

The AutoAuditorium* System – 10 Years of Televising Presentations Without a Crew

Michael H. Bianchi
Foveal Systems LLC
190 Loantaka Way
Madison, NJ 07940
MBianchi@Foveal.com

September 2009

Abstract

Making video telecasts and recordings of auditorium and classroom presentations has become much more common over the past 10 years, especially with the rapid improvements in video delivery via the Internet. Most of those videos are created “by hand” using at least one person to operate the camera (or cameras) and video editing equipment.

10 years ago the AutoAuditorium™ System was introduced as a fully automatic alternative to manually created video of such presentations. It combines a target-free, initialization-free Tracking Camera subsystem with a Slide Camera, additional cameras, and an automatic Director subsystem that runs a video mixer. Together they produce a television program of a presentation in real time.

This paper describes the system as high-level concepts and their implementation in hardware. It then recounts how the system has been used by several customers.

1. The AutoAuditorium System

The AutoAuditorium™ System was developed during the 1990s as a fully automatic, multi-camera system designed to televise auditorium and classroom presentations without a crew. The project was originally motivated by observing that most presentations, both formal and informal, had a very simple structure that could be readily captured by one person operating two or more remote control cameras and simple video editing equipment. In fact, the predecessor system was just that. Several were built and used. One, now over 15 years old, is still in use.

* ™ AutoAuditorium is a trademark of Telcordia Technologies, Inc. used under license

One of the major impediments to using manually controlled video production systems is the requirement of a human operator. Many talks are not captured because the required person is not available at the appointed hour.

Hence the manual system design was, by degrees, automated until the person was only needed to turn the system on and off. The evolution of that manual system into the AutoAuditorium System is recounted in [1]

1.1. The System In Brief

The AutoAuditorium System is an automatic video production system that can serve as the front-end to many technologies used to capture and distribute videos.

The Tracking Camera subsystem follows the presenter giving a talk and the Director subsystem produces a fully edited television program of the talk.

The program includes pictures that follow the presenter moving around the stage, pictures of the projected visual aids, picture-in-picture combinations of the presenter and the projections, and, occasionally, other views of the room. The resulting program is produced in real-time, so it can be viewed live by local and distant audiences as well as later by those who watch the recordings.

Either of the two subsystems can operate independently. When used together they cooperate.

A typical system consists of the components shown in Figure 1.

- *The Tracking Camera* subsystem automatically follows the presenter on stage as he or she gives the talk. It does not follow a target on the presenter. The presenter does not wear or carry anything. Nor is the person identified to the system. Any person walking “on stage” is the object of interest to the Tracking Camera.

A stationary Spotting Camera collects images of the entire stage area. See Figure 2. These images are ana-

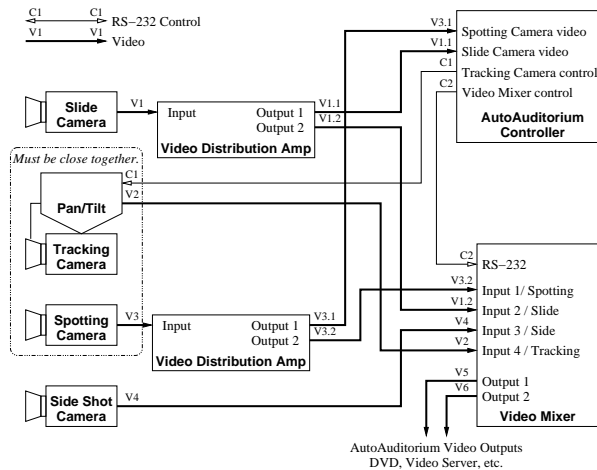


Figure 1. Schematic diagram of a typical AutoAuditorium System



Figure 2. The Spotting Camera image.

lyzed for the motion of the presenter on the stage. That analysis is then used to pan, tilt and zoom the robotic Tracking Camera.

The ability to zoom the Tracking Camera image allows it to accommodate how much the person is walking, gesturing, or even if there is more than one person on stage. See Figure 3.

- *The Slide image* is an image of the projected visuals. The source of that image may be either a camera or a scan converter. A “scan converter” is a device used to change the video signal that comes out of a computer, which is not electrically a *television* video signal, and turns it into one that is. The signal’s transformation includes changes in the number of scan lines, the number of pixels per scan line, the order in which the scan lines

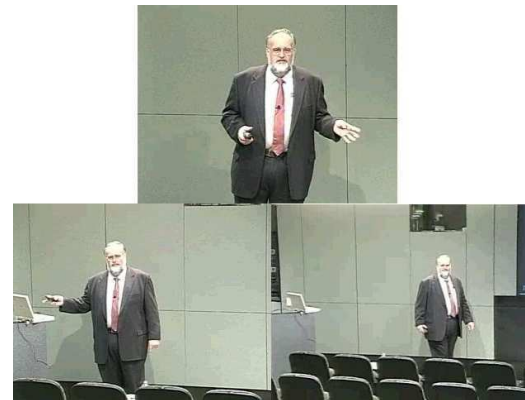


Figure 3. The Tracking Camera zooms to accommodate the person’s gestures and walking.

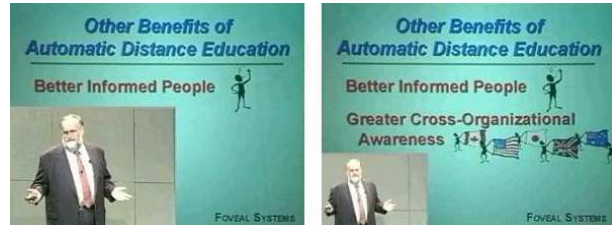


Figure 4. The Director adjusts the size of a picture-in-picture to the available space when it can.

are present in the signal and the frame rate. The resulting television signal is the same as would be delivered by a camera.

- *The Spotting Camera image of the entire stage area* is used both by the Tracking Camera subsystem to find the moving person and by the Director subsystem as a transition shot. See Figure 2.
- *Other camera images of the stage and audience* from different points-of-view are used as transition shots to keep the video editing from being too simple and thus boring.
- *The Director* subsystem runs the video mixer, thus creating the edited program from the various video sources. It detects changes on the projection screen and makes sure the viewers see new information quickly. It also constructs picture-in-picture shots of the instructor and the screen that are sensitive to the amount of blank space on the screen. See Figure 4.

The Director also uses transition shots when the program “slows down”, e.g. when the projected slide is still shown but has not changed in a long time.

The basic output is a fully edited television program as a baseband television signal; audio and video. It can be transmitted live by means of cable, the Internet, satellite dish, or other medium and it can be recorded to video tape, DVD, video server, or other format.

Originally, recordings on analog video tape were the most common means of distribution. Lately, the most common transmission media include recordings stored on video servers, live distribution within and between buildings, and recorded DVDs. The DVDs are rarely distributed, but are kept as a “safety” recording against a video server problem, and are used as the source material for publicly available recordings, such as those found on video collection web-sites: YouTube.com, VideoLectures.net, etc.

To be clear, the AutoAuditorium System is used as a *signal source* to many forms of video capture and video distribution systems. It, alone, does not capture nor distribute the programs.

The AutoAuditorium System works best with presentations primarily given by one person at a time, on stage, talking to a local audience, using projected visual aids. It can be installed both in large, formal rooms and in smaller, informal, more intimate rooms.

Because the system is built into the room, along with the sound and projection systems, and because the system is fully automatic and runs without a crew, it’s primary attribute is that it removes the need for a human operator. Once the system is turned on there is nothing for anyone to do until it is time to turn the system off.

One of the insights that motivated creating the AutoAuditorium System was that the costs of video equipment had decreased to the point that all the equipment needed to make videos could be economically installed permanently into the room.

In fact, permanent installation is a crucial technical requirement of the Tracking Camera. It also makes the start-up and shut-down time of a telecast a matter of minutes instead of the hours (or sometimes days) needed to install and remove portable equipment.

If the AutoAuditorium control is on the lectern televising a lecture is just a matter of pressing the **AutoAuditorium System Start** button and putting a DVD in a recorder or starting a video server recording.

1.2. Camera Image vs. Scan Converter Image

Shooting the Slide image from the projection screen with a camera has a trade-off when compared with using a scan converter.



Figure 5. A Slide Camera sees gestures in front of the screen that a scan converter misses.

The Slide Camera can see everything on the screen, just as the local audience does, but the scan converter can only see the signal sent to the projector. A presenter standing in front of the screen, pointing at something and saying, “This is the most important thing to remember!” is better understood in the video from the Slide Camera than with that from the scan converter. See Figure 5.

Moreover, the camera can see anything projected on the screen, no matter how it got there. We still occasionally hear the story of someone arriving with old presentation graphics on 35mm slides, or even overhead projector foils. When that happens, the AutoAuditorium Slide Camera still can “see” the images and the Director knows exactly what to do.

1.3. Television Programs vs. Two-Stream Side-by-Side Programs

Since about 2000, there has been a growing trend to make recordings of presentations and lectures that can only be seen properly on computers. The programs contain two independent image streams. One is of images from the projection screen captured either as video or as a sequence of still images, and the other is of the presenter giving the talk. There is also an audio stream, most often integrated with the presenter stream. The playback application that runs on the computer ensures that all three are synchronized.

Typically a person watching a two-stream recording is shown a small image of the presenter next to a larger image of the screen. See Figures 6 and 7.

Several AutoAuditorium customers have moved from the original “single stream television program” to this “two stream side-by-side program”. They use video from the

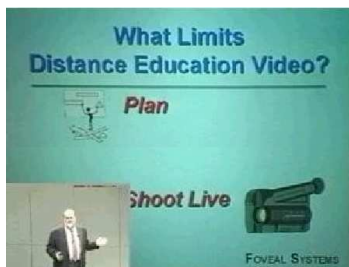


Figure 6. Single Stream Television Program

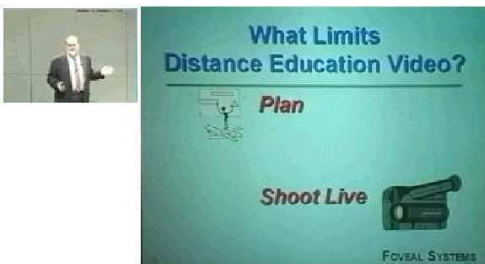


Figure 7. Two Stream Side-By-Side Program

Tracking Camera as the presenter stream and the projector's input signal as the slide stream.

There is a trade-off.

Many people watching the program prefer the side-by-side model because they can always see the current slide and the presenter. They do not have to settle for the AutoAuditorium Director's editing decisions.

On the other hand, the side-by-side images are utterly static, which can be tiring long before most talks are finished. To many people AutoAuditorium's editing, picture-in-picture shots and occasional transition effects between shots makes the program more watchable.

2. The First Installation Is Still Running

The following paragraphs describe the first commercial AutoAuditorium installation in some detail. We do this both to brag a bit that the original installation has proved to be reasonably robust and to illustrate that the original architecture supported changes in usage over the past decade.

The first commercially installed AutoAuditorium System is still in use at IBM Watson Research in Hawthorne, New York. Operational since January 2000, it consists of a personal computer with two video framegrabbers running the AutoAuditorium applications. The cameras are Panasonic AW-E600 3-chip cameras with Fujinon robotic lenses on Telemetrics analog pan/tilt heads. The video mixer run by the AutoAuditorium Director is a Panasonic WV-MX50.

Over the past 10 years there have been a number of repairs:

- One camera burned out and was replaced.
- A couple of power supplies became electrically noisy and were replaced.
- Microswitches in the pan/tilt heads and switches in the robotic lenses became intermittent and were replaced.
- The Spotting Camera (which is used to determine where the Tracking Camera should point) was upgraded from a single-chip to a 3-chip with very wide contrast range.

The software in this system has not changed much. Although improved versions of the AutoAuditorium applications have been installed, the operating system and computer remain the same, Red Hat 6.1 running a 2.2 Linux kernel on an IBM Pentium PC. Most of the software changes over the years have involved adjustments to various parameter settings read by the applications when they start.

The installation allows for manually produced video recordings using the same cameras and pan/tilt heads normally used by AutoAuditorium. The cameras are controlled by joysticks and the video mixer is run by hand.

The first installation was part of a newly constructed lecture hall that *has* changed over the decade. Most notably there have been three generations of projectors that got brighter and higher resolution with each upgrade. Originally the room was dimmed somewhat to make the projected images easier to read. With the latest projector, the image is very bright and the room lighting is not dimmed.

At their nearby Yorktown Heights, New York, lab, two other AutoAuditorium Systems serve three rooms. One system has a set of cameras in each of two rooms. A switch selects which camera set is connected to the system and which room profile is used by the software.

2.1. Common Usage

In 2008, those three AutoAuditorium Systems telecast 234 presentations, as both a single-stream television program and a two-stream side-by-side program.

A DVD of the television program is recorded as a "safety" copy.

When an event exceeds the seating capacity of the room the live television program is also sent to other meeting rooms and the cafeteria. Frequently the same program is sent to meeting rooms in other buildings, up to 14 miles away, and sometimes over video codec connections to other IBM locations worldwide.

The "side-by-side" program is made available as a live Internet stream and recorded to an in-house server for on-demand webcast access.

Eleven manual productions were made using the AutoAuditorium equipment in 2008. Manual operation is primarily used for programs that the current AutoAuditorium System does not handle well, such as panel discussions where there are several people sitting in a row on stage. They are also used when the final program requires that “polish” that only human operators can provide. Those programs may also go through a post-production editing process.

History shows that AutoAuditorium programs are seldom edited.

2.2. Uncommon Usage

IBM Research has another AutoAuditorium installation in San Jose, California. There they use the “side-by-side” program in an interesting way. Although the live program is available on the in-house network while the presentation is being given, they routinely also send it to a conference room where people watch and simultaneously work on other things, talk with each other, or make phone calls. If they did these things in the auditorium it would be considered anti-social, but in that conference room ...

3. The Cyber Classroom at UMich/Flint

The University of Michigan at Flint is largely a commuter school where many students have full-time jobs, so life impinges on them more than on the resident students.

The AutoAuditorium System is a key component of their “cyber classroom”. The vast majority of the classes taught in that room are recorded onto a video server. Some are in normal television format and others are in side-by-side format, at the instructors’ discretion. They report making 720 recordings in 18 months. “Four days a week we turn it on at 9 AM and turn it off at 9 PM.”

A paper by Turner and Farmer on the cyber classroom reports statistical evidence that having lecture recordings available to all students can improve their performance. [4]

3.1. Class, or TV, or Both, or Neither

Unlike many institutions, the students attending the courses in the cyber classroom do not have to declare if they will be physically present in class, or watching the classes on the Internet, both, or neither. Each can do what ever works for him or her, and what works can be changed without “permission.” A couple of recent stories illustrate the point. [5]

One student started out just watching the lectures online. After a week or two he decided that was not working for him. He changed to attending every class so he could interact with the instructor and the other students.

In another case a student had a family emergency where he felt he should visit the hospital as often as possible. The distances involved were such that he could not continue attending class. But he found that he could visit the hospital and still watch the recordings. He got an A in the course.

3.2. Study, and Review, and Cram

Interviews with several students who have taken the cyber classroom courses show that they use the recordings in a number of ways.

The most common way is to watch some of the recordings, especially the review sessions, just before tests. Another is to attend class whenever possible, but watch the recordings of any that are missed.

But there were other stories that show how individual the use can be.

A student with weak English skills attended every class and then rewatched it, occasionally pausing the playback so a friend could translate and explain what was just said. The student did well in the course and his English improved substantially.

A wheelchair bound student also attended every class, but had difficulty taking notes. Instead, he made his notes while watching and pausing the playback in his dorm room.

A student whose use of the recordings was infrequent did watch and repeat portions that were on topics where he had difficulty. “I must have watched this one section ten times, until I got it.”

Another said that he would rewatch some portion of almost every class he attended.

3.3. Student Presentations

It has long been recognized that recording students making presentations to the class and then playing them back helps build public speaking skills. In most rooms setting up for recording and then making the videos available to the students can be tedious and time consuming. One instructor pointed out that making student videos in the cyber classroom is “free”.

In the same vein, instructors also can critique themselves for free.

4. Distance Learning at MIT

The Massachusetts Institute of Technology has a long-standing series of distance learning courses where the video is produced by a person at a console. The operator is responsible for camera motion, zoom and focus, video and audio mixing, plus coordinating with and troubleshooting the remote locations. [3]

They originally had a different type of tracking camera that followed a battery powered target worn by the instructor. That arrangement was not popular because the instructors did not like wearing the target, the target battery would die inconveniently, and the tracking camera sometimes lost sight of the target. They often had to track the instructor by hand, which is both tedious and distracting from other tasks.

They find using a stand-alone Tracking Camera subsystem as a “camera assistant” in each of three lecture halls much more reliable.

5. Lecture Distribution at the Broad Institute

The Broad Institute in Cambridge Massachusetts has two rooms with high definition Tracking Cameras run as stand-alone subsystems. In each room the Tracking Camera video is combined with a high definition scan-converted image from the projector. The resulting live side-by-side programs are sent to colleagues in another building about a mile away. [2]

6. The Future

As of late 2009, all the existing full AutoAuditorium Systems (Tracking Camera plus Director) are NTSC standard definition systems. The oldest one is 10 years old as of January 2010.

Two of the Tracking Camera only Systems use high definition tracking cameras but the spotting cameras are standard definition, which means the algorithms run on standard definition images to drive the HiDef cameras. The hardware costs for high definition are higher, but expected to drop sharply as HiDef becomes the standard.

There is nothing special about processing high definition images. There are just more pixels to look at. Every-day workstation and high-end desktop computers have sufficient computation and bus speeds to keep up with the extra processing needed. The first HiDef AutoAuditorium System is now on the drawing boards to be installed in 2010.

7. Conclusions

The AutoAuditorium System is an automatic video production system that can serve as the front-end to many technologies used to capture and distribute videos of classroom and auditorium lectures. The Tracking Camera subsystem follows the presenter giving a talk and the Director subsystem produces a fully edited single-stream television program of the talk. The program, produced in real-time, includes pictures of the presenter, pictures of the projected

visual aids, picture-in-picture combinations of the presenter and the projections, and, occasionally, other views of the room.

Once built into a room, the cost of using the system to capture the next presentation is little more than the cost of the recording and transmission media.

Programs of presentations, lectures and classes produced by the system have proven to have benefits for both the presenters and viewers.

References

- [1] M. H. Bianchi. Automatic video production of lectures using an intelligent and aware environment. In *3rd International Conference on Mobile and Ubiquitous Multimedia (MUM2004)*, pages 117–123, October 2004. http://autoaud.com/AutoAud_MUM2004.pdf.
- [2] The Broad Institute
Conversations with Lars-Erik Siren, Executive Producer.
- [3] Massachusetts Institute of Technology
Conversations with Michael Leoncini, Distance Education Technology Administrator, Academic Media Production Services.
- [4] S. Turner and M. E. Farmer. Assessment of student performance in an internet-based multimedia classroom. In *2008 International Conference on Frontiers in Education: Computer Science and Computer Engineering (FECS08)*, July 2008. http://autoaud.com/FECS08_StudentPerformance.pdf.
- [5] University of Michigan at Flint
Conversations with:
Dr. Micheal E. Farmer, Associate Professor, Computer Science.
Theresa Kraft, Lecturer, Management Information Systems.
Judith Pinn Carlisle, Assistant Professor, Computer Science.
Dr. Yener Kandogan, Associate Professor of International Business, School of Management.
Randy Schapel, student.
Brian McBride, student.
George Croser, student.
Poulami Nandi, student.