

Social Science Information

<http://ssi.sagepub.com>

Cognitive technologies for establishing, sharing and comparing perspectives on video over computer networks

Roy Pea, Robb Lindgren and Joseph Rosen

Social Science Information 2008; 47; 353

DOI: 10.1177/0539018408092577

The online version of this article can be found at:
<http://ssi.sagepub.com/cgi/content/abstract/47/3/353>

Published by:



<http://www.sagepublications.com>

On behalf of:

[Maison des Sciences de l'Homme](#)

Additional services and information for *Social Science Information* can be found at:

Email Alerts: <http://ssi.sagepub.com/cgi/alerts>

Subscriptions: <http://ssi.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.co.uk/journalsPermissions.nav>

Citations <http://ssi.sagepub.com/cgi/content/refs/47/3/353>

Roy Pea, Robb Lindgren & Joseph Rosen

Cognitive technologies for establishing, sharing and comparing perspectives on video over computer networks

How can I tell what I think till I see what I Say?

(Forster, 1927, *Aspects of the novel*)

Abstract. *The rapidly increasing presence of digital video recordings and new communication capacities on the Internet has created new possibilities for collaboration in behavioral science research. Unfortunately, digital video has so far proven to be an unnatural medium for collaborative activity due to a lack of adequate tools that support joint analysis of a shared video record. We describe seven socio-technical design challenges that face what we term computer-supported collaborative video analysis. We also describe a software environment that we have created called DIVER that was designed to address these challenges and provide a platform for the fluid exchange of video data as well as the insights that the data elicit. The affordances of DIVER for supporting collaboration around video are grounded in the descriptions of two academic contexts in which the software played a central role in the group's video-based activities. The potential of DIVER to serve as a 'cognitive technology' is discussed.*

Key words. *Cognitive technologies – Collaboration – Digital video – Socio-technical design challenges*

Résumé. *Le développement accéléré des enregistrements en vidéo numérique et les nouvelles possibilités de communication offertes par l'Internet ont donné accès à de nouvelles opportunités de collaboration dans la recherche en sciences du comportement. Malheureusement, la vidéo numérique s'est montrée jusqu'à aujourd'hui un outil inhabituel dans les activités collectives en raison du manque d'outils adéquats permettant l'analyse*

conjointe d'un enregistrement vidéo partagé. Les auteurs décrivent ici sept défis socio-techniques pour la conception de ce qu'ils qualifient d'analyse vidéo collective assistée par ordinateur. Ils décrivent aussi l'environnement software qu'ils ont créé, le DIVER, pour relever ces défis et fournir une plate-forme permettant l'échange fluide de données vidéo et la visualisation des informations que ces données explicitent. Les 'affordances' offertes par le DIVER pour faciliter la collaboration autour de vidéos sont basées sur la description de deux contextes académiques dans lesquels le software a joué un rôle central dans des activités de groupe autour de la vidéo. Finalement, les auteurs discutent le potentiel du DIVER en tant que 'technologie cognitive'.

Mots-clés. *Collaboration – Conception et défis socio-techniques – Technologies cognitives – Vidéo numérique*

It has been a long-standing interest to understand how the cognitive technologies we invent – not only computers but symbolic systems of writing and mathematical notations – serve as instruments of cultural redefinition not simply by serving as *cultural amplifiers* (Bruner, 1966) of what we do, but by shaping who we are by *reorganizing* our activity systems (e.g. Pea, 1985, 1987; Pea & Kurland, 1987; Mills & Pea, 1989).

As the materialist history of Marx & Engels (1978) highlighted to such effect, humans are unique among species in fashioning their own nature, as their productive activities change when new resources and forms of society are invented. The Soviet psychologist L. S. Vygotsky developed this approach as he looked to how social life and mental life related to one another, asking: how do humans collectively produce new means for regulating their behavior? In his English-translated work, his characterization of such self-regulatory innovation is expressed as follows: 'The sign acts as an instrument of psychological activity in a manner analogous to the role of a tool in labor' (Vygotsky, 1978: 52). He argued that, because socially organized activities change in history, human nature is not a fixed but a changing category; and he characterized how tools that mediate symbolic and communicative spheres of activity are vital to these changes in human nature. In short, humankind is reshaped through a dialectic of reciprocal influences: our productive activities change the world, thereby changing the ways in which the world can change us. By shaping nature and the way our interactions with it and with one another are mediated, we change ourselves (Pea, 1985).

The proliferation of digital video recording, computing and Internet communications in the context of behavioral sciences research has opened up dramatic new possibilities for creating what we call 'digital video collaboratories' (Pea, 2006; Pea & Hoffert, 2007). In digital video collaboratories, scientists

work together to share datasets, tools, coding systems, analyses and other resources in order to advance collective understanding of behaviors such as learning and teaching interactions that are captured in digital video records. Video can be a powerful tool for examining the learning that occurs through human interaction (Jordan & Henderson, 1995). A large variety of interaction types may be portrayed in video data, with common contexts including K-12 classrooms, surgery and hospital emergency rooms, parent-child or peer-peer situations, informal learning encounters with peers on a playground, corporate workgroups, gesture displays in human interaction, uses of various products in their everyday environments (e.g. cars, computers, cellphones, household appliances, medical devices), and so on. For example, Goldman et al. (2007) provide a comprehensive volume of contributions to the uses of video research in the field of the learning sciences.

Yet video has been an unnatural medium to support collaborative activity such that collaborator insights can develop in concert and with a persistent record of their synergies. Video has also been dramatically under-utilized as a 'cognitive technology' that permits one or more persons to develop their reasoning and understanding of human interactions captured in video recordings and other time-based media or artifacts involved in cultural practices. Despite the importance of video as a research-data medium, video data circulates sluggishly if at all within scientific research communities. Researchers doing video analyses must ultimately revert to text-only displays for presenting results (obscuring connections between argument and evidence). It is rare for collaborators working at different sites to conduct a joint analysis of shared video records. Original data sources are typically not made available for re-analysis by other researchers (often leading to redundant primary-data collection efforts and the lack of valuable secondary-data analyses). These limits collectively impede the development of shared examples of exemplary video analyses that could serve training and socialization functions for novice researchers within a community of practice. Several federally sponsored workshops have documented the need for better tools that would enable powerful video capture and analysis technologies to support studies of learning and instruction (Lampert & Hawkins, 1998; MacWhinney & Snow, 1999; Pea & Hay, 2003; Derry, 2007).

Text provides for collaborative writing, but, when researchers or other learner groups want to work with video as a medium for interpretation, sharing, and commenting, their collective work is not well enabled. Our challenge is making *computer-supported collaborative video analysis* (CSCVA) into a commonplace collective practice, whether face-to-face or in distributed groups connected with mediating technologies. Currently we are a long way from achieving collective practices of video analyses in the social sciences,

studies of human–computer interaction and any other area where video serves as a valued means of data collection.

In this article we describe the socio-technical design problems facing CSCVA, in order to explain our rationale for and experiences with DIVER, a platform that supports video collaborations for several ongoing scientific and educational activities. The need is to define and overcome obstacles for the infrastructure of our research activities that use video data for studying learning and teaching. Our objective is to invent affordances of video interactivity to better support learning inquiries, and to do so by iteratively co-evolving our tools and practices. We shall consider how workflow scenarios with DIVER are meeting – and being challenged by – the CSCVA socio-technical design challenges with which we began, and the new issues that emerge in this co-evolution of tool and human activities.

Background and significance of the CSCVA problem

The advancement of augmentation tools for human activities generally has built on several decades of socio-technical design theory and practices. Scacchi notes how:

Socio-technical design is concerned with advocacy of the direct participation of end-users in the information system design process. The systems include the network of users, developers, information technology at hand, and the environments in which the systems will be used and supported. The process includes the design of the human–computer interface and patterns of human–computer interaction. (2004: 656)

With this view in mind, we ask: what are the primary socio-technical design issues for CSCVA that heed these concerns?

We characterize seven socio-technical issues that constitute core challenges affiliated with CSCVA. Virtually all the concerns outlined relate to the fundamental problem of *coordination* of attention, interpretation and action between multiple persons. Clark (1996) described as ‘common ground’ what it is people seek to achieve in the work they do to coordinate what they are attending to and/or referring to, so that, when comments are made, what these comments refer to can be appropriately inferred, or elaborated. In the learning sciences, ‘common ground’ is mostly used to examine collaborative or teaching–learning discourse and pointing, bodily orientation and joint visual regard to the focus of a conversation that is being analyzed for studies of learning or teaching (e.g. Barron, 2003). But it is not sufficient to focus just on the non-technology-mediated aspects of common ground; once we look at inscriptional systems (e.g. Latour, 1986) that lay down and layer symbolic records, such as text, diagrams or other paper-based representations, they,

too, become a focus of pointing and joint visual regard. One also often needs to refer to specific states of information display when using computer tools, so establishing a common ground for discourse and sense-making in a digital realm is important, too, including the dynamic medium of video.

1. The problem of reference. When analyzing video, how does one create a lasting reference to a point in space and time in the dynamic, time-based medium of a video record? In the here and now, one can point one's finger to a video display, encircle the area of interest and thereby highlight what will subsequently be commented on. Traditionally, the video medium has not provided the capability for a watcher of the video to point in such a way that *records* the locus of reference for a time-shifted or space-shifted (remote) audience of the referring act. It is noteworthy how important this is, even for an individual analyzing a video; absent a record of what one has pointed at, it may not be possible to later recall what aspects of the video were found to be interesting, problematic, compelling, or to which aspect a written comment should be attached. Hence solving the video-reference problem can serve as a personal memory aide, not only a collaborative one.

2. The problem of attentional alignment. How does one know that what is being referred to in a video record is understood by others? This problem, related to the first problem of reference, is one of co-reference, or 'attentional alignment'. It pertains to how it is that someone analyzing video and engaging in a referential act to some part of it can establish that the person being addressed by this referring act is attending to the video segment that is being referenced. Gesture and bodily orientation are often used in face-to-face efforts to organize attentional alignment (e.g. Goodwin, 1986). Such coordination is important since consequent dialog about a video referent can lead to misunderstanding and other conversational troubles if the listener has an alternative referent in mind. This process of 'deixis' – communicating one's referential perspective through pointing or otherwise specifying one's locus of attention for another to acknowledge and follow – requires a shared context.

3. The problem of creating video 'immutable mobiles'. In characterizing the power of written texts in contributing to the impact of science and technology after the 16th century, Latour (1986) developed the influential concept of inscriptions as external representations of ideas that serve as 'immutable mobiles', and which have these key properties: (i) inscriptions are mobile; (ii) they are immutable when they move; (iii) they are made flat; (iv) the scale of the inscriptions may be flexibly modified; (v) they can be

cheaply reproduced and spread; (vi) they can be reshuffled and recombined; (vii) one may superimpose different images of totally different origins and scales; (viii) they can be made part of a written text; and (ix) their 2-D character allows them to merge with geometry. Although his concerns were to provide a novel explanatory account of how science and technology took hold so powerfully, Latour's theory has had considerable influence in the digital-documents world (e.g. Levy, 2001), and its implications for conceptualizing video as an inscriptional medium are important as we consider how it can serve theory development and argumentation in the learning sciences. How can video be transformed into a flexible inscription system with similar properties for broad incorporation into human discourse (Stevens & Hall, 1997)?

4. The problem of effective search, retrieval and experiencing of collaborative video work. The researcher's work-life of video analysis using videotapes was challenging enough with physical tape. If we can solve the problems of allowing users to point to video, establish attentional alignment and create video as an immutable mobile medium, we generate new problems along the way. Users will create a vast array of persistent video + pointing + commentary digital objects. How will these objects be meaningfully organized such that one can search for and subsequently experience those video moments that matter to them?

5. The problem of permissions. How does one make the sharing and spreading of video and its associated interpretations more available, while maintaining control over sensitive data that should have protection for human subjects or digital rights management? Access and control issues depend on appropriate assignment of permissions with videos that have human-subjects conditions for informed consent. For this reason, it is important to have only approved individuals view such videos. Permissions may also be required because of digital-rights-management issues, common for film or television works but also applicable to user-generated video content. One may wish to allow only certain individuals to view videos, to create video annotations, or to make remixes of video assets. Subtle attribution issues emerge when creative actions are taken using secondary source material.

6. Integrating the insights of a collective group. What is the best way to harvest and synthesize the collective intelligence of a group engaged in collaboratively analyzing video? In the canonical interaction analysis methodology for group work on video, there is a hierarchical rather than heterarchical organization of analysis: a group leader controls the video

selection and playback, regulates participant contributions, audio-records the group's discourse, mines the group-work audiotape and then makes an executive summary of the analysis informed by group deliberations (Jordan & Henderson, 1995). The result of such a process is that there are many times when contributions from members are not being tapped for comparison and reflection, and it is possible that important perspectives and insights are excluded. How can CSCVA open up new kinds of research-activity structures that more effectively integrate participant contributions compared to traditional face-to-face video analysis, and consequently bring added value to the analysis?

7. The problem of establishing coherent multi-party video-anchored discourse in an activity system. Consider a face-to-face conversational interaction, as in a seminar, where the rules of discourse that sustain turn-taking and sense-making as people converse are familiar. Discourse analysis and studies of conversation have made great progress in uncovering the systematics of turn-taking and speech acts, and in accounting for the semantic and pragmatic coherency of discourse across turns and speakers (e.g. Goodwin & Heritage, 1990; Heath & Luff, 1993). In an academic setting, video-, film- and audio-recordings, as well as paper, may be used. Traditionally these are used asymmetrically, as the facilitator/instructor prepares these records to make a point or to serve as an anchor for discussion, and controls their play during the discourse. Computer-facilitated meetings for doing video analysis – where each participant has a computer and is network-connected with other participants and to external networks for information access – bring new challenges in terms of managing a coherent group discourse.

Illustrating the problems to be resolved

In an attempt to make these problems concrete, we discuss them in terms of two contexts in which there are ongoing efforts to engage in face-to-face and distributed CSCVA: (1) in several undergraduate *courses* with video-based instructional activities, and (2) in learning sciences *research* concerning informal learning.

1. Undergraduate courses utilizing video

Over the past several years we have worked with a number of college-level instructors at large research universities who had all made previous attempts

to enact video-based discussions into their lessons. These courses included a film-studies course and a Japanese language course at a West Coast university, and a film-production course at a Midwestern university. In each setting the goal was to present students with one or more video artifacts and have the students generate insights using an analytic frame provided by the instructor. For example, students in the film-studies course looked at two video clips, the 'Crispin's Day' speech in the film adaptation of Shakespeare's *Henry V* from the 1989 version directed by and starring Kenneth Branagh, and the same speech from the 1944 film version of the same play as directed and acted by Sir Laurence Olivier. The instructor's objective was to have students comparatively analyze how the same source text was translated to film by two different directors/actors. In past instantiations of the course, students had to make the comparison for this assignment by reconstructing the film events from memory for a written essay. In the CSCVA scenario, students are faced with the reference problem: in making their argument they have no way to explicitly refer to an important aspect of the video film-clip that they may have noticed in their initial viewing (e.g. an aggressive gesture by Branagh at a point where Oliver is subdued).

In the film-production course, students are also asked to analyze video clips; but in this case the clips are created by fellow student filmmakers. Students in this course are responsible for making their own experimental film available for others to view and for providing feedback on at least two other student films. How valuable one student's feedback is for another student depends largely on the issue of attentional alignment: how effective is the critiquing student in conveying to the film's creator the focus of the critique? For example, one student shot a music video for his film project, and another student commented: 'This is my favorite sequence. The transition between band members works well with the music and beat.' Without a persistent record of what precisely the critiquing student was looking at when she made this comment, it carries little meaning. The film-production course also exemplifies the problem of search and retrieval. As the critique of other students' films is a class assignment, the course instructor must be able to efficiently access the corpus of comments, sorted by author or target film, so as to evaluate the quality of the feedback.

In the Japanese language course, the instructor's desire is to engage her students in a discussion about the dialog styles used by the actors in a set of video clips she has collected. Specifically she seeks to elicit insights from her students that will illustrate a sophisticated understanding of conversational Japanese, and so she uses a diverse set of sources, ranging from pop-culture artifacts (e.g. Japanese soap operas and anime) to videotaped interviews with native Japanese speakers. Although these insights are not

formally assessed, it is important to the instructor that all students contribute to the discussion and that these insights build off one another. Structuring a productive multi-party discourse of this type, centered around a number of video records, can be difficult for an instructor to accomplish given the lack of precedent for video-anchored conversation.

2. *Learning sciences research: collaborative video analyses*

One of the studies being conducted as part of the LIFE Center (Learning in Informal and Formal Environments), an NSF-funded Science of Learning Center, investigates the informal use and learning of mathematics in the family environment. The research team, which consists of two faculty members and several graduate students, has been investigating the contexts and activities in which middle-school-aged learners and their families engage in mathematical problem solving. In interview sessions and observations lasting roughly two hours, the team has been working with over 20 families representative of California's population to identify the contexts and situations in which families participate where mathematical learning and practice take place. Videos from these interviews have been digitized and coding schemes have been developed that capture the nuances of family math *in situ*. The research team seeks to describe the resources family members use for recognizing and solving problems, characterize the structure of their mathematical activities, and analyze the social conditions and arrangements for their family-based mathematics practices.

In conducting its analysis of this data corpus, the research team engaged in face-to-face group meetings to review video records as well as in independent work in which a team member offered an analysis that could be reviewed by other members at a later date. The team also conducted collaborative analyses of the dataset with researchers at other institutions. Such an ambitious multi-party analysis effort demands effective solutions to the seven socio-technical design challenges, particularly the integration of insights and the establishment of coherent discourse conventions.

The DIVER software environment for video collaboration

DIVER (Digital Interactive Video Exploration and Reflection) is a software environment first developed to explore research uses of panoramic video records that encompass 360-degree imagery from a dynamic visual environment such as a classroom (Pea et al., 2004). The current version of the

DIVER platform allows a user to control a ‘virtual camera window’ overlaid on a standard digital video record such that the user can ‘point’ to the parts of the video they wish to highlight. The user can then associate text annotations with the segments of the video being referenced and publish these annotations online so that others can experience the user’s perspective and respond with comments of their own. In this way, DIVER makes it possible to create an infinite number of new digital video clips and remix compilations from a single source video recording. As we have developed a web-enabled version of DIVER that allows for distributed access and the annotation of digital video direct from handheld digital video recorders, our focus has shifted to supporting collaborative video analysis and the emerging prospects for digital video collaboratories. DIVER and its evolving capabilities have been put to work in support of collaborative video analysis for a number of diverse research and educational activities.

We refer to the central work product in DIVER as a ‘dive’ (as in ‘dive into the video’). A dive consists of a set of XML metadata pointers to segments of digital video stored in a database and their associated text annotations. In authoring dives on streaming videos via any web-browser, a user is directing the attention of others who view the dive to see what the author sees; it is a process we call ‘guided noticing’ (Pea et al., 2004). To make a dive using DIVER, a user logs in and chooses any video record that has been made available in the searchable database. A single dive can be constructed by a single user or by multiple users each of whom contributes their own interpretations that build off the interpretations of the other users. The target video can be viewed using standard video controllers. As the video plays, a user manipulates a virtual camera viewfinder (the rectangle inside the video window shown in Figure 1) on top of the video to focus in on a specific area of interest.

By clicking the ‘Mark’ button, the user saves a reference to a specific point in space and time within the video and places it within a data container – a single panel that resides inside the DIVER ‘worksheet’ on the right side of the interface (see Figure 1). Inside the worksheet, a mark is represented by a thumbnail image of the video record. Once the mark has been added to the worksheet, the user can comment on that mark by entering text in the panel. Panels are also created by clicking on the ‘Record’ button, an action that creates a pointer to an entire segment of the video, and stores the path taken by the virtual viewfinder during that segment. Like a mark, a recorded clip can be annotated by adding text within its associated panel on the worksheet. The DIVER user can replay the recorded video segment or access the point in the video referenced in a mark by clicking on its thumbnail image.



FIGURE 1
WebDIVER user interface

Note: The rectangle inside the video window represents a virtual camera viewfinder that is controlled by the user's mouse to make a movie inside the movie with the software.

Assuming they have appropriate permissions assigned by a system-administration process, multiple users can access a dive simultaneously, with each user able to add new panels or make comments on a panel that another user has created. Users are notified in real time when another user has made a contribution to the dive. Thus users may be either face-to-face in a meeting room or connected to the same webpage remotely via networking as they build a collaborative video analysis. In principle and in practice, there is no need for the users to be watching the same portions of the video at the same time. As the video is streamed to them through their web-browser, they may mark and record and comment at their own pace and as a function of their own interests. Collaborative video analysis activity using DIVER can be as purposeful or as emergent as participants choose to make it; constraints on focus, intent, duration of sessions and so on are not built into the technology but are a matter of negotiated social practice among collaborative team members.

CSCVA workflow with DIVER: Addressing the socio-technical design issues

With a brief sketch of DIVER and a presentation of examples in which research or learning groups have been using DIVER in support of collaborative video analysis, we now may reflect on the workflow steps and the activity systems in which the DIVER platform is playing an instrumental role. We have been learning a great deal about the unique affordances of this video-analysis environment for the nature of such work, and unearthing new challenges.

1. The problem of reference. DIVER provides a virtual camera viewfinder for a user to inscribe a video region of interest as their referential act (Stevens, Cherry & Fournier, 2002, use a pointing gesture symbolized by an image of a hand). Text annotation is used for free-field text interpretations of the referred-to video content. Virtual pointing to a video region and associated annotations provide the communication infrastructure for a dive author to make *link-addressable* references to the dynamic medium of video in their conversations.

2. The problem of attentional alignment. The method of virtual pointing to a circumscribed part of the space–time continuum in a video record enables distributed users to focus their attention on the same regions of the video as the basis of their interpretive work. With DIVER’s web-based methods, users can align attention to the parts of the video that are pertinent to their discourse regardless of whether they are synchronously or asynchronously connected. These properties of DIVER have enabled immensely productive and cumulative analyses of informal mathematics learning among the Family Math Project collaborating team members even as they work at different locations and at disparate times on their video-interpretive task activities.

3. The problem of creating ‘immutable mobiles’ from video recordings. Latour’s concept of written texts as immutable mobiles reviewed earlier has considerable applicability in considering how a dive establishes a video immutable mobile. The digital inscriptions provided by the lightweight metadata of a dive are mobile, immutable when they move (in that they preserve their character across geographical locations, computer platforms and different web-browsers), are made ‘flat’ (2-D), can be flexibly modified in scale (through projection), can be cheaply reproduced and spread (thanks to Internet standards), allow one to ‘superimpose different images of totally different origins and scales’ (in their digital document forms), can be made

part of a written text (through hyperlinking to dives) and may be merged with geometry thanks to their two-dimensionality.

4. The problem of effective search, retrieval and experiencing of collaborative video work. DIVER provides Google-like search for any term or phrase used in the full text of annotations, title or user name. A DIVER search returns a list of dive panels with the search term or phrase highlighted and the video keyframe for that panel. In terms of retrieval, clicking on such a list item or keyframe opens up the affiliated dive and enables the user to view the precise video regions in space/time to which the searched-for terms were applied. The value of the metadata tagging of a user community for research videos will grow tremendously as multiple researchers work with a dataset and develop cumulative analyses across projects. In the informal math-learning work, team members are finding it easy to compile the accumulated annotations of a multi-party research group on specific phenomena such as the use of props, symbol manipulation, emotion terms and the like. The result is the creation of composite codebooks that are dives resulting from all exemplars of a given type, identified across multiple researchers, and now able to be experienced in a sequential and consumable remix of the clips.

5. The problem of permissions. DIVER provides access to video records and their dives only for users granted the appropriate permissions. Administrative tools enable the formation of groups and the designation of whether specific video assets and dives can be viewed, copied, edited or deleted.

6. Integrating the insights of a collective group. Some of the most dramatic changes brought about by DIVER-enabled collaborative video analysis concern this design problem. Whereas sequential turn-taking is required in collaborative face-to-face video analysis work practices, parallel analyses can be carried out with DIVER, to cumulative effects. Multiple individuals can be streaming the same video file, looking at different parts of it at the same time, and making their dive recordings and annotations without control from the single group leader customary in video-interaction analysis sessions (Jordan & Henderson, 1995). Stopping and starting the video being played, setting in- and out-points to segments to be annotated, selecting different segments to compare in an analysis can all be carried out in parallel when DIVER is used by a distributed group of analysts. For its uses in computer-supported collaborative video analyses, DIVER thus shifts control from the group facilitator to individual participants, who can play the entire source or can use the cues from one another's ongoing analyses to focus

their attention on subparts of the video. In relation to scientific-inquiry processes, there is the added benefit that, as one posts conjectures in the form of DIVER annotations concerning the interpretations of an event, multiple analysts can seek out confirming or disconfirming evidence elsewhere in the video data records, and then post such evidence as comments or as new panels in that dive.

7. The problem of establishing coherent multi-party video-anchored discourse. Our experiences with DIVER for video analysis in collaborative groups have only recently been making headway on this dimension. DIVER has largely been used in an asynchronous manner for collaborative group work or in face-to-face meetings in which a projected screen is used to coordinate group attention. Workgroup participants take turns by ordinary social conventions in selecting dives they would like to share or ask questions about from other members of the workgroup.

A recent development is proving promising, however. To lower the skills of entry to DIVER video interactions, we have separated the roles of *authoring* dives from *observing* dives. A dive author (or authoring team) can send an email link to others with a live URL link that, when clicked upon, will open a DIVER server page (see Figure 2) that displays a DIVER ‘player’ by which the recipient can play and pause the video, and read the annotations



FIGURE 2

Illustration of a DIVER player for viewing the remix clips and comments from dive outside the WebDIVER authoring environment (e.g. on a webpage or blog)

made for each video segment by the dive's author – without the distractions of having to access the DIVER authoring environment. The DIVER author can also embed HTML code so that this lightweight DIVER 'player' will appear on a website or a blogging site such as Blogger. What is the result of such a separation of dive authoring and consumption? We are finding it easier for groups to establish a usage trajectory that begins with observing dives created by others, followed by being motivated through the experiencing of others' perspectives on video moments that mattered to them, to finally expressing their own analytic perspectives on video.

Towards a shareable corpus of multimedia social science data

The DIVER system distinctively enables what we call 'point-of-view' authoring of tours of existing video materials in a way that supports sharing, collaboration and knowledge building to establish a common ground of reference. We are in the early stages of documenting its uses as a cognitive technology for enhancing what can be achieved by such distributed activity systems as 'video research laboratories', comprising multiple researchers, multimedia datasets and video-analysis technologies. Addressing the seven socio-technical design challenges for computer-supported collaborative video analysis has been a fruitful beginning to this evolving adventure in making networked video more accessible to the kinds of conversational interactions and interlinking that we have grown accustomed to for text.

As in other scholarly domains of inquiry, there are fertile opportunities in education and in the learning sciences for developing new methods of work, knowledge creation and learning that leverage the collective intelligence of the field in ways analogous to the biological, health, earth and space sciences (e.g. Cerf et al., 1993; Finholt, 2003). Advances in the physical sciences depend upon a body of scientific data that can be shared, annotated, analyzed and debated by the community of scientists, as well as developments in the instruments and technologies that are integral to formulating, conducting and analyzing results from scientific investigations utilizing cyberinfrastructure (NSF, 2007). Today we do not yet have such a corpus of shareable video data of human interactions. The absence of a shareable corpus of social science data has hindered theory development in the human sciences and advances in the improvement of educational practices based on cumulative analyses of learning and teaching by multiple investigators. The DIVER project is devoted to filling this void by providing social and behavioral scientists with a software tool and platform for generating different perspectives on the rich multimedia data of human interactions, including those from learning and

teaching. The same technologies may also come to be useful for K-12 teachers who want to share their teaching insights and classroom expertise, and for researchers and other teachers alike to learn from.

Acknowledgments

The present article is an extended version of a paper prepared for the conference ‘Digital ethnography: analyzing activities in the information society’, Cognitive technologies seminar series, Fondation Maison des Sciences de l’Homme, Paris, France, December 2006. DIVER™, WebDIVER™, Dive™ and Guided Noticing™ are trademarks of Stanford University for DIVER software and services with patents pending. DIVER has been supported by grants from the National Science Foundation (#0216334, #0234456, #0326497, #0354453) and the Hewlett Foundation.

Roy Pea is a Stanford University Professor of the Learning Sciences and Director of the Stanford Center for Innovations in Learning (<http://scil.stanford.edu>). He has published over 140 chapters and articles on such topics as distributed cognition, learning and education fostered by advanced technologies, including scientific visualization, online communities, digital video collaboratories and wireless handheld computers (<http://www.stanford.edu/~roypea>). He is currently developing a new paradigm for everyday networked video interactions for learning and communications (<http://diver.stanford.edu>), and for how informal and formal learning can be better understood and connected, as Co-PI of the LIFE Center (<http://life-slc.org>), funded by the National Science Foundation as one of several large-scale national Science of Learning Centers. He was co-author of the 2000 National Academy Press volume *How People Learn*. He is a Fellow of the National Academy of Education, the American Psychological Society and the Center for Advanced Study in the Behavioral Sciences. In 2005–6, Roy Pea was President of the International Society for the Learning Sciences. He was also IBM Faculty Fellow in 2006. He received his doctorate in developmental psychology from Oxford as a Rhodes Scholar. He also serves as a Director for Teachscape, a company he co-founded in 1999 that provides comprehensive K-12 teacher professional development services incorporating web-based video case studies of standards-based teaching and communities of learners. *Author’s address*: Stanford Center for Innovations in Learning, Wallenberg Hall, Building 160, Stanford University, Stanford, CA 94305, USA. [*email*: roypea@stanford.edu]

Robb Lindgren is a doctoral candidate in Stanford University’s Learning Sciences and Technology Design program, where he studies the cognitive affordances of technologies for learning. He is a 4-year member of the DIVER project and has led efforts to implement this software platform in a wide range of professional settings, including teacher education and toy design. He is interested specifically in the ways technologies can offer new perspectives on phenomena such that they promote learning. Lindgren is currently completing his dissertation, which entails the study of how different perspectives offered in a virtual-world simulation environment affect learning outcomes. *Author’s address*: Stanford Center for Innovations in Learning, Wallenberg Hall, Building 160, Stanford University, Stanford, CA 94305, USA. [*email*: robblind@stanford.edu]

Joseph Rosen has over 15 years' experience as a programmer and engineer, developing digital media concepts into interactive prototypes and products. His expertise with digital video dates back to the late 1980s, when at Apple Computer he contributed to the time-based media experiments that were a precursor to QuickTime. He has been the Senior Research Software Engineer at the Stanford Center for Innovations in Learning since 2002. Rosen received a Master's degree in Interactive Telecommunications from NYU's Tisch School of the Arts (1989), where he was also an Adjunct Professor from 1990 through 1996. *Author's address:* Stanford Center for Innovations in Learning, Wallenberg Hall, Building 160, Stanford University, Stanford, CA 94305, USA. [email: joro@stanford.edu]

References

- Barron, B. (2003) 'When smart groups fail', *Journal of the learning sciences* 12(3): 307–59.
- Bruner, J. S. (1966) 'On cognitive growth', in J. S. Bruner, R. R. Olver & P. M. Greenfield (eds) *Studies in cognitive growth*, pp. 30–67. New York: Wiley.
- Cerf, V. G., Cameron, A. G. W., Lederberg, J., Russell, C. T., Schatz, B. R., Shames, P. M. B., Sproull, L. S., Weller, R. A. & Wulf, W. A. (1993) *National collaboratories: applying information technology for scientific research*. Washington, DC: National Academy Press.
- Clark, H. H. (1996) *Using language*. New York: Cambridge University Press.
- Derry, S. J., ed. (2007) *Guidelines for video research in education: recommendations from an expert panel*, prepared for the National Science Foundation, Interagency Education Research Initiative and the Data Research and Development Center. Arlington, VA: National Science Foundation. (available at: <http://drdc.uchicago.edu/what/video-research-guidelines.pdf>)
- Finholt, T. A. (2003) 'Collaboratories as a new form of scientific organization', *Economics of innovation and new technologies* 12(1): 5–25.
- Forster, E. M. (1927) *Aspects of the novel*. New York: Harcourt, Brace & World.
- Goldman, R., Pea, R. D., Barron, B. & Derry, S., eds (2007) *Video research in the learning sciences*. Mahwah, NJ: Lawrence Erlbaum.
- Goodwin, C. (1986) 'Gestures as a resource for the organization of mutual orientation', *Semiotica* 62: 29–49.
- Goodwin, C. & Heritage, J. (1990) 'Conversation analysis', *Annual review of anthropology* 19: 283–307.
- Heath, C. & Luff, P. (1993) 'Disembodied conduct: interactional asymmetries in video-mediated communication', in G. Button (ed.) *Technology in working order: studies of work, interaction, and technology*, pp. 35–54. London: Routledge.
- Jordan, B. & Henderson, A. (1995) 'Interaction analysis: foundations and practice', *Journal of the learning sciences* 4(1): 39–103.
- Lampert, M. & Hawkins, J. (1998) *New technologies for the study of teaching*, Report to the National Science Foundation from a workshop held 9–11 June 1998, Ann Arbor, Michigan. (NSF Grant #REC-9802685)
- Latour, B. (1986) 'Visualization and cognition: thinking with eyes and hands', in H. Kuklick & E. Long (eds) *Knowledge and society*, pp. 1–40. Greenwich, CT: JAI Press.
- Levy, D. M. (2001) *Scrolling forward: making sense of documents in the digital age*. New York: Arcade Publishing.

- MacWhinney, B. & Snow, C. (1999) *Multimodal studies of instructional discourse*, Report to the National Science Foundation from the Classroom Discourse workshop. Pittsburgh, PA: Carnegie Mellon University. (available at: <http://tinyurl.com/ys2rl8>)
- Marx, K. & Engels, F. (1978) 'A critique of the German ideology', in R. Tucker (ed.) *The Marx-Engels reader*, pp. 146–200. New York: Norton, 2nd edn. (orig. pub. 1846)
- Mills, M. I. & Pea, R. D. (1989) 'Mind and media in dialog: issues in multimedia composition', in K. Hooper & S. Ambron (eds) *Full-spectrum learning*, pp. 14–37. Cupertino, CA: Apple Computer, Inc.
- National Science Foundation, Cyberinfrastructure Council (2007) *Cyberinfrastructure vision for 21st century discovery*. (available at: <http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf> and downloaded 29 December 2007)
- Pea, R. D. (1985) 'Beyond amplification: using computers to reorganize human mental functioning', *Educational psychologist* 20: 167–82.
- Pea, R. D. (1987) 'Cognitive technologies for mathematics education', in A. Schoenfeld (ed.) *Cognitive science and mathematics education*, pp. 89–122. Hillsdale, NJ: Erlbaum.
- Pea, R. D. (2006) 'Video-as-data and digital video manipulation techniques for transforming learning sciences research, education and other cultural practices', in J. Weiss, J. Nolan & P. Trifonas (eds) *International handbook of virtual learning environments*, pp. 1321–93. Dordrecht: Kluwer Academic.
- Pea, R. D. & Hay, K. (2003) *CILT workshop on Digital Video Inquiry in Learning and Education, 25–6 November 2002*, Report to the National Science Foundation. Palo Alto, CA: Stanford Center for Innovations in Learning. (based on NSF grant #0124012)
- Pea, R. & Hoffert, E. (2007) 'Video workflow in the learning sciences: prospects of emerging technologies for augmenting work practices', in R. Goldman, R. D. Pea, B. Barron & S. Derry (eds) *Video research in the learning sciences*, pp. 427–60. Mahwah, NJ: Lawrence Erlbaum.
- Pea, R. D. & Kurland, D. M. (1987) 'Cognitive technologies for writing development', in L. Frase (ed.) *Review of research in education*, vol. 14, pp. 277–326. Washington, DC: AERA Press.
- Pea, R., Mills, M., Rosen, J., Dauber, K., Effelsberg, W. & Hoffert, E. (2004) 'The DIVER™ project: interactive digital video repurposing', *IEEE Multimedia* 11(1): 54–61.
- Scacchi, W. (2004) 'Socio-technical design', in W. S. Bainbridge (ed.) *The encyclopedia of human-computer interaction*. Great Barrington, MA: Berkshire Publishing Group.
- Stevens, R., & Hall, R. (1997) 'Seeing tornado: how video traces mediate visitor understandings of (natural?) phenomena in a science museum', *Science education* 81(6): 735–47 (special issue on informal science education).
- Stevens, R., Cherry, G. & Fournier, J. (2002) 'Video traces: rich media annotations for teaching and learning', in G. Stahl (ed.) *Proceedings of the computer supported collaborative learning conference*. Mahwah, NJ: Lawrence Erlbaum.
- Vygotsky, L. S. (1978) *Mind in society: the development of higher psychological processes*. Cambridge, MA: Harvard University Press.