Chapter 19

VIDEOCONFERENCING SYSTEMS AND APPLICATIONS

Sandra Brey and Borko Furht

Abstract

In this chapter we present an overview of videoconferencing technologies, systems, standards, applications, and commercial products with an emphasis on Internet-based videoconferencing. We begin with the history of videoconferencing and then we introduce fundamentals of audio and video technologies for videoconferencing. The videoconferencing standards, such as H320, H.321, H.323, and H.324, are described next. We discuss networks for videoconferencing and evaluate IP-based and ATM-based videoconferencing systems. Various videoconferencing applications are briefly described as well as a variety of commercial videoconferencing products and systems.

1. INTRODUCTION

Videoconferencing is the transmission of live video images and audio between two or more disparate participants. Once merely a figment of science fiction writers' imaginations, it is now used for both business and personal use, on a variety of different types of network media and with varying degrees of quality.

Conversations may be one to one (point-to-point) or one-to-many (multipoint), in simplex (one-way only), half-duplex (one way at a time, taking turns) or full-duplex (all parties are seen and heard simultaneously). A range of products is offered, which span a wide spectrum of applications from group (or room) based systems, to desktop videoconferencing systems, to less expensive (and lower quality) personal conferencing/videophone systems. Products are available which can convert a multimedia PC, or even a television set, into a videoconferencing workstation.

Networks used include ISDN, IP packet-data LANs, ATM networks, analog phone lines, and even the Internet. Of course, the quality of the video and audio obtainable depends in large part on the characteristics of the network used. Factors such as data throughput rate, delay, and delay variation, differ widely based on the type of network used. Several different standards have been developed which are optimized for use on each of the different network types.

Taken together, these segments add up to a huge, rapidly growing business. This is partly due to the explosion of videoconferencing via personal computer and the Internet. The videoconferencing market is expected to exceed 3 billion dollars annually by 2001, as illustrated in Figure 1.

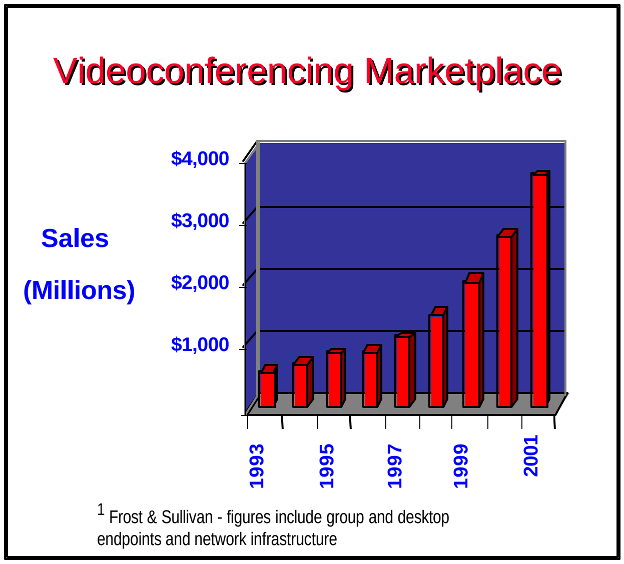


Figure 1. Projected growth in videoconferencing sales.

2. HISTORY OF VIDEOCONFERENCING

The first demonstrations of videoconferencing in the United States occurred in the 1920s. Other experiments were done in Europe in the 1930s. Put on hold during World War II, research began again in the mid 40s. However, it wasn't until 1964, at the World's Fair in New York, that Bell Labs presented the Picturephone, shown in Figure 2, to the world.

Figure 2. Advertisement for 1960's Picturephone.

This device compressed a video signal such that it could be carried over the digital equivalent of 100 telephone lines. It was designed to be a personal conferencing tool, used in conjunction with the telephone.

Nippon Electric Corporation and British Telecom developed group videoconferencing systems in the 70's. These early systems worked with analog signals, because networks deployed at that time could not provide the required bandwidth to transmit video signals in digital form.

In the early 70's and into the 80's, expensive room-based videoconferencing tools were the norm. The price of only the "codec" (coder/decoder) was $250,000. The additional costs for cameras, speakers, lighting equipment, microphones and other required equipment put videoconferencing out of the reach of all but the wealthiest companies.

The VLSI (Very Large Scale Integration) circuit technology breakthroughs made in the 80's had a major impact on the videoconferencing industry. The availability of cheap memory and powerful microprocessors allowed images to be stored and processed, steps necessary for data rate compression. This made it possible to reduce the bandwidth of the video signal to the point where it was able to be transported over available networks.

For years the videoconferencing market was plagued by high costs, low quality, and the inability of different products to work together. Due to innovations in networking and computer technologies, the quality has improved while the prices have continued to drop.

Videoconferencing was a tool used primarily by the upper echelons of the business world until 1990. Also, it was virtually unavailable on the desktop until 1994. VTEL, founded in 1985, offered the first PC-based system.

Two other pioneering companies in the field of video compression were Compression Labs Incorporated, founded in 1976, and PictureTel, established in 1984. PictureTel developed the first software based videoconferencing system to become commercially available. It required a mere 224 Kbps bandwidth, revolutionary at the time.

In December of 1990 an interoperability videoconferencing standard was accepted by CCITT (now ITU). This allowed compatibility of products among different vendors. It was designed mainly for the group-based systems.

Before standardization, all parties involved would have to be using the same manufacturer's product, on a similar network. Clearly, this was not compatible with the idea of videoconferencing between people from different organizations, such as for a sales presentation. Only those closely affiliated could be included. The adoption of standards has been a stimulus to the videoconferencing market, as the opportunities for and rewards of true interconnectivity become apparent. Today, people are able to videoconference over different network configurations, using different, standard compliant products.

The cost of videoconferencing tools continues to plummet. Indeed, some vendors have software products that can be downloaded for free from the Internet. This, coupled with the availability, and popularity of multimedia-enabled PCs with fast processors and plentiful memory, is pushing the videoconferencing market to the desktop, supplanting the room-based systems. One can find the rudimentary personal conferencing capabilities bundled with the operating system, preinstalled on the consumer personal computer.

More inclusive standards were ratified by the International Telecommunication Union's Telecommunications Standardization Sector (ITU-T) in 1996, providing for interoperability among devices operating over various networks. One of these standards, H.324, defines videoconferencing over standard phone lines (POTS). H.323 addresses the requirements of videoconferencing communications over local area networks (LANs) and the Internet. The adoption of these standards has set the stage for a skyrocketing PC-based personal conferencing market.

3. AUDIO AND VIDEO FUNDAMENTALS

The bandwidth requirements of uncompressed video data are immense; therefore it is necessary to perform compression prior to transmission over communication channels. The compression must occur in real-time to satisfy the strict timing constraints of the videoconferencing session.

The audio and video signals must be converted from analog form to digital form and stored digitally to be manipulated by the computer. The codec (COder/DECoder), the centerpiece of any videoconferencing system, performs the function of coding, decoding, compressing, and decompressing the video and audio.

3.1 HARDWARE vs. SOFTWARE CODECS

Codecs may be implemented in either hardware or software. Deploying hardware codecs on a network requires disassembly of each computer which becomes tedious when working with a large system of many computers.

A software codec is usually easier to deploy, but puts a burden on the host's CPU, potentially making it problematic to concurrently run other applications. Hardware codecs have their own processors, allowing them to relieve the host processor of such duties as call setup, compression and decompression, and echo-cancellation. Clearly, there are other factors to consider when determining which to use, such as RAM, the speed of the network, the quality of the video capture card, among others. For example, if the networking environment is a 28.8 POTS connection, a hardware codec won't offer any improvement.

Software codecs are far less expensive than hardware codecs. In fact, both Microsoft and Intel have versions, albeit minimally equipped, that can be downloaded free from the Internet. The hardware variety comes in a wide range of prices.

3.2 AUDIO CODING AND COMPRESSION

3.2.1 Pulse Code Modulation

Digitizing an analog audio signal is done through a process called Pulse Code Modulation (PCM). The sampling rate and the number of bits per sample are defined parameters of the process. The sampling rate is the number of samples per second. Bits per sample is the number of bits used to represent each sample value. Additionally, the number of channels can be one for monaural, two for stereo, and so on.

Digitizing the audio signal requires the analog audio signal to be passed through several stages. First, the signal is low-pass filtered to remove high frequencies. Any audio signals present, which have frequencies above 1/2 of the sample rate, will be "aliased", or translated to lower frequencies, and result in a distorted reproduction of the original sound. Due to the finite rolloff of analog filters, it is a practical necessity to sample at more than twice the maximum audio frequency of interest. Although humans can perceive sound frequencies between 20 Hz and 20 kHz, videoconferencing systems are typically designed to handle speech quality audio which encompasses a much smaller range of frequencies. Telephone quality audio extends from about 300 to 3300 kHz.

After filtering, the signal is sampled. The amplitude values of an analog audio signal, representing the loudness of the signal, are continuously varying in time. To encode this signal digitally, the amplitude value of the signal is measured (sampled) at regular intervals. According to the Nyquist Theorem, to have lossless digital representation of the analog signal, the sampling rate must be at least twice that of the highest frequency present in the analog waveform. This is termed the “Nyquist rate”.

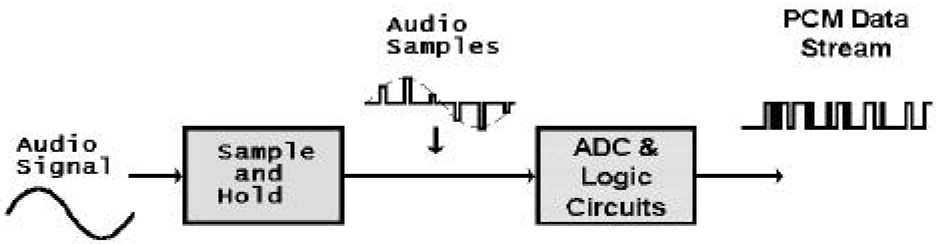


Figure 3. PCM encoder simplified block diagram.

The sampled value is then quantized. This requires the value to be mapped into one of a set of fixed values, which are binary coded for transmission. See Figure 3 for a diagram of the encoding process.

The errors, which result from this mapping of analog values to quantized levels, result in “quantization noise”. It follows that the level of quantization noise drops as the quantization levels become closer together. Therefore, more bits of quantizer resolution translate to less quantizer noise, and hence greater dynamic range.

Some implementations have a linear quantization. Other approaches may use a logarithmic quantization, which results in a type of audio compression, reducing quantization noise during quiet intervals without wasting bits at higher signal levels.

The total bit rate for a monaural PCM signal can be found by multiplying the sample rate by the number of bits/sample. A stereo signal would require twice the bit rate, etc.

Because of the fact that telephone quality voice transmission requires about 3 kHz of audio bandwidth and 256 quantum levels, a sample rate of 8 kHz and 8 bits per sample are commonly used, resulting in the 64 kbps channels used for ISDN and other phone applications.

3.2.2 ADPCM

Adaptive Differential Pulse Code Modulation is a compressed version of PCM, which requires a lower bit rate than standard PCM to transmit the same voice information.

In DPCM, previous PCM samples are used to predict the value of the current sample. It is possible to do this because of the patterns present in speech samples. This prediction algorithm is performed at both the transmitting and receiving end. The transmitter compares the actual sample to its predicted value, and computes the error. Because the error signal will have a lower variance than the original speech samples, it can be quantized with fewer bits than the original speech signal. This error signal is then transmitted.

Because the prediction algorithm is performed at the receiving end as well, the receiver knows what the predicted value is. It uses the error signal to correct the predicted value and reproduce the original sample. This predict-compare-adjust process is repeated for each input sample, reproducing the original PCM samples at the output.

The system is called “adaptive” because the prediction parameters and the quantization levels of the error signal can change dynamically depending on the rate of change of the sample values (i.e. signal level).

Many ITU-T videoconferencing recommendations include ADPCM encoding methods. Different flavors of ADPCM encoder/decoders vary in the way the predicted value is calculated and how the predictor or quantizer adapts to signal characteristics. This results in various levels of compression. Standards include G.721, G.722, G.723, G.726, and G.729. Various characteristics of these standards are summarized in Table 1.

Higher quality speech (50 Hz -7 kHz, 14 bit resolution) may be encoded by dividing the audio spectrum into two subbands and performing separate ADPCM coding on each. The technique is covered in G.722 and is called “Sub-Band ADPCM”. G.722 specifies three modes of operation: 64, 56 and 48 kbps.

3.2.3 LPC/CELP/ACELP

LPC (Linear Predictive Coding) is used to compress audio at 16 Kbps and below. An LPC encoder fits speech signals to a simple analytic model of the vocal tract. The signal is broken into frames, usually tens of milliseconds long, and best fit parameters for each frame are transmitted and used by the decoder to generate synthetic speech that is similar to the original. The result is intelligible but artificial sounding.

Plain LPC is not included in videoconferencing standards, but is the basis for CELP (Code Excited Linear Prediction) which is important for obtaining high audio compression rates in videoconferencing. CELP is quite similar to LPC. The CELP encoder does the same frame-based LPC modeling but then computes the errors between the original speech and the synthetic model and transmits both model parameters and the errors. The error signal actually represents indices in a “codebook” of “excitation vectors” shared by the encoders and decoder. Thus the error signal is very much compressed. It follows that the computational complexity and speech quality of the coder depend upon the search sizes of the code books, which can be reduced at the expense of sound quality.

CELP makes possible much higher quality speech at very low data rates. ITU-T Recommendation G.728 uses a variation of CELP, LD-CELP (Low Delay CELP). The compressed audio requires a bandwidth of only 16 kbps, but the encoder and decoder are quite computationally complex, requiring special hardware.

Table 1. Audio Standards G Family

ITU Standard	Year Approved	Algorithm Used	Bit Rate	Bandwidth (kHz)	Typical End-To-End Delay (ms)	Application
G.711	1977	PCM	48, 56, 64	3	<<1	GSTN telephony, H.323 & H.320 videoconferencing
G.723	1995	MPE/ACELP	5.3, 6.3	3	67-97	GSTN videotelephony, H.323 telephony
G.728	1992	LD-CELP	16	3	<<2	GSTN, H.320 videoconferencing
G.729	1995	ACELP	8	3	25-35	GSTN telephony, modem h.324 GSTN videophone
G.722	1988	subband ADPCM	48, 56, 64	7	<2	ISDN videoconferencing

3.3 VIDEO COMPRESSION

Since network bandwidth is in limited quantity, and video is inherently bandwidth thirsty, the choice of video compression technique takes on great importance.

Video is composed of a sequence of still images, called frames. The sequence of frames are presented at a rate that makes the motion of the depicted video scene appear fluid. The frame rate for television in the United States is 30 frames per second. The frame rate in a business quality videoconferencing session should be at least 15 frames per second. At lower rates the video will appear jerky.

Each frame of the video is digitally represented as a two dimensional matrix of pixels. Color images are composed of three image frames, one for each color component.

Video compression is typically lossy, meaning some of the information is lost during the compression step. The compression process takes advantage of the functioning of human vision, discarding information that is not perceptible.

Further compression can be achieved, further reducing the required bandwidth, but at the sacrifice of quality. The required level of quality will depend on the application.

Color space sampling and redundancy reduction is techniques common to most video codecs. Color space sampling is a technique used to reduce the amount of data that needs to be encoded. Because the human eye is less sensitive to chrominance information, an image encoded in YUV space can have the U and V components subsampled. In this way, these components will require one half, or less, of the bits required to encode the more important Y component. Redundancy reduction is also used to decrease the amount of encoded information. Intraframe encoding achieves compression by reducing the spatial redundancy within a picture. This technique takes advantage of the fact that neighboring pixels in an image are usually similar.

Further compression is achieved through interframe encoding, which uses the fact that neighboring frames in a sequence of images are usually similar, by reducing the temporal redundancy between frames.

3.3.1 Discrete Cosine Transform

Discrete Cosine Transform is a video compression technique that forms the basis for the two important video compression standards, H.261 and H.263. This compression algorithm is also used by the Joint Photographic Experts Group (JPEG) standard for still-image compression. DCT is an intraframe spatial compression technique that converts pixel values into their frequency-based equivalents.

The first part of the algorithm transforms the pixel value information into values in the frequency domain. The encoding process then codes the higher frequency components with less fidelity than the lower ones, since the human eye is more responsive to low frequencies than to high ones.

The DCT algorithm begins by dividing a frame into eight-by-eight blocks of pixels. The DC coefficient is the first value produced by the transform. Its value represents the average luminance for the entire block. The remaining 63 coefficients, called AC coefficients, are calculated, concentrating the low-frequency components in the upper left corner, and the high-frequency components in the lower right corner. The low frequency components describe detail shifts in the image such as edges and color changes, while the higher frequencies describe larger areas of uniform color. No information is lost during the DCT step of the algorithm; the data could be recovered by performing an inverse DCT.

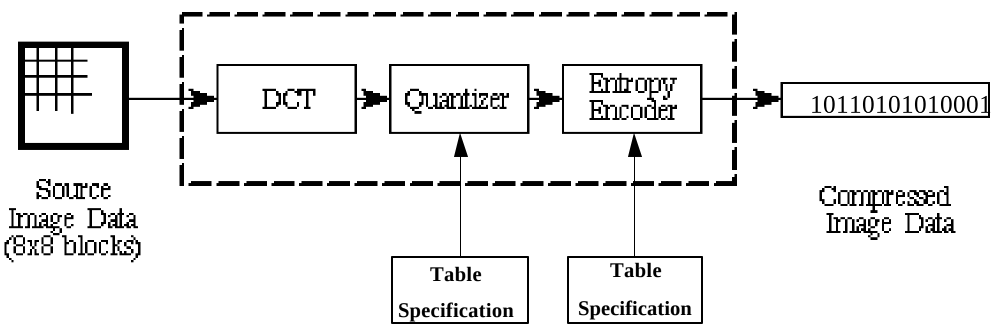


Figure 4. DCT encoder block diagram.

3.3.2 Quantization

The next step is quantization; it is the lossy step of the encoding process. First, an 8x8-quantization matrix is generated. The values in this matrix are calculated from a user defined quality factor. Each coefficient is divided by its corresponding value of the quantization matrix. The purpose of quantizing is to reduce the amplitude of the coefficients and to increase the number of zero value coefficients. The quantization matrix is designed so that the values increase diagonally right. This is so that the value of more of the higher frequency AC components, which are less visually significant, will go to zero.

3.3.3 Run-length and Entropy Encoding

This bit stream undergoes further compression through runlength and entropy coding. The run-length encoding removes long runs of zero valued coefficients. The entropy encoding process encodes the information efficiently based on statistical characteristics. Patterns that are statistically more likely to occur are encoded with less bits. Huffman entropy encoding is used in the H.261 and H.263 standard codecs.

The DC component undergoes differential encoding. It is encoded as the difference between the current DC coefficient and the DC coefficient of the previous block. Next, the AC coefficients are encoded in a zigzag sequence from the upper left of the block to the bottom right. The AC coefficients are then run-length encoded and entropy encoded to produce a binary coded stream. Figure 4 shows a block diagram of the DCT encoding process.

At the receiving end the decoder will reverse the encoding process. The decoder will be provided the same quantization and entropy tables that were used to encode the data.

This type of intraframe compression can achieve compression on the order of 15:1, with excellent quality. As compression increases, the quality of the image decreases.

3.3.4 Motion Compensation

A sequence of frames from a video is full of redundancy; there is little noticeable difference between subsequent frames. The purpose of interframe coding is to remove this temporal redundancy, and vastly improve the compression ratio. Motion Compensation is an interframe coding technique that involves working with 16 by 16 macroblocks to identify a group in the previous frame that best matches a group on the current frame. This difference between the frames is coded along with a vector that describes the offset of that group. This data is then entropy coded, to achieve yet further compression.

3.4 ITU-T VIDEO COMPRESSION STANDARDS

3.4.1 ITU-T Standard H.261

The H.261 is a video codec for audiovisual services at p x 64 kbps (p=1,2,3...30), and uses a DCT based codec. H.261 supports two image formats, Common Interchange Format (CIF), and quarter CIF (QCIF). CIF is 352 by 288 pixels; QCIF is 176 by 144 pixels. Support of QCIF is mandatory, and support of CIF is optional.

The standard does not define a required frame rate for compression, however for a codec to be H.261 compliant, it must decode up to and including 30 frames per second. The DCT algorithm is used for intraframe coding, but only certain frames of the video sequence are fully encoded in this way. For interframe coding, the H.261 standard calls for coding only the difference between a frame and the previous frame.

The more complex and processing-intensive motion compensation interframe coding is only optional for H.261 coders, however the H.261 decoder must be able to decode motion compensation information. DCT intraframe coding is mandatory for both the coder and decoder.

Typical compression ratios are around 80 to 100:1. However, compression ratios can go as high as 500:1, depending on the video.

3.4.2 ITU-T Standard H.263

The H.263 is a video codec for narrow telecommunications channels (less than 64 Kbps). The formal name of H.263 is “Video Coding for Low Bit Rate Communication.” It is a backward compatible refinement of the H.261 standard, adding many performance and error recovery improvements.

Like H.261 the support of the QCIF format is mandatory. H.263 also optionally supports several other formats. Sub Quarter Common Intermediate Format (sub-QCIF) is supported for very low-resolution images. Also supported are 4CIF, which has four times the resolution as CIF, and 16CIF, which has sixteen times the resolution of CIF. Table 2 presents a summary of the video picture formats.

Table 2. Video Picture Formats.

Picture Format	Number of Luminance Lines	Number of Luminance Pixels	Number of Chrominance Lines	Number of Chrominance Pixels	Supported in H.261	Supported in H.263
sub-QCIF	96	128	48	64	not supported	optional
QCIF	144	176	72	88	optional	mandatory
CIF	288	352	144	176	mandatory	mandatory
4CIF	576	704	288	352	not supported	optional
16CIF	1152	1408	576	704	not supported	optional

Where the H.261 standard was limited to full pixel precision for motion compensation, H.263 has required support of half-pixel precision. The half-pixel refinement greatly improves the picture quality, particularly in low-resolution video.

New to the H.263 standard are negotiable coding options offering improved performance. One of these options is the support of P-B frames used in interframe encoding. This is a technique that is also used in MPEG video. Although it is computationally more expensive, it allows for much higher compression, and therefore a potentially higher frame rate.

4. COMPONENTS AND FUNCTIONS OF A VIDEOCONFERENCING SYSTEM

A typical videoconferencing system and its components are shown in Figure 5. Videoconferencing stations are equipped with video and audio capture and compress subsystems, and decompress and display subsystems. The communication media can be POTS (Plain Old Telephone System), LANs, or WAN.

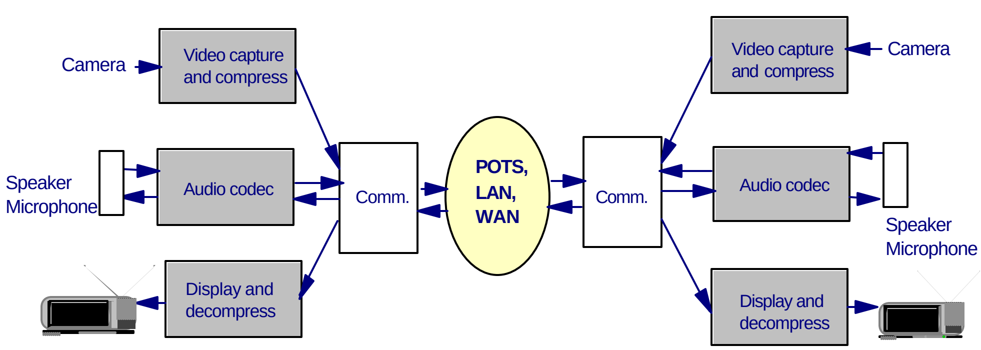


Figure 5. Components of a videoconferencing system.

The common software architecture of a videoconferencing system, shown in Figure 6, consists of videoconferencing application, middleware, video and audio codec and data, stream multiplexer and demultiplexer, and line drivers.

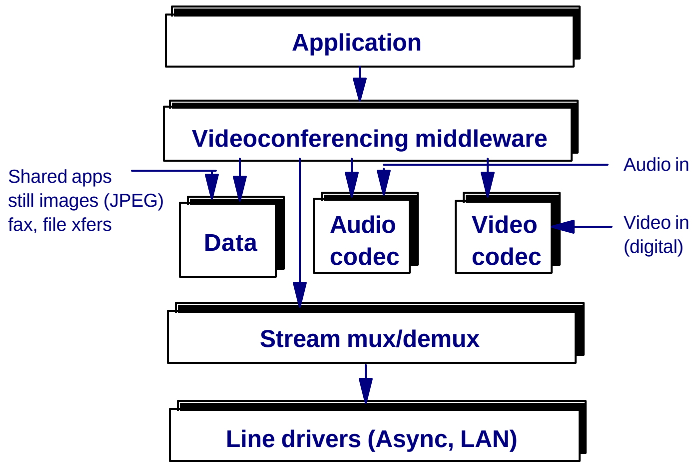


Figure 6. Common software architecture of a videoconferencing system.

Videoconferencing systems support a number of functions including:

Multipoint connection set up. The system should be able to negotiate for network resources and end-to-end conference capabilities.

Dynamic session control. The system should have an ability to add and delete participant to/from an existing videoconference. Such a change may require modification of the underlying network configuration.

Conference directory. The system should provide a conference directory service that will support conference registration, announcement, and query. The directory should contain various information such as: the title and the brief description of the conference, a list of participants, start and end time for the conference, audio and video coding schemes, their protocols and QOS requirements, and shared working space.

Automatic conference scheduling and recording. The conference scheduling function combined with resource reservation mechanisms will allow planning of network resources. Automatic conference recording is a useful function that does recording of conference sessions in a form of multimedia documents.

Conference termination. The system should be capable to release all reserved resources when the conference is complete.

5. VIDEOCONFERENCING STANDARDS AND NETWORKS

5.1 VIDEOCONFERENCING STANDARDS

The original, widely accepted videoconferencing standard was H.320, which defines a methodology for transporting videoconferencing traffic over ISDN. However, in 1996, additional standards for videoconferencing emerged - H.323, H.321, and H.324. These standards define methodologies for videoconferencing over other various networks such as POTS, IP networks such as LANs and the Internet, and ATM networks. Each of these standards brings with it certain capabilities and quality levels, illustrated in Figure 7. Each has advantages and disadvantages in videoconferencing transmission. The various characteristics of these standards are summarized is Table 3, and discussed in the following sections.

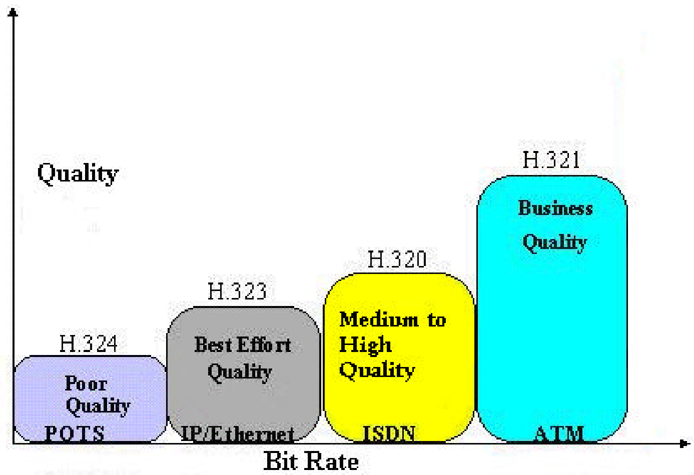


Figure 7. Quality of videoconferencing standards.

Videoconferencing involves the transmission of video, audio, and data. Data can be in the form of whiteboard data or shared application data, used in a collaborative conference. These different types of information have different reliability and delay variation requirements of the networks over which they are being transmitted.

Table 3. Characteristics of Various Videoconferencing Standards

	H.320	**H.321**	**H.323**	**H.324**
Approval Date	1990	1995	1996/1998	1996
Network	Narrowband Switched digital ISDN	Broadband ISDN ATM LAN	Non-guaranteed bandwidth packet switched networks	POTS, the analog phone system
Video	H.261 H.263	H.261 H.263	H.261 H.263	H.261 H.263
Audio	G.711 G.722 G.728	G.711 G.722 G.728	G.711 G.722 G.728 G.723 G.729	G.723
Multiplexing	H.221	H.221	H.225.0	H.223
Control	H.230	H.242	H.242 H.230	H.245
Multipoint	H.231 H.243	H.231 H.243	H.323	
Data	T.120	T.120	T.120	T.120
Communication Interface	I.400	AAL I.363 AJM I.361 PHY I.400	TCP/IP	V.34 Modem

Video and audio data are delay sensitive; they can't tolerate much deviation. They need to arrive at a constant rate with little variance to be presentable and intelligible. Large delays, known as jitter, will cause the picture to appear "jerky". The audio will sound unnatural if it encounters network delays. Video and audio are less sensitive to reliability.
Uncompressed video can tolerate some lost or corrupt frames without severely affecting the quality of the presentation. A corrupt frame will simply be replaced by the next frame of the video.

Compressed video may suffer from unreliable transmission. This is because in interframe encoding, a video compression technique, redundant information has been removed and just the differences between the frames are transmitted. This interdependency could cause problems in an unreliable transmission environment. However, video compression techniques have been designed to work around this problem. This can be compensated for by periodically sending complete information about a frame, even if the data in these blocks has not changed.

Conversely, general data is not sensitive to delay but is sensitive to reliability. An example is a data file that is sent over a network. Since it doesn't have the same strict timing constraints of the audio and video, it doesn't matter how long the file takes to get to its destination. But the data information in the file needs to be correct, as any transmission errors may render it useless.

5.2 NETWORK CONFIGURATIONS FOR MULTIMEDIA CONFERENCING

Several network configurations are used for multimedia conferencing (Figure 8):

- Fully distributed (mesh) network,
- Centralized (star) network,
- Double-star network, and
- Hierarchical network.

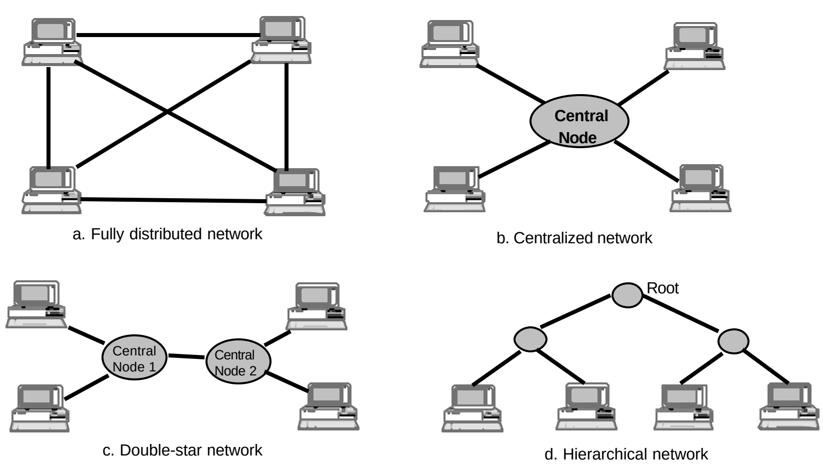


Figure 8. Network configurations for videoconferencing: (a) distributed network (mesh), (b) centralized network (star), (c) double-star network, and (d) hierarchical network.

The fully distributed network is based on multiple point-to-point links, and connects each participant with each other. Each participant sends the media directly to every other participant. The processing and mixing media is done at each individual participant's location. This configuration uses the highest network bandwidth and duplicates the work involved in mixing data at every participant's location. On the other hand, this system gives the best (shortest) delay time when considering media transmission and mixing. Its major disadvantage is that with increased number of conference participants, the point-to-point connections increase rapidly.

The centralized network consists of a central node connected to every participant of the conference. The central node acts as intermediate processor and performs media mixing. It received multimedia data from participants, mixes or processes the media, and then broadcasts the media back to participants. The centralized system has an advantage that the mixing and processing of media is only done once within the system. The disadvantage of the system is an increased delay time when transmitting composite media from one to another conference participant, since the intermediate processor must wait until all media is received before it begins media mixing and broadcasting.

The double-star network is extension of the centralized network. In this system, a central node from one star network is connected to another central node of another star network. The central node is used as a concentrator for several sites communicating via a single bi-directional link with several sites connected to a second central node.

The hierarchical network is another extension of the centralized network. The system consists of a series of intermediate nodes, with one root node and other as internal nodes in a tree structure. All intermediate nodes are capable of performing mixing and processing of data, while leaves in the tree are the conference participants. The multimedia data is sent up to the tree for mixing and processing. A mixer receives the media from multiple leaves or mixers below it, and then transmits the mixed media to the mixer above. The completely mixed data is generated by the root node and then broadcasted directly from the root to the leaves of the tree involved in the conference. This configuration reduces the network traffic, and therefore the system is capable of handling larger number of conference participants than either centralized or distributed networks.

Videoconferencing poses very strict timing requirements. The one-way end-to-end delay is defined from the time when the conference participant moves or speaks until the time the motion or sound is perceived by the other participants. This time should be less than 150 milliseconds, which gives the total round trip delay less than 300 milliseconds for maintaining conversation "face-to-face".

The total one-way time delay comprises of four major components: sending workstation operation, sending workstation transmission time, network delay, and receiving workstation, as illustrated in Figure 9.

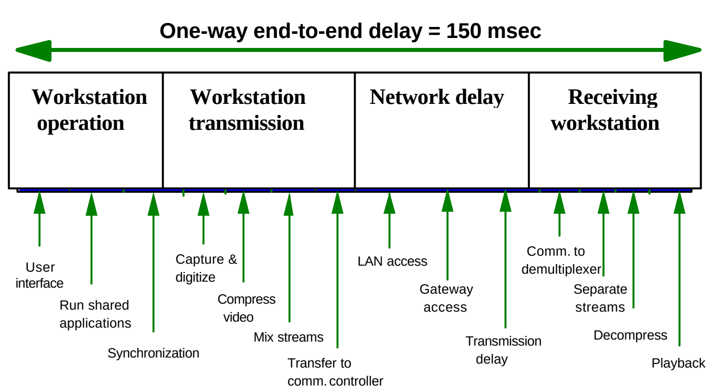


Figure 9. The one-way end-to-end delay in videoconferencing.

The delay between audio and video is very critical and limited to 20 to 40 milliseconds in order to provide good lip synchronization.

5.3 CIRCUIT SWITCHING vs. PACKET SWITCHING NETWORKS

Circuit-switched communication is a manner of data transmission where the communication path is established and reserved for the duration of the session. The bandwidth allocated for a session is used exclusively by it alone. The resources used by the session are freed, and available for other calls, at the end of the session. The dedicated bandwidth is an advantage for videoconferencing, providing predictable delays. However, circuit-switching transmission underutilizes network resources, as the dedicated bandwidth can go unused at times of limited activity.

Unlike circuit-switched transmission, a packet-switching environment has no dedicated bandwidth circuit set up; the bandwidth is shared with other network users. The information is divided into packets, and each packet is routed individually through the network. Since the packets may take different routes, they may arrive at their destination at different times and out-of-order. This variable delay in the delivery of information can cause problems in the quality of videoconferencing applications. Video packets received out-of-order may have to be discarded. Various protocols have been developed to try to overcome these inherent problems associated with packet switching such as RSVP, and RTP. These attempt to provide some quality of service over this type of transmission, and are discussed in later sections.

5.4 ISDN VIDEOCONFERENCING

Integrated Services Digital Network, ISDN, is a circuit switched end-to-end digital service. ISDN was designed to support the transmission of a wide variety of services, including voice, data, and video at high speeds over the public switched telephone network (PSTN). ISDN relies on 64 kbps channels, originally chosen to support digitized voice traffic.

Two access rates are defined for ISDN: Basic Rate Interface (BRI) and Primary Rate Interface (PRI). The user information is carried over bearer channels, known as B channels, each having a capacity of 64 kbps. The Basic Rate Interface provides 2 B-channels, while the Primary Rate Interface provides 23 B channels, in North America and Japan, and 30 B channels in Europe. A separate channel, the D-channel, is used for signaling. The D-channel has a data rate of 16 kbps in the BRI and 64 kbps in the PRI.

While a single BRI service is not sufficient to support a business quality videoconference with a frame rate of 15 frames per second, this service does satisfy requirements for many videoconferencing applications, such as desktop videoconferencing. A business quality videoconference will require at least 6 B channels with a data rate of 384 kbps (6 x 64). This setup is normally only seen in group based systems. In the past, ISDN videoconferences were just point-to-point connections. However, today it used for multipoint as well, by using a Multipoint Control Unit (MCU). Figure 10 illustrates a typical ISDN videoconferencing session.

Despite its reputation for difficult installation, lack of availability, and expense, many videoconferencing products on the market utilize ISDN. One of the main reasons for ISDN videoconferencing's broad acceptance is the timely emergence of ITU standards supporting it in 1990. The standards for other communication networks were not seen until 1996. However, with so many changes taking place, and the ratification of further standards for videoconferencing over such diverse networks as POTS, and IP based networks such as LANs and the Internet, ISDN's days as the dominant videoconferencing communications medium are over. While ISDN will continue to provide the quality required for business applications, POTS and Internet products have become more widespread among home and less serious users.

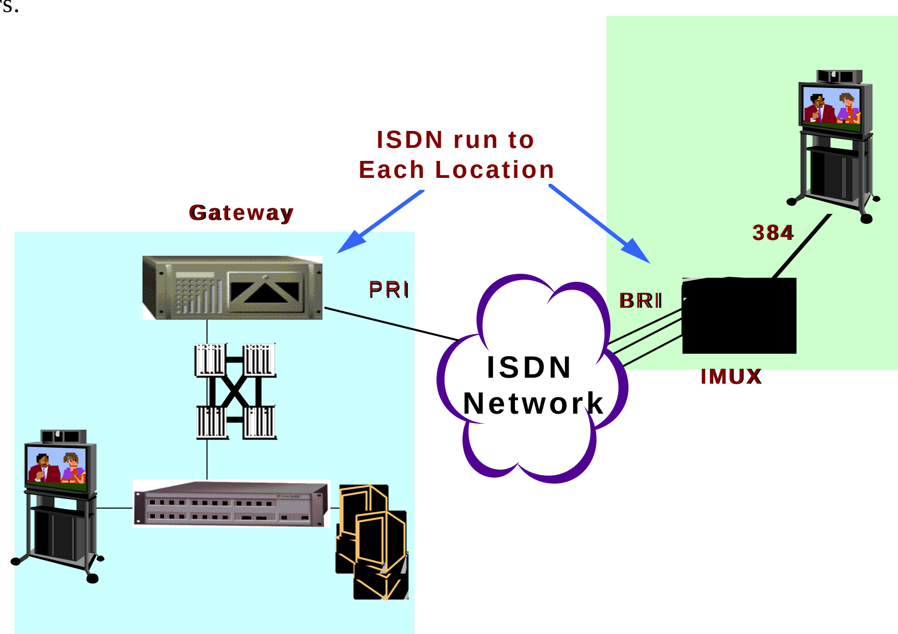


Figure 10. Example of an ISDN videoconference connection.

ITU-T Standard H.320

ITU-T's ratification of standard H.320, titled "Narrow-Band Visual Telephone Systems and Terminal Equipment", in December 1990, gave the videoconferencing industry a much needed jump-start. This standard was optimized for popular circuit-switched media such as ISDN. For the first time, a standard existed that made it possible to link equipment from different vendors in the same conference.

H.320 is an umbrella standard. It contains other standards concerning audio and video encoding and compressing, specifications for how calls are negotiated, how the data is multiplexed and framed, multipoint conferencing, data transmission, and communication interfaces.

5.5 PLAIN OLD TELEPHONE SERVICE (POTS) VIDEOCONFERENCING

The telephone system gradually began converting its internal connections, once purely an analog system, to a packet-based, digital switching system in the 1960s. Today, nearly all voice switching in the US is digital within the telephone network. However, the final link from the local central office to the customer site remains primarily an analog line, although it too is being replaced. This connection is often called Plain-Old Telephone Service (POTS).

The standard telephone system is the most widely pervasive network in existence. The availability of this existing infrastructure, combined with the latest CPU's, compression techniques, and advanced modem technologies, has brought videoconferencing to the home consumer. In May of 1996 the ITU-T ratified the H.324 standard for videoconferencing, which defines a methodology for transporting videoconferencing across POTS.

POTS videoconferencing has been technically difficult because of the low bandwidth of audio phone lines. However, with improved coding methods, there is sufficient bandwidth available to support audio, video and data sharing with this media. And the equipment is inexpensive and easily installed, relative to other existing methods.

Although most Pentium PC's can digitize local video at 15 fps and higher with little problem, quality suffers because the POTS modems cannot transmit the data fast enough to maintain the same frame rate. This typically restricts POTS-based videoconferencing to just a few frames per second. Thus, POTS videoconferencing does not approach the levels of quality required for even the most casual of business needs, and is better suited for home and recreational uses.

A POTS connection can be established either modem-to-modem, or over the Internet.
An Internet POTS connection requires the user to first establish a dial-up connection to the Internet. A call is then made to someone else on the Internet, who is also probably using a modem. The quality of the connection varies, as Internet traffic can severely compromise the quality of the videoconference. A typical frame rate is 2 to 5 frames per second, and basically looks like a series of stills. The advantage to this method is that a call can be made to anyone in the world for the price of accessing the Internet. A modem-to-modem connection can achieve frame rates of 5 to 10 frames per second, because the Internet traffic is avoided. But, long distance phone charges will apply if calling out of area.

POTS videoconferencing is finding acceptance in the consumer marketplace, due to the ubiquitous nature of the phone system. Also, recreational videoconferences do not require the same quality level of a live business meeting or a training class. Furthermore, today's PCs come multimedia equipped, with modems, are inexpensive, and readily available.

ITU-T Standard H.324

The H.324 ITU standard, titled "Multimedia Terminal for Low Bitrate Visual Telephone services over the GSTN", was the first standard to support point-to-point video and audio compression over analog POTS lines. Specifically, H.324 is designed to optimize videoconferencing quality over the low-speed links associated with the POTS system, typically operating at the speeds of modems - 28.8 kbps - 56 kbps. This standard allows users to interoperate across diverse endpoint such as ISDN, ATM, POTS, or mobile devices, and makes it possible to hold modem-based, POTS video calls that connect equipment made by different vendors. H.324 terminals may carry real-time voice, data, and video, or any combination, including videotelephony.

H.324 is an umbrella recommendation with a similar structure to H.320, its ISDN counterpart. See Section 4.7, Summary of Standards, for specifics on the various standards that H.324 supports.

5.6 IP-BASED VIDEOCONFERENCING

Networks are a fundamental part of today's information systems. They form the backbone for information sharing in enterprises, governmental and scientific groups. This shared information comes in a variety of forms such as e-mail and documents, files sent to colleagues, and real-time applications such as videoconferencing.

Local Area Networks (LANs) are commonly used on campuses and in companies to connect desktop computers together. At the physical layer, LAN's are frame-based and usually consist of 10 Mbps Ethernet, 16 Mbps Token Ring segments, or ever 100 Mbps Fast or Switched Ethernet.

Ethernet and Token Ring networks differ in the way that clients gain access to the transmission medium. 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 must wait a random amount of time before transmitting again. Token Ring is a network where a token is passed around and clients must gain access to the token before transmitting.

Internetworking is the process of connecting various networks together. These different networks may have different network technologies and different network protocols. In the 70's, the development of technologies and protocols for internetworking were initiated by the US Defense Advanced Research Projects Agency (DARPA). The communications protocol developed under this project, called the IP (Internet Protocol), allow applications the ability to talk to other applications, regardless of the underlying network technology. IP is actually a suite, or stack, of protocols. These standards have been adopted as the standard protocols for internetworking.

The *Internet* resulted as new wide area networks were created in the US and the rest of the world and became interconnected using the IP stack of protocols. The Internet (with a capital I) refers to a worldwide set of interconnected networks. The Internet is the most widely used universal information network.

Two different transport layer protocols were developed for use with IP. TCP (Transmission Control Protocol) will be familiar to most readers and is the protocol commonly used for Email, FTP, and Web Surfing. TCP/IP is capable of transporting large amounts of data between computers, with 100 percent accuracy, but can delay transmission and reduce throughput, as error checking occurs at every node in the transmission path. Any corrupt or lost packets are retransmitted, and if congestion is encountered, the packets are rerouted. Because TCP/IP has variable length packets, each device along the network needs to determine the length of each packet, it copies the whole packet into memory before forwarding it, slowing transmission.

Although these attributes are critical for effective data transmission, the are not conducive to the real time video and audio streams needed for videoconferencing, where data must arrive consecutively and in a timely manner, or it's useless. If lost packets were retransmitted, they would arrive too late to be of any use.

In addition to TCP, the IP allows for a second mode of data transfer, specified by the UDP (User Datagram Protocol). In contrast to TCP, UDP offers no guarantee of packet delivery, and packets can be dropped when the network is busy, with no way for any network element to ask for retransmission. It promises only "best effort" service quality. For video, it is preferable to miss a frame in order to maintain the overall flow.

There is, in essence, a tradeoff between reliability and timeliness; timeliness must win for IP videoconferencing, which abandons TCP for audio and video and instead specifies the UDP. However, TCP is used for data transfer, the control channel and the call signaling channel.

5.6.1 Disadvantages of Using IP for Videoconferencing

IP networks do not have an underlying QoS architecture on which to base video transport. Even switched Ethernet infrastructures are quite unpredictable in nature, especially on network trunk links and through routed backbones, and this leads to increased delay and jitter, greatly decreasing picture quality. This effect is most visible when using the Internet.

Video on Ethernet infrastructures also has the unfortunate side effect of permitting interaction between the data and the video traffic. The bandwidth requirements of the video traffic slow down the data traffic on the network, the best effort nature of the transmission may impact the quality of the communication. Management of network resources becomes a key element of IP videoconferencing standards, as bandwidth may be reserved for videoconferences or, in fact, held from them so that other network applications may still function.

Expensive, high-bandwidth networks must be obtained to reduce the likelihood of latencies and bottlenecks. Currently, the Internet simply does not offer the performance required for high quality videoconferencing. Often, video over the Internet looks more like a still picture, has little or no motion and broken audio, and is pretty much a useless business tool. However, for personal use, or more informal business uses, Internet videoconferencing apparently does have a place, because its popularity is growing to the point that it may make POTS-based standards irrelevant.

5.6.2 Requirements for IP Videoconferencing

IP videoconferencing can make sense as a corporate solution if the network is sufficient. The first consideration is available bandwidth because this determines audio and video quality. 10 Mbps Ethernet LANs are the minimum used in most companies today, and these have enough bandwidth to support desktop conferences. Multiple simultaneous calls require more bandwidth, and a 100Base-T backbone is the minimum requirement for handling many simultaneous calls. Larger organizations may require even more than 100M bit/sec backbone.

With a LAN offering significantly more bandwidth than ISDN, the video quality within a conference can be potentially much higher and can approach that of television. Communications advancements such as Fast Ethernet (100 Mbps) and ATM (155 Mbps) have increased available bandwidth, and Multicast technology has reduced network loading in conferences involving more than two participants.

However, as IP videoconferencing becomes more popular, existing networks may become bogged down with its traffic, degrading the quality of the video and audio, but also slowing other network applications such as web browsing. Most organizations have less bandwidth over the wide area network as compared to the LAN, so this problem will be grow especially acute if simultaneous multiple “external” videoconferences are required.

One alternative for communicating with distant locations is through ISDN gateways. These devices can convert LAN systems to H.320 ISDN and call out to circuit-switched locations. This gives users the flexibility of H.320 compatibility as well as the option of dial-up access to remote locations.

Despite the disadvantages of implementing videoconferencing over IP networks, the ready accessibility and low cost have moved the market in this direction.

5.6.3 RSVP

Ethernet's inherent lack of Quality of Service has prompted the designers of this network transport system to go back to the design board and attempt to retrofit Ethernet with QoS capabilities. This new protocol, called the Resource Reservation Protocol (RSVP), is a signaling system designed to enable the end points of a network transaction to request network bandwidth for particular communication streams and receive a reply indicating whether the request has been granted. Many IP-based videoconferencing products support RSVP, which is likely to become a widely accepted method of enabling existing network infrastructures to deliver some QoS- like capabilities.

However, RSVP cannot solve the fundamental QoS issues of frame based networks. Because it operates in the Ethernet environment, RSVP must deal with variable length frames. It is probable that small 'latency sensitive' frames carrying videoconferencing could easily be stuck in a buffer behind much larger 'data' frames. This is the case for 10Base-T, 100Base-X and Gigabit Ethernet.

RSVP is a simple signaling system, where every hop-by-hop link must be negotiated separately. There is still no end-to-end guarantee of a minimum service level. ATM, on the other hand, sets up an end-to-end connection with a specific QoS class that commences at call set-up and ends at call teardown.

5.6.4 RTP

In conferences with multiple audio and video streams, unreliable transport via UDP uses IP Multicast and the Real-Time Protocol (RTP) developed by the Internet Engineering Task Force (IETF) to handle streaming audio and video. IP Multicast is a protocol for unreliable multicast transmission in UDP. RTP works on top of IP Multicast, and was designed to handle the requirements of streaming audio and video over the Internet. A header containing a time-stamp and a sequence number is added to each UDP packet.

With appropriate buffering at the receiving station, timing and sequence information allows the application to eliminate duplicate packets; reorder out-of-sequence packets; synchronize sound, video and data and achieve continuous playback in spite of varying latencies. RTP needs to be supported by Terminals, Gateways, and MCUs with Multipoint Processors.

5.6.5 RTCP

The Real-Time Control Protocol (RTCP) is used for the control of RTP. RTCP monitors the quality of service, conveys information about the session participants, and periodically distributes control packets containing quality information to all session participants through the same distribution mechanisms as the data packets.

5.6.6 Multicast

In some videoconferencing applications it is necessary to send the same real-time video and/or audio streams to multiple destinations throughout the global Internet. Typically this would be accomplished by sending multiple streams of redundant packets, one for each destination. This can be very inefficient and slow.

RTP-based applications can use “IP multicast” capabilities in conjunction with the MBONE (Multicast BackBONE), a virtual network designed to facilitate the efficient transmission of video and audio signals simultaneously over the Internet. The network is composed of “islands”, Internet sites supporting multicast and linked by virtual point-to-point links called "tunnels". The IP multicast packets are encapsulated for transmission through the tunnels, so that they look like normal unicast datagrams to intervening routers and subnets. Once they reach the destination island, they are copied and forwarded to destinations as required.

5.6.7 ITU-T Standard H.323

In 1996, the ITU ratified the H.323 standard for videoconferencing over packet-switched networks, such as Ethernet and Token-Ring, and ultimately the Internet. The standard is platform independent and runs on top of common network architectures.

To ensure that critical network traffic will not be disrupted by videoconferencing traffic, the standard includes network traffic management capabilities and supports multicast transport in multipoint conferences.

H.323 defines four major components for a network-based communications system: Terminals, Gateways, Gatekeepers, and Multipoint Control Units, shown in Figure 11.

Terminals are the client endpoints on the LAN that provide the user interface. All terminals must support voice communications; video and data are optional. Also required of H.323 terminals is support of H.245, for negotiation of channel usage and capabilities, Q.931 for call signaling and call setup, Registration/Admission/Status (RAS), a protocol used to communicate with a Gatekeeper; and support for RTP/RTCP.

Gateways provide translation between H.323 conferencing endpoints and other terminal types, such as H.320 and H.324. Gateways are not required if connections to other networks are not needed, since endpoints may directly communicate with other endpoints on the same LAN. Terminals communicate with Gateways using the H.245 and Q.931 protocols.

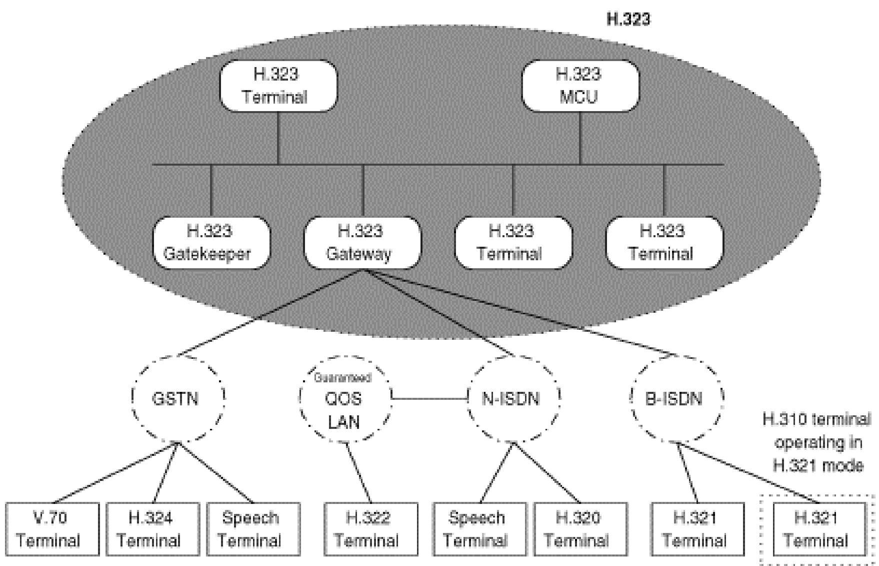


Figure 11. H.323 network and interfaces.

The **Gatekeeper** acts as the central point for all calls within its zone and provides call control services to registered endpoints. It translates addresses from LAN aliases for terminals and gateways to IP or IPX addresses, as defined in the RAS specification, and manages bandwidth. While a gatekeeper is an optional part of an H.323 system, all compliant terminals must support its use.

The **Multipoint Control Unit** (MCU) supports conferences between three or more endpoints, handling negotiations between all terminals to determine common capabilities for audio and video processing, and controlling multicast.

5.7 ATM-BASED VIDEOCONFERENCING

Unlike many specialized networks used in the past, Asynchronous Transfer Mode, ATM, is a networking technology that was designed at the outset to be service independent. The flexibility of ATM allows it to support a variety of services with different bit rates, burstiness, and acceptable delay characteristics. The bit rate can be constant for the whole transmission, or it can be variable over time. ATM satisfies all the various types of data, each with unique service requirements, found in videoconferencing such as video, audio and data.

ATM was designed to be efficient in the use of its available resources. All available resources are shared between all services, such that the optimal statistical sharing of the resources is achieved.

5.7.1 Basic Principles of ATM

ATM is based on the concept of asynchronous data transfer. It combines the best qualities of circuit-switched and packet-switched communication. ATM does not use fixed bandwidths as in circuit switching techniques such as ISDN. It is implemented by using virtual circuit packet switching technology with fixed sized packets, called cells. In the virtual circuit technique all the cells in a call follow the same route, and arrive in the correct sequence.

Each cell is 53 bytes in length, 48 bytes for the information field and 5 bytes for the header. To guarantee a fast processing in the network, the ATM header has very limited function. Its main function is the identification of the virtual connection by an identifier, which is selected at call set up and guarantees a proper routing of each packet. The information field length is relatively small, reducing the network queuing delays. This in turn leads to small delays and a small delay jitter as required by real-time applications such as videoconferencing. No processing is performed on information field of the cells inside the network, further reducing delays.

ATM is connection oriented. Before information is transferred from the terminal to the network, a logical/virtual connection is set. When a user requests a connection, it provides information to the network about its requirements such as peak cell rate, acceptable cell delay variation, sustainable cell rate, and its highest expected cell burst. If the network can support this connection without jeopardizing the promised quality of service for the existing connections, the connection is accepted. On admission into the ATM network, a virtual channel is set up between the end users, through the network.

ATM provides a Quality of Service (QoS) system that is vastly more sophisticated than that found in IP networks. While IP networks attempt to provide quality with protocols such as RSVP and RTP, it is not guaranteed. A constant good quality videoconferencing application with a fixed frame rate, no audio gaps, and lip synchrony, can only be guaranteed on a network with a guaranteed QoS. Although possible on dedicated LANs or unoccupied high-speed links, in general it requires ATM or ISDN connections, with guaranteed QoS.

5.7.2 ITU-T Standard H.321

H.321 is the ITU-T's standard defining the implementation of videoconferencing over ATM. It is an enhancement of the ISDN standard H.320, and is fully compatible with existing H.320 systems. Like ISDN, H.321 is implemented at transmission rates, which are multiples of 128 kbps (128 Kbps, 384 Kbps, 768 Kbps, etc.). The standard was designed to take advantage of the inherent QoS capabilities of ATM, delivering the highest quality videoconferencing.

Implementing an H.321 ATM-based videoconferencing is less costly and less complex than ISDN, for several reasons. ATM switches are significantly cheaper than ISDN switches. Also, in an ATM implementation, single gateways give centralized access to other networks, where an ISDN implementation requires separate IMUXs for each end-point ISDN user. Not only does this extra cabling required in ISDN implementations increase costs, it also makes the installation more complex.

H.321 is claimed to deliver high-quality videoconferencing at significantly lower costs and with greater flexibility than with ISDN based standards, yet to provide significantly higher quality levels than IP-based implementations. ATM's unique suitability to the transport of video, as well as voice and data, is making it the network of choice for many videoconferencing applications.

5.8 ITU-T STANDARD T.120 FOR DATA

ITU-T Recommendation T.120 is specified as the standard for data transfer for many different videoconferencing standards. The specification defines protocols and services for multipoint data conferencing, and 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 actually a series of recommendations listed below.

T.121: Generic Application Template (GAT). Defines a template as a guide for developers in managing T.120 resources.
T.122: Defines Multipoint Communication Services (MCS) available to developers.
T.123: Defines low-level of protocol stack for audiographics & audiovisual conferencing for various types of networks (including POTS, ISDN, and LANs).
T.124: Defines Generic Conference Control (GCC), the mechanism to set up and manage conferences mandatory for 'group' conferences.
T.125: Defines MCS data transmission protocol.
T.126: Multipoint Still Image and Annotation protocol (also known as T.SI). Defines the protocol used to provide interoperability with graphics data in applications such as whiteboarding, annotated image exchange, screen sharing and remote apps control.
T.127: Multipoint Binary File Transfer protocol (also known as T.MBFT). Defines the protocol used to support binary file transfer within a conference.
T.128: Defines Multipoint Application Sharing protocol (also known as T.SHARE).
T.134: Defines Multimedia Application Text Conversation protocol (also known as T.CHAT).
T.135: Defines how to achieve secure T.120 conferencing.

6. APPLICATIONS

Once used only for high profile, executive, business oriented room-based conferences, videoconferencing is now being used or has shown great promise to be beneficial in a variety of areas of human activity.

6.1 PERSONAL CONFERENCING

In the business community, users are using videoconferencing for spontaneous video-enhanced telephone-like communications from desktop to desktop. Some businesses even have "open camera" policies in the workspace. This makes it possible to see if another worker is in his cubicle and therefore (maybe unfortunately for him if he seeks to get any work done) available.

As the Internet becomes more and more ubiquitous, videoconferencing is sure to follow the trends in Internet telephony. Although the quality of video will be low (at least for the time being) it will not be long before many on-line users will commonly place a video call to their friends and loved-ones on-line, seen in Figure 12.

Figure 12. Example of personal videoconferencing session.

6.2 BANKING AND FINANCIAL SERVICES

Video conferencing technology is becoming a large part of their overall long-term company strategies for many banks and financial service providers, who are using it in areas such as home banking, customer kiosks, automated teller machines (ATMs), video tellers, video call centers, and in conjunction with on-line Internet banking services. Through the use of face-to-face video conferencing capabilities, banks are able to handle more business while improving convenience and flexibility for their customers, without losing the personal contact which their customers value. Customers gain the ability to speak face-to-face with specialists who can assist them in opening accounts, pre-qualifying for mortgages and automobile loans, and obtaining investment advice, even from their desktop.

6.3 DISTANCE LEARNING

Videoconferencing is becoming a key element of *distance learning* programs at many schools, and allows educators to create a seamless and real-time interactive learning environment, which can bridge vast distances and include multiple sites. It is used from kindergarten to graduate school.

In public schools, "virtual field trip" programs like the "Jason Project", created by Dr. Robert Ballard (the man who found the *Titanic*), promise to bring the excitement of live scientific exploration to millions of schoolchildren who could otherwise not participate. Ballard has taken his audience to the Rainforests of South America and deep into the sea. As schools are wired for Internet, this type of program may become commonplace. An example of a virtual field trip is seen in Figure 13.

Over 30% of the 3500+ higher education institutions in the US already use videoconferencing to deliver part of their course curriculum. The goals and benefits of implementing videoconferencing in a distance learning setting include increasing school enrollment, creating new revenue streams by reaching out to non-traditional and previously inaccessible student populations, and cutting delivery costs.

For the traditional classroom, videoconferencing enables institutions to import global expertise. It may also provide an effective way for teachers, administrators and students from different institutions to collaborate by sharing resources and information.

Figure 13. A virtual field trip using videoconferencing.

6.4 DISTANCE TRAINING

Most large corporations have determined that it is in their critical business interest for their workers to have access to the educational opportunities that will help them increase their knowledge level and skill sets. They know this is necessary to compete, and accordingly are investing large sums of money in workforce training and education. Today over $50 billion is spent annually on corporate training, and over $6 billion is spent annually on the purchase and maintenance of facilities and equipment for this purpose.

Videoconferencing is quickly becoming part of many companies' overall employee training programs, because it provides a way to offer high-quality training while cutting course delivery costs. Because employees do not have to leave their offices for extended periods of time, disruptions are minimized and short-term productivity is not sacrificed.

6.5 PERSONNEL RECRUITMENT

Videoconferencing can be used as a screening tool for corporate personnel departments. It allows face-to-face interviews to occur without the expense of flying each potential candidate to the job location.

6.6 HEALTH CARE

Videoconferencing is used in numerous clinical applications. The most common application is in radiology, but it is also used for cardiology, dermatology, psychiatry, emergency medicine, home health care, pathology, and oncology.

The healthcare industry is a rapidly growing and evolving market. As healthcare corporations merge and form huge networks, successful organizations will be those that can leverage the medical expertise and resources of their member institutions across multiple sites.

The use of videoconferencing increases operational efficiency through better use of staff resources, provides rapid information transfer, achieves better physician collaboration, and facilitates medical education, as illustrated in Figure 14.

Figure 14. A remote consultation.

Figure 15. MedLink bedside terminal.

Uses include patient interviews, remote medical examinations using remote sensors, medical education via “interactive rounds”, remote consultation with a specialist, and even such mundane uses as enabling hospital administration to negotiate medical supply contracts.

Videoconferencing is used in some hospitals to facilitate the care of newborns in the neonatal intensive care unit (NICU). These babies may be hospitalized for the first several months of their lives, see Figure 16. Videoconferencing systems are installed in the NICU, as well as the homes of families with newborns in the unit, enabling family members to see and talk to their babies and review their progress with doctors, all from their homes.

Once the baby is sent home with his or her nervous parents, the hospital staff use the system to monitor the progress of the baby and provide comfort and support to the parents. By using videoconferencing, the hospital can smooth the transitions of NICU babies from the hospital to their homes, while reducing overall costs.

Figure 16. Baby in NICU.

Psychiatrists can us videoconferencing technology to monitor their patients such as those with severe mental illnesses, personality disorders, and adjustment problems. For example, a psychiatrist employed at a state hospital may have patients who live over a wide area and who are unable to travel. Even hearing- or speech-impaired patients may be reached in this way, provided that the doctor knows sign language.

One example of videoconferencing used by the medical industry is the *MedLink* Mobile Videoconferencing Unit (by PictureTel), designed to operate at the patient’s bedside and allow him to speak with a remote specialist face-to-face. This system is shown in Figure 15. The MedLink delivers separate or simultaneous medical data along with videoconferencing to any of PictureTel's other products. It is based on the Venue 2000 Model 50 platform and is network independent. The network link can range from simple ISDN through T-1/E-1 or the unit can be attached to the local area network to take advantage of existing topologies.

6.7 LAW ENFORCEMENT AND SURVEILLANCE

Videoconferencing technology is used for remote video monitoring of secured locations, as well as for replacing human operators at monitoring locations, such as at drawbridges, which must be raised and lowered based on boat and car traffic. Figure 17 illustrates TeleEye, a remote videoconferencing surveillance system.

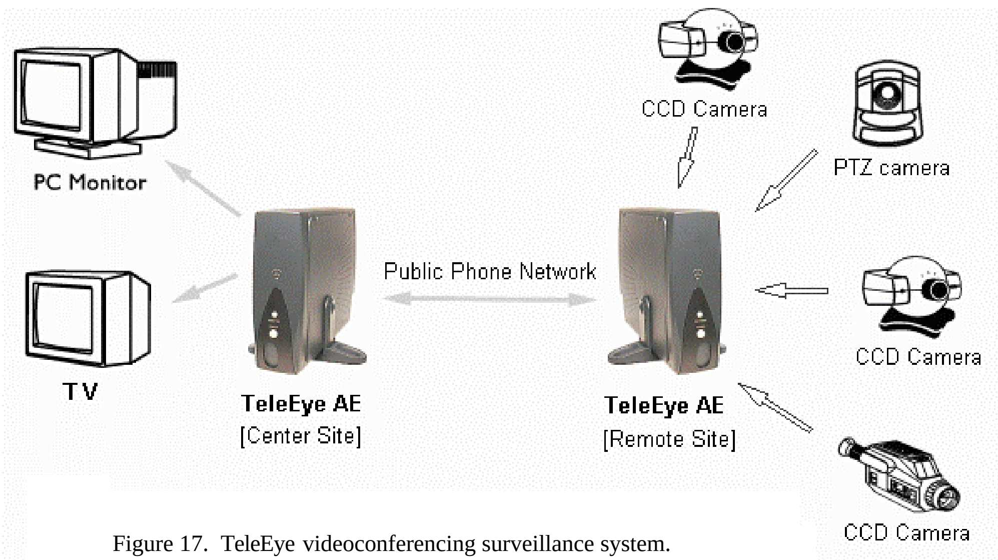


Figure 17. TeleEye videoconferencing surveillance system.

6.8 TELECOMMUTING

Some companies use videoconferencing to allow employees to work out of their homes. This is part of a trend termed telecommuting.

For example, IBM has put 95% of their US marketing and services personnel into telecommuting. They have been able to close or reducing the size of the field sales offices accordingly, and claim a 15% gain in productivity and a 40-60% savings in real estate per location. One benefit of the system is that it allows cheap customer follow-up “video visits”, which be cost-prohibitive if carried out in person.

6.9 LAW AND CRIMINAL JUSTICE

Many state and federal courts now use videoconferencing to arraign criminals, for appeals and parole hearings, and to provide telemedicine to inmates.

Lawyers use desktop system to hold meetings with their clients and to gather data from expert witnesses without incurring large expenses.

6.10 PRODUCT DEVELOPMENT AND MANUFACTURING

In today's global economy, it is not unusual for products to be developed at multiple locations. For example, a complex consumer electronics item may have the software developed overseas, the electronics developed at one location and manufactured at still another, all while the overall project was being managed at the companies home office.

In order for such a project to be completed while its market window is still open, communications and teamwork between the disparate design teams is critical.
Videoconferencing allows team members to interact as though they were collocated, sharing ideas, drawing pictures, and generally becoming comfortable with each other, while saving millions of dollars in travel expenses and lost travel time. Design and engineering teams can then conduct design reviews at a distance with full audio-visual support.

Once the product moves to manufacturing, complex mechanical problems can be communicated to remotely located experts, who can solve the problem quickly and get the production line up and running as soon as possible. This is critical, as factory downtime translates to dumping money in the trash

6.11 GOVERNMENT

State, local governments, and municipalities commonly use videoconferencing for conducting daily business such as meetings, interviews, public hearings, training, and press conferences. A cost benefit analysis conducted by the state of Utah showed it achieved a 69 percent cost savings in travel expenses for events in which videoconferencing was used. In Oregon, the state reported that it realized up to a 90 percent cost savings through the implementation of teleconferencing events versus face-to-face events. As taxpayers demand more performance for less money, videoconferencing is sure to become even more important to government function.

7. VIDEOCONFERENCING PRODUCTS

Videoconferencing is a many-tiered market. There are basically three classes of products available: group (or room) based systems, desktop videoconferencing systems, and the lower-end personal conferencing/videophone systems.

7.1 GROUP-BASED VIDEOCONFERENCING

The group-based systems, generally operating over ISDN networks, were the first to enjoy widespread market acceptance, and for years were the mainstay of the videoconferencing market. Today, however, the market for desktop and personal videoconferencing products is exploding, and overtaking the group-based market.

The group-based systems require the users to convene to a conference room equipped with videoconferencing equipment, as illustrated in Figure 18. The session may be multicast, and users at the other end may be either at group-based sites or desktop. The system can be enclosed in one cabinet with a monitor and camera on top. Alternatively, the products such as the codec, cameras, projection equipment, audio equipment, and lighting can be purchased separately to achieve the desired presentation quality. Figure 19 shows an example of group-based videoconferencing equipment.

Figure 18. Group-based Videoconferencing.

Figure 19. Group-based equipment.

Table 4 summarizes group-based videoconferencing manufacturers, products and equipment included, as well as the standards they operate under.

Table 4. Group-Based Videoconferencing Products.

Manufacturer	Product	Product Description	Camera	Mic	H.320	H.323	T.120
Intel	TeamStation System	Conference room workstation videoconferencing system, includes add-in card	P	P	P	P	P
NEC	VisualLink 384	Set-top device	P	P	P		
PictureTel	SwiftSite II	Portable, set-top group videoconferencing system	P	P	P		P
Sony	TriniCom Digital Meeting System	Parallel-port peripheral	P	P	P		P
VTEL	WG500 WorkGroup	Includes PC	P	P	P	P	P

7.2 DESKTOP VIDEOCONFERENCING

Because of the universal accessibility of the PC, desktop conferencing is appealing to a much broader customer base than the traditional group systems. Desktop systems are installed on workstations or PCs and connect over standard LANs, and operate under the H.323 standards. These systems take advantage of several features found in today's widely available multimedia PC, such as high-resolution monitors, powerful processors with MMX technology, speakers, and microphones. Videoconferencing packages are available that augment the multimedia PC with a camera and software. Figure 20 shows typical desktop videoconferencing equipment and Figure 21 shows a desktop videoconferencing session. Table 5 summarizes desktop videoconferencing manufacturers, products and equipment included, and the standards they operate under.

Figure 20. Desktop videoconferencing equipment.

Figure 21. A desktop videoconferencing session.

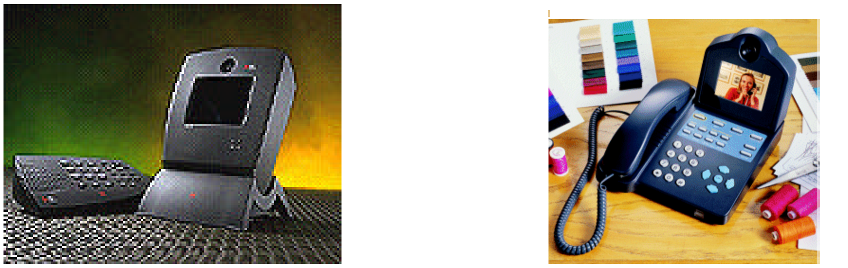
Table 5. Desktop Videoconferencing Products.

Manufacturer	Product	Product Description	Camera	Mic	H.320	H.323	T.120
Intel	ProShare Video System 500	Computer Add-in Card	P	P	P	P	P
PictureTel	Live200	Computer Add-in Card	P	P	P		P
PictureTel	LiveLAN	Computer Add-in Card	P	P		P	P
VTEL	SmartStation Desktop 384	2 Computer Add-in Cards	P	P	P	P	P
Zydacron	OnWAN350	Computer Add-in Card	P	P	P		P

7.3 PERSONAL VIDEOCONFERENCING WITH VIDEOPHONES

Videophones are also considered a personal conferencing tool, but are generally of lower quality than the desktop-to-desktop variety, and appeal to a different market segment. Videophones may be special telephones that include a small video screen, set-top boxes operating with the user's TV, but are most often implemented as software running on the PC. They communicate via POTS analog phone lines, and conform to the ITU-T's H.324 standards.

The dedicated telephone with integrated video screen, known as a desktop videophone, shown in Figure 22, is the most expensive option in the videophone arena, however installation and operation is simple. These devices simply plug into any analog telephone outlet and are ready to make a video call. Depending on the vendor, features such as adjustable picture quality, size and frame speed, electronic pan, tilt, zoom, snapshot, caller ID, and auto answer can be found. There is no software to install, or special wiring involved. These videophones are designed to work with other H.324-compatible videophones including computer-based videophones.



8X8 VC150 MM220 Videophone

Figure 22. Desktop videophones.

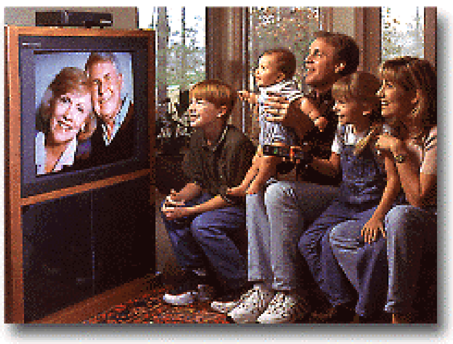


Figure 23. Comtrad C-Phone system.

Figure 24. Set-top videophone system.

The next, slightly less expensive, option is the set-top box device, shown in Figures 23 and 24. This device is about the size and shape of a cable TV converter box, and hooks up to the TV just like a VCR. It may be operated via remote control. The product includes a camera, a modem, and an H.324 industry standard videoconferencing codec. The system uses a television set to present the audio and video of the person being called, thereby making this solution more economical.

Figure 25. Videophone kit.

Figure 26. Intel videophone.

A third option available for videophone conferencing is available in the form of software that is installed on the user's PC. These products are similar to the desktop videoconferencing systems, except that these systems are operating over POTS. Table 6 shows a list of videophone manufacturers and products. This solution takes advantage of the existing capabilities of today's multimedia PC, making this it by far the most economical solution available. In fact, new PCs, with an Intel® Pentium® III processors, come preinstalled with the latest Intel® Video Phone software, shown in Figure 26. All that is required is a camera.

Most videoconferencing vendors provide the products as a kit, which includes the software, video capture board (optional), and a camera, shown in Figure 25.

The user can choose to make video phone calls over regular telephone lines or through the Internet. When used over regular telephone lines, the audio and video quality is much better than that of video phone calls made through the Internet. The quality of an Internet videophone call will vary depending on Internet traffic at the time of the call.

Many of these products support both H.323 and H.324 standards, and support a variety of broadband Internet connections, including cable modems, DSL, ADSL and LAN.

Table 6. H.324 Videophone Products.

Manufacturer	Product Number	Product Description
8x8	VC150	ViaTV Desktop Videophone
8x8	VC105	ViaTV Set-Top Videophone W/Camera
Winnov	Videum Conference Pro	Videoconferencing Kit
Winnov	VideumCam	Videoconferencing Kit
Panasonic	Eggcam	Desktop Video Camera w/Cu-Seeme
3Com	Bigpicture VideoPhone	PC Videoconferencing Kit, includes video capture card
Intel	Create and Share	PC Videoconferencing Kit

References

1. E. Brown, "Videoconferencing Systems," *New Media*, December 1998.
2. K. Nisenson, "Tune in to IP Videoconferencing," *Network World*, September 1998.
3. W. Wong, "Video Conferencing for the Enterprise," *Network Magazine*, April 1998.
4. T. Trowt-Bayard and Jim R. Wilcox, "Video Conferencing and Interactive Multimedia: The Whole Picture", Flatiron Publishing, March 1997.
5. M. Herman, "The Fundamentals of H.324 Desktop Videoconferencing," *Electronic Design,* October 1996.
6. C. Tristram, "Video Conferencing: A Work in Progress," *Network Magazine,* April 1998.
7. S. J. Bigelow, "The Many Faces of Videoconferencing," *Bay Area Computer Currents,* October 1998.
8. K. Cholewka, "IP VideoConferencing: Beat the Clock," July 1997, http://www.data.com/tutorials/video.html.

9. DataBeam Corp, “A Primer on the H.323 Series Standard, Version 2.0,” http://www.databeam.com/h323/h323primer.html.
10. CMU Artificial Intelligence Repository, “CELP: FS-1016 Code Excited Linear Prediction Coder,” http://www.cs.cmu.edu/afs/cs/project/airepository/ai/areas/speech/systems/celp/0.html.
11. Coding and Compression, http://www.cs.ucl.ac.uk/staff/J.Crowcroft/mmbook/book/node91.html.
12. FVC.COM, “Videoconferencing Standards,” December 1997, http://www.fvc.com/whitepapers/standards.html.
13. Cisco Systems, “Packet Voice Primer,” http://www.cisco.com/warp/public/cc/sol/mkt/ent/gen/packv_in.htm.
14. T. Needleman, “A Videoconference on Every POTS,” http://www.zdnet.com/pcmag/features/vidconf/pots.htm.
15. TeamSolutions (UK) Limited, “Video Standards and Terminology,” January 29, 1999, http://www.teamsolutions.co.uk/tsstds.html.