Screen-Based Multimedia Telephony

Paul E. Crouch
Jorge S. Rodriguez
William C. Schwartz

Today, telephone users are already realizing the multimedia dream as the AT&T VideoPhone 2500, Picasso,® and Vistium™ products are introduced into consumer's homes and onto office desktops. A major, common element of these products is a display screen that provides a visual perspective. In addition, end-users now have a choice about what they can view at any given instant. The quality of this choice depends on the product itself, the feature set being used, and the bit rate of the telephone transmission medium. This bit rate varies from the very low speeds of the public switched telephone network (PSTN) to the higher speeds of basic-rate ISDN and ultimately to the very high speeds of broadband and asynchronous transfer mode technologies. This paper discusses these factors, the overall architecture of the three AT&T products, as well as their rich sets of features. In addition, key markets and applications are highlighted.

Introduction

This paper discusses the dream of multimedia and how it is being realized today for telephone users as they employ the AT&T VideoPhone 2500, Picasso,® and Vistium™ products in their homes and offices (Figure 1).

The major, common element of these products is the use of a display screen that provides end-users with a visual perspective. A video camera, which is important to each of these products, enables end-users to maintain complete two-way visible linking. Only by employing a camera and display screen can the sense of sight (and to some degree, the sense of touch) be employed to provide full visual presence.

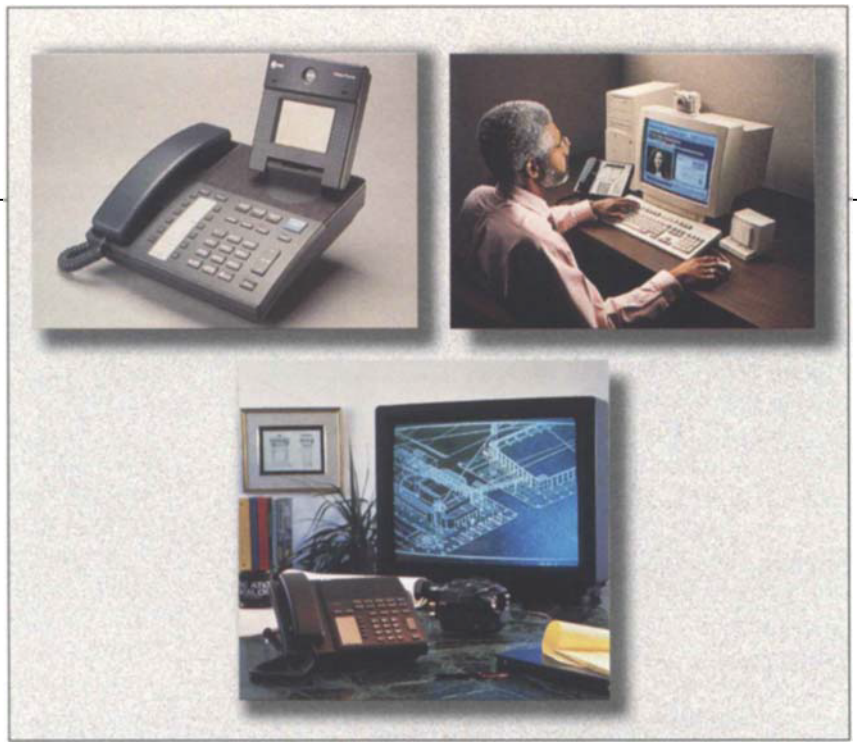
The general opinion is that the multimedia vision is realized only when visual connectivity, as well as data and application sharing, are combined with voice communication to create a sense of presence and togetherness, even though users are connected solely by a communications link. The window that the screen and camera provide create the sense of presence.

An added advantage of screen-based multimedia telephony is not just seeing, but also providing end-users a choice of what they can view at any given instant. The quality of this choice depends on the following three considerations:

- The selected product,
- The application of the selected product's feature set, and
- The bit rate of the telephone transmission medium.

Today's most commonly used transmission media range from the public switched telephone network (PSTN), which has very low bit rates, to basic-rate ISDN featuring considerably higher speeds. Ultimately, multimedia products will incorporate the very high bit rates of broadband ISDN and asynchronous transfer mode technologies, which can be carried over traditional copper cable, as well as fiber-optic cable and direct satellite links.

In addition to a discussion of the foregoing aspects of screen-based multimedia telephony, the paper also provides:



- A definition of screen-based multimedia,
- A technical description of a general multimedia platform,
- An overview of the common portion of the VideoPhone 2500 and Picasso system architectures,
- A report on the unique aspects of videophones,
- Details of the Picasso and Vistium product offerings,
- The standards background for multimedia,
- Quality factors for screen-based multimedia products, and
- A conclusion and perspective on the future.

Figure 1. Telephone users are employing the AT&T VideoPhone 2500 (left), Picasso® (center), and Vistium™ (right) systems in their homes and offices as the dream of multimedia becomes reality. A common feature of each system is the use of a display screen that provides end-users with a visual perspective. A video camera, furnished as standard equipment, enables end-users to maintain complete two-way visual linking. Only by employing a camera and display screen can the sense of sight (and to some degree, the sense of touch) be employed to provide a full visual presence.

Screen-Based Multimedia Communications

A significant opportunity and challenge for would-be product providers is defining multimedia functionality and furnishing it within the overall communications marketplace. At its root, the definition of screen-based multimedia telephony is *the performance of simultaneous tasks over communications links.*

Perhaps of greatest importance is simply sharing remotely, but this sharing should involve the total use of human senses as much as possible. It should include hearing, sight, and touch. The definition must also include information exchange—both verbal and visual—and embody true human feelings. Collaboration using audio, visual, and data media is also a significant consideration. The goal is to create an environment using a communications link in which each user or participant feels in the physical presence of all other call participants.

Multimedia communications requires specific product features to meet end-user needs. These features usually include voice, motion video, still image, general use data, and graphics provisioning. To provide users with necessary directions, instructions, and choices, text overlay of all visual media types is also required. To provide these elements over affordable and available transmission media requires signal compression without compromising quality.

Multimedia is global in scope and universal in its applications, some of which include person-to-person, person-to-group, group-to-group, and person-to-machine communications (for example, electronic linking facilitated by client servers, answering machines, and messaging services).

Customer groups include consumers at home; employees working at home; people working in business, commerce, or at their avocation; and individuals traveling for business, pleasure, or any other reasons.

The key to unlocking the multimedia communications marketplace is providing user applications that meet the vast and diverse needs of all customers at their

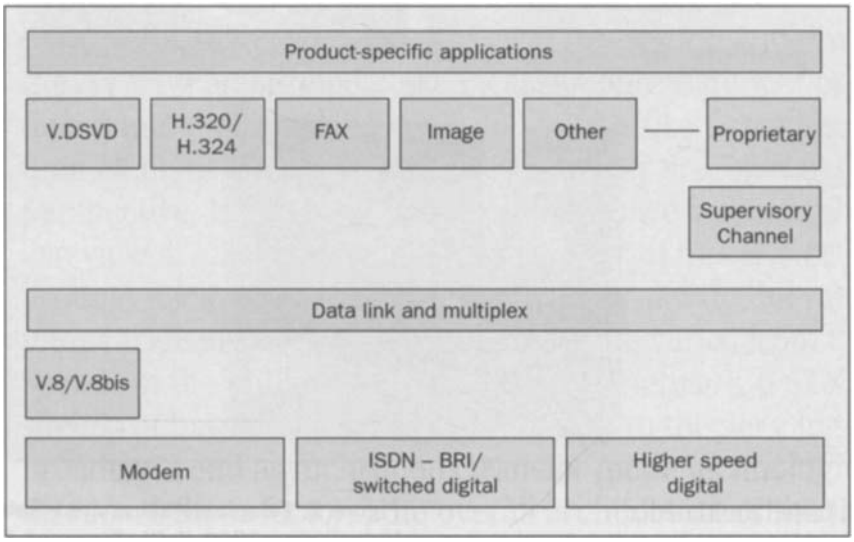


Figure 2. This hierarchical block diagram depicts the technical elements of the screen-based multimedia platform and provides a high-level view of its common elements. The choice of transmission media is shown at the lowest level. Data-linking and multiplexing functions are shown next. The two highest levels show some of the functions customers might expect. The topmost level suggests that many different products will be created in the future—those having specific features that match each market need and individual design intent.

point of need. Such applications and needs are endless and already include:

- Project planning;
- Bid preparation;
- Telecommuting;
- Provision of remote expertise for banking, customer service, maintenance and repair, and sales;
- Information banks and kiosks;
- Distance learning and training;
- Legal;
- Medical;
- Criminal justice; and
- Entertainment.

Actually, however, in spite of all of the applications that are already known, many more remain to be identified. All can benefit from further development.

Technical Elements of the Multimedia Platform

Wherever possible, AT&T strives to ensure that its communication products conform with international telephony standards, which are discussed later in this paper. Such standards are formally established by the International Telecommunications Union – Telecommunications Standardization Sector (ITU-T), the International Organization for Standardization (ISO), or the appropriate regional or national standards bodies.

Figure 2 shows the technical elements of a screen-based multimedia platform and also provides a high-level view of its common elements. The choices of transmission media are shown at the lowest level. Selections include the low-speed analog media of PSTN connections to the higher speeds of basic-rate ISDN and Switched 56 digital connections to the even higher speeds of T1 or primary-rate connections and beyond.

When a V.34 modem is used on an analog telephone line, the bit rate for the connection can range from 28.8 kb/s to as low as around 9.6 kb/s. The V.8 protocol (or V.8bis, when available) will be used to facilitate the

Panel 1. Abbreviations, Acronyms, and Terms

A/D—analog-to-digital
BRI—basic rate interface
BRICM—basic rate interface communications module
CELP—code-excited linear prediction
CELP+ —enhanced CELP
CIF—common intermediate format
D/A—digital to analog
DSP—digital signal processor
DSVD—digital simultaneous voice and data
HDLC—high-level data link control
ISA—industry standard architecture
ISDN—integrated services digital network
ISO—International Organization for Standardization
ITU-T—International Telecommunications Union –Telecommunications Standardization Sector
JPEG—Joint Photographic Experts Group
MVIP—multivendor integration protocol
NTSC—National Television Standards Committee
PAL—Paradox Application Language
PCM—pulse code modulation, a technique for digitizing speech by sampling the sound waves and converting each sample into a binary number
POTS—"plain old telephone service"
PSTN—public switched telephone network
Q factor—a parameter used to process an image during encoding
QCIF—quarter common intermediate format
RAM—random access memory
RGB—red/green/blue
S-VHS—Super-VHS, a video recording and playback system that uses a higher-quality VHS cassette and S-video technology
SRAM—static random access memory
Switched 56—a digital service at 56 kb/s provided by the local telephone companies and long distance carriers that works like a dial-up telephone system, only for digital data
TMI—telephone multimedia interface
VRAM—video random access memory
YUV—color encoding scheme for separating luminance and chrominance

crucial start-up periods for modem connections. Basic-rate ISDN and switched digital connections ordinarily provide bit rates of 64 and 56 kb/s, as well as 128 and 112 kb/s. Today, however, the higher speeds provided by ISDN primary-rate, T1, and sub-T1 connections are not commonly used, although the 384 kb/s rate has become common for high-quality video communication in room conferencing situations.

Data-linking and multiplexing functions are shown in Figure 2 just above the physical layer. The data-linking capability is used to ensure that the multimedia data are communicated error free over the connection. Several common data-linking choices are available, ranging from the packet-based link access procedure used in PSTN-based products and LANs to the H.221 procedure used for ISDN-based products. The multiplexing capability permits the various types of data (speech, video, end-to-end control, still image, file transfer, and so on) to be combined efficiently into a single data stream for conveyance over the transmission media.

The two highest levels of Figure 2 show the various functions customers expect in a multimedia product. The breadth of functionality suggested includes digital simultaneous voice and data (DSVD) applications; videophone and videoconferencing applications using H.320 and H.324 protocols; facsimile applications; still-image transfer applications; and other standards-based applications extending to proprietary uses and protocols.

The topmost level of Figure 2 suggests that many different products will be created in the future. Such products will have various, specific features that match each market need and individual design intent.

Some elements not shown in Figure 2 must be mentioned because of their relative importance. These elements include the user interface, display screen, and camera. Additional essential components include audio and speech interfaces, data-transmission interfaces, and end-to-end communication and control capabilities.

User interfaces differ in complexity from very simple telephone interfaces to the highly complex types mandated by the PC's vast functionality or by remote or wireless input. The variety of screens includes the very common CRTs, which are used as desktop monitors and television sets; and the flat screens used in videophones, laptop computers, and new products now being developed. Camera and video input sources include the inte-

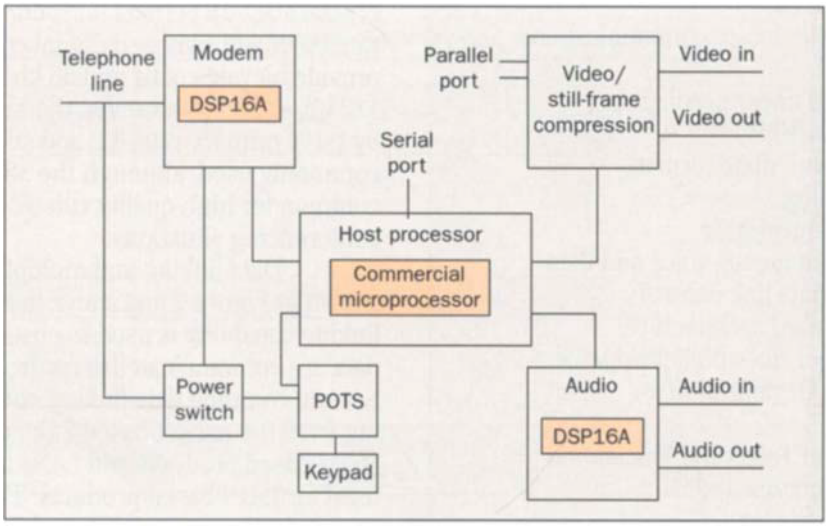


Figure 3. This architectural flow diagram illustrates the major subsystems that implement the AT&T VideoPhone 2500 and Picasso system functions. The Picasso system is based on the VideoPhone 2500 architecture.

grated cameras commonly found in videophones; camera modules used with PCs, camcorders, still-frame platters, and video disk players; and electronic input from computers.

Quite likely, natural-sounding speech will always be the most important attribute of any multimedia call involving voice transmission. With today's transmission media, unrelenting pressure is placed on speech and audio subsystem designers for greater compression and bandwidth, higher quality, and reduced transmission delays. Data interfaces, data-transport uses and mechanisms, and end-to-end applications are growing and diverging so rapidly that it is currently impossible to predict future multimedia usage.

VideoPhone 2500 and Picasso System Architecture

The Picasso system is based on the VideoPhone 2500 architecture.[1] Figure 3 shows a simple block diagram of the major subsystems that implement the VideoPhone 2500 and Picasso system functions. The rest of this section provides brief descriptions of each subsystem.

Host Processor. Overall control of each product is provided by the host processor, which also interfaces with all other subsystems. The host processor's main function is to gather compressed video, image, and audio data from the respective processors and reliably transmit these data and other system information through a modem.

In the *receive* direction, the host processor takes data from the far end via the modem, checks its integrity, then unbundles it into video, image, audio, and other elements. The host processor then forwards the data to the video, audio, and data processors, respectively. In addition, the host processor also handles the various user's requests invoked through key presses captured by the POTS processor. These requests are forwarded to the host for further processing.

The host processor is based on a Motorola MC68302* microcontroller. This device uses a Motorola 68000* chip as its central processing unit (CPU). The MC68302 also has several peripheral processing functions that handle both input and output, thus offloading these operations from the CPU.

Modem. The VideoPhone 2500 uses a proprietary modem[1] that provides bit rates of 16.8 and 19.2 kb/s. Future videophone products are expected to use the V.34 modem, which provides bit rates of up to 28.8 kb/s.

The Picasso system uses an industry standard V.32bis modem. V.32bis is a 2,400-baud, echo-canceling modem having a fixed carrier frequency of 1,800 Hz. This modem is capable of transmitting data at bit rates of up to 14.4 kb/s. Lower data rates (12, 9.6, 7.2, and 4.8 kb/s) are also supported if the start-up rate of 14.4 kb/s is not successful.

Both the VideoPhone 2500 and Picasso system modems are based on the AT&T Microelectronics high-speed modem chip set. This chip set includes a ROM-coded AT&T Microelectronics DSP16A digital signal

processor, the CAMIL2 interface logic chip, and the AT&T T7525 linear codec for analog-to-digital (A/D) and digital-to-analog (D/A) conversion, which facilitates interfacing to a telephone line.

Audio Processor. Compression and decompression of audio data is performed by the audio processor using a code-excited linear prediction (CELP) algorithm[2]. The coder operates at 6.8 kb/s and incorporates fractional-pitch prediction[3] and constrained stochastic excitation[4]. Due to these enhancements, the algorithm has come to be known as CELP+. The audio processor, like the modem chip set, is based on a ROM-coded DSP16A digital signal processor.

VideoPhone 2500 Video Subsystem. At a high level, the video subsystem provides the following basic functions: video preprocessing; "lossy" video compression, which uses discrete cosine transform processing and motion compensation; and post processing.[1]

Picasso Video Subsystem. All video-related functions are performed by the Picasso video subsystem. These functions include:

- Video I/O,
- A/D and D/A conversion,
- Frame capture,
- Image compression and decompression,
- Image storage and retrieval,
- Text overlay,
- PC interface, and
- Annotation.

The Picasso system supports the following two video I/O signal formats: National Television Standards Committee (NTSC) and Super-VHS (S-VHS). The composite input signal is color separated into luminance (Y), red chrominance (Cr), and blue chrominance (Cb). Each of these components, in turn, is digitized and stored in video random access memory (VRAM) for each video frame captured. The maximum spatial resolution is 512 x 480 pixels with 4:2:2 sampling, and eight bits per pixel are used. For the output signal, the reverse is performed. For image compression and decompression, the algorithm used is based on the Joint Photographic Experts Group (JPEG) ISO standard.[5]

The video subsystem is based on the AT&T Microelectronics DSP3210 processor. This chip fully supports 32-bit-wide integer and floating-point computations. It has access to 2 megabytes of memory for image compression, image storage, and image processing software. In addition, this subsystem contains 512 kilobytes (KB) of VRAM for image capture, and 128 KB of VRAM for overlay text and annotation data.

POTS/Keypad. The POTS and associated speakerphone circuitry provide the analog audio interface and power-fail operation. These circuits are based on the VideoPhone 2500. Similarly, the Picasso system's keypad is a modified version of the VideoPhone 2500 keypad. It supports additional keys for image-related functionality, and it also contains circuitry for infrared remote control.

Communication Protocol. Both the VideoPhone 2500 and Picasso system multiplex four different data types over the physical line. The data types include video/image, annotation (Picasso system only), audio, and control. The X.25 protocol is used to accomplish this task. X.25 provides for the reliable transmission of the data stream by performing error detection and retransmission. Audio frames are not retransmitted because of real-time constraints. Each data type has an X.25 virtual channel associated with it. The control channel is used for end-to-end control, status information, set identification, and capabilities exchange.

Videophones

The VideoPhone 2500, which was introduced in 1992, was the first product of its kind to operate over ordinary analog (PSTN) telephone lines. It is truly an international product, having been certified for operation in 38 countries. It has a built-in camera and a 3.3" liquid crystal display designed into a turret, permitting tilting and swiveling of both the camera and display.

Typical of high-end consumer telephone product, its feature set includes a 12-number memory, redial, flash, hold, microphone mute, push-button touch-tone/dial-pulse operation, and internal speakerphone. Its videophone features include motion video transmission, self view in both full-screen and picture-in-picture modes, display brightness control, a focus button that allows adjustment of the far-end videophone's operation to satisfy a user's preference, and a physical privacy shield for the camera.

The technology used in a PSTN videophone includes the following features:

- Video and speech compression;
- Color video camera system;

- Color display;
- Host processor functions;
- Telephone/speakerphone electronics; and
- User interface technology, which includes an attractive housing and handset design, keypad, and feature button technology.

The VideoPhone 2500 was originally designed for person-to-person communication in the consumer marketplace. The paradigm is one of "talking heads"—for example, person to person, relative to relative, and so on. Actual use has extended far beyond the consumer market to include applications in public telephone booths, businesses, physicians' offices, and commercial enterprises. Interest in the VideoPhone 2500 is also increasing for person-to-machine applications and interfaces including video and application servers, video answering machines, and video messaging.

New videophone configurations are also in demand. A follow-on product to the VideoPhone 2500 is being jointly developed by an AT&T Global VideoPhone Service licensee and AT&T. This product contains neither an integrated camera nor display. Rather, it features sets of I/O video and audio jacks for connection to external off-the-shelf cameras, displays, and video recorder equipment. The product is ideal for such applications as small group conferencing and surveillance, monitoring and security, as well as typical person-to-person use. It has both motion and still-frame video capabilities (compatible with the Picasso system), and it employs a V.34 modem to provide noticeably improved motion video quality.

Operation of an AT&T videophone is simple. To place a call, a user first lifts the handset or presses the *speaker* button to acquire dial tone. Then the user dials the desired number. For added convenience, both redial and memory dialing features are available. Once the called party answers and both parties agree to a video call, one user simply presses the *video* button and the digital videophone link is established through the modem. Either user can terminate the call by replacing the handset or by pressing the *speaker* button. To end the video link without terminating the audio connection, either party can press the *flash* button.

The Picasso System

Introduced in May 1993, the AT&T Picasso still-image phone allows the simultaneous transmission of voice and high-resolution color images over regular analog telephone lines connected to the PSTN. Users at both ends of a call can collaborate and discuss the images while using a mouse for pointing and annotating. Images can also be saved and retrieved from a PC by using the PC interface.

Basic Operation. From a user's perspective, the Picasso system was designed to have the look and feel of a regular telephone (see Figure 1). To complete a call, a user first lifts the handset or presses the *speakerphone* button and then dials the number. Once the called party answers, an image can be captured by pressing the *preview/capture* button. The image is transmitted by pressing the *send* button. To store a captured or received image, a user simply presses the *store/select* button. Up to 32 images can be stored internally. These stored images can be displayed and manipulated by using the Picasso system's extensive set of image processing features.

Features. The Picasso system has a rich set of features, providing users with the abilities to transmit, collaborate, and manipulate objects represented as captured images.

Annotation. The feature known as *annotation* allows users at both ends of a call to point and draw over the image being displayed on the screen, thus providing the ability to discuss details and highlight areas of particular interest. A user annotates the screen by using a mouse connected to one of the phone's RS-232 ports. Annotations can either be stored as part of the image or erased by simply double clicking a mouse button.

Image enhancement. A powerful set of utilities is provided to users of the Picasso system. This utility set improves an image's visual appearance and transmission speed. It includes:

- *Photographic enhancement.* Dark images can be brightened significantly by applying a technique known as histogram equalization with which the gray-scale values of an image are spread out while keeping their sum constant.
- *Text enhancement.* Images combined with text can be enhanced by sharpening the text using a technique known as edge detection.
- *Jitter-free operation.* Single-frame capture of moving objects may generate a flickering effect due to the interlaced mode of operation of most standard video

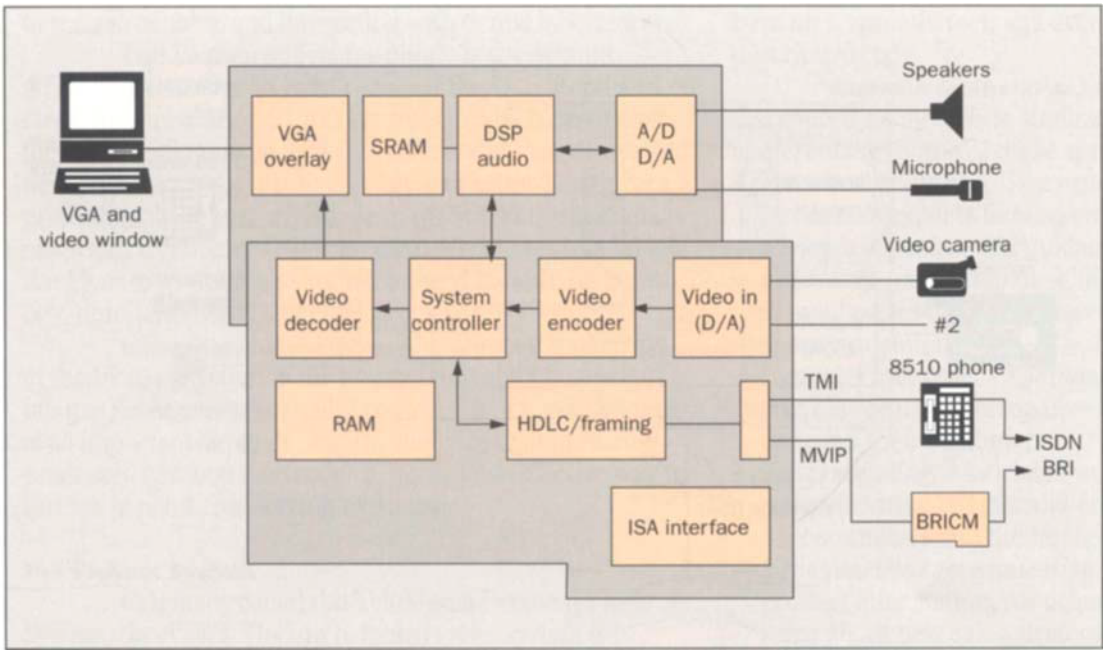


Figure 4. This drawing shows the architecture of the AT&T Vistium 1300 system, which is a two-board addition to a PC. The system provides video and audio signal compression and decompression. Speeds of up to 15 frames/s for common intermediate format (CIF) resolution images and up to 30 frames/s for quarter common intermediate format (QCIF) images are achieved. This system multiplexes the video and audio data streams together with an optional user data stream. It also supports the AT&T 8510 ISDN telephone, an ISDN adapter, or a Switched 56 adapter.

cameras. If an object is moving during capture, each half of the captured image may be slightly different due to the delay between the video fields. When the image is displayed, flicker results. This effect can be eliminated either by displaying only one field or by constructing a new frame through interpolation of the fields. Both methods affect the vertical resolution of the image. The interpolation technique, however, usually produces better image quality.

Image quality versus transmission time. The JPEG compression algorithm used in the Picasso system is a lossy process—that is, some information is irretrievably lost when the image is compressed. High compression results in more information loss, and the greater the loss, the more picture quality is affected. Hence, the amount of compression must be application dependent due to its effect on quality.

The Picasso system allows a user to alter picture quality in two ways. First, the compression ratio can be increased or decreased by choosing a *Q factor*, which is used to process the image during encoding. Second, a user can reduce the spatial resolution of the image being compressed, thus reducing the amount of data to be processed. The tradeoffs are reduced image quality and longer transmission time.

The Picasso system supports three image transmission speeds (and quality levels): fast (good), moderate (better), and slow (best). The slow speed, which achieves the highest quality level, uses the multiple Q factors method described in the preceding paragraph. Decimation to a lower resolution is performed, however, when using the moderate and fast speeds, which have lower quality levels.

PC interface. With this feature, Picasso system users have the ability to display, store, and print images received from the system on a Windows*-based PC. Similarly, a PC can be used as a source of images to be downloaded to the Picasso system for transmission. Connection to and from a PC is provided by means of a parallel port. Several common file formats are supported

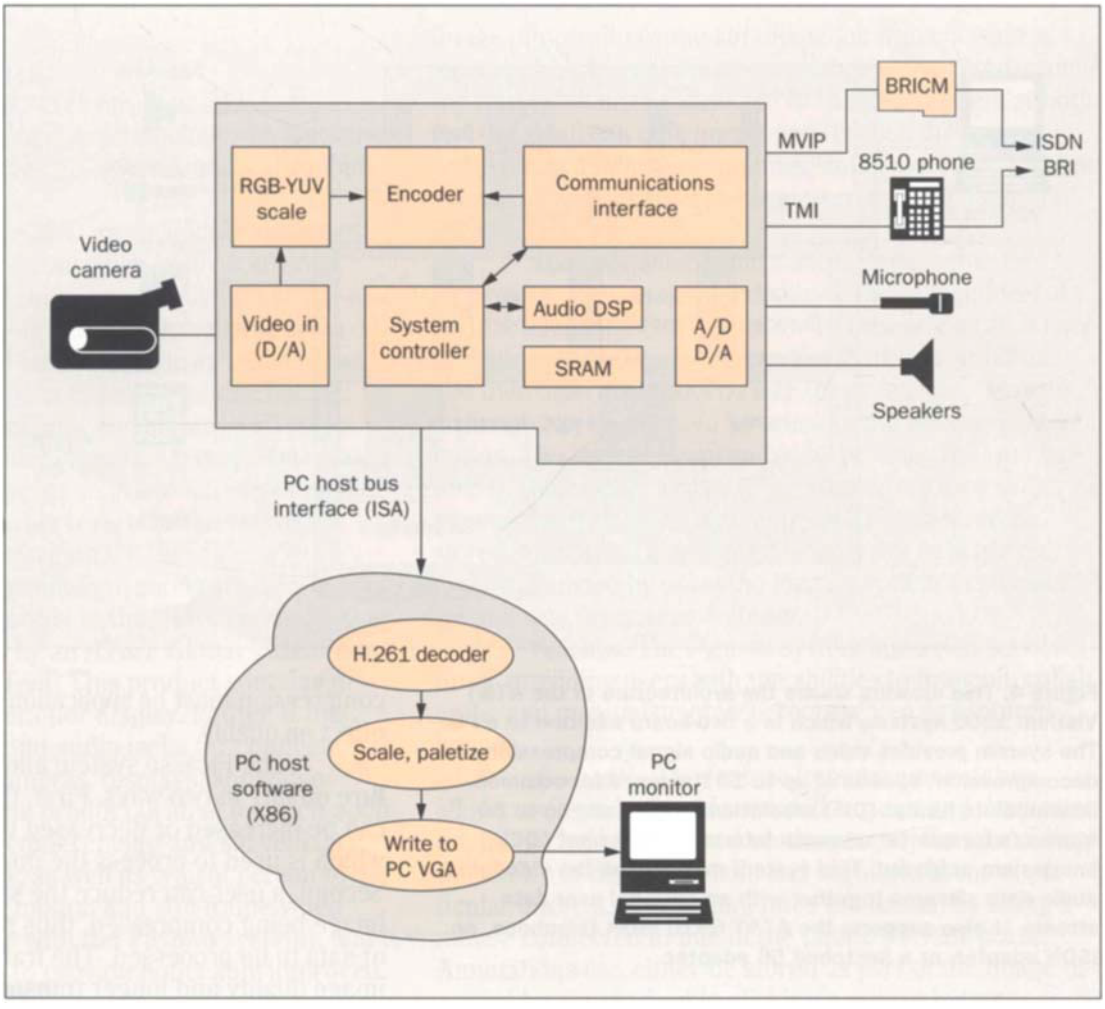


Figure 5. The architecture of the AT&T Vistium 1200 system, a single-card videoconferencing solution, is shown. Much of the system's structure and many of its components are similar to those of the earlier Vistium 1300 system. The Vistium 1200 system, however, differs in that it uses a PC software algorithm to decompress the video stream and a PC graphics subsystem to overlay the monitor's video image. Thus, the Vistium 1200 system requires less hardware and features a substantial price advantage over the Vistium 1300 system.

by this feature, including .BMP, .JPG, .PCX, .TIF, and .TGA.

Remote control. The Picasso system comes equipped with an infrared remote control that enhances ease of use during stand-up presentations.

Secure modem. Secure transmission of voice and images is also possible with the Picasso system. This feature allows a user to connect to the AT&T STU-III secure modem (model 1900 or model 1910) and exchange such encrypted signals.

Peripherals. The Picasso system is compatible with any device meeting the NTSC interface for composite or S-VHS video. Examples include video cameras, camcorders, still-image cameras, document cameras, VCRs, TV monitors, photo CDs, and video printers.

Markets and Applications. Since its market introduction in 1993, the Picasso system has been used in a wide variety of applications and in many different market segments around the world (it is now deployed in 23 countries). These market segments include creative services, advertising, retailing, health care, and engineering/manufacturing. The Picasso system is particularly well suited for applications in the health care and engineering/manufacturing markets, which are discussed next.

Health care. The medical community first became interested in the Picasso system for the purposes of remote information exchange and consultation. Applications have now been expanded to include remote training, emergency data transmitting from ambulances

to trauma centers, and interacting with dental laboratories.

The Picasso still-image phone is the second AT&T product ever to have received the U. S. Food and Drug Administration 510(k) clearance, which certifies it for use as a diagnostic tool. Therefore, the Picasso system now enables physicians to consult with other health care professionals almost anywhere in the world to aid in diagnosis and treatment. Other medical disciplines into which the Picasso system is being introduced include dermatology, ophthalmology, and emergency care.

Engineering/manufacturing. The primary application of the Picasso system in the engineering and manufacturing market segment is quality control. The Picasso system is an important aid to the improvement of manufacturing processes because it provides a highly cost-effective way to furnish in-plant engineering expertise.

The Vistium System

Originally called the TeleMedia Personal Video System, the AT&T Vistium personal video system was introduced in April 1992. The Vistium system product line has been redesigned twice with corresponding price reductions. Currently, the Vistium Solutions offering is certified in 28 countries. It features a set of versatile PC video conference cards, collaboration software, assorted communication interfaces, peripheral cameras, and audio devices.

The primary purpose of the Vistium system is to provide desktop conferencing. Many corporations are using this technique to maximize the effectiveness of their most valuable resource—people. This often means creating project teams or even virtual business units through a communications link. By meeting and conferring electronically regardless of geographical location, associates can successfully complete projects in their own offices without traveling to remote locations.

For collaboration, the Vistium system provides a shared electronic whiteboard over a communications link. When sharing applications, both parties can view and control a single Windows application on a PC and also send still images or files back and forth. A collaborative session begins with a simple phone call. Once connected, either party of the call opens a Windows application (that the other party might or might not have) and the choices begin. The parties can work on tasks together in real-time, manipulate graphics, add a banner heading, use the annotation tool to circle or highlight numbers on a spreadsheet, and even reword a graphic's descriptive copy.

Customer Applications. Vistium system users have discovered many unique applications for desktop videoconferencing. Some of these applications include:

- *Remote consultation.* Users can collaborate with subject matter experts located virtually anywhere for advice, assistance, and guidance.
- *Video help lines.* Help desk files, photographs, and manual pages can be accessed to solve problems or answer questions.
- *Customer and vendor relationships.* Video collaboration provides regular, personalized contact between vendors and their customers.
- *Telecommuting.* Associates working remotely can maintain visual contact and collaboration with other telecommuters and the home office.
- *Headquarters communication.* Employees can access product information, participate in training classes, or learn about new sales strategies.
- *Project team organization.* Project teams can be established and members can collaborate regardless of geographic locations.

Architecture. The Vistium system's architecture is based on established industry standards in videoconferencing, personal computing, and telephony. For videoconferencing, the Vistium system complies with the ITU H.320 videoconferencing standard. For personal computing, the system extends existing PCs with the addition of Industry Standard Architecture (ISA) bus cards and application software. For telephony, the system accommodates both ISDN and proprietary switched digital connections.

The Vistium product line has two components: the Vistium 1300 and 1200 systems, which are discussed in the following two subsections.

Vistium 1300 System. Figure 4 shows the architecture of the Vistium 1300 system, which is a two-board addition to a PC. As an H.320 videoconferencing product, the Vistium 1300 system provides video and audio signal compression and decompression. Speeds of up to 15 frames/s for common intermediate format (CIF) resolution images and up to 30 frames/s for quarter common intermediate format (QCIF) images are achieved.

The Vistium 1300 system multiplexes the video and audio data streams together with an optional user data stream. It also supports a direct connection to the AT&T

Table I. ITU-T standards for screen-based multimedia products

Multimedia Element	Transmission Medium		Notes
	PSTN	ISDN/SW 56	
Overall system	H.324		For an overall PSTN-based system
		H.320	For an overall ISDN-based system
Motion video	H.263		For video compression
	H.261		For interoperability with ISDN
		H.261	For video compression
Speech/audio	G.723		Dual-rate coder (63/53 kb/s)
		G.711	64-kb/s mu-law or A-law PCM
		G.722	7-kHz audio at 64, 56, or 48 kb/s
		G.728	16-kb/s low-delay speech coder
Physical interface	V.34		28.8-kb/s modem
	V.8 (V.8bis)		To facilitate modem start-up
		I.430	ISDN S/T line-interface specification
Data link and mux	H.223		Data link, multiplex, and error/flow control
		H.221	Data-transmission framing structure
System control	H.245		Capabilities exchange and system start-up
		H.242	Capabilities exchange and system start-up
Still-frame pictures	T.81		Uses the ISO JPEG standard
		H.261 annex	Uses the procedure defined in the H.261 annex
Additional data uses	T.120 series		Protocol family for transmission of other data types
		T.120 series	Protocol family for transmission of other data types

8510 ISDN telephone, an ISDN adapter, or a Switched 56 adapter. The system works in conjunction with an audio speakerphone, which provides sampling and mixing of voice inputs, as well as echo cancellation for hands-free operation. G.711 and G.728 audio are provided.

The Vistium 1300 system supports two separate video inputs and one video output in either NTSC or Paradox Application Language (PAL) format. The video input is converted in color format and fed to the encoder chip for compression. The compressed video received from the remote end is decompressed and converted into the NTSC/PAL format provided on the video output connector. The auxiliary video output is useful for large screen displays and video recorders.

The Vistium system was designed for digital networks, either ISDN or non-ISDN. For those areas not having ISDN service, the Dual V.35 interface allows connection to a non-ISDN digital switch. As mentioned earlier, the Vistium 1300 system provides a direct connection to the 8510 ISDN telephone for customers served by an AT&T 5ESS® central office switch or AT&T DEFINITY® PBX.

The Vistium 1300 system makes extensive use of current AT&T technology. For example, the video decoder, encoder, and system controller are elements of the AT&T Microelectronics AVP4000 chip set. The system also integrates the DSP3210 digital signal processor

into its design. The video and audio algorithms and application software were all developed jointly by AT&T Microelectronics and AT&T Global Information Solutions.

Vistium 1200 System. The most recent addition to the Vistium family is the Vistium 1200 system, a single-card videoconferencing solution. Figure 5 illustrates the system's architecture. Much of its structure and many of its components are similar to those of the Vistium 1300 system. The Vistium 1200 system, however, differs in two principal areas:

- A PC software algorithm decompresses the video stream, and
- A PC graphics subsystem overlays the video image on the monitor.

As a result, the Vistium 1200 system requires significantly less hardware and features a substantial price advantage when compared with the older Vistium 1300 offering.

The video performance of the Vistium 1200 system depends on the performance of the underlying CPU in the PC. On 486*-based PCs, the system can support seven to ten CIF frames/s; on Pentium*-based systems, it can support up to 15 frames/s.

Peripherals. Although the Vistium system operates with any NTSC or PAL camera, AT&T offers two camera options. Either unit is designed to be placed on top of a PC monitor. The first option is a standard, low-cost color camera having a manual focus adjustment. The second option is a pan-tilt-zoom camera having an autofocus feature. This camera comes with an infrared remote control for convenient adjustment, making it particularly useful for small group conferences.

The Vistium system also offers two telephony options. The 8510 ISDN telephone provides full business-set functionality in DEFINITY or 5ESS environments. The Vistium Audio Unit provides hands-free voice communication. Furthermore, when an optional second speaker is added, the Audio Unit serves as a multimedia PC speaker system as well.

Application Software. The Vistium system supports a rich environment of collaboration and phone-management capabilities, which include:

- A point-and-click directory for voice or video calls, complete with per-entry picture IDs;
- An on-screen telephone window that mirrors the 8510 ISDN phone's status and allows actions to be taken either by using the phone's keypad or the mouse and keyboard; and
- A videophone call-control window.

The video call control window provides the following features:

- Call initiation and termination;
- Contrast, brightness, and color control;
- Audio and video muting;
- Still-image capturing; and
- Video input selection.

The video call window is the "launching pad" for the Vistium system collaboration utilities. The collaboration utilities are also available as a shrink-wrapped software product.

Standards for Screen-based Multimedia Terminals

As shown in Table I, many major elements of screen-based multimedia terminals are now governed by relevant international standards. In addition, new standards are currently being developed for DSVD applications, which are very similar to the screen-based applications mentioned earlier. In the DSVD situation, however, the term "data" is considered in a very general sense and could include motion video data transmission when needed by the application.

Interoperability of all the previously discussed products is a paramount concern. To prevent the mandatory recoding of video and its accompanying excessive delay, the H.324 standard requires that both the H.261 and H.263 video encoding algorithms be provided within H.324. Recoding of audio may be required but it should not increase delay or degrade quality unacceptably. Since the DSVD standards are just now emerging, it is too soon to determine what the effects will be when DSVD products interact with those based on H.320 or H.324.

Quality Factors

The following quality factors are of primary importance in the planning and configuration of a screen-based multimedia system:

- *Speech characteristics*, which include bit rate, delay, bandwidth, and complexity needed for compression;
- *Video attributes*, which include resolution, frame rate, absence of compression artifacts, complexity, and encoding delay;
- *End-to-end delay*, which includes user spontaneity, full

duplex audio, and absence of harmful echo;
- *Synchronization* between video and audio;
- *Transmission time* for still-frame pictures, data files, and shared applications;
- *Ease of use* of the product;
- *Interoperability* with other similar services, products, or features;
- *Availability of applications* to support the expected services and features;
- *Display system properties*, which include screen size, resolution, contrast, repaint speed, color rendition, and viewability effects (absence of glare, viewing angle, and so on); and
- *Camera system specifics*, which include resolution, color performance, light control, and ease and extent of user control.

Ideally, individual products blend these and other important attributes to best meet intended uses.

Conclusion

The multimedia product market is just beginning to emerge. Thus, it is very difficult to predict the future. Several assumptions, however, appear certain. For example, the demand for ever lower prices and expanding feature sets will continue. In addition, increasing focus will be placed on software rather than hardware because of the following three governing factors:
- Lower product costs;
- Flexibility of design, modification, and reuse; and
- More rapid time to market.

As previously mentioned, satisfying users' needs for applications that meet special demands will ultimately create the universal multimedia marketplace. Lower prices will be obtained through improved integrated circuit technology and by moving functionality from hardware to software. The economies of large scale production as supported by the adoption of global standards will also play a crucial role in enabling lower costs, just as it has for the PC, FAX machine, and VCR markets.

Multipoint applications, including video and data conferencing, screen-based work sharing, and collaboration will become increasingly important. In addition, global networking interoperation utilities—for example, WorldWorx™ Solutions—will become more powerful and pervasive in resolving areas of incompatible interoperation not addressed by standards.

Over the next five years, the performance and feature sets of videophones and audiographic units will continue to improve and prices will continue to fall. The following considerations will encourage and accelerate these improvements and price reductions:
- The need to eliminate travel;
- The necessity to interconnect individuals and groups easily and quickly;
- The adoption of international standards;
- The emergence of similar services, such as DSVD; and
- The continuing feature growth and adoption of the multimedia-capable PC.

During this five-year period, PCs most likely will realize full multimedia capabilities within the main processing complex of the motherboard. Full-performance, user-friendly multimedia telephones—tailored for specific applications—will also be widely deployed. Thus, the future for screen-based multimedia products will continue to be both exciting and challenging.

Acknowledgments

The authors thank the members of the AT&T Global VideoPhone Services, AT&T Image Solutions, and AT&T Global Information Solutions multimedia workstation teams who worked tirelessly to develop and produce the products discussed in this paper. The support of these teams, as well as the efforts of the many individuals who contributed to the preparation and review of this paper, are also gratefully acknowledged.

*Trademarks

MC68302 and 68000 are trademarks of Motorola, Inc.
Windows is a trademark of Microsoft Corp.
Pentium and 486 are trademarks of Intel Corp.

References

1. S. H. Early, A. Kuzma, and E. Dorsey, "The VideoPhone 2500 – Video Telephony on the Public Switched Telephone Network," *AT&T Technical Journal*, Vol. 72, No. 1, January/February 1993, pp. 22-32.
2. M. R. Schroeder and B. S. Atal, "Code-Excited Linear Prediction (CELP): High-Quality Speech at Very Low Bit Rates," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Tampa, Florida, March 1985.

3. P. Kroon and B. S. Atal, "On Improving the Performance of Pitch Predictors in Speech-Coding Systems," *Advances in Speech Coding*, Kluwer Academic Publishers, Boston, 1991, pp. 321-328.
4. Y. Shoham, "Constrained Stochastic Excitation Coding of Speech at 4.8 kb/s," *Advances in Speech Coding*, Kluwer Academic Publishers, Boston, 1991, pp. 339-348.
5. "Digital Compression and Coding of Continuous-Tone Still Images," *ISO Draft International Standard 10918*, Part 1.

(Manuscript approved August 1995)

Paul E. Crouch *is a technical manager and the research and development director for videophones in the Global VideoPhone Systems research and development department at AT&T Bell Laboratories in Holmdel, New Jersey. He is responsible for the development of videophone and related multimedia products, as well as associated technologies and platforms. Mr. Crouch has two degrees in electrical engineering, a B.S. from West Virginia University in Morgantown, and an M.S. from Ohio State University in Columbus. He joined AT&T in 1969.*

Jorge S. Rodriguez *is research and development vice president in the Image Solutions Development department at AT&T Image Solutions, a Global Business Communications Systems (GBCS) venture in Somerset, New Jersey. He is responsible for all departmental activities, including the development of the Picasso system and its associated PC software applications. Mr. Rodriguez has a B.S.E.E. from Rutgers University in New Brunswick, New Jersey, and an M.S.E.E. from Princeton University, also in New Jersey. Mr. Rodriguez, a recipient of the AT&T GBCS Trailblazers Award and AT&T Bell Laboratories Distinguished Member of Technical Staff Award, joined AT&T in 1980.*

William C. Schwartz *is director of engineering in the Multimedia Products department at AT&T Global Information Solutions, Naperville, Illinois. He is responsible for the development of PC-based telephony, videoconferencing, and collaboration software products. Mr. Schwartz has a B.S. degree from Purdue University in Lafayette, Indiana, and an M.S. from the University of Michigan in Ann Arbor, both in electrical engineering. He joined the company in 1966.*