# Broadband Video Conferencing and Collaborative Technologies

## 2003 Workshops Background

Peter Marshall
Director, Network Applications

CANARIE
110 O'CONNOR ST, 4TH FLOOR
OTTAWA, ON
K1P 5M9

# BROADBAND VIDEO CONFERENCING AND COLLABORATIVE TECHNOLOGIES

## 2003 WORKSHOPS BACKGROUND

### WORKSHOP GOAL

The key goal is to bring the leading researchers from both academia and industry together to work on developing a consensus vision for broadband video conferencing and collaborative technologies in Canada. Not only will specific initiatives be discussed, but also because of our limited resources, it is important that these be woven into a focused vision of how best to improve the state of the art. It must be an innovative vision that is feasible to implement over the next three years.

### EXPECTED RESULTS

The workshop will produce a vision and recommendation paper for collaborative technologies. The paper will describe the areas where the Canadian research and education community could best concentrate its relatively limited resources over the next three years to make an effective and lasting contribution to this field. It must tackle and answer the thorny question of focus: Should the emphasis be on research and highly leading edge development, advancing the current state of the art, or removing the barriers (end-station, system integration, network infrastructure) for the wide-spread deployment of current technologies?

The recommendations must also take a stab at funding: Given the constraints of the CA*net 4 program, what is an appropriate level of funding for activities in the video conferencing and collaborative technologies area? What should the funding rules be (or how can they be improved without impinging on the original CA*net 4 funding agreement with Industry Canada)? The workshops will help us determine how much funding will be made available for these and other applications projects. It could range from $8M to $20M or even higher. They will also help to determine the timeframe for projects. CA*net 4 funding finishes on March 31, 2007, which gives us an upper bound but programs within that might well end earlier.

The results of the workshop will be consolidated with the results of a number of other workshops into a larger report and set of recommendations to be presented to CANARIE's Board of Directors. Additional comment and input will be solicited from Industry Canada and other funding agencies to arrive at a final set of recommendations and priorities that will fundamentally guide the definition of the CA*net 4 applications funding program.

### THE LARGER PICTURE

The context for this work includes the special aspects of the CA*net 4 network and the many projects that include various aspects of video conferencing and collaborative technologies that have been and continue to be funded through CANARIE and other agencies around the world. . A short description of the network is provided as Appendix A below. A brief and incomplete overview of existing and emerging video conferencing and collaborative technologies is included as Appendix B below.

Many studies and many people's personal experience have shown that, in general, adding "talking-head video" to an audio-mediated conversation doesn't add a great deal of value. This is especially true when previous face-to-face encounters have allowed the participants to get to know each other. When negotiating, it can help to have both audio and video, but most people find that a video mediated negotiation just doesn't work nearly as well as a face-to-face talk.

> *At the heart of e-science lies the need to support collaborations between geographically dispersed research teams.*
>
> *from* Research Networking in Europe*:*
> *ftp://ftp.cordis.lu/pub/ist/docs/rn/rn_brochure-part1.pdf*

It is clear that collaboration in the sense described in *Research Networking in Europe* involves much more than talking heads, much more than audio and video. What is being asked for is a truly workable environment where collaboration can easily and naturally take place. If this is important in for the compact space described as Europe it is ten times as necessary in a large dispersed country like Canada. Of course there is also a clear need to collaborate with our international partners primarily in the United States, Europe and the Pacific Rim.

The challenge, then, of this program is to find the special cases where either we don't have talking heads, where talking heads really do contribute or where the video is just part of a larger and well integrated collaborative environment. Smooth integration of existing technologies will be key to moving these technologies from the category of "neat toy" and into the "essential and expected tool" category. Alternatively, a project could show by demonstration, that previous studies are flawed since they used poor quality audio and video that is isolated from other environmental cues.

As with any collaborative technology, it is key that you actually have collaborators! For the adoption of any technology, especially in this realm, it is vitally important to include a community and solve a real problem for that community. Of course there is a chicken and an egg issue here: Scott Adams, through Dilbert, summed up one of the problems back in 1992 very well. There are more communities of users that have the facilities, but there's still a long way to go.

DOGBERT, LOOK, I GOT THE FIRST VIDEO PHONE IN THE CITY!
S. Adams
NOW WE WAIT FOR SOMEBODY ELSE TO BUY A COMPATIBLE VIDEO PHONE AND CALL US.
6/1
© 1992 United Feature Syndicate, Inc.
THE AMAZING THING IS THAT SOCIETY COULDN'T ADVANCE WITHOUT PEOPLE LIKE YOU.
I THINK I SAW SOME-THING.

DILBERT reprinted by permission of United Feature Syndicate, Inc.

One interesting and common goal for projects would be to create a collaborative environment that not only worked at a distance but for many tasks, turns out to be the preferred mechanism for collaboration even when people are local to each other. A key to this is the careful and well thought-out design of the integration between data collaboration tools and the real-time audio and video components.

The program and vision must fit into and be synergistic with the rest of the CA*net4 program. Specifically consideration must be given to the prominence of the emerging web services technologies, international standards and de facto standards and the unique features of the CA*net4 network (primarily its high speed, its limited reach and the availability of dedicated lightpaths) must figure strongly in the recommendations. For example, given that we can remove bandwidth restrictions, where are the other choke points? How do you make this work and how do you find out whether you have made it work or not?

## POSSIBLE FOCUS AREAS

The following list outlines a number of possible areas that a program in video conferencing and collaborative technologies might focus. It is by no means exhaustive, and just hints at some of the things on which the community may want to focus.

- Building new, more powerful and well integrated systems to support complex collaboration at a distance:
  - Human interface improvements to the AG model
  - Improvements to the reach of multicast signals – encouraging native implementation vs tunnels and bridges
  - Implement and test technical extensions to the AG model (extensions to V2.0) like:
    - Touch sensitivity of display surfaces
    - Implementation and usability testing of immersion through high-resolution video, surround sound, VR techniques
    - Higher quality audio (noise suppression, music-quality echo cancellation) and video (higher frame-rates, more resolution, better compression algorithms) integration
    - Integration of AG into a web-services-based system
    - Use of the AG in novel collaborative situations – beyond the scientific chat room
  - Experiments in collaborations that use immersive environments that enhance audio and video transmissions with virtual reality overlays and other data intensive connections
  - Security and privacy enhancements for existing collaboration tools to allow them to be used safely for sensitive applications
- HDTV & high-resolution sound as a mechanism for creating a compelling sense of presence and immersion at a distance

- Addressing the shortcomings and barriers to use of H.323, perhaps via a SIP-based solution, perhaps via the widespread adoption of agnostic technologies like VRVS. Some of the known problems in H.323 include:
    - Poor reaction to variable latency network
    - Complex signaling
    - Limited functionality (when compared to a modern business telephone)
    - Firewall and NAT barriers to use
    - Limited or weak integration to non-audio/video collaboration tools
    - Complex user interface especially in multipoint conferences
    - Expensive and complex multipoint conferencing facilities: specialized hardware multipoint control units (MCUs)
    - Lack of community (starting to be addressed via ViDeNET)
- Implementing a sustainable infrastructure for H.323 video conferencing including T.120 collaboration servers, a robust gatekeeper, directory and MCU infrastructure that links Canadian users to the rest of the world.
- Developing a set of modalities that describe when the various video conferencing and collaborative technology suites are most suitable for specific applications. (In the end, though, it is likely the community that surrounds a particular solution that is the largest contributing factor.)

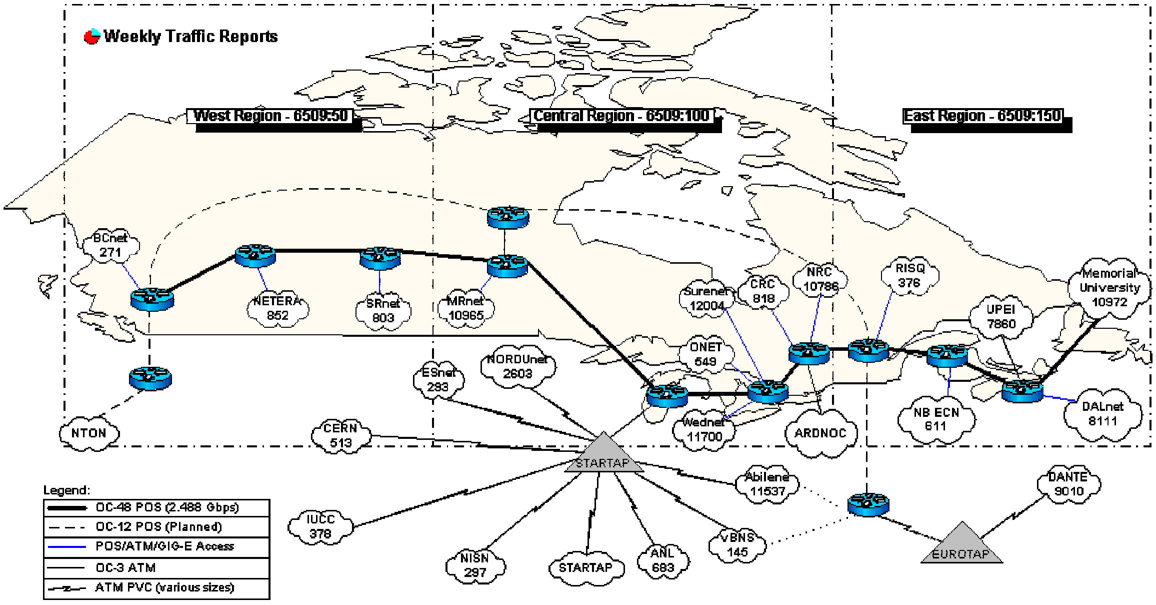
## KEY ISSUES, KEY QUESTIONS

The key is to pitch a very good case for putting lots of effort (and therefore lots of money) into this area. Of course this would be in competition with any other possible programs. One way to try to increase the appeal and so the viability of a particular approach is to link it into some of the other "hot" topics:

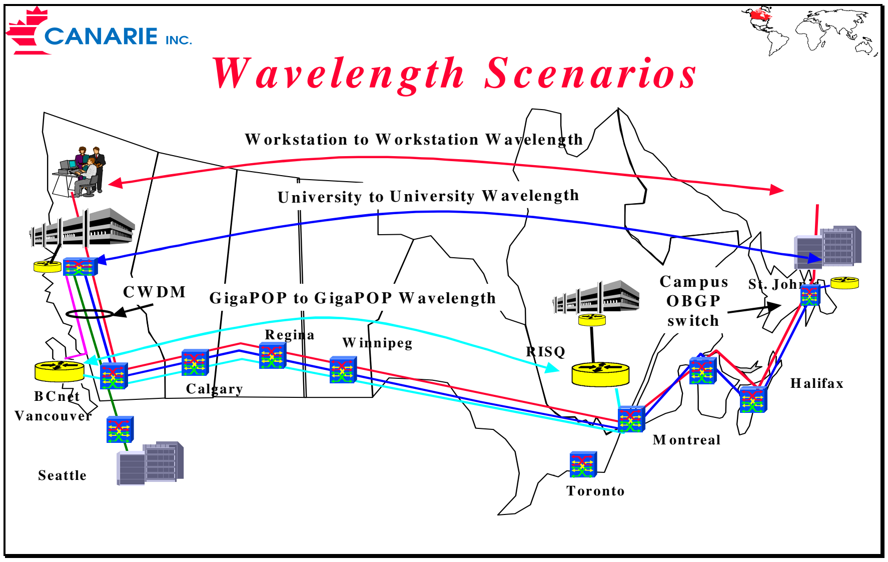
- Grid computing. (Access Grids, which, at least initially, had nothing much to do with Grids, except that they were sponsored by the people working on Grids and that there was some cross-over in personnel, seemed to benefit from the name.)
- *Lightpaths*. Do you have an application that would just work really well given a solid, nailed up connection, but would fail miserably if it didn't?
- Peer to peer systems (especially if they need a *lightpath*)
- Web services based. Make sure that any development (any directory, any modification to the AG, any system that moves from experimental to product has some concept of a web service built into it. This is clearly the methodology of the future and newly developed systems should conform.
- Synergy with CA*net4. Its reach, its speed, its *lightpath* capability.

## APPENDIX A:
## CA*NET 4, THE ESSENTIAL FACTS

CA*net 4 is Canada's 4th generation high-speed research and education data network that spans Canada from St. John's to Victoria with a touch-down (or two) in every province. It also maintains high-speed IP connectivity with similar networks in the US and around the world. Regional R&E networks in each province connect to CA*net 4 and in turn universities, colleges, research institutions and schools connect to the regional networks.



CA*net 4 can be distinguished from other high-speed networks by deploying *lightpaths* to provide private, point-to-point connections for special applications in addition to shared high-speed internet protocols-based connectivity.



Because *lightpaths* are not shared like a typical IP connection you can be guaranteed (to a much greater extent) of the characteristics of that connection. *Lightpaths* make it much simpler to predict latency and to reduce it by eliminating network hops between the end points. This is happy news for near-real-time applications like audio and video transmissions. It does come at the cost of universality. It just isn't feasible with the current architecture to have a point-to-point connection (like a telephone connection) to huge numbers of end-points. Also, due to the overhead of setting up such connections they're likely to be at a minimum 100Mb/s each.

CA*net 4 is funded through a March 2002 grant of $110 to CANARIE Inc. The grant terms specified that a nation-spanning, internationally connected network be designed, built and operated and that an applications funding program be established with a portion of the funds. The series of workshops that this paper is a part of is an early step toward establishing a funding program to satisfy this portion of the CA*net 4 grant.

## APPENDIX B:
## OVERVIEW OF BROADBAND VC AND COLLABORATIVE TECHNOLOGIES

### THE HERE AND NOW

This group includes those facilities and services that are up and running now and that have a reasonable installed base of users. Typically they're not fully functional in some dimension and further development and integration with other technologies will be necessary. Also there may be barriers to very large-scale adoption of these technologies that could be effectively addressed by some central network infrastructure.

**H.323**

Chief among the currently deployed technologies are the widely deployed H.323-based systems. The H.323 standard was developed from the ISDN-based (phone-line) H.320 protocol and it retains these telecom roots. The complexity of the translation to Internet Protocols can make it difficult to get sessions working reliably through simple firewall systems. End stations provide the usual user

interface to an H.323-based system. End-stations include both stand-alone systems from Polycom and Tanberg (that connect to a TV monitor or projection system) as well as PC-based (desktop and laptop) systems from Polycom, VCon and others. New forms are also appearing usually with the aim to make the systems easier to use. Typical of these systems is the telephone-like systems like the one pictured here from Aethra or D-Link's recently released DVC1000 a $US299 system that turns a home TV into a video conferencing station (just add Ethernet). Systems like these are in use daily at universities and businesses around the world. At CANARIE, H.323-enabled meetings have gone from being a novelty to being an almost ordinary daily occurrence in the past year.

In a typical H.323 network there are also gatekeeper computers that provide both access control and addressing functionality and multipoint control units (MCUs) that enable conference calls. The emerging ViDeNet standards for linking distributed gatekeepers into an addressing hierarchy now known as the Global Dialing Scheme (GDS) provides a good model for the infrastructure that would make it easier to deploy H.323 systems in Canada.

Most H.323 end-stations rely on a version of Microsoft's Netmeeting to implement T.120-based applications sharing. Without a T.120 hub, Netmeeting is only capable of point-to-point applications sharing. This poses a couple of problems:

1) Microsoft hasn't supported Netmeeting for a number of years. It is moving toward a SIP-based system and they've let Netmeeting languish.

2) T.120 is a general applications sharing protocol that relies on sending updates to screen images to the remote site. This is suitable for many applications like PowerPoint or MS-Word, but not the best approach for more graphically intensive applications (CAD/CAM, video-streaming).

H.323 systems are widely deployed at universities around the world. The community is large and it is growing quickly. The Internet2 currently supports H.323 through their Information Commons, a ViDeNet compatible service which consists of a large gatekeeper, scheduler and MCU farm run by the Ohio regional educational network (OARnet). CANARIE maintains a small MCU and gatekeeper that is linked into the ViDeNet network. Other MCUs on CA*net 4 are deployed at various universities across the country.

Even though open-source implementations of H.323 are appearing, it is typically a commercial product that transmits medium quality audio and video using modest network resources (typically between 400Kb/s and 2Mb/s). These systems are at the deployment, barrier removal and integration with collaboration tools stage of development. The chief technical barrier is probably the protocol's sensitivity to packet loss and jitter of the type typically found in a shared and partially loaded IP network.

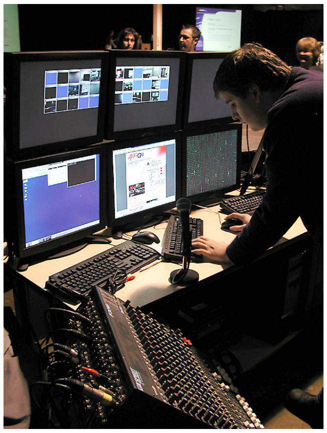
### MPEG2

In addition to H.323 systems, there are a number of existing, commercially provided video conferencing systems in use. Typically they are proprietary in some way and they have a limited number of users. MPEG2-based systems from Amnis, Star Valley (Litton) would fall into this category. Typically they provide better audio (stereo) and video than H.323 implementations and so they have niche markets in the remote music teaching area for example. The high-cost of the hardware and incompatibilities and non-integration of other collaborative tools has limited deployment. Typically such systems use between 3 to 8Mb/s.

## THE NEAR FUTURE

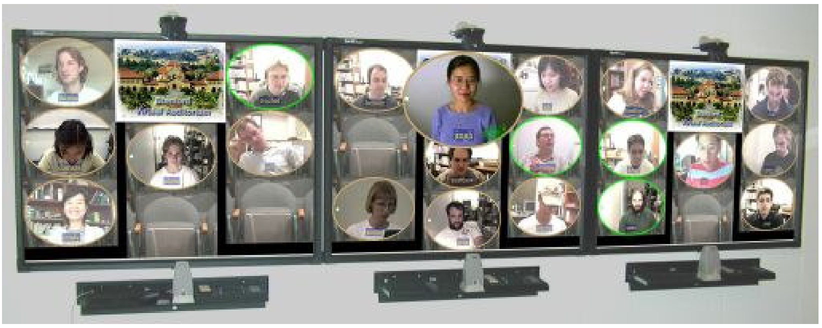
Typically these are emerging technologies or enhancements to the technologies described in the "Here and Now" group. They strive to improve on the existing systems along many dimensions: quality of transmission, richness of the environment and ease of use to name a few. Projects here may seem to be just too expensive or difficult for "real" deployment, but each can demonstrate a future (or a special circumstance) where those barriers no longer apply.

### Access Grid and Multi-window Systems



AG v2.0 is being released in beta this spring. Version 2.0 takes the multi-screen, multi-window approach and adds extra elements suggested by the experience and the lessons learned in the initial implementation of the Access Grid to build a more useful collaborative environment. There are still lots of usability issues to be addressed as well as the integration of innovative technologies to be done. One suggestion is that we take a web-services-based model as the basis for these modifications that could turn out to be AG v3.0 over the next few years.

The *Virtual Lecture Hall,* being developed at Stanford University holds out some promise for handling large numbers of connections. By using commodity cameras and microphones, pseudo-MPEG4 software that comes with the Microsoft Windows operating system, a single graduate student has built a rather scalable system that is very easy to install (java-based). The proprietary underpinnings are just an accident of the implementation and could be easily substituted if a similarly widely implemented suite of tools was available. Like other niche products both commercial and more research-based systems like *Isabel*, currently the community of use is rather limited.



**SIP-based**

There are a number of tools and services that are taking the Session Initiation Protocol as their base rather than using the H.323 suite. Microsoft, for one seems to have moved away from their attempt to implement a software-based H.323 system (NetMeeting) and is now concentrating its efforts in the SIP realm. There's a lot of work in the IP-telephony arena that should easily extend to video. It is important to keep the barriers between the audio-only and video enhanced conference to a minimum. Vendors of H.323 MCUs are now including SIP support in their products. There's some hope that as the SIP and related conferencing tools have grown out of the IP world and the IETF that they will suit the IP environment better than H.323 which has its roots in telephony and ISDN connections. This remains to be seen.

**VRVS**

Virtual Rooms Video System is not strictly a conferencing tool, but a web-based meeting structure that facilitates setting up and conducting meetings over the IP network. It is very flexible and supports clients as diverse as an access grid node, commercial H.323 end-point and MBone-tool based clients (that work in a non-multicast environment through a worldwide network of VRVS reflectors). It is a practical and pragmatic tool for scientists around the world. It is included in this section only because it hasn't escaped that limited scientific world.

**ISABEL**

The Isabel system, from at the Universidad Politécnica de Madrid has been in continual development since its first public use in 1993. It is a software-only system (like the AG and MBone tools) that runs on dedicated Linux (SuSE 8.1) systems and uses motion JPEG (or MJPEG) for video compression. By using MJPEG, a relatively simple frame-by-frame compression (similar to DV), both the complexity of the encoding/decoding and the latency can be minimized at the expense of bandwidth. Typical sessions use from 200Kb/s to 2Mb/s of bandwidth. Both multicast and relay-based (flow-server) distribution over non-multicast-enabled networks is supported.

IPv6 networks are supported by Isabel which makes it attractive for sites that might want to use either a native or tunneled IPv6 network as a mechanism to get around the restrictions of a NAT-based IPv4 infrastructure. (For example, if every end station has a unique IP address it is much easier to initiate a connection.

The Isabel application is more flexible than many other video-conferencing systems and it easily adapts from conference distribution to small, distributed group collaborations. The tools for collaboration beyond audio and video, tend to be fairly primitive (like the distribution of .GIF versions of MS-PowerPoint slides). Isabel has been extensively used for classroom and music teaching and collaboration in the LearnCanada, MusicGrid and VirtualClassroom projects in Canada. Until recently Isabel was freely distributed for research use but now there's a per-end-point licensing fee for the Isabel system being charged.

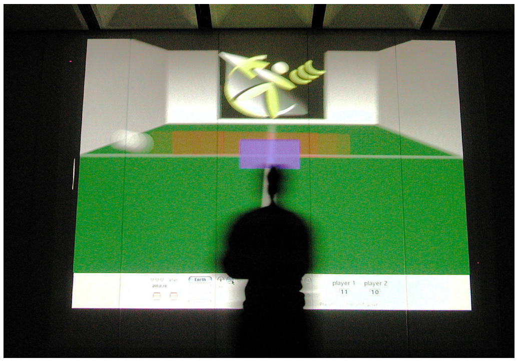
**CAVE™-like Immersive Environments**

CAVE™ and Immersa-desk are emblematic of a class of immersive environments that insert a person into a 3D simulated space. They can be used for collaboration of small numbers of people in remote locations but most work seems to have been with two plus observers. In these virtual worlds avatars typically represent remote participants and each collaborator is wired into a head position, body position and perhaps hand and finger motion systems. There can be a lot of "suiting up" involved. The data requirements can vary enormously, but since they are synthetic 3d worlds rather than real-life video these can sometimes be pre-loaded so that only the coordinates of the participants being exchanged in real-time.

### THE LEADING EDGE

These technologies and methods are very experimental and no one is really sure how or if they'll really work. Or if they do "work" that they're useful for collaborative work or communication. I'll just list a few of these with some short descriptions.

#### Enhanced Reality Systems



APR's *Moveable Feast* project is an example of trying to link two physically large and physically separated spaces and the people within those spaces via multimedia, IP network connections. There are other similar projects that usually come out of the dance or other arts communities. The keys are to link spaces and people using more than just video and sound. And then of course to explore what that means for how those people collaborate. How is this different from a CAVE™ environment? In a cave the primary concept is to move the body into a virtual space. With these systems the idea is to map a virtual space onto a physical one – "overlaying and enhancing physical objects and people present with a hyper-real DATA-SPACE created via projected video, robotic lighting, and multi-speaker sound"[1]. The bandwidth potential of such frameworks is limited only by the components included, but current (experimental) implementations use between 10 and 20Mb/s. Keeping latency low, especially when these environments are used for music and dance, is a key issue.

#### Tele-Haptics

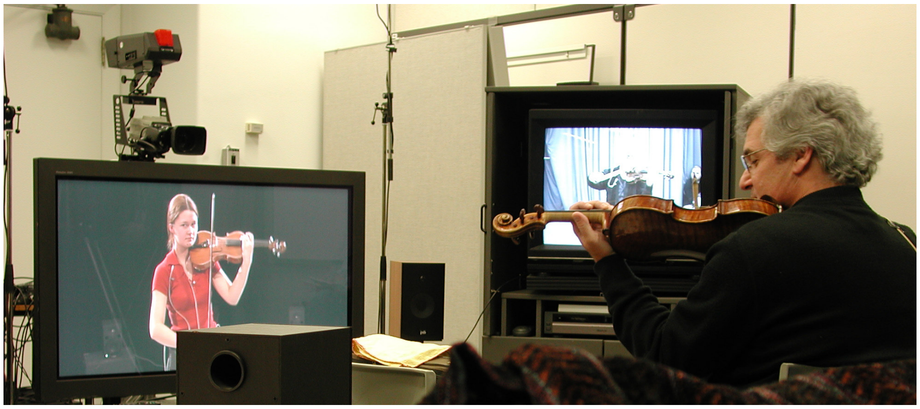
A promising technology for collaboration is represented by the new and relatively inexpensive force-feed-back systems that can be adapted to provide a touch-based sensory experience at a distance. These devices present a new dimension to a collaborative space. For effective collaboration these technologies need to be integrated into a more comprehensive collaboration environment. It is still early days.

For haptics, the keys are latency and reliable delivery rather than huge volumes of data. It is crucial for safety that position information and force information to arrive very quickly after a motion is made and that errors are not made that might move the tool in a dangerous way. To provide a smooth haptic experience most systems generate about 1000 updates

[1] From the Moveable Feast final report CANARIE ANAST project 2003: Also as part of the introduction on the project web page: http://www.positioning-research.com/feast/

per second (as compared with the typical 30 for video). Thus a system that included haptics might be a good candidate for a light-path experiment.

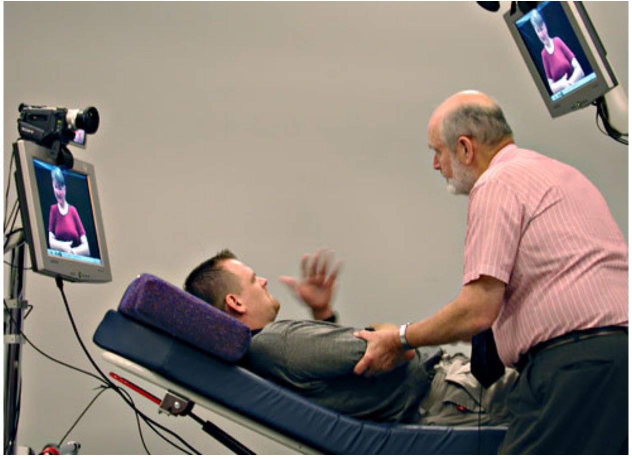
**HDTV and uncompressed SDTV transmissions including High Resolution Sound**



McGill and others have built and experimented with systems for using the high-speed research and educational networks to transmit high-speed and time dependent data like uncompressed SDTV and synchronized high-resolution (DVD-audio quality) surround sound over long distances. They've experimented with musical collaboration and teaching with these tools. The application tries to create the sense of presence through really good sound and a really good image. The obvious next step in these experiments is to use HDTV or at least moving images that have a higher resolution than standard definition TV to further enhance the experience. Key to any such work is to find the appropriate applications where the expense (by some measure) of these interactions makes it reasonable to use. But if we're looking to the future, we might be able to ignore some of these using the assumption that they'll get ever cheaper. Then the emphasis of the research moves to making these systems easier to use and more useful for a wider variety of applications. SDTV takes about 270Mb/s. HDTV at high quality (editable) would take 1.2 to 5Gb/s. Usable MPEG2 distributions of HDTV could be as low as 15Mb/s.

**DV over IP implementations**

Over the past few years, researchers in Japan, at Berkeley and at McGill have experimented with using cheap consumer DV-format camcorders as the codec for high-quality audio and stereo video transmission. The DV format, in theory, should have lower latency than more aggressive compression systems like MPEG2 or MPEG4 since DV has no inter-frame compression (and so each frame can be encoded, sent and decoded individually). In fact many consumer-grade hardware implementations (i.e. cameras) inexplicably introduce significant delay. Careful testing is required. A typical DV stream uses about 30Mb/s.



In the McGill/UNB remote video interpreting project partially funded by the CANARIE ANAST program, the same base-software as used for uncompressed standard definition video is used to quickly transport two streams of DV video from doctor and patient to a remotely located sign-language interpreter. DV allows a very crisp image, a very light and portable camera, software-based decoding and relatively low latency for good interaction. Could this system be effectively extended into other application realms?