Computer Supported Cooperative Work (2006) 15:257–279
DOI 10.1007/s10606-006-9027-y

Remote Collaboration Over Video Data: Towards Real-Time e-Social Science

MIKE FRASER[1,*], JON HINDMARSH[2], KATIE BEST[2], CHRISTIAN HEATH[1], GREG BIEGEL[1], CHRIS GREENHALGH[3] & STUART REEVES[3]

[1]*Department of Computer Science, University of Bristol, Merchant Venturers Building, Woodland Rd, Bristol, BS8 1UB, UK (Phone: +44-117-9545144; Fax: +44-117-9545208; E-mail: fraser@cs.bris.ac.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 & Learning Sciences Research Institute, School of Computer Science & IT, University of Nottingham, Jubilee Campus, Wollaton Road, Nottingham, NG8 1BB, UK*

Abstract. The design of distributed systems to support collaboration among groups of scientists raises new networking challenges that grid middleware developers are addressing. This field of development work, 'e-Science', is increasingly recognising the critical need of understanding the ordinary day-to-day work of doing research to inform design. We have investigated one particular area of collaborative social scientific work – the analysis of video data. Based on interviews and observational studies, we discuss current practices of social scientific work with digital video in three areas: Preparation for collaboration; Control of data and application; and Annotation configurations and techniques. For each, we describe how these requirements feature in our design of a distributed video analysis system as part of the MiMeG project: our security policy and distribution; the design of the control system; and providing freeform annotation over data. Finally, we review our design in light of initial use of the software between project partners; and discuss how we might transform the spatial configuration of the system to support annotation behaviour.

Key words: video analysis, e-social science, groupware, synchronous collaboration

1. Introduction

One of the most impressive developments in CSCW over the past decade or so has been the substantial corpus of naturalistic studies of work and collaboration and the ways in which tools and technologies feature in everyday practice (Hughes et al., 1992; Heath and Luff, 2000; Luff et al., 2000). The emergence of these workplace studies has provided, as Barley and Kunda (2001) describe, a distinctive contrast to certain trends within organisational analysis and the sociology of work and has helped to drive analytic attention towards the local and indigenous organisation of workplace activities. This corpus of research has also begun to address some of the methodological challenges that arise in undertaking fine grained, naturalistic studies of work in complex organisational environments, environments where participants

rely upon and utilise an intricate array of tools and technologies, objects and resources, in accomplishing their daily activities. Audio-visual recordings have increasingly provided an important resource for these ethnographies, augmenting more conventional field work and enabling researchers to examine, both alone and increasingly in collaboration with others, how various tools and technologies feature in the work and interaction of the participants themselves. Indeed, audio-visual recordings coupled with a range of other material resources, including field notes, official documents and records, computer logs, transcripts, and the like, play an important part in CSCW research and collaboration between researchers, research groups and in some cases practitioners. As yet, however there have been few studies of these forms of collaboration that have emerged within CSCW and despite a large literature on communication through video since the inception of the field (e.g. Fish et al., 1990; Gaver et al., 1992; Bly et al., 1993), relatively little attention has been paid to the ways in which we can support and enhance research activities with video data.

Nonetheless, there is increasing interest throughout computer science to investigate ways in which systems can be developed to support research work across geographical boundaries. The emergence of 'e-Science' in Europe, and the corresponding 'cyberinfrastructure' in the US, draws on primarily technical developments, such as the advent of Grid Computing (Foster and Kesselman, 1998) within the high-performance computing community and distributed resource and service discovery within distributed computing. These communities have been driven by the desire to see infrastructures which can support large-scale teams in analysing large-scale datasets across large numbers of processors. Projects have provided middleware which is capable of processing a complex range of academic and professional data, from virtual astrophysics observatories (Lawrence et al., 2002) to emergency team response modelling (Berry et al., 2005). The social sciences have more recently become of interest to developers in these areas, and what has been called 'e-Social Science' emerges partly because areas of quantitative social science also require fairly large-scale data processing, and partly because the methodological variability of the social sciences offers new challenges for the development of such infrastructures.

The scale of development in e-Science is impressive, with a quarter of a million pounds sterling having been spent in the UK programme alone, and $100 m per annum spent on Cyberinfrastructure in the US thus far, with substantial budget increases anticipated from this level. Having spent such large budgets on engineering the technical capability to distribute and share large volumes of scientific data, inevitably such communities are now facing 'usability challenges' in incorporating 'collaboration support' into their novel middleware infrastructures. For this reason, the e-Science community is turning to the social sciences for design guidance as well as users. Despite

early work developing ontologies of scientific processes to form workflow software that structure collaborative work (Bechhofer et al., 1999), research in the usability of e-Science is starting to uncover more detailed relationships between e-Science systems and the practices of those within target domains (Jirotka et al., 2004). Given the sheer scale of development of these systems, however, there is a pressing need to further investigate everyday practices of research across the social and physical sciences.

In this paper, we present an investigation into the collaborative analysis of video data as part of everyday research practice. We draw on these studies to inform the design of tools to support the real-time analysis of video data by distributed groups of social scientists.

2. Background

In recent years we have witnessed the emergence of a substantial body of research in the social and cognitive sciences concerned with the visual, material and spoken aspects of language, communication and social interaction. This body of research has built upon wide-ranging studies of talk, discourse and language use to address the ways in which objects and artefacts, tools and technologies feature in human conduct and social interaction. We have witnessed the emergence of 'workplace studies', naturalistic studies of complex organisational environments (Luff et al., 2000), of HCI and CSCW (Bannon, 1992), of research on computer-mediated communication (Turoff et al., 1982), and research on the ways in which talk and gesture is embodied within material resources (Hindmarsh and Heath, 2000). These developments reflect a broad range of methodological commitments and yet, in various ways are concerned with the fine details of conduct, communication and interaction. As a result, video is now used in a wide range of disciplines within the social and cognitive sciences and provides researchers with opportunities to capture versions of human conduct and subject them to repeated scrutiny using slow motion facilities and the like.

One of the affordances of video data is that it enables researchers to show colleagues their raw data in order to assess their analytic observations with others. It facilitates collaborative analysis in ways that other forms of (especially qualitative) social science data struggle to support. It should not be a surprise therefore that video-based research across the social sciences increasingly involves close collaboration between individual researchers and research groups. For example, there are number of leading research groups in the UK, Europe and United States, in various disciplines, who undertake video-based projects that involve team-based data analysis and collection. There are also an increasing number of inter-institutional projects, for example between laboratories across Europe, that involve the collaborative analysis of shared data corpora. This is exemplified by various projects in the

fields of HCI, CSCW and ubiquitous computing that involve close collaboration between social and computer scientists undertaking video-based research, and by various EU projects and networks that involve collaboration between social science teams. Aside from these formal arrangements there are widespread informal collaborations between video-based researchers, collaborations that involve co-present and collaborative data analysis workshops, both within the UK and between researchers in the UK and abroad. Therefore, video-based research often involves distributed teams of researchers engaged in collaborative analysis of shared data.

Of course, there exists a range of software to support synchronous distributed collaboration with documents, text, diagrams and other physical or computational objects which may feature in these research projects. Groupware for document editing has formed a core thread of CSCW research. Whilst not directly impinging on the use of video, clearly the policies and procedures introduced by these systems have a direct impact on understanding how remote groups might come together to work with articles and artefacts of interest. So, for example, Sun et al. (1998) discuss how consistency in world view of all participants might be maintained in the face of network delays and architectures which preclude truly synchronous work. Equally, Gutwin and Greenberg (1998) demonstrate the trade-off between a user's need to perform activities and their need to understand the activities of remote others. Such research highlights the importance of understanding difficulties associated with sharing data under the auspices of sharing interaction.

More recently, there has emerged a developing body of work on digital photography and the sharing of images over computer networks. So, for example, researchers are beginning to tease apart categories of the social use of digital images (Frohlich et al., 2002; Grinter, 2005; Kindberg et al., 2005), and draw on these uses to propose new ways of sharing and comparing images (Van House et al., 2005). Whilst systems to support the sharing of images between members of the public are developing, system support for real-time video sharing is less prominent. However, the increasing emphasis on video-based studies in the social and cognitive sciences has not gone unnoticed by software developers, and has led to a range of commercial products and prototype research technologies to support the work of video analysts. These tools primarily include enhanced forms of software designed to support general qualitative data analysis in the social sciences (e.g. Atlas-ti, Nvivo) or software specifically dedicated to the analysis of video in the social sciences (e.g. Observer, CIAO, DIVER, CEVA). However, the widespread use of text data in the social sciences means that many systems treat video as an add-on to software primarily concerned with textual materials, so that the analysis of the video itself is rather cumbersome. Some software, however, more closely attempts to support the distributed collaborative

analysis of video data by qualitative researchers. Our initial review of these programs highlights two particular areas for development.

- *Concealing Collaboration*. Some programs, such as Transana (Fassnacht and Woods, 2005), allow multi-user video data sharing at a database level. The software treats collaboration as an additional feature, being designed primarily for lone researchers working in isolation, rather than supporting the specific work of groups of distributed researchers or collaboratories. Any support for collaboration provided tends to be rather rudimentary – for example, by providing opportunities to attach notes to sequences or images from the video, or email comments. Furthermore, some prominent approaches to video-based research prioritise data sessions as a way of working, emphasising the value of real-time discussion and analysis of data. Therefore asynchronous support for collaborative work is insufficient, further tools to facilitate synchronous collaboration would be beneficial.
- *Meeting Facilitation*. A clear approach to the development of distributed systems has been to design for meetings as primarily about face-to-face talk – the visual channels are simply to provide 'back channel' information. Programs like Access Grid (Childers et al., 2000), and the more sophisticated work in the Memetic project (Buckingham Shum et al., 2006) carry forward this approach. However, the kinds of research meetings of concern here, namely data analysis sessions, demand that participants can all see and discuss video sources materials. They also involve other forms of mixed media, including transcripts, images and drawings. Existing e-Research tools to support meetings provide clumsy support for people to share, discuss and gesture over and around video data and associated materials, an issue that we will return to later. This encourages consideration of how real-time data control mechanisms relate to analytic purpose and practice. Given these constraints, we found no current dedicated tools that can adequately help remote groups of social scientists discuss and collaborate over shared video data in real-time. This encourages us to suggest that e-Social Science needs to consider traditional CSCW topics like distributed interaction. That said supporting real-time, distributed analysis of video materials brings new issues and challenges to the classic interests of CSCW in mediated collaboration. Furthermore, workplace studies may be in a position to offer shape to the kinds of research technologies that would fit with existing scientific and social scientific practice. However, despite a surplus of social science literature discussing method and methodology, there is very little work describing the everyday practices of doing research in the social sciences – especially geographically distributed research – which developers might draw upon. Therefore, to begin, we investigated how

analysts currently use video data in order to inform our design and development work.

3. Method

In order to explore current ways in which video analysts collaborate over and around video materials, we have conducted a series of qualitative interviews with expert video analysts from a range of disciplines. The interviews took place over a seven-month period and, in total, we have interviewed 27 individuals working in seven different countries. The interviews lasted between one hour and one and a half hours each. They were audio recorded and subsequently transcribed in full. The interviewees were selected as leading exponents of various forms of video analysis drawn from the fields of sociology, linguistics, anthropology, psychology and organisational studies. 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 and performance analysts. 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 and presentations. 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.

However, for the purposes of this paper we focus on one aspect of data sharing and collaboration around video data and present ways of supporting such data sharing amongst remote research groups. We focus exclusively on support for real-time analysis of video materials, in what are commonly called 'data sessions'. Data sessions are essentially collaborative video analysis sessions and are common features of everyday practice and training especially within the fields of ethnomethodology and conversation analysis. They do vary in form as we will discuss and are used to support different kinds of activities, from identifying themes deserving further analysis, refining transcriptions, focusing on analysis for a paper or more generally for developing the analytic skills of students and young researchers. Whilst there are various technological developments aimed at supporting the work of individual analysts and some concerned with asynchronous support for data sharing, there are no strong solutions to the problems of managing remote and synchronous analysis of video materials. To augment our interview materials concerning the organization of data sessions, we also undertook participant observation in a number of data sessions.

Building on from these studies, we are engaged in a software development programme to support the distributed analysis of video over computer networks. The MiMeG system has been written in Java, and makes use of the

Java Media Framework (Sun Microsystems, 1999). Given the possibility of a group using the system, we have designed the interface to the software to work on both desktop PCs and projected interfaces. The software provides synchronisation of multiple 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, multiple video streams (perhaps collected within the same time frame), associated materials, text transcripts, visual annotations and so on to be presented coherently as they become relevant to the sequence of events. Our development and use of software has been a useful method of both testing our designs and allowing us to reflect on the different interactional properties of conducting analysis across remote sites.

The following sections introduce our requirements gathering programme and illustrate how specific designs take these requirements into account. These investigations describe existing data analysis and data sharing practices as well as considering designs and limitations of tools that might support such practices for distributed analyses. There are three key issues arising through our studies that have informed the design of these tools. These issues include the work of *Preparation* for data sessions, through ethics approval, trust and collating associated materials; *Control* of data at run time and management of the application; and *Annotation* configurations and techniques that organise shared perspectives on, indeed help to constitute, emerging analytic phenomena.

3.1. PREPARATION WORK

Working together on the analysis of data cannot begin until all the interested parties have access to all the materials necessary. In this section, we outline the shareability and variety of materials which video analysis currently relies upon, and describe ways in which our system is designed to support preparation work, the beginning of the process of research.

Video can be highly sensitive data with consents and approvals governing its use and distribution. Ethical issues affect video analysts to differing degrees, dependent largely on the nature of their recordings. For example, those carrying out lab-based experimental research expressed fewer ethical concerns than those recording naturally occurring data. However, the issue is universally viewed as important. The existing ethical and legal frameworks regarding the collection and presentation of video are perceived as unclear, especially regarding video-based research. However, as one participant offered:

> 'I'm unhappy with a lot of the legalisms. I think it's more my own sense of having a responsibility to the [participants]'. (Interviewee #2)

This statement expresses the sentiment of many interviewees. Indeed, such responsibility extends concerns around sharing video data to the possibility that other researchers might be able to form different conclusions without interaction with the participants and first-hand experience of the setting. For these reasons, researchers are keen not to release control over the data.

These concerns have raised acute difficulties for the development of shared databases containing raw video data. The major reasons are outlined in the following interview comments:

> 'Some data is really sensitive and you just don't want to give it away...it will lead you to define different access, different databases, and security barriers with a strong difference between data that can be shared and used and data that has to remain confidential.' (Interviewee #1)

> 'It's not just about managing who's got the data, but what they will do with it. My worst nightmare is that I get a call from a school or parent saying "I saw my little Johnny was on the telly about your project and I didn't give permission for it".' (Interviewee #16)

Quite aside from ethics, more personal concerns with allowing others to access your data were discussed.

> 'I might say the biggest issue I think on collaborative data is the question of whether you trust other people so that you can be open to them with your ideas and getting credit for your hard work and that's a big issue.' (Interviewee #2)

The least ethically demanding situation in which public access to data might be allowed was via the data session, or a similar forum, in which a discussion of data is possible whilst the raw data do not leave the control of the owner. The crucial aspect expressed here was that control was retained and discussion could be moderated by the owner of the data.

Researchers possess a range of existing strategies for controlling who views and can copy material. In addition to data sharing in data sessions, many of our respondents were engaged in work with colleagues at institutional, national and international levels. They are uncomfortable with making video available in ways that could risk the security of those materials. Thus, they tend to post hard copies of data to one another (either on DVD or tape). This can involve significant amounts of copy time if the raw data is to be copied. Therefore a standard 'workaround' is to simply send 'collection tapes' (featuring significant clips of data), but this has the effect of narrowing potential analytic foci. Nonetheless, these data distribution routes are trusted and used, whereas electronic transmission is generally avoided by our interviewees. In order to discuss materials colleagues would either communicate via

email or through preliminary drafts of papers or by arranging face-to-face meetings for data sessions, which are necessarily less common in international (or even national) collaborations.

To address these complex practical, personal and ethical demands, it might be possible to implement equally complex access control, encryption and security methods to preserve data over a network. Aside from the technical difficulties in achieving completely secure networking, and the time taken to work through and adopt additional security mechanisms at each site, the primary problems with this approach are (a) it takes the control of data security out of the hands of the researcher and puts it into the hands of the developer, who has no direct vested interest in keeping particular items of data safe; and (b) researchers would no longer be able to articulate the details of security mechanisms to stakeholders, including writing consent forms, applying to ethics committees and, most importantly, working with those being recorded.

As a result, our design for the MiMeG system assumes that each user has a local copy of the digital video corpus for that data session on their machine, which is distributed via the existing external trusted channels already employed by the community, rather than over the computer network. Time indices into the data are instead transmitted over the network, which indicate at what point the application should currently display that video. The decision to rely on existing channels of data distribution means that we can rely on existing ethical, legal and research practice to form part of distributed data sessions, and on researchers to 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. Effectively, it means that our system is as secure for video data as the use of the same machines without our system.

Our approach also circumvents a major issue with the real-time transmission of video. High-resolution video transmission would significantly increase the bandwidth requirements of the infrastructure. Even with continuous high-quality networking between all sites, it is likely that delivery of the video data would be unpredictable at best. 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 temporal order and interactional significance of events and, more importantly, references to those events conveyed between sites through talk and action (Ruhleder and Jordan, 2001; Gutwin et al., 2004).

Whilst we have suggested that data sessions bring together researchers to analyse video data, analytic work also involves the juxtaposition and discussion of a variety of associated data and materials. For example participants will routinely draw on documents produced to chart, map or transcribe action unfolding on the video. The most common of these is some sort of transcript of the talk by participants featured on-screen. Depending on the

type of research this can range from 'soundbites' through to detailed phonetic transcripts. There may also be: 'indigenous materials' relevant to the analysis, such as documents taken from the scene – log books, record cards, computer print-outs – or physical artefacts such as instruments or tools; photographs of elements of the scene – e.g. signs, whiteboards, technologies; documentary materials that relate to the setting, such as pages from manuals or textbooks that describe standard procedures or rules for settings such as this one; or sketches or diagrams produced during the data session to clarify the standard ecology of the setting or the character of the tools and technologies in use. Finally, participants often have to hand multiple camera viewpoints. So two, three, four or even five recordings of a scene may have been taken and provide different angles and perspectives with which to piece together adequate descriptions of the action.

The range of additional materials at hand will depend upon the research domain and the analytic interests of the researchers. However, there can be quite an assembly of items that will be shown and discussed at different points within the data session. Indeed researchers sometimes even leave the session to collect these additional forms of data if they become relevant to emerging analytic discussions. It should be noted that whilst these materials are physically distributed in different ways in the setting, some (such as the core materials) are presented on shared displays; others, such as copies of verbal transcripts, are often given to all those attending the session; and others still, such as documents or artefacts from the research domain, are passed around at particularly relevant moments.

Clearly it will be problematic to distribute physical artefacts over a network. For digital or digitised materials, we have provided for the association and juxtaposition of materials by creating an extensible architecture to allow the simultaneous viewing of different media types, based on Mime types. Viewers can be created for any data type (the application currently supports video, images, text and some log data types) and, using a multi-window interface, each viewer can be juxtaposed, sequenced and replayed alongside others. Additionally, instances of data types may be arbitrarily synchronised with each other. The researcher defines a point of intersection between media streams (such as two videos with overlapping timeframes) and the application generates the necessary timeline.

The video viewer renders video data of various formats, allowing the analyst to control the video stream via a set of simple VCR-like controls and time-slider provided by the control window. The text viewer component renders text, typically containing transcriptions, and allows on-screen editing. As with all viewers text can be synchronised with video, allowing entries to be added at a particular point in time. Our text-based viewer serves as a public transcript display, and while it does not impose a transcription alphabet it provides for the time-stamping of (descriptions of) events. Selecting these

descriptions or transcriptions will move the timeline to that point, and therefore move the video data to that moment. We have also provided an Axis Web service connected to a relational database over the Web to store and retrieve transcripts across different data sessions. As transcripts change and evolve, so we provide the ability to merge multiple transcripts over time and to build up a time-stamped transcript set of analytic work over multiple data sessions.

3.2. MATTERS OF CONTROL

Co-present analysis of video requires a researcher to shuttle around the dataset; usually the researcher who has primary responsibility for that data, although control can be passed over to others for a variety of specific reasons. In this section, we look at who controls data in a data session, how control is managed, when control is transferred and how control events are requested.

During a data session, the nature of the equipment used normally demands that one participant takes control of the video playback for the duration of the session. Most frequently, control of the video falls to the owner or deliverer of the data; that is, the person who brings the data to the session. This individual's first hand experience of the data – and most likely the research setting – is seen as most relevant. They will be responding to most questions about the data, so they also control the video. On occasion, control of the video switches to others. However, this is relatively rare and often relates to emerging problems in communicating which elements to view:

> 'When I get domineering and there's something that I really want to see, I can control it' (Interviewee #2)

Controlling the video and requesting parts of the video to be seen again are not without their problems:

> 'I [find] it frustrating that I am the one who controls the data and who decides exactly when to go back and how far. And that can suit or not suit the other participants.' (Interviewee #1)

In particular, the ability to locate the right moment on a video becomes a problem shared by all participants. Rather than requiring a solution, however, some argue that this problem holds many advantages for data session practitioners. As one research participant suggested:

> 'I feel that there is an advantage in having to ask somebody to do it for you, because you then have to account for why you want to skip to that part of the tape...I expect it makes things more explicit, justified, and might even help the analysis process if you have to explain why you're about to move the talk on to something else.' (Interviewee #16)

So, whilst there are difficulties associated with describing moments on a tape, this practice can encourage participants to justify their reasoning and their developing analytic interests and concerns. Having one system, with one controller, also means that the conversation is organised with regard to a single individual encouraging a single strand of discussions rather than multiple segmented conversations.

In order to manage control within our software, we needed to provide some scalable event-passing communication between applications. We wanted the networking to be robust for people to join and leave data sessions, but also to give maximum data throughput between clients to minimise the variation latency between different machines. Our real-time application events are enabled by Equip, event-based middleware designed to support distributed interactive systems through the sharing of data among distributed heterogeneous applications (Greenhalgh, 2002). 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 publishing and subscribing to event notifications in a conceptual network 'data space'. In effect, this means that our software is optimised to transmit control data in synchronous groupware-like latencies (minimising the latency issues mentioned above), but is also able to allow sites to arbitrarily join and leave online data sessions without disrupting data flows between the other participant sites.

Using this infrastructure, we wanted to reinforce the notion that typically data is brought to a data session provided and controlled by a particular researcher. MiMeG is therefore structured in a single Master, multiple distributed Slave configuration (Figure 1), with control of the video stream resting with the Master application, who then leads the analysis session. Reflecting the fact that another researcher may want to request control or control events, any Slave site can be selected from the Master application to take control of the video. The use of voice over IP provides opportunities for remote participants to talk to one another, negotiate issues of control and the like.

Control events are one of two major categories of events communicated between Master and Slave applications, shown in Figure 2 (the second category is annotation events which are discussed in the next section). Control events represent instructions published by the Master application, and subscribed to by Slaves in order to enforce play, pause, rewind and so on across all applications.

Testing shows that control events are typically of a low frequency and incur negligible communication overhead. Using this approach, we have retained the individual's control over the data in a single Master multiple Slave configuration. Initially the Master application has control over the data, and others can request events, retaining the use of articulation found in co-present analysis. We have also preserved the potential for passing control

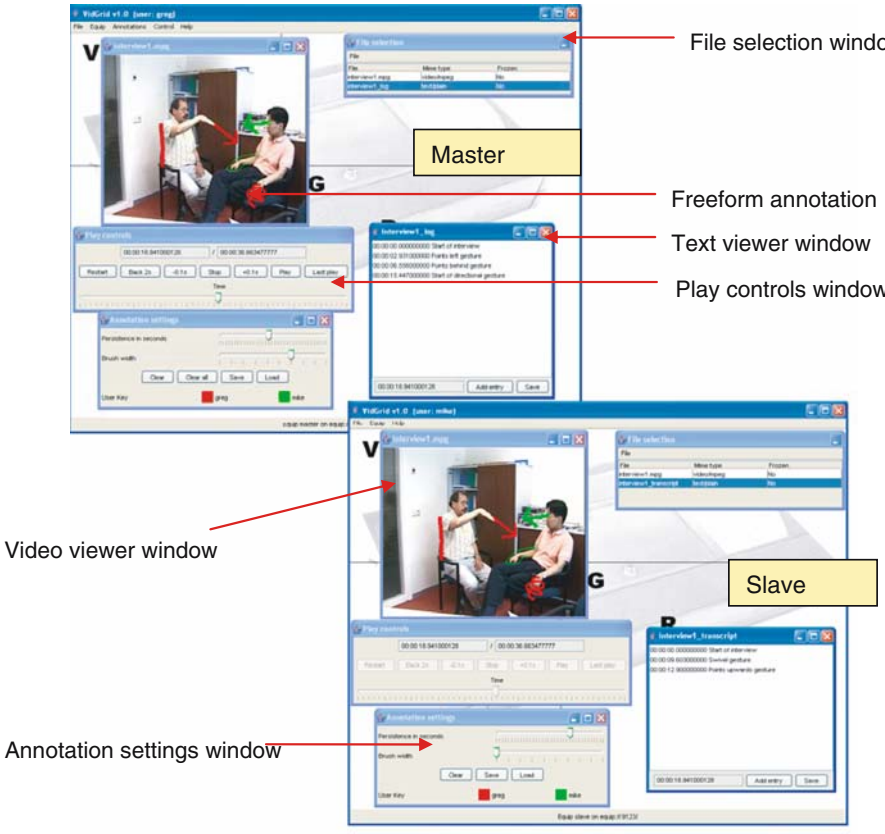

Figure 1. Master and slave interfaces to the MiMeG system.

by enabling the Master to transfer controls to other sites, although the Master still remains in control of the control.

3.3. ANNOTATION AND PERSPECTIVE

Annotations are the other form of event distributed in MiMeG between Master and Slave applications (also shown in Figure 2 above). Within the e-Science community the term 'annotation' has become synonymous with 'metadata', with the arduous creation of textual descriptive notes to ease the automated processing of the data they describe. However, work by Goodwin (1994) considers the ways in which annotation is a key feature of everyday occupational practice. He describes the various practices that underpin the development of 'professional vision' for practitioners, practices that involve the use of coding schemes, highlighting practices and graphical representations for example. Whilst he focuses on these practices in the work of

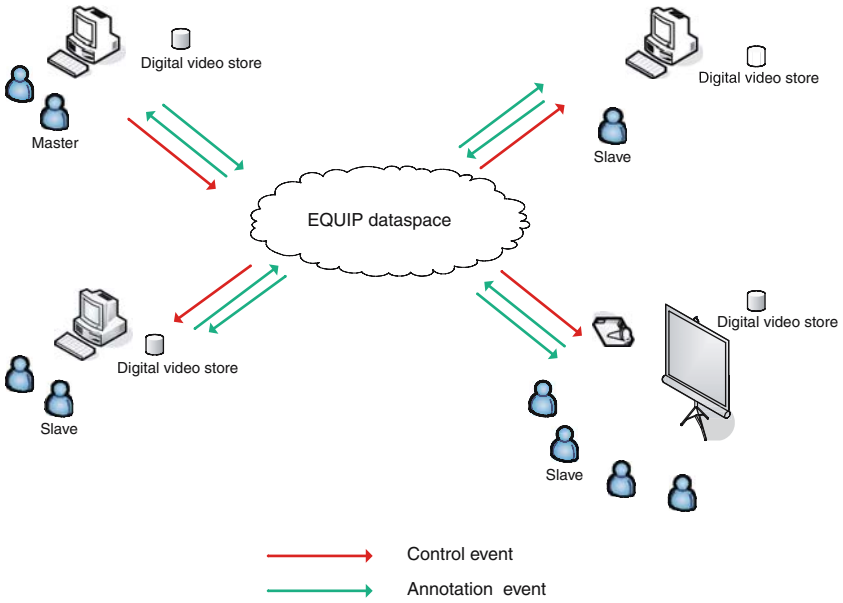

Figure 2. Communicated events in a typical MiMeG data session.

archaeology and courtroom interrogation, he also notes how they are equally evident in work of social science. We start this section by broadly considering practices of reference and highlighting in data sessions and move on to how this understanding has informed our use of annotation in MiMeG.

One of the major concerns within data sessions is to organize a shared seeing or shared perspective on features of the video materials, 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 even in identifying where to start and stop the video so as to best reveal phenomena.

Also participants use various forms of embodied conduct to reference features on screen. 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.

There are a number of broad practices for revealing phenomena in the video data. The most common resource of this type is clearly pointing out a

feature on-screen as it appears on-screen. 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)

They would point to different people in the image and in many images would demarcate regions of the scene or artefacts 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 image is on display. Matters of reference are considerably more complicated when the phenomena are not available in a static image but rather in a dynamic series of images. 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 a cumbersome practice, but as mentioned earlier can also refine an analytic issue through discussion.

Transcripts can provide an important 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 coordination of perspectives 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. These mimicking gestures are in many ways not concerned with providing 'exact copies' of on-screen conduct, but rather are designing to render both the relevant action visible and the analytic point that is being made about that action. Thus they tend to exaggerate or transform the on-screen conduct. These gestures may be later used again to make a further point about that action.

These various embodied practices of revealing phenomena are critical as participants progressively highlight conduct of interest and then develop preliminary characterisations of its organisation. Therefore rather than treating annotation as an activity to be performed for purposes of managing data and metadata, they are treated as critical for supporting communicative practice. Video data in particular requires a more subtle interpretation of real-time annotation than can be predetermined and structured, for example

through the use of an ontology of existing practice. There are, however a number of ways of supporting annotation. For example, systems such as Vannotea (Schroeter et al., 2003) allow annotations to be shared in real time whereby regions of interest within the video can be highlighted using pre-determined shapes and media associated with those regions. The means of annotation is not sufficiently flexible to convey a trajectory of production. More interactional detail is required to provide a real-time communicative resource to help participants to convey their emerging analytic perspectives. Such difficulties have led us to closely consider ways in which annotations can be designed to be flexibly constructed and co-constructed in interaction.

To this end, annotation data are represented in our software as a set of individual points making up each freely drawn line. These represent freeform annotations made over a video stream by any of the distributed users, who can all publish and subscribe to annotation events. 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. We anticipate here a balance between the ability to perceive the production of a stroke at remote sites and the latency in perceiving that stroke at all, whilst losing the ability to understand and use the way in which strokes are produced in order to embody perspective.

There are also some general issues about the spatial organisation of a data session. As can be seen in Figure 3, participants often form a horseshoe shape around a television or a projection image. Clearly the geographical distribution of participants in remote data sessions will transform this assembly. Indeed we have produced software that works over different configurations, both public projected displays and privately used desktop

Figure 3. An example data session – the participants assemble around a television, with one attempting to illustrate a point at a distance.

systems. An individual could participate in a data session, using a mouse, headphones and a microphone. However, given many data sessions involve two or more groups of researchers, we have primarily experimented with projecting the interface to provide for multiple analysts at each 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. MiMeG interprets events representing a pen's position as mouse events, 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 4.

There are interesting challenges in storing and retrieving these freeform annotations which – contrary to traditional synchronous/asynchronous categories – are important for using in subsequent sessions to recall to points or processes that were previously achieved with the data. We have implemented an axis webservice to store and retrieve annotations in relational database. The storage of flexible annotations to identifiable and retrievable categories is

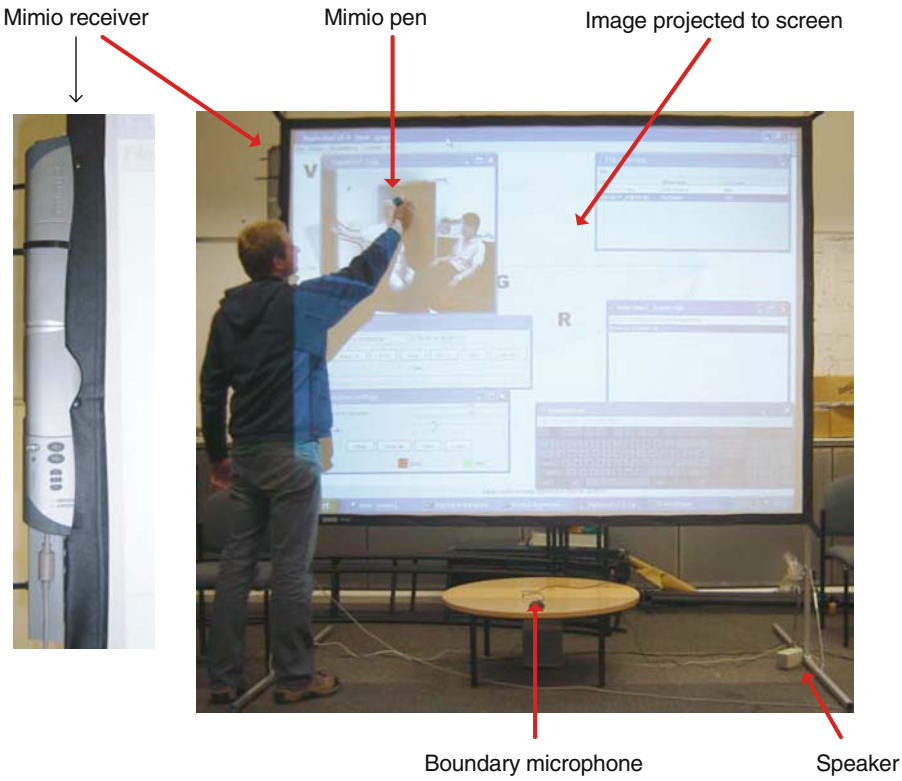

Figure 4. MiMeG projected display, showing Mimio receiver placement.

another area in which previous work in the CSCW field might be brought to bear (e.g. Dourish et al., 1999), given that the fewer predetermined metadata the annotation has, the greater the challenges in identifying memorable categories against which they might be stored. Thus far our interface stores scribbles according to the log-in used, project identified at start-up, video data file annotated and time/date. This provides some means to browse for previous annotations according to occasion as well as project or owner, for example by retrieving the annotations of the previous data session as a means to remind the researchers of the process by which they reached previous interim conclusions. However, such benefits need to be balanced against the potential difficulties that researchers may encounter accessing additional ports through institutional firewalls.

4. Early experiences with MiMeG

Over the past year, we have conducted a number of trials with the MiMeG software between project partners at sites within the UK. Whilst data collected from the use of the system will be the subject of longer-term scrutiny, here we reflect on issues which arise from our initial use of the system between Bristol, London and Nottingham.

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 100 Mbit/s national-scale network. The result is that simultaneous and virtually instantaneous 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 sensibly juxtaposed with a reference in talk to that visual feature. This is partly due to avoiding streaming of video data in real-time, and indicates this is a sensible and simple approach in the first instance.

We have encountered well-understood problems of networking systems and software. For example, we have had to contend with differing use of firewalls and networking security in place at our respective sites. Here, however, we note particular issues that relate to our studies of co-located data session practice: our use of display technologies, and our ability to convey analytic perspective.

The use of projected interfaces has highlighted the importance of the display to a group in sharing perspectives on data. We have initially used front-projection screens to conduct data sessions. The shadows cast on the screen obscure the area of the application being used, generating difficulties for the researcher attempting to use the system and the co-located analysts attempting to view the screen. In addition, the use of a shared projection means that researchers have to physically approach the screen to convey their

perspective on the data. Using MiMeG, the screen is a locus for both site-to-site communication and for analytic data. Annotating the screen significantly increases the researchers' physical activity be enforcing contact with and movement around the display. Data sessions therefore involve high levels of activity, swapping marker pens and changing places. We anticipate a number of possible configurations to address this issue.

Firstly, we might move towards flat surface use, such as table-top displays that may reduce levels of physical activity whilst increasing the potential for multi-participant access to the application. Secondly, we might consider annotation methods which do not require direct access to the screen, through individual analysts' use of devices which can be used to interact with the application. For example, we might provide a tablet PC with private/public regions for interface activity, or we might consider the use of physical artefacts such as paper-based transcripts which also provide access to digital scribbles (e.g. the Anoto pen system). However, there are two further issues relating to ways of indicating features during video playback that are noticeable with the use of freeform annotations.

Firstly, the use of strokes over video data alters significantly when annotating a paused frame versus annotating at playback. The annotation of a single frame allows relatively straightforward reference to features 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, annotation during playback 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 adequately preparing remote sites that someone is about to produce an annotation. 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, the trajectory of the gesture, 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 visualisation 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 capture the movement of participants and relate those movements to sequences within the video data.

5. Conclusions and future work

Following previous work on coherently sharing artefacts over distributed systems (e.g. Hindmarsh et al., 2000; Luff et al., 2003; Kirk et al., 2005), we have found that corresponding issues exist in distributed analysis. We take the difficulties described in initial trials with annotating data to be a result of prioritising data transmission over the relationship between the analysts' bodily conduct and the data. This step was taken to circumnavigate the known problems with using only video as a medium of communication when working with objects (e.g. Gaver et al., 1993), and to shift the attention to video as a focus for collaboration.

However, our brief presents somewhat different challenges than supporting individual to individual work with objects. Because groups of researchers will typically be co-located as well as distributed, such attempts to represent remote participants will need to be sensitive to the production of artefact-centred analytic behaviour within the context of both local and remote groups. Additionally, the data artefacts in question have temporal as well as spatial properties, and therefore references can be to the development of a sequence rather than simply a single feature. Such complexities will provide distinctive challenges for supporting remote collaboration over data.

Importantly, we have started to investigate the relationships between real-time research support and collaborative work with data. The few frequently used e-Research systems which provide real-time group-to-group work such as the Access Grid (Childers et al., 2000) are targeted at meeting support. Although such systems may also provide distributed visualizations or

presentations, there is little design consideration of the ways in which participants might come together to analyse these representations together. Real-time collaborative research on unknown problems and answers is an order of magnitude more complex a situation in which to achieve interactional coherence between sites.

Whilst our work in this area is exploratory, the overwhelming finding from this process has been that maintaining coherent access to researchers perspectives' on artefacts and data is absolutely critical to analytic interaction. Where we have seen troubles in analysis, they have not simply derived from commonly held beliefs about the ability of social scientists to operate technologies. Although there are clearly training issues (as with all groups), many social scientists, particularly those within CSCW, have studied difficulties with operating technologies in depth and well understand their way around software use. Rather, analytic troubles stem from difficulties with designing systems that adequately provide the opportunity to establish mutual access to the technologies, both physical and digital, that provide the loci of analytic discussion. These issues bear strongly upon concerns within CSCW on the design of technologies to support real-time remote collaboration, and it is our conclusion that the challenge of distributed research rests more in supporting coherent mutual configuration than it lies in difficulties of recruiting social scientists into technical enterprises.

Acknowledgements

This work reported here forms part of by the VidGrid and MiMeG projects. The VidGrid Project was funded by the ESRC under the e-Social Science pilot projects initiative, award numbers RES-149-25-0013 and RES-149-25-0013-A. MixedMediaGrid (MiMeG) is an ESRC e-Social Science Research Node, award number RES-149-25-0033. We very much appreciate the participation of all of our interviewees. We would also like to thank for their comments and contributions our colleagues in MiMeG, members of the distributed National Centre for e-Social Science (NCeSS), the anonymous reviewers of this paper and the special issue editors.

References

Bannon, L. (1992): Perspectives on CSCW: From HCI and CMC to CSCW. In *Proc. International Conference on Human-Computer Interaction*. ACM Press, pp. 148–158.

Barley, S. and G. Kunda (2001): Bringing Work Back In. *Organization Science*, vol. 12, 76–95Informs.

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

Berry, D., A. Usmani, J. Torero, A. Tate, S. McLaughlin, S. Potter, A. Trew, R. Baxter, M. Bull and M. Atkinson (2005): FireGrid: Integrated Emergency Response and Fire Safety Engineering for the Future Built Environment. In *Proc. UK e-Science Programme All Hands Meeting (AHM 2005)*, 19–22 September, Nottingham, UK.

Bly, S., S. Harrison and S. Irwin (1993): Media Spaces: Bringing People Together in a Video, Audio, and Computing Environment. *Communications of the ACM*, vol. 36, no. 1, pp. 27–47.

Buckingham Shum, S., R. Slack, M. Daw, B. Juby, A. Rowley, M. Bachler, C. Mancini, D. Michaelides, R. Procter, D. De Roure, T. Chown and T. Hewitt (2006): Memetic: An Infrastructure for Meeting Memory. In *Proc. 7th International Conference on the Design of Cooperative Systems*, Carry-le-Rouet, France.

Childers, L., T. Disz, R. Olson, M.E. Papka, R. Stevens and T. Udeshi (2000): Access Grid: Immersive Group-to-Group Collaborative Visualization. In *Proc. 4th International Immersive Projection Technology Workshop (IPT 2000)*. Iowa State University Press.

Dourish, P., W.K. Edwards, A. LaMarca and M. Salisbury (1999): Presto: An Experimental Architecture for Fluid Interactive Document Spaces. *ACM Transactions on Computer-Human Interaction*, vol. 6, no. 2, ACM press, pp. 133–161.

Fassnacht, C. and D. Woods (2005): *Transana v2.0x*. http://www.transana.org. Madison, WI: The Board of Regents of the University of Wisconsin System.

Fish, R., R. Kraut and B. Chalfonte (1990): The VideoWindow System in Informal Communications. In *Proc. CSCW '90*. ACM Press, pp. 1–11.

Foster, I. and C. Kesselman (1998): *The Grid: Blueprint for a New Computing Infrastructure*. Morgan-Kaufmann.

Frohlich D., A. Kuchinsky, C. Pering, A. Don and S. Ariss (2002): Requirements for Photoware. In *Proc. CSCW '02*. ACM Press, pp. 166–175.

Gaver, W.W., T. Moran, A. MacLean, L. Lvstrand, P. Dourish, K. Carter and W. Buxton (1992): Realizing a Video Environment: EuroPARC's RAVE System. In *Proc. CHI '92*. ACM Press.

Gaver, W., A. Sellen, C. Heath and P. Luff (1993): One is Not Enough: Multiple Views in a Media Space. In *Proc. INTERCHI '93*. ACM Press, pp. 335–341.

Goodwin, C. (1994): Professional Vision. *American Anthropologist*, vol. 96, 606–633.

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

Grinter, R.E. (2005): Words about Images: Coordinating Community in Amateur Photography. *Computer Supported Cooperative Work*, vol. 14, no. 2, pp. 161–188.

Gutwin, C. and S. Greenberg (1998): Design for Individuals, Design for Groups: Tradeoffs between Power and Workspace Awareness. In *Proc. CSCW'98*. ACM Press, pp. 207–216.

Gutwin, C., S. Benford, J. Dyck, M. Fraser, I. Vaghi and C. Greenhalgh (2004): Revealing Delay in Collaborative Environments. In *Proc. CHI 2004*. ACM Press, pp. 503–510.

Heath, C. and P. Luff (2000): *Technology in Action*. Cambridge University Press.

Hindmarsh, J. and C. Heath (2000): Sharing the Tools of the Trade: The Interactional Constitution of Workplace Objects. *Journal of Contemporary Ethnography*, vol. 29, no. 5, pp. 523–562.

Hindmarsh, J., M. Fraser, C. Heath, S. Benford and C. Greenhalgh (2000): Object-Focused Interaction in Collaborative Virtual Environments. *ACM Transactions on Computer-Human Interaction (ToCHI)*, vol. 7, no. 4, pp. 477–509.

Hughes, J., D. Randall and D. Shapiro (1992): Faltering from Ethnography to Design. In *Proc. CSCW'92*. ACM Press, pp. 115–122.

Jirotka, M., R. Procter, M. Hartswood, R. Slack, A. Simpson, C. Coopmans and C. Hinds, (2005): Collaboration and trust in healthcare innovation: The eDiaMoND Case Study. *Journal of Computer Supported Cooperative Work*, vol. 14, no. 4, pp. 369–398.

Kindberg, T., M. Spasojevic, R. Fleck and A. Sellen (2005): I Saw This and Thought of You: Some Social Uses of Camera Phones. In *Extended Abstracts of CHI 2005*. ACM Press, pp. 1545–1548.

Kirk, D., A. Crabtree and T. Rodden (2005): Ways of the Hands. In *Proc. ECSCW 2005*. Paris, France: Springer, pp. 1–21.

Lawrence, A. and the Astrogrid consortium (2002): The AstroGrid Project: Powering the Virtual Observatory. In A. S. Szalay (ed.): *Proc. SPIE*, vol. 4846, SPIE, pp. 6–12.

Luff, P., C. Heath, H. Kuzuoka, J. Hindmarsh, K. Yamazaki and S. Oyama (2003): Fractured Ecologies: Creating Environments for Collaboration. *Human Computer Interaction*, vol. 18, 51–84.

Luff, P., J. Hindmarsh and C. Heath (eds.) (2000): *Workplace Studies*. Cambridge University Press.

Ruhleder, K. and B. Jordan (2001): Co-Constructing Non-Mutual Realities: Delay-Generated Trouble in Distributed Interaction. *Computer Supported Cooperative Work*, vol. 10, no. 1, pp. 113–138.

Schroeter, R., J. Hunter and D. Kosovic (2003): Vannotea – A Collaborative Video Indexing, Annotation and Discussion System For Broadband Networks. In *K-CAP 2003 Workshop on Knowledge Markup and Semantic Annotation*, Florida, October 2003.

Sun, C., X. Jia, Y. Zhang, Y. Yang and D. Chen (1998): Achieving Convergence, Causality-Preservation, and Intention-Preservation in Real-Time Cooperative Editing Systems. *ACM Transactions on Computer-Human Interaction*, vol. 5, no. 1, pp. 63–108.

Sun Microsystems (1999): *Java Media Framework API Guide version 2.0* [on-line] http://www.java.sun.com/products/java-media/jmf/2.1.1/guide/index.html, verified 30/11/2005.

Turoff, M., S.R. Hiltz and E.B. Kerr (1982): Controversies in the Design of Computer-Mediated Communication Systems: A Delphi Study. In *Proc. CHI 1982*. ACM Press, pp. 89–100.

Van House, N., M. Davis, M. Ames, M. Finn and V. Viswanathan (2005): The Uses of Personal Networked Digital Imaging: An Empirical Study of Cameraphone Photos and Sharing. In *Extended Abstracts of CHI 2005*. ACM Press, pp. 1853–1856.

Virtual Ink Corporation (2005): *Mimio Xi Specifications* [on-line] http://www.mimio.com/meet/xi/docs/Xi_winmac_specsheet.pdf, verified 30/11/2005.