Room with a Rear View

Meeting Capture in a Multimedia Conference Room

Patrick Chiu, Ashutosh Kapuskar, Sarah Reitmeier, and Lynn Wilcox
FX Palo Alto Laboratory

Using advances in audio and video technology, capturing meetings can be much more than recording slide presentations and videotaping speakers. The FX Palo Alto Laboratory combined note-taking software and behind-the-scenes network hardware to create an unobtrusive multimedia conference room capable of capturing a range of presentation styles.

Using high-quality multimedia recording to capture meetings, presentations, and lectures requires a substantial infrastructure. Since meetings often take place inside a room (or rooms virtually connected as in a teleconference), it makes sense to tightly integrate the system into a conference room. Exploiting digital audio and video technology [1-9] to make the room itself a computing environment supports the recording of its activities and provides content for documentation, communication, and collaboration.

Meetings span a range of informational and group activities. Examples include staff meetings, design discussions, project reviews, videoconferences, presentations, and classes. Usually handwritten notes, augmented with presentation material that's either hand copied or obtained from the speaker form the basis of meeting records. Some cases require more detail, so the meeting is recorded on audio or video. Those who attended the meeting along with those who missed it use these records for review, retrieving facts and details. Sometimes their activities are more involved. They study the records, prepare reports, and create meeting summaries.

This article focuses on applying multimedia to meeting capture and note taking with digital video and ink in a physical conference room. Just as pen and paper notes augments an individual's memory, a multimedia conference room that assists in producing meeting records augments the group memory of collaborative work. There are other opportunities for using technology and media in meeting rooms, including software to facilitate meeting processes, and in this broader area of computer supported meeting rooms a substantial body of research exists—for example, see Nunamaker et al.[10] and Stefik et al.[11]

FXPAL conference room

At the FX Palo Alto Laboratory (FXPAL), we have a multimedia conference room equipped for meeting capture with room cameras and microphones, videoconference cameras, and a large display rear-video projector (see Figure 1). A variety of roomware (a term coined by Streitz et al.[12]) facilitates the capture, display, and transfer of multimedia information.

We aim to support multimedia meeting capture and note taking in a flexible, seamless, and unobtrusive manner in a public conference room used for everyday work. Figure 2 shows a blueprint of the room. Typical conference room furniture with tables and chairs are arranged in a U-shape at the center of the room. Wireless pen-based notebook computers, which may be freely positioned and moved around the room, serve as unobtrusive devices for note taking and interacting with the digital world.

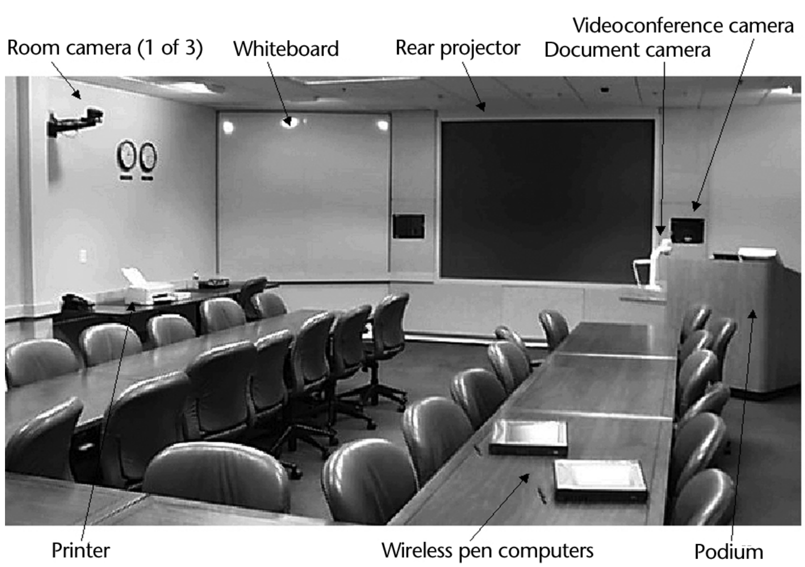

Figure 1. The FXPAL conference room.

In contrast to an earlier generation of computerized meeting rooms with bulky workstations built into fixed tables,[10,11] participants can interact more comfortably in our room, and with less intrusion by the technology.

A flush-mounted large screen (120-inch) rear-video projector displays presentation material on the room's front wall. Any of the following may be used to feed presentation material into the rear projector: a PC workstation, a visitor's laptop, a VCR, a DVD player, or a document camera. The document camera folds up and retracts into a podium drawer. A visitor may bring a laptop and plug it into a connector at the podium. Controls on the podium let the presenter select the media source for the rear projector (see Figure 3). It also has a thin LCD display, a keyboard and a mouse connected to the workstation in the control room. The control room houses all of this hardware out of sight, and is walled off from the conference room with its own entrance (see the right of Figure 2).

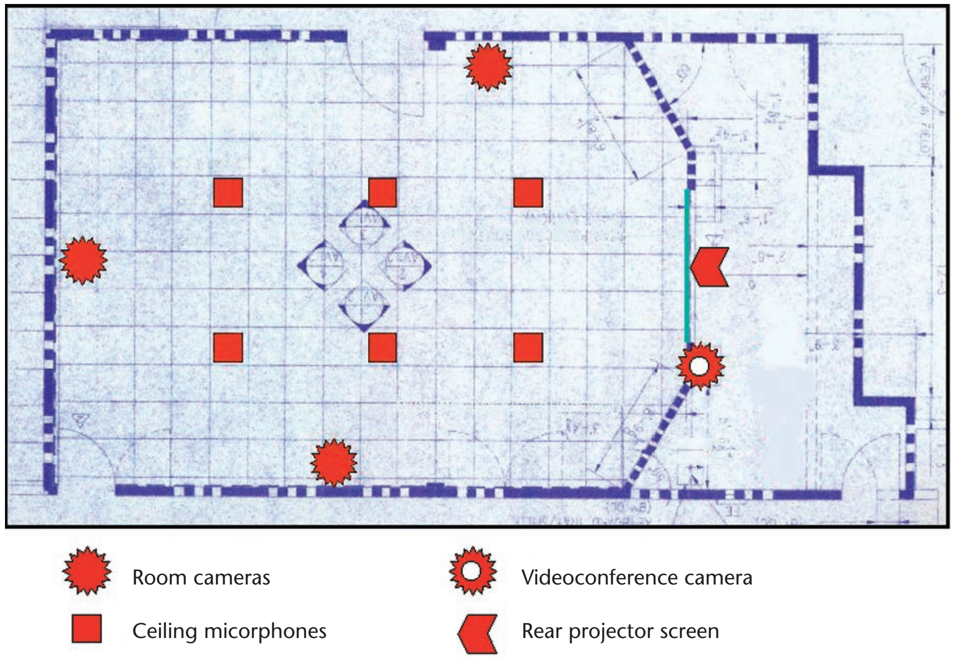

Figure 2. Blueprint of conference room.

The room contains three computer controllable cameras plus a videoconference camera for capturing and transmitting room activity. A room camera can also be used to obtain whiteboard images. Six ceiling microphones handle the audio. We combine their inputs into a single audio stream and mix them together with the video. A high-bandwidth, IEEE 802.11-compliant 11-Mbit wireless system provides network connectivity. A small ink jet printer produces color hard copies of notes or presentation material.

From the control room, the room cameras may be tilted, panned, and zoomed. We have presets programmed for different meeting types. For example, in a presentation meeting, we might aim one of the side cameras at the speaker, the other side camera at the participants, and the back wall camera for a wide-angle shot at the whole room. When we need higher quality production, a person sits in the control room and directs the cameras.

The underlying medium for capturing all types of visual images in this setup is video. The room video cameras record the room activity and the scribbles on the whiteboard, the video conferencing system provides images of a remotely connected room, and the rear-video projector captures the presentation material. Video records a variety of visual information from a meeting seamlessly and flexibly. However, a trade-off between versatility

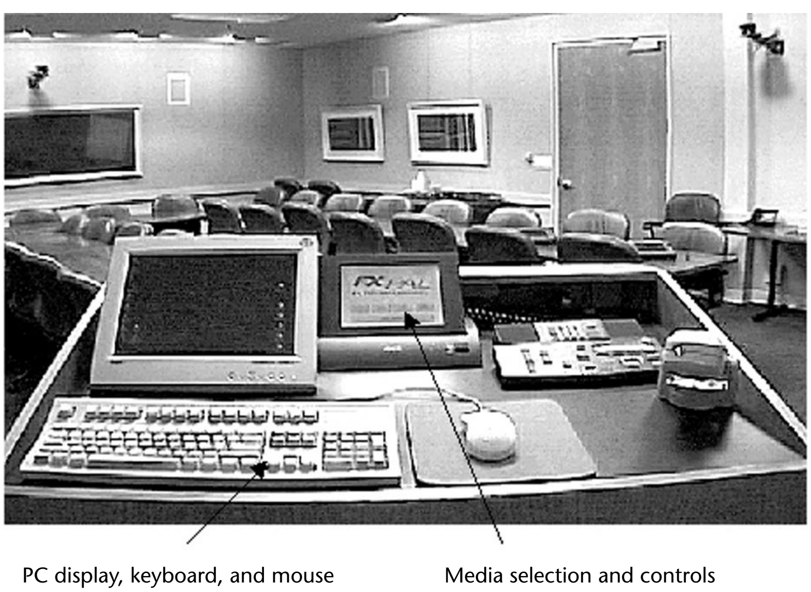

Figure 3. Behind the podium at the front of the room.

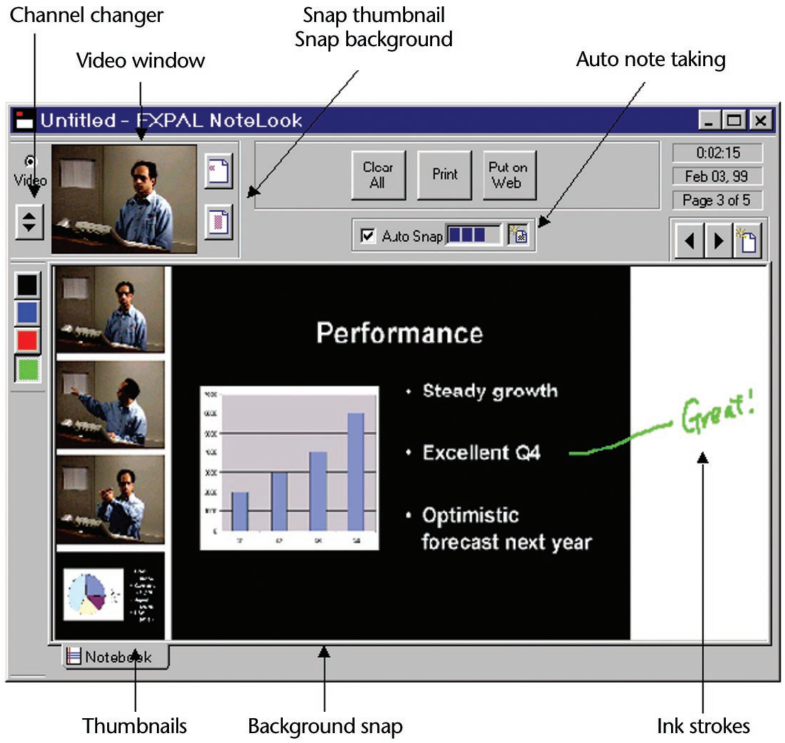

Figure 4. NoteLook application screen shot showing (from upper left-hand corner) the channel changer, video window, snap thumbnail, snap background, auto note-taking, ink strokes, background snap and thumbnails.

and fidelity exists, which we'll discuss in later.

Meeting capture and note taking

Simply recording a meeting isn't enough; it's important to provide tools to index, access, and browse the captured meeting. You can obtain indexes naturally by using the notes from meeting participants. Notes made on pen-based computers can be time-stamped automatically, and later used to locate the interesting points in the multimedia recording. This technique, pioneered at EuroPARC,[13] has been used in digital ink and audio systems for intellectual property management meetings[6] and classroom lectures.[1]

Existing research systems for note taking with digital ink and audio/video may be categorized into ink and audio notebooks,[1,8,14,15] electronic whiteboard and audio sytems,[1,5,6,9] and video annotation systems.[16] Unlike these multimedia note-taking systems and video annotation systems, ours lets users interactively incorporate live video and images from the cameras and presentation material into the notes.

We call our digital video and ink note-taking application NoteLook. It runs on the wireless pen-based notebook computers in the room (see Figure 1). The NoteLook client is a descendant of the FXPAL Dynomite ink and audio standalone application.[15] Figure 4 shows a screen shot of the NoteLook application. Users can incorporate images from the room cameras, the videoconference cameras, and the rear projector into the note pages. They can annotate these images and write freeform notes with digital ink.

Users can monitor the live video in the upper left corner of the application. Around this video window are several buttons for interacting with the video. Pressing the Video button on the left makes a connection to the NoteLook server and initiates video recording and transmission to the clients. The left bottom button changes the video channels. We normally operate with two channels—one for the room activity from a preselected room or videoconference camera, and one for presentation material shown on the rear projector. Usually, the preselected camera is a room camera pointed at the speaker. The right top button snaps the image in the video window as a thumbnail into the note page's margin. When NoteLook snaps a sequence of thumbnails, they're laid out one below another (see Figure 4). Underneath the thumbnail button is a button for snapping in a large background image. Snapping a new background image overwrites any existing background image on the page.

NoteLook uses the interaction model YCAGWYS (you can always get what you see). Images of the room activity and the presentation material can be captured in real time as the user sees them. NoteLook accomplishes this seamlessly by using video as the underlying medium.

Furthermore, NoteLook has a feature for automatic note taking. In this mode, when the presenter puts up a new slide on the rear projector, the system automatically detects it and places it as the background of a new page. NoteLook then appends this page to the stack of note pages. The system also places a sequence of thumbnails from the room cameras in the margin. When users turn to that page, they can annotate the images with ink. This feature relieves the users of snapping in many slides during a presentation.

At the end of the meeting, users can print copies of the notes or put them on the Web for easy access and sharing.

Accessing and browsing captured meetings

You can browse captured meetings indexed with NoteLook notes on the Web. Figure 5 shows a sample. A NoteLook application generates the HTML pages. The thumbnails, background snaps, and ink strokes on the Web pages have links to the recorded video. These links are based on the time stamp made during note taking. The video playback correlates to those times and is played back in a separate application window. We integrated the NoteLook Web pages with a video playback application developed at our lab called the Metadata Media Player.[17]

The Web pages have several standard navigational features. A weekly meeting notes Web page lists the NoteLook sessions. For each session, a table of contents page shows all the pages as reduced images. Clicking on one of the reduced images brings up that page. Each page may be zoomed in or out with a range of five different magnification levels.

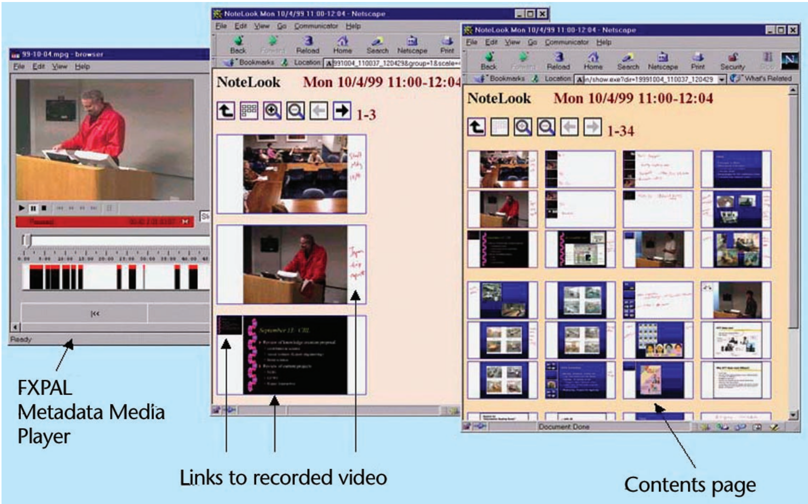

Figure 5. Web access to captured meetings and notes. The FXPAL Metadata Media Player (left), links to the recorded video (middle), and contents page (right).

Media management and system architecture

We designed the NoteLook client as a lightweight application. However, it requires a substantial video infrastructure to obtain adequate quality images of the room activity and presentation material. To deal with this lightweight-heavyweight trade-off, we off-load most of the video processing and media management to the NoteLook servers and switchers. While the space in the conference room is relatively clutter free (see Figure 1), many pieces of the system lie outside the room, hidden away from meeting participants.

The NoteLook system architecture connects the various components inside and outside of the room by employing both wired and wireless networks (see Figure 6). The white rectangle in the illustration corresponds to the interior of the conference room. The key pieces consist of the NoteLook clients, servers, and switchers for video source management. We describe the interplay of these along with other components in more detail below.

The NoteLook system is auto-configurable, extensible, and scalable. The clients and servers configure themselves automatically using resource discovery techniques. Each video channel corresponds to a server associated with a set of sources.

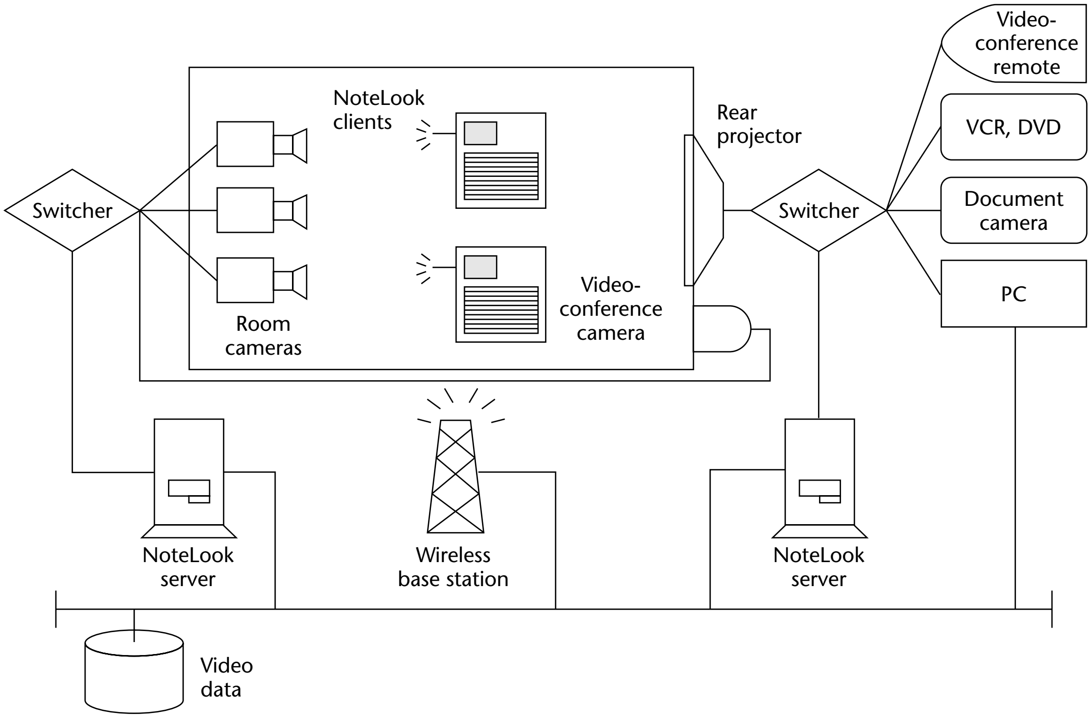

Figure 6. NoteLook system architecture.

The NoteLook servers take video and audio inputs, process them, transmit the output to the NoteLook clients, and store the data for later retrieval.

Adding and removing servers or channels doesn't require modifying existing clients, and multiple clients can be supported by multicasting.

The set of sources that feed into a channel typically relate to each other. For example, the presentation material channel comes from the set of sources that feed into the rear-video projector. The switchers have the capability to either manually or automatically select the desired source.

Another trade-off in the NoteLook system is that of versatility and fidelity. While video provides a versatile way to capture room activity and presentation material, simply aiming a camera in the room at the large display to capture presentation material (as in the Streams[3] system) has problems. It can cause skewing and inadequate resolution. In addition, speakers sometimes block parts of a displayed slide. Another approach without employing video is to force speakers to use a specific image file format like GIF for their presentation slides, as in the Classroom 2000[1] and Cornell Lecture Browser[7] systems. This requires extra preparation time and is troublesome for visiting speakers.

Our more flexible system captures images of the presentation material shown through the rear-video projector. Presentation material may come in a variety of forms—PowerPoint slides or Web pages from a computer, paper or plastic transparencies, overhead slides via the document camera, whiteboard via a room camera, video clips from VCR or DVD, and so on. While the rear projector video feed is versatile enough to capture images of different presentation materials, it doesn't always provide the highest quality images. For example, by the time an image of a PowerPoint slide travels from a PC's video output through the plumbing (which may contain various splitters and scan converters) and reaches the rear projector, the captured image degrades so that the text on the slide is difficult to read.

To deal with this problem, images in our system can be taken at the projector or closer to the device handling the presentation material. We use a smart-source-management component to identify the highest fidelity source available for capturing images. The source-management component operates automatically in real time and interfaces with the switchers and a room automation and control system. We use a commercial AMX system, which is programmable and provides hooks for turning a piece of equipment on/off and checking its on/off status and input/output sources.

Here's how this works. When the rear projector displays PowerPoint slides running from the PC workstation, the source-management component directs the server to get the images from the PC by screen snap (that is, the PC's screen bitmap, not the PC monitor video signal, nor the rear-projector video signal). When speakers supply their own laptop, the server must get its images further downstream from the rear projector's video signal with some unavoidable degradation in fidelity. This results in obtaining the best images while keeping video-source management hidden from users.

The NoteLook servers take video and audio inputs, process them, transmit the output to the NoteLook clients, and store the data for later retrieval. When users initiate a session by pressing the Video button on the client application, it broadcasts a request for service. The servers respond and identify themselves, and establish connections.

To conserve wireless bandwidth, video is transmitted to the clients at a highly reduced frame rate (0.5 frame per second). During a meeting, users don't necessarily need full-motion video for note taking since they're present in the room watching the live action. A software component that runs on the servers handles automatic note taking and analyzes the video data. When the speaker puts up a fresh slide, the software detects it and packages it along with a sequence of thumbnails of room images. The server sends these to the client for creating a new note page.

For more details on the NoteLook system, and a full discussion of related meeting capture systems, see Chiu et al.[18]

User experience

As part of our design process, we conducted a user study over a six-week period with 13 meetings. These meetings consisted of presentations, staff meetings, and Japanese language classes. The user study tested the system's robustness and provided insights that led to a number of refinements and features.

We observed instances where it's important that the speakers use their own presentation material format and not be distracted by the devices in the room. One of the presentations was a job interview talk, and another was a briefing by a former head of a lab. These speakers wouldn't likely have used a system that requires extra preparation, and distractions wouldn't have been tolerable because of the serious nature of these events. In the 13 sessions, we used and captured a variety of presentation material seamlessly—PowerPoint slides (sometimes embedded with video clips), Web pages, a custom JPEG slide viewer, transparencies on the document camera, scribbles on the whiteboard, and videotapes. From interviews with the users, we found that the system and the wireless pen notebook computers were minimally intrusive to the presenters and the participants in the room.

With any kind of recording technology, privacy concerns exist. We have a policy of asking speakers for permission to record them and none of them objected. Members of the lab are familiar with the conference room and know that the cameras and microphones are live.

We observed two distinct note-taking styles—*image-centric* and *ink-centric*. The image-centric note takers took many background snaps and thumbnails. They produced a set of note pages that resembled a full set of the speaker's slides with annotations. In contrast, the ink-centric note takers had more handwritten notes in ink supplemented by some images, and fewer pages of notes overall. It's possible that other styles may emerge as more users take notes with the system over long periods of use. During testing we noticed it took more work for users performing image-centric style note taking to manually snap each new image into NoteLook note pages. We developed the automatic note-taking feature to relieve them from this repetitive task.

We currently use the meeting capture capabilities of our multimedia conference room in many of our meetings. We regularly capture our seminars and staff meetings, and we've begun to explore remote meetings via videoconference. In addition, people outside the lab have made requests to use the facility for capturing meetings and workshops. Several issues have arisen such as security and privacy of the recorded data, accessing multimedia recordings from printed notes, and providing better integration with lightweight tools on the Web for note taking and browsing of meeting records. We plan to explore these issues further, as well as design new applications for multimedia conference rooms. **MM**

Acknowledgments

We thank Sara Bly, John Boreczky, John Doherty, and Andreas Girgensohn for their valuable help on this project.

References

1. G.D. Abowd et al., "Investigating the Capture, Integration and Access Problem of Ubiquitous Computing in an Educational Setting," *Proc. CHI 98 Human Factors in Computing Systems*, ACM Press, New York, 1998, pp. 440-447.
2. P. Chiu et al., "Meeting Capture in a Media Enriched Conference Room," *Proc. CoBuild 99*, Lecture Notes in Computer Science (LNCS), Vol. 1670, Springer-Verlag, Heidelberg, Germany, 1999, pp. 79-88.
3. G. Cruz and R. Hill, "Capturing and Playing Multimedia Events with Streams," *Proc. ACM Multimedia 94*, ACM Press, New York, 1994, pp. 193-200.
4. L. He et al., "Auto-Summarization of Audio-Video Presentations," *Proc. ACM Multimedia 99*, ACM Press, New York, 1999, pp. 489-498.
5. S. Minneman et al., "A Confederation of Tools for Capturing and Accessing Collaborative Activity," *Proc. ACM Multimedia 95*, ACM Press, New York, 1995, pp. 523-534.
6. T.P. Moran et al., "'I'll Get that Off the Audio': A Case Study of Salvaging Multimedia Meeting Records," *Proc. CHI 97*, ACM Press, New York, 1997, pp. 202-209.
7. S. Mukhopadhyay and B. Smith, "Passive Capture and Structuring of Lectures," *Proc. ACM Multimedia 99*, ACM Press, New York, 1999, pp. 477-487.
8. S. Whittaker, P. Hyland, and M. Wiley, "Filochat: Handwritten Notes Provide Access to Recorded Conversations," *Proc. CHI 94*, ACM Press, New York, 1994, pp. 271-276.
9. C. Wolf, J. Rhyne, and L. Briggs, "Communication and Information Retrieval with a Pen-Based Meeting Support Tool," *Proc. Computer-Supported Cooperative Work 92*, ACM Press, New York, 1992, pp. 322-329.
10. J.F. Nunamaker et al., "Electronic Meeting Systems

to Support Group Work," *Comm. ACM*, Vol. 34, No. 7, 1991, pp. 40-61.
11. M. Stefik et al., "Beyond the Chalkboard: Computer Support for Collaboration and Problem Solving in Meetings," *Comm. ACM*, Vol. 30, No. 1, 1987, pp. 32-47.
12. N. Streitz, J. Geiβler, and T. Holmer, "Roomware for Cooperative Buildings: Integrated Design of Architectural Spaces and Information Spaces," *Proc. CoBuild 98*, LNCS, Vol. 1370, Springer-Verlag, Heidelberg, Germany, 1998, pp. 4-21.
13. M. Lamming and W. Newman, "Activity-Based Information Retrieval: Technology in Support of Personal Memory," F.H. Vogt, ed., *Proc. Information Processing 92 (Personal Computers and Intelligent Systems)*, Vol. 3, Elsevier Science, Amsterdam, 1992, pp. 68-81.
14. L. Stifelman, *The Audio Notebook: Paper and Pen Interaction with Structured Speech*, PhD thesis, MIT Media Lab, Cambridge, Mass., 1997.
15. L.D. Wilcox, B.N. Schilit, and N. Sawhney, "Dynomite: A Dynamically Organized Ink and Audio Notebook," *Proc. CHI 97*, ACM Press, New York, 1997 pp. 186-193.
16. K. Weber and A. Poon, "Marquee: A Tool for Real-Time Video Logging," *Proc. CHI 94*, ACM Press, New York, 1994, pp. 58-64.
17. A. Girgensohn et al., "Facilitating Video Access by Visualizing Automatic Analysis," *Proc. Interact 99*, IOS Press, Amsterdam, pp. 205-212.
18. P. Chiu et al., "NoteLook: Taking Notes in Meetings with Digital Video and Ink," *Proc. ACM Multimedia 99*, ACM Press, New York, 1999, pp. 149-158.

Patrick Chiu is a researcher at FX Palo Alto Laboratory, where as a member of the Smart Media Spaces group he helps design and build applications for meeting capture and note taking. His current research interests include multimedia applications and content analysis, computer-supported cooperative work, and user interfaces. He received a PhD from Stanford University in mathematics and graduated summa cum laude from UC San Diego with a BA in mathematics.

Ashutosh Kapuskar is the founder of DialW.com, a Silicon Valley startup that builds products to access the Internet from cell phones, PDAs, and other wireless handheld devices. His current areas of interest are transactions, multimedia, and user interfaces. He has developed products for conferencing and network management. He has worked as a software consultant for FX Palo Alto Laboratory, Xerox PARC, Xerox LiveWorks, Novell and Sun Microsystems. Ashutosh has a BS in electrical engineering from Nagpur University, India.

Sarah Reitmeier is a user interface designer at Documentum, where she's responsible for user research, interaction design, and improved usability for the company's new Web content-management tools. Her current area of interest is XML-based authoring systems and interfaces. She was a summer intern at FX Palo Alto Laboratory before receiving a MS from the School of Information at the University of Michigan. She also has a MA in English from the University of Chicago.

Lynn Wilcox is a staff scientist at FX Palo Alto Laboratory, where she manages the Smart Media Spaces group. Her current research interests are in audio and video editing, indexing, and retrieval. She worked at Xerox PARC for 10 years on various projects involving audio and image word-spotting and segmentation. She has a PhD in mathematical sciences from Rice University.

Readers may contact the authors at FX Palo Alto Laboratory, 3400 Hillview Ave., Bldg. 4, Palo Alto, CA 94304, e-mail at {chiu, kapuskar, reitmeier, wilcox}@pal.xerox.com, http://www.fxpal.xerox.com.