Technical University of Denmark

DTU

Virtual seminar room-modelling and experimentation in horizontal and vertical integration

Forchhammer, Søren; Fosgerau, Anders; Hansen, Peter Søren Kirk; Hansen, Steffen Duus; Jensen, Ole Riis; Sharp, Robin; Sørensen, John Aasted

DTU Library
Technical Information Center of Denmark

VIRTUAL SEMINAR ROOM

MODELLING AND EXPERIMENTATION IN HORIZONTAL AND VERTICAL INTEGRATION

S. Forchhammer[§], A. Fosgerau[†], P.S.K. Hansen[+],
S.D. Hansen[+], O.R. Jensen[§], R. Sharp[‡], J.Aa. Sørensen[+]
[†]Research Center COM, [‡]Dept. Information Technology,
[+]Dept. Mathematical Modelling, [§]Dept. Telecommunication
Technical University of Denmark

Abstract - The initial design considerations and research goals for an ATM network based virtual seminar room with 5 sites are presented.

INTRODUCTION

The objective of this paper is to discuss the initial specifications and research goals for the design of a Virtual Seminar Room. The project is carried out in a cooperation between 4 Departments at the Technical University of Denmark and the Department of Information and Media Science at Århus University, [6]. The basic observation behind the design of the Virtual Seminar Room is, that besides the constant growth in available bandwidth for transmission in communication networks, many networks either already offer or are developing technologies to give Quality of Service (QoS) guarantees. This means that applications not only will support transmission of coded audio and video, but can be designed in a known and well-controlled network environment, which enables the applications to provide, in a broad sense, a high and reliable quality presented at the Human Computer Interface. This abstraction from technical details and -limitations also means that the system is suitable for research in less technical areas, such as pedagogical aspects of remote education using virtual seminar rooms.

In the following, the concept of modelling and experimentation in horizontal integration represents the classical methodology of algorithm development, comparison, and implementation in a certain area e.g. image or audio compression. The system discussed in this paper is also aimed at modelling and experimentation in vertical integration, which offers the possibility of modelling the effect on selected pedagogical properties, of certain design specifications e.g. related to the fundamental site interconnection network, thus providing opportunities for Quality of Service modelling. Thus, in the framework of this paper, vertical integration represents the cross disciplinary aspects of the project.

The Virtual Seminar Room consists of 5 small rooms (sites) interconnected by a network. The initial system design is partitioned into the following com-

ponents: The Human Computer Interface, The Real Time Stream Control, The Video Stream, The Audio Stream, and The Site Interconnection Network.

THE HUMAN COMPUTER INTERFACE

Initial Specification The Virtual Seminar Room consists of five sites, where the seminar leader and participants have suitable audio, video, and computer devices, by means of which they can communicate with the other participants. They also have access to a common whiteboard, on which they can (one at a time) write and draw, and to which they can point using a virtual pointer. The leader has special privileges to control the communication between sites. He/she can also run demonstrations and display problems for the participants to solve as exercises.

This means that each site must have full video and audio communication, with slide projector and recording equipment. Furthermore there will be computer facilities for handling the whiteboard, as well as possibilities for switching between local group work and full collaboration between all the sites. The computer system shall allow for central session control and monitoring from any of the sites (chosen at session start by the leader).

Research Goals The new technology offers many new possibilities for teaching environments, and it is important to use the technology in such a way that people can concentrate on the communication and the teaching process, not just on mastering all the gadgets. The project aims at developing a framework for user interfaces in distributed, on-line education and collaboration, stressing usability and Quality of Service 'across' the technology. The framework should lead towards a formal description of such interfaces, to be used in future modelling of distributed educational systems, [2].

THE REAL TIME CONTROL

Initial Specification The purpose of the real-time control is to organise and schedule the multiple audio and video streams presented at each site. As in all mixed media systems, the temporal relationships between each individual site's audio and video streams have to be preserved in order to maintain the required Quality of Service, expressed in terms of the quality of sound and pictures and their synchronisation. But due to the multi-user interactive aspects of the system, it is also necessary to preserve the temporal relationships between streams arising from different sites, so as to give the users, as far as possible, the illusion of working in the same room. In addition to the exchange of interactive video/audio, each system also has to deal with the white board, slides, computer animations, films and other demonstration equipment required in the teaching activity in question, and these facilities also place real-time requirements on the system.

The system is not a static one, and an important feature of the stream control is *admission control*, to ensure that the system never takes on more activities than it can deal with satisfactorily. This is closely related to and must be integrated into the concept of *session control* at the user interface level.

Research Goals A principal goal of this research is to develop real-time scheduling techniques which permit scheduling of multiple multi-mode streams from interactive and pre-recorded sources, in order to achieve a Quality of Service, as experienced by the user, which is satisfactory for the desired teaching situation. Most current real-time scheduling techniques concentrate on CPU scheduling, but for this application real-time scheduling of other system resources, such as discs and buses, will need to be taken into account in the implementation of the Virtual Seminar Room, where all parts of the system will be very heavily stressed. Thus this work will build on the body of research performed in recent years on so-called Quality of Service based Operating Systems [11, 1]. An important aspect of this work is to relate the parameters of the scheduling algorithms to the quality of service requirements provided by the user, and expressed in terms of frame rate, pixel density, colour rendering, audio bandwidth and distortion and similar user-oriented measures of quality. Experiments are currently in progress to determine the required relationship.

THE VIDEO STREAM

Initial Specification We aim at high quality video (PAL TV quality), resolution 720x576 pixels, 25 picts/sec. such that easy identification of a person including gesticulations and body language is possible. We propose basing the video coding for compression on the current MPEG or H.26x standards based on DCT techniques [3] with tuned parametrisation or modifications to obtain low en- and decoder delay. To obtain high quality a bandwidth of 4-8 Mbit/sec. is suggested.

Video quality must be scalable, such that video can be stored in a selected resolution and quality, and such that sites with high bandwidth connections can communicate with low bandwidth sites. Layered coding is one way to obtain this feature. If necessary, e.g. in systems which do not feature high level Quality of Service, the decoder may implement strategies to recover from package loss to assure Graceful Degradation.

In a virtual seminar room large screens are desirable, as are functions such as random access to any part of a video stream for (e.g.) zoom and selective storage.

Research Goals To allow interactive video communication in a two (or more) way system, low encoder and decoder delay is important. This feature is not very well supported in the current MPEG standards, except when simply not using B frames, which leads to a trade-off between compression rates,

image quality and delay. Thus work has to be invested in the development of low complexity methods for using temporal information in the video stream, in order to maintain both high compression rates, high image quality, and low delay. E.g. the use of long GOPs (Group of Pictures) will be investigated.

Another important aspect of the work will be to investigate the possibilities of implementing high speed MPEG-4 object based encoding for different elements within the picture, e.g. to compose a virtual room based on video from several sites. Different high-speed methods for improving the quality of the decoded images will also be investigated.

THE AUDIO STREAM

Initial Specification The overall requirement of the audio stream processing [6] is a high sound quality with clear intelligibility and speaker identification obtained through a natural and unrestricted capture. This requirement prohibits the use of personal microphones, so the sound capture is obtained by using a multi-microphone system, positioned near the screens. Due to the far field sound capture, it is necessary for the system to do suppression of reverberations, annoying impulse background noise from papers, cups, etc., and of wideband noise.

Furthermore, it is desireable to use a layered audio coding based on MPEG or G.7xx standards. A voice/sound activity detection can be used to improve transmision and decoding efficiency. Thus sites with no activity only transmit information about the background noise, which is used to generate comfort noise at the receiving sites, i.e. typically only one decoder corresponding to the speaker active site is running. The decoder must implement strategies to recover from packet loss, ensuring some degree of Graceful Degradation.

The virtual seminar room audio is then created by combining audio signals from all units into a stereo signal, such that the sound from each site comes from different spatial directions. This should lead to a comfortable listener identification of the speakers in the virtual seminar room.

Research Goals The aim is to achieve robust multi-microphone sound capture with adaptive directivity constraint. This gives some degree of dereverberation, noise reduction and localization of the sound source(s). This can then be combined with the mandatory echo cancellation, and algorithms for reduction of broadband, and impulsive noise [4], [9]. An important research goal here is to devise a unified representation of these problems.

It is expected that the echo cancellation requires design of a subband algorithm, so integration with MPEG like algorithms may be another possibility. Furthermore the prospectives of combining perceptual properties of noise reduction [10], and MPEG coding must be investigated. Finally it is to be investigated if it is possible to combine the source localization, echo cancellation, noise reduction, and coding into one transform based framework.

THE SITE INTERCONNECTION NETWORK

Initial Specification The network should provide connectivity from any site to all other sites. To avoid a central unit for distributing the signals or duplication of the streams, the network must support multicasting. The network must support the fairly high bit-rates produced by the audio and video encoders, i.e. 1-10 Mbit/s per station. This requirement is fulfilled by using SDH (STM-1) [7] transmission to provide a 155 Mbit/s capacity to each site. Furthermore, the network must also support Quality of Service to ensure that the service delivered from the network has the required quality, e.g. in terms of bit-rate transferred without loss of data.

The basic infrastructure has been designed around an SDH STM-4 ring network. The SDH network will be connected to ATM [8] switches to implement a packet switched network supporting Quality of Service. To access this network, the IP protocol will be used in the terminals to provide a standardised interface to the application.

Research Goals The network must implement and maintain a multicast distribution path for each station. This will require signaling and eventually an algorithm to determine when a reconfiguration is advantageous or required due to bandwidth limitations.

Quality of Service must be implemented on the network and methods for exchanging Quality of Service requirements between the application and the network must be established. Scaleable coding will be used for audio and video to enable graceful degradation and operation over a large range of bit-rate. This means that the various streams must be mapped into different service classes on the network. It would also be advantageous to be able to dynamically allocate more bandwidth to the site that is "active" (speaking loudest).

In the initial setup, ATM is only used to provide QoS transmission. Clearly, it would be interesting to run IP directly over SDH, but this requires IP routers to support QoS guarantees. The RSVP IP protocol extension may be used to reserve capacity on the network. The possibility for establishing a low-end interface without QoS guarantees to the Internet should also be investigated.

To interface the application to the network an RTP-like [12] protocol will be considered to transmit the individual audio and video streams in a non-multiplexed format. Other requirements are minimal protocol overhead and a synchronisation mechanism that supports the multi-site setup.

ACKNOWLEDGMENTS

References

[1] R. Baumgartl, M. Borriss, H. Härtig, Cl.-J. Hamann, M. Hohmuth, L. Reuther, S.Schönberg, and J. Wolter. "Dresden Realtime Operating System," in *SDA '98: Proceedings of the Workshop on System-Design Automation, Dresden*, March 1998.

[2] P. Dewan and R. Choudhary 1992. "A High-level and Flexible Framework for Implementing Multiuser User Interfaces, *ACM Trans. Inf. Syst.* Vol. 10, 4 (Oct.), 345-380, 1992.

[3] Jerry D. Gibson, *et. al.*, "Digital Compression for Multimedia", Morgan Kaufmann Publishers, Inc., 1998.

[4] P. S. K. Hansen, *Signal Subspace Methods for Speech Enhancement*, Ph.D. Thesis, Department of Mathematical Modelling, Technical University of Denmark, 1997.

[5] http://www.dscc.dk/

[6] http://www.it.dtu.dk/˜dtu-rtmm/

[7] ITU-T Recommendation G.707, "Network Node Interface for the Synchronous Digital Hierarchy (SDH)".

[8] ITU-T Recommendations I-series on "B-ISDN", ATM Forum Publications.

[9] S. H. Jensen, P. C. Hansen, S. D. Hansen and J. Aa. Sørensen, "Noise Reduction by Truncated QSVD," *IEEE Trans. Speech and Audio Processing*, vol. 3, no. 6, pp. 439–448, 1995.

[10] K. T. Petersen, S. D. Hansen and J.Aa. Sørensen, "Speech Quality Assessment of Compounded Digital Telecommunication Systems, Perceptual Dimensions," in *ICASSP*, Vol. 2, pp. 1375–1378, 1997.

[11] Ragunathan Rajkumar, Chen Lee, John Lehoczky, and Dan Siewiorek. "A Resource Allocation Model for QoS Management," in *IEEE Real-Time Systems Symposium 18, San Francisco, California*, pp. 298–307. IEEE, December 1997.

[12] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson "RTP: A Transport Protocol for Real-Time Applications", RFC 1889 IETF (The Internet Engineering Task Force) http://www.ietf.org/rfc/rfc1889.txt, January 1996.