INFORMATION PAGE

Cat's Cradle: Working with other people in overlapping real and virtual worlds through tangled strands of visual and other media

Mike Robinson, Samuli Pekkola
Department of Computer Science
University of Jyväskylä
PO Box 35, SF-40351, Jyväskylä, Finland
+358-14-603095/603245
mike@cs.jyu.fi, samuli@studio.it.jyu.fi

Dave Snowdon
Xerox Research Centre Europe
6, chemin de Maupertuis
38240 Meylan, France
+33 (0)4 76 61 5180
Dave.Snowdon@xrce.xerox.com

Abstract: Experience of small and medium enterprises shows successful collaborative projects are a function of human and technical networking. Both would be enhanced by the provision of 'virtual offices' that allow for file sharing and for awareness of the presence and activities of others. An advanced Virtual Reality multi media and 'virtual office' project is described. Some new issues arise when combining file handling and exchange on the World Wide Web, Virtual Reality, video, audio. Technical issues concern the best local and transport architectures. The ability to change and switch between workgroup boundaries with multiple media raises inportant social, interface, and technical questions. One of these is the relation between the 'spatial model' and the 'address model' at the local and network architecture levels.

Keywords: Virtual Office, VR, multi-media, person-awareness, separate channel architecture

Paper Type: Research & Technical Report

Contact Author: Robinson

Paper size: A4 12-point

Cat's Cradle: Working with other people in overlapping real and virtual worlds through tangled strands of visual and other media[1]

Mike Robinson, Samuli Pekkola
Department of Computer Science
University of Jyväskylä
PO Box 35, SF-40351, Jyväskylä, Finland
mike@cs.jyu.fi,samuli@studio.it.jyu.fi

Dave Snowdon
Xerox Research Centre Europe
6, chemin de Maupertuis
38240 Meylan, France
Dave.Snowdon@xrce.xerox.com

Abstract: Experience of small and medium enterprises shows successful collaborative projects are a function of human and technical networking. Both would be enhanced by the provision of 'virtual offices' that allow for file sharing and for awareness of the presence and activities of others. An advanced Virtual Reality multi-media and 'virtual office' project is described. Some new issues arise when combining file handling and exchange on the World Wide Web, Virtual Reality, video, audio. Technical issues concern the best local and transport architectures. The ability to change and switch between workgroup boundaries with multiple media raises important social, interface, and technical questions. One of these is the relation between the 'spatial model' and the 'address model' at the local and network architecture levels.

Keywords: Virtual Office, VR, multi-media, person-awareness, separate channel architecture

Introduction

'Cats Cradle' is a children's game in which all ten fingers act as posts for a complex 'cradle' made of a single strand of wool. If the tension is too great or too little, or if one loop is dropped, the whole construction unravels. The situation is remarkably similar to the one faced by modern enterprises juggling multiple complex projects. It is also remarkably similar to the situation faced by designers of infrastructure and applications to support multi-media working. Many lessons can be learned from the last decade of CSCW research, but the technologies are changing so rapidly that practice is once again marching ahead of theory.

In this paper we will draw on one study of office work, and many established principles from CSCW to derive a design basis for a 'populated virtual office'. We will argue that in real office work — while document handling and file access is important — ongoing interactive access to other people is central to successful coordination, and the successful interleaving of multiple tasks and projects. The current generation of virtual offices gives excellent support for file handling, and this will be discussed in the next section. The next generation of World Wide Web interfaces and browsers also offer advanced, distributed file handling. But, disappointingly, access to other people is hardly supported at all.

[1] We wish to acknowledge the support of Nokia, Telecom Finland, TEKES. And our thanks to Lucy Suchman (Xerox PARC Palo Alto), Leigh Star & Geof Bowker (University of Illinois), and Andrew Clement (University of Toronto) for drawing our attention to important issues of Boundary Management at the Documentation as Performance Program VIVA Workshop: 14-15 May 1997

Many media are currently available for use in the Internet and World Wide Web for supporting *communication with*, rather than *awareness of* other people. New and improved video and audio applications appear almost daily. The first public internet phone connection was recently opened between the USA and Colombia. Applications increasingly support text with the audio and video conferencing (e.g. Intel's Pro-Share and CUSeeMe[2]) and Microsoft's NetMeeting supports distributed document editing. All of these applications use a 'phone call' metaphor. The connection is opened and then closed. They do not offer a 'permanent place' of the sort that Stone (1996) has argued is essential for the formation of Internet communities — and by implication, the maintenance of ongoing collaborative and cooperative project relationships and interactions.

Many have argued that Virtual Reality does provide a 'virtual place', and an adequate way of supporting awareness of other peoples presence in virtual locations. It has a significant advantage over video, in that it is far more lightweight (needs much less bandwidth). It is therefore inherently more suitable for heterogeneous networks in which both bandwidth, and the capacity and age of client machines can be very variable. Current VR applications have been successfully and widely used as 'social spaces' e.g. Habitat (REF) and Community Place (Lea et al. 1997). Some, such as DIVE (Carlsson and Hagsand 1993; Hagsand 1996) and Onlive! Traveller[3] support audio. DIVE is the only application so far to try to provide some support for rudimentary document handling from VR, of the sort that will be needed in virtual offices. Nevertheless, current VR systems are not designed with work (collaborative or otherwise) in mind. Nor are their distributive network architecture designed to be compatible with the most effective ways of distributing video and audio streams. Both these factors raises issues that are still to be resolved.

In this paper, we sketch an application, VIVA — an advanced Virtual Reality multi media and 'virtual office' demonstrator, that is intended to show the potential for integrating person awareness and multiple media with document handling interfaces and applications. Experience with the demonstrator shows the need for offering users complex boundary management tools. There are good reasons in CSCW for assuming that the boundaries of each medium (VR, video, audio, document, text, etc.) will need to be set independently or in combination. VIVA has adopted a 'separate channel' architecture to support this at the transport level, and at the level of user interface. This raises an issue of the relation between the 'spatial model' (Benford and Fahlén 1993; Snowdon et al. forthcoming) and the more traditional 'address model'. The spatial model originated to deal with the changing perspectives and possible interactions generated by movement within a 3D world. It has considerable appeal as a general model for the control of multiple media, and has been adopted for this by e.g. Berlage (Forthcoming). The address model is a generic term for the selection of a group of people from lists, as with email, or with access permissions. Both models have advantages, both involve a trade-off between ease of use and complexity, and each has different implications and difficulties for architecture and networking.

Office Work and Virtual Office Work

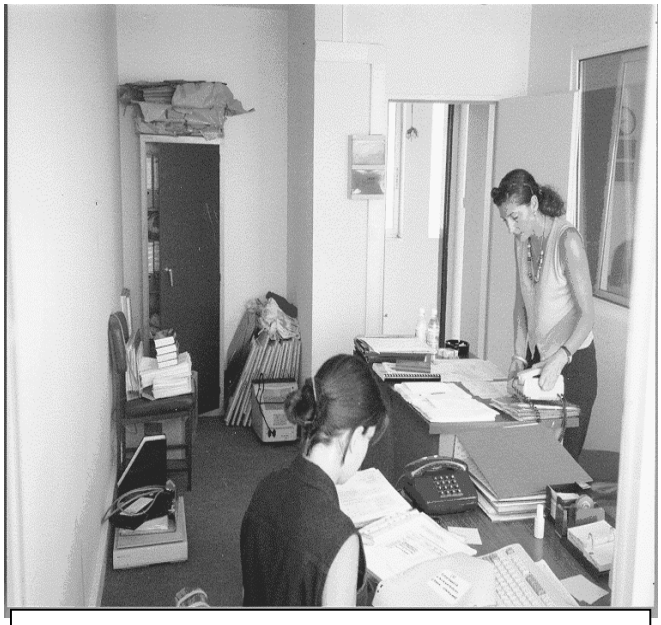
A comparative study of office work in 10 European countries (England, Ireland, Belgium, The Netherlands, France, Germany, Denmark, Finland, Spain, Greece) has recently been completed (Robinson and Hinrichs 1997). All the offices were concerned

---

[2] From White Pine Software (http://www.cu-seeme.com)

[3] From Onlive! Technologies (http://www.onlive.com)

with “local employment creation” in one form or another. The organisations ranged from small voluntary groups, often with less than 5 paid workers, to offices with 30 or more workers handling projects and contracts for local authorities. The largest employed more than 400 workers, and, as well as local projects, conducted multi-national consultancy (China, Europe, South Africa, and many others). All the offices were concerned with running local projects, and many took part in national and international projects as well. Many of the offices and organisations were linked in human networks, with a coordinating office, meetings, and regular exchanges of staff and experiences. A fairly typical small (Greek) office is shown in Figure 1. We reproduce the photograph to illustrate the sort of very ordinary, familiar, and unromantic office situation that we have in mind when designing Internet support.



Researchers visited offices in each country for up to a week, and examine the day-to-day, and minute-by-minute work activities — with a special interest in the levels of Internet use, the applications that were found useful, and for what. Many conclusions were drawn from the work studies, and from a more general survey (Robinson and Hinrichs 1997). One of the main ones was that “the *constantly changing and developing projects, project partners, and inter-LEI cooperation* compel the conclusion that only the Internet has sufficient “reach” to serve as the base communication space for LEI’s. In other words, private networks (with fixed memberships, and fees) do not provide sufficient flexibility. This has led us to focus our own CSCW work on the Internet (and in particular the easier-to-use WWW) rather than private networks or even LAN’s.

Another conclusion was that small organisations (5 people or so) were not handicapped in respect to Internet access or staff ability to use it. The latter was often greater than in larger organisations (30 people or so). Although direct Internet connections and Local Area Networks provide ‘better access’ in a technical sense, they are also demanding on staff time, and access skills tend to centre on experts (or a small computer department), and do not diffuse to the staff as a whole. Small organisations used the Internet just as much as the larger, and were also early adopters of other technologies, such as mobile phones/GSM’s. They differed in the type of access, which was usually a modem connection. This led us to a concern with support for *heterogeneous* networks. Many research institutions focus on state-of-the-art machines and networks, and neglect slower, older client machines, and low bandwidth connections. We believe that usable systems for cooperative working need to take such heterogeneity into account. Another powerful argument for this position can be found in studies of a large distributed scientific community (Star and Ruhleder 1996). Generally this means that ‘graceful degradation’ is a feature virtual offices should support. Both high-end workstations with fast connections and modem/mobile devices with limited processing power and bandwidth need to be able to link coherently to the same applications.

A third conclusion was that human networks were more important than technical networks, but that a combination of the two was dramatically more effective.

The success of each office studied could be broadly measured in terms of the number of projects it could deal with. This is illustrated in Figure 2. Here we see that even relatively small offices manage to deal with an average of 3 international and 4 national, as well as many local projects — given human and technical networking. However technical networking alone, without the accompanying human networking does not achieve anything. This emphasises that *work practices at the human level* need to be seriously taken into account by virtual office designers — confirming much ethnographic work in CSCW. Technologies and technical fixes alone do not change anything, even where they can be 'sold'.

Projects (Type &Average no.)	**Human** & **Technical** Network	**Human** Network only	**Technical** Network only	Without Network
International	2.8	1.8	0.0	0.0
National	3.9	2.6	0.0	0.7
Local	9.5	4.2	1.4	1.0

Figure 2. Projects managed by European LEI Offices.

We have distilled three main qualitative principles (Pekkola and Robinson 1997) from CSCW to inform the design of virtual offices and facilities to be supported: flexibility; common artefact; and "beyond being there".

flexibility (Keil-Slawik 1992; Robinson 1993) Application constraints (over and above those constraints of the medium itself) should *not* be based on anticipating use. For instance, it is usual that audio needs to be higher quality than (i.e. have priority over) video to maintain smooth communication (Tang and Isaacs 1993) — but it is a mistake to "hardwire" this into the application, since there are circumstances were the opposite will apply (e.g. graphics reviewing). This indicates that virtual workspaces should be explicitly available for tailoring by users. In general, the requirement for flexibility and tailorability reflect a large volume of CSCW research on the importance of contingent and ad hoc activities - for example (Gerson and Star 1986; Suchman 1987; Bowers et al. 1995) discuss "articulation work", "the work to make the system work", and "situated action" respectively.

> "*Without an understanding of articulation, the gap between requirements and the actual work process in the office will remain inaccessible to analysis. ..... When the articulation of the work is deleted in representations of that work, the resulting task descriptions can only be uneasily superimposed on the flow of work.*
>
> It will always be the case that in any local situation actors "fiddle" or shift requirements in order to get their work done in the face of local contingencies. We argue here that *such articulation* is not extraneous to requirements analysis, but *central to it.*"
>
> (Gerson and Star 1986)

common artefact (Robinson 1992) was originally explained in the following way:

> Mundane everyday life provides many examples of common artefacts. Many of them work so well that they are taken for granted, inconspicuous, and appear trivial. Most of them are not considered to have any great "technological altitude". Yet they contain important lessons for the design of computer applications to support people working together.
>
> The keyrack behind the reception desk in a hotel is one example. It a place for guests to leave and collect their keys. They can also see which other guests are in or out, and leave messages in the pigeon-

holes. Hotel staff use it to communicate with their colleagues, for instance by leaving bills, faxes, etc. that have to be given out to guests. The placing of keys, or the contents of pigeonholes, may convey information in themselves, or may be the subject of questions or discussion. Generally, certain operations are considered legitimate, while others are not. For example: usually only the receptionist can place keys or messages; keys have to be hung over appropriate numbers; etc. The keyrack is also a model of the hotel — a mapping between the number of hooks and the number of rooms. The model may include spaces for administrative areas, such as the manager's office, or the kitchen. This rudimentary model of the hotel also allows the keyrack to be regarded as a "template" that can be filled in. Putting a key on a hook "augments the model" with information about whether the guest is in or out. Additionally, a glance at the keyrack in the late evening gives an overview of the hotel occupancy.

Yet this keyrack is not foolproof, nor is it "active". There is nothing to prevent keys being hung in the wrong places, or lost. And it can be used in idiosyncratic ways. A guest might request that their key might be hung under someone else's key (when the second person returned they would be given both keys, and have access to both rooms). The room owner would of course know where the key was, and could ask for it if necessary..... Such use could violate the example of legitimate operations noted above. Conversely, there are certain things about a keyrack that are fixed, like the positions of the hooks. It is simply not possible to hang a key between two hooks, as there is nothing to hang it on. So potential uses are a result of physical properties, local conventions and rules, and situational activities.

With respect to computer applications — and in particular, the relation between communication and document handling, between human and technical networking — it was then said:

Neither situated action not articulation work imply that coordination can be accomplished by speech *alone*. The concrete work situation and its common artefacts are critical in grounding conversation and spoken exchanges. They provide essential context for explicit communication.[4] "Double Level Language" is a phrase intended to catch the idea that implicit communication (through artefacts) and explicit communication (speech, ad hoc notes) are not alternatives, but complementary and mutually supportive.

> "In general it can be said that any non-trivial collective activity requires effective communication that allows both ambiguity and clarity. These ideas of ambiguity and clarity can be developed as the 'formal' and 'cultural' aspects of language as used by participants in projects and organizations. 'Computer support' is valuable insofar as it facilitates the separation and interaction between the 'formal' and the 'cultural.' Applications and restrictions that support one level at the expense of the other tend to fail.
>
> The formal level is essential as it provides a common reference point for participants. A sort of 'external world' that can be pointed at, and whose behaviour is rule-governed and predictable.
>
> The 'cultural' level is a different type of world. It is an interweaving of subjectivities in which the possible and the counterfactual […] are as significant as the 'given.' The formal level is meaningless without interpretation, and the cultural level is vacuous without being grounded." [(Robinson 1991): 43]

In addition to the need for Double Level Language, virtual office design focuses on "peripheral awareness" (Heath and Luff 1991) and "shared material" (Sørgaard 1988). *Peripheral awareness* (AKA "at a glance" or "out of the corner of the eye" awareness, and "overhearing") is the finding that much coordination is achieved by knowing, in a contextual and effortless way, what activities colleagues and workmates are engaged in. Peripheral awareness of colleagues and other workers in the same place has been shown to be essential to successful work in venues as diverse as London Underground Train Con-

---

[4] Ethnomethodology and Conversation Analysis have demonstrated the *indexical* character of spoken exchanges. "This" and "that" are obvious examples of words that do not have meaning outside specific contexts. It now appears that the majority of human interchanges need an awareness of the particular situation in which they are happening in order to be comprehensible.

trol rooms (Heath and Luff 1991; Luff et al. 1992), Aircraft Control rooms (Goodwin and Goodwin 1996), the London Stock Exchange (Heath et al. 1993), and Nuclear Power Plants (Kasbi and Montmollin 1991).

It is widely believed that VR/video workspaces can support such peripheral awareness (Dourish and Bly 1992; Huxor 1998). From the point of view of virtual offices, the question arises of boundaries within which peripheral awareness is supported, and outside which it is excluded. Who should be "in the space?" — and especially if the space can be textured as VR, as video, as audio, or as a document store, "who should be in what space?"

Shared material is shorthand for the fact that much communication happens through shared artefacts, not just through dialogue. This is especially important for video conferencing, where, with Nardi (Nardi et al. 1993) and Saffo (Saffo 1997) we believe the design and development path should concentrate on accurate and appropriate representations of objects and processes rather than improved images of dialogue partners ('talking heads'). Who is doing the sharing of which image at any moment, and how, again raises the questions of awareness and setting boundaries.

"Beyond Being There" *(BBT)* is a *quality principle* theoretically identified by Hollan and Stornetta (1992) in a paper that also has important general implications for VR and Mixed Reality Applications. It argued that simulating face-to-face co-presence was the objective of most tele-application designers: to produce environments that were as close as possible to "being there". This does not parallel experience. A phone call or an email is often better, more effective, or more appropriate than a visit to another's office or a conversation. The authors argue that each medium has its own affordances, and that mere approximation to face-to-face is a bad design objective, is unattainable, and does not mirror experience. We take the point further, and believe that BBT is a *quality principle* for CSCW applications. An application has quality if and only if there are circumstances in which people *prefer* it to physical presence, which then becomes just one more item on the menu. The greater potential awareness of others, and tractability of virtual over physical boundaries is cause for optimism on the quality of applications.

Current Virtual Offices

In this section we describe briefly some of the functionality to support document handling and access that is already available. File and access facilities have been discussed elsewhere in the CSCW community and we base ours on BSCW (Bentley et al. 1997; Robinson and Hinrichs 1997). An example of a BSCW workspace is given as Figure 3. Several other products are now available[5] with similar functionalities, plus electronic brainstorming, voting, etc. At this level, there is no support for peripheral (or direct) awareness of others, but there is flexibility of document handling (shared material) and membership change (boundaries), as in the following list:

- structured sets of files and facilities accessible by multiple people regardless of location
- permission structures for accessing and editing files
- available change histories of objects (supporting member awareness of ongoing changes)
- tailorable interfaces and ability to change file structures
- member lists

[5] E.g. eRoom (Instinctive Technology), Facilitate.com (Facilitate.com Inc), involv Intranet (Changepoint), Teamwave Workplace (Teamwave), Instant Team-Room (Lotus)

- ability to invite new members, and remove existing members
- "visitor friendliness": the ability of existing member to invite visitors without a) centrally administered permissions, and b) without granting access permission to the visitor beyond the duration of the visit (not currently offered in available systems)
- ability to set up new workspaces/offices

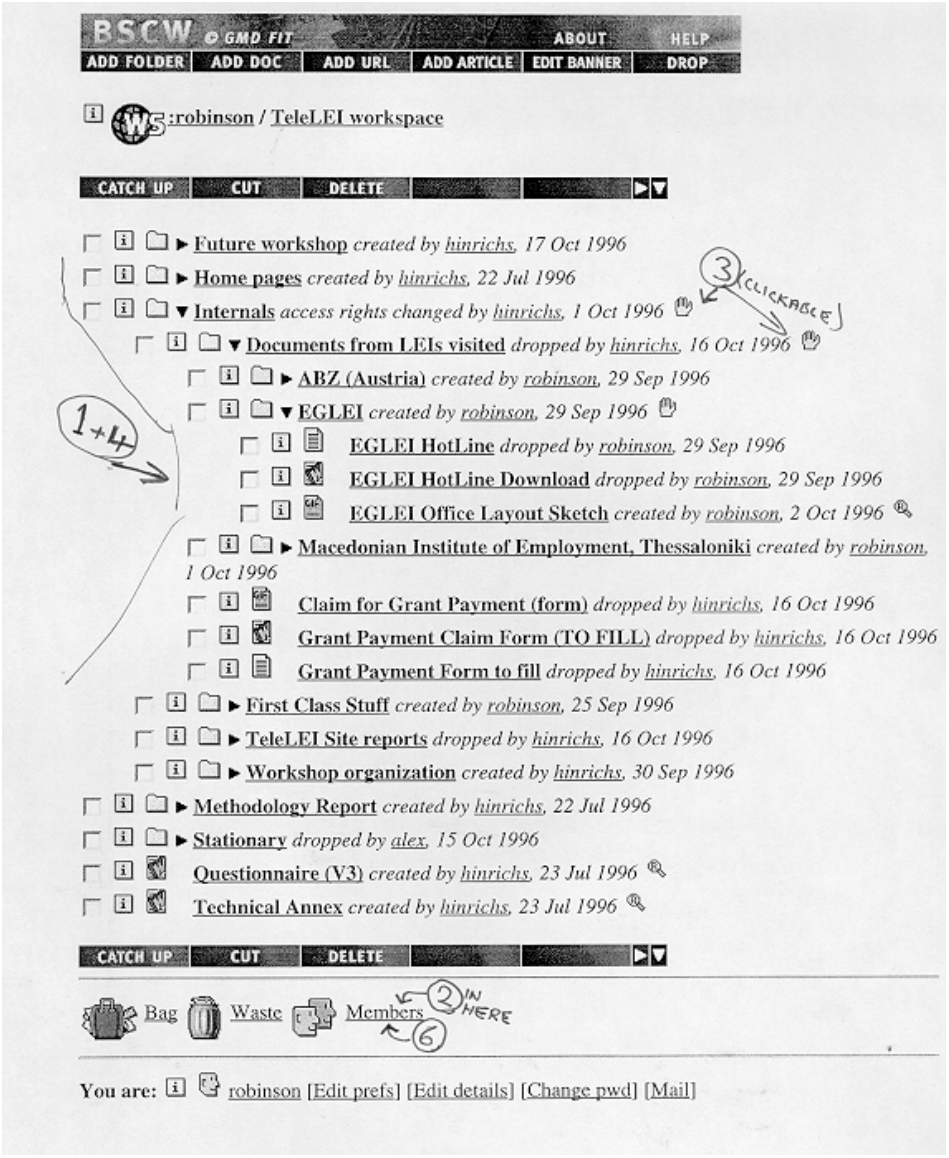


Figure 3: A BSCW Virtual Office as used by members of the TeleLEI Project (Robinson and Hinrichs 1997) from 10 European countries.

Person Awareness

The current generation of virtual offices (as noted above) do not support awareness of the presence of others. The July '97 issue of Byte magazine (Halfhill, 1997) produced a comprehensive survey of the projected new generation of web browsers. The functionality, in terms of document access and document handling, progresses in line with virtual office design. The headline implicitly claims a revolution, whereby Graphical User Interface is replaced by Network User Interface: "*Good-Bye GUI, Hello, NUI*" The article describes the "NUI's" — usually new browsers — emanating from the likes of Apple, IBM, Lotus, Microsoft, Netscape, Oracle, Sun, and others. Some of the more generic features heralded by the NUI are summarised in Figure 4.

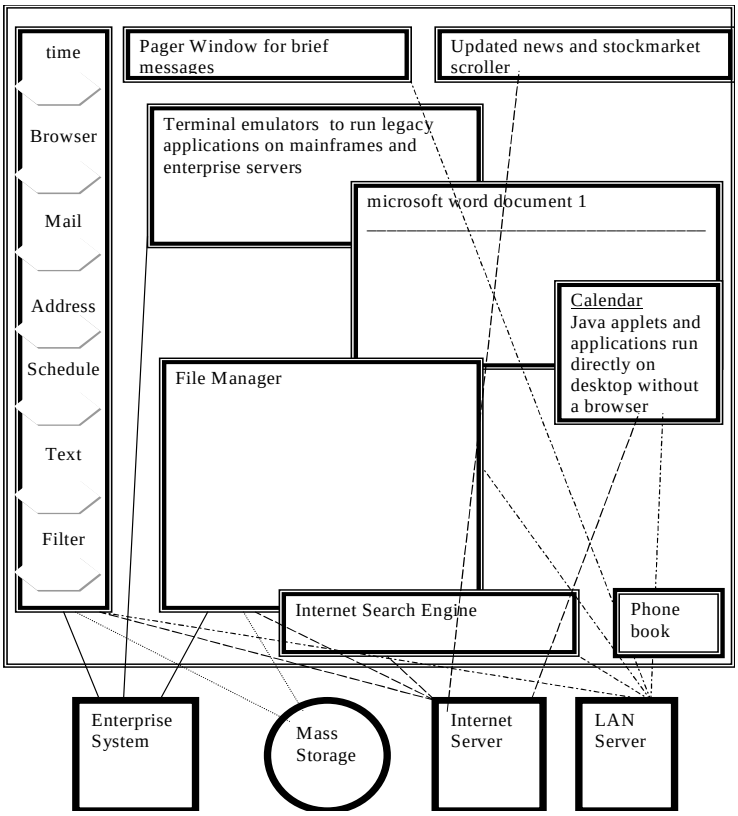


Figure 4. Network User Interface Schema.

NUI features include that the user sees consistent views of files/resources on local and remote drives (own disk, LAN, Internet server); that Windows, Mac, and UNIX applications all run (whether on the local system or remotely on servers); Java Virtual Machine and just-in-time compiler are standard components of the native OS; and multiple connections to multiple local and remote servers are well supported.

We note that the NUI suffers from a critical weakness. It is document-centric to the point of docuphilia. There are no people. There is almost no support for conversation or human interaction. A brief message window may be an advance, but it is not a big one. Yet as Mintzberg (1979) and others (Marx 1867; March and Simon 1958; Mathiassen 1987; Moran and Anderson 1990; Bannon 1991; Nardi and Miller 1991; Grudin 1994; Mogensen 1994; Dourish 1995) amply demonstrate in different ways, almost all human activity is collaborative, carried out in a socio-organisational context, and shaped in situ with reference to others. We concur, and believe future NUI's and virtual offices must at least support "knowing who is in the office".

With this in mind, we designed and implemented a PersonAwareness Engine for the Web. Technically it is lightweight and relatively simple. A "beeper" is downloaded to the client with the WWW page that notifies the WWW server of the start-time and end-time of the visit, and some other small items of information. From the log-file thus generated it is possible to monitor visit duration information. More important, it is relatively easy to inform other visitors who is on the same page at the same time, and to support the awareness of workgroup members, managers, and administrators of who is where in the Virtual Office (WWW site), even if they are using different programs and applications. Similar work is underway in other research centres e.g. Ulm University[6] in Germany. As soon as it becomes commonplace to know who is in the same office (or WWW page), all the currently available synchronous communication applications (VR, video, audio, etc.) will need to be linkable to support conversation and discussion as well as awareness.

Mixing Media

VIVA has identified the following general requirements for VR and multimedia communication in a virtual office.

- ability to attach comment, post notes, and send email

[6] http://www.cobrow.com/pages/

- multi-way Interactive Audio and Video: the ability to open multiple video or audio channel with other people/places through, within, or without the VR
- multi-way Interactive Text ("Chat") with graphics: the ability to open a text/graphics window/channel with one or more people who are "in the office" at the same time[7].
- simultaneous media use. VIVA should allow the maintenance of multiple simultaneous media and modalities (audio/ video/ VR conferences/ interactions, remote video streams, file accesses)
- multi-language support & independence of hard and software platforms are long term objectives
- scalability. VIVA will be upwardly and downwardly scaleable (see below).

The communications modalities listed are intended to be available from a VR interface (See Figure 5) or separately, allowing maximum flexibility, and supporting peripheral awareness and shared material through or bypassing the VR. While we are not convinced of the merits of "talking head" video conferencing we do believe it is necessary to provide support for it — in line with our philosophy of not hardwiring interaction styles. Rather than traditional video conferencing, it seems more likely that remote cameras will be used to support appropriate images of working objects and contexts e.g. (Ishii and Kobayashi 1993; Nardi et al. 1993; Kuzuoka et al. 1994). In this case, Saffo's (1997) image of " cameras aimed at everything everywhere, watched over by machines, and only occasionally examined by people" is likely to evoke a demand for multiple *open* video channels. This raises questions of the boundaries and accesses to each channel — and possibly of maintaining, switching between, and managing multiple peripheral awarenesses!

Figure 5. A screenshot of the VIVA prototype.

Network Architecture

The question of handling a number of boundaries and awarenesses is a challenging task. One possible solution is the separate channel architecture which, among other things, provides a mechanism for managing each medium and its boundaries separately. Generally, the architecture does not specify or suggest any transmission technique or network architecture, but provides a *conceptual* infrastructure. This section identifies some of the issues involved in applying the separate channel architecture to the virtual office design.

Each medium has different requirements from the computer and underlying network. Video requires a lot of bandwidth and relative low reliability while actions taken in pure VR have almost opposite needs. Reliability and latency requirements as well as the number of destination hosts may vary both statically (from one start-up time to another) and dynamically (i.e. while the application is running). The features of the network become

[7] We note that text "chat" is not a poor alternative to audio or video connections. There are occasions when text is better than audio — for instance when people have different mother languages and little speaking practice in the language being used (Robinson and Hinrichs 1997)

more and more important in dealing with crashes and overloads, and a great variety of different equipment. Our solution is isolate each medium in the separate channel architecture. We are currently discussing with Telecom organisations how this might be reconciled with current data compression techniques that increasingly combine media. Before presenting our network architecture, we shall briefly review a number of existing systems against which we can contrast our approach.

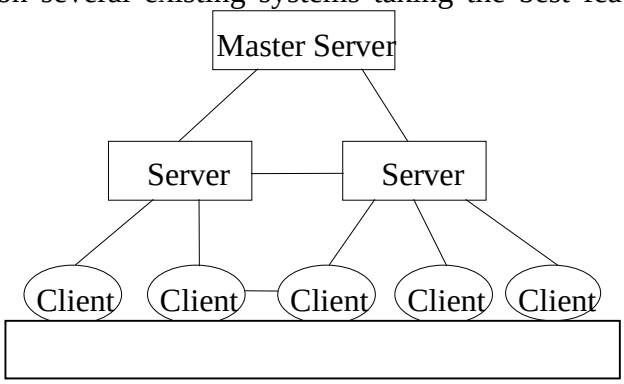
Current VR systems use many different network architectures and transmission techniques. For example DIVE (Hagsand 1996), MASSIVE-1 (Greenhalgh and Benford 1995) and MASSIVE-2 (Benford et al. 1997) are all based on server-less peer-to-peer architectures, although they use very different approaches of sharing the environment. The other extreme are client-server architectures, where a centralized server serves the clients. Examples of these are many VRML2 based systems (Brutzman et al. 1995; Husberg et al. 1997; Lea et al. 1997) . Both face problems of scalability.

One way to increase scalability is the heavy use of multicast protocols, such as IP-MULTICAST (Deering 1989), and cleaver partitioning of the environment. For example MASSIVE-2 allows spatial subdivision of a world into regions, each region having its own set of multicast groups, each used for different media. Partitioning also allows some specific settings for each region, e.g. a visual appearance of the content of the region might vary whether the observer is inside or outside. A member of a region can see the contents of the region, but an observer at the distance may see only an aggregate view of the contents (or nothing depending on the settings). This reduces network load and increases performance, since an observer at the distance gains only an overall image of the content (both visual and audio) which is necessary to transmit. Naturally the content of the region clears up when an observer encloses the region.

By partitioning the environment some social and technical problems arise. Greenhalgh and Benford (1997) drew out many significant issues of regions which should be flexible and natural to use, and are able to scale to different technical platforms (including computers and networks). MASSIVE-2, the system they used, improves scalability but does not answer to question of mixing VR with many different communication and collaboration tools.

Each system mentioned above has problems in handling multiple media and scaling both to network demand and to external resources (ordinary telephone or mobile networks, cable TV, etc.). Since all media are already separated from each other, the separate channel architecture provides natural starting point and platform for different network architectures.

The VIVA Virtual office has an architecture which is very flexible, since it is based on several existing systems taking the best features from each one. The basic idea is similar to NetEffect (Das et al. 1997). A master server manages overall infrastructure, client servers (there could be one or many) are connected to the master and other client servers, and coordinate the action occurring within a certain region in virtual reality. Each client is connected to client server. Figure 6 illustrates a basic structure of the architecture. This kind of tree-hierarchy is best for media which are not bandwidth or low latency demanding, but require high reliability (e.g. VR movement



information, non-verbal interaction). However, it cannot provide good response for media with direct communication and interaction (e.g. audio, video), since the servers form bottlenecks. To handle this kind of situation, the original NetEffect architecture needs some modification. Performance can be improved by allowing both the clients and the client servers to communicate directly with each other. So the VIVA network architecture combines the NetEffect, client-server and peer-to-peer architectures with the use of multicast protocols to ensure flexibility. In this, the VIVA approach has some similarities with the use of the spatial trader in MASSIVE-1 since once processes are put in contact with one another they are free to communicate directly.

Another modification is needed when some external applications are added or new networks interfaces are created. This is because ordinary telephone or mobile networks do not necessary fit to this architecture, and their management is totally different. If the connection is made from each client, the invoicing is easy, but the lack of hardware to make the physical connection to telephone line creates a problem. On the other hand, if the connection to telephone network is provided in any server, the invoicing is almost impossible, since finding the real user behind the nick names and connections from public computers is difficult.

When adding any external media to traditional VR, the importance of boundaries increases. Not only the heterogeneity of computers and networks hardware, but also any additional media related hardware, have different limitations and requirements. Therefore the boundaries set in VR are not necessary valid with any other media. For instance, walls and long distances limit awareness in real life, but the use of external media (telephone, TV, email, etc.) still provides a medium for communication. In virtual reality systems similar rules for external tools, their management and boundaries between them apply.

Implementation

The implementation of a technically and socially complex system with the multiplicities discussed earlier is extremely hard. The VR requirements of scalability and immediacy are enormously complicated by external communication devices. Some new methods have been developed, while many features and methods have been learned from existing systems.

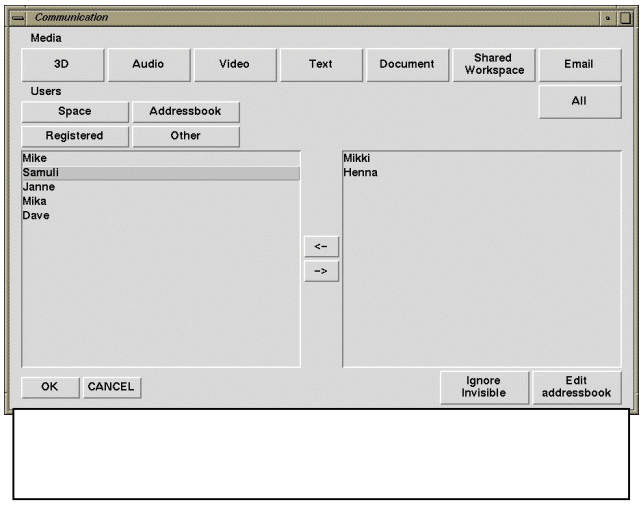
In mixed reality applications, interpersonal communication can occur entirely in the VR, between people in the VR and people not located in the VR, and outside the VR entirely. This lays the cornerstone of the VIVA Virtual Office design. The 3D user interface provides a central administrator tool, which controls all other communication applications. Aside from management, VR provides quasi-natural awareness of other people and their actions —the user is able to see, hear and communicate with the people within the same area.

From the technical point of view, two different approaches have been used to manage the communication between participants. The “spatial model of interaction” (Benford and Fahlén 1993) is used for interaction within the VR, while the “address model” is used for interaction through some other medium. The “address model” is simply the ability to select/filter-out individual and grouped recipients/senders. To manage and provide smooth, intuitive control for the user when combining these models is extremely problematic from the technical and interface points of view. Despite this, the need to provide user friendliness and realistic *ad hoc* user activity, forces this kind of control mechanism to be the most important issue in the whole system. Communication with

multiple media and external tools overrides but does not negate the importance of plain VR and the spatial model.

To manage “expression” (interaction, communication, awareness) in VIVA, both user interface and technical implementation are classified as either: direct communication or monitoring between users — who you are talking with, looking at, etc. and vice versa (*focus*); indirect communication between users — who is allowed to be aware of, overhear you, etc. and vice versa (*nimbus*); and *override* nimbus/focus[8] sets — allowing, amongst other things, the ability to establish private communication with a limited number of people, without being overheard, etc. An example user interface and it's complexity can be seen from the Figure 7 which includes a large number of different media and several ways of grouping people (VR workspaces, non-VR virtual workspaces, personal address books, etc.) As can be seen, all media and people groupings are separated, so both technical (e.g. different network architectures, and transport protocols) and social boundary management can be applied at many granularities. Consequently, the scalability and flexibility of the whole system increases since it can be adapted to many different situations.

However, separating media and people groupings creates some new problems. Since the user is able to modify her existence towards other people and other people's presence *independently for each single medium*, some confusing and contradictory situations may occur. For instance, if one sets herself invisible, e.g. disables her 3D representation and becomes as a ghost, other participants may not want to communicate with her since they do not know with whom they are interacting. Therefore the simple option to get rid of such ghosts is provided. The "Ignore Invisible" button (see Figure 7) is essentially a ‘reciprocate invisibility’ command, and means that audio, video, VR ghosts are exorcised from your world: you don’t see them and vice versa.



Another complicated case arises when sharing a single medium with several groups of people. For example, in traditional communication situations a person is able to handle a phone conversation while simultaneously discussing and even overhearing any other possible sources of sound e.g. (Luff 1997). Swapping between different groups of people and media (e.g. from phone to face-to-face or another phone conversation) is “natural”, relatively uncomplicated, and does not require much special effort. These kind of situations are much more complex with mixed reality applications, because of the number of media with different boundaries, and because the audience is not to hand.

Most of the communication media are not problematic, since they can be produced in easily understandable and distinguishable ways (e.g. video and text messages in own windows). The critical issue is audio, since both VR environments and private communication may produce a lot of different sounds and noises. VIVA manages the situation by supporting the damping of VR sounds (both incoming and outgoing) when a private call

8 ‘focus’ and ‘nimbus’ are both taken from the spatial model (Benford and Fahlén 1993)

is being made. This allows the overhearing of presence and possibility to swap easily between the private call and the VR. However, sometimes total muting of an environment is needed for private conversation (so it is also supported), but then the awareness of the environment soundscape is naturally lost.

Being aware of the media which other users may be using at any moment is almost impossible, especially since the number and quality of media available for communication may change depending on time, and on location of users. This kind of problem does not occur so often when using traditional table-top computers in static locations, but becomes significant with mobile connections. With a mobile phone, the quality of service varies over a wide range from excellent to awful depending on the weather, location, etc. The importance of quality increases when transferring video (bandwidth requirements) or data which requires high level of reliability (movements in VR). Therefore every medium cannot be used all the time, and the awareness of usable media have to be provided to the user. To achieve this, more cooperation with telecomms network operators is needed, since all necessary detail is not yet publicly available.

Summing up and concluding

Multimedia and Virtual Reality office applications to support commercial activity over the Internet are in their infancy. Yet, with increasing availability of fast, wide bandwidth connections, it is predictable that: the use of continuous media (audio and video) and complex graphical applications (VR worlds and interfaces) will increase; that applications in different media will need to be interoperable, and configurable into office like constellations; that networks, especially wide area networks, will continue to be composed of links and machines of widely differing ages, quality, and power; and that with increasing wide area, wide band, multi-modality and multi-media communication, social and technical means for the construction, reconstruction, and management of accesses and boundaries will be needed.

Ethnographic studies of office work from CSCW give some indications about the type of feature that might be needed in multi-modal office-like structures for commercial work over the Internet. In general, there is a lack of design principles as technologies are developing and accumulating more rapidly than experiences. Because of this, many intuitive leaps are necessary, and the grounding in work processes has to be post hoc, as recommended in Participatory Design and Evolutionary Prototyping.

The experience of building and discussing around the uses, architecture, and functionality of a prototype multi-modal demonstrator Internet Office with a VR interface has enabled us to focus more clearly on the issues involved. Technical and social questions of access and boundaries have been considered. We are convinced by the merits of a “separate channel” architecture. This provides robustness in the face of heterogeneous and unevenly resourced networks. More than this, it provides a path to graceful degradation, where for instance mobiles can be linked with powerful workstations. A “separate channel” architecture is also supportive of, and consistent with providing users, individually and collectively, with the ability to manage complex combinations of social, organisational, and media accesses and boundaries.

The complexity of separate boundary management of many media, including the VR itself, over many distinct, possibly overlapping workgroups gives *flexibility* but raises serious questions of interface and ease-of-use. One path out would be to regard the “address” model as a general case of the “spatial model”. Proximity of users would be imaged by focus-nimbus driven from video, audio, etc. as well as VR. This raises the possibility that the manipulative conveniences of the spatial model (changing proximity by

movement) might be retrofitted to audio video communication. This would start to meet the *beyond being there* criterion for proximity quality discussed earlier, and would help align the affordances of technical boundaries with working practices. This is the subject of ongoing research. In the meanwhile, we hope we have started to raise the topics of mixed realities and their boundaries as a subject of critical enquiry.

References

Bannon, L. (1991). From human factors to human actors. Design at Work: Cooperative Design of Computer Systems. J. Greenbaum and M. Kyng. London and New Jersey., Lawrence Erlbaum.

Benford, S. and L. Fahlén (1993). A Spatial Model of Interaction in Large Virtual Environments. Proceedings of the Third European Conference on Computer Supported Cooperative Work - ECSCW'93. 13-17 Sept. Milan, Italy. G. de Michelis, C. Simone and K. Schmidt. Dordrecht, Kluwer Academic Publishers**:** 109-124.

Benford, S., C. Greenhalgh, et al. (1997). Crowded Collaborative Virtual Environments. ACM CHI'97.

Bentley, R., W. Appelt, et al. (1997). "Basic Support for Cooperative Work on the World Wide Web." International Journal of Human Computer Studies: Special Issue on Novel Applications of the WWW **46**(6): 827-846.

Berlage, T. and M. Sohlenkamp (Forthcoming). "Visualizing common artefacts to support awareness in computer-mediated cooperation." Computer Supported Cooperative Work: The Journal of Collaborative Computing.

Bowers, J., G. Button, et al. (1995). Workflow from within and without: Technology and cooperative work on the print industry shopfloor. Proceedings of the Fourth European Conference on Computer Supported Cooperative Work - ECSCW'95. 10-14 Sept. Stockholm, Sweden. H. Marmolin, Y. Sunblad and K. Schmidt. Dordrecht, Kluwer Academic Publishers.

Brutzman, D. P., M. R. Macedonia, et al. (1995). Internet Infrastructure Requirements for Virtual Environments. VRML Symposium, San Diego, CA, USA.

Carlsson, C. and O. Hagsand (1993). "DIVE — A platform for multi-user virtual environments." Computer & Graphics **17**(6): 663-669.

Das, T. K., G. Singh, et al. (1997). NetEffect: A Network Architecture for Large-Scale Multi-User Virtual Worlds. VRST'97: ACM Symposium on Virtual Reality Software and Technology, Lausanne, Switzerland. Sept. 15-17 1997, ACM.

Deering, S. (1989). Host extensions for IP multicasting, RFC1112.

Dourish, P. (1995). "Developing a reflective model of collaborative systems." ACM Transactions on Computer-Human Interaction(in press).

Dourish, P. and S. Bly (1992). Portholes: Supporting Awareness in a Distributed Work Group. Proceedings of ACM CHI'92, ACM.

Gerson, E. M. and S. L. Star (1986). "Analyzing Due Process in the Workplace." TOIS **4**(3): 257-270.

Goodwin, C. and M. Goodwin (1996). Formulating Planes: Seeing as a situated activity. Communication and Cognition at Work. Y. Engestrom, & Middleton, D. N.Y., Cambridge University Press**:** 61-95.

Greenhalgh, C. and S. Benford (1995). "MASSIVE: A Virtual Reality System for Tele-conferencing." ACM Transactions on Computer Human Interaction (TOCHI) 2nd(1): 239-261.

Greenhalgh, C. and S. Benford (1997). Boundaries, Awareness and Interaction in Collaborative Virtual Environments. 6th. International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises (WET-ICE). June 18-20., Cambridge, Mass. USA, IEEE.

Grudin, J. (1994). "Groupware and Social Dynamics: Eight Challenges for the Developers." CACM **37**(1): 92-105.

Hagsand, O. (1996). "Interactive Multiuser VE's in the DIVE System." .

Heath, C., M. Jirotka, et al. (1993). Unpacking Collaboration: The Interactional Organisation of Trading in a City Dealing Room. Proceedings of the Third European Conference on Computer Supported Cooperative Work - ECSCW'93. 13-17 Sept. Milan, Italy. G. de Michelis, C. Simone and K. Schmidt. Dordrecht, Kluwer Academic Publishers.

Heath, C. and P. Luff (1991). Collaborative Activity and Technological Design: Task Coordination in London Underground Control Rooms. ECSCW '91. Proceedings of the Second European Conference on Computer-Supported Cooperative Work. L. Bannon, M. Robinson and K. Schmidt. Amsterdam, Kluwer Academic Publishers**:** 65-80.

Hollan, J. and S. Stornetta (1992). Beyond being there. CHI '92: Striking a Balance, Monteray, CA., ACM.

Husberg, F., M. Blomqvist, et al. (1997). Shared Platform Independent VRML Worlds. XIV IMEKO World Congress: New Measurements - challenges and visions: Topical Workshop on Virtual Reality, Tampere, Finland.

Huxor, A. (1998). The Role of 3D Shared Worlds in Support of Chance Encounters in CSCW. *Forthcoming* Collaborative Virtual Environments (CVE'98).

Ishii, H. and M. Kobayashi (1993). ClearBoard: A Seamless Medium for Shared Drawing and Conversation with Eye Contact. Readings in Groupware and Computer Supported Cooperative Work: Assisting human-human collaboration. R. M. Baecker. San Mateo, CA, US, Morgan Kaufmann**:** 829-836.

Kasbi, C. and M. d. Montmollin (1991). Activity Without Decision and Responsibility: The Case of Nuclear Power Plants. Distributed Decision Making. Cognitive Models for Cooperative Work. B. B. J. Rasmussen, J. Leplat. Chichester, John Wiley & Sons**:** 275-283.

Keil-Slawik, R. (1992). Artefacts in Software Design. Software Development and Reality Construction. C. Floyd, Zuellighoven,H., Budde,R., & Keil-Slawic,R. Berlin, Springer.

Kuzuoka, H., T. Kosuge, et al. (1994). GestureCam: A Video Communication System for Remote Collaboration. CSCW 94: Transcending Boundaries, Chapel Hill, North Carolina, USA, ACM.

Lea, R., Y. Honda, et al. (1997). "Virtual Society: Collaboration in 3D Spaces on the Internet." Collaborative Computing **6**: 117-140.

Luff, P. (1997). Computers and Interaction: The Social Organisation of Human-Computer Interaction in the Workplace. Department of Sociology, University of Surrey.

Luff, P., C. Heath, et al. (1992). Tasks-in-Interaction: Paper and Screen Based Documentation in Collaborative Activity. Proceedings of ACM CSCW'92 Conference on Computer-Supported Cooperative Work**:** 163-170.March, J. G. and H. A. Simon (1958). Organizations. New York etc., John Wiley.

Marx, K. (1867). Das Kapital. Zur Kritik der politischen Ökonomie. Gesamtausgabe (MEGA). K. Marx and F. Engels. Hamburg. **II/5**.

Mathiassen, L. (1987). Systems, processes, and structures: A contribution to the theoretical foundation of information systems development and use. System Design for Human Development and Productivity: Participation and Beyond. P. D. e. al. Amsterdam, Elsevier North-Holland**:** 49-61.

Mintzberg, H. (1979). The Structuring of Organizations. A Synthesis of the Research. Englewood Cliffs, New Jersey, Prentice-Hall.

Mogensen, P. (1994). Challenging Practice - an approach to cooperative analysis, Computer Science Dept, Aarhus University.

Moran, T. P. and R. J. Anderson (1990). The Workaday World as a Paradigm for CSCW Design. Proceedings of ACM CSCW'90 Conference on Computer-Supported Cooperative Work**:** 381-393.

Nardi, B., H. Schwartz, et al. (1993). Turning Away from Talking Heads: The Use of Video-as-Data in Neurosurgery. Proc. INTERCHI '93, Amsterdam, 22-29 April, ACM.

Nardi, B. A. and J. R. Miller (1991). "Twinkling lights and nested loops: distributed problem solving and spreadsheet development." Int. J. Man-Machine Studies **34**: 161-184.

Pekkola, S. and M. Robinson (1997). The Jyväskylä studio a mixed reality telematic worksite for CSCW. BBCC International Conference Broadband Communications for Research and Education, Moscow. June 17-18, 1997, Russian Academy of Sciences.

Robinson, M. (1991). "Double-Level Languages and Co-operative Working." AI & Society **5**: 34-60.

Robinson, M. (1992). Computer Supported Cooperative Work: Cases and Concepts. Readings in Groupware and Computer Supported Cooperative Work: Assisting Human-Human Collaboration. R. Baecker. San Mateo, CA, Morgan Kaufmann.

Robinson, M. (1993). Design for unanticipated use .... ECSCW '93 (3rd. European Conference on Computer Supported Cooperative Work), Milan, Italy, Kluwer.

Robinson, M. and E. Hinrichs (1997). Study on the supporting telecommunications services and applications for networks of local employment initiatives (TeleLEI Project): Final Report. Sankt Augustin, Germany, GMD, Institute for Applied Information Technology (FIT), D 53754.

Saffo, P. (1997). "CACM." CACM "The Next 50 Years" **40**(2 (Feb.97)): 93-97.

Snowdon, D., C. Greenhalgh, et al. (forthcoming). "A Review of Distributed Architectures for Networked Virtual Reality." Virtual Reality: Research, Development and Applications.

Sørgaard, P. (1988). Object Oriented Programming and Computerised Shared Material. Second European Conference on Object Oriented Programming (ECOOP '88), Springer Verlag, Heidelberg.

Star, S., Leigh and K. Ruhleder (1996). "Steps towards an Ecology of Infrastructure." Information Systems Research **forthcoming**.

Stone, A. R. (1996). The War of Desire and Technology at the Close of the Mechanical Age. Cambridge, Mass., MIT Press.

Suchman, L. (1987). Plans and situated actions. The problem of human-machine communication. Cambridge, Cambridge University Press.

Tang, J. and E. Isaacs (1993). “Why Do Users Like Video? Studies of Multimedia-Supported Collaboration.” CSCW **1**(3).