Teletutoring over the BETEL Network

Yu-Hong Pusztaszeri[1], Jean-Pierre Hubaux[1], Mireille Goud[2],
Ernst Biersack[3], Philippe Dubois[3]

Abstract

We will describe the multimedia teletutoring prototype jointly developed by EPFL and Eurecom in the context of BETEL, using the state of the art technology available in 1993. With the aide of high quality videoconferencing and shared workspace tools, the actual experiment was successfully carried out at the end of the year over one of the first 34 Mbps Trans-European ATM networks 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 scenarios, together with limitations and future enhancements of this prototype. Focus is placed on the interactive audio and video communications part of the application.

1. Introduction

The trend in today's telecommunication networks is migrating 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 communication.

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. The BETEL project, funded by the

[1] EPFL, DE-TCOM, Eclublens, CH-1015, Lausanne, yhp@tcom.epfl.ch, hubaux@tcom.epfl.ch.

[2] EPFL, SIC, Eclublens, CH-1015, Lausanne, goud@sic.epfl.ch.

[3]Institut Eurecom, 2229, Route des Crêtes, B.P. 193, F-06904 Sophia Antipolis, erbi@eurecom.fr, dubois@eurecom.fr.

European Commission and the Swiss Government, 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 Mbps international ATM networks.

Two innovative applications were designed to satisfy specific user needs and were demonstrated over the BETEL platform (see figure 1). The first application, Teletutoring, involved interactive multimedia communications between students at the Institut Eurecom in Nice, France and a teacher at Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland. The other was a meta-computing application for sharing computer resources between the European Laboratory for Particle Physics (CERN) in Geneva and the National Institute of Nuclear Physics and Particle Physics (IN2P3) in Lyon [1, 2].

Fig. 1. BETEL: Europe's first operational high-speed transmission platform

The goal of this paper is to give an overview of the BETEL teletutoring application. Section 2 describes the BETEL teletutoring platform, scenarios and building blocks, while section 3 is devoted to an interactive audio and video communication tool

developed for this prototype. Finally, section 4 discusses the limitations and future enhancements of the system.

2. An Overview

The BETEL network infrastructure is based on ATM technology, and supports FDDI LAN interconnection, ATM transfer service, and AAL 3/4 data service [3]. The FDDI LANs at EPFL and Euercom were interconnected to the BETEL network by means of Cisco routers. Figure 2 shows the teletutoring network infrastructure.

The network topology at Eurecom is more complex than at the EPFL site. Not only an FDDI ring was dedicated for high speed audio and video transmission, but also an Ethernet was used to connect student workstations to the BETEL platform, and to support connection control and shared workspace data communications at Eurecom. Because the shared sessions generate 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 used an analog audio/video switch. All cameras, microphones, monitors, and loudspeakers were connected to the switch, which was controlled by a dedicated software driver to establish and release audio, video and data connections.
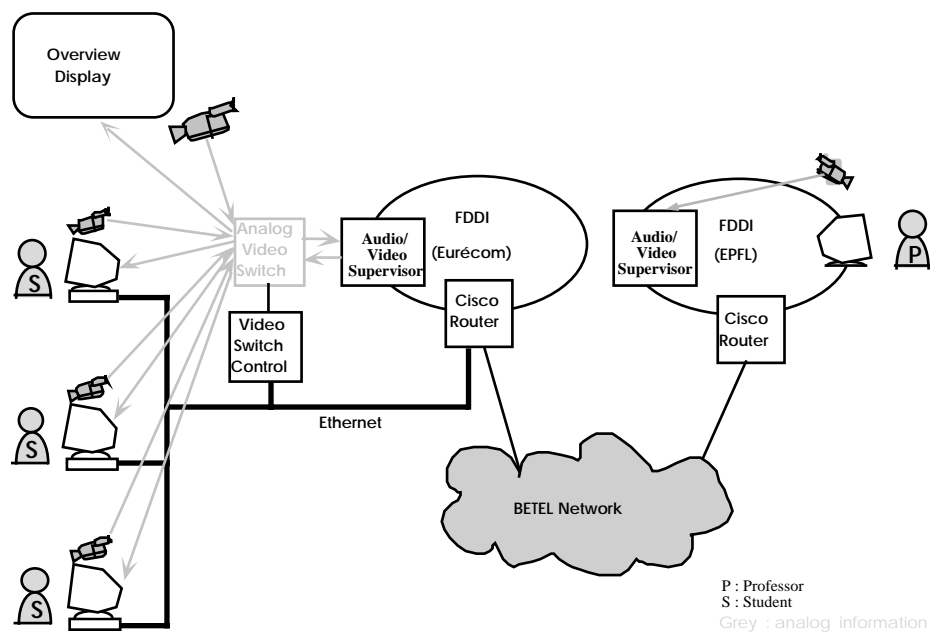

Fig. 2. Teletutoring network infrastructures at EPFL and Eurecom

The teletutoring application uses videoconferencing and shared workspace tools to allow interaction between a teacher and a group of remotely located students. The teacher at one site can teach a class or supervise students in their individual work in

another site while the students can seek assistance from their teacher located several hundred kilometers away. A basic teletutoring scenario is shown in figure 3 and 4.

The overall teletutoring application is based on two main services:
- videoconferencing, which enables the teacher and students to see / talk to each other.
- shared workspace, which allows both student and teacher to collaborate on problem solving.
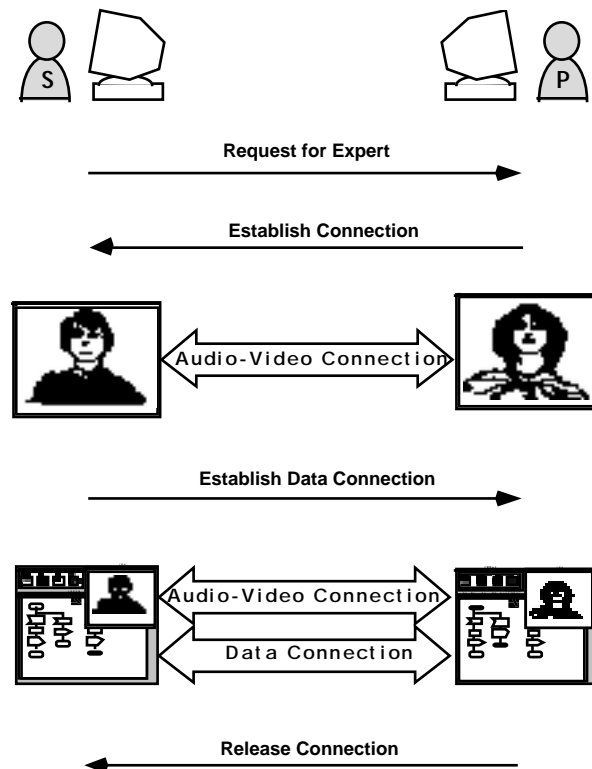

Fig. 3. Teletutoring Scenario

Fig. 4. Professor at EPFL interacting with his student at Eurecom

Owing to the technical constraint of using commercially available products, and limited to a time constraint of one year (January to December 1993), the choice of the platform was confined by the hardware and software available on the market then. The building blocks are summarized as follows:

- Hardware:
    - HP 9000/700 workstations
    - SUN Sparc 10 stations
    - Parallax video acquisition board
    - Echo cancellor

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

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 a workspace which could be shared between the teacher and the students using SharedX, a Shared Workspace Manager provided by HP.

The User Interface[6], shown in figure 5 and 6, was developed on the HP platform to provide an intuitive and user friendly access to the Audio-Video Supervisor (AVS) and Shared Workspace Manager. On the other hand, AVS was implemented on the Sun platform.

Moreover, high audio quality has proven to be essential in the teletutoring application. Because of large latency in the BETEL teletutoring network, audio quality was constantly hampered by echoes. Improving room acoustics was one of the two methods used to minimize the ambient noise and reverberation, and thus the echo. Another was to build echo cancellors based on adaptive filtering techniques [6]. These echo cancellors, located at both sites (EPFL and Eurecom), were used to reduce echoes.
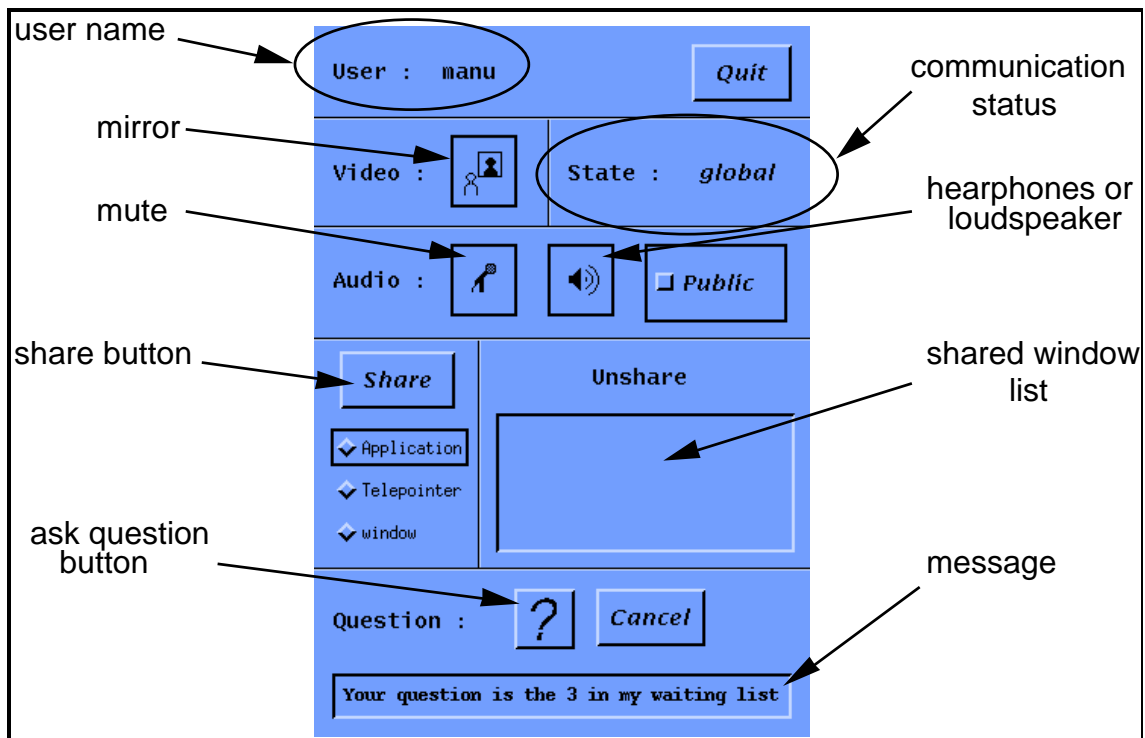
user name

User : manu

Quit

communication status

mirror

Video :

State : global

mute

hearphones or loudspeaker

Audio :

☐ Public

share button

Share

Unshare

shared window list

◆ Application
◆ Telepointer
◆ window

ask question button

Question :

?

Cancel

message

Your question is the 3 in my waiting list

Fig. 5. User interface for students

user

global state button

mirror

classroom or Picture In Picture

view classroom

mute

View : manu

Quit

Video :

ClassRoom

Question

Audio :

☐ Public

manu
philippe
christian

Next   Delete all

Share

Unshare

philippe : Question
christian : Question

◆ Application
◆ Telepointer
◆ window

connected students list

student question list

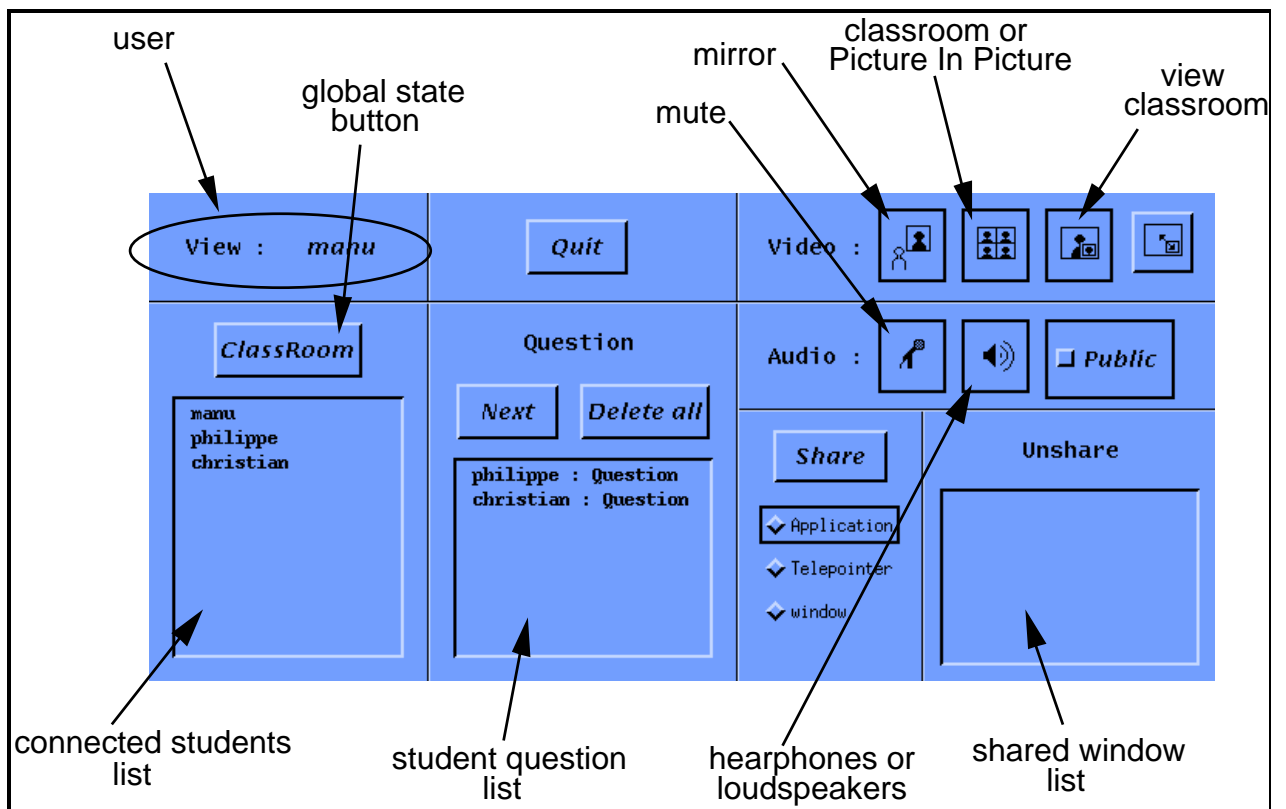hearphones or loudspeakers

shared window list

Fig. 6. User interface for teacher

3. Audio - Video Supervisor

AVS provides real-time audio and video acquisition and end-to-end transmission. Under the supervision of AVS, audio and video signals from analog sources are digitized and encoded, and then are transmitted to a remote station via the BETEL network. At the receiver end, the data are decoded, and video images are reconstructed and displayed while audio is being replayed. Audio and video signals are handled in a similar fashion (see figure 7).
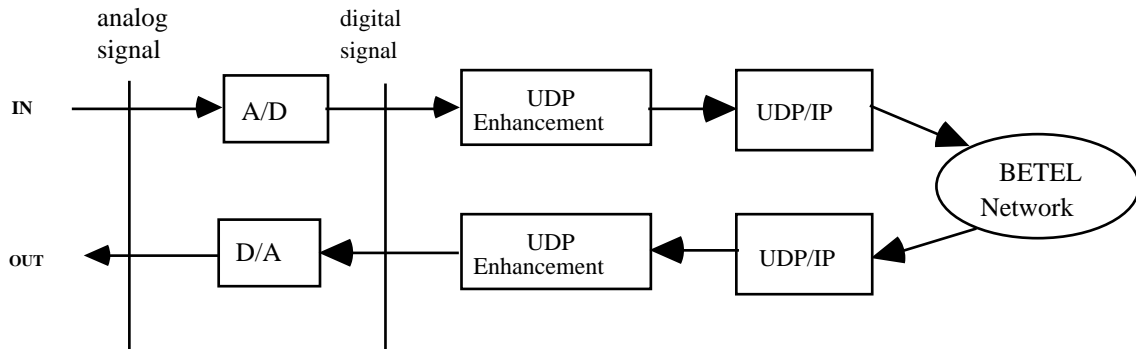

Fig. 7. AVS processing pipeline

3.1 Video acquisition

The current AVS implementation used the Parallax XVideo board. This board was the only hardware available on the market that permitted real-time video compression and decompression at a reasonable frame rate. Details about this board and its performance are given in [4].

The Parallax board can handle analog video input and output in various standards (PAL and NTSC) and formats (VHS, YUV and RGB). 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 frame buffer of the XVideo then compressed / decompressed by the

JPEG Image Compressor of 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 the SpeakerBox of the Sun Sparc10 station. The SpeakerBox audio peripheral provides an integral monaural speaker and microphone, and 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 Network issues for audio/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 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 discard data according to their priority in case of congestion
- synchronisation
    - inter-medium
    - intra-media
- adaptive rate-based flow control

The transport layer protocol is restricted to TCP and UDP since IP was imposed by the Cisco router. 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 networks with large bandwidth-latency products. The BETEL network is one such networks which has a large bandwidth-latency product. The sliding-window flow control with credit allocation does not allow to use the full bandwidth of the BETEL platform. Retransmission in TCP significantly increases end-to-end delays and is unsuitable for interactive audio and video data transport service. Hence, in this context, window based flow control and error control mechanisms found in TCP create problems for real-time audio and video transmission.

On the other hand, UDP does not guarantee in sequence delivery, and does not have any error or flow control; the lack of retransmission in UDP makes it a better candidate to transport real-time audio and video data. However, some endsystem enhancements added to UDP are needed. For instance, video frames in general are larger than 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 ordered frames have to be detected. The minimum UDP enhancements are loss detection and packetization (including segmentation and reassembly for video frames). Therefore, UDP/IP protocols with endsystem enhancements were used to transport audio and video data.

3.4 Implementation Issues

AVS was designed to guarantee best 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 the 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, and so were the segment number and total segments in a given video frame also included for video frames. A typical video packet size is about 4 Kilobytes and the minimum MTU[4] in the BETEL network (excluding the Ethernet segment) is also 4 Kilobytes, whereas audio frames have 128 audio samples to ensure low delay and loss rate.

The audio quality was the most important factor in the design of this teletutoring prototype. Voice is still the most common and effective means of communication,

---

[4]Maximum Transfer Unit

although eye contact and other facial information are also important. Hence, some steps were taken to ensure good 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 audio quality as did the Macintosh microphone, so the latter was used, together with semi-professional microphones (connected to line in port of the SpeakerBox). In addition, echo cancellors 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 cancellor used could not treat audio with sampling frequency higher than 8KHz. Therefore, audio was restricted to telephone quality. Audio data in AVS were not compressed to reduce processing delay and to ensure good audio quality, as the BETEL network bandwidth was no bottleneck in this application.

Another challenge consisted in justifying the 34 Mbps high-speed links for this teletutoring experiment. Why do teletutoring applications and other interactive multimedia application need broadband networks nowadays? The most obvious answer is video quality. High quality digital video signals demand large bandwidth. Since the XVideo board is one of the most performance real-time compression hardware available, understanding the parameters influencing the video quality and the performance of the XVideo board is essential. A performance evaluation study [5] showed that the bottleneck came mainly from the video acquisition board, and to a minor extent from the current implementation which was not able to exploit the inherent parallelism of the different operations involved in capturing and transmitting video.

Also, AVS consists of four independent processes. Each process is dedicated to either receiving or sending video and audio data. Transmission of audio and video is hence independent. In order to overcome this bottleneck, a Sun Sparc10 station was dedicated for either to transmit or to receive video sequences. Therefore, four Sparc10 stations equipped with Parallax boards were involved in the video acquisition and transmission of this prototype.

4. Limitation and future enhancements

The implementation philosophy of BETEL was to integrate currently available technology and build a demonstrator within one year. The teletutoring prototype unearthed several shortcomings in the original design, and 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 build-in synchronisation and rate-control mechanisms implemented in the videoconferencing system. Therefore, enhancements are needed in the following areas : multipoint and multiplatform teletutoring configurations, and system support for interactive real-time teletutoring applications.

4.1 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 has 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 part of 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 same platform. The BETEL teletutoring prototype could then be ported to multiple platforms.

4.3 Audio Quality

Audio quality is most important in teletutoring, and was hampered by echo, which was a serious problem because of large latency audio experienced in the BETEL links. Several echo cancellors 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 is to design new echo cancellation algorithms which support CD quality audio and can be used with mixing devices. This would benefit mostly teletutoring with geographically dispersed students.

4.3 Scalability

The current prototype is limited to a one-to-one interaction. The teacher can interact with one student at a time, although his image and voice can be broadcast to everyone in the classroom. Two or more students cannot engage in a discussion. All video and audio signals have to be transported via point-to-point audio and video connections. It would be useful, if either a 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.

4.4 System support for teletutoring

One of the long term solutions to the performance problem is to implement endsystem 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 busses and to minimize data movement and data copying.

The UNIX operating system is not adequate to support real-time services. The clock resolution and scheduler of the Sun OS 4.1.3 illustrate this point. Since its clock resolution is not higher than 20 milliseconds, it is difficult to implement any efficient audio-video synchronisation mechanisms. In addition, the scheduler cannot give higher priority to real-time media.

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

5 Conclusion

This report shows that it is feasible to build a high quality teletutoring prototype using the technology available in 1993. The bottleneck of this prototype was at the endsystem level, particularly in the video acquisition board. 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, not using any explicit audio-video synchronisation and adaptive rate-based flow control mechanisms, were used to transport real-time audio and video data. More robust and realistic teletutoring scenarios will be realized in a multipoint and multiplatform environment in the framework of European ATM pilot experiment starting in July 1994, for example, a Europe-wide distributed M. Sc. program with distributed lectures, classrooms and campuses.

References

[1]   Martin, O. H., "Broadband Exchange over Trans-European Links (BETEL)", Proc. SMDS Conference, Amsterdam, November 1993.

[2]   Martin, O. H., "A Perspective on the shift/BETEL Project", the 15th Speedup Workshop on "Visualization and Networking", Lugano, Switzerland, on March 17-18, 1994.

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

[4]   XVideo Technical Overview Release 1.0, 1991.

[5]   Gaspoz, J.P., "Performance evaluation of a Video Conferencing system using SPARCstation 10 and Parallax board", BETEL internal document, 1-EPFL-006, September, 1993.

[6]   Pusztaszeri, Y. H. et al, BETEL Teletutoring Internal Report, EPFL-Eurecom internal report, March, 1994.