Hypotheses
Formally stated, our three visualization techniques were evaluated against no technique and against each other for response accuracy and time. We defined the following hypotheses:

H1: Each visualization will reduce response time compared to no visualization (V0 vs V1, V0 vs V2, V0 vs V3).
H2: Each visualization will reduce response time compared to the last (V1 vs V2, V2 vs V3).
H3: Each visualization will increase response accuracy compared to no visualization (V0 vs V1, V0 vs V2, V0 vs V3).
H4: Each visualization will reduce response accuracy compared to the last (V1 vs V2, V2 vs V3).

Method

Participants
Twelve participants (5 female) participated in the study. The average age of the participants was 28 years, ranging from 20 to 40. All participants had normal or corrected-to-normal vision. We attempted to recruit participants with height as close as possible to 1.80m. This was the height of the researcher who recorded the tracking data for the experiment (see below). We hoped similar heights would reduce the effect of varying arm reach on performance. Participants’ mean height was 1.77m (min=1.69m, max=1.89m, σ=0.07m). To check the impact of height on our findings, we performed a Pearson Correlation on participants’ height against individuals’ response times, and found no correlation (r=-0.356, p=0.256). Each of the twelve participants was tested for all ten data points over all four conditions, resulting in 120 data points per condition, 480 in total.

Set-up and Materials
We wrote additional software to record and replay 3D ultrasound data. Our system created logs by recording ultrasound readings from the mobile ultrasound receiver and then writing them to a file. It was then possible to re-inject these logs into the MiMeG software as if they were live ultrasound data to simulate the actions of a remote user. This approach allowed us to reproduce movement visualizations for each participant, without the artificial impression that randomly generated data might give. We recorded forty sets of ultrasound-tracked movements around a large projection screen ending in a dot being placed on a random point on the screen (see procedure below). The recorded movements included ‘false start’ movements towards the display and near-display ‘hovering’. This was to avoid predictable linear positioned and paced trajectories towards the display and therefore make the task more realistic and challenging. The movement data was recorded independently of the visualization which would be used to display it, and then each visualization condition V0 (i.e. no visualization) to V3 was randomly assigned ten sets of the movement data.

![Figure 6. Experiment system configuration](image)
Clients as normal due to the additional complexity this would introduce into the experiment.

Dependent variables of response time and response accuracy were automatically calculated and logged on the participant’s computer. We recorded distance in pixels between the simulated dot produced by the recording and a dot the participant placed on their screen at a constant resolution. We also recorded time elapsed (to the accuracy of the computer’s clock, i.e. nanoseconds) between the simulated dot event and the participant’s dot placement. Of course, there were latencies in the communication pathway to the logging computer (via the electronic whiteboard receiver over USB2 to machine 2). These latencies were insignificant in comparison to the measurements, and stable for the experiment. We piloted the study prior to conducting the full experiment, which raised no unforeseen problems.

Procedure
Participants were given the adapted electronic whiteboard maker and shown how it could be used to make marks on the screen in front of them. They were told that someone else standing at an equivalent screen in another room was also going to be making marks on their screen. They were also told that some visual representations might give hints about the other person’s movements. Before the task started, the participant was shown visualizations of their own pen to help them understand how someone else’s visualization reflected their movement.

The task set was to play a simple turn-taking game called ‘spot the dot’. The participant stood in a room and was told another person stood in another room which was kitted out with the same system. The other person (actually the ultrasound recordings) would place a number of dots on their screen with their pen. The participant was instructed to ‘spot’ these dots on their screen and use the marker pen to put their own dot on their projection screen as close to the first dot and as quickly as possible.

Participants were informed they could take a break or stop the experiment at any time, although none did. They were allowed to hold the pen using either hand (all chose their dominant hand). They were told they could stand wherever they felt was most appropriate in front of the screen for the task. The twelve participants experienced the four visualization conditions in a counterbalanced order, to mitigate learning effects. No trial time limit was imposed, although all participants completed the task within 20 minutes.

Results
Response Time
Overall response time means varied substantially between participants across conditions (from 1.03 to 2.09 seconds). Nonetheless, despite the individual variation across conditions, there was remarkably little relative variation within conditions. The mean response time for each visualization condition ($V_0$ to $V_3$) is shown in Figure 7. A repeated measures (within subjects) Analysis of Variance (ANOVA) was performed on the overall response time data. A main overall effect of Visualization type was observed, $F(3,357)=24.81$, $p<0.001$. Evaluation between pairs using the Bonferroni correction for multiple comparisons showed highly significant response time differences between $V_0$ and $V_1$ ($p<0.001$), $V_0$ and $V_2$ ($p<0.001$) and between $V_0$ and $V_3$ ($p<0.001$). All three visualizations produced reliably faster responses than no visualization, providing strong evidence for hypothesis H1. Indeed, the mean response time for no visualization (2.37s) was just over one second slower than the mean across the other conditions (1.32s). The standard deviation in the no visualization condition (2.236s) was much larger than the other conditions (0.5s, 0.488s, 0.495s).

We also compared the mean response times for visualizations against one another again correcting with Bonferroni. We found significant response time differences between $V_1$ and $V_2$ ($p<0.01$) and between $V_1$ and $V_3$ ($p<0.01$). There was no significant difference between $V_2$ and $V_3$ ($p>0.05$), so H2 overall remained unsupported, although we can say that for reducing response time Orb and Streamer were both better than Crosshair, but Streamer was no better than Orb.

\begin{figure}[h]
\centering
\includegraphics[width=\columnwidth]{image7.png}
\caption{Mean response time for each visualization condition, with confidence interval at $p=0.05$.}
\end{figure}

Response Accuracy
The mean response accuracy for each visualization condition is shown in Figure 8. We also performed a repeated measures ANOVA on the overall response accuracy data. Despite a marginally higher mean distance under $V_0$, no main overall effect of Visualization type was observed, $F(3,357)=0.31$, $p>0.05$, providing evidence against H3 and H4. This result suggests that response accuracy was not affected by visualization presence or type.
DISCUSSION

Remote Representation Effects

Our evidence for H1 indicates that any of our selected visualizations makes a highly significant difference to time taken to respond. In turn, the use of any display trajectory visualization improves the ability of distributed users to anticipate the conclusion of a remote action, and therefore to take their own turn more quickly. This is important evidence that showing remote display trajectories is critical to coordinating with remote users. There are some outlier V0 cases where the dot was simply not noticed at all for a time leading to a much greater standard deviation for this condition. Individuals’ mean response times seemingly varied depending on the tactics employed for the experiment (e.g. whether to follow the visualization closely with the pen hovering above the screen, or to stand back and observe in readiness). Nonetheless, the findings show that tactical variation did not affect proportional within condition effects of the use of a visualization.

Our results partly supported H2. Orb and Streamer visualizations were both significantly better for response time than the Crosshair visualization. We may associate the worse performance of the Crosshair with the difficulty of noticing and gauging its fade representation of distance. This accounts for the significantly better performance of both representations which use radius to represent distance.

We can further say that there was no significant response time difference between Orb and Streamer visualizations. This implies that the use of both fade and radius to represent distance in the Orb condition was redundant, as the Streamer condition only used radius for this purpose. It also implies that the display of historical data in the Streamer condition is redundant, at least for enhancing response time. We suggest, however, that these additional mappings may have uses in supporting the richness of collaborative activities beyond turn taking, and we return to this issue below.

Finally, we found no difference for accuracy between any conditions. Our impression is that these results occurred because the video content of the MiMeG application was not included to help experimental control, and therefore the position of the dot had no contextual relevance. Under normal use conditions, on-screen points might indicate a feature in the video data under inspection, and therefore have positional relevance that dots in empty space do not.

Our response time findings indicate that the use of radius to represent distance of the display trajectory from the screen is a useful solution, as it makes the pace of the trajectory towards the board easy to parse. Both visualizations using this strategy map distances closer to the screen with smaller radii. This gives the impression that as the 2D tracking system kicks in, the visualization turns from a tiny circle into the point of the pen’s first on-screen annotation. Unfortunately, despite the opacity used in the Orb visualization, the reverse is also true. When the pen is a large distance from the screen – potentially reflecting times of user disinterest – the visualization is capitalizing the most screen real estate, and fulfilling least well of the designs our original objective of obscuring as little of the data as possible. We therefore note that these display trajectory visualizations will operate optimally with smaller numbers of users, reflecting other groupware visualization effects [5, 6].

Just Give Me A Second

The basic human visual stimulus response time has long been accepted to be around 0.2 seconds [4]. The lowest mean response time of any participant in our experiment was just over one second. This 0.8 second difference reflects the various latencies beyond basic reaction time in participants’ responses. The first of these is the time to notice the dot. The second is the physical distance to traverse to that dot on the display. The third is the technical event transmission (this is effectively zero compared with the other factors). Both the first two latencies are reduced by using the visualizations.

Firstly, noticing the dot is easier because the pace and position with which the dot will appear are far more predictable from the trajectory’s display, and can be better anticipated. Secondly, and less obviously, the presence of a visualization also reduces the traversal distance to the screen because, as we repeatedly observed during the experiment, participants stand further back from the display in the No Visualization condition to gain a wider visual perspective on the entire screen. In cases where they did not adopt this approach, some response times were even worse as the dot might be initially missed completely if the participant attended a different area of the display than that in which the dot appeared.

Our experiment shows that mean response time without visualization is just over one second slower than with any visualization. We suggest that cumulative one-second latencies will have a significant impact on turn-taking and therefore task focus. One second represents a typical ‘pause tolerance interval’ at which Jefferson [17] shows that co-present interaction repair may be instigated. In distributed settings, such next-turn latencies caused by interface design [14] or by network delay [30] have been shown to disrupt users’ ability to coordinate their activities. Typically, talk is
used in these situations to repair or preempt the latency, but forcing explicit focus on coordination at the cost of the task. Reducing such latencies with trajectory techniques will therefore significantly affect overall task accomplishment.

**Richness and Predictability**

We have looked at the ability of display trajectory visualizations to support the mechanics of coordination between users. Our experiment showed that mapping radius to distance was appropriate, but that the inclusion of ‘trails’ in the Streamer visualization had no significant effect on response time. However, studies investigating visualizations of interaction history such as telepointer traces show that these are useful for gestural interpretation, particularly in light of real-world groupware effects such as jitter and latency [11].

Our interpretation of this apparent disparity is that *predictability of activity and richness of interaction* are not codependent groupware properties. Predictability and richness place different demands on visualizations of remote activities. Gutwin and Penner’s study would indicate that our Streamer visualization may have additional useful effects over the Orb visualization in supporting rich forms of gesturing, particularly under network delay conditions, because it shows interaction history, whilst also retaining the benefits described in this paper. Combining these areas to enhance both the process and the content of collaborators’ distributed work is an important next step. To explore richness issues, we plan to use the ultrasound logging system described in this paper to reconstruct an archive of tracked marker pen movement. This archive could then be correlated with video recordings and explored using techniques comparable to those outlined in [3]. Such an approach would also allow us to inspect social scientists analyzing video data under realistic everyday conditions.

**CONCLUSION**

It is possible to hear the term *display trajectory* as a pathway in space that represents a route to a predetermined destination. Yet the trajectory of producing a real annotation for others over a display is not of a predetermined character. It is a kind of referential practice, which studies of co-present interaction have repeatedly shown to have complex features which are designed moment-by-moment with regard to recipients’ (re-)actions and to referential targets. We have observed examples where the varying pace at which a gesture is produced is itself indicative of the character of a movement noticed in a video. It is problematic to transmit the complete character of how data is seen by an analyst without the ability to directly remotely embody that character rather than transform it, and we have not attempted that here. Our future work will rather look at the extent to which users can configure simple visualizations such as those we have introduced to share complex insights. Such considerations will become increasingly important for understanding the local context in which groupware of many kinds operates, from simple video conferencing to emerging mobile communications. At the cutting edge of these forms of distributed coordination, seconds really do matter.

**ACKNOWLEDGMENTS**

Thanks to our colleagues on the MiMeG node of the National Centre for e-Social Science (ESRC grant RES-149250033) which funded this work, and to our colleagues on the EPSRC-funded Equator IRC for their support.

**REFERENCES**


18. Kirk, D. and Stanton Fraser, D., Comparing remote gesture technologies for supporting collaborative physical tasks, in Proc. CHI 2006, pp. 1191-1200, ACM


29. Rodden, T., Rogers, Y., Halloran, J. and Taylor, I., Designing novel interactional workspaces to support face-to-face consultations, in Proc. CHI 2003, pp. 57-64, ACM.


