D. James Gemmell and
C. Gordon Bell

Noncollaborative Telepresentations Come of Age

New technology now allows us to attend business conferences without physically attending the conferences. Is this cost-cutting, time-saving practice the next killer app?

At this moment, thousands of people are criss-crossing the globe to attend business presentations. Most of them will spend more time traveling to the presentation than they would in the presentation itself. Also at this moment, thousands of people are not at presentations they need or want to attend due to scheduling conflicts, crises, or budget constraints. Similarly, presenters are busy re-presenting the same material to people who could not attend their last presentation. Now technology exists for saving time and money on presentation attendance to allow a wider audience to benefit from presentations they would otherwise miss. Scheduling, transportation delays, and costs can be eliminated through *telepresentations*—a presentation in which the presenter and/or some of the audience members are not physically present but are telepresent, that is, in a different location and/or at a different time. Consider the last presentation with an audience of 300–2,000 that you attended and ask: "Wouldn't you rather have attended via your desktop at some more convenient time and with the ability to fast-forward through it and occasionally even replay it?"

Before explaining what this article is about, it is important to be clear on what it is *not* about. This article is not about meeting or collaboration settings where significant interaction between all group members is expected. Nor is it about the social and economic changes due to the introduction of the technology—changes that may be more important than the technology. The larger spectrum of teleconferencing [6, 9], the in-depth social aspects of teleconferencing [1, 4], and especially the practice of telemeetings are outside the scope of this article. Furthermore, we do not intend to disparage the side effects of technical conferences or call for their elimination.

This article examines how telepresentations can be used in courses, conferences, lectures, product introductions, and general informational meetings; that is, any presentation not requiring substantial group interaction. It is about how the Internet, intranets, and plain old telephone service (POTS) can reduce the need for special videoconference facilities. The equipment costs to achieve this are negligible and already exist in

Table 1. Bandwidth and storage for various media

Format	Typical rate	MB/hr	Transmit over	Media to store
Text (script @ 120 words/min, 6 char/word)	96bps	0.04	POTS	1 floppy
Slides (AKA presentation graphics or overheads)	[a]	1	POTS	1 floppy
Voice (compressed)	8Kbps[b]	3.5	POTS	3 floppies
Video (H.263)	< 28.8 Kbps	12.7	POTS	10 floppies Zip MiniDisc
Video (H.261)	128Kbps[c]	56	ISDN	Zip MiniDisc
High-quality audio (uncompressed)[d]	1411Kbps	620	TI/LAN	CD
Video (MPEG-1)	1.5Mbps	675	TI/LAN	CD
Video (MPEG-2)	4Mbps	1800	T3/LAN	DVD

[a] Overhead data can be extremely bursty (without elaborate prefetching). Therefore, a rate figure may be misleading. 64Kbps bursts are adequate for slides consisting of 2D graphics and text. With prefetching, rates as low as 10Kbps are adequate.

[b] There are voice compression schemes with rates of 4Kbps and lower, but of the schemes we have actually listened to, 8Kbps represents the lowest rate that still maintains good quality voice.

[c] H.261 supports multiples of 64Kbps. We use 128 here for comparison. There is not expected to be much quality difference between 28.8Kbps H.253 and 64Kbps H.261.

[d] MPEG-1 at 256Kbps is virtually indistinguishable from uncompressed CD-quality audio (16-bit stereo @ 44.1KHz). However, high-quality compressed audio hasn't caught on in a big way yet, so we list uncompressed values.

most organizations. Nearly all PCs have adequate to excellent sound capabilities; and a reasonable microphone, camera, and video-capture facility can be added for a few hundred dollars. From Table 1, note that a one-hour presentation with slides, voice, and an H.261 talking-head video could be stored easily on a CD that costs only a couple of dollars to produce. A one-hour presentation with slides, voice, and H.263 video could be transmitted over ISDN, or over POTS if the video is omitted.

Using technology to assist the delivery and recording of presentations for later delivery is not new. Telepresentation technology has been gradually evolving, from 35-mm slides or filmstrips advanced according to cues in a tape-recorded talk, to television and VCRs. Television has been used for presentations in a number of distance education settings (e.g., the Stanford Instructional Television Network, or National Technological University). However, video has a number of drawbacks as a presentation medium, and specialized videoconferencing solutions tend to be expensive. In this decade, the advent of MBONE multicasting tools [5] using the Internet introduced the ability to deliver presentations consisting of audio, video, and slides to large audiences at their desktops. However, much like the Internet of a few years ago, the MBONE has been primarily the domain of researchers using high-end workstations, and nothing like a tool for the masses. Concurrent with the development of MBONE technology, desktop preparation of presentations has progressed to the point where nearly all presenters prepare their own presentations in a desktop publishing fashion.

Table 2. Remote presenter to a group in a room or to desktops

	Audience in One Room		Presented to Desktops	
	Present	Telepresent	Present	Telepresent
Presenter–Audience	–	✓	–	✓
Audience–Audience	✓	–	–	✓

What is new is that telepresentations have reached a level of convenience and economics that will cause an explosive growth in use for everything from taking minutes of meetings to formal presentations of all kinds. The telepresentation paradigm we envision is the ubiq-

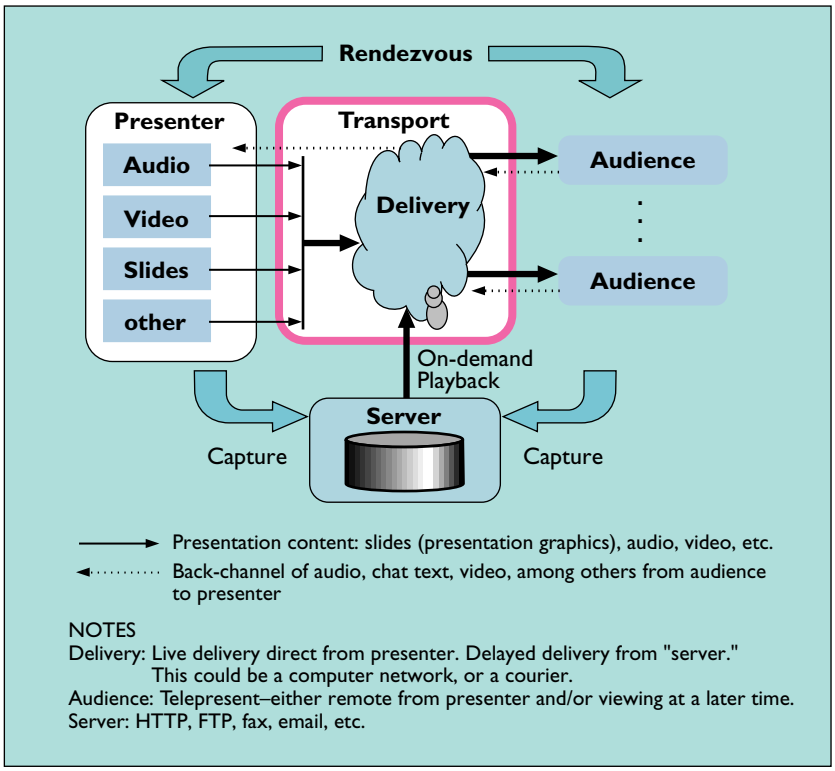

Figure 1. Structure of a telepresentation

uitous broadcasting and recording of all types of formal presentations so they can be viewed in real time or on-demand. We believe telepresentations are the next "killer app" because of the ease and low cost to produce, broadcast, capture, and deliver high-content presentations that can be viewed anywhere, anytime.

Figure 1 shows the general structure of a telepresentation including a server for delivering a presentation at another time. For live telepresentations, it is important to consider which group interactions are telepresent, and which are physically present. Is the presenter telepresent to the audience? Is the audience physically present with each other? For example, the presenter may be remote, (the presenter-audience relationship is telepresent), while the audience-audience relationship is present. Or, a presenter may remotely give a presentation to a group, each of which are viewing the presentation on their own desktop. In this case, both present-audience and audience-audience is telepresent (See Table 2). Finally, there may be a mix. A presenter may remotely give a presentation simultaneously to a group in a single room and to people at their desks. Each case has different social implications and technological requirements.

Goals for Telepresentation Technology

THERE ARE MANY POSSIbilities as to what makes up a telepresentation. Some are more critical than others. First, let's go over what the ideal telepresentation would include, and then consider how critical each element is, and how difficult each is to achieve.

The ideal telepresentation includes the following communication channels, in order of content importance:

- *Slides:* Displays with text, graphics, images, including animation and special effects. (We use the word "slides" to mean not just overheads, but also the richer presentation graphics.)
- *Audio:* The speaker's voice ("the talk") and audio such as sound tracks and sound effects
- *Video:* The "talking head" of the speaker and video clips (snapshots)
- *Back-channels:* These provide the speaker with feedback about the audience status together with chat, questions, comments, and a way of recording action items, among others things
- *Slide comments and "the talk" script:* When a presentation is viewed "on-demand" from a server, having the talk in script form allows the viewer to read versus hear the talk

The slides and video are two complementary visual portions of the presentation. Several ways of dealing with slides and video should be possible. They could be displayed in individual windows the user can size and place according to their own desire. Ideally, this would include the ability to use two monitors. In the case of a single monitor, the presenter should be able to size and place video and slides relative to each other. For example, at some points, the video window may be entirely hidden to give the slide the full display. At another time, the video window may be full screen, with the slides window occupying the top-right corner, like is commonly seen in a

news broadcast. This visual mixing capability is an important feature in creating compelling and visually appealing presentations. Furthermore, having several cameras always live is desirable. Ideally, a conference would have a couple of cameras pointed at the speaker and several aimed at the crowd. The visual mixing should allow selection of the desired video feed and special effects-like fades. However, each individual raw feed could be made available for viewing (both live and from a server).

In the ideal case, all the raw and mixed media would be transmitted and recorded simultaneously. The recorded material could be browsed in a number of ways. Time-based browsing would support VCR or CD-player type controls, such as rewind, play, fast-forward, or jump to time offset. Fast-forward could be used in such a way that it is still slow enough for the presentation to be intelligible: In this way, a one-hour presentation could be viewed in, say, 30–45 minutes. Pitch shifting on the audio track is required to eliminate the "chipmunk" effect. Logical browsing would allow the viewer to go to a particular slide or to search for a keyword. Once at a particular slide, the viewer could switch to time-based and start playing the presentation.

The back-channel applies to live transmission.[1] It allows the audience to communicate with the presenter. The back-channel is important for fielding questions, but also has implications for speakers who feed off the audience emotions. Some speakers may feel hesitant telling jokes if they cannot hear the laughter. Visual cues may also be important for the speaker in deciding the pace of a presentation. A number of back-channels should be made available:

- *Video* lets the presenter see the audience to get a feel for it and understand its attentiveness
- *Audio* for questions and comments
- *Text chat* lets all participants type and see what each other is typing. This is invaluable, especially for debugging the session itself (e.g., if the audio is experiencing difficulties)
- *Whiteboard* takes chat a little further and allows participants to do arbitrary drawings. It is useful to be able to superimpose the drawings over the slides to facilitate discussion of items on the slides

[1]Annotations added later to the presentation could be considered a back-channel of sorts, but it is really the live feedback we are concerned with here.

Necessary Features

Now that we have established what we want for telepresentations, let's be pragmatic. For the next 3–5 years, there are both social and technical hindrances to achieving the ideal. Additionally, not all the features of our ideal telepresentations are equally important. Slides and high-quality audio to carry the talk constitute the essentials for typical presentations. Other elements are not as critical.

Consider the back-channel. From a social point of view the back-channel may not be very important, especially for very large meetings. Visual cues coming back from the audience can be misleading. Many people who are concentrating intently on technical material will look away, doodle, or even close their eyes. Furthermore, for every speaker who enjoys having a crowd to play off, there are speakers who are nervous or distracted by the people. Such speakers may enjoy the lack of a back-channel and find their presentations actually improve in that context. Technically, it is very difficult to support back-channels for large audiences (that is, they "don't scale easily"). With large audiences, floor control becomes extremely difficult. Based on these social factors, and the technical difficulties involved in back-channels, we consider back-channels important for small room audiences of less than 50, but not essential for very large presentations.

Table 3. Important elements of a telepresentation

Critical media:
Audio (the talk)
Slides (presentation graphics with animation)
Important for adding feeling of presence:
Pointing / scribbling / animation on slides
Some talking head (quality and frame rate not critical)

For recorded presentations, there is no live back-channel, although it could clearly be possible to edit or to add annotations to the presentation, and to initiate new communication (email, voice mail, or perhaps a meeting). One might be tempted to conclude that if there is no back-channel, live transmission could be omitted and only on-demand delivery provided. However, there are two reasons to continue live presentations, even without live back-channels: *urgency and timeliness* (such as recent results, or news), and *cost* because it

is cheaper to multicast/broadcast a presentation rather than supporting unicast connections for each viewer.

Video is desirable but not critical. Most of the content is in the slides and audio (talk). Video is a very poor way to transmit presentation graphics with text and figures, as it fixes them into a bitmap (which is usually small due to bandwidth constraints). It often employs lossy compression that reduces legibility. Typically, the video is the speaker's talking head. This headshot is desirable because it adds an important connection with the speaker. However, unless the presenter is lecturing on dance steps and demonstrating them, it is unlikely the videoshot will convey any information at all about the subject of the presentation. Furthermore, video is either bandwidth-expensive, or else is of marginal quality.[2]

We would assert that the only critical media in a presentation are the presentation graphics and audio. In order to add life to the presentation, and a feeling of presence with the presenter, it is highly desirable to add pointing, scribbling, and animation, and to display some kind of talking head no matter how low the resolution or frame rate. Occasional still shots improve the feel of a presentation by an order of magnitude over one without any headshots at all. (Table 3 summarizes the important elements of a telepresentation.)

Practical Issues

HAVING ESTABLISHED WHAT IS DESIRABLE and what is necessary in telepresentations, we now must examine how feasible it is to achieve these goals. A number of practical issues, both technical and social, must be faced to produce a telepresentation.

Slides can be sent or stored in any of a large number of formats, ranging from faxes to online computer presentation graphics including documents of various sorts. HTML pages are attractive due to their ubiquity. However, current HTML pages are *documents*.[3] An HTML page may take several pages to display unless they are well hyperlinked into real pages. For a presentation, the presenter wants to show a certain image to the audience without worrying about the point to which they all must scroll.[4] A page-based display is what most of us understand to be a presentation.

A simple paged-based display can be achieved using bitmaps. However, the problem with bitmaps is that they fix the display size, and scaling them to other sizes sacrifices quality. For this reason, many CD-ROM and game titles work with the lowest common denominator of 640x480, and cannot use the full screen of a higher-resolution display. Furthermore, if the content is primarily text, lines, and polygons, then storing it as such will result in a far more compact representation than even the best bitmap compression.

One form of bitmap compression is to transmit the bitmaps as compressed digital video. Clearly, the poorest way to transmit slides is to point a video camera at them and send video. Videotaping of technical presentations from an overhead projector has established the notion that video presentations are very difficult to give and record. Resolution ranges from annoying at best to unreadable at worst. If a bitmap slide can be fed directly to the video compression input without using a camera (e.g., Precept's SlideCast), then the situation improves somewhat. Many video compression schemes will show an initially blurry slide that becomes progressively clearer. However, at best this gets one back to the same point as if image compression had been used. The only reason it is an alternative is that video is available and compressed bitmaps are not.

Page-based formats supporting text, 2D graphics and embedded bitmaps are well suited to presentations. Postscript and various graphics standards, like X Windows or Windows metafiles (WMF) would fit the bill. Adding support for animation and special effects, as is done in Microsoft PowerPoint or Lotus Freelance, takes the concept of a slide even further and makes for extremely compelling presentations (as one would hope from a product designed for presentations). Supporting animation and special effects requires efficient support from the underlying format. It doesn't take many bits to define a polygon and say "fly from (x1,y1) to (x2,y2)." It can take a lot of bits to retransmit the polygon for every point along its path. Similarly, defining an effect that fades from one slide to another changes what could be a very

[2]H.263 allows video over POTS modems. The frame rate and image quality are enough to create the feeling of connection with the party on the video, but are insufficient for much else.

[3]As of HTML 3.2, 2D layout is not part of the standard. 2D layout is on the agenda to be added soon. (See section entitled "The Future")

[4]You could, of course, synchronize scrolling, but this would involve having each party with the same sized display region—which gets us right back to page-based display.

expensive operation into a very cheap one. Therefore, many whiteboard-type programs (e.g., wb of the MBONE) or application-sharing programs (e.g. Microsoft NetMeeting) are adequate for static slides but would require major changes to efficiently support animation and special effects.

In some scenarios, any sort of live transmission may be impractical. In this case, slides can be present in their entirety to each audience member.

Table 4. Transmitting, capturing, and encoding audio and video

Media	Issue
Audio	Microphone selection
	Microphone placement
	Sound card selection
	Input level (solved with an external audio compressor)
	Compression selection and paraqmeter setting
Video	Camera selection
	Capture card selection
	Lighting (more important than the camera)
	Compression selection and parameter setting

During the presentation, the presenter would instruct the audience members when to advance to the next slide. Slides in this instance can be in any format including fax or paper copies. While certainly not ideal in terms of timeliness and production cost, this scheme works in many cases.

The first item to be faced in producing a presentation is to create media worth transmitting and/or recording. While there are many software packages for producing great slides with relative ease, capturing high-quality audio and video can be frustrating and time-consuming. This aspect is often last in the thoughts of the presenter, but turns out to be the most important.

Audio is a critical channel for a presentation. Hours can be spent on tweaking software settings only to find out the microphone level is set too high. Selecting the right microphone, matching its impedance with the sound card, placing it effectively, and getting the input level right is nontrivial. Indeed, it is an art for serious audio engineers. Likewise, lighting is an art for cinematographers, and is critical to producing video. The addition of a couple of well-placed lights will make the perceived quality of the video increase dramatically.

Table 4 enumerates some of the issues in capturing and encoding audio and video. Correctly matching equipment (for example, finding a microphone with the correct impedance for your sound card) can be a real problem, so the trend to bundling all the necessary equipment together should only increase. We believe that some form of automatic gain control on the microphone is almost a necessity.[5] Similarly, when one gets frustrated enough with computer telephony, using the plain old telephone for voice is a fine solution and very hard to beat! If audio and video capture is to become as simple as using the telephone, then package solutions that include all the A/V capture equipment, complete with audio gain control and video lighting will eventually need to be on the market.

One aspect of video capture that can be annoying is the lack of eye contact from the speaker to the viewer. This happens particularly when the presentation is coming from a desktop, with the camera mounted above or beside the monitor. The presenter will commonly be looking at the slides and/or the back-channel on the monitor, giving the impression to the viewer that the presenter's gaze is fixed away from them. This lack of eye contact immediately makes the presentation less engaging—all public speakers are taught the importance of eye contact. To solve this, we may see dual cameras using a stereo view to construct a view as if from the center of the monitor. Alternately, replacing the monitor with a translucent or partially reflective projection screen allows some options in restoring eye contact.

A step beyond the capture is parameter settings for compression. Many products currently on the market present the user with a large, bewildering array of possibilities. Other products give almost no options. This is an area where software will mature, and do more for the user to make compression setup easier and more powerful.

While capture and compression can be painful, we do not want to give the impression it is impos-

[5]With microphone level (i.e., the power of the signal being sampled by the A/D converter) too low, signal-to-noise gets out of hand (besides the fact that on the receiving end they may not be able to amplify enough to hear the signal). With the microphone level too high, distortion will result (which is often mistaken for software trouble). We have found that using an analog limiter between the microphone and the A/D converter can make a tremendous difference. The use of such limiters is common at radio stations where they, too, want to keep the signal as "hot" as possible without clipping.

sible. Large conferences will probably have skilled A/V staff who can handle these problems. With some effort, do-it-yourselfers can master the production of telepresentations. However, it is not reasonable to expect the speaker to also be the engineer until hardware and software improves. We expect the ease of use to rapidly improve over the next few years as telepresentations become more pervasive.

Network Latency, Bandwidth, and Scalability

HAVING SUCCESSFULLY CAPTURED AUDIO AND video, and having produced the desired slides, all the media will need to be transmitted. This raises the issues of network latency and bandwidth. Like most Cyberspace apps, the cost and availability of bandwidth is a key limit to the growth of telepresentations. For presentations with no back-channel, latency is not a critical issue. Using dedicated phone lines or ATM allows bandwidth to be reserved. On the other hand, Internet transmission, which can be much cheaper, does not allow for bandwidth reservation. The RSVP proposal [2] would add bandwidth reservation to the Internet, but it will be some time before RSVP could become ubiquitous; indeed, its adoption is not yet assured.

For networks like the Internet, which cannot guarantee bandwidth, care must be taken to deliver audio. Software should make *intelligent* responses to lack of bandwidth. For example, it is preferable to lose video before losing audio. Port number conventions can be used to have routers drop video packets before audio packets. Sending software should be able to recognize the situation and adapt—the most simple case being to reduce video to make way for audio, and in the extreme, getting into layered (or hierarchical) transmission schemes [7]. Even with the best of schemes, the Internet cannot be relied on for continuous audio or video transmission of any given quality. Therefore, for some telepresentations, a telephone conference call may remain the audio network of choice for the near future.

Scalability—the ability to handle very large audiences—is a serious problem for telepresentations. The back-channel is difficult to accommodate with a large audience, because of floor control (such as who gets to speak) issues. Note that this is not strictly a technical problem. It is difficult to accommodate audience feedback in a large stadium just as much as it is over the network. If everyone in a stadium speaks at once you have unintelligible crowd noise. If everyone on a network sends at once you have congestion problems. Besides back-channels, *side-channels* may also be desirable, where a small number of participants set up a private session which is often short lived. This is analogous to attending a large conference and chatting with a few people about a particular item of interest. A major problem for side-channels is identifying the interested parties and bringing them together

Standards relevant to telepresentations

Standard	Description
H.263	Video compression for POTS rates (<28.8Kbps). Encoding somewhat computationally expensive – requires about a 90MHz Pentium to encode >1fps in real time. Decoding can be done on a 486.
H.261	Video compression for ISDN rates (n x 64Kbps). Less computationally expensive than H.263.
G.711, G.722, G.723.1, G.728, G.729	Audio compression. Targeted for bitrates of <64Kbps, 48/56/64Kbps, 5.3/6.4Kbps, 16Kbps, and 8/13Kbps, respectively. Can be encoded in real time.
T.120	Data communications, including file transfers, image data, and image annotations.
MPEG-1, MPEG-2	Video compression. Encoding is computationally expensive, so to date these standards have been used for on-demand video, not live conferencing. MPEG-1 gives TV-like quality at 1.5Mbps. MPEG-2 gives very high-quality (HDTV) at 4Mbps (6Mbps sometimes used for high motion footage, like sports). The video standard also includes audio. MPEG-1 audio at 256Kbps is comparable with CD-quality, at 64Kbps it is comparable with AM broadcast.
H.323	Overall standard for videoconferencing over ethernet or token ring. Uses H.261, H.263, G.711, G.722, G.728, G.729, G.723, T.120.
H.320, H.321, H.324	Similar to H.323, but intended for use on N-ISDN, B-ISDN, and POTS/Mobile Radio, respectively.

Table 5. Product characteristics for live delivery of telepresentations

	Net Meeting	NetShow Live	PowerPoint	IP/TV	CuSeeMe	MBONE tools	VXtreme
Slides			✓	Screen grab		✓	
Audio/voice	✓	✓	(capture)	✓	✓	✓	✓
Video	H.263	✓		✓	✓	H.261	✓
Whiteboard	✓		over draw		✓	wb	
Chat	✓				✓		✓
App-share	✓						
Live capture			✓			RTP record for audio and video[a]	✓
Communication channels	IP, POTS	IP	IP	IP	IP	IP	IP, POTS
Rendezvous	✓	✓	manual	✓	✓	sdr	✓
1 – 1	✓		✓		✓	✓	✓
n – n (n <7)	✓		✓		✓	✓	
1 – N		✓		✓		✓	✓

(many conferences hold "birds of a feather" sessions for such a purpose). Work on back-channels and side-channels for large presentations is still in the research phase.

Even without a back-channel, sending a presentation to a large audience is still difficult. Attempting to unicast to all from the presenter (or a server) can easily exceed the capacity of the sending unit, or the network bandwidth available. IP multicast is a scalable method for transmitting data [3]. Schemes can be adopted using *reflectors, repeaters* or a tree structure of participants passing on the data. However, these schemes require significant setup effort and/or lack stability in the case of many members rapidly joining and dropping out.

Social Issues

We have already mentioned the social impact of a back-channel for the speaker to view and/or hear the audience's reaction. Furthermore, some speakers may feel awkward addressing a camera, regardless of back-channel capabilities. Whether the speaker is alone in the room as they speak can be significant.

When a back-channel exists, there is the question of how remote viewers feel about using it. In this case, it is more important to consider the presence/telepresence between audience members than between an audience member and the speaker. People typically desire the ability to glean information about the rest of the audience before taking the floor. They are after feedback that their timing is appropriate and that their remarks or questions are being well received. If the speaker is physically present with some number of the audience, remote audience members can feel quite intimidated about participating. Experience with lectures from the University of California at Berkeley shows this participation almost never happens [8]. On the other hand, our own experience with remote speakers presenting to a room full of people is that the audience in this setting is just as confident as if the speaker had been present. Of course, the culture of the participants is an important factor in the social impact of telepresence [4].

Where We Are in 1997

At the present time, scalable solutions for large, live audiences are still difficult to obtain. IP multicast is far from ubiquitous, and getting on to the MBONE [5] can be a lot of work in setting up tunnel software. Similarly, substantial work is required to set up reflector-type solutions (like CuSeeMe uses). The H.32x and T.120 stan-

dards don't scale well, and, of course, neither do telephone conference calls. Essentially, the difficulty in achieving scalable telepresentations is lack of infrastructure, although the situation is changing. For the near term, live telepresentations are likely to be limited to 100s. Nevertheless, given that each receiver could be a large display viewed by a group, this can be less limiting than it sounds. What is currently difficult is giving presentations to a large number of desktops (and including back-channels makes it even more difficult).

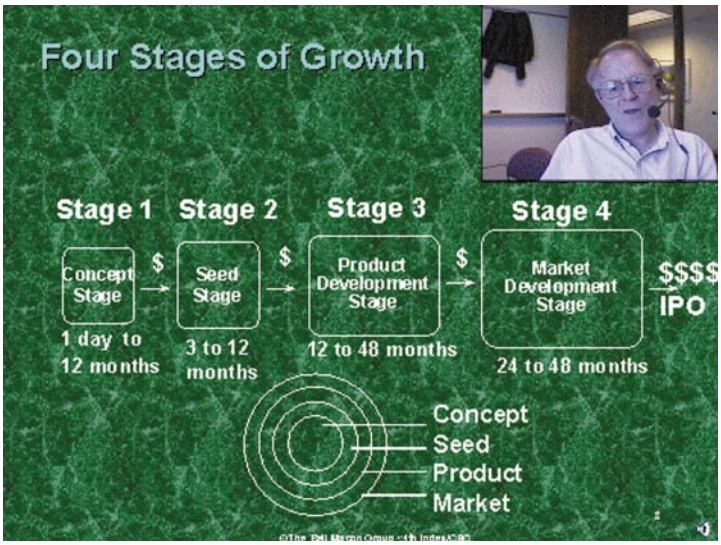

Figure 2. Gordon Bell's telepresentation

It is possible to have successful presentations today by simply faxing or mailing slides to each member. Online, one could use an application sharing as in NetMeeting, display grabbing as in Precept's SlideCast, postscript slide transmission as is the MBONE wb tool, or presentation graphic transmission (including animation and effects) as in PowerPoint conferencing. Note that with application sharing or display grabbing, any application can be used to create the slides. Without the help of this software, the presenter could also simply instruct participants to aim their browsers at certain Web pages or some other online documents.

For audio, the telephone still remains attractive in quite a few scenarios. H.323 audio is available and will become more and more common. Also, proprietary audio schemes like the numerous Web phones (and the audio included in video products) work reasonably well over the Internet. All of these solutions are inexpensive. POTS is very hard to beat.

There are quite a number of relatively inexpensive live digital video solutions available, including CuSeeMe, Intel ProShare, Intel Video Phone, and VDOPhone. A camera and capture card can be obtained for $500 or less, and the software is also reasonable. Furthermore, these products work over ISDN or POTS, so the connection is not too expensive. Stored video can be distributed via storage media like CD-ROM, or online using http, or specialized streaming servers like VDOLive or VXtreme Web Theatre. Early software codecs for H.263 video (designed for POTS) required a CPU like a Pentium in the 125–166Mhz class. Emerging codecs can encode using a 90Mhz Pentium and decode on a 486. As this software becomes available and faster machines become mainstream, we predict a significant increase in the ubiquity of video.

Product announcements, and the experiences we and others have had, show that telepresentations are becoming more and more common and practical. Consider the following true stories:

Table 6. Product characteristics for on-demand delivery of telepresentations

	Net-Show on-demand	StarWorks	PowerPoint	VXtreme
Slides			✓	✓
Audio	✓	✓	✓	✓
Video	bit map	✓		✓
HTML	URL links		output	✓
Delivery	streaming, file	streaming	html (www), file	streaming

- In March 1996, Yale University's School of Public health announced "the first regularly scheduled global Internet channel for scientific research." The audience first downloads the audio and accompanying slide presentations, and an hour later participants meet in a global chat room.
- Silicon Graphics records product introductions including technical training courses using their Showcase presentation tool for either live or delayed viewing. In this way, both conferences and training are completely decentralized to the desktop.

- In 1996, Microsoft, VXtreme, Progressive Networks, and NetScape each announced and/or shipped products that combine audio with synchronized Web pages and/or bitmap images, delivered from streaming servers. "Shows" can be viewed at rates as low as 28.8Kbps.

Our own experience with telepresentations has been positive; indeed, they prompted this article. We have experience doing telepresentations in several ways. The MBONE is one that most academicians use, but few corporations support due to lack of experience with multicast and the past lack of tools for PCs.[6] Jim Gemmell spoke from California to a conference in France using the MBONE tools over IP multicast. The MBONE tools include audio, video, and a whiteboard tool that can transmit Postscript slides (the audio and video can be stored and later replayed). A number of other speakers were remote. In addition to the attendees at the French site there were telepresent audience members viewing via MBONE software. Many presentations are regularly given on the MBONE, for example, the Berkeley Multimedia and Graphics Seminar.[7]

Gordon Bell has given voice/overhead presentations that Morgan Stanley sponsors for the last five years. Atttendees are prefaxed overheads and a telephone conference call carries the audio. This includes a controlled audio back-channel for questions. Audio tapes are available for a time-shifted audience.

Bell gave a telepresentation from our Bay Area Research Center (BARC) in San Francisco lab to fellow Microsoft employees in Redmond, Washington. CuSeeMe was used for two-way audio and video and PowerPoint's Presentation Conference tool was used for the presentation graphics. The audio, video, and graphics were transmitted over Microsoft's corporate network (aka intranet). Bell's talking-head video, was placed in the upper right corner, in an area reserved on each slide. He viewed the remote audience in a window on his display. While the transmission was happening, PowerPoint 97 was capturing the audio for a stored version. Since the stored version does not support video, selected frames from the video feed were put into the presentation to give a "feeling of presence" when viewing at a later time. Figure 2 shows a slide from the stored presentation. The live presentation was declared a success by all involved—especially Bell, who did not spend the eight hours traveling to give a one-hour talk. No drawbacks were experienced from having a remote presenter, and Bell found giving the telepresentation still gave audience feedback, yet allowed him to better focus on the talk. This experience was very encouraging because no special infrastructure was required to produce both a live and stored telepresentation. The presentation can either be ftp'd or directly viewed from a server (http://www.research.microsoft.com/gbell).

Tables 5 and 6 show the characteristics of various telepresentation technologies that can be used for live and on-demand. Note that in our own telepresentation both Powerpoint and CuSeeMe were required to handle the presentation, recording, audio (talk and back-channel for questions), and video (talking-head and audience back-channel).

The Future

Over the next year or two we expect to see multicast infrastructure widely deployed within organizations, to the extent that IP multicast [3] is assumed to be supported in any IP network. This will allow massively scalable presentations to take place to millions of participants.

Animations and 2D layout are proposed to go into future versions of HTML,[8] which will allow anyone with a browser to view high-quality telepresentations. For live presentations, standards for driving browsers (going beyond simply pointing them at a Web page, by also indicating when an animation should occur) will also need to be written.

Another important development for telepresentations is the need for *layered* media encoding, a method whereby the media is encoded into a number of separate layers (streams) that can be combined to provide a quality level proportional to the number of layers used. For example, the first layer may be a very low-bandwidth encoding designed for POTS. The first and second layers might use a combined bandwidth of an ISDN line, and when combined will yield higher quality (Note: the second layer is unusable without the first). Further

[6] Readers interested in Windows MBONE tools, see http://www.research.microsoft.com/BARC/mbone

[7] See http://bmrc.berkeley.edu/298

[8] 2D layout is close to becoming a defacto standard, and animation via scripting is already being done.

layers could exist for even more quality at the expense of bandwidth. Layered encoding is important because it allows clients with a variety of bandwidth requirements to obtain their material from a single source. Layered encoding is just making the transition from the research lab to the commercial product, and should be expected to play an important role in the future.

Finally, the growth in the use of telepresentations will lead to more mature software and capture hardware. This hardware and software will enable the production of basic telepresentations without turning tens of knobs and reading complex instructions about the software options. Streaming servers for later viewing of presentations will become more and more common. Browsing of stored telepresentations will become a powerful experience. Time-based browsing with VCR-like controls will be enhanced with logical browsing and playback at increased speeds (with pitch-corrected audio).

We propose that all technical conferences both transmit and capture their presentations to allow telepresent speakers and attendees. We are not proposing the elimination of real presence at technical conferences! Real presence at technical conferences is important, beneficial, and will continue. But the telepresent option can be added to technical conferences at low cost. Adding the telepresent option to presentations allows the presentations to reach a much wider audience as University Video Communications [10] has demonstrated by videotaping various ACM and IEEE conference talks and tutorials. Currently, the nearest thing to a telepresent experience of a technical conference presentation has been to read the associated paper. However, the formal paper is not the presentation. Very often, new results have been obtained since the writing of the paper, and are discussed in the presentation only. In many settings, the paper consists only of crude notes. We would argue that the presentation has its own value, distinct from the paper, and distinct from the side effects of actual attendance (such as offline conversations). And a stored presentation offers entirely new value by allowing browsing and perhaps even condensed viewing.

ACM97 was a good example of a telepresentation conference. Held in San Jose last month, Precept multicast the conference live on the Internet MBONE. Microsoft and VXtreme captured the conference for on-demand and CD publication. (To view on-demand versions of ACM97, see http://www.research.microsoft.com/barc/telepresence/)

Conclusion

Today, telepresentations are practical and low-cost. They offer enormous advantages by making presentations that are delivered via slides and talking heads widely available at low cost by allowing presenters and viewers to be telepresent. Over time, we expect presentations ranging from courses to conferences to all be easily viewed via the Web so that one can "be there, while being some place else at same time." We do not care to speculate about the social consequences of the availability of this information. Similarly, we do not believe the telepresentations apply to all situations using graphical presentations and a talking head as in the need for extensive collaboration. However, we do believe that telepresentations are interesting enough to put them forward as a candidate for a future killer app. C

References

1. Ackerman, M.S., Starr, B. Social activity indicators for groupware. *Computer 29*, 6 37–42.
2. Braden, R. (Ed), Zhang, L., Berson, S., Herzog, S., Jamin, S. Resource ReSerVation Protocol RSVP). Version 1 Functional Specification. Internet draft. Internet Engineering Task Force, Nov. 5, 1996.
3. Deering, S. Host Extensions for IP Multicasting, RFC 1112, May 1988.
4. Dustdar, Schamram, and Hofstede, G.J. Videoconferencing across cultures—A conceptual framework for floor control issues.
5. Erikson, H. MBONE: The Multicast Backbone. *Commun. ACM 37*, 8 (1994) 54–60.
6. Kouzes, R.T., Myers, J.D. and Wolf, W.A. Collaboratories: Doing science on the Internet. *Computer* (Aug. 1996), 40–46.
7. McCanne, S., Jacobson, V., and Vetterli, M. Receiver-driven layered multicast. In *Proceedings for ACM SIGCOMM '96*. (Aug. 1996, Stanford, Calif.), 117–130.
8. Rowe, L. Private correspondence.
9. Schooler, E.M. Conferencing and collaborative computing. *Multimedia Syst. 4*, (1996), 210–225.
10. University Video Communications Distinguished Lecture Series http://www.uvc.com

D. James Gemmell (jgemmell@microsoft.com) is a researcher at Microsoft's Bay Area Research Center in San Francisco, Calif.

C. Gordon Bell (gbell@microsoft.com) is senior researcher at Microsoft's Bay Area Research Center in San Francisco, Calif.