

## **ABSTRACT**

RETTINGER, LEIGH ANNE. Desktop Videoconferencing: Technology and Use for Remote Seminar Delivery. (Under the direction of Dr. Thomas K. Miller III.)

The purpose of this research project has been to investigate the current state of desktop videoconferencing technology and to evaluate the potential effectiveness of this technology for delivering interactive seminars to a remote audience. Enabling technologies for desktop videoconferencing are discussed, including the compression of audio and video data and the transmission of this data over various communication channels. Features of currently available desktop videoconferencing systems and emerging interoperability standards are addressed. Non-technical aspects of desktop videoconferencing systems such as the value added by audio, video, and integrated computer applications and the importance of usable user interfaces are discussed. To demonstrate the use of desktop videoconferencing for distance learning, a weekly seminar class was broadcast, using the Internet MBone, to remote participants located throughout the state of North Carolina. Delivery of this demonstration project and observations about this project are discussed in detail. This demonstration project is judged to be a successful demonstration of the potential application of desktop videoconferencing for distance learning. Recommendations for future projects of this nature are made.



**DESKTOP VIDEOCONFERENCING:  
TECHNOLOGY AND USE FOR REMOTE SEMINAR DELIVERY**

by

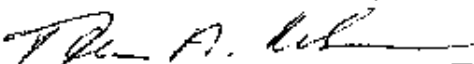
**LEIGH ANNE RETTINGER**

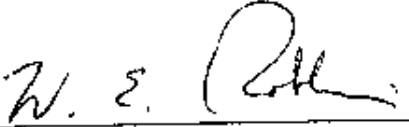
**A thesis submitted to the Graduate Faculty of  
North Carolina State University  
in partial fulfillment of the  
requirements for the Degree of  
Master of Science**

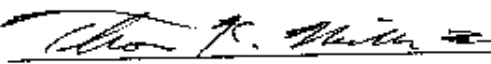
**COMPUTER ENGINEERING**

**Raleigh**

**July 1995**

  
\_\_\_\_\_  
Dr. Arne A. Nilsson

  
\_\_\_\_\_  
Dr. Woodrow E. Robbins

  
\_\_\_\_\_  
Dr. Thomas K. Miller, III  
Chair of Advisory Committee

**APPROVED BY:**

---



[TABLE OF CONTENTS]

## **DEDICATION**

Dedicated to my parents, who have provided me with support emotionally and financially throughout this long long journey called my college career.

## **BIOGRAPHY**

Leigh Anne Rettinger was born on May 27, 1967 in Cumberland, Maryland. Her family moved to Raleigh, North Carolina in 1968 when her father took a job with International Business Machines in Research Triangle Park.

Leigh Anne attended E.C. Brooks Elementary School, Crosby 6th Grade Center, Carroll Junior High School, and Sanderson High School. In August 1985, she started her undergraduate work at North Carolina State University. In May 1991, after six long years of school and cooperative educational experiences at Northern Telecom and International Business Machines, she graduated Magna Cum Laude with a Bachelor of Science degree in Electrical Engineering. She then worked for a year at Westinghouse in Baltimore, Maryland, before returning to NCSU to begin work on her Master of Science degree in Computer Engineering.

In August 1995, Leigh Anne will start work with Intel Corporation in Hillsboro, Oregon. She will be working in the Personal Conferencing Group on the development of Intel's ProShare videoconferencing software.



[TABLE OF CONTENTS]

## ACKNOWLEDGEMENTS

Many thanks go out to the following people who provided immeasurable help during this project:

Bobby Gia Pham, NCSU Computing Center  
John Bass, NCSU Computing Center  
Phil Emer, NCSU Computing Center  
David Smith, NCSU Engineering Computer Operations  
Lee Collins, NCSU Video Communications Services  
David Nelson, NCSU Video Communications Services  
Phil Johnson, NCSU Video Communications Services  
Darren Ley, NCSU Video Communications Services  
Cary Burnette, North Carolina A&T State University  
Dr. John Kelly, North Carolina A&T State University

Thanks to my fellow SUCCEED/Deliverable Team 5 colleagues:

Dr. Scott Midkiff, Virginia Polytechnic Institute and State University  
Rhett Hudson, Virginia Polytechnic Institute and State University  
Dr. Niyazi Bodur, University of North Carolina at Charlotte  
Dr. John Grant, University of North Carolina at Charlotte  
Jeric Newby, University of North Carolina at Charlotte

Funding for this project was provided in part by SUCCEED, The Southeastern University and College Coalition for Engineering Education.

Special thanks to my advisor and committee chair Dr. Tom Miller, who has been a constant source of support and inspiration.



[TABLE OF CONTENTS]

# TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF ABBREVIATIONS AND TERMINOLOGY

### 1. INTRODUCTION

- 1.1 Background
- 1.2 Objective
- 1.3 Organization of the study
- 1.4 Scope and limitations
- 1.5 Research methodology

### 2. TECHNOLOGY

- 2.1 Audio
  - 2.1.1 Audio sampling
  - 2.1.2 Audio quantizing
  - 2.1.3 Digital audio compression techniques
    - 2.1.3.1 mu-law and A-law PCM
    - 2.1.3.2 ADPCM
    - 2.1.3.3 LPC and CELP
- 2.2 Video
  - 2.2.1 Color theory
  - 2.2.2 Video formats
  - 2.2.3 Video delivery
  - 2.2.4 Digital video compression techniques
    - 2.2.4.1 MJPEG
    - 2.2.4.2 ITU-T Recommendation H.261
    - 2.2.4.3 CellB
    - 2.2.4.4 nv
    - 2.2.4.5 CU-SeeMe
    - 2.2.4.6 Indeo
- 2.3 Communication channels
  - 2.3.1 Circuit-switched communication channels
  - 2.3.2 Packet-switched communication channels
  - 2.3.3 Broadband ISDN
- 2.4 Desktop videoconferencing systems
  - 2.4.1 General features
  - 2.4.2 Modes of conferencing
    - 2.4.2.1 POTS conferencing
    - 2.4.2.2 Switched 56 conferencing
    - 2.4.2.3 ISDN conferencing
    - 2.4.2.4 LAN conferencing
    - 2.4.2.5 Internet conferencing

- 2.4.2.6 Multicast Backbone (MBone) conferencing
- 2.4.3 Interoperability standards
  - 2.4.3.1 ITU-T Recommendation H.320
  - 2.4.3.2 ITU-T Recommendation T.120
  - 2.4.3.3 PCS
  - 2.4.3.4 ITU-T Recommendation H.324

### 3. HUMAN COMPUTER INTERACTION AND SOCIAL ISSUES

- 3.1 Lecture mode verses collaboration mode
- 3.2 Benefits of moving to the desktop model of videoconferencing
- 3.3 The value of audio
- 3.4 The value of video
- 3.5 The value of integrated computer applications
- 3.6 Problems to overcome
- 3.7 The Forum prototype: A step in the right direction

### 4. DEMONSTRATION PROJECT: INTERACTIVE SEMINAR USING DESKTOP VIDEOCONFERENCING

- 4.1 Choosing the class and the medium
- 4.2 Infrastructure
  - 4.2.1 Hardware
  - 4.2.2 Software
  - 4.2.3 Network topology
- 4.3 Seminar description
  - 4.3.1 Seminar dates
  - 4.3.2 Broadcasting the seminar
  - 4.3.3 Quantitative measurements
  - 4.3.4 Qualitative measurements
  - 4.3.5 Observations from the seminar broadcasts
- 4.4 Other MBone broadcasts
  - 4.4.1 EPA workshops
  - 4.4.2 "Good Morning America" demo

### 5. CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

### 6. REFERENCES

### 7. APPENDICES

- 7.1 World Wide Web desktop videoconferencing product survey
  - 7.2 Remote seminar participant feedback questions
  - 7.3 MBone participants for 03/22/95 and 04/18/95 EPA broadcasts
-



[TABLE OF CONTENTS]

## LIST OF TABLES

- Table 2.1 CIF and QCIF image formats
- Table 2.2 Service requirements for various data types
- Table 2.3 H.320 series of recommendations
- Table 2.4 T.120 series of recommendations
- Table 2.5 H.324 series of draft recommendations
- Table 4.1 Sun SPARCstation 5 stereo audio inputs and outputs
- Table 4.2 DEC 3000 Model 300 Series AXP audio port pin-outs
- Table 4.3 MBone software tools used in the seminar broadcasts
- Table 4.4 Seminar dates and descriptions

## LIST OF FIGURES

- Figure 2.1 S-video connector
- Figure 2.2 DCT-based encoding
- Figure 2.3 Block coefficients and the zig-zag sequence
- Figure 2.4 CellB encoding
- Figure 2.5 WWW accesses to desktop videoconferencing product survey
- Figure 2.6 Multicast delivery: The illusion
- Figure 2.7 Multicast delivery: The reality
- Figure 4.1 SunVideo input ports
- Figure 4.2 Audio and video flow
- Figure 4.3 NCSU MBone network topology
- Figure 4.4 Session Directory (sd): Main window and "New" popup
- Figure 4.5 Network Video (nv): Main window and video popup
- Figure 4.6 Visual Audio Tool (vat): Main window and "Menu" popup
- Figure 4.7 Whiteboard (wb): Main window
- Figure 4.8 'nv' statistics from NC A&T for the 04/12/95 seminar
- Figure 4.9 'vat' statistics from NC A&T for the 04/12/95 seminar
- Figure 4.10 'nv' and 'vat' statistics from UIUC for the 04/18/95 EPA Broadcast



[TABLE OF CONTENTS]



## LIST OF ABBREVIATIONS AND TERMINOLOGY

ADPCM	Adaptive Differential Pulse Code Modulation
ATM	Asynchronous Transfer Mode
BISDN	Broadband ISDN
BNC	Bayonet, Non-Continuous
bps	bits per second
BTC	Block Truncation Coding
CellB	Sun video compression method
CIF	Common Intermediate Format
CSMA/CD	Carrier Sense Multiple Access/Collision Detection
DCT	Discrete Cosine Transform
DVI	Digital Video Interactive
DT-5	Deliverable Team 5
FDDI	Fiber Distributed Data Interface
fps	frames per second
FST	Fast Slant Transform
ftp	file transfer protocol
GSTN	Global Standard Telephone Network
GSM	Groupe Speciale Mobile
Hz	Hertz
IETF	Internet Engineering Task Force
IGMP	Internet Group Management Protocol
IMTC	International Multimedia Teleconferencing Consortium
IP	Internet Protocol
ISDN	Integrated Services Digital Network
ITU	International Telecommunication Union
ITU-T	Telecommunication Standardization Sector of the ITU
JPEG	Joint Photographic Experts Group
LAN	Local Area Network
LPC	Linear Predictive Coding
CELP	Code Excited Linear Predictive Coding
MBone	Multicast Backbone
MCU	Multi-Conferencing Unit
MJPEG	Motion JPEG
NCSU	North Carolina State University
NC A&T	North Carolina Agricultural and Technical State University
NC-REN	North Carolina Research and Education Network
NTSC	National Television Standards Committee
PAL	Phase Alternation Line
PCM	Pulse Code Modulation
PCS	Personal Conferencing Standard
PCWG	Personal Conferencing Working Group
PSTN	Public Switched Telephone Network
QCIF	Quarter CIF
S-Video	Separate Video
SUCCEED	Southeastern University and College Coalition for Engineering Education
TCP	Transmission Control Protocol
ttl	time to live
UDP	User Datagram Protocol
URL	Uniform Resource Locator
WWW	World Wide Web



[TABLE OF CONTENTS]

## **1. INTRODUCTION**

This paper discusses various aspects of desktop videoconferencing and describes an experiment which utilized this technology to deliver a weekly seminar class to participants distributed throughout the state of North Carolina.

### **1.1 Background**

Room videoconferencing has been used for some time as a means to deliver interactive classes to geographically distributed audiences. One example of the successful use of room videoconferencing for class delivery is the North Carolina Research and Education Network (NC-REN). Teleclassrooms and conference rooms across the state of North Carolina are linked together through NC-REN's private microwave facilities. Video Communications Services at North Carolina State University (NCSU) and similar facilities throughout the state produce programs that can be sent to any of the 19 universities, medical schools, and research organizations that are connected via NC-REN.

Advances in computer technology such as faster processors and better data compression schemes have made it possible to integrate audio and video data into the computing environment. A new type of videoconferencing, desktop videoconferencing, has become possible. Unlike room videoconferencing, which requires specially equipped rooms with expensive hardware, desktop videoconferencing can be achieved by adding software and hardware to standard desktop computers.

One benefit of desktop videoconferencing is the convenience of not having to physically move to a special location. Another benefit is the ability to incorporate data from other desktop computer applications into the conference. Desktop videoconferencing systems typically cost a few thousand dollars to set up, which is significantly less expensive than room videoconferencing systems which typically cost a minimum of \$40,000 to set up. [22]

### **1.2 Objective**

In accordance with the goals of Deliverable Team 5 (DT-5) of the Southeastern University and College Coalition for Engineering Education (SUCCEED), the main objective of this research project was to implement a demonstration of electronic connectivity, using desktop videoconferencing, for educational delivery and interaction among NCSU and other SUCCEED institutions. To achieve this objective, existing technology was surveyed, infrastructure was put into place, and a demonstration project was performed and evaluated.

### **1.3 Organization of the study**

Chapter 2 discusses the enabling technology for desktop videoconferencing. Chapter 3 discusses human factors issues involved with desktop videoconferencing. Chapter 4 describes the demonstration project of delivering an interactive seminar using desktop videoconferencing. Chapter 5 discusses conclusions drawn from the demonstration project and gives recommendations for future work.

### **1.4 Scope and limitations**

This paper presents a broad overview of enabling technologies and human factors aspects of desktop videoconferencing. For more detailed discussion of these topics, the reader is encouraged to follow the references.

### **1.5 Research methodology**

The Internet has been a valuable resource for obtaining information about this topic. Much of the material referenced in this paper is available on the Internet. In the reference section, URLs (Uniform

Resource Locators) have been included that are current at the time of publication. Due to the changing nature of the Internet and especially the World Wide Web, there is no guarantee that this information will continue to be available at these locations.

Valuable sources of information about this topic have been the videophone mailing list (videophone@es.net), the MBone mailing list (mbone@isi.edu), the IETF Remote Conference mailing list (rem-conf@es.net), the CU-SeeMe mailing list (CU-SeeMe-L@cornell.edu), and several newsgroups including: comp.dcom.videoconf, comp.speech, comp.multimedia, misc.education.multimedia, rec.video.desktop, comp.compression, comp.compression.research, alt.education.distance, and comp.groupware.



[TABLE OF CONTENTS]

## **2. TECHNOLOGY**

This chapter discusses the enabling technology for desktop videoconferencing. Audio and video must be captured from their analog form and stored digitally to be manipulated by the computer. Uncompressed, this data would require massive amounts of bandwidth to transmit, therefore the data is compressed before it is sent over communication channels. All this must happen in real-time to facilitate communication and interaction.

This chapter begins with a discussion of audio and video encoding and compression. Then, various communication channels are discussed. This chapter concludes by describing currently available desktop videoconferencing systems and emerging standards which allow these systems to interoperate.

### **2.1 Audio**

The frequency of sound waves is measured in Hertz (Hz), meaning cycles per second. The human ear can typically perceive frequencies between 20 Hz and 20 kHz. Human voice can typically produce frequencies between 40 Hz and 4 kHz. [6] These limits are important factors to remember when discussing digital audio encoding. Desktop videoconferencing systems are typically designed to handle speech quality audio which encompasses a much smaller range of frequencies than the range perceptible to humans.

Audio is delivered to computer equipment through various types of connectors. The author has come in contact with phono connectors, 1/8 inch mini phono connectors, 3/32 inch sub-mini phono connectors, and even an audio interface that is pin-compatible with a telephone handset. The moral to this story is that a trip to Radio Shack may be necessary to connect analog audio equipment to a computer.

Digital audio data is usually described using the following three parameters: sampling rate, bits per sample, and number of channels. The sampling rate is the number of samples per second. Bits per sample is the number of bits used to represent each sample value. Number of channels is one for mono, two for stereo, etc.

#### **2.1.1 Audio sampling**

An analog audio signal has amplitude values that continuously vary with time. To encode this signal digitally, the amplitude value of the signal is measured at regular intervals. This is called sampling. According to the Nyquist theory of signal processing, to faithfully represent a signal of a certain frequency, the sampling rate must be at least twice that of the highest frequency present in the signal. [17] Using this theory, sampling is lossless since the original signal can be reconstructed based on the samples. To avoid aliasing distortion, the signal is low-pass filtered to remove any high frequencies that can not be represented by the sampling rate.

Using Nyquist's theory, 8 kHz is a sufficient sampling rate to capture the range of human voice (40 Hz to 4 kHz) and 40 kHz is a sufficient sampling rate to capture the range of human hearing (20 Hz and 20 kHz). In practice, typical sampling rates range from 8 kHz to 48 kHz. [16]

#### **2.1.2 Audio quantizing**

Sampled values representing the amplitude of the signal at the sample time are quantized into a discrete number of levels. The number of levels depends on how many bits are used to store the sample value. For digital audio, this precision usually ranges from 8 bits per sample (256 levels) to 16 bits per sample (65536 levels). [16] Quantization induces error into the data because no matter how many bits of precision are used, it is impossible to represent an infinite number of amplitude values with a finite

number of increments. [17]

Uniform pulse code modulation (PCM) encoding is an encoding method where the quantizer values are uniformly spaced. Uniform PCM is an uncompressed audio encoding format, however some other PCM formats such as mu-law or A-law PCM use quantizer values that are logarithmically spaced, effectively achieving a degree of compression (discussed in section 2.1.3.1).

### **2.1.3 Digital audio compression techniques**

Uncompressed digital audio can require a large amount of bandwidth to transmit. There are many techniques used to compress digital audio. Some of the techniques commonly used in desktop videoconferencing systems are described below. Typically these are techniques that can achieve real-time compression and decompression in software or inexpensive hardware. Some techniques apply to general audio signals and some are designed specifically for speech signals.

#### **2.1.3.1 mu-law and A-law PCM**

With PCM encoding methods, each sample is represented by a code word. Uniform PCM uses a uniform quantizer step spacing. By performing a transformation, the quantizer step spacing can be changed to be logarithmic, allowing a larger range of values to be covered with the same number of bits. There are two commonly used transformations: mu-law and A-law. These transformations allow 8 bits per sample to represent the same range of values that would be achieved with 14 bits per sample uniform PCM. This translates into a compression ratio of 1.75:1 (original amount of information:compressed amount of information). Because of the logarithmic nature of the transform, low amplitude samples are encoded with greater accuracy than high amplitude samples.

The mu-law and A-law PCM encoding methods are formally specified in the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) Recommendation G.711, "Pulse Code Modulation (PCM) of voice frequencies." The mu-law PCM encoding format is common in North America and Japan for digital telephony with the Integrated Services Digital Network (ISDN). The A-law PCM encoding format is common with ISDN in other countries. [16] G.711 is one of the audio standards specified in H.320 (discussed in section 2.4.3.1). Note that at 8 kHz, 8 bits per sample, and 1 channel, mu-law or A-law PCM requires a bandwidth of 64 kbps.

#### **2.1.3.2 ADPCM**

PCM encoding methods encode each audio sample independently from adjacent samples. However, usually adjacent samples are similar to each other and the value of a sample can be predicted with some accuracy using the value of adjacent samples. For example, one simple prediction method is to assume that the next sample will be the same as the current sample. The ADPCM (Adaptive Differential Pulse Code Modulation) encoding method computes the difference between each sample and its predicted value and encodes the difference (hence the term "differential"). [16] Fewer bits (typically 4) are needed to encode the difference than the complete sample value. Encoders can adapt to signal characteristics by changing quantizing or prediction parameters (hence the term "adaptive"). ADPCM typically achieves compression ratios of 2:1 when compared to mu-law or A-law PCM. [6] Differences among different flavors of ADPCM encoders include the way the predicted value is calculated and how the predictor or quantizer adapts to signal characteristics.

Many desktop videoconferencing systems use ADPCM encoding methods. The ITU-T has several recommendations defining different ADPCM methods (G.721, G.722, G.723, G.726, G.727). One of the audio encoding methods specified by H.320 (discussed in section 2.4.3.1) is G.722, "7 kHz

Audio-coding within 64 kbit/s," which uses SB-ADPCM encoding (Sub-Band ADPCM). The G.722 encoder samples at a rate of 16 kHz with 14 bits precision. With the SB-ADPCM method, the frequency band is split into two sub-bands (higher and lower) and the signals in each sub-band are encoded using ADPCM. G.722 has three modes of operation: 64, 56 and 48 kbps. With the 56 or 48 kbps modes, the additional 8 or 16 kbps of bandwidth (assuming a 64 kbps communication channel) can be used for other data.

### **2.1.3.3 LPC and CELP**

There are some encoding methods designed specifically for speech. By using models of the characteristics of speech signals, these encoding methods can achieve good results for speech data. However, these methods usually do not work well for non-speech audio signals. [16] Two encoding methods designed for speech signals are LPC and CELP.

A LPC (Linear Predictive Coding) encoder fits speech signals to a simple analytic model of the vocal tract. The best-fit parameters are transmitted and used by the decoder to generate synthetic speech that is similar to the original. [23] A standard that utilizes simple LPC encoding is U.S. Federal Standard 1015 which requires a bandwidth of 2.4 kbps. Also, GSM (Groupe Speciale Mobile) encoding uses a variation of LPC called RPE-LPC (Regular Pulse Excited - Linear Predictive Coder with a Long Term Predictor Loop). GSM began as a European cellular phone speech encoding standard. GSM compresses 160 13-bit samples (2080 bits) to 260 bits which is an 8:1 compression ratio. For 8 kHz sampling, this means GSM encoded speech requires a bandwidth of 13 kbps. [23]

A CELP (Code Excited Linear Prediction) encoder does the same vocal tract modeling as an LPC encoder. In addition, it computes the error between the input speech data and the model and transmits the model parameters and a representation of the errors. The errors are represented as indices into a common code book shared between encoders and decoders. This is where the name "Code Excited" comes from. The extra data and computations produce a higher quality encoding than simple LPC encoding. [23] A standard that utilizes simple CELP encoding is U.S. Federal Standard 1016 which requires a bandwidth of 4.8 kbps. Also, ITU-T Recommendation G.728, which is one of the audio encoding formats specified by H.320 (discussed in section 2.4.3.1), uses a variation of CELP, LD-CELP (Low Delay CELP). G.728 requires a bandwidth of 16 kbps and is quite computationally complex, requiring special hardware.

## **2.2 Video**

Video is a sequence of still images. When presented at a high enough rate, the sequence of images (frames) gives the illusion of fluid motion. For instance, in the United States, movies are presented at 24 frames per second (fps) and television is presented at 30 fps.

Desktop videoconferencing uses video as an input. This video may come from a camera, VCR, or other video device. An analog video signal must be encoded in digital form so that it can be manipulated by a computer. To understand digital encoding, it helps to understand some background information about analog video, including basic color theory and analog encoding formats.

### **2.2.1 Color theory**

The human eye has three types of color photoreceptor cells called cones. Because of this, three numerical components are necessary and sufficient to represent a color. [19] Color spaces are three dimensional coordinate systems whose axes correspond to three color components. Different color spaces are useful for different purposes and transformations translate data from one color space to

another.

The color encoding systems used for video are derived from the RGB color space. RGB is an additive space which uses combinations of Red, Green, and Blue primaries. The RGB system is transformed to other systems that allow video encoding techniques to exploit the characteristics of human color perception.

Brightness and color information are treated differently by the human visual system. Humans are more sensitive to changes in brightness than changes in color. Because of this, a special component is used to represent brightness information. This component is called luminance and is denoted by the symbol Y. In video encoding, the nonlinear version of luminance termed luma is used and denoted by the symbol Y' (the prime symbol meaning nonlinear). The remaining two components are used to represent color and are called chrominance. These chrominance components are called color differences and are the Blue and Red components with luma removed, (B'-Y') and (R'-Y'). Therefore R', G', B' space is transformed to Y', (B'-Y'), (R'-Y') space. The matrix version of this transform is shown in Equation 2.1. [19] The values of R', G', and B' range from 0 to 1.

$$\begin{bmatrix} Y' \\ B'-Y' \\ R'-Y' \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.299 & -0.587 & 0.886 \\ 0.701 & -0.587 & 0.114 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} \quad (2.1)$$

With the color separated in this way, the color differences (B'-Y') and (R'-Y') can be subsampled with no visible results. This allows the same visual information to be encoded in less bandwidth.

Various different color systems have been defined (Y'PbPr, Y'CbCr, PhotoYCC). All these systems are derivatives of Equation 2.1 with different scaling factors. Y'PbPr is used for component analog video. Y'CbCr is used for component digital video and in digital encoding/compression schemes such as JPEG and MPEG. PhotoYCC is used in Kodak's Photo CD format. [19] The notation YUV is often used generically to refer to a color space represented by luminance and two color differences.

### 2.2.2 Video formats

There are two widely used formats for analog video: NTSC and PAL. Each format is described briefly below.

The previous section discussed how video signals are encoded in three separate components. However, NTSC and PAL are composite formats which combine these three components into one signal. First the two color difference components are combined into a single chroma signal using a technique called quadrature modulation, and then the luma and chroma are combined using a technique called frequency interleaving. [18]

NTSC (National Television Standards Committee) format is used in the Americas and Japan. This standard was approved in 1953 by the Federal Communications Commission (FCC) for commercial broadcasting. NTSC format has a resolution of 525 lines per frame and 60 (really 59.94) interlaced frames per second (60Hz). With interlacing, two fields make up a complete a frame. Consecutive fields contain even and odd lines of the frame, resulting in 30 frames per second. However, not all 525 lines are visible. Only 483 lines contain the video information. [14]

PAL (Phase Alteration Line) format is used in Western Europe and Australia. PAL format has a resolution of 625 lines per frame and 50 interlaced frames per second, resulting in 25 frames per second.

There exists a third video format used in France, Russia and Eastern Europe called SECAM (SEquential Couleur A Memoire, meaning sequential color with memory) It has the same resolution as PAL, but the video information is encoded differently. The author has not seen this format associated with desktop videoconferencing.

### 2.2.3 Video delivery

Video is commonly delivered in composite NTSC or PAL format through phono or BNC (Bayonet, Non-Continuous) connectors. Either format can also be delivered by an S-Video (Y/C) connector. S-video (S as in separate) delivers the luma and chroma components separately. The Y and C signals, if summed together, will form a legal NTSC or PAL signal. [18] S-Video provides a sharper image with better color separation. A S-Video connector is shown in Figure 2.1.

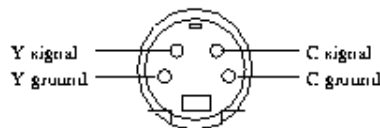


Figure 2.1 S-video connector

### 2.2.4 Digital video compression techniques

Analog video is digitized so that it may be manipulated by a computer. Each frame of video becomes a two dimensional array of pixels. A complete color image is composed of three image frames, one for each color component.

Uncompressed images and video are much too large to deal with and compression is needed for storage and transmission. Important metrics of compression are the compression ratio and bits per pixel (the number of bits required to represent one pixel in the image).

Video compression is typically lossy, meaning some of the information is lost during the compression step. This is acceptable though, because encoding algorithms are designed to discard information that is not perceptible to humans or information that is redundant. There are some basic techniques common to most video compression algorithms, including color space sampling and redundancy reduction.

Color space sampling is an effective technique used to reduce the amount of data that needs to be encoded. If an image is encoded in YUV space, the U and V components can be subsampled because the the human eye is less sensitive to chrominance information.

Redundancy reduction is another technique used decrease the amount of encoded information. *Intraframe* encoding achieves compression by reducing the spatial redundancy within a picture. This technique works because neighboring pixels in an image are usually similar. *Interframe* encoding achieves compression by reducing the temporal redundancy between pictures. This technique works because neighboring frames in a sequence of images are usually similar.

Some important video encoding and compression techniques related to desktop videoconferencing are discussed below.



### 2.2.4.1 MJPEG

JPEG is an encoding standard for still images developed by the Joint Photographic Experts Group. Although designed for still images, with special hardware it is possible to encode and decode a series of JPEG images in real-time to achieve motion video. This use of JPEG encoding is typically referred to as Motion JPEG or MJPEG. However, no official MJPEG standard exists.

There are four defined modes of operation for JPEG: sequential, progressive, lossless, and hierarchical. Typically only sequential mode is implemented and therefore this is the only mode discussed in this paper.

JPEG utilizes an intraframe spatial compression technique, Discrete Cosine Transform (DCT) encoding. This technique is used by other encoding methods such as H.261 (discussed in section 2.2.4.2). The basic steps of DCT-based encoding are shown in Figure 2.2 and discussed below. This discussion of JPEG assumes the encoding of a single component image (grayscale). For multiple component color images, the component information is interleaved. JPEG encoding is color space independent, though systems may choose to convert images to YUV space and subsample the chrominance components. [1]

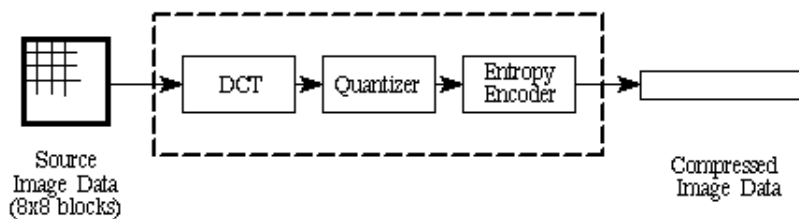


Figure 2.2 DCT-based encoding

The first step of the encoding process is to perform a DCT on 8x8 blocks of component samples. This step transforms the information into the frequency domain. The output of this step is 64 DCT coefficients. The DCT has the effect of concentrating most of the information in the 8x8 block into the upper left hand corner. The average value of the block, called the DC component, is the upper left hand coefficient. The remaining coefficients are called AC coefficients. No information is lost during the DCT step. The data is merely transformed to another domain and can be recovered by performing an inverse DCT.

The next step of the encoding process is quantization. The DCT coefficients are divided by an 8x8 quantization matrix. This matrix is designed to reduce the amplitude of the coefficients and to increase the number of zero value coefficients. [1] The quantization step is the lossy step of the encoding process.

After quantization, a bit stream is formed from the block. The DC coefficient is encoded as the difference between the current DC coefficient and the DC coefficient of the previous block. The AC coefficients are encoded in a zig-zag sequence from the upper left of the block to the bottom right. Figure 2.3 illustrates an 8x8 block of coefficients and the zig-zag sequence. The AC coefficients are run-length encoded and entropy encoded. The run-length encoding removes long runs of zero valued coefficients. The entropy encoding (Huffman encoding or optionally arithmetic encoding) encodes the information efficiently based on statistical characteristics. Patterns that are statistically more likely to occur are encoded with shorter code words.

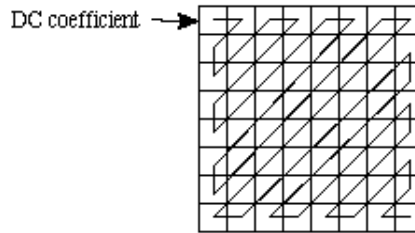


Figure 2.3 Block coefficients and the zig-zag sequence

The decoding process is basically the reverse of the encoding process. The decoder must have access to the same quantization and entropy tables that were used to encode the data.

JPEG encoding typically achieves compression on the order of 10:1 to 20:1 [1] Higher compression results in poorer image quality. A user-configurable quality parameter is usually available that allows a compression vs. quality tradeoff. This parameter usually causes a scaling of the quantization matrix.

#### 2.2.4.2 ITU-T Recommendation H.261

ITU-T Recommendation H.261 is a video compression standard designed for communication bandwidths between 64 kbps and 2 Mbps, measured in 64 kbps intervals. This technique is also referred to as "px64" where "p" ranges from 1 to 30. H.261 was designed primarily for videoconferencing over ISDN and is specified by H.320 (discussed in section 2.4.3.1).

H.261 utilizes both intraframe spatial and interframe temporal encoding. In intraframe encoding mode, DCT-based spatial compression is used. In interframe encoding mode, motion compensation is performed to compute the differences between frames. The differences, usually of small magnitude, are then DCT encoded. [1]

Two picture formats, CIF (Common Intermediate Format) and QCIF (Quarter CIF) are defined. QCIF operation is mandatory, while CIF operation is optional. QCIF is usually used for low bit rates, such as  $p < 3$ . Images are composed of three color components, Y and two color differences Cb and Cr (YCbCr corresponds to a transform of YUV space). The color difference components contain half the amount of information as the luminance component (for each 4 blocks of luminance information encoded, only 2 block of chrominance information is encoded.) Table 2.1 shows the lines per frame and pixels per line for CIF and QCIF.

Intraframe encoding works essentially as with JPEG. 8x8 blocks are DCT transformed, quantized and run-length/entropy encoded. In interframe encoding mode, a prediction for blocks in the current frame is made based on the the previous frame. If the difference between the current block and the predicted block is below a certain threshold then no data is sent. Otherwise the difference is calculated and DCT transformed, quantized and run-length/entropy encoded.

	CIF lines/frame	CIF pixels/line	QCIF lines/frame	QCIF pixels/line
<b>Luminance (Y)</b>	288	352	144	176
<b>Chrominance (Cb)</b>	144	176	72	88
<b>Chrominance (Cr)</b>	144	176	72	88

Table 2.1 CIF and QCIF image formats

The quantizing step determines the amount of information that is sent, more information meaning better image quality. H.261 encoders adjust the quantizer value to achieve a constant bit rate. [12] If the transmission buffer is close to full, the quantizer step size will be increased, causing less information to be encoded and poorer image quality. Similarly, when the buffer is not full, the quantizer step size is decreased, causing more information to be encoded and better image quality. Because of this quantizer adjustment, rapidly changing scenes will have poorer quality than static scenes.

### 2.2.4.3 CellB

Cell is a non-proprietary compression technique developed by Sun Microsystems. [15] There are two types of Cell compression, CellA and CellB. CellA is an asymmetric technique, requiring more computation for compression than decompression. CellB is similar to CellA but is more computationally symmetric, making it more suited to real-time use such as for videoconferencing.

CellA takes RGB as input while CellB takes YUV as input. Each requires that the image width and height be divisible by four. The input image is divided into 4x4 groups of pixels called cells. The Cell encoding technique is based on a method called Block Truncation Coding (BTC). The 16 pixels in each cell are encoded by a 16 bit mask and two 8 bit intensities. CellA and CellB differ in the way the intensities are chosen. In CellA, the values are indices into a colormap. In CellB, the values are indices into vector quantization tables. Default tables are part of the CellB specification.

Figure 2.4 shows a CellB cell and the cell code that is formed after encoding. The U/V field represents the chrominance component. The Y/Y field represents two luminance values. If a pixel's bit is set in the bit mask, the pixel is rendered as <Y1,U,V> and if the bit is not set the pixel is rendered as <Y2,U,V>. This means that for each encoded cell there are two possible luminance values and one possible chrominance value. CellB encodes 16 pixels (384 bits) using 32 bits. This is a 12:1 compression ratio.

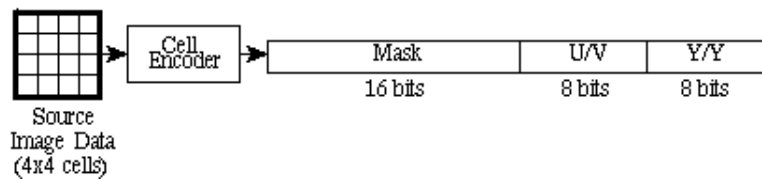


Figure 2.4 CellB encoding

CellB is basically an intraframe encoding technique. However it does support a simple form of interframe coding by defining a skip code which indicates that a certain number of cells should be skipped, meaning their value has not changed.

An advantage of Cell encoding is that the decoding process is similar to character fonting in a color framebuffer. The character display process for a framebuffer takes as input a foreground color, background color and mask indicating which color to use at the pixel. Since this is a very important function of a windowing system, it is often implemented as a display primitive in graphics accelerators. Therefore, Cell encoding can take advantage of existing hardware primitives to achieve efficient decoding. [15]

### 2.2.4.4 nv

Network Video (nv) is an Internet videoconferencing tool developed at Xerox/PARC. It is the most commonly used video tool on the Internet Mbone (discussed in section 2.4.2.6). The native nv encoding technique utilizes spatial (intraframe) and temporal (interframe) compression. The first step of the nv algorithm compares the current frame to the previous frame and marks the areas that have changed significantly. Each area that has changed is compressed using transform encoding. Either a DCT or a Haar wavelet transform is used. The nv encoder dynamically selects which transform is used based on whether network bandwidth (use DCT) or local computation (use Haar) is limiting performance. The DCT is desired since it almost doubles the compression ratio. The output of the transform is quantized and run-length encoded. Periodically, unchanged parts of the image are sent at higher resolution, which is achieved by eliminating the quantization step. Typically, nv can achieve compression ratios of 20:1 or more. [7]

#### **2.2.4.5 CU-SeeMe**

CU-SeeMe is an Internet videoconferencing tool developed at Cornell University. It utilizes spatial (intraframe) and temporal (interframe) compression, with a few twists to optimize performance on a Macintosh, its original platform. CU-SeeMe represents video input in 16 shades of grey using 4 bits per pixel. The image is divided into 8x8 blocks of pixels for analysis. New frames are compared to previous frames, and if a block has changed significantly it is retransmitted. Blocks are also retransmitted on a periodic basis to account for losses that may have occurred in the network. Transmitted data is compressed by a lossless algorithm developed at Cornell that exploits spatial redundancy in the vertical direction. The compressed size is about 60% of the original (a 1.7:1 compression ratio). The CU-SeeMe encoding algorithm was designed to run efficiently on a Macintosh computer, and operates on rows of 8 4-bit pixels as 32-bit words, which works well in 680x0 assembly code. The default transmitting bandwidth setting for CU-SeeMe is 80 kbps. This number is automatically adjusted on the basis of packet-loss reports returned by each person receiving the video. About 100 kbps is required for fluid motion in a typical "talking heads" scenerio. [20]

#### **2.2.4.6 Indeo**

Indeo is a video compression technique designed by Intel. It evolved from DVI (Digital Video Interactive) technology. Indeo starts off with YUV input, U and V subsampled 4:1 both horizontally and vertically. Indeo supports motion estimation, using the previous frame to predict values for the current frame and only transmitting data if the difference is significant. Transform encoding is done using a 8x8 Fast Slant Transform (FST) in which all operations are either shifts or adds (no multiplies). Quantization and run-length/entropy encoding are used as in previous algorithms. Indeo specifies that the encoded bit stream be a maximum of 60% of the input data, therefore compression is guaranteed to be at worst 1.7:1.

### **2.3 Communication channels**

There are different types of communication channels available to transport desktop videoconferencing data. These channels can be classified as either circuit-switched or packet-switched. Each type of channel has advantages and disadvantages when evaluating its suitability for the transport of data characteristic for desktop videoconferencing.

Different types of data have different service requirements. Some data types are sensitive to delay, while other types are sensitive to reliability. Table 2.2 outlines different data types and their sensitivity to delay and reliability. "Yes" indicates that the data type is very sensitive and can not tolerate much deviation. "No" indicates that some deviation is tolerable.

	<b>Data</b>	<b>Voice</b>	<b>Image</b>	<b>Video</b>
<b>Delay Sensitive</b>	No	Yes	No	Yes
<b>Reliability Sensitive</b>	Yes	No	Yes	Yes/No

Table 2.2 Service requirements for various data types

Generic data is not sensitive to delay but is sensitive to reliability. An example is a data file that is sent over a network. It does not matter how long the file takes to get to its destination, but the information in the file is expected to be correct. Voice data is sensitive to delay but is not sensitive to reliability. Voice data must arrive at a constant rate with little variance for it to be intelligible, but some loss of information is acceptable. Still image data is not sensitive to delay but is sensitive to reliability. Incorrect image data may be noticeable in the form of visual artifacts, but delivery time is not crucial. Video data is sensitive to delay and large delays will be obvious by a jerky picture. Uncompressed video data is not sensitive to reliability since if one frame is lost it will immediately be replaced by another frame. However, compressed video, which uses intraframe and interframe encoding techniques, is sensitive to reliability since redundancy has been removed and the effects of data loss may propagate. This is an important thing to consider when sending video data across unreliable communication channels. Some video compression techniques compensate for this sensitivity to data loss by periodically sending complete information about a frame. For example both nv and CU-SeeMe, which were designed to operate over the unreliable Internet, have a refresh mechanism to periodically send blocks of data even if the data in these blocks has not changed. This allows receivers to recover from stale blocks that may result from packet loss. It also allows receivers that have joined an in-progress session to receive a complete picture.

Desktop videoconferencing can involve all the data types discussed above. Audio and video data is sent among participants. Other types of data that may be sent are whiteboard data or shared application data. Some of this data requires reliable transmission while some requires timely transmission.

### **2.3.1 Circuit-switched communication channels**

Circuit-switched communication is a method of data transfer where a path of communication is established and kept open for the duration of the session. A dedicated amount of bandwidth is allocated for the exclusive use by the session. When the session is completed, the bandwidth is freed and becomes available for other sessions.

Advantages of circuit-switched communication for desktop videoconferencing are that dedicated bandwidth is available and the timing of the data delivery is predictable. A disadvantage of circuit-switched communication for desktop videoconferencing is that sessions are primarily point-to-point and require expensive multi-conferencing units (MCUs) to accommodate multipoint conferences. Also, dedicated bandwidth is wasteful during periods of limited activity in a conference session.

### **2.3.2 Packet-switched communication channels**

Packet-switched communication is a method of data transfer where the information is divided into packets, each of which has an identification and destination address. Packets are sent individually through a network and, depending on network conditions, may take different routes and arrive at their destination at different times and out-of-order. No dedicated bandwidth circuit is set up as with circuit-switched communication. Bandwidth must be shared with whomever else is on the network.

An advantage of packet-switched communication for desktop videoconferencing is the capability to more easily accommodate multipoint conferences. A disadvantage is the unpredictable timing of data delivery, which can cause problems for delay sensitive data types such as voice and video. Video packets received out-of-order may have to be discarded. Audio packets can be buffered at the receiver, re-ordered, and played out at a constant rate, however this induces a delay which can be detrimental to interactive communication.

### 2.3.3 Broadband ISDN

Broadband ISDN (BISDN) has the potential to solve the problems encountered with circuit-switched and packet-switched communication. Asynchronous Transfer Mode (ATM) is the data link layer protocol that is commonly associated with BISDN. ATM combines the best qualities of circuit-switched and packet-switched communication. ATM can support different data transmission speeds, multiplex signals of different data types, and provide different classes of service. [14] These capabilities will satisfy the service requirements of the different types of data possible with desktop videoconferencing. BISDN and ATM show great promise for the future, but their deployment at this time is limited.

### 2.4 Desktop videoconferencing systems

The author has compiled a survey of desktop videoconferencing information. This survey has been available on the World Wide Web (WWW) for approximately one year at the URL <[http://www2.ncsu.edu/eos/service/ece/project/succeed\\_info/dtvc\\_survey/](http://www2.ncsu.edu/eos/service/ece/project/succeed_info/dtvc_survey/)>.

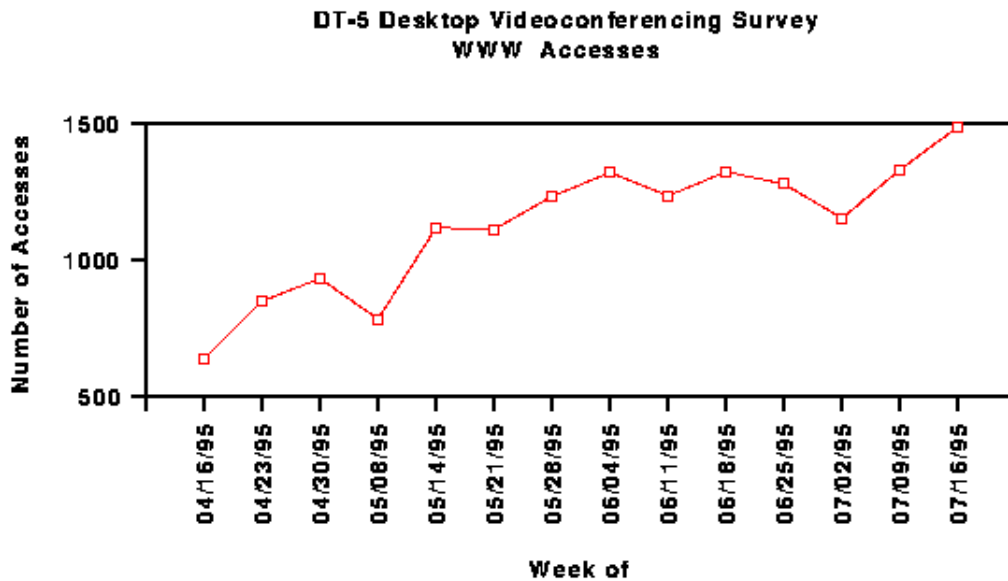


Figure 2.5 WWW accesses to desktop videoconferencing product survey

The information in this survey is rapidly evolving due to the explosive market of desktop videoconferencing products. The list has become quite popular on the Internet. Figure 2.5 shows the number of accesses to this information plotted by week for the months of April through July, 1995. The survey was cited in the January 1995 issue of *Computer* magazine as a "URL that provides a good review of desktop videoconferencing tools." [24] The survey was also cited in an answer in the "Multimedia Q&A" section of the July 1995 issue of *Multimedia World* magazine.

### **2.4.1 General features**

There are three major platforms for desktop videoconferencing products: Intel-based personal computers running Microsoft Windows or IBM OS/2, Apple Macintosh computers, and Unix-based workstations running the X Window System. Unfortunately, there is currently very little interoperability among products and platforms. Products are evolving towards conformance to the emerging desktop videoconferencing interoperability standards (discussed in section 2.4.3).

All systems require hardware that captures and digitizes the audio and video. Video is typically input in NTSC or PAL formats. Most systems have some sort of graphical user interface that assists in making connections to other parties, usually utilizing the paradigm of "placing a telephone call." Many products allow you to store information about other parties in a phone book or Rolodex format. Systems commonly have controls to adjust audio volume, picture contrast, etc. Many systems have controls that allow you to adjust the transmitted bandwidth for video to minimize traffic on a network.

An additional feature found in most systems is a shared drawing area usually called a whiteboard which is analogous to the whiteboards found in many conference and classrooms. These whiteboards commonly allow participants to import other graphics such as images and to make annotations. Whiteboards are good for simple sketches, but fine detail is difficult to achieve using a mouse. Many systems allow an easy way to transfer files between participants. Some systems allow application sharing, which enables a participant to take control of an application running on another participant's computer. The usefulness of application sharing is often demonstrated with an example of sharing a spreadsheet or word processor program to facilitate group collaboration.

### **2.4.2 Modes of conferencing**

Desktop videoconferencing systems communicate in a variety of different ways. These ways are described in the following sections.

#### **2.4.2.1 POTS conferencing**

POTS (Plain Old Telephone Service) is the basic telephone service that provides access to the public switched telephone network (PSTN). This service is widely available but has very low bandwidth (typical modem speeds are 9.6 kbps, 14.4 kbps or 28.8 kbps). There are very few desktop videoconferencing products that attempt to operate at these rates. Work is in progress on H.324, an interoperability standard for videoconferencing products operating at a rate of 28.8 kbps (discussed in section 2.4.3.4).

#### **2.4.2.2 Switched 56 conferencing**

Switched 56 is a digital service that delivers 56 kbps of bandwidth. It is similar to ISDN except that its channels are of a lower bandwidth. Switched 56 is an older technology and is being replaced by ISDN. Typically desktop videoconferencing products that operate over switched 56 can also operate over ISDN.

#### **2.4.2.3 ISDN conferencing**

ISDN (Integrated Services Digital Network) is a digital service. There are two access rates defined for ISDN, Basic Rate Interface (BRI) and Primary Rate Interface (PRI). Basic Rate Interface provides 2 data channels of 64 kbps (B-channels) and one signaling channel of 16 kbps (D-channel). There are many desktop videoconferencing products on the market that utilize ISDN BRI. However, problems exist with access to ISDN because it is not available in all areas. Even when it is available, it can be

non-trivial to configure correctly. Primary Rate Interface provides 23 or 30 B channels of 64 kbps and one D channel of 64 kbps. ISDN PRI is expensive and therefore not really applicable for desktop videoconferencing.

Because ISDN channels offer 64 kbps of bandwidth, standards and compression algorithms have been designed around that number. 64 kbps has become somewhat of a magic number for videoconferencing.

#### **2.4.2.4 LAN conferencing**

Local Area Networks (LANs) are common on campuses and in companies to connect desktop computers together. At the physical layer, LANs usually consist of 10 Mbps Ethernet, or 4 or 16 Mbps Token Ring segments. The difference between Ethernet and Token Ring is how clients gain access to the medium for transmission. Ethernet is a Carrier Sense Multiple Access with Collision Detection (CSMA/CD) network where clients transmit data and listen to detect collisions. If a collision occurs, the client waits a random amount of time before trying to transmit again. Token Ring is a network where a token is passed around and clients must gain access to the token before transmitting.

Desktop videoconferencing products support a variety of network protocols such as TCP/IP, Novell IPX/SPX, NetBIOS, and Appletalk.

#### **2.4.2.5 Internet conferencing**

LANs provide connectivity among a local community. The Internet connects LANs to other LANs. The protocol developed to interconnect various networks is called the Internet Protocol (IP). Two transport layer protocols were developed with IP, TCP and UDP. TCP (Transmission Control Protocol) provides a reliable end-to-end service by using error recovery and reordering. UDP (user datagram protocol) is an unreliable service making no attempt at error recovery. [2]

Desktop videoconferencing applications that operate over the Internet primarily use UDP for video and audio data transmission. TCP is not practical because of its error recovery mechanism. If lost packets were retransmitted, they would arrive too late to be of any use. TCP is used by some videoconferencing applications for other data that is not time sensitive such as whiteboard data and shared application data.

#### **2.4.2.6 Multicast Backbone (MBone) conferencing**

The Multicast Backbone, or MBone, has been called a virtual network because it is layered on parts of the Internet. [3] Using the MBone, it is possible to transmit audio, video and other data in real-time to multiple destinations throughout the global Internet. This section gives a technical background on the MBone. More information about the MBone can be found in Chapter 4 which describes in depth a seminar class that was broadcast using the MBone.

The MBone originated in 1992 from an experiment to transmit live audio and video from meetings of the Internet Engineering Task Force (IETF). The name was inspired by the name of the European backbone network "EBONE." Recent estimates describe the size of the MBone to be approximately 20,000 users on 1500 networks in 30 countries [11].

To understand how the MBone works, it is important to understand the difference between unicast and multicast. Unicast is a point-to-point transmission of data. To achieve a one-to-many transmission, separate copies of the data must be sent by the source to each destination. Multicast enables a more efficient way to deliver the same data to multiple destinations. Figure 2.6 shows an example of a source transmitting to three destinations.



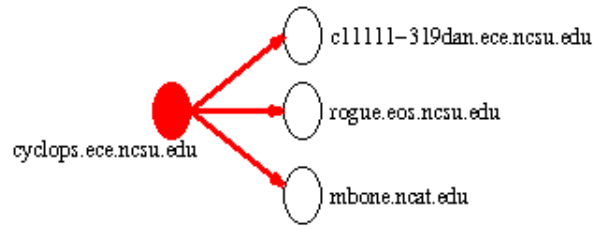
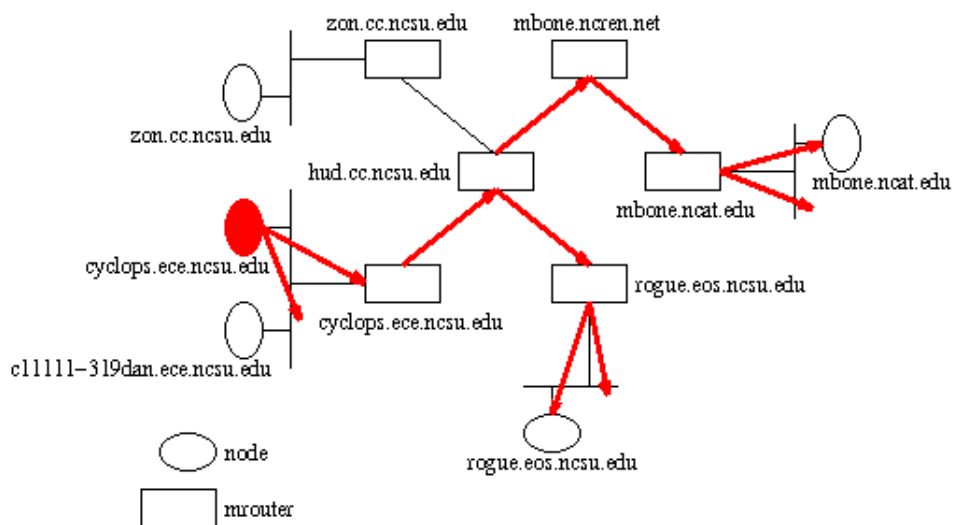


Figure 2.6 Multicast delivery: The illusion

Multicast has been implemented on local area networks such as Ethernet and Fiber Distributed Data Interface (FDDI) for some time. [13] In 1989, Steve Deering specified an extension to the Internet Protocol to support multicasting known as RFC (Request For Comment) 1112. [5] This extension allowed for the concept of multicast to be extended to include the entire Internet. Basically, the Mbone is composed of "islands" supporting IP multicast that are connected by "tunnels" of point-to-point links. [3]

With IP multicast, data is transmitted to a "host group." [5] Groups are specified by a class D IP address in the range of 224.0.0.0 to 239.255.255.255. Hosts that wish to receive the data join the host group. It is then the responsibility of the multicast routers to efficiently deliver the sender's data to all receivers that have joined the group. The Internet Group Management Protocol (IGMP), is used by multicast routers to determine what groups are active on a particular subnet. [5] There are several different routing protocols that multicast routers can use to construct efficient delivery trees from sender to receivers [4], including DVMRP (Distance Vector Multicast Routing Protocol), MOSPF (Multicast Open Shortest Path First) and PIM (Protocol Independent Multicast). The problem is that these routing protocols have not traditionally been available on production routers. The workaround for this problem is "tunneling." A multicast router, typically a Unix workstation running a routing daemon, encapsulates the multicast packet inside a regular IP packet and sets the destination to another multicast router. To intervening routers and subnets, the data looks like a normal unicast packet. [3] Figure 2.7 gives a more realistic view of what actually happens during multicast routing.



## Figure 2.7 Multicast delivery: The reality

Most newer versions of Unix support multicast; older versions require some modifications to the kernel. Software (usually in source and binary form) is available free via ftp (file transfer protocol). Practical bandwidth guidelines have been set for the MBone. Some links could be saturated if a large amount of MBone traffic flowed. These guidelines are a maximum of 500 kbps and 250 packets per second of MBone traffic. Scheduling of broadcasts is coordinated through the rem-conf@es.net mailing list. The MBone is still an experimental technology, but this is starting to change. Most major router vendors now support IP multicast. [11]

The challenges of transmitting audio and video over the Internet has led to the development of a new transport protocol proposed by the Audio/Video Transport working group of the IETF (IETF-AVT working group). RTP (Real-time Transport Protocol) provides support for sequencing, timing, and quality of service reporting for point to point or multipoint. Most of the commonly used MBone tools implement some version of RTP as do some commercially available tools such as Communique! (InSoft), InPerson (Silicon Graphics), and ShowMe (Sun Microsystems).

### 2.4.3 Interoperability standards

Interoperability means that products from different vendors can communicate. To accomplish this goal, standards are required. There are several standards groups working towards producing and promoting standards for desktop videoconferencing.

The ITU (International Telecommunication Union) is an agency of the United Nations. It is a world-wide organization within which governments and private companies coordinate the establishment and operation of telecommunication networks and services. The ITU-T is the Telecommunication Standardization Sector of the ITU and has developed standards for audio, video, videoconferencing and data conferencing, primarily over ISDN. The ITU-T is working cooperatively with the IETF to extend its videoconferencing standards to include packet-switched networks, but no formal standards have been adopted yet.

The IMTC (International Multimedia Teleconferencing Consortium) is a non-profit corporation founded to promote the creation and adoption of international standards for multipoint videoconferencing and document conferencing. IMTC's emphasis is on multimedia teleconferencing, including still-image graphics, full motion video and data teleconferencing. The IMTC basically promotes the standards adopted by the ITU including H.320 (discussed in section 2.4.3.1) and T.120 (discussed in section 2.4.3.2).

The PCWG (Personal Conferencing Working Group) is another group promoting standards and interoperability in the videoconferencing industry. The IMTC and PCWG share many of the same members and many of the same goals. The PCWG is developing the Personal Conferencing Standard (PCS) (discussed in section 2.4.3.3) which will be compatible and based on the ITU-T standards.

#### 2.4.3.1 ITU-T Recommendation H.320

ITU-T Recommendation H.320 [9] is titled "Narrow-Band Visual Telephone Systems and Terminal Equipment." This recommendation was approved in March of 1993. Narrow-band bit rates are defined to range from 64 kbps to 1920 kbps (64 kbps x 30). This recommendation was designed primarily for ISDN. A typical ISDN Basic Rate Interface that might be available for desktop videoconferencing is two 64 kbps channels (p=2) or 128 kbps total bandwidth.

H.320 is a series of recommendations. These recommendations and their titles are shown in Table 2.3. H.261 video requires bandwidths in multiples of 64 kbps. There are several bandwidths defined for audio. G.711 operates at 64 kbps which is one entire ISDN B channel. This leaves only 64 kbps for video and other data, assuming ISDN BRI. G.722 operates at 48/56/64 kbps. G.728 operates at 16 kbps. G.711 and G.728 both support frequencies of 3 kHz. G.722 supports frequencies of 7 kHz. H.221, H.230 and H.242 specify how calls are negotiated, how the data is multiplexed and framed, etc.

Video codec	<b>H.261:</b> Video codec for audiovisual services at p x 64 kbit/s
Audio codec	<b>G.711:</b> Pulse code modulation (PCM) of voice frequencies <b>G.722:</b> 7 kHz audio-coding within 64 kbit/s <b>G.728:</b> Coding of speech at 16 kbit/s using low-delay code excited linear prediction
Frame structure	<b>H.221:</b> Frame structure for a 64 to 1920 kbit/s channel in audiovisual teleservices
Control and indication	<b>H.230:</b> Frame-synchronous control and indication signals for audiovisual systems
Communication procedure	<b>H.242:</b> System for establishing communication between audiovisual terminals using digital channels up to 2 Mbit/s

Table 2.3 H.320 series of recommendations

#### 2.4.3.2 ITU-T Recommendation T.120

ITU-T Recommendation T.120 is titled "Transmission Protocols For Multimedia Data." This recommendation defines multipoint transport of multimedia data. T.120 enables participants to share data during a conference. This data could be whiteboard data or another type of data such as a binary file.

T.120 is a series of recommendations. These recommendations and their titles are shown in Table 2.4. T.123 defines connections for various types of networks (including POTS, ISDN, and LANs). T.122 and T.125 define the multipoint communication. T.124 provides the mechanism to set up and manage conferences. T.127 defines file exchange. T.126 defines viewing and annotating shared images and application sharing. Some of these recommendations have been formally approved by the ITU-T while others are in the draft stage and expected to be approved soon.

<b>T.121</b>	Generic Application Template (T.GAT)
<b>T.122</b>	Multipoint Communication Service for Audiographics Conferencing: Service Definition
<b>T.123</b>	Protocol Stacks for Audiographic and Audiovisual Teleconference Applications
<b>T.124</b>	Generic Conference Control
<b>T.125</b>	Multipoint Communication Service Protocol Specification
<b>T.126</b>	Multipoint Still Image and Annotation Protocol
<b>T.127</b>	Multipoint Binary File Transfer
<b>T.128</b>	Audio Visual Control for Multipoint Multimedia Systems

Table 2.4 T.120 series of recommendations

### 2.4.3.3 PCS

The Personal Conferencing Specification (PCS) is being developed cooperatively by members of the Personal Conferencing Working Group (PCWG). PCS defines the infrastructure to provide interoperable point-to-point and multipoint communications of audio, video, and data conferencing.

PCS 1.0 specified Indeo video compression and GSM audio compression. Future versions of PCS will be designed to be interoperable with H.320 and T.120.

### 2.4.3.4 ITU-T Recommendation H.324

ITU-T Recommendation H.324 is titled "Multimedia terminal for low bitrate visual telephone services over the GSTN." The H.324 series of recommendations (formerly called H.32P) define real-time audio, video and data transfer over V.34 modems on the GSTN (Global Standard Telephone Network). Work on these documents is still in progress. The various documents will be up for decision in November 1995 and, if decided, will go to ballot and eventually become ITU-T recommendations. The H.324 recommendations and their titles are shown in Table 2.5.

Video codec	<b>H.263:</b> Video coding at rates less than 64 kbit/s
Audio codec	<b>G.723:</b> Speech coder for multimedia telecommunications transmitting at 5.3/6.3 kbit/s
Control	<b>H.245:</b> Multimedia system control
Multiplex	<b>H.223:</b> Multiplexing protocol for low bitrate multimedia terminals

Table 2.5 H.324 series of recommendations

A V.34 modem has a total bandwidth of 28.8 kbps. G.723 has two audio bandwidth modes: 5.3 kbps and 6.4 kbps. This leaves either 23.5 kbps or 22.4 kbps for video and overhead. G.723 also has a silence suppression mode so that the audio bandwidth can be used for other data when no audio is being transmitted.



[TABLE OF CONTENTS]

### **3. HUMAN COMPUTER INTERACTION AND SOCIAL ISSUES**

When evaluating the use of desktop videoconferencing, there are important non-technical aspects to consider. Specifically, systems must support the way people work or they will not be successful. [8]

#### **3.1 Lecture mode versus collaboration mode**

There are two distinct paradigms to consider with desktop videoconferencing: lecture mode and collaboration mode. These modes differ by the type and amount of interaction that takes place among the participants. It is important to understand the difference between these modes because each require videoconferencing systems to support different types of interactions.

Lecture mode is typically a one-to-many interaction. There are distinct and unequal roles of the participants. There is typically one lecturer and multiple students. The lecturer is in control of the conference. The lecturer may ask for interaction from the students in the form of questions or discussion. Students may indicate their desire for interaction by raising their hand. Lecture mode utilizes asymmetric communication among the participants.

Collaboration mode is typically a many-to-many interaction. Each participant is a peer who participates in the conference equally, although in some cases there may be a facilitator who manages the conference and keeps the agenda flowing smoothly. Collaboration mode utilizes symmetric communication among the participants.

Users will be frustrated if the system does not support the mode of interaction they desire. For instance, in lecture mode it may be difficult for students to indicate they wish to ask a question if the lecturer can not see the student or if the interface does not provide a way to "raise a hand."

#### **3.2 Benefits of moving to the desktop model of videoconferencing**

Room videoconferencing systems typically offer two way real-time audio and video. In addition, they usually have the capability to send high quality still images to remote sites. However, surveys of room videoconferencing system users have identified additional desired features such as a shared drawing area, ability to connect multiple sites, and ways to incorporate computer applications into the conference. [21] These types of features can be provided with desktop videoconferencing systems. Perhaps the most important aspect of desktop videoconferencing is not that it is on the desktop but that it is integrated into the computing environment that the user is already familiar with. This opens up the possibility for data conferencing as well as videoconferencing.

Room videoconferencing systems have scheduling and booking problems. Time slots sometimes have to be booked well in advance. With desktop systems, more impromptu and informal interaction can take place. Users will be more likely to use a system if they have easy access to it. However, this can be a disadvantage since on the desktop there are likely to be more distractions than in a conference room setting (for example, incoming email, phone calls, etc.)

#### **3.3 The value of audio**

The importance of good quality audio in a conference can not be overstated. Since not many of us can read lips, effective communication can not occur without intelligible audio. Audio delay can make interaction difficult. Audio that is not synchronized with video can be distracting. However, some studies suggest that users prefer having audio with minimal delay over having audio in sync with video if a noticeable delay is imposed [21].

### **3.4 The value of video**

Intuitively, it seems that video adds value to a conference. Video enhances communication by creating a sense of presence. Video allows for communication through gesturing. Objects can be shown to other participants by holding them up in front of the camera. Video from auxiliary sources (such as a VCR) can be included in a conference. Video allows for interpretation of what is going on in the environment of other participants. For example, long pauses may not be perceived to be unusual if video information gives some indication to the meaning of the pause (for instance, the other person is looking for a particular slide in a stack of papers). [21] One of the author's personal observations is that it seems easier to concentrate on what the lecturer is saying (i.e. stay awake) if visual information is present. Perhaps this is a side-effect of a generation that grew up with television.

### **3.5 The value of integrated computer applications**

Anyone that has tried to explain something to someone over the phone (for example, give directions) has probably experienced the desire for some sort of shared drawing surface to supplement communication with sketches and annotations. Most desktop videoconferencing applications have a shared whiteboard capability. In the lecture environment, it is very helpful to have a good view of the speaker's written materials. It is also very helpful to be able to save a copy of the visual material and/or print it out. In conference mode, feedback on visual materials such as annotations is very useful.

Application sharing is another useful feature of desktop videoconferencing. A common example used to illustrate this capability is participants collaboratively editing a spreadsheet or word processor document.

### **3.6 Problems to overcome**

It takes time to compress video and audio and transmit it. This lag can contribute to a loss of interactivity experienced with videoconferencing systems. There is a learning curve involved with effectively utilizing new tools. Some people are not yet computer literate and may be wary of using desktop videoconferencing. Unforeseen circumstances are bound to happen, especially when computers are involved. The new replacements for excuses such as "my dog ate my homework" may be "the network was down" or "my computer crashed in the middle of the lecture."

### **3.7 The Forum prototype: A step in the right direction**

Some research has been done by Sun Microsystems toward developing an application that is suited to delivering interactive presentations to distributed audiences. This research prototype is called Forum. [10] Forum attempts to address many of the problems that were encountered during the demonstration project described in Chapter 4.

Forum is specifically designed to facilitate lecture mode interactions. Roles of lecturer and student are clearly defined, and the two types of participants have different user interfaces and different capabilities.

Students receive audio, video and slides from the lecturer. They are able to interact with the lecturer through a poll, a spoken question, or written comments. Lecturers receive audio and a still snapshot of a student asking a question. Both the lecturer and students can see a list of who is in attendance and the results of polls.

Some valuable features found on this prototype are 1) the ability for students to queue up to indicate they wish to ask a question, 2) the ability for students to "raise their hands" by the poll function, 3) the

ability for students to send in written comments without disrupting the lecturer, and 4) the ability for students to send messages to other students. These features are valuable because they increase the amount and ease of interaction between the lecturer and students as well as interaction among students.



[TABLE OF CONTENTS]

## **4. DEMONSTRATION PROJECT: INTERACTIVE SEMINAR USING DESKTOP VIDEOCONFERENCING**

The purpose of this demonstration project was to test, demonstrate and evaluate the desktop computer model of distance learning. Live audio, video, and slides were sent to remote participants throughout the state of North Carolina using the Internet MBone (technical aspects of the MBone are discussed in section 2.4.2.6). This project successfully demonstrated that desktop videoconferencing technology can be used to deliver interactive presentations to participants at distributed locations.

### **4.1 Choosing the class and the medium**

The Engineering Entrepreneurs class was identified as a suitable class to use for this distance learning experiment. In the past, this class has been taught at NCSU and North Carolina Agricultural and Technical State University (NC A&T) with funding from SUCCEED. Students in this class work in teams to conceive, design, develop and market products. An important part of this class is the weekly seminars given by speakers on entrepreneurial issues. Previously these seminars were presented at NCSU, videotaped, and sent to NC A&T to be viewed by students there. By offering these seminars live to the students at NC A&T, it was hoped to demonstrate that the capability of real-time interaction contributed positively to the effectiveness of the seminars.

MBone was chosen as the medium for the broadcast of the seminars for several reasons. MBone was already available at NCSU, and though NC A&T did not yet have MBone capability, it was determined that set-up time and cost for NC A&T was reasonable. Using MBone, the seminars could be broadcast to the entire state of North Carolina and not just NC A&T.

The original plans changed slightly when the NC A&T Engineering Department decided not to offer the Entrepreneurs Class in the Spring 1995 semester. However, NC A&T agreed to assist with the experiment by setting up the necessary equipment and arranging for remote viewers to provide feedback about the broadcast. This arrangement was acceptable, since the main purpose of the project was to demonstrate the use of this technology.

### **4.2 Infrastructure**

The following sections describe the hardware, software and network infrastructures that were utilized in this experiment.

#### **4.2.1 Hardware**

Both NCSU and NC A&T used Sun SPARCstation 5 workstations in this experiment. Each of the workstations had a SunVideo real-time video capture and compression board as well as built-in audio capability.

The SunVideo board captures, digitizes and compresses video signals in NTSC, PAL or S-Video format. One input port is available for S-Video format and two input ports are available for NTSC/PAL format. Figure 4.1 illustrates the input ports on the SunVideo card. The SunVideo card supports several different video compression techniques: CellB, JPEG, MPEG-1 (I frames only) and MPEG-1 (I and P frames). The SunVideo board does hardware compression only, the decompression is done in software. [15]

The Sun SPARCstation 5 workstations have built-in audio support. Audio may be input using either the microphone input port or the line-in input port. Table 4.1 details the audio input and output port characteristics for the Sun workstations.



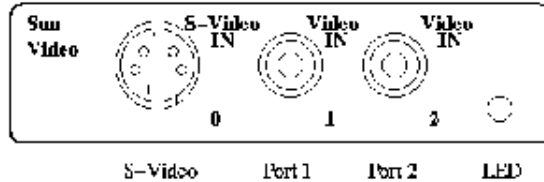


Figure 4.1 SunVideo input ports

Description	Characteristics
Microphone Input	SunMicrophone II, 5 Vdc supplied on the tip via a 2.2 kohm resistor
Microphone Output	15 mV, 0.6-1 kohm impedance
Headphones Output	1 V typical, 2.4 V maximum; 16-1 kohm impedance
Line-In	2 V typical, 4 V maximum; 5-50 kohm impedance
Line-Out	1 V typical, 2.4 V maximum; 5-50 kohm impedance

Table 4.1 Sun SPARCstation 5 stereo audio inputs and outputs

Audio and video for the seminars originated from the NCSU Park Shops studios. Audio and video signals were sent over fiber and switched to the lab in 319 Daniels Hall where "cyclops.ece.ncsu.edu", the Sun SPARCstation 5 workstation is located. This configuration allowed high quality audio and video signals to be received from Park Shops without having to physically move the transmitting workstation to that building.

Another workstation, "rogue.eos.ncsu.edu", was set-up in the Park Shops conference room adjacent to the studio. This workstation, a DEC 3000 Model 300 Series AXP, was configured to receive the broadcast and feed the audio signal to an audio mixing board and back into the studio. This configuration provided a way for audio from the remote participants to be heard by the speaker and the local audience. Table 4.2 shows the audio pin-out configuration for the DEC workstation. Figure 4.2 details the entire audio and video flow described above.

Description	Pin Number
Audio Input A	1
Earphone Interface 1 (Audio Out)	2
Earphone Interface 2 (Audio Out Return)	3
Audio Input B (Ground)	4

Table 4.2 DEC 3000 Model 300 Series AXP audio port pin-outs

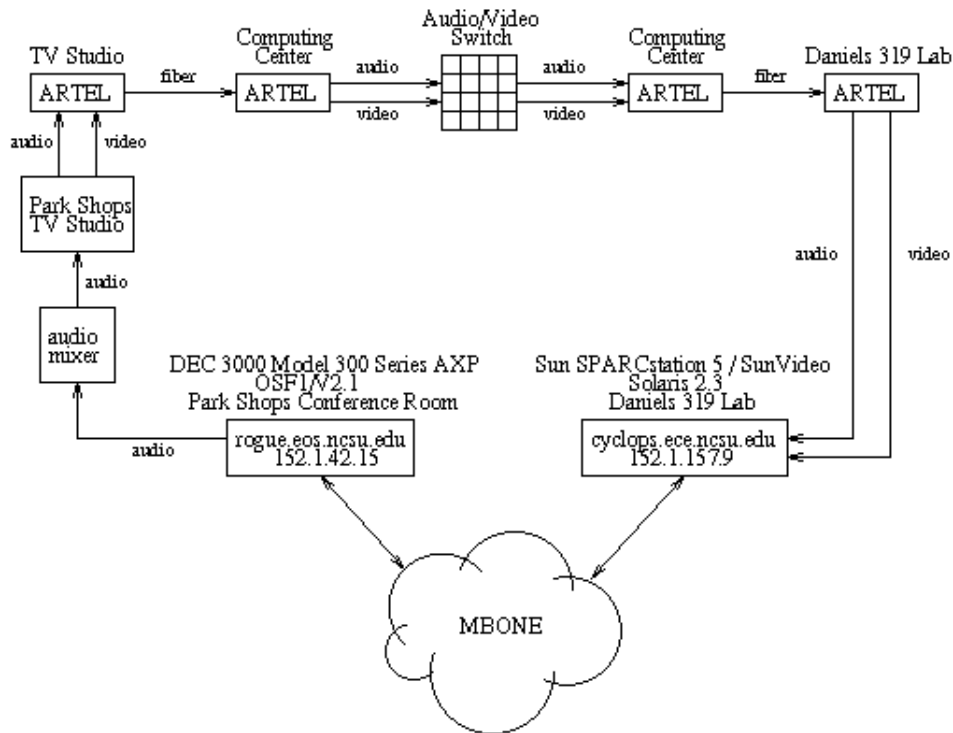


Figure 4.2 Audio and video flow

#### 4.2.2 Software

Four software tools that are commonly used for MBone broadcasts were used for this experiment. These tools are shown in Table 4.3. All are available free from various ftp sites.

Application	Description
Session Directory (sd 1.14)	Tool that lists available MBone sessions
Network Video (nv 3.3beta)	Videoconferencing tool
Visual Audio Tool (vat 3.4)	Audioconferencing tool
Whiteboard (wb 1.59)	Shared Whiteboard tool

Table 4.3 MBone software tools used in the seminar broadcasts

#### 4.2.3 Network topology

Because MBone audio and video sessions can consume a large amount of bandwidth, care needs to be taken to minimize the effect of the traffic on other users of the network. The network topology at NCSU that is relevant to MBone is described below and illustrated in Figure 4.3.

NCSU has a workstation, "hud.cc.ncsu.edu," acting as the main multicast router for campus. A multicast tunnel is configured between "hud.cc.ncsu.edu" and "mbone.ncrn.net", a workstation operated by the campus Internet provider NC-REN. Fan-out to other machines on campus is provided by "hud.cc.ncsu.edu" which is directly connected to the high-speed campus FDDI fiber backbone via an FDDI Ethernet switch. This configuration effectively gives "hud.cc.ncsu.edu" a dedicated bandwidth of 10Mbps. Other machines acting as multicast routers include "cyclops.ece.ncsu.edu" and

"rogue.eos.ncsu.edu."

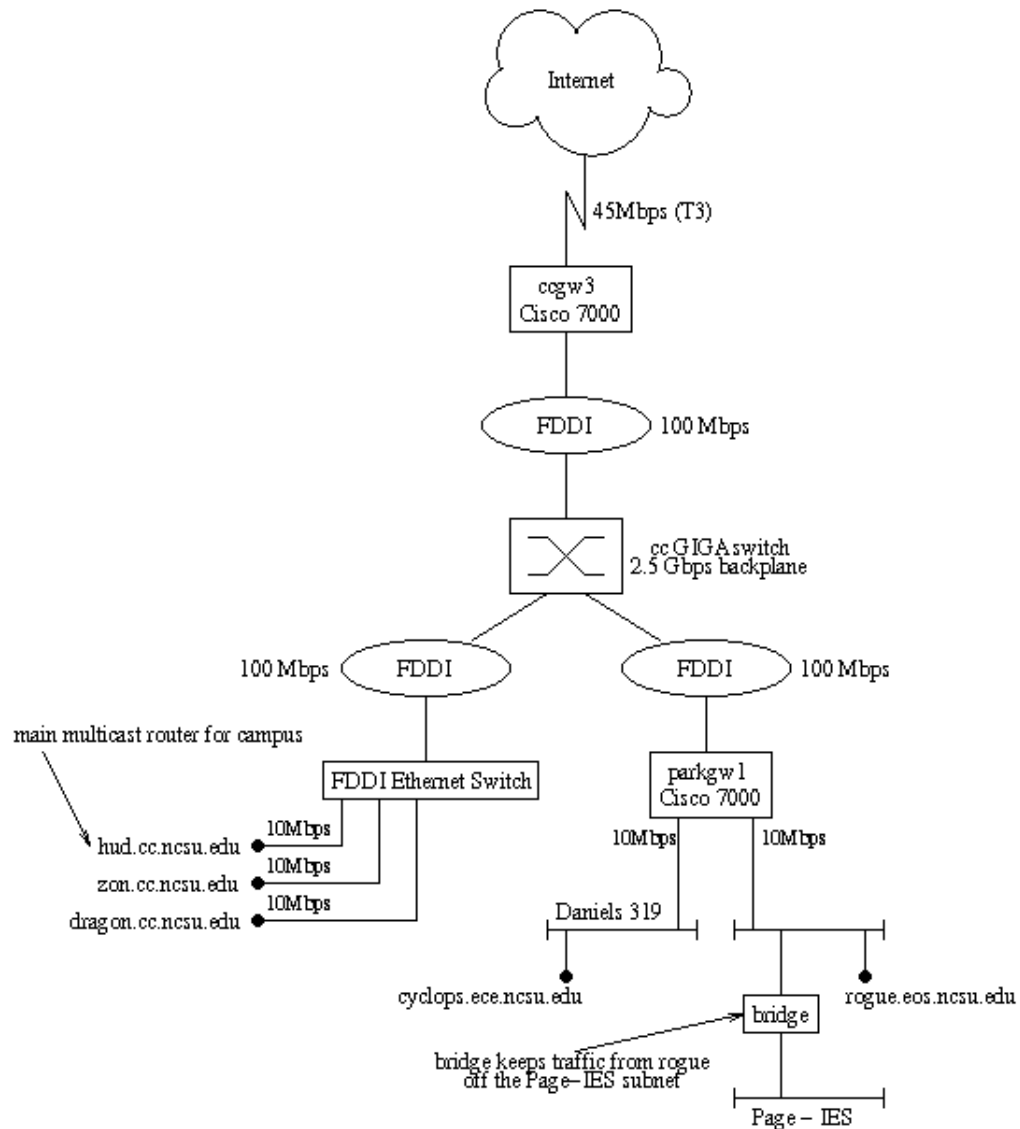


Figure 4.3 NCSU MBone network topology

The workstation "cyclops.ece.ncsu.edu" originated the MBone seminar broadcasts. This workstation resides in the Daniels 319 lab which is segregated onto a private subnet. This configuration allowed the generation of MBone traffic without concern about negatively impacting other network users.

The workstation "rogue.eos.ncsu.edu" was configured to receive the broadcasts and feed the audio back to the studio. This workstation resided in the Park Shops conference room adjacent to the studio. Since the subnet shared by the studio and the Industrial Extension Service (IES) in Page Hall was already heavily loaded, there was some concern about the extra traffic that the MBone broadcast would create. To work around this problem, a bridge was added that isolated rogue from the Page-IES network.

An important note is that since "hud.cc.ncsu.edu" tunnels to both "cyclops.ece.ncsu.edu" and "rogue.eos.ncsu.edu," two streams of multicast traffic pass through the campus fiber backbone and the parkgw1 router. This redundant traffic was not enough to cause any network problems, but is something that may need to be addressed in the future.

### 4.3 Seminar description

The following sections give a detailed description of the seminars that were broadcast in this experiment. Use of the MBone tools is described. Quantitative and qualitative evaluations of the experiment are outlined.

#### 4.3.1 Seminar dates

A total of seven seminars were broadcast between February and April of 1995. The dates and descriptions are shown in Table 4.4. The April 5 seminar was not a live broadcast due to class being cancelled for that day. As a substitute, a tape of Tim Seaver from MCNC discussing technical aspects of the MBone was played. April 12 was the first seminar that students from NC A&T participated in. It was also the first seminar where interactive audio feedback capability from participants to the studio was available.

Date	Description
February 22	Scott Daugherty: Executive Director, North Carolina Small Business and Technology Development Center
March 1	Dennis Dougherty: Founding partner, Intersouth Partners
March 29	Student Mid-term Presentations
April 5	Tim Seaver: Re-broadcast of a seminar about the MBone
April 12	Walter Daniels: Co-Founder of Daniels & Daniels Law Firm
April 19	Jim Poitras: Founder, President and CEO of Integrated Silicon Systems
April 26	Dr. Dave Van den Bout

Table 4.4 Seminar dates and descriptions

#### 4.3.2 Broadcasting the seminar

Each seminar was announced using 'sd'. This tool acts as the "TV Guide" of the MBone, displaying currently available sessions. A new session is created by pressing the "New" button which causes the "New" popup window to be displayed. In the "New" popup window, there are settings for the session name, description, address, scope (ttl), and lifetime (how long 'sd' will announce the session). In addition, certain media such as audio, video and whiteboard can be associated with a session. Unused address, port, and RTP channel id numbers are selected by 'sd', but these values can be changed if desired. Figure 4.4 shows the main 'sd' window and the "New" popup window.

All seminars were broadcast with audio and video. In addition, one seminar was broadcast with audio, video and whiteboard. Sessions were created in 'sd' several hours before the seminar actually began so that remote participants could see the session listed in their 'sd' ahead of time. The scope of the sessions was set to ttl=63, limiting the sessions to North Carolina sites only.

To start a session, a remote viewer finds the desired session in 'sd', highlights the session with a click of the mouse, and selects the "Open" button. The appropriate tools for that session will be automatically

started. For example, if a session has audio and video associated with it, Visual Audio Tool (vat) and Network Video (nv) will be invoked.

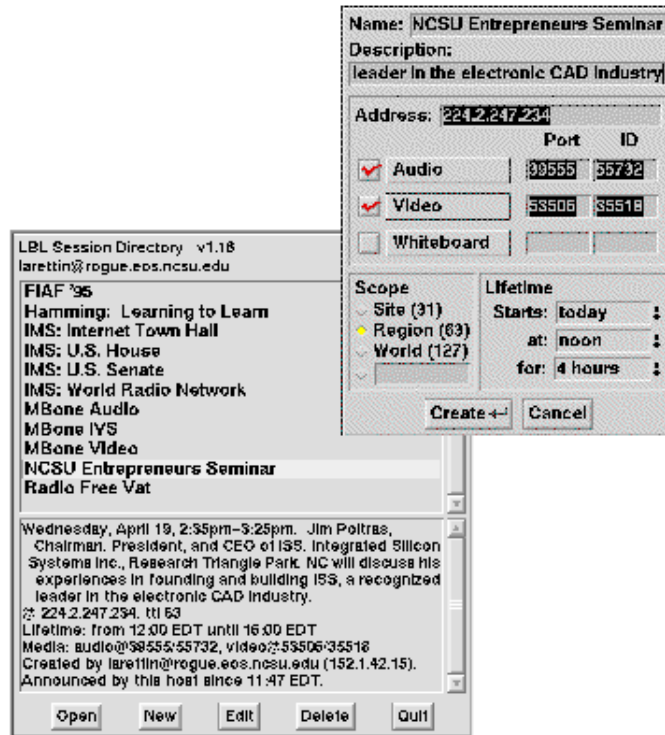


Figure 4.4 Session Directory (sd): Main window and "New" popup

Video was transmitted using Network Video (nv). The correct video input port was chosen, and the video encoding method was changed from the default "Native NV" to "Sun CellB." "Sun CellB" encoding allowed us to send approximately 3-3.5 frames per second, where "Native NV" encoding only resulted in around 1 frame per second. A third encoding format "CU-SeeMe" was not used because it encodes in black and white only. The "Max Bandwidth" slider, which ranges from 1-1024 kbps, was kept at the default of 128 kbps. The defaults "Medium" and "Color" were also maintained. The "Name" field was changed from the default of username@hostname to something more descriptive like "Entrepreneurs Seminar." Figure 4.5 shows the main 'nv' window and the video popup window that is displayed by clicking once on the video icon in the main window.

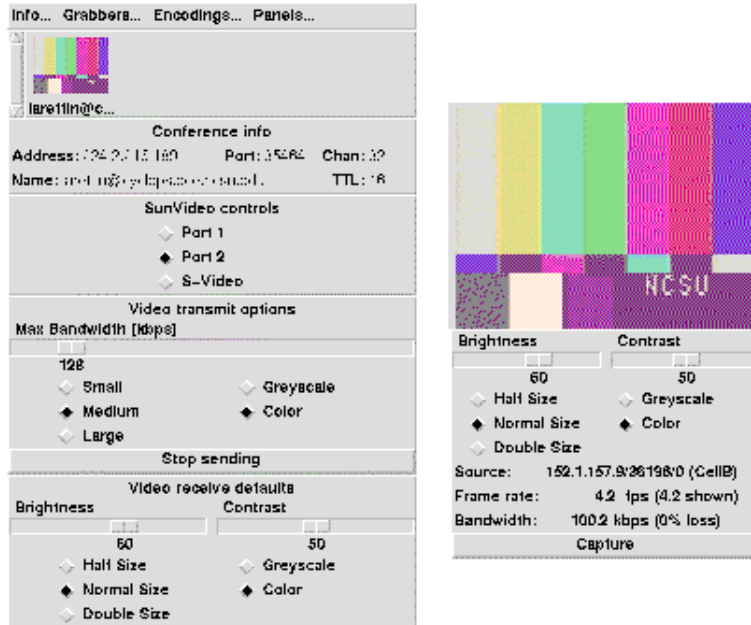


Figure 4.5 Network Video (nv): Main window and video popup

Audio was transmitted using Visual Audio Tool (vat). The default audio input was changed from microphone mode to line-in mode, and the gain level slider was moved to its minimum value. Pressing the "Menu" button brings up the "Menu" popup window, where several options were changed. First, the automatic gain control for the input audio was disabled by clicking the "Mike" button and manually setting the value to -6 dB. The value of -6 dB was selected after determining experimentally that this resulted in a good transmitted audio level. Silence suppression was left on, since this mode allows for interactivity. It was noted that if silence suppression was turned off, the transmitting site would keep the focus of the audio session making it difficult or impossible for other sites to send audio. For a non-interactive scenario however, disabling silence suppression at the transmitting site should result in better audio quality since silence suppression sometimes tends to clip off the beginning or ending of an audio stream.

On the receiving end, "Lecture Mode" resulted in better received audio quality than "Conference Mode." In "Lecture Mode" a larger receive buffer is maintained in an effort to reorder packets that arrive out-of-order. "Lecture Mode" resulted in approximately a 5 second delay between the time the audio was sent from the transmitting site and the time the audio was played back at the receiving site. This mode makes interactive feedback difficult because of the large delay factor. "Conference Mode" is recommended for interactive discussions. Figure 4.6 shows the main 'vat' window and the "Menu" popup window.

For the April 12 seminar, Whiteboard (wb) was used to transmit the speaker's slides to the remote audience. In the other seminars, any slides that were used were visible in the video window, but 'wb' offers several advantages over video slide transmission. Slides in the video window are sometimes difficult to read due to the video encoding. Particularly with "Sun CellB" encoding, the slides end up looking quite "blocky." Using 'wb' the slides are clear and very easy to read. Also, the remote audience can use 'wb' to print copies of the slides to their local printer.



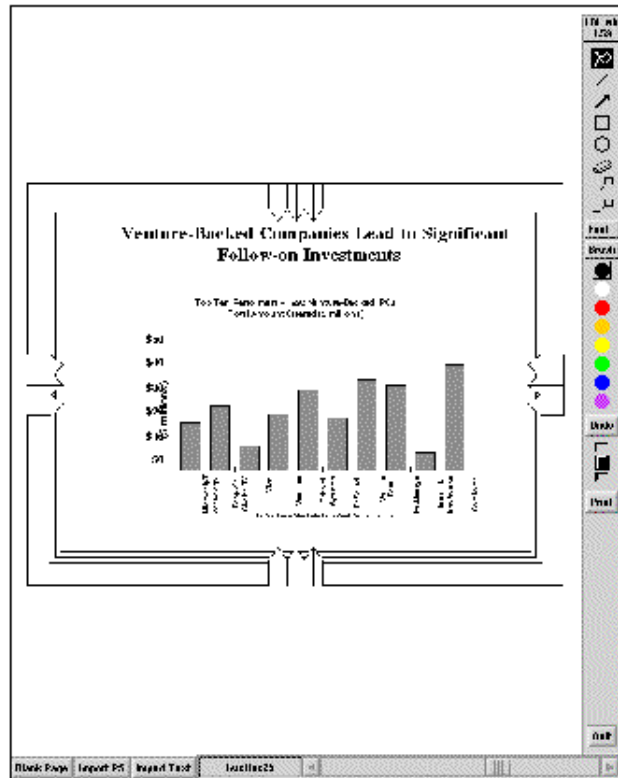


Figure 4.7 Whiteboard (wb): Main window

### 4.3.3 Quantitative measurements

Both 'nv' and 'vat' have mechanisms to report statistics about the data received. These statistics were used to judge the quality of the seminar broadcasts.

The 'nv' video window displays the incoming and displayed frame rate, the incoming bandwidth, and the approximate network loss averaged over the last few seconds. These values can be seen in the video popup in Figure 4.5. The statistics that seem to give the best indication of quality are incoming frame rate and network loss.

Using 'vat' there are two ways to obtain statistics. By typing "s" in the 'vat' main window, current statistics are sent to standard out. This information includes the number of packets received, number with bad headers, number with wrong id, number with bad encoding, number lost, number 'vat' has dropped, number of duplicate packets, number arrived out of order, and the current playout value. The statistics that seem to give the best indication of quality are the number of packets lost and the number of packets that arrived out of order. A second way of obtaining statistics is to press the shift key and the left mouse button while holding the mouse pointer over the transmitting site in the 'vat' main window. This will bring up a popup window showing the statistics described above along with a strip chart plotting the selected statistic. The three columns of numbers show the difference between the last update and the current update, a smoothed average of the first column, and the aggregate statistics since 'vat' started.

To evaluate the quality of the received data, statistics were collected at NC A&T. The above statistics were recorded at 5 minute intervals during the seminar broadcast.



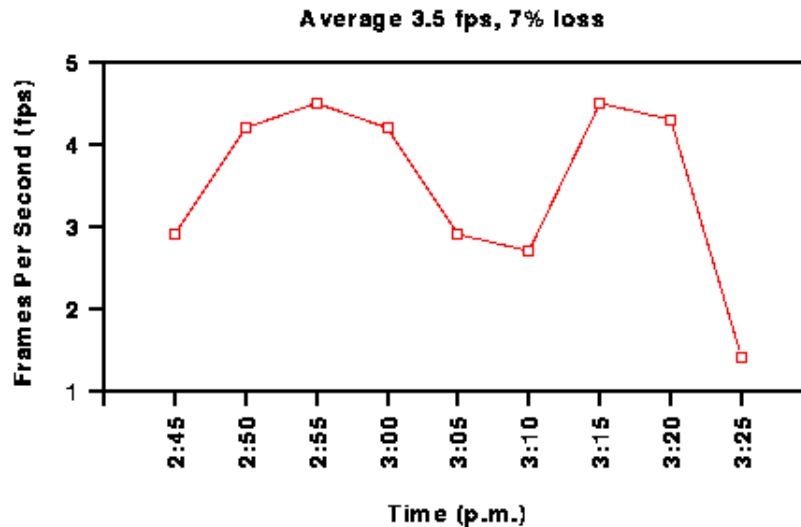


Figure 4.8 'nv' statistics from NC A&T for the 04/12/95 seminar

Figure 4.8 illustrates some results of the 'nv' statistics. An average frame rate of 3.5 frames per second (fps) was observed at NC A&T for this particular seminar. This value ranged from 1.4 fps to 4.5 fps. This range is due to the video encoding method. A higher frame rate is achieved when there is less movement in the encoded scene (for example, when the video shows only a slide that is not moving). Lower frame rates are achieved when there is movement in the scene, such as when the speaker paces back and forth, or rapidly gestures with their hands.

Figure 4.9 illustrates some results of the 'vat' statistics. An average of 3.1% of the audio packets were lost and 2.9% reached the destination out of order and needed to be rearranged. It is interesting to note the correlation between the packets lost and packet arrived out of order. Observing the graph, the two lines follow each other very closely. This seems to support the hypothesis that at this snapshot of time, network congestion was affecting the two metrics similarly. With less network congestion, it is expected that more packets would arrive safely at their destination and in the correct order. With more network congestion, it is expected that more packets would be lost or arrive out-of-order. An additional statistic not shown on the graph is the received playout delay. NC A&T had "Lecture Mode" set in 'vat' and experienced an average of 5.2 seconds playout delay.

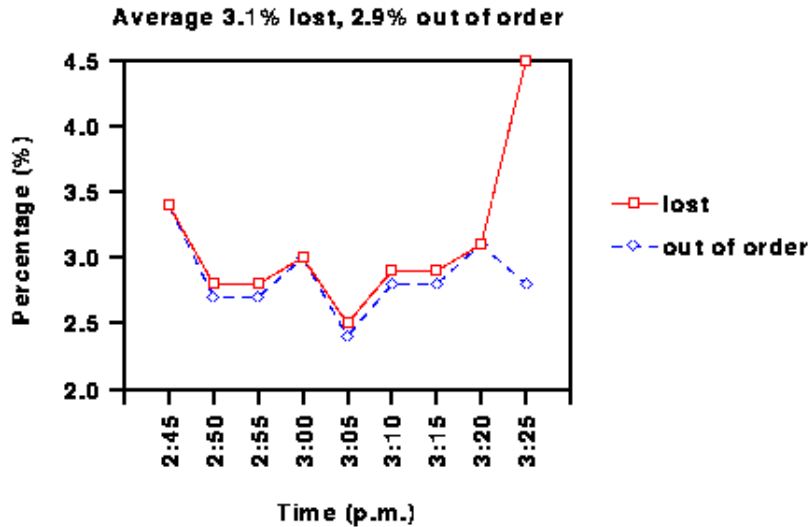


Figure 4.9 'vat' statistics from NC A&T for the 04/12/95 seminar

#### 4.3.4 Qualitative measurements

An attempt was made to gather qualitative feedback from students at NC A&T and other people who participated in the seminar broadcasts. The questions used to elicit feedback can be found in Appendix 7.2. Because the amount of feedback actually received was minimal, no broad conclusions can be made. However, it is the author's opinion that similar results would have been seen with a larger data sample.

The quality of each of the media (audio, video, and whiteboard) was rated acceptable or excellent by all survey respondents. Audio was rated to be essential to the conference, video was rated either useful or marginal, and whiteboard was rated essential. Participants were either somewhat interested or indifferent about the topic of the broadcast. When multiple people watched the broadcast at the same workstation, they tended to talk about matters unrelated to the broadcast. Though the speaker explicitly asked for questions from the remote audience, none of the remote audience members asked any questions (though one participant tried unsuccessfully). Participants thought that they would be more likely to ask questions in this environment or that it would not make a difference. Participants were optimistic about attending a class this way in the future, saying either they definitely would or that they would consider it.

#### 4.3.5 Observations from the seminar broadcasts

This demonstration project illustrated the feasibility of delivering an interactive class using desktop videoconferencing technology. Some observations from the demonstration project are described below.

Audio interaction can be difficult. Even when the speaker explicitly asked for questions from the remote audience, interaction was impaired by the delayed audio at the receiving sites. In some cases, the playout delay was as high as 5 seconds. This extremely long delay made an interactive exchange difficult or impossible. It is important for receiving sites to configure their audio tool to minimize the playout delay if interaction is desired. Using the 'vat' tool, setting the mode to "Conference" produced a minimal delay which was typically about 1 second.

It was confusing if several remote audience members tried to ask a question at the same time. Improved user interfaces (such as the type implemented by the Forum prototype discussed in section 3.7) that facilitate audience members queuing up to ask questions can improve this type of interaction greatly.

Audio echo can be a problem. A delayed echo at the transmitting site can occur if the audio is sent over the network and a remote workstation receives the audio and feeds it back into its microphone. The original audio is then sent back to the transmitted site. This problem can be alleviated by having the audio output at the receive site set to a headphone set instead of the workstation's speaker. Also, delayed echo can be eliminated by echo cancellation schemes.

The whiteboard can be a valuable way to transmit written materials to the remote audience. The received materials are high quality, easy to see, and can be printed out. However, it is sometimes difficult to get a speaker's materials ahead of time, and not all flavors of PostScript were supported by the whiteboard tool. During the seminars, the whiteboard was sometimes used for informal communication among participants when the audio channel was not operative or when speaking over the audio channel would have caused a disruption of the seminar presentation.

The use of desktop videoconferencing applications can lead to a screen management problem. Typically there are separate windows for audio, video, and whiteboard. Additional windows may exist for statistics displays. The monitor screen can become cluttered and users may spend a lot of time moving, resizing, and iconifying windows.

Even slow frame rate video seems to add value to the conference. The frames rates experienced with this demonstration project were usually in the range of 2-5 frames per second. However, video still seemed to create a sense of presence and give visual clues.

It can be valuable to be able to do other things at the workstation while receiving a broadcast. For instance, it could be useful to look up some additional information by telnetting to the library, or bringing up a page in a web browser. However, being at the desktop workspace can also be a disadvantage. Distractions such as phone calls and email can divert attention away from the broadcast. John Grant, a participant of the seminar broadcasts at the University of North Carolina at Charlotte, commented that a red light outside his office, indicating that he was participating in a conference, would have been helpful to prevent distracting interruptions.

#### **4.4 Other MBone broadcasts**

In addition to the broadcasts of the Engineering Entrepreneurs class, several other events were broadcast. Two broadcasts were done in cooperation with the EPA and one broadcast was a demo for the ABC morning show "Good Morning America."

##### **4.4.1 EPA workshops**

The EPA workshops were already scheduled to take place in the NCSU Park Shops studios and be sent over satellite to various sites. These seminars were broadcast on the MBone with a ttl value equal to 127 which is a global scope. The first broadcast, "Computer Based Tools to Assist in Air Quality Modeling," took place on March 22 1995, 12 noon to 4 pm. A total of 33 MBone participants tuned in for some or all of the broadcast. The second broadcast, "Electronic Access to EPA and Other Environmental Information," took place on April 18 1995, 12 noon to 4 pm. A total of 42 MBone participants tuned in for some or all of the broadcast. Viewers were present from all over the United States as well as the United Kingdom, Finland, France, The Netherlands, New Zealand, Spain, Mexico, and Italy. A listing of the MBone participants for these two broadcasts can be found in Appendix 7.3.

A viewer from the University of Illinois Urbana-Champaign (UIUC) tracked statistics on the quality

received at his site during the April 18 broadcast. From these statistics, it can be concluded that the quality experienced locally at NCSU and within North Carolina is comparable to the quality received at other sites. Figure 4.10 shows the 'nv' and 'vat' statistics collected at UIUC.

#### 4.4.2 "Good Morning America" demo

On May 23 1995, a demo was arranged for the ABC morning show "Good Morning America." During this broadcast, the ability for audio interaction was successfully demonstrated. A scan converter was set up to scan the image of the workstation screen in the Park Shops conference room. The workstation display was visible to the speaker and the local audience. Three questions were successfully posed from the Mbone audience. This was the most successful broadcast in terms of demonstrating the potential of this medium for remote interaction.

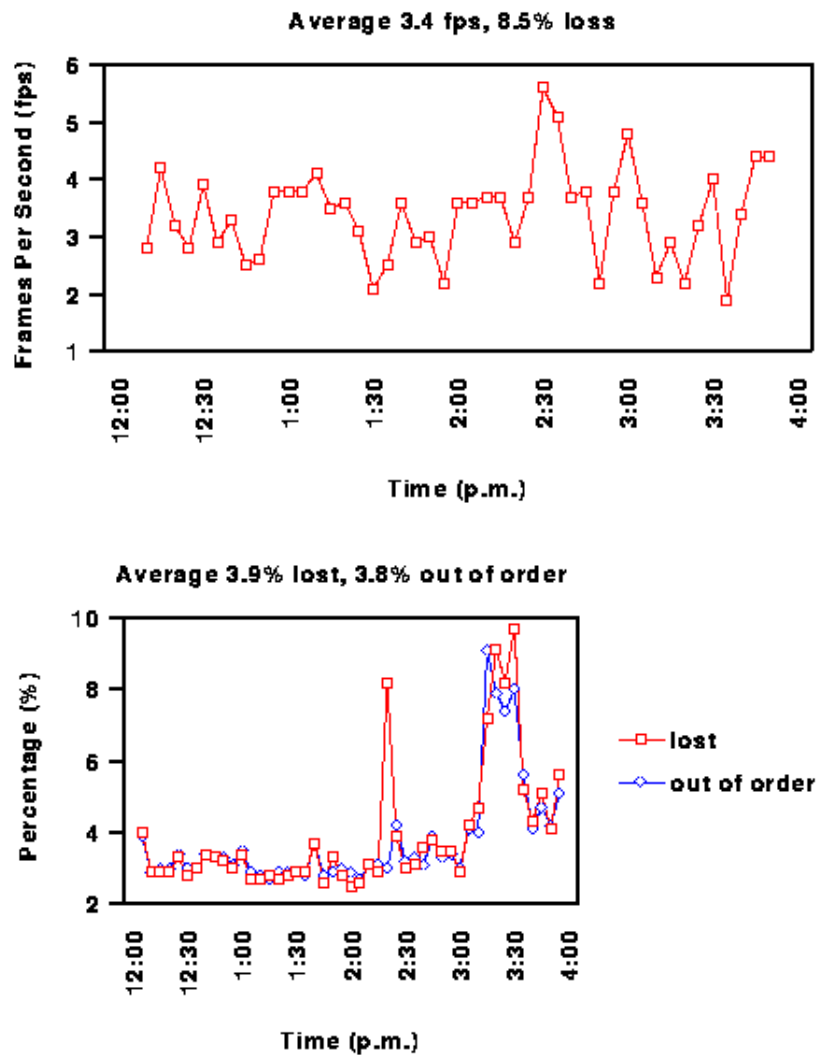


Figure 4.10 'nv' and 'vat' statistics from UIUC for the 04/18/95 EPA Broadcast



[TABLE OF CONTENTS]

## 5. CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

Desktop videoconferencing has great potential in the area of distance learning. This paper described a demonstration project which used desktop videoconferencing to deliver a weekly seminar class to remote participants that were located throughout the state of North Carolina. This demonstration project illustrated the feasibility of delivering classes using this technology and also illustrated some of the problems that need to be addressed to improve the effectiveness of this mode of educational delivery.

This paper discussed important technical aspects of desktop videoconferencing. Data compression is important since audio and video data require a large amount of bandwidth for transmission. Data compression techniques vary in their quality, amount of bandwidth required, and computational complexity. There are two major types of communication channels available to transmit the data: circuit- and packet-switched. Circuit-switched channels such as ISDN offer dedicated bandwidth and predictable timing of data delivery but do not easily support multipoint communication, which is required for the type of lecture mode seminar illustrated by the demonstration project. Packet-switched channels, either local (LAN) or wide area (Internet), more easily support multipoint communication but do not provide predictable timing of data delivery. Desktop videoconferencing systems have requirements for timely data delivery and reliable data delivery. For example, audio and video data require timely delivery while other types of data, such as whiteboard data, require only reliable delivery. The tools used in the demonstration project compensate for the limitations of packet-switched channels by buffering audio data at the receiver and using modified video encoding techniques that frequently send updated video information in case some information has been lost in the network. BISDN and ATM show promise for solving some of the problems encountered with both circuit- and packet-switched channels for desktop videoconferencing data. Emerging interoperability standards that allow systems to communicate with each other are also very important to the future of desktop videoconferencing.

This paper also discussed important non-technical aspects of desktop videoconferencing. The successful use of these systems requires understanding the benefits of having these systems on the desktop and the value added by audio, video, and integrated computer applications. Since these are communication systems, user interfaces should facilitate easy interaction among participants.

The demonstration project seminar was delivered using the Internet MBone. MBone is well suited to large, distributed presentations. MBone uses network bandwidth efficiently and allows participants to join and leave a conference easily. Remote seminar participants received audio, video and, in some cases, whiteboard information. Interactive audio communication was possible among the lecturer and remote participants. The MBone tools allowed interactions to take place, however better tools that incorporate improved user interfaces could greatly increase the ease of interaction. The largest problem encountered during the demonstration project was the difficulty in facilitating effective audio communication.

This demonstration project should serve as inspiration for continued experimentation with this technology. Future broadcasts should involve a larger number of remote participants and participants that are actually enrolled in the class. Classroom-type lectures should be supplemented with "video office hours" held between the instructor and the remote students. This project illustrated the feasibility of desktop videoconferencing for educational delivery. Future projects should focus on the most effective way to use this technology.

---



[TABLE OF CONTENTS]

## REFERENCES

- [1] R. Aravind et al., "Image and Video Coding Standards," *AT&T Technical Journal*, Vol. 72 No. 1, January/February 1993, pp. 67-89.
- [2] D. Bertsekas and R. Gallager, *Data Networks*, Prentice-Hall, Inc., 1992.
- [3] S. Casner, "Frequently Asked Questions (FAQ) on the Multicast Backbone (MBONE)," December 22 1994, <<ftp://isi.edu/mbone/faq.txt>>.
- [4] S. Casner, "Getting on the Mbone: Videoconferencing Over the Internet," slides from a presentation given at the NLUUG Spring Conference, April 13 1995, <<ftp://ftp.isi.edu/mbone/nluug-slides.ps.Z>>.
- [5] S. Deering, "Host Extensions for IP Multicasting," Internet Request For Comment 1112, August 1989, <<http://ds.internic.net/rfc/rfc1112.txt>>.
- [6] A. Daniel et. al., *Introduction To Video And Audio Compression Techniques*, Course Notes 6, ACM Siggraph 1994, Orlando FL, July 1994.
- [7] R. Frederick, "Experiences with real-time software video compression," July 22 1994, <<ftp://parcftp.xerox.com/pub/net-research/nv-paper.ps>>.
- [8] S. Gale, "Desktop video conferencing: technical advances and evaluation issues," *Computer Communications*, Vol. 15 No. 8, October 1992, pp. 517-525.
- [9] ITU-T Recommendation H.320, "Narrow-Band Visual Telephone Systems and Terminal Equipment," March 1993, <[http://www.itu.ch/itudoc/itu-t/rec/h/h320\\_23397.html](http://www.itu.ch/itudoc/itu-t/rec/h/h320_23397.html)>.
- [10] E. Isaacs et al., "A Forum for Supporting Interactive Presentations to Distributed Audiences," *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW) '94*, Chapel Hill NC, October 1994, pp. 405-416, <<http://www.sun.com/tech/projects/coco/papers.html>>.
- [11] V. Jacobson, "The Mbone - Interactive Multimedia on the Internet," slides from a presentation given at U.C. Berkeley (and on the Mbone), February 17 1995, <<ftp://ftp.ee.lbl.gov/talks/vj-ucb-feb17.ps.Z>>.
- [12] M. Liou, "Overview of the Px64 kbit/s Video Coding Standard," *Communications of the ACM*, Vol.34 No. 4, April 1991, pp. 60-63.
- [13] M. Macedonia and D. Brutzman, "Mbone Provides Audio and Video Across the Internet," *Computer*, Vol. 27 No. 4, April 1994, pp.30-36, <<ftp://taurus.cs.nps.navy.mil/pub/mbmg/mbone.html>>.
- [14] D. Minoli and R. Keinath, *Distributed Multimedia Through Broadband Communications*, Artech House, Inc., 1994.
- [15] Sun Microsystems, Inc., *SunVideo 1.0 Users's Guide*, 1993.



- [16] D. Pan, "Digital Audio Compression," *Digital Technical Journal*, Vol. 5 No. 2, Spring 1993, pp. 28-40, <<ftp://ftp.digital.com/pub/Digital/info/DTJ/mm-05-audio-compress.ps>>.
- [17] K. Pohlmann, *Principles of Digital Audio*, Howard W. Sams & Co., 1985.
- [18] C. Poynton, *The Science of Digital Color*, Course Notes 26, ACM Siggraph 1994.
- [19] C. Poynton, "Frequently Asked Questions about Colour," May 28 1995, <<http://www.inforamp.net/~poynton/Poynton-colour.html>>.
- [20] M. Sattler, "CU-SeeMe Desktop Videoconferencing," WWW page, June 1995, <<http://www.indstate.edu/msattler/sci-tech/comp/CU-SeeMe/>>.
- [21] J. Tang and E. Isaacs, "Why Do Users Like Video? Studies of Multimedia-Supported Collaboration," Sun Microsystems Laboratories Inc. Technical Report TR-92-5, December 1992, <<http://www.sun.com/tech/projects/coco/papers.html>>.
- [22] K. Taylor and K. Tolly, "Desktop Videoconferencing: Not Ready for Prime Time," *Data Communications*, Vol. 24 No. 5, April 1995, pp. 64-80.
- [23] G. van Rossum, "Frequently Asked Questions: Audio File Formats," January 2, 1995, <<ftp://ftp.cwi.nl/pub/audio/AudioFormats.part1>> and <<ftp://ftp.cwi.nl/pub/audio/AudioFormats.part2>>.
- [24] R. Vetter, "Videoconferencing on the Internet," *Computer*, Vol. 28 No. 1, January 1995, pp. 77-79.



[TABLE OF CONTENTS]

## APPENDICES

### 7.1 World Wide Web desktop videoconferencing product survey

[http://www2.ncsu.edu/eos/service/ece/project/succeed\\_info/dtvc\\_survey/](http://www2.ncsu.edu/eos/service/ece/project/succeed_info/dtvc_survey/)

### 7.2 Remote seminar participant feedback questions

We are evaluating MBone as a way to do distance learning. Your feedback as a participant will be very useful to our evaluation.

Date: \_\_\_\_\_

1. How many other MBone sessions have you watched?

- None
- 1-2
- 2-5
- 5-10
- >10

2. Please rate the quality of the following media received at your site:

Audio Video Whiteboard

- Excellent
- Acceptable
- Poor
- Did Not Receive

3. Please rate the importance of the following media to the overall broadcast:

Audio Video Whiteboard

- Essential (broadcast would be meaningless without)
- Useful (enhanced the quality of the broadcast)
- Marginal (could have been successful without)
- Useless (totally unnecessary)

4. Please rate your interest in the subject matter of the presentation:

- Very interested
- Somewhat interested
- Indifferent

5. How many people watched the broadcast at the same workstation? \_\_\_\_\_

If more than one person watched the broadcast at the same workstation, did you:

- talk amongst yourselves about the content of the broadcast
- talk amongst yourselves about things unrelated to the broadcast

not talk at all

6. Did the speaker explicitly ask for questions from the remote Mbone participants?

Yes

No

Don't remember

7. Did you ask a question?

Yes

No

8. Do you think you would be more or less likely to ask a question with this type of broadcast:

More likely

Less likely

9. Would you be interested in attending a class this way in the future?

Yes

Maybe

No

Please feel free to add any additional comments below, including any ways you think this type of broadcast could be improved:

### **7.3 Mbone participants for 03/22/95 and 04/18/95 EPA broadcasts**

#### **EPA Workshop 03/22/95**

Total Mbone Participants = 33

(.edu=6)

gb@darmok.cs.utah.edu [155.99.208.1=darmok.cs.utah.edu]

jhm@blackhole.cs.utah.edu [155.99.208.51=blackhole.cs.utah.edu]

jbass@zon.cc.ncsu.edu [152.1.1.234=zon.cc.ncsu.edu]

jeff@hubsrv.alaska.edu [137.229.5.5=hubsrv.alaska.edu]

scolber@labrep [128.140.2.90=labrep.mathcs.emory.edu]

steve@wintermute.ucsd.edu [132.239.1.8=wintermute.ucsd.edu]

(.com=6)

deleon@cat.hpl.hp.com [15.25.48.203=cat.hpl.hp.com]

jec@mars.philabs.philips.com [130.140.55.36=mars.philabs.philips.com]

Dan Molinelli (TRW) (dan.molinelli@trw.com) [129.4.53.11=mpeterson.sp.trw.com]

test@mcast [132.197.160.10=h132\_197\_160\_10.gte.com]

Ray Sasaki (TRW) [129.4.53.10=bigmac.sp.trw.com]  
Vinay Kumar (EIT) <vinay@collage> [192.100.58.17=collage.eit.com]

(.gov=11)

Mark Bolstad (bolstad@vislab.epa.gov) [134.67.66.49=blad.rtpnc.epa.gov]  
crtb@midnight.dcrtnih.gov [156.40.112.29=midnight.dcrtnih.gov]  
dlb@sirius [134.67.66.12=sirius.rtpnc.epa.gov]  
dna@almond.nesc.epa.gov [204.46.19.36=almond.nesc.epa.gov]  
lpz@nautique.epm.ornl.gov [128.219.8.92=nautiquee.epm.ornl.gov]  
mjamshi@sorrento.arc.nasa.gov [128.102.80.44=sorrento.arc.nasa.gov]  
rdew@dude.dcrtnih.gov [128.231.129.47=dude.dcrtnih.gov]  
root@nepal.ccd.bnl.gov [130.199.88.1=nepal.ccd.bnl.gov]  
toy@fred.rtpnc.epa.gov [134.67.66.47=fred.rtpnc.epa.gov]  
wgl@isobar [134.67.67.217=isobar.rtpnc.epa.gov]  
wgl@tempest [134.67.67.221=tempest.rtpnc.epa.gov]

(.mil=1)

tplunke@sparhawk.nswc.navy.mil [128.38.42.73=sparhawk.nswc.navy.mil]

(.net=3)

crobson@andy.dia.atd.net [134.207.65.12=andy.dia.atd.net]  
Michael Bowen - CERFnet <mike@dover.cerf.net> [192.153.156.129=dover.cerf.net]  
dboehlke@spot.MR.Net [137.192.240.24=spot.mr.net]

(.uk=1)

le@moose.doc.ic.ac.uk [146.169.22.44=moose.doc.ic.ac.uk]

(.fi=1)

setala@kala.mdata.fi [192.98.43.122=kala.mdata.fi]

(.fr=1)

visio@cismmedia.univ-lyon1.fr [134.214.180.2=cismmedia.univ-lyon1.fr]

(.nl=1)

Axel.Belinfante@cs.UTwente.NL [130.89.16.45=utis107.cs.utwente.nl]

(.nz=1)

busa057@cow.mang.canterbury.ac.nz [132.181.155.22=cow.mang.canterbury.ac.nz]

(?=1)

melohn@fddi-sluggo [192.26.75.10] (192.26.75.10: Non-existent domain)

## **EPA Workshop 04/18/95**

Total MBone Participants = 42

(.edu=15)

David Meyer (UO) [128.223.60.116=frostbite-falls.uoregon.edu]

Greg Bothun (University of Oregon) [128.223.21.15=ruminant.uoregon.edu]  
smcgrew@theskye.uoregon.edu [128.223.20.84=theskye.uoregon.edu]  
Jack Callahan (West Virginia U.) [157.182.112.20=hopper.cs.wvu.edu]  
J.P. Suchoski (MTU) <jsuchosk@mtu.edu> [141.219.21.222=rift.geo.mtu.edu]  
Lenny Tropiano (ARL:UT) <lennyt@arlut.utexas.edu> [129.116.152.28=narn.arlut.utexas.edu]  
ram@shiva.bus.utexas.edu [128.83.153.47=shiva.bus.utexas.edu]  
Michael Van Norman (UCLA) [128.97.226.253=unagi.library.ucla.edu]  
dbl@verdi.cc.gatech.edu [130.207.3.248=verdi.cc.gatech.edu]  
dsr@lms100.lms.cornell.edu [128.84.170.170=lms100.lms.cornell.edu]  
jlohmar@rodin.beckman.uiuc.edu [128.174.209.158=rodin.beckman.uiuc.edu]  
stan@glorius.cs.uiuc.edu [128.174.240.11=glorius.cs.uiuc.edu]  
mark@marmot.usc.edu [128.125.62.109=marmot.usc.edu]  
niester@hotbox.math.lsa.umich.edu [141.211.60.51=hotbox.math.lsa.umich.edu]  
teledemo@cait-sun1.ucdavis.edu [128.120.134.76=cait-sun1.ucdavis.edu]

(.com=5)

JJ Krawczyk (Bay Networks) [192.32.174.133=maggie.wellfleet.com]  
belfiore@fox.delphi.com [199.93.0.243=fox.delphi.com]  
dpan@yamaha.engr.sgi.com [192.48.150.8: Non-existent domain]  
jledy@reliance.milford.ibm.com [192.231.8.132: Non-existent domain]  
moline@achtung.sp.trw.com [129.4.51.2=achtung.sp.trw.com]

(.gov=3)

Steve Elbert (Ames Lab) [147.155.30.3=elbert.ameslab.gov]  
dlb@sirius [134.67.66.12=sirius.rtpnc.epa.gov]  
xxseub@wyvern.lerc.nasa.gov [139.88.27.24=wyvern.lerc.nasa.gov]

(.mil=3)

Sally Ross (NCTS Washington) [198.116.62.7=empire.nctsw.navy.mil]  
don brutzman@nps.navy.mil [131.120.63.31=fletch.stl.nps.navy.mil]  
gillespi@dave.nrl.navy.mil [132.250.15.129=davenet-rtr.nrl.navy.mil]

(.net=1)

crobson@matt.dia.atd.net [134.207.64.14=matt.dia.atd.net]

(.uk=4)

George.Howat@Edinburgh.ac.uk [129.215.200.11=aten.ucs.ed.ac.uk]  
GMRC, Uni of Edinburgh [129.215.168.111=leonardo.ucs.ed.ac.uk]  
Graeme.Wood@Edinburgh.ac.uk [129.215.200.48=scorpio.ucs.ed.ac.uk]  
jk@rodent [193.62.83.68=rodent.ukerna.ac.uk]

(.es=3)

Ignacio Martinez (Fundesco, Madrid) [130.206.16.254=piquio.fundesco.es]  
root@astor.urv.es [193.144.16.4=astor.urv.es]  
root@tweety.urv.es [193.144.16.185=tweety.urv.es]

(.fi=3)

Jukka.Orajarvi@otol.fi (Oulu, Finland) [193.64.224.10=titan.otol.fi]

Vesa.Ruokonen@lut.fi (Vesa@IRC) [157.24.23.33=sauron.it.lut.fi]  
setala@kala.mdata.fi [192.98.43.122=kala.mdata.fi]

(.fr=2)

etienne@marquise.in2p3.fr [134.158.16.126=marquise.in2p3.fr]  
oliver@gameboy.3r.enst-bretagne.fr [192.44.77.93=gameboy.3r.enst-bretagne.fr]

(.mx=1)

macbeath@noc [140.148.1.16=noc.pue.udlap.mx]

(.it=1)

147.155.30.3 [131.175.2.11=iclsp20.cilea.it]

(?=1)

Soochon Radee [192.188.104.98: Non-existent domain]



[TABLE OF CONTENTS]