A Principled Design for Scalable Internet Visual Communications with Rich Media, Interactivity, and Structured Archives

Ron Baecker
Knowledge Media Design Institute
and Computer Science Department
University of Toronto

Abstract

In contrast to video conferencing, webcasting supports scaleable Internet visual communications, yet it is typically viewed as an ephemeral one-way broadcast medium. We present a principled design for interactive webcasts that are accessible both in real-time and retrospectively. We derive system architecture and functionality from project goals, results from the video communications literature, and observations of prototype implementations in real webcasts. The ePresence system is scalable, interactive, and able to support presenters and engage remote audiences with rich media. It also provides automatically derived, structured, navigable, and searchable archives for the retrospective use of webcasts.

1 Introduction

1.1 The Problem

Some computer-supported cooperative work technologies (Baecker, 1993) support real-time collaboration; others are designed for asynchronous use. Because distributed co-workers are rarely available concurrently, asynchronous tools such as email, mailing lists, threaded discussions, and organizational memories are the collaboration technologies currently in greatest use.

Yet asynchronous tools rarely succeed in establishing the sense of immediacy, interactivity, and shared purpose that results from face-to-face meetings. Our goal is to build a scalable Internet technology infrastructure that enables effective remote attendance at events, both concurrent and retrospective, with maximum engagement, interactivity, and support for community.

Audio/web conferencing and multipoint videoconferencing are two methods typically used for real-time communication, collaboration, and knowledge sharing over the Internet.

Audio/web conferencing (see, e.g., www.webex.com) allows the real-time multipoint transmission of voice and slides. Yet it lacks the media richness, sense of presence, and ability to engage participants that is afforded by video and other dynamic media.

Internet desktop video conferencing (see, e.g., www.microsoft.com/windows/netmeeting/) supports real-time multipoint audio and video communications as well as shared workpaces. Yet it still does not provide reliable Internet video performance, and is not scalable to large numbers of participants.

Our approach is based on a third kind of technology, webcasting (Wainhouse Research, 2002). *Webcasting* is the Internet broadcasting of streaming media so that it can be viewed via a Web browser on a personal computer. Webcasting is scalable to large numbers of participants, but is typically a one-way broadcast medium that is not interactive.

Another problem is that today's streaming media platforms do significant buffering in order to provide smooth media delivery to viewers despite the vagaries of Internet data transmission. Thus there typically are delays of 10 to 30 seconds between when events happen and when they are viewed. This is another challenge to interactivity.

1.2 A Possible Solution

To facilitate scaleable communications and knowledge sharing at a distance, we have initiated work to make Internet visual communications:

- scalable
- engaging, delivering rich media
- interactive
- accessible in real-time and via archives
- useful for knowledge building and sharing.

We have developed a viable and innovative webcasting infrastructure called ePresence (Baecker, 2002; Baecker, et al., 2003). This currently includes support for video, audio, slide, and screen broadcasting; slide review; moderated chat; private messages; the submission of questions; and the automated creation of structured, navigable, searchable event archives.

1.3 Applications and Significance

Sample applications include the use of Internet broadband transmission for distance learning (e.g., continuing medical education), presentations by global corporations (e.g., shareholder and analyst meetings), and briefings for the public (e.g., delivering health and safety information).

These applications are vital in a post-Sept. 11 and post-SARS world. For example, Wainhouse Research reported in September 2002 the results of a survey with over 700 respondents. More than 40% of people in the U.S. workforce were taking fewer trips; more than 70% were interested in alternatives to travel. In 2001, even before Sept. 11, both Jupiter Media Metrix and the Yankee Group forecast over U.S.$3B of annual business spending on streaming media by 2005.

2 Review of Relevant Past Work

A useful review of research on video-mediated communication and desktop videoconferencing is Finn, et al. (1997).

Webcasting is increasingly used for knowledge dissemination by universities and corporations. Stanford has been delivering video distance education for over 25 years, and began work with Internet distribution in the mid-90s (Cordero, et al., 1996). The Berkeley Multimedia Research Center's Internet Broadcasting System (BIBS) webcasts over 20 classes each semester; the university has recently adopted its technology as an integral part of the university's course delivery infrastructure (Rowe, et al., 2001). USC's School of Engineering's Distance Education Network now webcasts over 150 courses per year (http://den.usc.edu/). Medical faculties also are increasingly webcasting "Grand Rounds" for continuing medical education (see, e.g., Hsiung, 2000, http://psychiatry.uchicago.edu/grounds/).

With its pioneering Forum system (Isaacs, et al., 1994, 1995) in the early 90s, Sun Microsystems showed that providing seminars to a distributed audience via streaming media over a corporate intranet increased attendance by more than a factor of two. Audience members were enthusiastic because it gave them more flexibility in attending talks without leaving their desks and while doing other work. Lecturers preferred a face-to-face setting, primarily because they missed the face-to-face interaction with a live audience. This was true despite a number of innovative interactive features including audience questions submitted by voice or text and audience polling.

The Knowledge Media Institute at the Open University initiated their KMi Stadium research project with the goal of staging large-scale live events and on-demand replays over the Internet (Scott and Eisenstadt, 1998). A variety of technologies have been used over the years to support real-time Internet delivery of audio, graphics, slides, audience text questions, and sometimes video. Recent applications have included a virtual degree ceremony (Scott and Mason, 2001) and webcast communications to health care personnel (Scott and Quick, 2002).

Work at Cornell (Mukhopadhyay and Smith, 1999) focuses on the automatic generation of lecture archives consisting of synchronized and edited audio, video, images, and text. Their Lecture Browser is now under further development by Berkeley. The Authoring on the Fly System at the University of Freiburg (Hurst, et al., 2001) explores both the synchronous real-time transmission of such material and the creation of archives of recorded presentations.

Georgia Tech's eClass (nee Classroom 2000) is a ubiquitous computing (Abowd, 1999) high-technology classroom that combines source material, annotations, Web snapshots, and hand-drawn notes into a digital library of captured educational experiences (Abowd, et al., 2000). Projects at Xerox PARC (Moran, et al., 1997) and Fuji Xerox Palo Alto Lab (Chiu, et al., 2000) focus on meeting capture in a conference room.

The CMU digital video library Informedia project (Wactlar, et al., 1999) makes integrated use of techniques from speech, image, and video processing and from information retrieval. Techniques applied include shot detection, key frame selection, face and colour detection, and video OCR. To help users when large numbers of sequences are retrieved, headlines, thumbnails, filmstrips, and video skims are displayed. Video digests are also produced, including diagrams emphasizing word relationships, timelines showing trends over time, and maps showing geographic correlations (Christel, 1999).

There is overlap in goals between our work and several research projects at Microsoft Research. The Flatland system applies webcasting in distance education (White, et al., 2000), and incorporates mechanisms allowing questions from the audience and discussions among audience members. It received enthusiastic support from two lecturers, but serious unhappiness from a third. Their TELEP system (Jancke, et al., 2000) allows lectures to be webcast, provides novel methods for local attendees to be aware of remote attendees, and allows remote attendees to interact with others and the speaker.

Bargeron, et al. (2001) present technology to allow asynchronous collaboration around video archives. They show that students will annotate and ask questions if instructors will answer the questions. Cadiz, et al. (2000) focus on the advantages of allowing remote attendees to watch and discuss recorded lecture videos together, and the use of various communication modalities to facilitate discussions even if the remote attendees are not in the same room. Tiernan and Grudin (2001) compare individuals working solo, pairs working face-to-face, and pairs working asynchronously who communicate via annotations and instant messaging.

These projects are inspired by a classic study (Gibbons, et al., 1977) and a more recent one (Smith, Sipusic, and Pannoni, 1999) that demonstrate remarkable benefits from physical or virtual collaborative viewing of videotaped or videoconferencing instruction.

ePresence will be compared to some of these significant research projects in Table 2, which appears in Section 7 of this paper.

3 Design Requirements

ePresence design requirements result from project goals and observations of our prototype implementations in real webcasts, grounded in results from the video communications literature. We organize the requirements into 5 categories: **P**(articipants), **M**(edia), **I**(nteractivity), **A**(rchives), and **S**(ystem), and use them to derive implications for system architecture, functionality, and user interface.

P1: Design with the needs of various classes of participants in mind.
Participants (stakeholders) involved in a webcast include the speaker, possibly a moderator, members of a local audience, and members of a remote audience. Remote viewers may be viewing in real-time or retrospectively. Isaacs, et al. (1994, 1995) presented early evidence that different stakeholders could have dramatically different reactions to webcasts.

The speaker: A prime design requirement is to avoid new restrictions, obligations, or stresses on speakers.

The moderator: To protect the speaker, we assign new tasks and responsibilities for moderators.

Local attendees: Although events could be webcast to a remote audience only, they typically also involve local attendees. These individuals expect their experience not to be unduly degraded by the webcast.

Remote attendees (real-time): Remote attendees have the most to gain from the webcasts.

Remote attendees (retrospective): Because most participants in an event are not available at the correct time, the number of retrospective attendees is potentially greater than the number of concurrent remote attendees.

P2: Support scalability.
Our goal is to provide video-mediated communication with significant scalability, i.e., the ability to reach hundreds or thousands of viewers. We use webcasting because it is a scalable technology and video conferencing is not.

P3: Support a variety of devices, operating systems, media platforms, and bandwidths.
To allow the potential of scalable transmission to be realized, we need to support a wide spectrum of devices — desktop, laptop, and mobile; operating systems — Windows, Mac, and Linux; media platforms, e.g., Real, Windows, and QuickTime; and bandwidths — high-speed and modem. Appropriate device support includes simple procedures for users to test the suitability of their platforms for viewing events.

P4: Support both local and remote audiences.
We seek to support a wide variety of events with as large a potential attendance as possible. This implies that we need to include remote audiences. We also include local audiences because their presence, reactions, and body language provide useful feedback to speakers (Mane, 1997).

P5: If local and remote audience needs conflict, inconvenience local attendees slightly but not significantly to support remote attendees.
Providing good audio and video to both local and remote audiences can be difficult. For example, providing high-quality remote audio necessitates questions being asked using a microphone. This is a slight inconvenience, and is imposed on the local audience. On the other hand, lighting to achieve good video renditions of a speaker can impair the slide contrast in a poorly designed lecture room, so video production values must sometimes be sacrificed.

P6: Design the room to support the needs of the speaker, the audience, and the webcast.
Our experience, and that of the group at Berkeley (Rowe, et al., 2001) emphasizes the following:

- Having adequate space and movement for speaker, audience, production crew, and equipment for A/V capture, switching, mixing, encoding, and streaming
- Providing microphones and amplification to provide quality audio to both local and remote audiences
- Supporting lighting that allows for a bright display of projected slides and a visually attractive rendering of the face and movements of speakers and panelists.

P7: Do not make slide display dependent upon receiving a digital version in advance, nor upon adding software to the speaker's laptops.
No matter how much we plead, many speakers will not send us their slides ahead of time. Some work on their slides until the hour before their talk; some are leery about accepting strange software on their machines. The implications for system organization is that we must be able to intercept an analog data stream on its way to the lecture hall's data projector and scan convert that into digital slides for transmission to viewers.

P8: As speakers have difficulty attending to both local and remote audiences, plan for a significant role for a moderator.
Microsoft experience's with the TELEP system (Jancke, et al., 2000) is that speakers were only slightly aware of their remote audience, despite large-screen projection of still and video images of remote attendees.

We also have found that speakers have difficulty in observing audience details such as the location of portable microphones. A lecture hall is a busy environment; speakers need to concentrate on their material. We therefore currently use the moderator as an interface between the remote audience and the speaker. As we shall see below, the moderator can also mediate communications among members of the remote audience.

M9: Ensure quality sound even at the expense of sacrificing quality video.
Video conferencing literature stresses the importance of avoiding delayed or degraded audio even at the expense of reducing video quality (Finn, et al., 1997). We too have observed in our webcasts the lack of resilience exhibited by remote viewers if sound quality is not first rate. High quality audio has therefore been a key goal, implying the need for audio mixing and amplification controlled independently for local and remote audiences.

M10: Do not force speakers to use Powerpoint slides as their only audiovisual aids.
Many commercial and research systems only support Powerpoint presentations. We also support Web tours, opportunistic Web surfing, and screen captures of software demos via "remote desktops".[1]

M11: Emphasize delivery of quality slides and screen capture more than video.
Although this depends upon the application, a speaker's slides or screen captures is usually the most important visual component of a technical talk. The largest area of screen real estate should therefore be devoted to slides or screen capture, and video relegated to a lesser role.

M12: Enhance the sense of presence using high-quality cinematography.
Despite the greater importance of audio quality and slide display for communicating information, many studies confirm the importance of the video channel for motivating, facilitating, and enhancing collaboration (Finn, et al., 1997). Our design therefore includes multiple cameras, video switching and special effects generation, and careful cinematography to engage remote attendees and enhance the sense of presence.

[1] This currently works on Windows clients only.

I13: Support interactivity.
Although we do not have symmetrical video as does video conferencing, and our users see an event roughly 5-15 seconds behind real-time due to network and buffering delays, we do include an integrated public chat and private messaging facility. We shall discuss below current research intended to enable voice questions and to further enable interactivity and increase engagement.

I14: Allow slides to be independently controllable.
Microsoft Research reports (Jancke, et al., 2000) that remote attendees focus on the speaker from 44% to 56% of the time. They also spend significant time reading or doing other work. 17 of 30 respondents in a survey of our remote attendees cited a need to multitask as a reason for their viewing from their office. Remote viewers should therefore be afforded the ability to page forwards and backwards through slides already presented by the speaker, a capability not available to those in the lecture hall.

I15: Afford remote viewers easy Web access to relevant material.
Once we acknowledge that remote attendees need not always be in lock step with the speaker, we can design so that they can take advantage of this in other ways. For example, we allow users to examine relevant Web pages, including those recommended by the speaker.

A16: Make events available retrospectively through video archives.
Webcasts from Berkeley's BIBS system (Rowe, et al., 2001) were typically viewed retrospectively, with the heaviest use occurring before exams. Viewers of archived videos at Microsoft (He, et al., 1998) number roughly 40% of those who originally watched the talk live in the lecture hall. Viewings continue even one year after a talk. Data to be presented below shows an even greater preference for archived over live viewings of ePresence webcasts.

A17: Do not constrain video archives to be viewed linearly from beginning to end as is the case with live broadcasts.
57% of 33,000 viewing sessions of lecture archives by 9000 Microsoft users (He, et al., 2000) are shorter than 5 minutes; only 17% are longer than 30 minutes. Only 10% of Berkeley's views are for the whole talk (Rowe, et al., 2001).

We go further than the Cornell/Berkeley Lecture Browser and produce archives that are:

- *structured*, in that a talk should be divided into and accessible via an outline of its major logical chunks
- *navigable*, in terms of these chunks, and via the slides
- *searchable*, at least for key words, using the text in the slides or better yet the audio track of the lecture.

A18: Allow archives, like live webcasts, to be viewable interactively and with the capability for annotations and discussion..
Our interpretation of the results from Bargeron, et al. (2001), Cadiz, et al. (2000), and Tiernan and Grudin (2001), all cited above, is that we should augment the chat over live webcasts in two ways. The chat should (optionally) be included in the recorded webcasts. The system should also support annotations and dialogue on top of viewings of the archives.

A19: Support archive construction that is as automatic as possible.
Our experience is that the largest demand for archives is shortly after the event; it is therefore important to prepare the archives as quickly as possible. Given the automation of archive construction, ePresence archives can be ready a few minutes after the conclusion of an event.

S20: Provide for logging and data collection.
Effective iterative design of real software requires ongoing collection and analysis of user experience data. In addition to questionnaires, interviews, and participant observation, it is therefore essential to log and analyze real system data, such as chat messages and questions to the speaker.

General requirements
Our experience viewing webcasts is that it is a cognitively complex task, requiring full attention to understanding the speaker despite physical remoteness, distractions, and the temptation to multitask. Our system and interface design must therefore opt for simplicity rather than features.

We have encountered great diversity in ePresence machine platforms, operating systems, web browsers, and media engines (requirement P3). Our implementation strategy therefore minimizes our dependence on client-side applications that are harder to make portable and robust.

Finally, our ePresence system is intended as an infrastructure for research on technical and social science issues in eLearning (see examples below). We therefore developed a system architecture with maximum malleability and extensibility.

4 Iterative Design and Testing

The project began with our webcasting a series of lectures in the spring of 2000 before we wrote a line of code. We learned about the logistic complexity of supporting both local and remote audiences (Requirement P4), and about the importance of production values, especially of audio quality (Requirement M9). We learned about the inability to get speaker materials in advance (Requirement P7), and the need for interactivity (Requirement I13). Individuals who had missed some lectures stressed the need for archives (Requirement A16).

We built the first version of ePresence in the summer of 2000, and began to use it in webcasting a lecture series throughout the 2000-1 academic year. Insights into useful improvements to the lecture hall (Requirement P6) were communicated to the room's owner, and some improvements made 2001-2. A new building was available in 2002-3, and we finally had a lecture hall that was reasonably suitable.

New versions of the system were introduced frequently. Iterations were based on design meetings informed by participant observation, interviews, and surveys of the system in use. We learned how to produce good audio quality for both local and remote audiences (Requirement M9); how to achieve acceptable quality slide transmission (Requirement M11), despite our inability to get them in advance (Requirement P7); and how to enhance the sense of presence by using two cameras, cinematography, and video switching (Requirement M12).

We struggled with supporting multiple recipient platforms (machines X operating systems X browsers X media engines), and finally introduced online testing procedures in 2001 (Requirement P3). We learned how important the moderator's role was (Requirement P8). We saw hints of how even rudimentary interactivity could support the formation of community among webcast viewers.

Figure 1. A screen shot from a live webcast.

5 The ePresence System

5.1 Functionality

ePresence functionality currently includes video, audio, slide, and screen broadcasting; slide review; integrated moderated chat; private messaging; question submission; and the automated creation of structured, navigable, searchable event archives.

5.2 User Interface

The current interface to access live webcasts is illustrated by the screen snapshot in Figure 1. The video window and its controls are in the upper left; the slide window and its controls are in the upper right; the chat system is at the bottom. Slide controls allow a remote viewer to review any slide already presented by the speaker. The chat system supports public chat, private messages, and questions to the speaker. Web links can also be sent by the speaker and synchronized with the video. (Requirement I15). The "remote desktop" button enables transmission of live 600X800 screen capture streams of live demos from the presenter's computer.

The archives interface allows retrospective navigation and browsing through a webcast using an outline of the logical structure of the talk and its slides and live demo sessions (Figure 2, right side). Slide titles are picked up automatically from Powerpoint in case it is used; the outline is input by the moderator during the talk and if need be updated afterwards using the ePresence Producer (see below). Archive viewers can also navigate by a timeline (Figure 2, bottom). We also allow searching based on key words in the slides when Powerpoint is used.

There is also a separate registration and systems check procedure so that potential viewers can ensure technology compatibility in advance (Requirement P3).

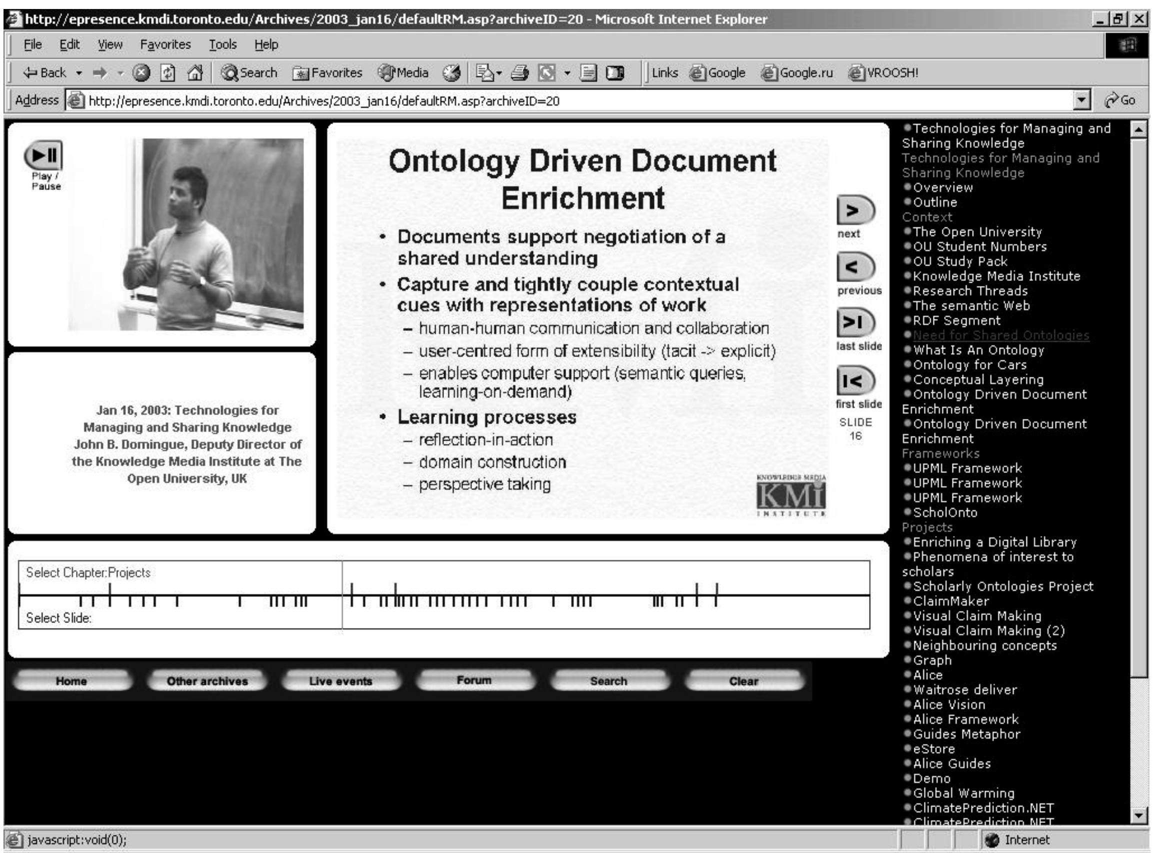

Figure 2. A screen shot from the archive interface.

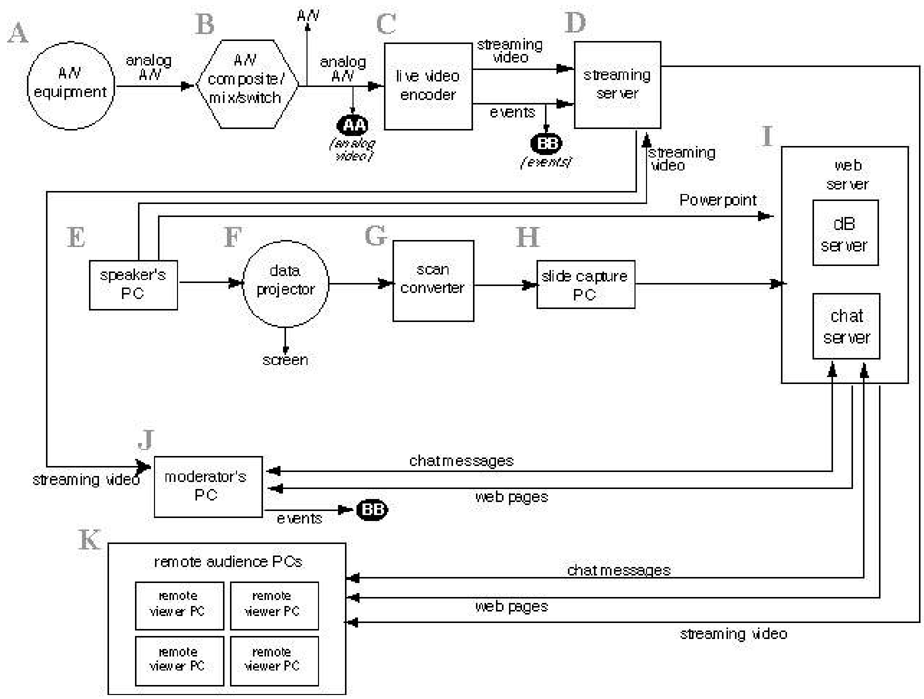

Figure 3. ePresence real-time webcasting system diagram.

5.3 Architecture and Implementation

This is illustrated in Figure 3 (real-time webcasting architecture) and Figure 4 (archive creation process). System components are labeled by capital letters.

Our hardware consists of:

- analog equipment for capturing live audio and video, including cameras and microphones (A)
- analog gear for audio-video compositing[2], including amplifiers, video switcher, and audio mixer (B)
- live video encoding machine(s) (C)
- streaming server(s) (D)
- speaker's laptop (E), data projector (F), scan converter (G), and slide capture PC (H) for projecting and capturing the speaker's "slides"
- Web server machine, incorporating database and chat servers (I).
- moderator's laptop (J)
- a number of remote participant desktop machines (K)
- archives video capture machine (L) and storage (M)
- archives assembly machine (N).

The live streaming server (D) uses the standard Helix® server software from Real Networks and/or Microsoft Windows Media Services. The live encoders (C) are based on the Real Producer SDK[3] and Microsoft WME9 SDK. Both support embedding event information into the live stream in real time, controlling slide queues, providing a web-based time service for the moderator using Simple Object Accessing Protocol (SOAP), and controlling the event and video capturing processes.

Slide capture is currently done by an operator-initiated trigger which grabs a scan converted representation (G) of the data projector's output (F).[4] If presenters allow us to put our software on

[2] The A/V compositor should be reimplemented digitally and output as an MPEG4 stream (Gibson, 2001).

[3] Output of the encoder is currently sent over the campus network to the streaming server, which then broadcasts video streams "to the world".

[4] This can be automated by a scene differencing algorithm.

their machines, screen captures may be transmitted as a separate high-resolution video stream; the screen as shown in Figure 1 is automatically reformatted to provide a larger area for the "remote desktop."

The ePresence web server software (I) can be described as a set of ASP.NET applications and XML Web services that generate the content for remote users depending on client platform, preferences, and role. An authentication program checks the user profile stored in the central database and generates appropriate HTML output. The system currently recognizes three types of users: remote participant, moderator, and administrator.

The chat server is implemented as a set of XML Web services and server-side scripts. The message queue is controlled by the server. On the client side the chat interface is implemented with web forms. The server supports four message types — public chat, private messaging, questions to the speaker, and moderator's announcements. Message refreshing is done using an embedded SOAP client. The number of remote participants is virtually unlimited.

The database server allows execution of SQL queries to support social science research.

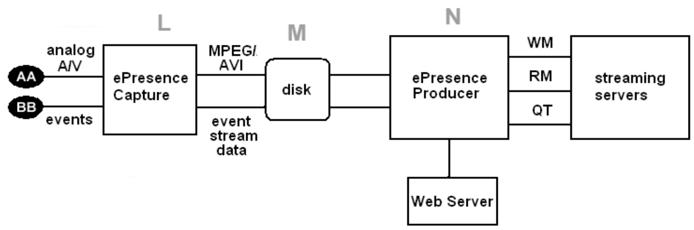

Figure 4. ePresence archive creation system diagram.

The capturing software (L) consists of two parts — video capture and events capture. The video record is encoded in the MPEG1 or AVI media formats; the events information is saved in XML. The event capturing application is a Web service that receives RPCs from the administrator, moderator, slide operator, and live encoding software, so the framework can be described as a scalable distributed system.

ePresence Producer (N) allows capturing the video, corrections and updating event information and timestamps by editing the event stream log file before encoding the video into Quicktime, Real, and Windows media formats. The editor stores video frames and lists of events sorted by time and grouped by type (e.g., "Slides", "Chapters", "Keywords").

The software also supports encoding and publishing the archives on the web. XML to HTML conversion is done with an embedded XSLT engine, which makes it easy to change the archive look and feel with minimal programming and HTML coding. The resulting web archive generated by ePresence Producer is a set of database records, media files, html, and ASP frames which together represent chapters, slides, video, a search form, and an interactive timeline.

In the case of PowerPoint slides, their titles are automatically extracted and inserted into the presentation outline (Figure 2, right). Slide text is also automatically extracted allowing the archives to be searchable.

The ePresence client system (K) is a set of frames rendered in the web browser and java applets. The video frame contains the player plug-in, which is also a script command interpreter. The script commands sent by the encoding software allow synchronizing live events such as slide changing to the video stream. The slide frame is a script that changes the slide by command from the player object, and also allows browsing and enlarging the slides and reviewing of live software demos.

6. System Uses and User Experience

Surveys of both local and remote attendees were administered during the 2000-1 lecture series.

Local attendees (19 responses) liked interacting and networking with people and the sense of community they experience by attending physically. Half felt that the experience was altered due to the remote attendees. They were mostly positive about this, despite some disruptions, and despite having less time to ask questions. Receptivity may have been increased by the novelty of the technology and to the excitement of having a video production crew at the event.

Comments included: "KMDI was truly a link to the outside world" and "The importance of the events felt higher". Some also attended remotely, citing the ability to multitask and the directness of the experience.

Remote attendees (30 responses) included 20 who were satisfied or extremely satisfied, and 6 who were dissatisfied or extremely dissatisfied, citing usability and learning problems. Remote attendance was chosen for reasons including the need to multitask (17), the inability to afford the travel time (12), and not living in the area (9). Almost all viewers liked the ability to interact with other viewers through the integrated chat facility.

	a.m. Day1	p.m. Day1	a.m. Day2	p.m. Day2
Content	11	5	13	16
Technology	116	112	44	41
Administration	38	21	13	10
Social	30	1	28	30
Other	18	19	13	14

Table 1: Categorizing chat messages over the four half-days of WebForum 2001

In November of 2001 we webcast WebForum 2001, the Millennium Dialogue on Child Development (http://www.webforum2001.net/). This consisted of 8 talks and panel discussions that took 12 hours over two days. There were 150 attendees in 6 spaces throughout the host building in Toronto, many in remote viewing locations within the building. 15 to 20 remote viewers were located elsewhere in North America.

Over 600 public and private chat messages were exchanged during WebForum 2001. Table 1 shows how the composition of the chat messages changed over the two days. Of particular interest is the increase in the percentage of messages related to the content of the sessions, from an average of 4% on day 1 to 13% on day 2, and in the percentage of social messages, from 8% on day 1 to 26% on day 2.[5]

Fifteen remote real-time attendees for our 2001-2 ten-lecture series filled out an online survey form. Ten said they were "satisfied" or "extremely satisfied" with the webcasts; five did not answer the question.

We had roughly 300 live attendees at lectures and almost 200 remote real-time attendees during the 2001-2 series. We added the Web archive subsystem partway through series. Although we did not actively publicize its availability, we have had over 3200 hits to the series archives in the year and a quarter since then.

[5] White, et al. (2000) report that messages went from 27%:62%:11% content:technology:social to 60%:14%:26% over the last 3 sessions of their course.

7. Summary and Evaluation

Our work may be distinguished relative to the body of work described in Section 2 in that:

- Speakers are not forced to use Microsoft PowerPoint; the system supports the transmission both of slides and rich media including screen captures of live software demonstrations
- Dialogue among remote viewers and questions to the speaker happen with an integrated chat facility
- The system produces automatically structured, navigable, and searchable video archives
- The "client" software for viewing webcasts supports multiple operating systems, browsers, and media platforms, and may be used with connection speeds as low as 56K.

A comparison of ePresence to the most interesting academic research systems appears as Table 2.

Perhaps the most important achievement is the creation of a flexible, modular, extensible infrastructure for exploring frontiers of collaboration technologies for distance learning. Thus ePresence is a solid foundation for future research, to be described in Section 8 below.

8. Future Work

There remain many research challenges and opportunities, which may be characterized in terms of **P**(articipants), **M**(edia), **I**(nteractivity), **A**(rchives), and **S**(ystem).

	ePresence	Sun Forum	KMi Stadium	Berkeley BIBS	Microsoft TELEP
Status	In experimental use	No longer active	In production use	In production use	No longer active
Audience	In-house+remote	Remote only	Remote only	In-house+remote	In-house+remote
Media richness	Audio, video, Powerpoint, live slide capture, live remote desktop	Audio, video, slides	Audio, video, slides	Audio, video, Powerpoint	Audio, video, Powerpoint
Interactivity	Public chat, private messages during live events, text questions, surveys/quizzes at log-out only, remote speaker slide control	Private messages, voice and text questions, audience polls and other feedback mechanisms	Public chat, text questions, audience polls and other feedback mechanisms	Text questions	Public chat, private messages during live events, text questions, polls during event, list and images of attendees visible to speaker
Archives	2-level navigable structure defined by outline and slides; timeline; archive search using text in slides	No archives	Archives as linear video	1-level navigable structure defined by slides; timeline; archive search using text in slides	Archives as linear video
Viewer platforms	Windows, Mac; Explorer, Navigator via Internet	Sun hardware via corporate intranet	Windows, Mac via Internet	Windows, Mac via Internet	Windows, Explorer only via Internet

Table 2: A comparison of research webcasting systems

8.1 Providing More Participant Access

ePresence currently is only available to users of desktop or laptop computers with resolutions of 1024X768 or higher. We need to remove this restriction.

More specifically, we intend to support hand-held mobile access to allow local participants to participate in the chat. We shall study the impact of lecture attendees typing messages into their PDAs or cell phones during a talk.

We also intend to enhance scalability by investigating new coding and streaming methodologies such as MPEG21 (Fassbakk, et al., 2001).

8.2 Enriching the Media

We seek to enhance the engagement and sense of presence experienced by remote participants, and to bridge the distance between local and remote participants. We are interested in how spatial (split screens) and temporal multiplexing (cuts and dissolves) enhance presence in webcasts. We are interested in how learning, attention, appeal, and stress vary with video quality.

8.3 Improving Interactivity

We have begun work towards reducing the delay between events and receipt of events. Upgrading from the Real server to the new Helix server has enabled reduction in this delay from 25 to 30 second to 5 to 15 seconds. Our goal is a delay no longer than 5 seconds. We will then introduce voice over IP and allow questions to be spoken as well as typed.

We also seek to implement the ability to switch video transmission from webcasting to conferencing in case only a few sites are involved, thus removing the delay totally and allowing video to be multi-directional.

We plan the addition of threaded discussions over the archives. We conjecture and intend to test if this environment will encourage and support the formation of a "community" of online participants. We are particularly interested in how the online discourse enhances viewer understanding, and on how this depends upon the use of public chat and private messaging, and the integration of real-time chat during an event with later discussions over archived video.

We intend to study the value of allowing remote viewing by groups as well as individuals. Gibbons, et al. (1977) showed that individuals viewing a video in a group learned more than those attending the live class who in turn learned more than those viewing a video by themselves. Smith, Sipusic, and Pannoni (1999) have also reported similar conclusions using videoconferencing. We seek to learn if these results also apply to webcasts.

8.4 Enhancing the Archives

We plan a study of how viewers use structured, navigable, searchable archives, and research on further automating the production of structured searchable archives. Goals include automatically recognizing key words in the audio track and using natural language processing to find topics.

We plan to integrate the Expresto Creator digital video authoring, editing, and production capability (Baecker, et al., 1996; Baecker and Smith, 2003) into ePresence, which will allow the easy addition of titles, special effects, and editing of archived productions.

8.5 Improving the System

The major cost of using ePresence is now the camera operators and audiovisual technicians to produce quality webcasts. We plan research aimed at automating these functions, leveraging the work of Machnicki and Rowe (2002), Rui, et al., (2003) and Kapralos, et al. (2003a,b).

Finally, we have discovered, as did Scott and Eisenstadt (1988), that there is a great need for flexibility in eLearning webcasting systems. Canned products rarely meet the great variety of needs encountered in different situations. We are therefore investigating the feasibility of releasing our software "open source." Our hope in so doing is to enable the formation of a community of quality institutions and individuals who can collaborate on the research and development needed to fully exploit the potential of the ePresence infrastructure.

9. Summary and Conclusions

We have presented a principled design for an interactive webcasting system that allows flexibility in the materials used by the speaker, that engages real-time remote viewers, and that allows retrospective viewing of automatically computed, structured, navigable, searchable archives. The design has evolved through an iterative, user-centred design process and through the interplay of project goals, results from the literature of video communications, and observations of the use of prototype implementations in real use. There remain a rich set of social science research questions about how people use such technology and how it impacts their communication, engagement, and learning.

Acknowledgements

We are grateful for research support to the Bell University Laboratories of the University of Toronto, the Natural Sciences and Engineering Research Council of Canada, and the IBM Centre for Advanced Studies. Primary participants included Gale Moore, Ph.D., co-director of the ePresence lab; Peter Wolf, system architect and developer; Maciek Kozlowski of Videotelephony, Inc. (Ottawa), consultant; Eric Smith and Russ Schick, programmers; Anne Postic, interface designer; Kelly Rankin, webcast producer; and Ben Mazzotta, leader of the technical production crew. We also thank collaborators Dan Keating, Anita Zijdemans, and Paul Smith.

About the Author

Ronald Baecker is the Bell University Laboratories Chair in Human-Computer Interaction, Professor of Computer Science, and founder and Chief Scientist of the Knowledge Media Design Institute of the University of Toronto. Baecker is an active researcher, lecturer, and consultant on human-computer interaction and user interface design, user support, software visualization, multimedia, computer-supported cooperative work and learning, the Internet, and entrepreneurship and strategic planning in the software industry. He has published over 100 papers and articles on topics in these areas and is also the author or co-author of four books and two published videotapes. He is the co-holder of 2 patents and 1 patent pending, and has founded and run two software companies.

References

1. Abowd, G.D., Harvel, L.D., Brotherton, J.A. (2000). Building a Digital Library of Captured Educational Experiences. *Kyoto International Conference on Digital Libraries*, 395-402.
2. Abowd, G.D. (1999). Classroom 2000: An experiment with the instrumentation of a living educational environment. *IBM Systems Journal* 38(4), 508-530.
3. Baecker, R.M. (1993). *Readings in Groupware and Computer Supported Cooperative Work: Software to Facilitate Human-Human Collaboration*. Morgan Kaufmann.
4. Baecker, R.M. (2002). Highly Interactive Webcasting with Structured Archives, Poster presentation. *Proc. CSCW'02*. Conference Supplement, 175-6.
5. Baecker, R.M., Moore, G., & Zijdemans, A. (2003). Reinventing the Lecture: Webcasting made Interactive. *Proc. HCI International*

2003, Lawrence Erlbaum Associates, Volume 1, 896-900.

6. Baecker, R.M., Rosenthal, A., Friedlander, N., Smith, E., & Cohen, A. (1996). A Multimedia System for Authoring Motion Pictures, *Proc. ACM Multimedia 96*, 31-42.
7. Baecker, R.M. and Smith, E. (2003). Modularity and Hierarchical Structure in the Digital Video Lifecycle. *Proc. GI'2003*, Halifax NS, 217-224.
8. Bargeron, D., Gupta, A., Grudin, J., Sanocki, E., Li, F. (2001). Asynchronous Collaboration Around Multimedia and Its Application to On-Demand Training. *Proc. HICSS-34*.
9. Cadiz, JJ., Balachandran, A., Sanocki, E., Gupta, A., Grudin, J., Jancke, G. (2000). Distance Learning Through Distributed Collaborative Video Viewing. *Proc. CSCW 2000*, 135-144.
10. Chiu, P., Kapuskar, A., Reitmeier, S., Wilcox, L. (2000). Room with a Rear View: Meeting Capture in a Multimedia Conference Room. *IEEE Multimedia October-December*, 48-54.
11. Christel, M.G. (1999). Visual Digests for News Video Libraries. *Proc. ACM Multimedia 1999*, 303-311.
12. Cordero, C., Harris, D., Hsieh, J. (1996). High Speed Network for Delivery of Education-on-demand. *Multimedia Computing/Networking, Proc. SPIE 2667*.
13. Fassbakk, E., Manzanares, P., Yago, J.L., & Perkis, A. (2001). An MPEG21 framework for streaming media. *Proc. IEEE Fourth Workshop on Multimedia Signal Processing*, 147-152.
14. Finn, K., Sellen, A.J., and Wilbur, S. (Eds.) (1977), *Video-Mediated Communication*. Lawrence Erlbaum.
15. Gibbons, J.F., Kincheloe, W.R., Down, K.S. (1977). Tutored Videotape Instruction: A New Use of Electronics Media in Education. *Science* 195, 1139-1146.
16. Gibson, J.D. (2001). MPEG-4 and MPEG-7. In Gibson, J.D. (Ed.), *Multimedia Communications: Directions and Innovations*, Academic Press, 117-128.
17. He, L., Grudin, J., Gupta, A. (2002). Designing Presentations for On-Demand Viewing. *Proc. CSCW2000*, 127-134.
18. Hurst, W., Maass, G., Muller, R., Ottmann, T. (2001). The "Authoring on the Fly" System for Automatic Presentation Recording. *Proc. CHI'2001 Extended Abstracts*, 5-6.
19. Hsiung, R. C. (2000). Grand Rounds: The Next Generation. *Bulletin of the Canadian Psychiatric Association* 32(1), 9-13.
20. Isaacs, E.A., Morris, T., Rodriguez, T.K. (1994). A Forum for Supporting Interactive Presentations to Distributed Audiences. *Proc. CSCW'94*, 405-416.
21. Isaacs, E.A., Morris, T., Rodriguez, T.K., Tang, J.C. (1995). A Comparison of Face-to-Face and Distributed Presentations. *Proc. CHI'95*, 354-361.
22. Jancke, G., Grudin, J., Gupta, A. (2000). Presenting to Local and Remote Audiences: Design and Use of the TELEP System. *Proc. CHI'2000*, 384-391.
23. Kapralos, B., Jenkin, M. R. M., & Milios, E., (2003a) Audio-visual Localization of Multiple Speakers in a Video Teleconferencing Setting. *Int. J. Imaging Systems and Technology* [to appear]
24. Kapralos, B., Barth, A., Ma, J., & Jenkin, M. (2003b) A System for Synchronous Distance Learning, *Proc. VI'2003*, Halifax, NS.
25. Machnicki, E. and Rowe, L.A. (2002). Virtual Director: Automating a Webcast. *Proc. SPIE Multimedia Computing and Networking*, January 2002.
26. Mane, A. (1997). Group Space: The Role of Video in Multipoint Video Conferencing and Its Implications for Design. In Finn, K., Sellen, A.J. and Wilbur, S. (Eds.), op. cit., 401-414.
27. Moran, T.P., Palen, L., Harrison, S., Chiu, P., Kimber, D., Minneman, S., van Melle, W., Zelweger, P. (1997). I'll Get that Off the Audio: A Case Study of Salvaging Multimedia Meeting Records. *Proc. CHI' 97*, 202-209.
28. Mukhopadhyay, S., and Smith, B. (1999). Passive Capture and Structuring of Lectures. *Proc. ACM Multimedia 99*, 477-487.
29. Rowe, L.A., Harley, D., Pletcher, P., Lawrence, S. (2001). BIBS: A Lecture Webcasting System. BMRC Tech. Report June 2001, Univ. California, Berkeley.

30. Rui, Y., Gupta, A., & Grudin, J. (2003). Videography for Telepresentations. *Proc. CHI'2003*, 457-464.
31. Scott, P. J., & Eisenstadt, M. (1998) Exploring Telepresence on the Internet - the KMi Stadium Webcast Experience. In M. Eisenstadt and T. Vincent (Eds.). *The Knowledge Web*. Kogan Page.
32. Scott, P. and Mason, R. (2001). Graduating Live and On-line: The Multimedia Webcast of the Open University's Worldwide Virtual Degree Ceremony. *Proc. Webnet 2001*, Orlando Florida, AACE.
33. Scott, P. and Quick, K. (2002). Technologies for Electronically Assisting Nursing Communication. *Proc. IADIS 2002*, Lisbon Portugal.
34. Smith, R.B., Sipusic, M.J., and Pannoni, R.L. (1999). Experiments Comparing Face-to-face with Virtual Collaborative Learning. *Proc. CSCL 1999*, Palo Alto CA., 558-566.
35. Tiernan, S.L. and Grudin, J. (2001). Fostering Engagement in Asynchronous Learning through Collaborative Multimedia Annotation. *Proc. Interact 2001*.
36. Wactlar, H.D., Christel, M.G., Gong, Y., Hauptmann, A.G. (1999). Lessons Learned from Building a Terabyte Digital Video Library. *IEEE Computer*, Feb. 1999, 66-73.
37. Wainhouse Research (2002). The Business Case for Webcasting: A Comparison of Conferencing Services and Benefits. www.wainhouse.com/files/papers/wr-bizcase4webcasting.pdf.
38. White, S.A., Gupta, A., Grudin, J., Chesley, H., Kimberly, G., Sanocki, E. (2000). Evolving Use of a System for Education at a Distance. *Proc. HICSS-33*.