Distributing Data Sessions: Supporting remote collaboration with video data

Mike Fraser[1], Greg Biegel[1], Katie Best[2], Jon Hindmarsh[2], Christian Heath[2], Chris Greenhalgh[3], Stuart Reeves[3]

[1] Department of Computer Science, University of Bristol, Merchant Venturers Building, Woodland Road, Bristol, BS81UB, UK.

[2] Work, Interaction and Technology Group, Department of Management, King's College London, Franklin-Wilkins Building London SE1 9NH, UK

[3] The Mixed Reality Laboratory & The Learning Sciences Research Institute, School of Computer Science & IT, University of Nottingham, Jubilee Campus, Wollaton Road, Nottingham, NG8 1BB, UK.

Email: *fraser@cs.bris.ac.uk*

Abstract. The design of distributed infrastructures to support remote collaboration among groups of social scientists raises new computational and networking challenges that Grid developers are currently targeting. Beyond such technical goals, however, the e-Science programme as a whole is increasingly recognizing the critical need for a comprehensive understanding of ordinary day-to-day work in the sciences. We have investigated one particular area of collaborative social scientific work – the analysis of video data. This paper discusses current practices of social scientific work with digital video; describes the resulting requirements for distributed video analysis systems; and outlines our initial programme of infrastructure and interface development to address these requirements as part of the VidGrid project.

Introduction

Recent developments in e-Science research have begun to investigate how computational networks, and in particular Grid systems (Foster and Kesselman, 1999), might be used to support and enhance remote collaboration between groups of scientists. It is proposed that features of Grid technologies, such as distributed resource discovery and 'collaboration support' might enable new forms of scientific work. As such, remote problem solving across social and physical sciences is receiving widespread attention from computer science developers. In practice, collaboration support across the Grid has taken a number of forms, including the development of ontologies of scientific process to form workflows that structure collaborative work (Bechhofer et al., 1999), or the use of multiparty video centres such as Access Grid Nodes (AGNs) to enable remote meetings with data visualizations and presentations (Booth et al., 2002). Nonetheless, significant issues remain for such technologies to provide coherent support for distributed social sciences.

In parallel with such developments, the growth in video analysis in the social sciences has engendered a number of systems designed to support the work of video analysts. However, many of these systems have treated video as an 'add-on' to more conventional text-based analysis software and many more treat collaboration between analysts as an 'add-on' system feature. It

seems critical to us to provide tools that treat video as a principal form of data and collaboration as an assumed form of analytic work.

This paper presents our first round of requirements gathering with a range of users of analogue and digital video across the social sciences. We then describe the implications of our requirements for designing systems to support distributed video analysis, and show how our design for the VidGrid prototype software addresses these issues. We finally explore the limitations of our approach and indicate future directions in which we plan to take our work.

Studies

There is a need to analyse and understand existing work practices in the social sciences in order to tailor e-Social Science technologies for use. Therefore we are undertaking two parallel forms of work to inform the design of our demonstrator tools.

Firstly, we are engaged in video-based interactional studies of collaborative video data analysis sessions, which are commonly known as 'data sessions'. These data sessions involve multiple participants viewing video materials together to work up preliminary analytic issues and themes. We are analysing a number of data sessions to identify key interactional resources that underpin their organisation.

Secondly, we have completed a series of qualitative interviews with expert video analysts from a range of disciplines to explore the ways in which they share data. The interviews took place over a seven-month period and, in total, we have interviewed 26 individuals working in 7 different countries. The interviewees were selected as leading exponents of various forms of video analysis drawn from the fields of sociology, linguistics, anthropology, psychology, education and management. In addition we interviewed a small number of video analysts working in occupations outside of the social sciences in order to draw on their practices and experiences – these included ergonomists, film editors, communications experts, performance analysts and sports scientists. The interviews were organised so that participants were encouraged to tell a story of their data from the point of its collection, through the process of lone and group analysis, to its inclusion in papers. They were designed to gather information about the entire data process so that the full scope of activities and requirements of the analysts might be reflected in the study. Our analysis of this interview data is continuing.

Here, we highlight two key issues that arise that have informed the design of tools to support remote data sessions. These concern the problems of sharing perspectives on video and the impact of different technological configurations on the data session.

Embodying Perspective

One of the major concerns within data sessions is to organize a shared seeing or shared perspective on the scene, such that emerging phenomena can be identified and discussed. The phenomena of interest might relate to the subtle interplay of talk and the body, maybe the shape of a gesture during a turn at talk or the glance of one individual during the utterance of another. Thus the phenomena of interest can be fleeting and slight, placing significant interactional demands on the data session participants to highlight them for others. This can lead to difficulties for colleagues to agree even where to start and stop the video.

However participants use various forms of embodied conduct to reference features on screen and over the course of a few seconds. The challenge is greater than two people discussing a document for example, as there are multiple recipients in the room, the referrer is often some distance from the screen (although in cases of extreme difficulty participants will often step up to the screen) and the video is dynamic – it is not simply static image – so features of interest are

often on-screen for only a moment or two. Nevertheless the respondents emphasised the benefits of being in the same room as the others:

'I think…if you're sitting there next to each other, [you're] tuned in, in a way, rather better with other people.' (Interviewee #21)

Indeed there are a number of broad resources and practices for indicating phenomena on the video data. The most common resource of this type is that of demonstrative reference to the screen to locate objects or activities. For example, one interviewee explained how they introduced data by starting with a still image around which they would provide some background information about the nature of the scene displayed on screen:
'If it's my data, I'll usually give some kind of overview so that everybody else knows the same thing. So, you know "This is a family, this is a kid of eight, they've just been to the gym…"That's relevant to this piece, to give that kind of ethnographic background.' (Interviewee #10)

Figure 1. An example data session – the participants assemble around a television, with one attempting to illustrate a point at a distance.

They would point to different people in the image and in many images would demarcate regions or artifacts in the scene to familiarise others with the context for the video recording. Pointing at features on screen is not tied to the start of sessions, but occurs to support various activities. This is most readily available to participants when a relevant static images is on display. Matters of reference are considerably complicated when the phenomena are not available in a static image or are only available in a dynamic image. Therefore participants routinely coordinate referential activities through requests to the video controller to rewind and play and stop at just the moments most appropriate to illustrate an analytic point. This can be cumbersome, but can also refine an analytic issue through collaborative involvement.

Transcripts can also provide a significant resource to encourage others to find relevant moments in the action. By drawing attention to particular parts of a textual transcript, participants can encourage others to notice action that occurs around the words or utterances that feature at those moments in the transcript. Such work can be crucial in reaching a shared perspective and transcripts often form the basis for coordinating of talk and work in the data session.

Another way in which participants may try to convey a phenomenon is through mimicking a gesture or movement that features on screen. For example, where a participant wishes to emphasize the swiftness of a research subject's movement across screen, they might move their hand quickly and in the same direction, whilst discussing this point. This gesture may be later used again to make a further point about that action. These mimicking gestures are in many ways not concerning with providing 'exact copies' of on-screen conduct, but rather are designing to render visible both the relevant action and the analytic point that is being made about that action. Thus they tend to exaggerate or transform the on-screen conduct. These various embodied practices of revealing phenomena are critical as data sessions progressively highlight one or two actions of interest and then develop preliminary characterisations of their organization.

Technologies of Perception

Currently there are no consistent standards in the presentation technologies used in data sessions. Whilst some continue to work with video players and televisions, others are using laptops and projectors. Interestingly, differences in the technologies of presentation lead to differences in the organisation of sessions. For example, inexperience with the technology may lead to changes in who controls the video playback. As one interviewee explained, visitors without laptops would have their data transferred from a DV-tape to a Mac laptop prior to their data sessions. A number of visitors had no experience with Macs, and this meant that they deferred responsibility for control to someone who did.

Also the use of computers to present data can alter the ways in which materials are distributed among the group. Whilst the presence of a transcript remains as important as ever, on-screen transcripts are increasingly used. Some highlighted the benefits of the co-location of transcript and video on a single shared display, referring to the increased ability to make links between the two:

'I think it's ludicrous to think the typed version of the transcript can capture everything, even with images, because it doesn't capture the full range of intonation…so I like an electronic transcript where you can both read it and play the stuff at the same time.' (Interviewee #2)

However, the interviewee also noted the distinctive benefits of being able to write around, annotate and otherwise transform a written transcript. The interviewee refers to the use of a paper-based transcript by the conversation analyst Harvey Sacks, which has a variety of handwritten notes all over it, and about it:

'Now, notice what Sacks was able to do, making all these notes that are locating graphically the contrast between on and off, two different things. Then you're having all these handwritten notes on top of it.' (Interviewee #2)

The personal annotation of materials and opportunities to juxtapose comments with portions of transcripts are impoverished by on-screen transcripts. As such, the transcript no longer forms a part of both the private and public realm of activities possible within the data session, but has become a part of only the public realm. The use of computers to play materials provides increased flexibility of presentation in various ways. The range of video files available often allows more spontaneous discussions of data as new clips or files can be drawn into the session as unexpected lines of inquiry emerge.

'You can look at the video, you can see the themes within the video, you can look at related videos, related texts, related photographs. You can click on anything, actually.' (Interviewee #19)

With tape-based materials, it is far more cumbersome to bring along a large range of videotapes and thus the opportunities for such flexibility in shifting between data is more unlikely. Also, opportunities to loop portions of video provide novel possibilities for video analysts:

'I use … this for data sessions because it's interesting that it allows you to replay. For example, here I was interested in one particular phenomenon; that was the fact that he had finished to say something. He was taking this plan and putting it aside. So it does this in a small prose you can really show the movement … it's not magic (laughs) but once you have done this work, it's really nice to go through and 're-go through' for a certain place, so it's, I began to use it as a presentation tool and, as erm, data session tool' (Interviewee #1)

Arising Issues

These studies have raised a series of issues that are informing the development of our demonstrator tools. Here, we have focused on two key aspects of these studies:

Embodying Perspective: Analytic work undertaken in data sessions rests upon the mutual availability and intelligibility of various visual resources. This routinely involves identifying features on moving or still video images or interrelating aspects of the associated materials artefact with the video. In our development work we are using these studies to consider the resources participants will need to work with (talk about, gesture over, make sense of) the data hand whilst working in groups remotely.

Technologies of Perception: It is clear from our studies that different configurations of tools and technologies that work in support of the data session present different challenges and constraints for users. For example, the paper transcript affords private annotation whilst the electronic transcript makes it easier to clarify problems and make agreed changes. In designing new solutions, we are using our studies to make decisions about how to balance private versus public displays, individual versus common controls, as well as what sort of control functionality is critical (e.g. looping).

Sharing Video with VidGrid

We are currently producing a prototype system for distributed collaborative video analysis based upon Access Grid-style distributed projected interfaces. Our design allows real-time collaborative sessions that mirror the traditional 'data sessions' undertaken within existing video analysis practice. Typically a range of resources are brought to bear within these proceedings – not just video, but associated materials such as transcriptions and ethnographic materials collected from the scene of data collection. Thus far we have focused on developing the real-time component of our system which enables shared annotation and juxtaposition of digital video and associated materials. At this stage we have decided not to grid-enable the digital video under scrutiny itself due to the ethical and legal implications and security overhead that this approach might introduce. Rather, we have decided to allow traditional, existing and trusted physical channels of data distribution to remain, and then provide software which is able to make use of this data across a grid environment.

Our prototype tool provides multimodal annotation of a video corpus between distributed sites, allowing coordinated navigation of the corpus. It is written entirely in Java, using the JMF API (Java Media Framework, 2005), and is currently deployed on a Windows desktop platform although it will run on any platform supporting JMF. JMF itself supports replay of a number of different media formats, and the tool has been successfully tested with MPEG-1 and AVI video files. Session information and overlay annotations made on the video stream are persisted via XML, and the tool makes use of the Xerces SAX parser library (Xerces, 2005) for XML file manipulation. The types of annotation of the video file currently provided by the tool are (i) textual transcription alongside the video data; and (ii) freeform mark-up directly onto the video stream itself. Audio communication permitting conversation between analysts is currently provided via existing voice teleconferencing channels, although we plan to include video views of participants as well as data. In the following sections, we describe the underlying infrastructure and interface, and relate their design to our analytic requirements.

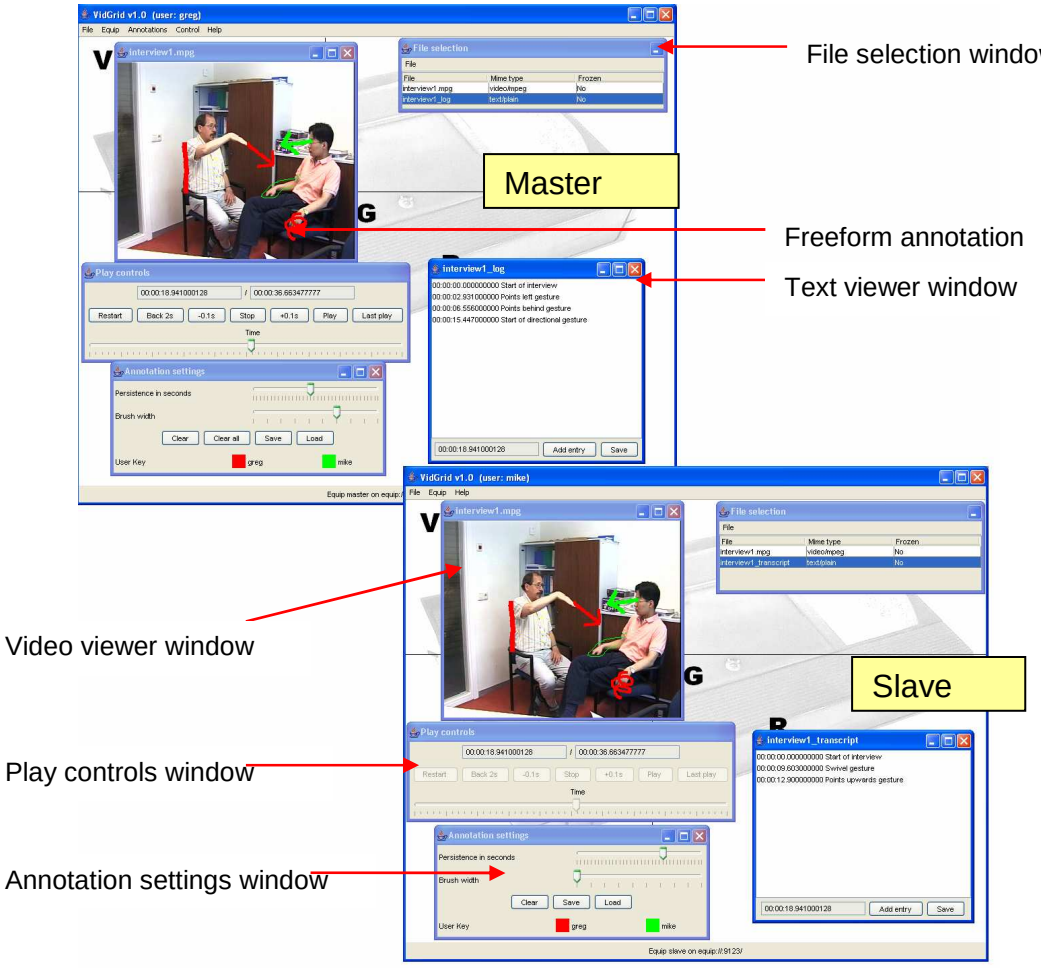

Figure 2. Master and Slave interfaces to the VidGrid system

Infrastructure and Communications

Communication between analysis clients is enabled by Equip, event-based middleware designed to support distributed interactive systems through the sharing of data among distributed heterogeneous applications (Greenhalgh, 2002). Equip provides indirect, loosely coupled many-to-many communication pathways between distributed application components via asynchronous event notifications. In contrast to a traditional synchronous point-to-point style of communication as in a client/server model, all communication in Equip is performed via event notifications to a conceptual network 'data space'. In effect, this means that it can be used to transmit data in synchronous groupware-like latencies, but is also able to allow sites to arbitrarily join and leave online data sessions without disrupting data between the other participants. Furthermore, it allows us to gather status information on the data session, allowing snapshots and histories of data sessions to be stored and retrieved at a later time.

We wanted to reinforce the notion that typically data is brought to a data session and controlled by a particular researcher. VidGrid is therefore structured in a single master, multiple distributed slave configuration, with control of the video stream resting with the master application, who then leads the analysis session. Nonetheless, reflecting the fact that another researcher may want to request control to emphasise a particular point, any slave site can be selected by the master to take control of the video at any point during the session.

Communication between components of VidGrid is composed of two major categories of events, illustrated in Figure 3.

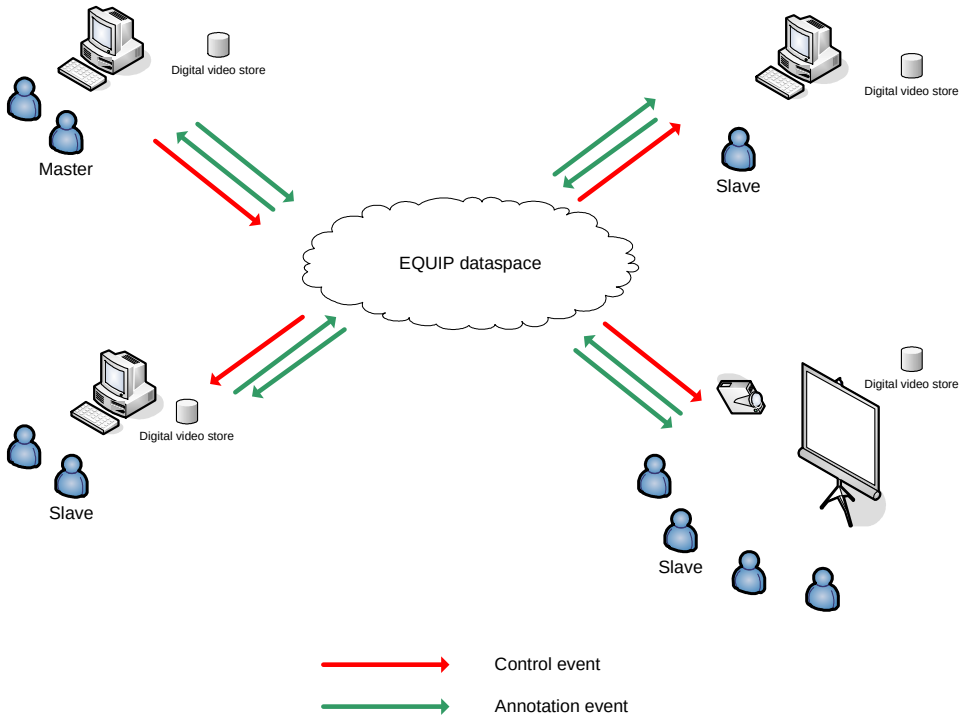

Figure 3. Communicated events in a typical VidGrid data session

- **Control events** – represent instructions published by the current master, and acted upon by all slave clients. Examples of control events include instructions to play or pause the video, and to restart the session.
- **Annotation events** – represent freeform annotations made on top of a video stream by any of the distributed users. All users may publish annotation events to the data space, and all users subscribe to annotation events. We will further explore annotation below.

Testing shows that control events are typically of a low frequency and incur negligible communication overhead, whereas annotation events are more frequent. VidGrid does not provide transmission of audio between clients, rather leveraging freely available Voice over IP (Skype, 2005), using boundary microphones and speakers for group audio support.

An important consideration of the system is that each user has a local copy of the digital video corpus for that data session, which is distributed via the existing external trusted channels already employed by the community, rather than over the network. For us, this approach circumvents major technical and ethical issues alike. Firstly, the real-time transmission of video would significantly increase the bandwidth requirements of the infrastructure. It is likely that, even with continuous high-quality networking between all sites, real-time transmission of video data would be at best unpredictable. Such latencies would affect the causality and/or quality of video playback, and would most likely vary these between multiple sites. Such problems would disrupt the social order and relevance of events, and more importantly, references to those events conveyed between sites through audio and/or annotation events (Ruhleder and Jordan, 1999, Gutwin et al., 2004). Secondly, our decision to rely on existing channels of video data distribution means that we can rely on existing ethical and legal practice to form part of distributed data sessions. In addition to avoiding the need for complex on-line access control and secure channels, data distribution can be controlled in ways which allow researchers to understand and decide when, how and where video data is distributed based on their detailed knowledge of the consents and agreements associated with particular items of data, and therefore independently of the distributed data sessions themselves.

User interface

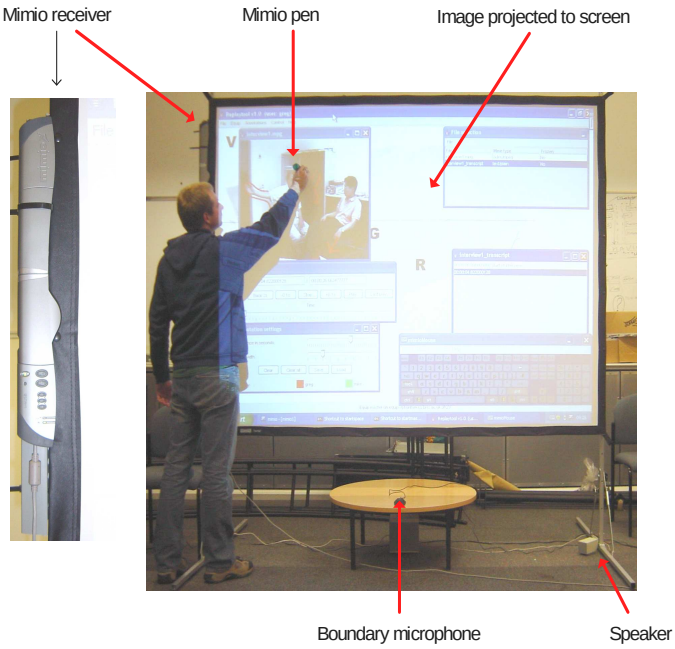

Figure 4. VidGrid user and projected display, showing Mimio receiver placement

All the VidGrid applications have been written in Java, and make use of the Java Media Framework (JMF) API[1] that provides video codecs for manipulation of MPEG-1 and AVI file types. Here, we encounter a trade-off between the diversity of video codecs used by different communities of practice, and the diversity of platforms used by researchers within those communities. The use of Java provides a nominal level of platform independence, allowing researchers' existing heterogeneous machine architectures and operating systems to be incorporated. However, JMF currently precludes some common video formats, for example QuickTime videos often used by researchers with Apple platforms. We have yet to incorporate video codec extensibility into our system.

The VidGrid user interface, illustrated in Figure 4 is implemented as a Multiple Document Interface. A major advantage of lightweight Swing components is that they specify a *glass pane* component, which acts like a transparent glass sheet over the windows. We make use of a custom glass pane under which the video stream is rendered to provide a transparent area on which freeform annotations may be scribbled. Both master and slave users of the application are presented with similar interfaces, illustrated in Figure 4. The most significant difference is that the master application has active video control buttons, whereas controls are deactivated in the other interfaces. Transferring control activates the particular slave's controls and deactivates those of the master ensuring only one user has control at any point in time. The master is able to clear all annotations made by all users, as well as control whether annotations are transmitted in real-time, or as packaged stroke objects.

Annotations

A single user sitting at a desktop screen could participate in a data session, using a mouse, headphones and a microphone. However, given many data sessions involve groups of researchers, we have experimented with projecting the interface to provide for multiple analysts at a single site. The projected interface incorporates a low-cost ultrasonic pen based input system (Virtual Ink, 2005), which uses a combination of infrared light and ultrasound emitted by a handheld pen to determine the pen's position relative to a stationary receiver. VidGrid interprets events representing a Mimio pen's position as mouse events (Mimio 2005), allowing an analyst to control the projected display, and also to make freeform scribbling annotations over a video window with the pen, as illustrated in Figure 3.

Annotation data are represented as a set of individual points making up each freeform line drawn. Communication of these freeform annotations is via individual event notifications per pixel drawn. We anticipated the network load of per-pixel events to be significant, so also created an option for packaged per-stroke transmission. The effect of using packaged strokes is that users only receive freeform annotations as a set of individual strokes, but with significantly lower

1 http://java.sun.com/jmf

communication overhead. We anticipate here a balance between the ability to perceive the production of a stroke at the remote site and the latency in perceiving that stroke at all. From our studies of co-located data sessions, it is crucial for an analyst to understand and use the way in which strokes are produced win order to embody and convey their perspective on the data. Such capabilities would be diminished by per-stroke transmission.

Viewer components

VidGrid provides single time-point synchronization of multiple time-related data and media streams, each of which is rendered by a particular software component supporting a common temporal navigation interface. Effectively, we want to support a range of data, such as multiple video streams (perhaps collected within the same time frame), associated materials, text transcripts and so on to be presented. Coordinated navigation of multiple time-related data requires a common underlying time model shared amongst viewer components. The time model adopted by the application is based on the JMF time model (Sun, 1999) which keeps time to nanosecond precision – although interestingly only at varying multi-nanosecond-scale intervals. All types of viewer in the application adopt this time model to ensure common time is established within the application.

The current version of the application provides separate viewer components to render video and text files, which may be arbitrarily synchronized with each other. The researcher defines a point of intersection between media streams (such as two videos with overlapping timeframes or a transcription of conversation within a video file), and the application generates the necessary timeline. Our design is extensible, using an abstracted viewer type, allowing the incorporation of viewer components for additional media and data types in the future (for example, we anticipate the need for images, screen captures, sensor log files and so on).

Video viewer

The video viewer component renders video data, allowing the analyst currently in control of the session to control the video stream via a set of simple VCR-like controls and time-slider provided by the control window, shown in Figure 3. Any analyst may make freeform annotations on top of the video stream, with annotations made by different sites, being differentiated by colour identified in an annotation settings window. Each site may also dynamically alter the width and persistence (how long to overlay the annotation on the video stream) of their strokes. Currently, individual sites are able to clear their own annotations, whilst the master may remove all annotations. Annotation removal is conducted via the data space, propagating to all instances of the application. Annotations may be saved and loaded locally, using a custom XML schema.

Text viewer

The text viewer component renders text, typically containing transcriptions, and allows in-band editing. The text viewer can be synchronized with video media, allowing entries to be added at a particular point in time, and navigation through transcription entries is synchronized with the video playback. Text files containing transcriptions are currently stored locally and transcriptions are not communicated amongst other users in the way that annotations are. This approach retains a locality of transcription, reflecting the fact that typically, analysts will annotate their own transcripts in recognition of, or in preparation for, analytic agreements on changes, additions, or points of importance.

Reflections on use

We have conducted trials with VidGrid, initially between rooms within the same institute and more recently across multiple sites within the UK. Whilst data collected from these and further trials will be the subject of further scrutiny, here we reflect on initial issues which arise.

We have encountered well-understood problems both with groupware-style systems and with the use of video data. For example, we have had to contend with varieties of video codec and their conversion at multiple sites so that all participants can view a particular piece of data. We have also had to solve occasional differences in the way in which Java operates on different systems. There have been further differences in the use of firewalls and networking security in place at the respective sites. These issues are illustrative of the difficulties which much e-Science work faces in standardization and sysadmin burdening which are the subject of current discussion within the e-Science programme's usability community. However, here we plan to revisit two particular issues from our studies of co-located data session practice: our use of technologies for display; and how perspective is collaboratively embodied in analytic work.

Re-embodying Perspective

Technically, we have achieved a reasonably low-cost set-up which functions well across multiple sites. Our trials suggest that annotation data can be transmitted in per-pixel mode with literally imperceptible latency over a 10 Gbps national-scale network. The result is that Voice over IP conversation and per-pixel production of a stroke gesture are possible in conjunction. So, for example, the circling of a feature of interest over the video data can be cogently juxtaposed with a reference in talk to that feature. Nonetheless, we have started to focus on two further issues relating to conveying a reference which jar with co-located analytic practice.

Firstly, the use of strokes over video data alters significantly when annotating a paused frame with annotating at playback. The annotation of a single frame allows consistent discussion over the feature of interest. However, during this process, participants tend to forget that the annotations they are producing have a variable persistence value which will result in those strokes continuing over subsequent frames. On playback, this persistence becomes noticeable, and as the frames change, the annotation loses its relevance whilst maintaining its presence. We might automatically reduce pause-frame annotations to very low persistence levels, but then those strokes would be barely visible during real-time playback of the sequence. Furthermore, annotating at playback time introduces its own set of problems. Whilst persistence levels are more naturally configurable, the production of the strokes themselves is not, given each stroke has a particular start time and lifetime. For example, drawing an arrow to point at some feature results in two strokes being used – one for the line and one for the arrowhead. The line of the stroke will typically be produced first, and therefore disappear first before the arrow head. We might address such issues by introducing particular shapes such as arrows as defined annotation options, but at the cost of both increasing interface complexity, and potentially reducing freeform flexibility.

Secondly, we have noticed the difficulty of producing strokes such that others at remote sites can identify features of interest in a video. Despite the use of real-time per-pixel strokes, there are aspects of annotating data for others which are lost by only transmitting screen-contact gesture and audio. Particularly, whilst co-located researchers are able to see the analyst prepare to produce a stroke in front of the screen, researchers at remote sites are only aware of the stroke *at the time it is being produced*. It turns out that understanding the ways in which the display is approached, and the particular region of data is homed in on, is crucial to the organisation of perspective. As with many CSCW applications, audio becomes fall-back channel on which researchers begin to rely for the preparation of a stroke. To alleviate such problems, we plan to start conveying some notion of where the annotating devices are with respect to the display, perhaps through tracking of the annotating pens' positions around the intervening space and appropriate visualization at remote sites. Finally, and perhaps most importantly, the use of on-screen annotation precludes much of the imitation and exaggeration of behaviour within data that we identify in co-located data sessions. It is highly problematic to convey the very character of how data is seen by an analyst, for example the way in which a head is moved or a gesture is

produced, without the ability to directly embody that character rather than translate it into strokes. Our future work, therefore will start to investigate ways in which we might also configure sensors to automatically capture the body movement of participants and relate those movements to sequences within the video data. Such attempts will, however need to be sensitive to the production of analytic behaviour within the context of both local and remote groups.

Technologies of Perception revisited

Our use of projected interfaces has highlighted the importance of the display to a group in sharing perspectives on data. We have initially used available front-projection screens to conduct data sessions. These have two clear problems. The first is the flexibility of the screen, which causes difficulty with stability when pressing the electronic whiteboard marker onto the screen strongly enough to generate an ultrasonic signal. The screen is not sufficiently taught to prevent quivering in the surface, making it difficult to maintain the position of the pen accurately. Secondly, the shadows cast on the screen obscure the very region of the application being used, generating difficulties both for the researcher attempting to use the system and the co-located analysts attempting to view the data. These difficulties could be solved by combining rear projection with the use of a solid surface screen. Unfortunately, there are few rear-project solutions which both use solid materials and hold a reasonable image. We plan to start experimenting with various acrylic and semi-opaque glass possibilities.

Further, we have begun to note the clear difference between levels of activity in co-located data sessions and with distributed projected interfaces. Using VidGrid, the projection screen becomes a window through which two interrelated activities are occurring: both analyst communication and analysis. Communication, in analytic talk and annotation, as well as the well-known overhead of additional interaction repair required in distributed collaboration, exponentially increases the activity with and around the display. Combined with the scale of the projection screen itself, this makes data sessions perspiration-inducing; and while contributing to researcher fitness, we anticipate different configurations, such as table-top displays will reduce levels of physical activity and increase the potential for multi-participant access to the application.

Future Work and Conclusions

Our studies have indicated a range of directions which are required for distributed video analysis. We have generated an initial front-end and networked system which takes appropriate perspectives on some of these issues. Beyond the next stage of development, which will be to enable Equip dataspace events for providing persistent corpus analysis, there is considerable work to be undertaken before the subtleties of our requirements gathering have been addressed. We summarise these developments below.

Our initial VidGrid prototype circumvents the ethical issues associated with distributing video data over the grid. We have effectively proposed an interim solution which promotes the ethical status quo, a strategy which should not be discounted in general approaches to distributing video data. At the least, even though access control and security mechanisms have been a key element of Grid middleware development (Foster et al., 1998), commensuration is required between (for example) implementations of GSI security layer authentication and ethical consent and constraint. Only then will networking of digital video data itself become practical.

We have implemented a token-passing control mechanism in which a single data session master retains control of multiple data session clients, and that server control can be moved between those clients on request. The challenges of jointly or concurrently controlling real-time distributed applications are delineated by Gutwin and Greenberg (Gutwin and Greenberg, 1998). We intend to develop such approaches based on evidence of requirements in experimenting with prototype control further. In addition, a persistent annotation system supporting the intertwining

of video data control from collection cradle to publication grave will require further significant consideration of how systems embed ownerships and relationships.

Given that studies using (sequences of) video data are made relevant to analysts through associated materials, we have implemented the ability to juxtapose and temporally synchronise textual sequences with video streams. However, there is much further work in introducing a range of distributed materials, such as photographs, background video, written notes and so on. We also anticipate that synchronization may be required in different axes than time, for example the spatial relationships between collection point of materials.

Finally we have discussed how embodying perspective in both remote and local domains is of key importance to analytic work. We propose that there are significant benefits to be gained by rendering data of the relationships between embodied activities and the local environment to be used in the remote representation of activities. In a companion paper (Fraser et al., 2005), we discuss how such data might be obtained to uncover these relationships, providing the foundations for distributed and collaborative production of remote data analysis.

Acknowledgements

VidGrid is supported under the e-Social Science pilot projects programme by the ESRC, award number RES-149-25-0013-A.

References

Bechhofer, S., Stevens, R., Ng, G., Jacoby, A. and Goble, C. (1999). Guiding the User: An Ontology Driven Interface, in *Proc. UIDIS'99*, pp. 158-161.

Booth, S., Brooke, J., Caldwell, K., Carver, L., Daw, M., De Roure, D., Flavell, A., Galvez, P., Gilmore, B., Hughes, H., Juby, B., Judson, I., Miller, J., Newman, H., Osland, C. and Rogers, C. (2002). Multi-Site Videoconferencing for the UK e-Science Programme, *UK e-Science Technical Report Series*, UKeS-2002-04.

Foster, I., Kesselman, C., Tsudik, G. And Tuecke, S. (1998). A Security Architecture for Computational Grids, in *Proc. 5th Conference on Computer and Communications Security*, pp. 83-92, 1998, ACM.

Foster, I. and Kesselman, C. (eds.) (1999). *The Grid: Blueprint for a New Computing Infrastructure*, Morgan Kaufmann, 1999.

Fraser, M., Biegel, G., Hindmarsh, J., Heath, C., Reeves, S. and Greenhalgh, C. (2005). Object Focused Interaction in e-Social Science, short paper in *Proc. Workshop on Social Aspects of Scientific Collaboration, International Conference on e-Social Science 2005*, June 22nd-24th 2005, Manchester, UK.

Greenhalgh, C. (2002). EQUIP: a Software Platform for Distributed Interactive Systems, *Equator IRC Technical Report Equator-02-002*, Nottingham, 2002.

Gutwin, C. and Greenberg, S. (1998). Design for Individuals, Design for Groups: Tradeoffs between Power and Workspace Awareness, in *Proc. ACM Computer-Supported Cooperative Work*, pp. 207-216, 1998, ACM.

Java Media Framework (2005). http://java.sun.com/jmf, verified 01/05/05

Skype (2005). http://www.skype.com, verified 01/05/05

Xerces (2005). http://xml.apache.org/xerces-j/, verified 01/05/05

Mimio (2005). http://www.mimio.com, verified 01/05/05