# Creating a Demarcation Point Between Content Production and Encoding in a Digital Classroom

Kevin Almeroth, Sami Rollins, Zhongnan Shen
Dept of Computer Science
University of California
Santa Barbara, CA 93106-5110
*{almeroth, srollins, szn}@cs.ucsb.edu*

Bruce Bimber
Dept of Political Science
University of California
Santa Barbara, CA 93106
*bimber@polsci.ucsb.edu*

**Abstract**: Incorporating a significant amount of technology into a classroom is an important, but extremely difficult task. In this paper we describe the next generation of the UCSB digital classroom, called the Collaborative Technologies Lab (CTL). The primary goal of the CTL is to investigate the challenges of deploying technology for technology's sake. We feel that without the ability to deploy a large amount of technology and offer robust functionality, the whole idea of enhanced learning environments becomes marginalized. Therefore, it is critical to understand the challenges of deeply embedding technology in a classroom. Based on experience in building and using the CTL, in this paper we propose a demarcation point between program production and content encoding. This demarcation point offers the advantage of breaking a large technical problem into smaller, easier-to-solve problems. The demarcation point also has the advantage of being able to separate complex functions along common lines of expertise. One skilled in production need not worry about the technology of the encoding systems and one skilled in technology need not worry about the ascetics of program production.

## 1. Introduction

Without a doubt, incorporating technology into classrooms is an important evolutionary step for both learning and for technology[12,13]. Classrooms and learning are key focal points in education. Even while education has an increasing role in our society, the roots of the basic paradigm have remained intact. From single-room school houses to state-of-the-art research universities, the atomic unit of instruction, the classroom, has remained essentially unchanged. Whether it is the basic nature of learning and communication or a more ingrained social behavior, the concept of gathering together and listening for the purpose of learning is a comfortable environment. However, with technology this is all changing.

With the use of technology, learning is no longer confined to a single room offered by a single instructor using a handful of rudimentary presentation aids. Technology has enabled students to participate in real-time from remote sites. Technology has enabled lectures to be recorded, archived, and accessed at any time. And finally, technology has enabled teaching to extend beyond the walls of a single room and pull content and experience from remote locations nearly instantaneously. For all of the benefits and change made possible by technology, there are an equal number of challenges.

From a technocratic perspective, one of the most fundamental challenges in bringing technology into the classroom is managing complexity[5]. As more technology is brought into the classroom, it requires more expertise, more personnel, and more time to successfully provide promised functionality. Therefore, providing a rich set of classroom services necessarily requires a significant infrastructure. As the size of the project grows, project management becomes a key component. And beyond this, the next evolutionary step is to create specific, specialized responsibilities for each person helping to make the infrastructure run. At some point, the infrastructure is finally so complex that no single person can understand everything that is happening. At this point, the complexity takes over, and effectiveness collapses under the weight of complexity.

Therefore, our goal is to study how to manage complexity, how to divide tasks effectively, and how to create the robust use of technology in classrooms. As part of an NSF CISE Infrastructure Grant, the UCSB

Digital Campus incorporates technology into the everyday lives of faculty, staff, and students. The digital classroom, also know as the Collaborative Technologies Laboratory (CTL), is a smaller part of this infrastructure. Because we are primarily technologists, our goals with the CTL have been largely focused around how to effectively use technology. One of the primary goals has been to incorporate as much technology into the classroom as possible. This goal was created not just for the sake of integrating and using technology, but with the broader vision of understanding where our ability to offer advanced classroom services breaks down. What we have found and what we report on in this paper is a discussion of one of the most important problems and our solutions for how to solve it.

As classroom technology increases, one of the hardest challenges is keeping complex functions divided into manageable pieces. One of the primary causes of complexity is that understanding how to produce a professional video program requires a completely different set of expertise than does trying to encode and transmit a real-time, high-quality video stream. We propose to address this problem by first making sure we have a broad range of expertise, e.g. content production, video direction, video compression, etc. Second, where possible we attempt to divide the technological aspects of the infrastructure along logical boundaries. Understanding the root of the problems has been learned only through a large amount of trial-and-error, but we feel we are finally beginning to grasp the problem and understand the potential solutions.

In this paper, we describe an architecture that creates a "demarcation point" between content production and content encoding. "Content production" includes all of the functions necessary to create a high-quality audio/video program. It includes functions like camera control, shot selection, lighting, etc. "Content encoding" includes all of the steps necessary to take an analog program stream and digitize it, encode it, compress it, and transmit it to a set of remote receivers or store it in an archive repository. Our demarcation point is an attempt to specifically separate the tasks of production and encoding. Separating these functions has a number of advantages and almost no disadvantages. Conceptually, dividing these functions along these lines creates logical divisions for breaking expertise. In our experience, we have found very few students who are knowledgeable in both production and editing as well as computing. Therefore, some students can be trained to handle lighting and cameras while other students are trained to run the encoders and transmitters. Even in the case where personnel is limited, the demarcation point makes logical sense and is therefore easier to manage. The one potential disadvantage is the added complexity in creating the demarcation point. However, in our efforts to date, we have been able to solve most problems.

The remainder of this paper is organized as follows. Section 2 presents related work and an overview of the CTL's capabilities. In Section 3 we describe our demarcation architecture. Section 4 offers lessons learned from the design and implementation of this architecture. The paper is concluded in Section 5.

## 2. Background and Related Work

The idea of creating a demarcation point between production and encoding is not specifically new. However, there does not seem to be any significant literature on the topic. Most of the insight we have found is created based on experience in trying to run a technology-based classroom. Therefore, most of the related work can be construed as related *efforts* and does not necessarily appear in press anywhere.

While the idea of a demarcation point is somewhat intuitive, in many cases the focus of related work has been one or more narrowly focused topics. The idea of a demarcation point is based on combining two functions requiring almost completely different skill sets. Therefore, when presenting related work, we must consider both narrowly focused efforts as well as broader, but high level, efforts.

Many of the related projects also have the goal of developing technologically-rich classroom environments. Most also focus on providing some additional functionality. Efforts at Stanford attempt to provide remote participants with most of the same capabilities and experiences as an in-class participant[3]. The Stanford project uses a new encoding mechanism to handle handwritten text, slides and figures in the classroom and records the classroom session as a Synchronized Multimedia Integration Language (SMIL) presentation available for on-demand viewing. The Multimedia Digital Classroom (MDC) system[7] aims to provide important tools to support a distributed interactive education environment. The system consists of a classroom management tool and teaching applications including the Multimedia Internet Browser, Shared White Board, and WWW-Based Note-Taking. The Virtual Classroom[6] provides a portable hardware solution, an easy-to-use software toolset, and easy-to-follow guidelines on how to create an interactive, web-based electronic classroom. The Classroom 2000 project at Georgia Tech[1,2] tries to build a system for the automatic generation, integration, and visualization of media streams in a classroom. It pays particular attention to the variety of media streams that can be captured, the time integration of independent streams, and techniques for providing robust interfaces to visualize and search across multiple media streams. The Intelligent Classroom[4] tries to show that a software system can actually determine what people want to do in the classroom and do it for them. It uses cameras and microphones to sense a user's actions and then use the intentions it infers from those actions to decide what to do to best cooperate with the user. The Smart Remote Classroom[11] deploys a hybrid application-layer multicast protocol and an adaptive content delivery scheme to allow large numbers of users to access virtual classroom content.

The Collaborative Technologies Lab (CTL) was first conceptualized in late 1997. At the time, researchers were beginning to incorporate technology, in particular multimedia capture and synchronization, into classrooms. Unfortunately, it took several years to find adequate space and sufficient funding to undertake the effort in earnest. Since then our goal has been to develop an infrastructure that supports a phased deployment of functionality[9,10]. The phases are:

1. Presentation Facilities
2. Webcasting Studio
3. Remote Collaboration
4. Lecture Replay

Currently the CTL is capable of supporting most of the functionality it was originally intended to provide. We have three projectors that can project from essentially any source: presenter laptop, other presenter devices, video from remote sites, etc. The front of the CTL is shown in Figure 1(a). There are also three video cameras: one to capture anything on the whiteboard or screens, one to capture the presenter, and one to capture the audience. In terms of audio capability, we have a wireless microphone for the presenter and table-top microphones for the audience. We also have mounted speakers to project any audio source. We have a sophisticated lighting system (described below and shown in Figure 1(a)), and the ability to capture content written on the whiteboard and synchronize it with an audio stream. In terms of encoding capability, we can create audio/video streams of any rate in Real Media format and/or Windows Media format. The classroom is also set up to send audio and video to the Access Grid (http://www.accessgrid.org/) in a variety of formats and at a variety of rates. We also have a real-time system to encode and store an MPEG audio/video stream to disk. Our "control room", which includes all elements of room control and encoding, is shown in Figure 1(b). In practice, not all of this technology works properly all of the time. Some of it is relatively robust and some we are continuing to work on and improve.


(a) (b)

Figure 1: The Collaborative Technologies Lab: front view, and "control room.

As a result of our attempts to deploy as much functionality as possible in the CTL, we have created a management nightmare for ourselves. Our goal now is to develop mechanisms to abstract the functionality along logical divisions. Our belief is that by doing this we can lower the training barrier for new personnel and make the entire process more efficient. Therefore, we have created the concept of a demarcation point to separate production from encoding. This architecture is described in the remainder of this paper.

## 3. Demarcation Architecture

The motivation for creating an architecture with a demarcation point between content production and encoding is motivated by several factors. First, the functionality of the system should be divided such that the complexity of an immense task is divided into smaller, manageable pieces. Second, the dividing line should be created between functions such that each piece falls into an area of common expertise. In other words, since it is unlikely to find someone skilled both in production and encoding, we separate the functions into two tasks. And finally, by dividing the task, we bring awareness to the problem.

As a result, we have defined an architecture for technologically enhanced learning that logically divides the problem into two major tasks. Our architecture is shown in Figure 2. The most interesting aspects of this figure are the following: (1) what falls into the "production" category, (2) what falls into the "encoding" category, and (3) what falls into the "control" category along the divide. Each of these three categories are described in the following subsections.

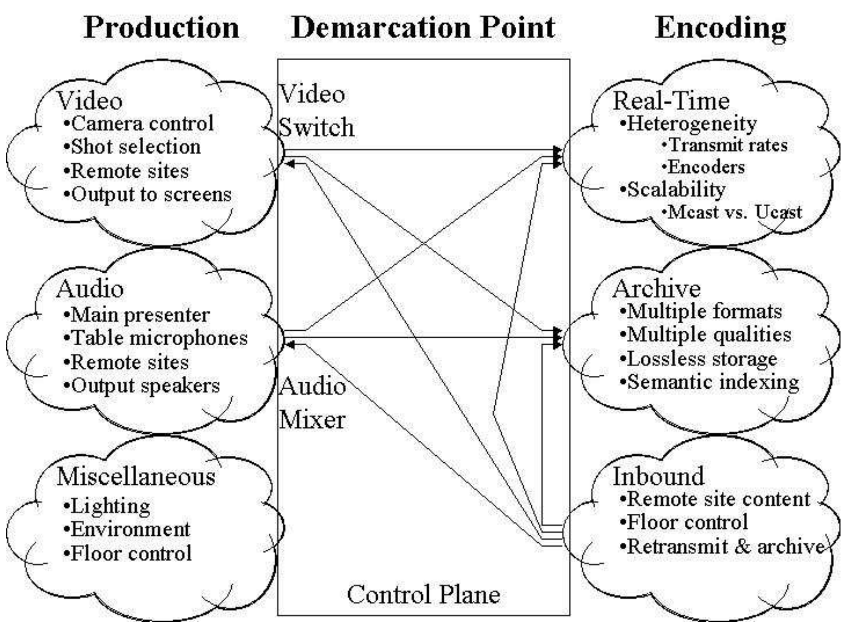


Figure 2: The demarcation architecture.

### 3.1 Production

Production includes all of those functions associated with creating a high-quality audio and video program. Essentially what has to happen is that the classroom needs to be turned into a television studio. There needs to be correct camera shots from appropriate camera angles, high-quality audio, and miscellaneous functions like correct lighting and floor control. Next, we describe each item in more detail.

Of primary importance is audio and video production. Video production requires all of the pieces necessary to create a visually appealing presentation. Cameras need to be mounted in the right location. Feeds from the different cameras need to be selected such that the right camera is used at the right time. The guiding principle is that the "right" angle is the one that captures the primary activity directly. Therefore, cameras must first all be controlled and directed. Second, the right camera selection must be made. Of course, the problem becomes even harder if multiple video streams are simultaneously selected and transmitted. Many of these concerns are not unlike traditional television broadcasts. Given that most such events use large crews to accomplish the function of video production, expecting one person (or even zero people in a truly automated environment) to accomplish the same task at the same level of quality is a bit extreme. In the CTL we use cabling and remote control software to move each of the cameras. We then use a rudimentary production board and a television to produce a single video stream. We have not yet developed a solution to the problem of producing multiple video streams in parallel for truly interactive applications. In truth, good production is more art than science and requires a great deal of skill to do well.

Audio production, like video production, requires talent to do well. The challenge with audio is finding the right balance between generated audio signals and amplification of those signals. The major problem to overcome is feedback loops that form between output devices and input devices. For example, if a presenter's voice is broadcast over speakers, then the table-top microphones must perform echo cancellation. A harder problem is the situation when a remote site transmits audio that is broadcast over the speakers. Because of additional delay introduced by Internet transmission, echo cancellation becomes almost impossible. And yet the hardest problem of all is receiving audio from a remote site, broadcasting it in the main classroom while at the same time relaying the transmission to all remote sites *except* the one that originated the broadcast. All of this functionality requires a level of technological sophistication and audio engineering expertise that we do not yet have. Therefore, we must focus our efforts on making sure that the things we can do, we do well.

Aside from audio and video, there are a number of miscellaneous functions that go into production. These may seem mundane but they contribute a significant amount to overall quality. Through our experience in the CTL, we have come to realize that lighting is one of the most critical functions. For a group whose focus was on technological functionality, having to deal with lighting was less than interesting. The goal was to be able to create a video stream that neither washed out the projection screen or whiteboard nor put the presenter in shadows. Given the further complication that one screen was always a whiteboard, two could either be screens or whiteboards, and the fourth was always a screen (see Figure 1(a)), we eventually needed to design a lighting system with six "zones". The first three zones are for the three flat screens/whiteboards along the front. If the projector is being used, the lighting in that zone is turned off, otherwise the lights are turned on to illuminate the whiteboard. The other three zones are to light the presenter as (s)he stands in front of each of the screens. Further complicating the setup was the fact that the first three zones differed from the second three in terms of the kind of lighting required. The screens/whiteboards required a softer, more diffuse light while the second three zones required brighter, more direct lighting. All of these were lessons we learned in the process of installing the lighting. Finally, having six different zones creates yet another system to properly configure before putting on an event[5].

The other miscellaneous issue that is a significant challenge in the CTL is floor control[8]. "Floor control" primarily concerns determining how remote sites participate in the main site. If there is a question from a remote participant, how should it be handled? Using video on the side screen (see Figure 1(a)), a presenter might "see" a hand go up signaling a question, but often the situation is more complicated than this; for example, if there are many remote sites or if there are presenters at remote sites, etc. Floor control is still very much an open issue in collaborative research and we feel it is not particularly well addressed in our classroom. As for its location in the demarcation architecture, it really falls into both production and encoding. As a result, it also becomes one of the few functions that crosses the demarcation point. The challenge from a production point-of-view is to incorporate the audio and video of remote sites seamlessly into the classroom. Given that remote sites can either be active participants, passive listeners, or change between these two, possibly frequently, the challenge is to make remote participants feel like they are in the room and keep local participants from being distracted from the presence of remote sites.

We believe there is significant advantage to be had if all of the functions of production can be accomplished before there is any thought given to what should be done with the content. By focusing only on a single manageable task, a better-quality program stream can be created to then feed to the encoding system(s). We now discuss the challenges of this next function.

### 3.2 Encoding

The encoding system we have in place is truly flexible. However, it has become so cumbersome that it is very difficult to make it work as needed. There are two basic problems. First, there are too many different systems and standards available for encoding *streamed* versus *archived* content. "Streamed" content is typically used for real-time, interactive sessions, and "archived" content is typically used for non-realtime, after-the-fact viewing. Not only are there proprietary formats (Real, Windows, Quicktime, etc.) in addition to open standards (H.261, H.263, motion JPEG, MPEG-1, MPEG-2, MPEG-4, MPEG-7, etc.), but there is bandwidth heterogeneity; there is the need to stream and to archive (typically using different formats); and then there are different modes of interaction/tools (Access Grid, videoconferencing systems like Polycom, and desktop streaming). This large set of parameters creates an immense set of possible stream formats and characteristics. In today's world, if a room is not versatile enough to send in a variety of formats, it is likely not to be useable to a majority of groups who want to use it.

The second problem, related to the first and a good motivation for our demarcation architecture, is that there is a disconnect between the types and number of audio/video streams provided by the production side of the system and the needs of the various encoding systems. For example, on a given day for a given event, if we want to transmit using Real Media and also transmit to the Access Grid, the problem arises that Real wants one audio/video stream but the Access Grid wants one audio stream and three separate video streams. Trying to manage the production side in such a way to produce a single stream *and* three separate streams (with one designated primary and the others as secondary) is nearly impossible.

Aside from dealing with the various format issues, there is also a significant challenge in how to deal with incoming audio and video from remote sites. The first requirement is that they be passed to the production side of the architecture for possible display on one of the screens. For example, video might need to be displayed and audio played if a remote site asks a question. Additional media might need to be projected if the presenter is actually at a remote site. And of course, there are floor control issues that bridge the demarcation point. The second requirement is that audio and video from remote sites might need to be incorporated into the archive stream or re-transmitted to other remote sites. A straightforward, but not completely effective solution we use in the CTL is to treat remote sites like just another audio and video input. This works well for many of the functions but we still have issues to solve with echo cancellation for the audio and integration of encoded remote video and unencoded local video.

Finally, there are number of scalability issues and possibilities for value-added services. Scalability is critical when there are many remote sites watching. One-to-many communication, or multicast, is a widely recognized solution for delivering one stream to many receivers simultaneously. By using multicast, load on the encoding system to deliver multiple streams is reduced to delivery of only a single stream. The only drawback of multicast is a lack of widespread deployment. In the encoding process there is also the possibility of adding support for value-added services like inter-stream synchronization, indexing, etc. The ability to add these services is a matter of having the technology and being able to make it work robustly. There are many possibilities for value-added services and numerous opportunities for additional work.

### 3.3 Control

The demarcation point is essentially what we call the "control plane". It exists at the interface between the production and encoding components. Programming created by the production side is provided as a source to the encoding systems. Through this interface also flows the content from remote sites. This content is decoded and sent back to the production side for display and/or sent back to the encoding systems for retransmission to remote sites or archived for later use.

The two functions that are the responsibility of the control plane are: (1) deciding what content should be displayed in the classroom and what content should be sent to which encoding systems, and (2) deciding what audio inputs (presenter microphone, table-top microphones, remote site audio, etc.) should be sent to what outputs (speakers, various encoding systems, etc.). For video, a hierarchy of video switches is used. For audio, a mixing board is used. Unfortunately, these two sets of equipment are not sufficient to support the variety of requirements demanded of a fully functional demarcation architecture. These needs and our attempts at implementing such an architecture are described in the next section.

## 4. Recommendations

In the last section, we proposed a demarcation architecture and described each of the components. In this section we describe efforts we have made and efforts we plan to make in order to make the CTL better adhere to this architecture. To start, we have already made a number of improvements to the production, encoding, and control components. We have also done a number of things to help separate the functions of production and encoding. Together, this first set of efforts include:

1. **Use separate machines for each encoding system**. By using different machines for each type of encoding, we make the encoding system much simpler. Figure 1(b) shows the large number of machines and monitors in use. Unfortunately, this solution can be limiting if a classroom architecture does not have access to space or equipment. While this solution solves some problems, it does not solve others. Each encoding system typically requires expertise in some specific software, and using the large collection of hardware/software can be complex.
2. **Minimize the number of encoding systems**. By eliminating infrequently used encoders and using software to duplicate and re-encode a high bit rate stream into a low bit rate stream, we can reduce the total number of encoding systems. The limitation here is that there is a one-for-one tradeoff in complexity versus flexibility. We are in the process of developing mechanisms to reduce this tradeoff.
3. **Use separate machines for each control system**. In order to reduce the personnel requirement, much of the functionality in the CTL is done via remote control. We have also attempted to centralize this control into as few machines as possible. The primary functions are remote control of our three cameras, shot selection, audio mixing, and video switching. Through experience we have learned that control from one or a couple of machines is not the ideal setup. The problem is that the control system provides visual cues. Switching between multiple control systems is cumbersome and inefficient. For example, at one point we had the audio mixer control software running on the same machine that controlled the cameras. Every time an adjustment was needed on the mixer, the camera control software would have to be minimized and the mixing software raised. This constant switching made monitoring difficult. As a result, our “control room” (see Figure 1(b)) has one machine for each control system.

While we have solved a number of problems, there are still things to be improved. Therefore, we propose a set of additional efforts we plan to undertake. These include:

1. **Better utilize function-specific technologies**. Instead of using PCs for encoding, there are now hardware specific solutions that are essentially plug-and-play. These solutions require very little configuration and hence are more reliable than Windows-based or even Unix-based PCs. Two companies offering this kind of solution are Vbrick (http://www.vbrick.com/) and NCast (http://www.ncast.com/).
2. **Create a set of defined stream types**. We would like to create a more manageable number of defined stream types. For example, “fixed shot of presenter video and his/her audio”, or, “wide audience shot with zoomed focus on question askers, if any”. By creating a formal interface between program production and content encoding, we can reduce to a manageable set the number of audio/video streams. By defining streams in terms of what they capture, and limiting the number of choices, we can simplify production requirements; reduce the number of choices given to the encoding system; and simplify the interface between the two components.
3. **Pre-sets for control systems**. The control systems themselves are quite complex. Even though they provide significant control, it comes at a cost. One solution is to create pre-sets so that a person who wants to use the room in a certain way can simply make sure the right pre-sets are set.
4. **Establish a labeled patch board**. A significant amount of complexity is created because we have wires running every direction. Even though we have tried to implement good wiring behavior, we often find ourselves moving wires around. This creates confusion because the set-up changes. And since we never plan on moving things around, no wires are properly labeled. Often times we

find ourselves spending a significant amount of time debugging problems caused by short-sited rewiring. By creating a patch board of programming and documenting what each wire is we can easily adapt to different (and new) encoding systems. This idea also meshes well with our plan to create a set of defined stream types.

5. **More and better trained personnel**. At some point, suggested improvements become unrealistic. Given the problems we have in getting funding for personnel, this improvement is unlikely to happen, but it is certainly worth mentioning. It would certainly solve many problems!

## 5. Conclusions

In this paper we have described the motivation for a demarcation architecture. The need is rooted in the challenge of a digital classroom becoming too complex to be manageable or always capable of delivering its promised functionality. The ability to operate a digital classroom, to manage its technology, and to limit complexity are all well known problems. The focus of this paper is to design an architecture that will help divide the technology into more manageable pieces. By using a demarcation point to separate the functions of program production and content encoding, we hope to create more stability and better flexibility. We have begun to implement some of the steps necessary in our own classroom, the Collaborative Technologies Lab (CTL).

On a grander scale, we feel that while there has been a tremendous amount of effort and attention given to technology in classrooms, there has been minimal forward progress towards the long-term goal of having a significant impact on teaching and education. We feel that one of the primary reasons for this is that the curve for deploying and using technology is too steep. Hence there is a need for more and better mechanisms to compartmentalize complexity. We believe the demarcation architecture is one step in the right direction.

## References


[1] J. Brotherton, J. Bhalodia, and G. Abowd, “Automated Capture, Integration, and Visualization of Multiple Media Streams”. *IEEE International Conference on Multimedia Computing and Systems*, (San Antonio, Texas, USA), pp. 54-63, July 1998.

[2] G. Abowd, "Classroom 2000: An Experiment with the Instrumentation of a Living Educational Environment", IBM Systems Journal, vol. 38, num. 4, pp 508-530, 1999.

[3] S. Deshpande and J.N. Hwang, “A Real Time Interactive Virtual Classroom Multimedia Distance Learning System”, IEEE Transactions on Multimedia, vol. 3, num. 4, pp. 432-444, December 2001.

[4] D. Franklin, "Cooperating with People: The Intelligent Classroom", *National Conference on Artificial Intelligence*, (Madison, Wisconsin, USA), July 1998.

[5] J. Humfrey, S. Rollins, K. Almeroth, and B. Bimber, “Managing Complexity in a Networked Learning Environment”, UCSB Technical Report, December 2002 (submitted for publication).

[6] A. Krukowski and I. Kale, “Virtual Classroom”, *IEEE International Conference on Advanced Learning Technologies*, (Madison, Wisconsin, USA), pp. 279-282, August 2001.

[7] K.-C. Lee, et. al., “Design and Implementation of Important Applications in a Java-Based Multimedia Digital Classroom”, IEEE Transactions on Consumer Electronics, vol. 43, num. 3, pp. 264—270, August 1997.

[8] R. Malpani and L. Rowe, "Floor Control for Large-Scale MBone Seminars", *ACM Multimedia 97*, (Seattle, Washington, USA), November 1997.

[9] S. Rollins and K. Almeroth, "Deploying and Infrastructure for Technologically Enhanced Learning", *World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED MEDIA)*, Denver, Colorado, USA, June 2002.

[10] S. Rollins and K. Almeroth, "Lessons Learned Deploying a Digital Classroom", UCSB Technical Report, December 2002

[11] Y. Shi, W. Xie, G. Xu, “Smart Remote Classroom: Creating a Revolutionary Real-Time Interactive Distance Learning System”, *International Conference on Web-Based Learning*, (Hong Kong, China), August 2002.

[12] B. Shneiderman, M. Alavi, K. Norman, and E. Borkowski, "Windows of Opportunity in the Electronic Classroom", Communications of the ACM, vol. 38, num. 11, pp. 19--24, November 1995.

[13] D. Tsichritzis, "Reengineering the University", Communications of the ACM, vol. 42, num. 6, pp. 93--100, June 1999.