The Provision of Intelligent Agent-Based Enhanced Multimedia Network Services

Marc Beacken
Robert L. Doran
Michael J. Geller
Julia Hirschberg

This paper describes some of the latest work in developing infrastructures that will provide customers with multimedia information. The paper discusses the concept of providing enhanced multimedia network services based upon intelligent agents (IAs), and the attributes of a user interface. Examples of emerging and future intelligent agent technologies are discussed, and delivery options for future enhanced multimedia services are examined. By evaluating trends in computer, networking, and multimedia processing technology, an approach is derived for creating and deploying enhanced multimedia network services. The paper concludes with examples of key components of the proposed infrastructure.

Introduction

We are all familiar with AT&T's television commercials describing some of the enhanced services that its network will provide in the not too distant future. For the public to accept these services, the human interface must be natural, simple, and enjoyable to use. This paper explores the development of *intelligent agents (IAs)*, also called *personal assistants*—the primary approach being taken to create this natural user interface. This approach has users interact with programs, databases, and even other users through *software agents*, which act as personal assistants to customize and automate certain functions for the user. These agents quietly perform simple routine tasks that might be tedious, repetitive, or burdensome to the user, such as monitoring stock performance or filtering e-mail.

Significant technological resources are required to implement intelligent agents and the enhanced services that they provide, including high-quality multimedia data. For example, a teleshopper probably will not purchase a dress before seeing a high-quality "real-life" display, such as a broadcast quality video.

However, users are accessing data with a variety of personal computer equipment having varying degrees of capability, so it will be up to the network to account for these hardware and software differences and often translate data for the user from one form to another. Similarly, network-based and third-party databases from which information is obtained will be using different storage methods, such as different compression techniques, so network equipment must quickly translate the data to the appropriate form for delivery, regardless of either the customer's equipment or database storage system.

We describe some of the latest work in developing infrastructures that will provide customers with multimedia information. This paper first discusses both the concept of providing enhanced multimedia network services based upon intelligent agents and the attributes of a user interface for the intelligent agents. Then it looks at examples of emerging and future intelligent agent technologies and the delivery options for future enhanced multimedia services. By evaluating trends in computer, networking, and multimedia processing technology, an approach is then derived for creating and deploying enhanced multimedia network services. The paper concludes with examples of key components of the proposed infrastructure.

IA Interfaces for Multimedia Services

An important goal of current research

in user interfaces is providing intelligent interfaces to data, processes, and services, so that people can find and do—naturally—what they want, when they want, and without having to specify more than they want. Users should be able to interact in ways that are simple and intuitively sensible to them, even though they are getting and transmitting information using multiple modalities.

Users should be able to interact with resources via speech or text in their own natural language, or via a stylus or pen with natural gestures. They should be able to receive information in ways that make the information easy to interpret, given their own requirements and the hardware and software resources available. And, while users should be able to specify preferred forms of presentation, they should not have to determine for themselves the best medium for the presentation or the details to achieve it.

One way to provide users with such capabilities is by using intelligent agents in user interfaces. While the research area is still being defined, there are several criteria for intelligent agents:

- **Goal orientation**: Users should be able to state *what* they want accomplished, rather than *how* they want it to be accomplished. The intelligent agent, in turn, should determine how to perform the necessary actions, without user supervision.
- **Cooperation**: Intelligent agents should interact with users in specifying goals where necessary and providing helpful feedback as appropriate, rather than just displaying simple, and often uninformative, error messages.
- **Expression**: Users should be able to formulate requests without having to learn an artificial query language or be limited by the constraints of a menu system. Spoken language, text input, or graphical interfaces are often the preferred media for interacting with intelligent agents.
- **Adaptability**: Intelligent agents should adapt to different users, both through direct requests and by "learning" the patterns of a user's behavior.
- **Autonomy**: Intelligent agents should have some independent decision-making capabilities and should be able to choose among several strategies for accomplishing a user's goals.
- **Integration**: The interface through which users and agents communicate should be understandable, consistent, and tuned to the task at hand. Users should not have to remember details of the task in order to specify their goals.
- **Security**: Intelligent agents often deal with information at varying levels of privacy, such as scheduling meetings, providing electronic mail, and analyzing financial information. Agents must be able to handle such information, including "learning" as they do so, in a secure way.

Intelligent agent applications that meet some of these criteria are just beginning to be developed as research prototypes. Agent applications now entering the commercial market, such as Microsoft's "Bob"* or Wildfire's messaging assistant, tend more to represent simple personifications of existing user software functionality, although these agents will become more "intelligent" in the near future.

The next section discusses some of the research prototypes for intelligent agent applications developed at AT&T and elsewhere, to give a flavor for the work currently under way. Some additional areas for research also are discussed.

Prototype Applications of IA Technology

Intelligent agent interfaces have been developed in research environments to perform such essential but mundane tasks as filtering e-mail, scheduling appointments between individuals or among groups of people, locating expertise within a large organization, retrieving (sometimes called "mining") information on the Internet, and managing personal messaging. While humans can perform these functions, the tasks are tedious and time-consuming and, as such, are ideal for intelligent agent applications.

Etzioni and Weld's "Internet Softbot,"[1] for software robot, provides an integrated interface to global Internet services via an intelligent agent. This intelligent agent uses Internet resources such as "archie," "gopher," and "netfind" to enable users to access personnel and information sources by specifying high-level requests.

For example, the softbot can handle requests such as "Send the budget memos to Mitchell at CMU," which involve determining which "Mitchell" the user is referring to, which document to send, which "Mitchell/CMU" address is the appropriate one, what medium to use in sending the document, what security measures are appropriate, and what to do under certain contingencies, such as "Mitchell" being out of town. Security and reliability issues are a particular focus of this research effort, as well as the importance of collaborating with the user to come up with a reasonable plan for satisfying the user's goals.

Designing agents that blend transparently into normal work environments is the goal of work on software agents that schedule visitors. In Kautz et al.,[2] the "Visitorbot" prototype employs both task- and user-specific agents to plan a visitor's presentations and meetings, using e-mail and a graphical interface to allow users to specify their constraints. Versions of this system are currently used to perform other mundane but time-consuming tasks, such as ordering books and orchestrating pizza orders for group lunches.

Intelligent agents that "learn" how to perform a task by watching users perform it themselves are exemplified by the electronic mail reader agents proposed by Lashkari et al.[3] This prototype e-mail filter employs personal agents for each user, and these agents use machine learning techniques not only to learn from their "own" user, but also to learn from other users' agents in cases where their user has not provided them with sufficient training material.

For example, user A's personal agent may learn that A always archives mail sent to a certain mailing list in a particular archive and may offer to perform this task automatically. When situations arise that A's agent has not previously observed—for example, when A joins a new mailing list—A's agent may ask other users' agents what behavior their users exhibited in such circumstances, and offer to provide the same actions automatically for A.

The ability to learn from particular users and to generalize from other users' behavior is important and useful property, allowing for the automation of customizing agents for individual users and permitting agents to collaborate in suggesting useful behavior to their users.

Enabling users to specify what their intelligent agents are to do for them in a simple and effective way is the focus of an AT&T Bell Laboratories researcher's work on intelligent agents in a messaging domain. This end-user programming approach to the problem of customizing agents permits users to program their messaging agent graphically, while an agent manager subsystem analyzes the proposed rules for potential interactions with previously defined rules, and warns users if conflicts are detected.

For example, a user may initially specify that messages from "Mom" and "Dad" are always to be put through immediately. Subsequently, the user may define the notion of "late night phone calls" and tell the messaging agent that these should not be put through but handled by voice mail. However, these rules potentially conflict: What should the messaging agent do with "late night phone calls" from "Mom" and "Dad"? In this proto-

Panel 1. Abbreviations, Acronyms, and Terms

AAL1 and AAL5 — ATM adaptation layers
ATM — asynchronous transfer mode
DCT — discrete cosine transform
DMA — direct memory access
DRAM — dynamic random access memory
DSP — digital signal processor
EMMI™ interface — AT&T broadband multimedia access interface
FIR — finite impulse response
fps — frames per second
GOPS — giga (billion) operations per second
H.261 — CCITT (ITU-TSS) standard for video conferencing applications that specifies the compressed-data syntax for encoded video over Px64 ISDN lines
HDTV — high definition television
IA — intelligent agent
JPEG — Joint Photographic Experts Group
MM — multimedia
MOPS — mega (million) operations per second
MPEG — Motion Picture Experts Group
NTSC — National Television Systems Committee
PC — personal computer
PVP — programmable video processors
RISC — reduced instruction set computer
STT — set-top terminal
TCP/IP — transmission control protocol/Internet protocol
VCP — video controller and processor
VRAM — volatile random access memory

type system, such conflicts are flagged and the user is prompted for a new rule to resolve the two conflicting rules.

Agents that handle "referral chaining," the problem of efficiently locating people with particular types of knowledge or expertise, without inundating a user with irrelevant requests for information, are the focus of work by Kautz et al.[2] and other Bell Laboratories researchers. "Pat" is an animated agent that tries to handle "referral chaining." Pat assumes the availability of a database that lists people and their skills, using the AT&T common skills language, and proposes members of new project teams based upon skill specification and availability.

Kautz et al. address the question of how constrained referral chaining might be accomplished without the explicit maintenance of a skills database. They propose mechanisms for inferring an individual's areas of expertise by browsing technical memoranda and other personal documentation that is available online.

The ability to infer information an individual human agent would not automatically have access to differentiates these intelligent agents from those that merely perform repetitive tasks that people could perform for themselves.

Future Directions of IA Interfaces

The field of intelligent agent interfaces is exciting a good deal of commercial interest as well. A major problem, though, is finding applications where technology exists to permit agents to take over certain human activities or to go beyond them. In general, tasks that are repetitive, require patience, detailed bookkeeping, and continuous monitoring appear to be likely targets for intelligent agent applications.

For example, monitoring large amounts of data regularly to look for changes in trends or particular events, such as checking stock prices, scanning online want ads, or browsing Internet locations, are tasks for intelligent agents. As online catalog shopping becomes available, wide-ranging, automatic browsing of catalog items according to specified user constraints—for example, finding a fringed suede jacket for under $300 that comes in an extra-large size, a collapsible pet carrier that meets airline specifications for safety, an out-of-print book, or an interactive keyboarding tutorial—could be likely applications for intelligent agents.

As messaging software becomes more sophisticated, it also seems likely that enhanced messaging and scheduling services via intelligent agents will become increasingly popular. The personification of intelligent agents to enhance the user's collaboration may require the blending of a real-time intelligent agent with a visual rendition of the intelligent agent's results. In this environment, synthetic and live video, animation, and graphics must be dynamically composed. New multimedia network infrastructures and services will, of course, be required to create this environment.

Realization of Intelligent Services

We have discussed the desired attributes of a user's interaction with an intelligent agent and speculated on future directions of intelligent agent functionality. Next we consider *how* these new services are now delivered and upon *which* platforms. In order to do so, trends in computers, networks, and multimedia processing technology are considered.

Computer Trends. Today, many pundits hold that the operating system, application, and platform wars have been fought and won—by Microsoft and Intel. Further, they argue that the personal computer (PC) is the network access device of choice for all future multimedia service offerings—starting with voice telephony.

Historically, the computer industry has been driven by one dominating company for relatively long periods of time. It would be reasonable to conclude that the PC will, therefore, take on an ever increasing user access role for everything from "plain old telephone service" to advanced multimedia services and applications.

We need only look and observe other key technologies and applications—available today—that support the case for this evolution. Armies of service and content providers are busy coding applications for PCs, and there are few barriers to entry for a service from a technology perspective. Indeed, applications such as telephone, multimedia conferencing, collaboration environments, and applications sharing are available today.

However, despite its ever-growing strength, the PC today is not ideally situated as an access device. First, it is difficult to configure and to use. Anyone who has powered up a PC just to play a CD-ROM appreciates the simplicity of just hitting the "on" button of a regular audio CD player.

Second, the PC's lack of interoperability and its overly layered, interacting complexity also often defeat its use for networked multimedia applications.

Third, the cost of software, the problems of keeping software up to date and functioning, and the array of overlapping product offerings also cause great angst within the user community.

Fourth, although 30 percent of the nation's households include at least one PC, it still is used primarily for games and word processing. For multimedia services to become commercially attractive to the largest possible audience, they must be easier to access than they currently are by a PC.

Others profess a more "balanced" perspective, pointing out that a mix of PCs, set-top terminals (STTs), and yes, even telephones will be used as access devices for some time to come. Again, a look at current or emerging offerings reveals a wide variety of delivery platforms fashioned as STTs. An STT is an interactive device on top of a television set that permits viewers to interact with a service via the television. In many instances, the nature of the STT attachment to the network is quite proprietary, at key levels of the communications stack. The effect by the vendor is to tightly control the selection of service and content providers, as well as to control the providers of end-user access platforms.

Thus, while PCs today may lack the capabilities necessary to become ubiquitous multimedia network access devices, the proprietary nature of STTs also are out of step with industry trends toward open systems, and they certainly lack the processing capabilities that the trends in multimedia suggest will be necessary in the future.

Networking Considerations. Within a few years there probably will be even more intense competition among a larger number of major carriers to provide traditional and enhanced local and long distance services. In addition, computer software industry leaders are poised to offer enhanced network services. Extensive deployment of new broadband transport, cell-based switching technologies, and updated pricing methods by the carriers and others threaten to cannibalize the market and render the existing circuit-switched infrastructure irrelevant to all services but basic telephony. Clearly, network providers must be able to adapt.

The end-user, having become used to locally attaching to CD-ROM drivers, expects the same services on a global proportion via a "sea-of-servers." The ability to access information on a global scale and to collaborate with others and their intelligent agents—*without* regard to the underlying transport specifics—is essential to users. The notion of ubiquitous communications—anytime, anywhere—originally pioneered by the telephone, is being extended by multimedia demands to bring the distant world much nearer to the user.

Thus, it becomes clear that if the network can provide multimedia services almost as simply as it is to use a telephone or an STT, while transparently providing the necessary processing capabilities, it will be in a position to make multimedia a commercial success.

Multimedia Processing Technology. Price and perceived value are two other important determiners for the successful adoption of enhanced multimedia network services. Processing video and audio signals in support of advanced services is demanding and costly from a computational point of view. Yet many applications are pushing their platform "envelopes," requiring faster processing, memory, and input-output.

In some situations, an end user may not demand near-broadcast quality video. In others, such as entertainment or lengthy distance learning sessions, an end user may demand TV quality video and high fidelity stereo audio—or better. Contrary to many claims, PCs today are not able to meet the processing needs of these high-end applications. Nor do many PC users have the skills for this purpose, although it must be noted that in five years, software and hardware advances could change the situation.

Next Generation of Signal Processing

There is emerging, for example, a new generation of programmable multimedia digital signal processing chips (MM DSPs). These MM DSPs are replacing the first generation fixed-function variety found in set-top boxes and video conferencing coder-decoders. As the numbers of video and audio standards that must be supported grow, these MM DSP devices become increasingly important. The introduction of these low-cost MM DSPs enable the introduction of software-based, reasonably priced, dynamic-feature mixing.

One would expect the evolution of these devices to follow the evolution of other DSP devices. Thus, we can assume that, eventually, the need for distinct types of MM DSP devices will be reduced, as the equivalent functionality is incorporated on PC and STT main processors and input-output architectures are enhanced for multimedia data. However, for high-end performance needs, this will not occur for some time.

The new generation of MM DSPs are incorporating the necessary features and functionality to process both audio and video signal types. The programmable video processor (PVP) portion of these next generation devices embodies significant flexibility features beyond the current generation. Several PVP features are described below.

Programmable Video Processors. High-performance programmable video processors, providing a platform for video compression and decompression for a variety of popular video compression standards, are becoming available from several chip vendors. The programmable video processor, an evolution of the previous generation of hard-coded video processing devices, has the advantage of being able to perform a variety of standard compression algorithms using the same hardware device.

The typical PVP is based on a high-performance programmable core processor with several special-function hardware units, often operating autonomously and in parallel. These devices are used in coder-decoders and will be applied to emerging devices like digital video camcorders.

The processing power needed to perform real-time video processing algorithms, particularly high-levels of compression, cannot be attained by today's general purpose processors. Even when the general purpose processor's capabilities reach the level necessary to do a single video-stream compression and decompression at a specific resolution, the PVP will still be a better platform for even higher resolution processing and delay-sensitive applications, such as

- High-definition television (HDTV);
- More computationally intensive compression algorithms;
- Multiple video-stream server applications;
- Still image compression and decompression; and
- Other computationally stressing real-time processing tasks.

Evolution of the PVP. The programmable video processor is a great improvement over hard-coded video processors. Hard-coded solutions were created because, given a particular technology, greater performance typically can be achieved with a hardware-only solution, as compared to a hardware/software solution. The main disadvantage of hard-coded solutions is that a new device needs to be designed for each video compression/decompression algorithm—a particularly difficult situation since, in the rapidly evolving video processing area, new compression standards are continuously emerging.

Several techniques have been developed to assist the hard-coded processor. One is the use of *parameterized control* over the compression/decompression algorithm. For example, a Joint Photographic Experts Group (JPEG) processor may allow different image resolutions, selectable quantization tables, and table-driven Huffman coding.

Another technique for lower development costs for hard-coded processors was the development of video processing *building blocks*, since many video compression techniques are built upon similar algorithms, such as the discrete cosine transform. This approach led to hard-coded processors being built from individual devices, resulting in a low level of integration and a high and recurring design cost.

Typical PVP Architecture. The programmable video processor builds on the idea of video processing building blocks. A programmable control-core, surrounded by various hardware-assist blocks, with some blocks operating autonomously and in parallel, comprise the typical PVP architecture.

The major blocks of a typical programmable video processor are:

Programmable control core. This is typically a high-speed, general purpose reduced instruction set computer (RISC) used to control the other PVP blocks. Not involved with data movement and operations on a pixel-by-pixel basis, it generally controls video data flow and makes processing decisions on a video macroblock level.

Multi-pipe arithmetic unit. This block is used for high-speed arithmetic processing such as discrete cosine transform (DCT), quantization, and macroblock differencing for estimating interframe motion. It usually has a local high-speed memory or register file.

Memory interface. A wide interface to dynamic random access memory (DRAM) or variable random access memory (VRAM), 32 or 64 bits wide, it has direct memory access (DMA) features to use maximum data transfer rates to and from the arithmetic unit local memory or register file.

Huffman Encoder/Decoder. A hardware logic block, it is used to perform variable length encoding and decoding of the quantized DCT coefficients. Although the actual encoding values differ, this process is used in H.261, JPEG, and Motion Picture Experts Group (MPEG-1 and MPEG-2) coding schemes.

Video Interface. The video interface provides the necessary real-time interface to video encoder and decoder circuitry from the local memory or the register file. Pre- and post-filtering options are included to provide the needed interpolation or decimation to match video signal representations. To accomplish this, logic for horizontal subsampling, or finite impulse response (FIR) filtering, and temporal (frame-to-frame) filtering on video inputs, as well as horizontal oversampling and noise reduction of video outputs, is provided.

Comparison of Processors

General purpose processors lack the hardware assist blocks and multiple memory ports found in a PVP. The PVP technology will itself be driven towards higher levels of integration and lower cost due to its integration into consumer video appliances, such as camcorders, TVs, and STTs.

The processing power needed to perform video processing algorithms in real time, particularly compression, cannot be attained by today's general purpose processors. Even when the general purpose processor's capabilities reach the level necessary to perform a single video-stream compression and decompression at a specific resolution, the PVP is still a better solution.

This combination of video controller and processor (VCP) hardware, although specialized for video processing, provides tremendous processing power. For example, Integrated Information Technology Inc. rates its VCP processor at 4 giga-operations per second (GOPS). Because of their specialized memory interfaces and bus structure, video processors are capable of transferring large amounts of video data to the various processing blocks of the specialized programmable video processors.

For a general purpose processor to perform equivalent video compression/decompression algorithms as a video processor, it would need to perform approximately 4 GOPS. Today's general processors, however, are capable of several hundred mega-OPS (MOPS). If general purpose processor performance continues to increase at a similar rate as indicated in the recent past, the processors will reach a comparable level of performance with VCPs within 10 years.

In summary, the programmable video processor provides a solution today for real-time compression and decompression of full-frame rates of 30 frames per second (fps), full-resolution video images. The programmability allows various compression algorithms to be performed on the same processing platform. General processors will eventually have enough processing power to challenge the programmable video processor in today's applications. However, more processing intensive applications and multiple video-stream server applications will allow the programmable video processor to remain an important device for video compression and decompression for some time to come.

Creating the Services Infrastructure

From the discussion above, we assert that many users desire an affordable, simple, low-maintenance, and personalized portal into enhanced multimedia network services. Collaboration with other people or intelligent agents and exploiting information that is widely distributed and independent of the underlying transport are essential.

Processing capacity must be sufficiently robust to support the intelligent agents' activities in tailoring information presentations to the user and accepting requests from the user. Information objects span audio, video, graphics, images, and data types. For example, video composition may be performed for presentation to the user, while user video input may be processed for gesture recognition.

A network infrastructure supporting multimedia may be constructed by focusing on affordable end-points with intelligent agent user interfaces, and attached network intelligence close to users at the "edges" of the network, to support intelligent agents on behalf of many users. In doing so, AT&T may create differentiators from the competition by providing end-to-end enhanced multimedia network services to the end-user.

Examples of Infrastructure Building Blocks

This section reviews existing components of the proposed infrastructure and speculates on the future of these components.

Figure 1. AT&T's EMMI™ access interface, to the right of the workstation, provides broadband multimedia access that is independent of any type of attached host workstation, whether it is based on UNIX,* MS-DOS* with Windows,* or Macintosh* operating systems.

Access: Endpoints and Network Edges. Essential for widespread acceptance of multimedia services is the provision of broadcast-quality video, CD-quality audio, and easy-to-use application-specific collaboration facilities. A glimpse of the multimedia quality required may be gleaned from the early multimedia applications of the broadband asynchronous transfer mode (ATM) infrastructure.

Let's look at just one such device, AT&T Advanced Technology Systems' EMMI™ broadband multimedia network access interface (see Figure 1). The EMMI operation is independent of any specific type of attached host workstation, regardless of whether it is based on the UNIX,* MS-DOS* with Windows,* or Macintosh* operating systems. The EMMI interface has attributes of both a network *edge device* and a customer *endpoint device* and, thus, is called a network *edgepoint.*

The EMMI interface directly accepts National Television Systems Committee (NTSC) video, line-level stereo audio inputs, and also computer data from an attached host. Similarly, the EMMI interface generates NTSC video, CD-quality stereo line outputs, and computer data to an attached host. The EMMI interface appears to an attached host as a host network interface, providing standard TCP/IP transport over ATM,

The EMMI video compression technique is based upon the JPEG motion compression algorithm. Audio and video information is mapped directly onto ATM using ATM adaptation layers AAL1 and AAL5, respectively. The EMMI interface is controllable from the local host, from a remote EMMI host, or from within the network itself.

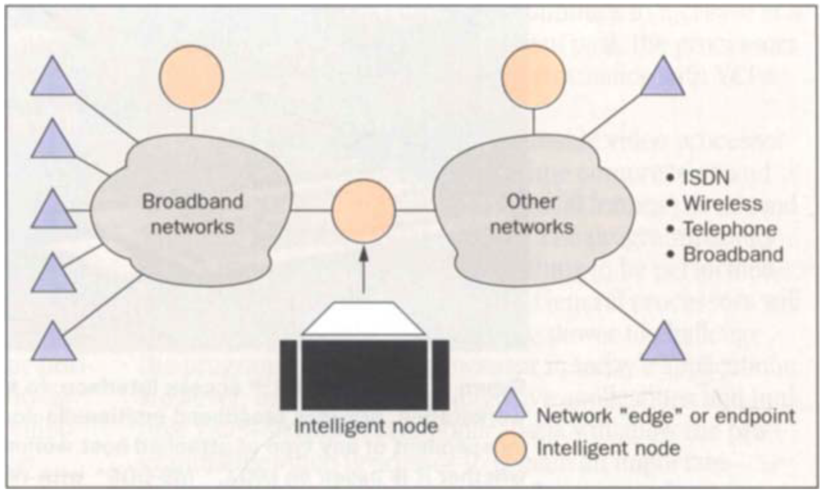
Because of it's high fidelity video and audio, the EMMI interface is initially being used in broadband multimedia experiments, trials, and production systems for telemedicine, distance learning, and business and scientific applications. It is likely that these high-end applications may provide a hint of the expected performance needs for future wide-spread use of enhanced multimedia network services.

Network Infrastructure. The required underlying network infrastructure must provide for a seamless integration of networks with different transport capabilities carrying a variety of "standard" multimedia information streams and control protocols. Processing is required within the infrastructure to support the intelligent agent capabilities described above.

Transcoding and multimedia composition among various multimedia representations, to meet endpoint or transport constraints, may be considered a special case of *intelligent object* or *intelligent data filtering.* An example of this concept of intelligent data filtering is the case where data representation is transformed based upon user preferences, network capacity, and real-time constraints.

The filtering may be trivial, such as sending an icon instead of an entire image, or more complex, such

Figure 2. An intelligent node is a network infrastructure that provides a seamless integration of multimedia information, flexible multimedia processing (video composition and audio bridging), and multimedia session capture, edit, and access.



as extracting critical parameters of interest to the user from large amounts of transmitted information. The filtering may evolve over time, based upon the system "learning" due to prior user interactions.

A network infrastructure that provides a seamless integration of multimedia information, flexible multimedia processing (video composition and audio bridging), and multimedia session capture, edit, and access is called an *intelligent node* (see Figure 2.). The intelligent node incorporates a multiprocessor architecture, featuring the next generation in programmable multimedia signal processors to create a platform for enhanced multimedia network services. Such a node is not like the switched-centered nodes of traditional telephony, but is a server-based application associated with the network and tightly coupled in function to the customer.

The intelligent node provides a powerful session reservation, scheduling, and control system using extensible World Wide Web-based cross-platform user interfaces. User presentation of the multimedia streams produced by the intelligent node is intended to scale from single-display multimedia workstation environments through large-scale multiple-display systems with several channels of audio.

The services provided by the intelligent node facilitate its role as a dynamic intermediary between the application and currently available network capability. The intelligent node services accept the application and session-specified multimedia requirements for fidelity or quality of service and negotiate with the underlying network resource management agents on how to best accommodate the specified needs. Only if a reasonable negotiation of communication parameters cannot be concluded will application involvement be necessary.

Conclusion

We have described some of the critical work being performed to develop intelligent agents, as well as some of the underlying multimedia infrastructures being developed to implement them and the services they perform. Ultimately they will make interacting with the AT&T network natural and simple enough so that the many enhanced services envisioned, and those yet to be envisioned, will become a commercial reality.

Some of the prototype intelligent agent applications were reviewed in order to give the reader a feel for where we are in the maturation process, as well as for some of the research directions being undertaken. It is clear that to achieve the goals of simple and natural interaction, and to provide creative new services, AT&T's network must interact with its users with video, audio, data, and images.

It is also clear that AT&T's network must be able to deal with a wide variety of user equipment that is capable of different levels of performance and adhere to multiple standards and data formats. We've described some of the new key technologies that are being developed to handle multiple mediums, as well as some of the new technology that allows us to process video data in a very flexible manner.

This new technology is being incorporated into new equipment, which is capable of interacting with a variety of communications and data, such as compression standards, and is flexible enough to deal with a spectrum of user capabilities. A particular example of a high-quality multimedia access device, AT&T's EMMI multimedia interface, was briefly introduced. We can expect to see continued efforts on both natural and simple-to-use intelligent agent network user interfaces and the underlying infrastructure for providing enhanced network services.

(Manuscript approved August 1995)

Acknowlegements

We express our thanks to Loren Terveen and also to Allen E. Milewski and Steven H. Lewis, Bell Laboratories researchers whose work on intelligent agents, as well as the work of those cited as references, was beneficial in the preparation of this paper.

*** Trademarks**

Bob is a registered trademark of Microsoft Corporation.
Macintosh is a registered trademark of Apple Computers.
MS-DOS is a registered trademark of Microsoft Corporation.
UNIX is a registered trademark of Novell in the United States and other countries, licensed exclusively through X/Open Company Limited.
Windows is a registered trademark of Microsoft Corporation.

References

1. Oren Etzioni and Daniel Weld, "A softbot-based interface to the Internet," *Communications of the Association for Computing Machinery*, Vol. 37, No. 7, July 1994, pp. 72-76.
2. Henry Kautz, Bart Selman, Michael Coen, Steven Ketchpel, and Chris Ramming, "An experiment in the design of software agents," *Proceedings of the Twelfth National Conference on Artificial Intelligence*, Vol. 1, July-August 1994, pp. 438-443.
3. Yezdi Laskari, Max Metral, and Pattie Maes, "Collaborative interface agents," *Proceedings of the Twelfth National Conference on Artificial Intelligence*, Vol. 1, July-August 1994, pp. 444-449.

Marc Beacken *is technical manager of the Broadband Multimedia Network Infrastructures and Services Group at AT&T Bell Laboratories QUEST/ACT organization in Whippany, New Jersey. His group is responsible for developing new multimedia transport, infrastructures, and service platforms. He joined the company in 1983. He has a B.S.E.E degree from Rutgers University in New Brunswick, New Jersey, and an M.S.E.E degree from Rensselaer Polytechnic Institute in Troy, New York.*

Robert L. Doran *is a member of technical staff in the Advanced Multimedia Communications Department at AT&T Bell Laboratories QUEST organization in Whippany, New Jersey. He is responsible for systems and electrical design for several multimedia products. He joined the company in 1987. He has a B.S.E.E. degree from Wikes University in Wilkes-Barre, Pennsylvania, and a M.S.E.E. degree from Stanford University in Palo Alto, California.*

Michael J. Geller *is director of New Business Development at AT&T Advanced Technology Systems in Whippany, New Jersey. His organization is responsible for creating new, diversified business through the application of advanced technology systems, products, and services within government markets and selective commercial markets. He joined the company 1978. He has a B.S. degree in mathematics from Rensselaer Polytechnic Institute in Troy, New York, an M.S. degree in mathematics from the University of Connecticut in Storrs, and is a Ph.D. candidate in mathematics at Polytechnic University in Brooklyn, New York.*

Julia Hirschberg *is head of the Human-Computer Interface Department at AT&T Bell Laboratories in Murray Hill, New Jersey. Her organization works on speech synthesis, computational linguistics, and human-computer interfaces. She joined the company in 1985. She has a B.A. degree in history from Florida Presbyterian College in St. Petersburg; M.A. and Ph.D. degrees in history, and an M.A. degree in comparative studies, all from the University of Michigan in Ann Arbor; and M.S. and Ph.D. degrees in computer science from the University of Pennsylvania in Philadelphia.*