Transparent Computer Shared Cooperative Workspace (T-CSCW) Overview

John C. Checco

Abstract:
The purpose of this paper is to define the architecural specifications for creating the Transparent CSCW model as a consumer based product.

Rev #	Date	Description
0.10	01-10-95	Initial Draft
0.20	01-25-95	Research References
0.30	02-01-95	Multi-Point Architectures
0.40	03-03-95	Details about Connectivity
0.50	04-10-95	Prototype Overview, Logic Flow, Acceptance Requirements, and I/O Control Commands
1.01	05-30-95	Proposed Possible Camera Video as Grayscale
2.01	04-20-96	Replaced dependency upon ISDN with suggestions for other technologies (HDSL).
3.00	08-13-96	Split document into Overview, Architecure, Functional, and Prototype specifications.

Table of Contents

Transparent Computer Shared Cooperative Workspace (T-CSCW) Overview

John C. Checco
John_Checco@compuserve.com

1 Introduction

Currently there is a trend in the marketplace for providing telecommuting capabilities to a large sector of the workforce. The wide acceptance of telecommuting relies on the ability for people to work as effectively from a remote site as they do from their normal work site. Because computers have become such an integral part of many people's work, remote connectivity can be provided through a combination of computer hardware and software applications, known as desktop-conferencing. The speed of many of these desktop-conferencing packages relies on the ability for phone companies to provide a wide base of relatively low-cost communication lines. For RBOCs, the broad acceptance of products which maximize the existing copper infrastructure (such as ISDN, HDSL and ADSL) relies on the acceptance of effective applications for telecommuting. Before any assumptions can be made about the use of telecommuting, a brief overview of the different forces shaping our evolving workforce is needed.

The most effective level of work relationships to note is **collaboration**. Collaboration is communication which results in synergy. This exists in many forms such as face-to-face meetings, phone-conversations, conferences, seminars, and even video-conferencing. The spectrum that defines this force runs from low-contact to high-contact.

Another variable to consider is **telepresence**, a force which decreases the remoteness of communication. Remoteness is not physical distance but perceived distance. Visual communications usually promotes higher telepresence than audio alone. For example, a phone conversation between two local offices may prove to be more remote that a video-conference halfway across the world. How effectively one projects their presence determines the amount of telepresence to a particular receiver. Also note that the word “presence” does not have a clear definition with regard to communication models.

Physical distance, though, does have a significant role in communication. Distance is that definite line that determines whether physical manipulation can exist in a collaborative session. That is, until computers allowed digital communication. Fax machines, electronic mail, electronic bulleting boards, computer-based networks, and other digital communication methods brought forth the idea of **tele-commuting**, the ability to work at a distance through digital communications.

It is not until very recently that the ideas of tele-commuting were applied to increasing the telepresence in collaboration, and the idea of "computer supported collaborative workspaces", or **CSCW** was developed. CSCW fosters the use of computers to promote manipulation of shared objects in collaborative sessions.

Factors contributing to what I believe to be effective CSCW environments lie in several distinct areas: social, technical, and anthropological.

1) Social: What are effective forms of collaboration in the electronic information age?

   How can telepresence be increased through communication?

2) Technical: Can we learn from the current attempts at CSCW?

3) Anthropological: Are the driving forces of CSCW development user- or technology-based?

2 Social, Technical, and Anthropological Aspects of CSCW

2.1 Collaboration in the Electronic Information Age

According to **Newton's Telecom Dictionary** [Newton, 93]: *Collaboration involves two or more people working together in real-time, or in a "store-and-forward" mode. Applications will enable a group of people to collaborate in real-time over the network using shared screens, shared whiteboards, and video conferencing....*

This has several implications. First, it means that there may be a substantial time factor involved, whether it is solving a problem or just brainstorming. Secondly, it also means that some sort of control dialog between the parties must take place -- i.e. the notion of **sharing**. Finally, it is imperative that all parties have a complete mental model of the shared workspaces. A truly complete model of collaboration incorporates a wider range of multimedia (and multimodal) devices than today's technology can provide [Cruz, Gomez and Wilner, 91]. Hence, the implementations of current collaborative environments are already compromised by both technological and social restrictions.

2.2 Increasing Telepresence in Communication

At the desktop level, telepresence in collaboration should be "projecting yourself into someone else's **workplace**". Projection may be of several forms: a projection of a face, an environment, an object of discussion, or an entire workspace. The question may not be "Is my presence there?", but "*What degree of presence do I have there?*".

Each type of projection may vary from an instance (e.g. still picture) to continuing comunication (e.g. live audio and video) depending on the task. As an example from everyday life (without technology complicating the scenario), if Beanie and Cecil are collaborating over a proposal, they may choose to meet in Beanie's office, Cecil's office, somewhere they are both familiar with, somewhere neither is familiar with, or just by phone. They may choose to bring Beanie's copy of the proposal, Cecil's copy, etc. They may choose to markup the original, or a copy of the proposal. By meeting in Beanie's office, Beanie is inherently projecting all the objects of his office onto Cecil -- obviously a strategic move on Beanie's part! Yet, if they talk over the phone (each having a separate copy of the proposal), there are several problems. Only verbal opinions are being projected, and may not be record by both sides. Also, there is the liability that the two copies of the proposal are not synchronous to begin with, making the collaborative effort that much more difficult.

In essence, the more numerous and detailed the projections, the higher the degree of telepresence. Conversely, excessive projections as well as too much detail may divert collaborative efforts. Effective telepresence in CSCW needs to fulfill three distinct roles: *content projection* (e.g. shared workspaces, applications, or whiteboards), *control projection* (e.g. verbal commands or shared input devices), and *context projection* (e.g. annotations, document cameras, or personal video).

2.2.1 Content Projection:

Content projection identifies the base starting point for a collaborative session. Educational videotapes, for example, are basic methods of projecting content to a large audience.

With any communication medium, content is the yardstick that maintains focus and synchronization of progress throughout the session. Focus prevents communication drifting. Synchronization prevents mis-communication. This is imperative for enforcing the concept of *What-You-See-Is-What-I-Think-You-See* [Smith, 92]. Because content projection is an important factor in communication, it can be stated that this synchronization is time-critical.

Currently, many computer companies are implementing standards (e.g. OLE, DDE, OpenDoc) to deal solely with effective content projection.

2.2.2 Control Projection:

Control projection defines the processes by which changes are made in a collaborative session. It may also define the roles in a turn-taking environment. In accordance with the educational example above, classrooms provide the student the ability to interrupt the flow of a lecture either verbally or with a raise of the hand.

The concept of control projection is analogous to the *coordination model of groupware* [Ellis and Wainer, 94]. However, it is expanded to define telepresence of any model, not just groupware. Control may be projected by one's verbal commands as well as explicit control over objects that is relevant to the shared workspace.

Current computer-based remote desktop applications (e.g. Carbon Copy, PC-Anywhere, KopyKat, Poly/PM, etc.) use control projection mechanisms to allow a user to control an unmanned computer. However, these applications do not fare well in a collaborative environment.

2.2.3 Context Projection:

Context projection clarifies or enhances actions defined by the control projection. It is meant to provide understanding, not content, to a collaborative session. Contextual queues help users form better content as well as more precise control. In an educational setting, a student or instructor may share their past experiences to provide insight to the current lesson being taught.

Context projection is the element of CSCW that should enhance effective communication over other collaboration methods. However, many implementations of context projection in current CSCW applications conform new technologies to existing communication models. This is hardly understood, given the wealth of information about increasing telepresence in the CHI and Human Factors research communities.

2.3 Current Implementations of CSCW

The decision of which communication model to implement needs to be reconsidered in parallel with the varying types of collaboration. In essence, the control projection must be clearly aligned with the content projection mechanism. For example, if a collaborative session uses computers to promote social interaction, the content is both visual and aural. Ideally, the control and context should also be visual/aural. The importance of video then far outweighs the necessity of a digitally-shared workspace. On the contrary, if the content being projected is an object of computer-based application, control projection should include the use of that computer application as part of the collaboration. With any method of collaboration, the paradigm of same-time different-place must be applied to the shared workspace.

Systems such as Intel's ProShare®, AT&T Vistium®, and PictureTel's PCS-50® use ISDN for transporting video talking heads and simulating application sharing. In all these examples, application sharing consists of binary data that allows a client to "see" the host application on their computer desktop. But in essence, remote users are only seeing a bitmapped representation of the host application. The host's computer generates these bitmaps and the client's computer displays this bitmap. Some packages allow the bandwidth ratio to change on demand; however, this usually results in loss of video frames, application refresh, or audio depending on the specific application priorities.

2.3.1 Advantages:

Traditional groupware consists of a video link for a talking head, and a specialized application for sharing documents or graphics. Additionally, a second video source can be used for displaying documents or objects. Newer systems break the barrier of traditional groupware by allowing multiple users to collaborate using any software application. Most implementations also provide the ability to exchange data files between users. The use of ISDN provides greater bandwidth which can be used for speedier and more complex transmissions than either POTS or direct LAN connections. Also, ISDN is widely available in most regions of the country.

2.3.2 Disadvantages:

Current implementations view the presence of video "talking heads" as important as the shared workspace from a bandwidth perspective (up to 80% can be used for video transmission). The real value of video is the detail conveyed through gestures and gaze. Smaller video images may drastically decrease the sense of telepresence [Prussog, Muhlbach and Bocker 94]. Yet these implementations present the video images in a space too small for any gesture interpretation. In addition, the desktop real-estate must still be split between the user image and shared workspace. This occlusion may prove to be more distracting than no video window at all.

Because of the limit of transmitting data over POTS lines, optimizations were made with these systems to capture the shared application window into bitmaps at specified intervals. These full window bitmaps were then compared to the last instance of the window, and a differential bitmap is generated (much smaller than a full window bitmap). This data is sent to the client workstation, where it is combined with the client's last full window bitmap, and then displayed on the client desktop. With this existing architecture, it is still the host and client computers doing the bulk of the preparation. The new bottleneck exists in the computers' abilities to process the bitmapped *picture* of the application "instantaneously". The methods that were optimized to make use of slow line speeds now become the bottleneck under ISDN.

Finally, the processing overhead associated with current implementations restricts desktops from using groupware to share resource intensive applications. How effective is the communication between a remote computer which refreshes its screen through bitmaps (approximately 0.5 fps) and a host computer running a CAD, 3D rendering, or video editing package? Would these packages even be able to run, given the resources needed to run the groupware itself?

As a simple experiment, two PS/2-based 486 machines were timed running AutoCAD 12 for Windows via the remote desktop control package, Poly/PM2. Running under a TCP/IP transport, the average rate of screen refresh was approximately once per 3 seconds (~0.30 fps) while input communication (mouse and keyboard) would lag anywhere between 1 to 10 seconds. This inconsistent delay can be attributed to the burst mode capabilities of the TCP/IP protocol and ethernet adapters. This same setup running under a 56K null-modem connection had a screen refresh rate of up to once every 5 seconds (~0.50 fps) while input communication was always maintained within a 1 second window. This can be attributed to the linear fashion that data is transmitted.

2.4 Drivers for CSCW Development, User- or Technology- Based?

What is the future of CSCW? Are the developers of CSCW applications following a "waterfall" method of development, or is it a truly iterate process? Are the collaborative needs of users driving the technology, or *is the technology driving -- and limiting -- collaborative behavior*?

CSCW, although still in its infancy, may become the standard of communication for tomorrow. How much of this growth will be dependent on *changes in human behavior* versus *changes in system interfaces*? An engineer would tend to think that human behavior must adapt, whereas an ergonomist would demand that system interfaces must adapt.

A system interface that emulates a model of face-to-face cooperation is a double-edged sword. Groupware has been designed to project as much as possible from one environment into another in the hope that the two worlds will seem as one. However, it is the mechanism of projection itself that becomes the hinderance to effective telepresence.

Even at the current point of groupware implementations, the merits of talking heads in a collaborative situation has been argued; yet it is easily demonstrated that details about a person's gestures and gazes are not fully communicated through these current implementations of talking heads [Heath and Luff, 91]. And the real estate issue not only diverts a user's attention improperly, but these small video queues may not be effective enough to grab the user's attention.

Consider also that all packages for desktop collaboration have the same basic interface. Other methods of context projection must be explored from in more realistic environment.

Current CSCW implementations require users to evolve their behavior to new methods of communication. However, CSCW application developers must realize that additional new methods of communication must evolve as users define what becomes necessary to communicate effectively. On a discussion of computer mediated communication (CMC), researchers observed that ". . . subtle changes in community norms and expectations are underway because the CMC technologies have given voice to interests not previously heard in the traditional channels of communication" [Pickering and King, 92]. RBOCs must build a unique CSCW product that creates new needs, not just satisfies existing needs, because the existing needs will change as people become familiar with the concept of CSCW.

3 Proposed Solution

An effective solution must not only provide telecommuters with a seamless sense of place but also a seamless integration of a shared desktop. The proposition that is being constructed here is two-fold: use the bulk of the bandwidth and the speed of independent hardware-based video transmission to enhance the sense of a shared workspace, and replace the talking heads model with a full-screen overlayed image of the remote person to promote the effective use of gestures [Heath and Luff 91] and possibly gaze awareness [Ishii, Kobayashi and Grudin 92].

3.1 Video Transport

In comparison to conventional solutions, this solution should be compatible with other hardware for video-conferencing. It should allow the use of H.320 and H.261 standards for transporting audio and video. However, if both ends of the connection use this hardware, custom compression should be used to achieve compression rates of at least 25 frames per second of 640x480 resolution 64-level grayscale video.

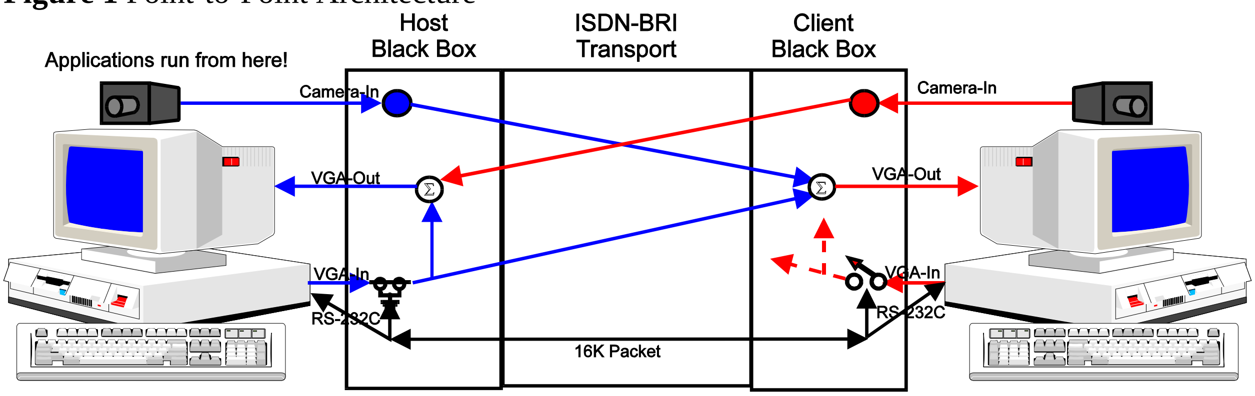
3.2 Desktop Sharing (Transport)

It is the method by which desktop sharing is accomplished which defines the unique content and control projections that make this proposed solution novel compared to other current solutions. Conventional methods use *statically defined* channels to send and receive audio, video and computer input. Most solutions use the video-conferencing standards (such as H.320 for video) to transport these channels. *However, the proposed solution requires an additional channel that may act as either a video input stream or a video output stream to transport the host desktop image to be shared.* Since this additional channel is transporting non-standard (high resolution) video, custom compression must achieve at least 5 frames per second of a maximum 1280x1024 resolution 256-level color video.

3.3 Point-to-Point Architectures

Focusing on a single point-to-point session, there are two stages of communication. The first allows social interaction by only transporting each user's video to the remote user. This model does not allow any application sharing at all. Each user would see the other superimposed upon their own desktop (so they may converse without disturbing any work a user may be doing). The second stage is known as a desktop sharing mode, whereby one user must be designated as the host -- providing the desktop to be shared -- and the other serves as a client. From that point on, each user would be looking at a single desktop with the other user's video superimposed on it. For the desktop itself, there would be no visual differences. Both host and client would be able to view and manipulate the desktop in the same manner.

Figure 1 Point-to-Point Architecture



A simple diagram of a possible architecture is shown in **Figure 1**. The simplicity of this implementation is apparent because the architecture is not computer intensive. Therefore the only platform dependent piece is the *control projection* software managing the computer input from both users.

3.3.1 Advantages:

By transporting the desktop as externally manipulated video, the computers' overhead for capturing the desktop as a bitmap, doing a binary differential comparison, transporting the differential bitmap as data, and having the receiving computer decode and display this differential bitmap is no longer incurred. This leaves more timeslices, memory, and processing power to the user's tunning applications. A screen refresh rate of 10 fps would be an *improvement of over 50 times* the refresh rate of current software-only solutions.

By allowing up to a full-screen interpolated overlay of the connected person, gesture interpretation is enhanced, gaze awareness may be communicated, and the shared workspace is no longer occluded. Research has demonstrated that this method of video overlay onto a workspace does not interfere with the perception of workspace [Ishii, Kobayashi and Grudin 92]. It is also known the larger video images may drastically improve the sense of telepresence [Prussog, Muhlbach and Bocker 94] and may also improve the value of gesture [Heath and Luff 91], which may, in turn, help users direct attention appropriately.

It is important to note that the size, position and fade level of each image (workspace and video) should be controlled by each user. This allows a user to explicitly set the level of overlay that causes the least distraction. Each user may then find their own level of comfort with this display system.

An entire host desktop is displayed at all locations without the requirement that all locations contain the applications nor the data being shared. Although this is true for most implementations, current bitmap implementations do not allow the transfer of desktop video outside of the machines internal video memory. Applications that use external video overlays (such as video editing systems) *could not be shared under the current architectures*.

In addition to offloading screen sharing from the confines of the desktop CPU, providing a small software-based input translator for each platform, collaborations can take place over heterogenous desktops. Imagine the power of being able to collaborate on a peer's workspace without having the same platform as that peer! This is possible with current implementations, but has yet to be demonstrated. Also the sociological implications of heterogenous collaboration could be explored.

3.3.2 Disadvantages:

Many of the video compression components needed to implement such a system are not available. Video mixers and digitizers are designed to handle broadcast standards such as NTSC and PAL, not custom resolutions. Also H.320 compression only specifies resolutions of 160x120 at 30 fps (or 320x240 at 15 fps), so to build in both H.320 compatibility and custom compression capabilties creates a more complex system internally and externally (with device handshaking).

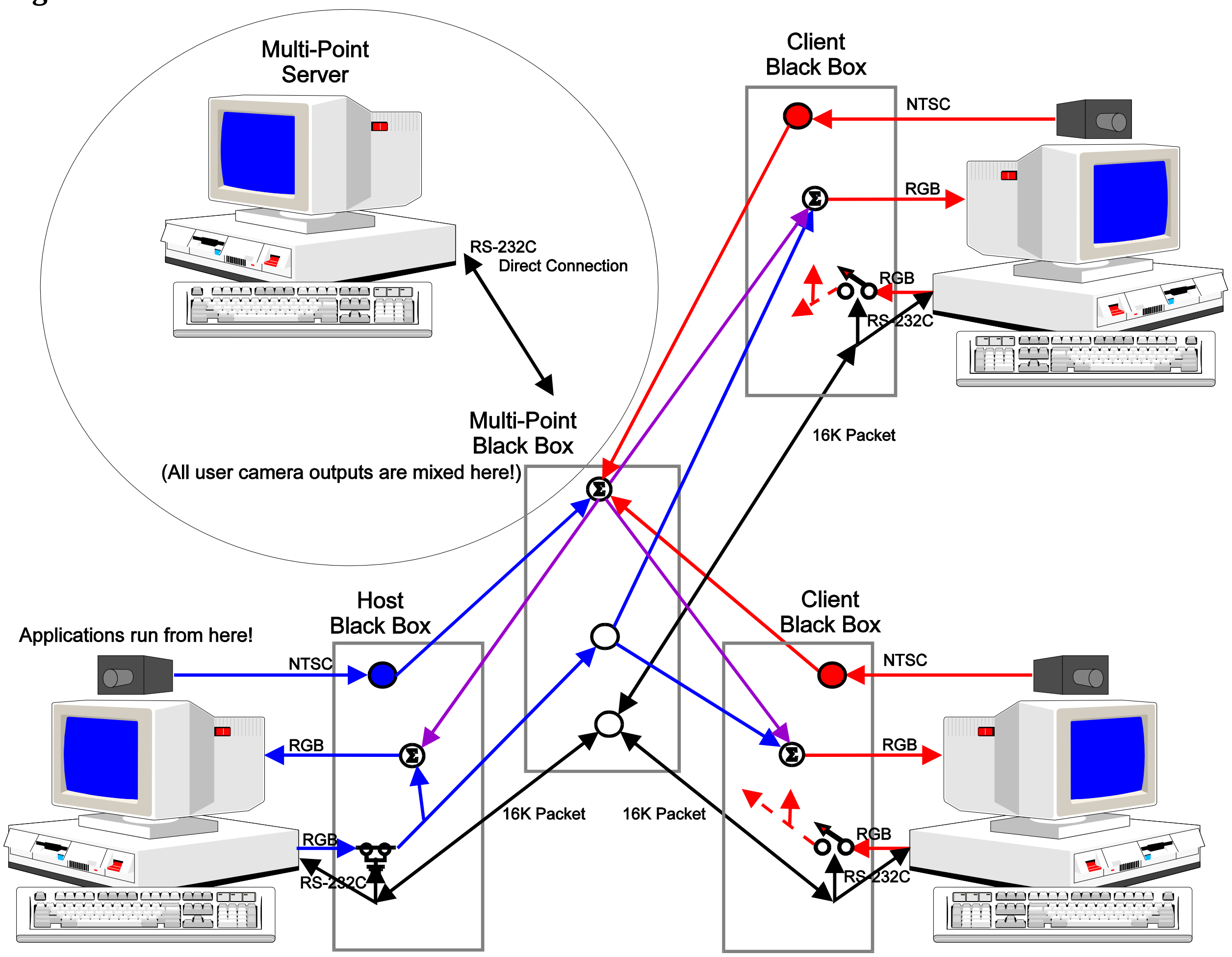
This solution does not yet consider the limitations of sharing a host desktop that has a higher resolution than the client's SVGA monitor can physically handle. From a technical level, sending a signal higher than the monitor can handle will permanently damage the monitor. The CSCW box can be made to hold different monitor IDs (and their respective capabilities) in an EEPROM, or installation software can write a specific monitor's capabilities to the EEPROM. But not all monitors send an ID, and having the EEPROM written through installation software for a single specific monitor reduces its portability. Once this technical problem has been overcome the question remains, from a user level, on how a higher resolution image can be effectively displayed on a lower resolution monitor. There could be image interpolation, image panning or image lockout.

In conferencing mode, each user is sending and receving a camera video signal. But, in desktop sharing mode, the host is sending 2 video signals (shared-desktop and host camera) and receiving 1 video signal (client camera). The client, on the other hand, is sending 1 video signal (client camera), but receiving 2 video signals (shared-desktop and host camera).

3.4 Multi-Point Architectures

Although multi-point architectures are not part of the formal proposal, it is necessary to show several possible architectures. This allows a design of a point-to-point architecture that can be expanded to include multi-point with as little change as possible.

Figure 2 Multi-Point Architecture



The architecture shown in **Figure 2** requires all users in a CSCW session to connect to a central multi-point server. As with the point-to-point architecture, any user may assume the role of application host that provides the desktop to be shared. The multi-point server receives the host desktop and transports it to all the clients. The mixing of multiple camera inputs and controlling of multiple input devices is handled by the multi-point server. The multi-point server also mixes all the incoming video signals into a single output (probably showing all the video sources in a matrix), and coordinates the traffic of remote user input to the host desktop.

4 Prototype Proposal

The two main puposes for a prototype are to demonstrate the advantages of a hardware-intensive solution as well as to explore the usefuleness of gesture (and possibly gaze awareness) with a novel method of content projection.

4.1 Technical Goals:

1) Can a desktop image be seamlessly converted to video, *digitized, encoded, transported, decoded* and displayed?

2) Can all aspects of a remote desktop be successfully manipulated?

3) Can standard user input be seamlessly transported?

    a) via Software Drivers?

    b) via Hardware Redirection?

4) Is data transfer necessary, and how can it be accomplished?

5) What design and implementation issues are there for heterogenous support?

4.2 Marketing Goals:

1) Is this method of groupware more preferable to use than traditional methods?

2) Is there a desire to have heterogenous support?

4.3 Research Goals:

1) What are the perceived differences between manipulating a remote desktop versus a local desktop?

2) How does application manipulation vary from local to remote desktops?

3) How does application manipulation vary from single to shared workspaces?

4) Using a matrix of collaborative tasks -- highly versus sparsely social, highly versus sparsely application intensive -- several psychological and social concepts can be evaluated, such as:

a) Are gestures communicated, and how do they affect commmunication [Heath and Luff 91]?

b) Increased problem solving with computer-mediated meetings [Fish, Kraut, Root, Rice 92].

c) Change in communication habits [Hulbert, Jones 92].

d) Equalization with computer-mediated decision making [Dubrovsky, Kiesler, Sethna 91].

e) How can gaze-awareness be enhanced [Ishii, Kobayashi, and Grudin 92]?

5) With which task models does the seamless desktop metaphor start to fail? This comparison uses tasks with the following parameters: highly computer processing intensive applications (i.e. graphics applications where lower frame rates are more noticeable) versus document editing, highly application interactive tasks versus occasional application interactivity, and highly social tasks versus sparsely social tasks.

6) How does the ability to provide heterogenous access affect collaboration?

5 Conclusion

Goals for developing a new CSCW strategy:

1) Develop a system that does not have any technological bottlenecks in allowing users to collaborate. Participants should not be limited to any number of applications to share. They should not be required to have identical hardware or software. They should be using someone else's workspace as if it were their own -- **with no difference in look or feel.**

2) Develop a system that implements the ideas behind the gesture enhancements described by Heath and Luff, and possibly the gaze awareness concept described by Ishii, Kobayashi and Grudin. Gesture communication and gaze awareness through overlayed facial video provide more than just detail, they provide effective context projection directly aligned with the content being projected.

Once these two goals have been accomplished, testing can begin comparing the various elements of CSCW described previously against a set of well-defined breeds of tasks. From this testing, many questions about human interaction through telepresence can be answered. At best, implementation of this strategy can be the next "killer application". At worst, it will be the start of a new iteration in the studies of Human Computer interaction.

6 References

Ellis, C. and Wainer, J. "A Conceptual Model of Groupware". *CSCW 94 Proceedings*. ACM, New York, N.Y., 1994, pp.79-88.

Newton, H. *Newton's Telecom Dictionary* (Sixth Edition). Telecom Library Inc., New York, N.Y., 1993.

Ishii, H., Kobayashi, M. and Grudin, J. "Integration of Inter-Personal Space and Shared Workspace: ClearBoard Design and Experiments". *CSCW 92 Proceedings*. ACM, New York, N.Y., 1992, pp.33-42.

Fish, R., Kraut, R., Root, R. and Rice, R. "Evaluating Video as a Technology for Informal Communication". *Bellcore Technical Memorandum TM-TSV-019988*. Bellcore, Morristown, N.J., 1992.

Reveaux, T. "Videoconferencing Moves to the Desktop". *New Media Magazine*, November 1992.

Dubrovsky, V., Kiesler, S., and Beheruz, N. S., "The Equalization Phenomenon: Status Effects in Computer-Mediated and Face-to-Face Decision Making". *Human Computer Interaction* (Volume 6), 1991, pp. 119-146.

Heath, C. and Luff, P. "Disembodied Conduct: Communication Through Video in a Multi-Media Office Environment". *Communications of the ACM*, 1991, pp. 99-103.

Smith, R. "What You See Is What I Think You See". *SIGCUE Outlook* (Volume 21 #3). ACM, 1992, pp.18-23.

Cruz, G., Gomez, L. and Wilner, W. "Tools to Support Conversational Multimedia". *GlobeCOM*, IEEE, 1991, pp. 2064-2068.

Pickering, J. and King, J. "Hardwiring Weak Ties: Individual and Institutional Issues in Computer Mediated Communication". *Communications of the ACM*. ACM, New York, N.Y. 1992.

Tucker, T. "From the Boardroom to the Desktop: Videoconferencing's Getting Personal". *TeleConnect*, September 1993, pp. 50-56.

Prussog, A., Muhlbach, L. and Bocker, M., "Telepresence in Videocommunications". *HFES 94 Proceedings*. Human Factors and Ergonomics Society, Santa Monica, CA, 1994, pp.180-184.

ITU-T Recommendation H.120 "Line Transmission of Non-Telephone Signals: Codecs for Videoconferencing Using Primary Digital Group Transmission", March 1993.

ITU-T Recommendation H.261 "Line Transmission of Non-Telephone Signals: Video Codec for Audiovisual Services at p x 64 kbits", March 1993.

ITU-T Recommendation H.320 "Line Transmission of Non-Telephone Signals: Narrow-Band Visual Telephone Systems and Terminal Equipment", March 1993.

(end of document)