Multimedia Teletutoring over a Trans-European ATM Network

Yu-Hong Pusztaszeri*, Maurice Alou*, Ernst W. Biersack**, Philippe Dubois**, Jean-Paul Gaspoz*, Pascal Gros**, Jean-Pierre Hubaux*

*Department of Electrical Engineering
Swiss Federal Institute of Technology, Lausanne (EPFL)
CH-1015 Lausanne, Switzerland

**Institut Eurecom,
2229 Route des Crêtes, B.P. 193,
F-06904 Sophia Antipolis, France

Abstract. We describe the multimedia teletutoring application jointly developed by EPFL and Eurecom in the context of the first 34 Mbits/s Trans-European ATM network interconnecting sites in France and Switzerland. This network was called the Broadband Exchange over Trans-European Links (BETEL). The aim of this paper is to describe the BETEL teletutoring platform and its building blocks, together with performance evaluation, limitations and future enhancements of this prototype. Focus is placed on the interactive audio and video communication part of the application.

1 Introduction

The trend in today's telecommunication networks is a migration towards Broadband Integrated Service Digital Networks (B-ISDN) to support integrated high-speed data, voice and video communications. Asynchronous Transfer Mode (ATM) is the packet switching and multiplexing technique chosen for B-ISDN to provide services with different Quality of Service (QoS) requirements. Meanwhile, new video and audio coding standards are emerging, and many commercial products, both hardware and software, are now available to integrate audio and video with conventional digital data communications.

With this in mind, the European Parliament launched the DIVON program (Demonstration of Interworking Via Optical Networks) in 1992 to prepare and promote ATM technology and new B-ISDN services in Europe. The BETEL project, funded by the European Commission and the Swiss Federal Office for Science and Education, was one of the four projects in this program. The aim of BETEL was to run user driven applications over one of the first 34 Mbits/s international ATM networks in the World.

A teletutoring application was demonstrated between EPFL and Eurecom. It used videoconferencing and shared workspace tools to allow interactions between a teacher at EPFL and a group of students located at Eurecom. The teacher at one site could teach a class or supervise students in their individual work in another site while the students could seek assistance from their teacher located several hundred kilometers away.

The goal of this paper is to give an overview of the BETEL teletutoring application. Section 2 describes the BETEL teletutoring platform and its building blocks, while Section 3 is devoted to an interactive audio and video communication tool developed for this prototype and Section 4 contains a brief description of echo cancellers built for this experiment. Section 5 gives detailed performance studies of this prototype. Finally, Section 6 discusses the limitations and future enhancements of the system.

2 An Overview of the BETEL Teletutoring Application

2.1 The BETEL Network

The BETEL network infrastructure shown in Figure 1, is based on ATM technology, and supports FDDI LAN interconnection, ATM transfer service, and AAL 3/4 data service [1]. The FDDI LANs at EPFL and Eurecom were interconnected to the BETEL network by means of Cisco routers.

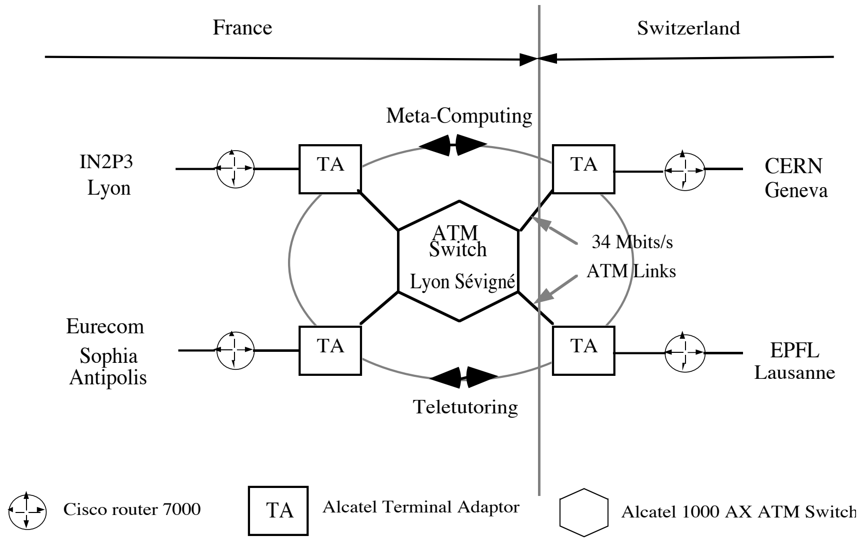
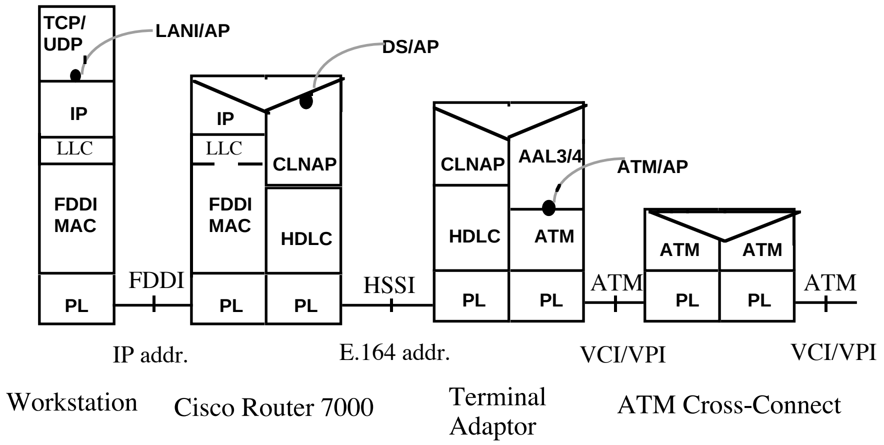


Fig. 1. BETEL: Europe's first operational ATM network

The end system protocol stack was imposed by the Cisco routers. Standard Internet protocol suite and FDDI protocols were used. The protocol stacks at the interworking units and workstations are given in Figure 2. The Switched Multimegabit Data

Service (SMDS) and point-to-point ATM connections were implemented but multiplexing of several Virtual Channel Connections (VC) over a Virtual Path Connection (VP) was not available in the BETEL network. A separate VP connection was thus established between each BETEL user sites, forming a fully meshed VP network.



CLNAP : ConnectionLess Network Access Protocol
HSSI : High Speed Serial Interface
HDLC : High-level Data Link Control
AP : Access Point
PL : Physical Layer

Fig. 2. Protocol stacks in the BETEL teletutoring network

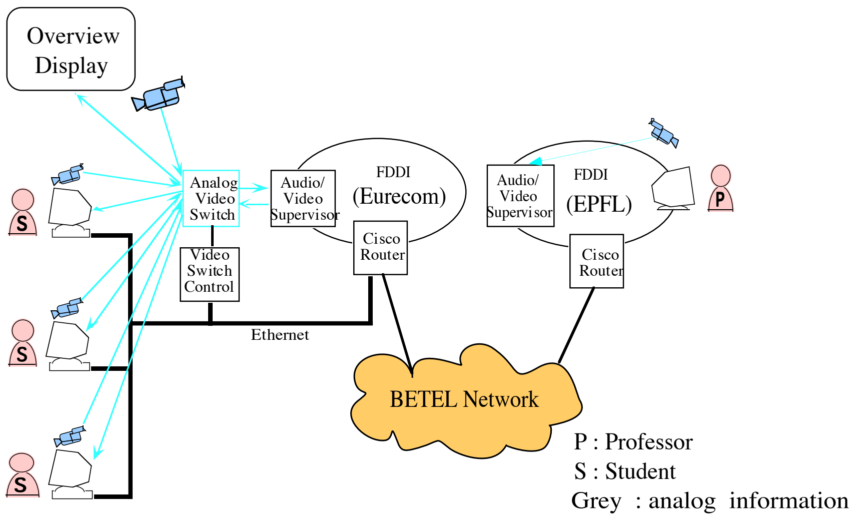


Fig. 3. Teletutoring network infrastructure at EPFL and Eurecom

2.2 Teletutoring Network Infrastructure at EPFL and Eurecom

The teletutoring network infrastructure at EPFL and Eurecom is shown in Figure 3. The network topology at Eurecom is more complex than at EPFL. At Eurecom, an FDDI ring was dedicated for high speed audio and video transmission, and an Ethernet was used to connect student workstations to the BETEL platform to support connection control and shared workspace data communications. Because the shared sessions generated relatively low data rate, it was not required to connect student workstations to the FDDI ring.

Moreover, the distribution of audio and video signals at the student site at Eurecom used an analog audio/video switch. All cameras, microphones, monitors, and loudspeakers were connected to the switch, which was controlled by dedicated software to establish and release audio, video and data connections. Using the existing analog infrastructure at the site provided a cheap implementation strategy since video compression hardware was not required for each student workstation.

2.3 Application Building Blocks

Commercial workstations were used to build this prototype. Sun Sparc10 stations equipped with Parallax boards were used for audio and video acquisition and transmission, while Hewlett-Packard (HP) workstations were used as workspaces which could be shared between the teacher and the students using SharedX, a Shared Workspace Manager provided by HP.

Connection Control and User Interface modules [2] provided an intuitive and user friendly access to the Audio-Video Supervisor and Shared Workspace Manager. Moreover, echo cancellers were used to reduce echoes generated in the BETEL teletutoring network. The building blocks were as follows:

- Hardware:
  - HP 9000/700 workstations
  - SUN Sparc10 stations
  - Parallax video acquisition board
  - Echo canceller

- Software Modules:
  - User Interface
  - Connection Control
  - Shared Workspace Manager
  - Audio - Video Supervisor

3 Audio - Video Supervisor (AVS)

AVS provided real-time audio and video acquisition and end-to-end transmission. Under the supervision of AVS, audio and video signals from analog sources were digitized and encoded, and then were transmitted to a remote station via the BETEL network. At the receiver end, the data were decoded, and video images were reconstructed and displayed while audio was being replayed. Audio and video signals were handled in a similar fashion (see Figure 4).

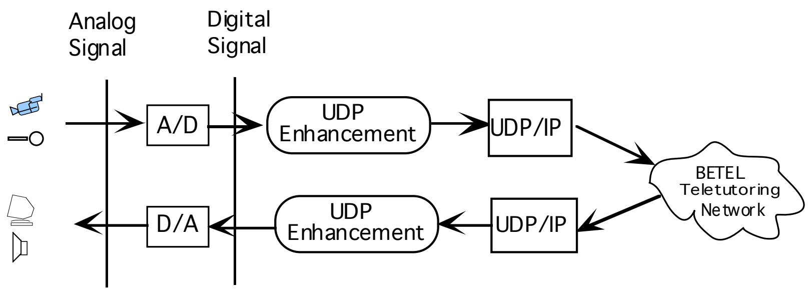


Fig. 4. AVS processing pipeline

3.1 Video Acquisition

The current AVS implementation used the Parallax XVideo board. This board was the only hardware available on the market which permitted real-time video compression and decompression at a reasonable frame rate. Details about this board and its performance are given in [3].

The Parallax board can handle analog video input and output in various standards (PAL and NTSC) and formats (YUV, RGB, super VHS, and Composite). The video signals are first digitized, then compressed by the XVideo board based on the JPEG standard before being sent to the network. On the receiver side, the digital video signals are decompressed, converted to the analog signals and displayed on the receiver's monitor.

The programming interface that came with this board uses an extension of the X11 library called XVideoToolkit. Its functionality is exploited by AVS through an extended X-Window server, which provides access to Parallax's graphical accelerator and frame buffer, and supervises video digitization and compression / decompression.

One of the major drawbacks of this board resides in digitized video images, which have to be first stored in XVideo's frame buffer then compressed by the JPEG Image Compressor on the XVideo board. Hence, video images cannot be compressed without being first displayed locally.

3.2 Audio Acquisition

Audio streams are digitized, recorded and played by a SpeakerBox on the Sun Sparc10 station. The SpeakerBox audio peripheral provides an integral monaural speaker and microphone, stereo line in/out and headphone connections. This SpeakerBox supports different audio qualities and encoding techniques. Moreover, it has a programmable audio device interface.

3.3 Networking Issues for Audio and Video Transmission

Real-time audio and video transport service imposes several performance requirements on the network. Since both audio and video sources produce continuous data streams, not only do their temporal relationships have to be satisfied, but they also require a large network bandwidth. In summary, interactive audio and video data generated by the teletutoring application impose the following network requirements:

- guaranteed high throughputs
- bounded end-to-end delay and delay jitter
- low loss and error rates
- connection-oriented service, i.e., in sequence delivery
- support for real-time data service
  - higher priority for real-time data
  - selective discarding of data according to their priority in case of congestion
- synchronization
  - intra-medium
  - inter-media
- adaptive and preventive rate-based flow control

The transport layer protocol was restricted to TCP and UDP since IP was imposed by the Cisco router in the BETEL network. TCP provides reliable end-to-end connection oriented transport service while UDP and IP support best effort services based on connectionless techniques. The Internet protocol suite was designed for point-to-point non-real-time data service and has many difficulties to meet the network performance requirements demanded by the teletutoring application.

TCP/IP is unsuited for a network with a large bandwidth-latency product. The BETEL network is one of such networks. The sliding-window flow control with credit allocation does not allow to use the full bandwidth of the BETEL network. Retransmission in TCP significantly increases end-to-end delays and is unsuitable for interactive audio and video data transport service. Hence, window based flow control and error control mechanisms found in TCP create problems for real-time audio and video transmission, as in the BETEL context.

On the other hand, the lack of retransmission mechanism makes UDP a better candidate to transport real-time audio and video data. Since UDP does not guarantee in sequence delivery and does not have any error or flow control, some end system enhancements added to UDP are needed. For instance, video frames in general are larger than the UDP datagram limit (9 Kilobytes), thus video frames need to be segmented into smaller frames before being sent to a UDP socket and be reassembled together at the receiver end. In order to make the reassembly process efficient, missing frames and out of order frames have to be detected. Minimum UDP enhancements are loss detection and packetization (including segmentation and reassembly for video frames). Therefore, UDP/IP protocols with these enhancements at the end system were used to transport audio and video data in AVS.

3.4 Implementation and Performance Issues

AVS was designed to guarantee optimal performance. This was done with as little data movements and copying as possible. Only minimum UDP enhancements were implemented. Data were packetized (segmented if necessary) and sent to a UDP socket without any buffering and copying. A small header was added to each packet. The frame sequence information (for loss sequence detection and sequence check) was put in the header. For video frames, the sequence number and the total number of segments were also needed in the header. A typical video packet size was about 4 Kilobytes, and so was the MTU in the BETEL network (excluding the Ethernet segment since no audio and video information was distributed across Ethernet using UDP). On the other hand, audio frames had 128 audio samples (one byte per sample in this context) to ensure low delay and loss rate.

The audio quality was the most important factor in the design of this teletutoring prototype. Voice is the most common and effective means of human communication, although eye contact and other facial information are also important. Some steps were thus taken to ensure optimal audio quality. The use of smaller audio frames was one. Another was the use of high quality sound equipment. The Sun microphone did not give as good an audio quality as did the semi-professional microphones, so the latter was used (connected to line in port of the SpeakerBox). In addition, echo cancellers were necessary to reduce echoes generated by the large round trip delay in the BETEL network. Further, it was impossible to use CD quality audio supported by the SpeakerBox because the echo canceller used could not treat audio with sampling frequency higher than 8KHz. Therefore, audio was restricted to telephony quality. Audio data in AVS were not compressed to reduce processing delay to ensure good audio quality, as performance studies in Section 5 show that the performance bottleneck of this prototype is not in the network.

On the other hand, one may ask why teletutoring applications and other interactive multimedia applications need broadband networks nowadays? The most obvious answer is video quality. High quality digital video signals demand a large bandwidth. Since the XVideo board is one of the most preferment real-time compression hardware available, it was used to have the optimal video performance. In addition, using a pair of Sun Sparc10 stations to send and receive video frames did not obtain high video frame rates and video bit rates and cannot fully exploit the 34 Mbits/s high-speed link for this teletutoring application. In order to overcome this problem, a Sun Sparc10 station was dedicated to either transmit or receive video sequences. As AVS consists of four independent processes, each process is used for receiving or sending video and audio data. Transmission of audio and video is hence independent. Therefore, four Sparc10 stations equipped with Parallax boards were involved in the video acquisition and transmission of this prototype.

4 Echo Cancellation Through Adaptive Filtering

The adaptive filtering technique [4] used the Least Mean Square (LMS) algorithm, which has the advantage that no prior knowledge of the room impulse response is required. The adaptive filtering shown in Figure 5, consists of two distinct steps:

- estimating an filtering error

$$e(k+1) = y(k+1) - \vec{g}^{T}(k) \cdot \vec{x}(k+1) \qquad (1)$$

- updating the coefficients $\vec{g}(k)$, using the error e(k+1).

$$\vec{g}(k+1) = \vec{g}(k) + Ke(k+1)\vec{x}(k+1) \qquad (2)$$

Where

$$\vec{x}(k) = \begin{bmatrix} x(k) \\ x(k-1) \\ \cdots \\ x(k-N+1) \end{bmatrix} \qquad \vec{g}(k) = \begin{bmatrix} g_0(k) \\ g_1(k) \\ \cdots \\ g_{N-1}(k) \end{bmatrix}$$

x(k) represents the sample at instant k,
$\vec{x}(k)$ represents the vector of the N most recent samples at time k
$\vec{g}(k)$ is the vector of the N filter coefficients at time k
y(k) is the sample coming from the microphone at time k
e(k) is the error at time k
K represents the adaptation step

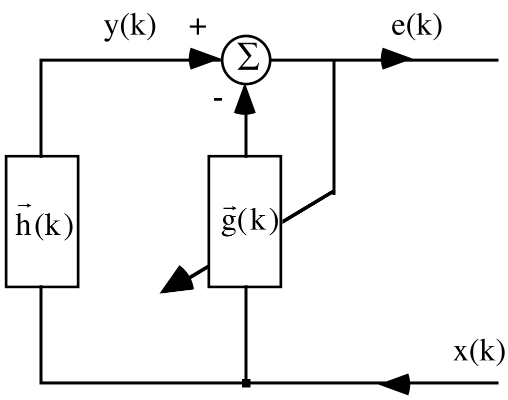


Fig. 5. Feedback of e(k) on the filter coefficients with emphasis on the adaptation process

The convergence time and stability of this system depend on the value of the adaptation step K, which depends on the length of the filter N and on the input signal power [2].

The measured echo attenuation was 20 dB in a given test room and the signal decreased by 20 dB in 102.5 milliseconds. The performance of this echo canceller is closely related to the room acoustics.

5 Performance Evaluation of the Teletutoring Platform

5.1 Measured raw TCP/IP and UDP/IP Performance Without Applications

The purpose of this study is to estimate the upper bound in performance which is available for applications running on top of TCP and UDP. The TCP/IP and UDP/IP throughputs were measured (using ttcp tool) both in the local FDDI environment and on the BETEL teletutoring platform [2].

The measured performance in the local FDDI LAN environment is shown in Figure 6. UDP/IP (without UDP checksum) could achieve 58.6 Mbits/s throughput for a message size of 4K bytes but with a maximum of 20% losses while TCP/IP had a maximum throughput of 40 Mbits/s for the same message size.

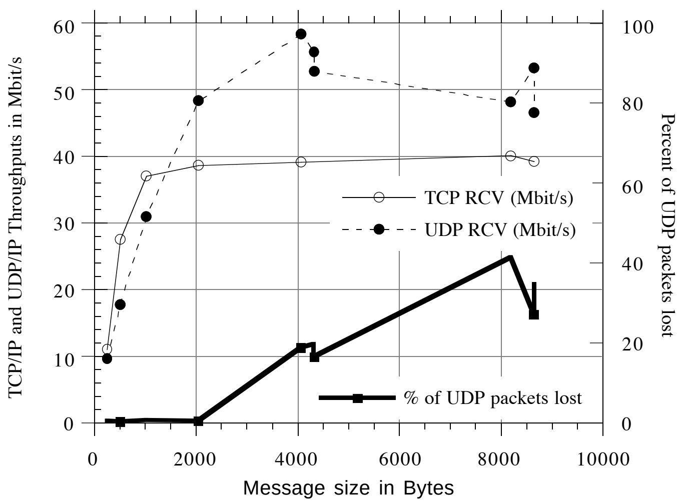


Fig. 6. Measured performance between two Sun Sparc10stations over FDDI

On the other hand, TCP/IP throughputs stabilize at around 8.4 Mbits/s in BETEL teletutoring network for message sizes above 2000 bytes. The Round Trip Time (workstation back to back) of the EPFL-Eurecom BETEL link is about 12 milliseconds for 64 byte messages and 17 milliseconds for messages of 1024 bytes.

These results show that the BETEL teletutoring network has a large bandwidth-latency product and that UDP is unreliable and suffers from losses but UDP can attain higher throughputs than TCP. Hence, at least a 8.4 Mbits/s of bandwidth on top of UDP/IP is available for real-time audio and video data communications between EPFL and Eurecom using the BETEL teletutoring network.

5.2 Measured Video Performance

Video communications demand a large bandwidth and their peak performance is likely to be limited either by the network or by the end system. Understanding the parameters influencing video performances can help us to obtain the best video quality. By measuring video bit rates and frame rates, we can gain an objective insight into video performance and hence identify the bottleneck of this prototype. The performance measurements in this study were taken between two Sun Sparc10 stations in an FDDI LAN at EPFL.

First, video bit rate depends on the Q factor (ranging from 25 to 1000) which is used by the XVideo board to determine the quantization level and to control the compression factor. The higher the Q factor and the larger the compression factor, the lower the video bit rate and video quality. Another important factor is the video frame rate, defined as the number of video frames captured and played per second. Figure 7 shows the relationship between the measured unidirectional video bit rates using tcpdump [5] and the number of frames captured per second when only a quarter of PAL resolution was used and the Q factor was at 50. The video bit rate increases proportionally to video frame rate.

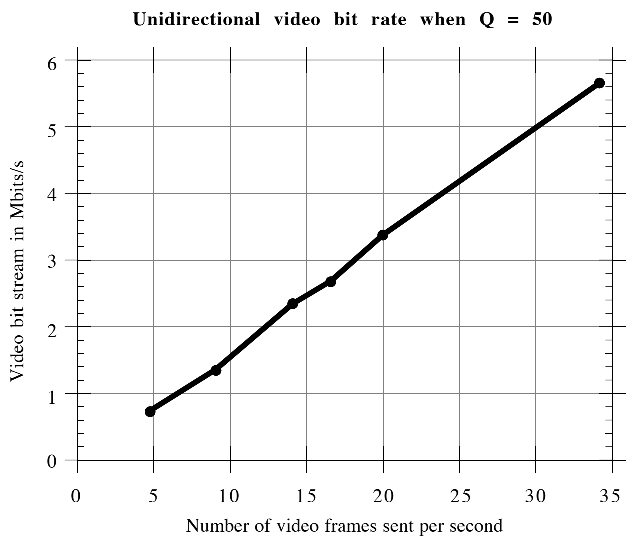


Fig. 7. Measured unidirectional video bit rate between two Sun Sparc10 stations over FDDI using UDP/IP

When we pushed the system to its maximum capability, we can reach 34 video frames captured and sent per second at a video bit rate of 5.7 Mbits/s. Similarly, transmitting the full PAL resolution video images of with the same Q factor, the system can reach only 12 video frames per second and can generate a video bit rate of 5.8 Mbits/s. As a consequence, the measurements from section 5.1 show that the bandwidth available for video communications in BETEL teletutoring platform is larger than the maximum video bit rate which the end system can deliver. It is clear that the performance bottleneck lies in the end system and not in the network.

6 Limitation and future enhancements

The implementation philosophy of BETEL was to integrate currently available technology and build a demonstrator within a year. The teletutoring prototype inherited the limitations of the current technology. This teletutoring experiment used a hardware dependent, point-to-point configuration (i.e., EPFL-Eurecom), and used the UNIX operating system and the Internet protocol stacks. There was no built-in synchronization and rate-control mechanisms implemented in the videoconferencing system. Therefore, enhancements will be needed in the following areas: multipoint and multiplatform teletutoring configurations, and system support for interactive real-time teletutoring applications.

6.1 Hardware Dependency

The hardware dependency could be relaxed as the video compression and decompression hardware and shared workspace tools were progressively made available. The release of a Parallax board for the HP platform had been announced for the Spring of 1994. AVS may then be easily ported to the HP platform since the Parallax boards (both for Sun and HP platforms) are using the same C-cube chips which are based on the JPEG compression/decompression standard. In addition, AVS can be also modified to use other video compression hardware, for instance, those based on the MPEG standard when they become available. Moreover, Sun has recently released a commercial product called ShowMe2, a competitor to SharedX. Unlike its predecessor (ShowMe), ShowMe2 can be used to share applications, allowing both videoconferencing and shared workspace tools to be integrated on the Sun platform. The BETEL teletutoring prototype could then be ported to multiple platforms.

6.2 Audio Quality

Audio quality is of paramount importance in teletutoring, but in this context it was hampered by echo, which was a serious problem because of large latency that audio experienced in the BETEL link. Several echo cancellers were designed to cancel this effect, but these devices could only process one speaker at a time and created problems when used with audio mixing devices. An audio enhancement would be to design new echo cancellation algorithms which support CD quality audio and can be used with mixing devices. This would benefit mostly a teletutoring scenario with geographically dispersed students.

6.3 Scalability

The current prototype is limited to point-to-point communications. A teacher can interact with one student at a time, although his image and voice can be broadcast to everyone in the classroom. Students may not engage in a discussion with each other. All video and audio signals have to be transported via point-to-point audio and video connections. It would be useful if either the teacher could simultaneously supervise several students from different sites, or several teachers from different sites could interact together. Thus, a fully meshed digital multipoint videoconferencing is needed.

6.4 System Support for Teletutoring

One of the long term solutions to the performance problem is to implement end system support and network support for interactive multimedia applications, such as teletutoring. A better multimedia workstation architecture is needed to sustain transmission of large amounts of data through the system buses and to minimize data movement and copying.

The UNIX operating system is not adequate to support real-time services. The scheduler granularity of Sun's OS 4.1.3 illustrate this point. Since its clock resolution is not higher than 20 milliseconds [5], it is difficult to implement any efficient audio-video synchronization mechanisms. In addition, the scheduler cannot give a higher priority to real-time data.

The Internet protocol suite cannot guarantee a high throughputs and a bounded delay and delay jitter which are required for teletutoring applications. Thus, network protocols supporting QoS requirements are needed here. Moreover, the current standard network protocols do not support a multicast service which is essential in the multipoint teletutoring configurations. Furthermore, end systems should at least support audio-video synchronization and adaptive rate-based flow control.

7 Conclusion

With the aid of high quality videoconferencing and shared workspace tools, the BETEL multimedia teletutoring application was successfully demonstrated at the end of 1993 over the first 34 Mbits/s Trans-European ATM network. Echo cancellers were essential in ensuring high quality audio in this experiment. This paper showed that it is feasible to build a high quality teletutoring prototype using the technology available in 1993. The performance bottleneck of this prototype was at the end system level, particularly in the video acquisition. The UNIX operating system and BETEL protocol stacks provided best effort services which were not ideal but satisfactory to support the BETEL teletutoring application. UDP/IP were used to transport real-time audio and video data, without any explicit audio-video synchronization and adaptive rate-based flow control mechanisms. More robust and realistic teletutoring scenarii will be realized in a multipoint and multiplatform environment in the framework of the Pan-European ATM pilot experiment starting in July 1994. Examples are distributed classrooms, teleseminars and archiving and retrieval of multimedia documents.

Acknowledgments

Sincere thanks to the BETEL teletutoring project team, in particular, Mireille Goud, Marc Furrer, Xavier Garcia, Bruno Dufresne, Andrea Basso and Marco Mattavelli from EPFL and Morris Goldberg, Antonio Fernando Vaquer Mestre and Revital Marom from Eurecom.

References

[1] Le Moan, Y., "Data Transfer Service Specification", BETEL internal document, CIT-5, April 1993.

[2] Pusztaszeri, Y. H., "Teletutoring over a Trans-European Broadband Network", EPFL-Eurecom Technical Report, June, 1994.

[3] XVideo Technical Overview, Parallax Graphics, Inc., Release 1.0, 1991.

[4] Kunt, M. Techniques modernes de traitement numérique des signaux, Collection Electricité, Presse Polytechniques et Universitaire Romandes, 1991, pp. 175-190.

[5] Jacobson, V., Leres, C., and McCanne, S., TCPDUMP(1), 1989-1993.