

Applications of Scalable Multipoint Video and Audio Using the Public Internet

Robert D. Gaglianello & Marah F. Rosenberg
Bell Laboratories, Lucent Technologies

rdg@lucent.com & mfr@lucent.com

Abstract

This paper describes a scalable multipoint video system, designed for efficient generation and display of high quality, multiple resolution, multiple compressed video streams over IP-based networks. We present our experiences using the system over the public Internet for several “real-world” applications, including distance learning, virtual theater, and virtual collaboration. The trials were a combined effort of Bell Laboratories and the Gertrude Stein Repertory Theatre (TGSRT). We also present current advances in the conferencing system since the trials, new areas for application and future applications.

Keywords: Multipoint Video Conferencing, Video Compression, Motion JPEG, Virtual Theatre, Production Industry, CSCW

Introduction

Bell Laboratories created several application scenarios with educators and artists. In each instance the academic and artistic goals as well as the technical implementation was different. By conducting trials in typical environments we have gathered valuable information about the viability of high quality video over the public Internet. This is part of a larger project that is currently integrating new models of electronic collaboration with web-based services. (Ginsberg, Hodge, Lindstrom, Sampieri, Shiau, 1998)

The Gertrude Stein Repertory Theatre (TGSRT, located at: 15 West 26th Street, New York, New York) was our main collaborator in these trials. The theater was founded in 1990 with the intent to expand and revitalize performing arts. The theater has a history of pursuing multicultural and international collaborations — from training, education and productions. The theater also has a history of partnering with cultural and educational institutions. In pursuing their mission, TGSRT has developed a strategy aimed at a global infrastructure of relationships as well as technology. TGSRT presented an interesting challenge, to visually collaborate around the world.

The Montage IP-based video conferencing and collaboration was the best solution for the collaborations and classes for several reasons. It provided to us a scaleable, extensible and multipoint international solution. One of the major collaborations occurred with artists in Russia. An ISDN or other Phone Call based system for video conferencing is extremely cost prohibitive. Additionally, a solution with dedicated lines would require that meetings always occur in the same place. This cuts into the flexible and spontaneous nature of artistic collaboration. The machines used in these collaborations were often placed on wheels and moved from location to location within a university or rehearsal studio. This gave the artists and arts educators the flexibility they needed to express themselves digitally. This IP video system allows for multiple frame-rates, frame-sizes and bit-rates. Participants in a conference can tailor the video that they receive to a fine level of detail. Each group can have a different amount of bandwidth and still converse with each other, at the bit-rates, frame-rates and frame sizes that their network can support.

The trials discussed within focus on using web-based collaboration in the context of arts production and arts education. In creating web-based collaborative spaces, Lucent benefits from experimentation outside of the laboratory environment and gathers valuable information about the process of on line collaboration.

Material published as part of this journal, either on-line or in print, is copyrighted by the publisher of Informing Science. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission and payment of a fee. Contact Editor@inform.nu to request redistribution permission.



Figure 1
Multiple Windows at Multiple Resolutions

Technology Overview

Montage (Gaglianello & Cash, 1995) is capable of generating multiple independent bit-streams corresponding to different resolutions of a single video source and is able to decode up to 32 such streams from multiple sources and display those multiple videos simultaneously. Montage provides this functionality by simple additions to standard Motion JPEG (MJPEG) codecs designed for handling a single video stream (Wallace, 1991; Civanlar & Gaglianello, 1998). The current implementation consists of a MJPEG decoder along with an SVGA graphics controller on one PCI card, and a MJPEG multiple resolution encoder and CD-quality audio I/O on a second PCI card. The compression techniques used by Montage are inherently low latency (less than 48 ms. for encoding and decoding), making the system ideal for highly interactive applications.

The encoder generates four 30 frame/sec independent resolution compressed streams from a single NTSC or PAL input video source. The resolutions are full-size (640x480), 1/4 size

(320x240), 1/16 size (160x120), and 1/64 size (80 x56). Figure 1 shows the 1/4 and 1/16 size resolutions.

The video-input section has two NTSC inputs that can be switched in software. For example, a user can encode and transmit from a camera as well as a video scanner, VCR or even another camera without having to plug and unplug cables. Also located on the encoder is a bi-directional "Sound-Blaster" like audio interface capable of generating and playing back CD-quality stereo sound. The audio section has two line level stereo inputs and a stereo microphone input. Again, input audio selection is software controlled.

Figure 2 shows how communications between different Montage stations runs over any TCP/IP based network infrastructure, such as IP over ATM, Ethernet, and IP over ISDN. Video can be scaled to the appropriate bandwidth by varying the resolution, the frame rate and the image quality of transmitted video. By judicious choices of these three parameters, it is possible to support a wide range of bandwidths. (128 Kb/sec to 155 Mb/sec)

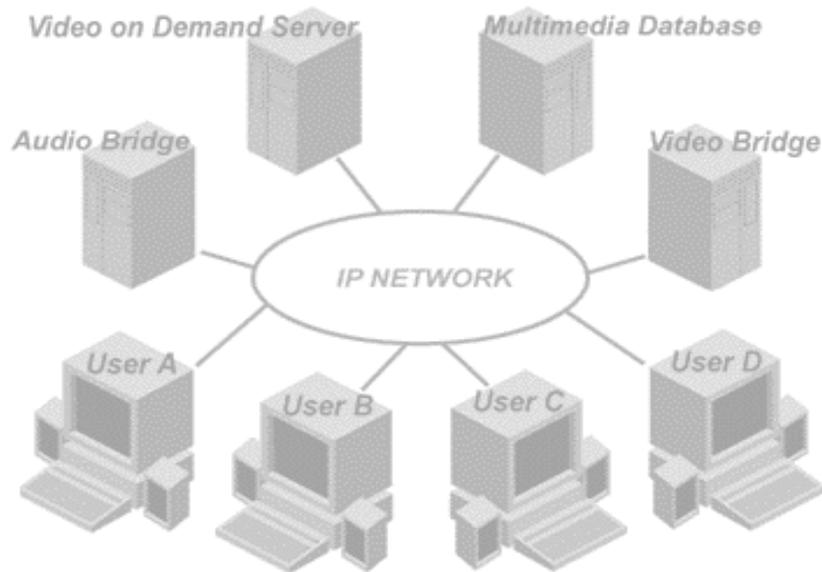


Figure 2
G2 Operates Over Any IP Network

In 1998, the second-generation board was developed (G2). The G2 board was designed to lower the cost of each board and include the video and audio encoding / decoding on a single PCI card. The G2 board provides encoding features similar to the first board, such as low latency encoding, multiple stream simultaneous encoding and support for multiple resolutions and band-

widths.

The VGA controller that was integrated into the first board has been removed. This creates greater flexibility for display, because any commercial VGA card can be installed on the machine (Figure 3). The video streams from the G2 are overlaid on

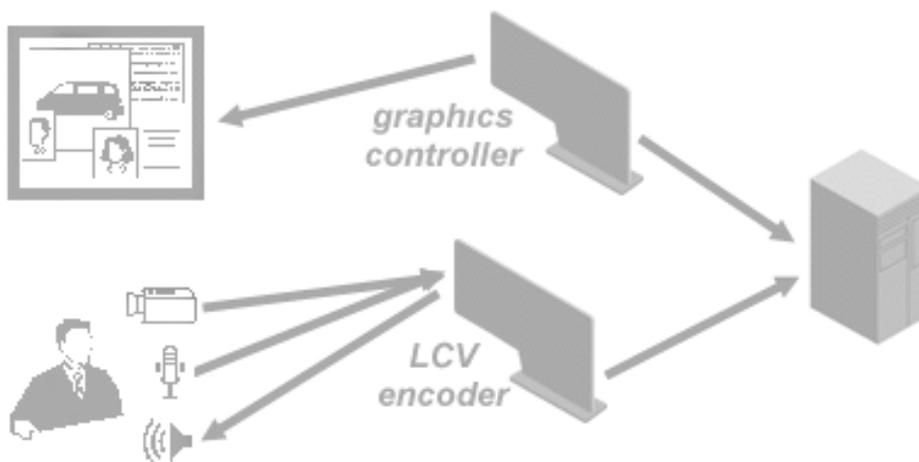


Figure 3
G2 Uses the existing graphics controller

Scalable Multipoint Video

the rest of the video using a pass-through cable. The G2 provides S-Video and Composite video inputs as well as stereo in/out. In addition to providing MJPEG video and Stereo quality Audio, the G2 board can also be used as a generic video capture card that may be used with other encoding schemes, such as H.263 (ITU-T H.262, IAO/IEC 13818-2, 1994) and Video for Windows applications. (Video for Windows is a registered trademark of the Microsoft Corporation. For more information, visit <http://www.microsoft.com/>)

The software has evolved to support the ITU standards T.120 (ITU-T Recommendation T.120, 1996) and T.127 (ITU-T Recommendation T.127, 1995) for data collaboration and file transfer. Other conferencing features have been implemented as well, including the support for an LDAP (Wahl, Howes, Kille, IETF 1997) directory service. These changes have provided new opportunities for experimentation in different industries. The Video Production/Post Production Industry need for broadcast quality video has presented an interesting case study that is discussed in detail later in the document.

The Trials

University of Georgia, Athens, April 1997

The Gertrude Stein Repertory Theater was invited to teach a one-week intensive course on acting techniques at the University of Georgia, Athens. Because of pre-existing engagements, the theater's historian and eastern movement experts in New York City were unable to travel to Georgia. Instead, the co-directors and founders of the theater, Cheryl Faver and John Reaves, both frequent lecturers at Yale University and New York University Theater Departments, decided to teach the class. Kelvin Chai, an actor trained in the traditional form of Chinese Opera and Rachel Shteir, the theater's historian were able to participate via videoconference. Montage enabled theater students to interact with theater professionals who are typically too busy to leave their work to lecture outside of Manhattan.

Project Discovery, June-August 1997

Project Discovery is an Internet-interactive distance-learning course in professional film and television production developed by many of the foremost leaders in the fields of film, television and communications technology for high school students in the United States. Each of the two phases consists of a ten-week, tuition-free, extracurricular summer course of virtual classes, interactively taught over the Internet using Montage.

The class was taught over the Internet to high school juniors and seniors from Jacksonville Florida, with little or no computer experience, as seen in Figure 4. Janet Nepris, an NYU professor located in New York City interacted with the students in Florida as they developed a writing style and possible scripts. Over the



Figure 4
Project Discovery, View from Florida

Internet we connected two Montage Units, one in NYC and the other in Jacksonville, Florida. Our first week of trials we were able to achieve an average of 15-20 fps at QCIF. After time, our performance degraded to less than 8fps at QCIF. After some investigation we discovered that we were losing packets at a specific router in Washington, DC. Montage, in conjunction with the TraceRoute utility of the Internet was able to pinpoint a network weakness that has been repaired. More information about Project Discovery is available at <http://www.owf.org/discovery>.

St. Petersburg, Russia at 55 Broad Street, July 1997

A multipoint Montage session was set up to demonstrate a digital rehearsal, where Actors in New York and Russia worked with a New York director. Designers in New Haven were able to show costume renderings and set models. By using a lipstick camera (very small and thin) in New Haven, the designers were



Figure 5
St. Petersburg, Russia at 55 Broad Street,

able to show the action of scenes, such as descending a staircase, to the actors in both locations. Figure 5 illustrates the Russian and American Actors, with the help of a translator, rehearsing a bilingual scene from a play.

Over the Internet we connected 3 Montage Units. The units were located at Yale University in New Haven, CT, USA, Russian Academy of Dramatic Arts, St. Petersburg, Russia and New York City. In each location the available connection to the LAN was utilized, as seen in Figure 6.

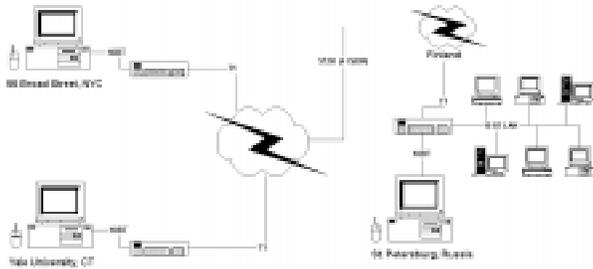


Figure 6. Montage at three locations.

Each of the sites was able to connect with the other 2 participants using QCIF video resolution. The average frame rate was 6-8fps. The end result was 3 QCIF windows on the desktop while using a shared application.

University of Southern California (USC), September 1997

This “cross-country casting call” demonstrated the possible uses of videoconferencing to the California entertainment industry. The casting agent was in California and his assistant was in New York with potential actors. The New York actors read from a script to audition for a part in a television sitcom. They took criticism and direction from the casting agents in California.

The Montage in NYC was connected via a T1 line to their Internet service provider (ISP) and the Montage in California was connected via USC’s Local Area Network to the Internet. An average of 10 fps at CIF was achieved at both sites, with clear audio. This was by far the most successful trial. This showed how important a good network connection and bandwidth is to the performance of IP based video conferencing.

Binghamton University, State University of New York, September 1997

The next distance learning experiment closely involved The Gertrude Stein Repertory Theatre, NY. TGSRT was commis-

sioned by the university’s drama department to develop an innovative integrated digital environment for teaching and learning theater history. With the theater, Lucent Technologies developed a database-derived website to be used over the Internet during the classes. Montage workstations were installed so that the theater experts in New York could teach the classes being held at Binghamton University.

The classes were scheduled for 9 am, EST. We found that performance quickly degraded in the minutes past 9:00 am. All of our tests performed before 9 am generated much better results. This seems to be an artifact of the volume of university and Internet traffic that occurs in the first hours of the morning workday. Tests before 9 am achieved CIF windows at 8 fps and above with 16 kHz ADPCM audio. Tests after 9 am achieved QCIF windows at 4 fps and less with inaudible audio, regardless of the compression scheme. Too many packets were lost between the two sites to create a cohesive videoconference. To gain a new perspective, we created a similar situation between Binghamton University and Yale University. We found that a videoconference of at least 8 fps, CIF size with 16 kHz ADPCM audio was achievable at many different times. Tests are still ongoing with the university as they upgrade parts of their network.

Yale University, October 1997

Montage was used over the public Internet to connect scholars at Yale University with a group in St. Petersburg Russia. Yale University sponsored a symposium on Russian Theater. Maria Valensky is the grand daughter of a famous Russian director Meyerhold. Ms. Valensky lived with her grandfather in St. Petersburg until the government arrested him. Ms. Valensky talked to the Yale audience about her grandfather’s arrest via Montage videoconferencing. She also shared other personal stories about Meyerhold’s work, and her role as the librarian and curator of a museum dedicated to his theatrical works.

At Yale University, we connected to the Internet via the University Administrative Local Area Network, which was then connected to the Internet via a T1 line. In Russia, we used the RUSNET – Russian Network connection to the Internet that was roughly equivalent to 384 KB. Even though they have a T1 to the Internet within Russia, the connection through Finland was significantly less. However, our connection was terrific, a consistent CIF frame at 5 fps with clear audio.

Click 3x, 1998-ongoing

Click 3x is a production and postproduction firm with multiple locations across the United States. They specialize in CG and CEL animation, visual effects and composite as well as editorial work for film and television. (CG and CEL are types of animation. CG stands for Computer Generated, and CEL refers to a

Scalable Multipoint Video

more traditional, non-digital animation technique.) Because of their need for near broadcast-quality video and a diverse client and talent base the G2 system has been implemented as a video and data conferencing solution. The G2 system software integrates support for a third party MPEG1 (ISO/IEC JTC 1/ SC 29/ N071, 11172, 1992) encode and decode solution from Optibase. (Optibase is a leading provider of hardware and software that is based on the MPEG standard. Optibase's products are designed for creating and distributing business-quality digital video over networks, in custom-designed solutions and on diverse digital media such as DVD discs. For more information about Optibase, visit <http://www.optibase.com>.) Click 3x encodes the clips to be discussed prior to the meeting and then uses the G2 system to play back the clips in real time during a videoconference. The MPEG video is available as another window in the conference or displayed separately on an NTSC monitor. The videoconference participants are encoded in MJPEG or H.263.

Generally only two or three locations are communicating together at any given time. The connections are established over the company's Virtual Private Network. They discovered that they are only getting approximately 75% of their T1 bandwidth in each location reliably. The G2 system does give them the ability to review material in real time without the costly travel costs. More information on Click 3x can be found at <http://www.click3x.com>.

Conclusions

The experiences of these trials using the Montage and G2 systems over the public Internet have shown that it is possible to obtain useful and compelling results for the participants. By collaborating with different sets of potential users, different strengths and weaknesses are found. We are continuing to research the system in different situations. Some work is best when completed face-to-face, other work can easily be accomplished with the G2 conference and collaboration solution. The trials here have been fairly successful in accomplishing their goals of evaluating the systems for distance education and col-

laboration in the fine and production arts. Experiments continue in other fields, such as tele-medicine. It is only through such collaborations that it is possible for technologists to become more aware of current problems that can be solved with current technology.

References

- "Coding of moving pictures and associated audio - for digital storage media up to about 1.5 Mbits/s" (1992) ISO/IEC JTC 1/ SC 29/ N071, 11172, International Standard, November 1992
- "Data Protocols for Multimedia Conferencing", (1996) ITU-T Recommendation T.120, July 1996.
- "Generic Coding of Moving Pictures and Associated Audio Information: Video" (1994) Recommendation ITU-T H.262, IAO/IEC 13818-2, Draft International Standard, November 1994.
- "Multipoint Binary File Transfer Protocol", (1995) ITU-T Recommendation T.127, August 1995.
- Civanlar, M. R. & Gaglianella, R. D. (1998) "Efficient Multi-Resolution, Multi-Stream Video Systems Using Standard Codecs", Journal of VLSI Signal Processing.
- Gaglianella, R. & Cash, G. (June 1995) "Montage: Continuous Presence Teleconferencing Utilizing Compressed Domain Video Bridging", IEEE ICC95, Seattle, Washington, June 1995.
- Ginsberg, A., Hodge, P., Lindstrom, T., Sampieri, B. & Shiau, D. (1998) "The Little Web Schoolhouse: Using Virtual Rooms to Create a Multimedia Distance Learning Environment", Proceedings of ACM MULTIMEDIA 98, Bristol, England.
- Wahl, M., Howes, T. & Kille, S. (1997) "Lightweight Directory Access Protocol (Version 3)". IETF Internet Draft, RFC 2251, December 1997.
- Wallace, G. (1991) "The JPEG Still Picture Compression Standard," Communications of the ACM, Volume 34, Number 4, April 1991, pp. 30-44.

