ELSEVIER

Annotations for streaming video on the Web: system design and usage studies

David Bargeron *,[1], Anoop Gupta [1], Jonathan Grudin [1], Elizabeth Sanocki [1]

Microsoft Research, Redmond, WA 98052, USA

Abstract

Streaming video on the World Wide Web is being widely deployed, and workplace training and distance education are key applications. The ability to annotate video on the Web can provide significant added value in these and other areas. Written and spoken annotations can provide 'in context' personal notes and can enable asynchronous collaboration among groups of users. With annotations, users are no longer limited to viewing content passively on the Web, but are free to add and share commentary and links, thus transforming the Web into an interactive medium. We discuss design considerations in constructing a collaborative video annotation system, and we introduce our prototype, called MRAS. We present preliminary data on the use of Web-based annotations for personal note-taking and for sharing notes in a distance education scenario. Users showed a strong preference for MRAS over pen-and-paper for taking notes, despite taking longer to do so. They also indicated that they would make more comments and questions with MRAS than in a 'live' situation, and that sharing added substantial value.

Keywords: Video annotation; Multimedia annotation; Streaming video; Distance learning; Workplace training; Asynchronous collaboration

1. Introduction

There has been much discussion about using streaming video on the World Wide Web for workplace training and distance learning. The ability to view content on-demand anytime and anywhere could expand education from a primarily synchronous 'live' activity to include more flexible, asynchronous interaction [20]. However, key parts of the 'live' educational experience are missing from on-demand video environments, including the comments and questions of other students, and the instructor's responses. A crucial challenge to making on-demand video a viable educational tool is supporting this kind of interactivity asynchronously.

A Web-based system for annotating multimedia Web content can satisfy many requirements of asynchronous educational environments. It can enable students to record notes, questions, and comments as they watch Web-based lecture videos. Each note can be stored with meta data which identifies the precise time and place (in the video) at which the annotation was created. Students can share annotations with each other and instructors, either via the system or via email, and email recipients can be presented with a URL which will take them back to the part of the video which contextualizes the annotation. Students can watch their own and others' annotations scroll by as the video plays. Annotations can be used

[1] E-mail: {davemb, anoop, jgrudin, a-elisan}@microsoft.com
* Corresponding author.

as a table-of-contents to allow jumping to relevant portions of the video. By allowing students to organize their annotations for public as well as personal use, such a system can go 'beyond being there'. Most note-taking is personal, but the communication and information-management capabilities of the Web can potentially promote greater interactivity than a traditional classroom.

We have implemented a prototype system called MRAS (Microsoft Research Annotation System) which implements much of this functionality. We present the MRAS architecture, its functionality, and its user-interface. MRAS is a Web-based client/server framework that supports the association of any segment of addressable media content with any other segment of addressable media. The underlying MRAS framework can be used for virtually any annotation task, for instance taking notes on 'target' documents of any media type (such as Web pages, 3D virtual reality simulations, or audio recordings), supporting automated parsing and conversion routines such as speech-to-text and video summarization, or even composing dynamic user-specific user interfaces. Since MRAS supports the storage and retrieval of media *associations* without placing restrictions on the media themselves, it represents a first step toward a distributed authoring platform in which multimedia documents are composed at runtime of many disparate parts. The current client-side user interface is specifically tailored for annotating streaming video on the Web.

In this paper we focus on the use of MRAS in an asynchronous educational environment featuring on-demand streaming video. In addition to describing the MRAS system design, we report results from preliminary studies on its use. We compare taking notes with MRAS to taking notes with pen and paper. We look at how users built on each others' comments and questions asynchronously using MRAS, similar to the way discussions evolve in 'live' classroom situations. We examine issues in positioning annotations for video and how users track previously made annotations. Finally, we examine the differences and similarities in how users create text and audio annotations with MRAS.

While we focus here on the use of MRAS in an educational context, we observe that Web-based annotations on multimedia content have much broader applicability. When content represents captured audio-video for presentations and meetings in corporate settings, for instance, annotations can enhance collaboration by distant and asynchronous members of the group, and can enhance institutional memory. Similarly, in a product design scenario, user-interface engineers can capture videos of product use and then create annotations to discuss with developers the problems the end-users are experiencing. Annotations also offer a novel way for authoring multimedia content, where the instructions on how to put together the final content are not hardwired and stored with the original 'base' content, but are assembled dynamically based on the available annotations and the end-users interest profile. Indeed, there are few Web-based scenarios involving collaboration where annotations would not enhance the situation significantly.

The extensive literature on annotation systems primarily addresses annotation of text [5,12,15]. The few systems concerned with annotations of video have mainly focused on the construction of video databases [3,21] rather than on promoting collaboration. We discuss the relationship between our work and earlier projects in Section 5.

The paper is organized as follows. Section 2 describes the architecture, functionality, and user interface of the MRAS system. Section 3 presents results of our personal note-taking study, and Section 4 presents results of our shared-notes study. We discuss related work in Section 5, and we present concluding remarks in Section 6.

2. MRAS system design

In this section we present the MRAS architecture, including how it interfaces to other server components (database, video, Web, e-mail), the protocols that provide universal access across firewalls, and how it provides rich control over bandwidth use for annotation retrieval. We show how MRAS supports asynchronous collaboration via annotation sharing, fine-grained access control to annotation sets, threaded discussions, and an interface to e-mail. We discuss the unique user-interface challenges and opportunities involved in annotating video, including precise temporal annotation positioning, video

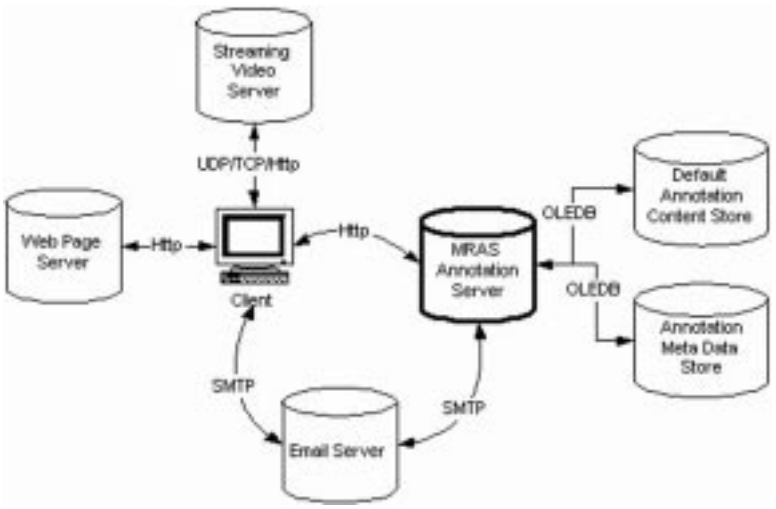

Fig. 1. MRAS system overview.

seek by annotation position, and annotation tracking while a video is playing.

2.1. MRAS system overview

When a user accesses a Web page containing video, the Web browser contacts the Web server to get the HTML page and the video-server to get the video content. If there are annotations associated with the video in the Web page, the user can contact the MRAS Annotation Server to retrieve them. Fig. 1 shows the interaction of these networked components. The MRAS Annotation Server manages the Annotation Meta Data Store and the Native Annotation Content Store, and communicates with clients via HTTP. Meta data about target content is keyed on the target content's URL. The MRAS Server communicates with E-mail Servers via SMTP, and can send and receive annotations in e-mail. Since the display of annotations is composed with target media at runtime on the client, location and user access rights for target content are not restricted.

2.2. MRAS system components

Fig. 2 shows the MRAS system components and illustrates how they work together. MRAS is composed of a server, clients, and stateless client–server communication via HTTP.

2.2.1. MRAS Server

The MRAS Annotation Server's activities are coordinated by the Multimedia Annotation Web Server (MAWS) module, which is an ISAPI ('Internet Services API') plug-in that runs in the Microsoft IIS ('Internet Information Server') Web server. Client commands arrive at the server in the form of HTTP requests which are unpacked by the Httpsvcs modules and forwarded to the Annotation Back End (ABE) module. ABE is the main system module, containing objects for accessing annotation data stores, composing outgoing e-mail from annotation data, processing incoming e-mail from the E-mail Reply Server, and making method calls to the server from the client side. The E-mail Reply Server is an independent agent which monitors the server's e-mail inbox and processes e-mail messages when they arrive. If a message contains the necessary information to be considered an annotation, the E-mail Reply Server converts the e-mail to an annotation and inserts it into the annotation data stores.

Annotation meta data and 'native' annotation content (i.e. annotation content authored in the MRAS user interface, which for now is either unicode text or wave audio) are stored in Microsoft SQL 7.0 relational databases. Since a clear distinction between meta data and content is maintained, however, annotation content need not be stored in the native annotation content store. Instead, it can be stored anywhere as long as the annotation server can retrieve it (i.e.

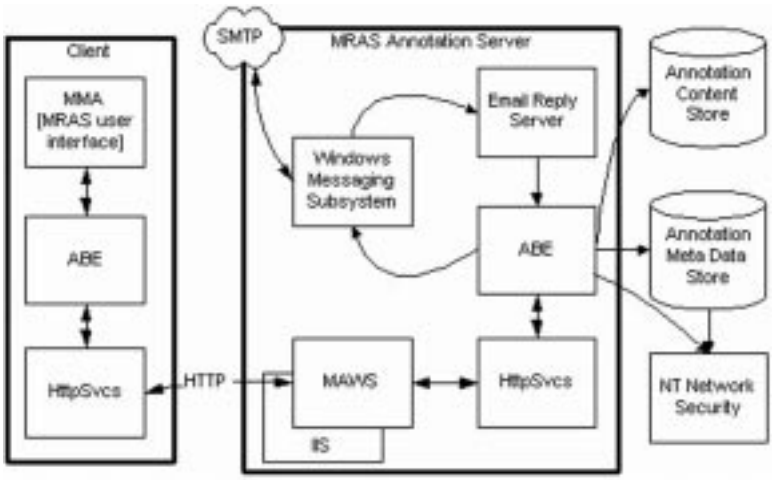

Fig. 2. MRAS system components.

as long as it is addressable). Alternative storage locations include other Web servers or other databases, for instance, and this reveals the possibility that annotations can themselves be third-party content. In addition, nothing in the underlying MRAS architecture restricts the media type of annotation content. Although the MRAS user interface prototype presented here only supports text and audio annotations on digitized video, the underlying framework supports associating any type of electronic media with any other type of electronic media.

User identification is authenticated on the server by referring user logon credentials to the underlying network. Once the user is authenticated, her access rights are enumerated by comparing the network user groups to which she belongs with the MRAS access rights stored in the Annotation Meta Data Store. Access rights are keyed on annotation sets, which group annotations into organizational units. In this way, groups of users are given specific access to groups of annotations. Annotation sets therefore can be private to an individual (in a group consisting of one person), restricted to a particular user group, or made public to all network users.

2.2.2. Client–server communication

Communication between MRAS clients and servers consists of HTTP request/response pairs wherein commands are encoded as URLs and Entity-Body data is formatted as OLE Structured Storage documents. OLE Structured Storage is a flexible binary standard for storing arbitrarily complex composition documents, but is not widely supported on the Web, so we plan to convert our wire data format to XML.

HTTP is used as the primary client–server protocol since it is widely supported on the Internet, and because most firewalls allow HTTP traffic through. Firewall neutrality is important since it allows collaboration among users of MRAS across corporate and university boundaries.

Due to the nature of the HTTP protocol, client–server communication is always initiated by the client. Although a push model of annotation delivery is possible, where users register their interest in a particular annotation set and the annotation server in turn delivers updates as the set is modified by other users, this has not yet been implemented in MRAS. Instead, the MRAS client ‘Query’ functionality (described below) allows users to retrieve annotations created after a particular time, and this opens the possibility for regular client-driven polling of the server.

2.2.3. MRAS client

The last piece of the MRAS system is the client, where the HttpSvcs module handles communication with the server and ABE translates user actions into commands destined for the server. The MRAS user interface is encapsulated in the Multimedia Annotations module (MMA), which hosts ActiveX controls that display an interface for annotating streaming video on the Web. The next section is devoted to an

Fig. 3. MRAS toolbar. The top-level toolbar is displayed at the bottom of the Web browser window. It allows the user to logon to an MRAS server, retrieve existing annotations, add new annotations, and see annotations that have been retrieved from the server.

in-depth discussion of MRAS system functionality and how it is manifested in the MRAS client user interface.

2.3. MRAS system functions

Fig. 3 shows the MRAS toolbar, the top-level user interface supported by the ActiveX controls in the MMA module. This toolbar is positioned in a Web browser just below the Web page display, or it can be embedded anywhere in a Web page. After logon to a MRAS server, users choose the annotation activities they want to perform from this toolbar.

2.3.1. Adding new annotations

Fig. 4 shows the dialog box for adding new annotations. When a user presses the 'Add' button on the top level annotations toolbar (Fig. 3), this dialog box appears. As shown, it currently supports adding text and audio annotations. If there is a video in the current Web page, then when the dialog comes up, the annotation's target position is set to the current position in the video's timeline, and the video pauses. Pausing is required when adding an audio annotation due to hardware constraints on most PCs: there is usually only one sound card available for all applications, which the video uses for its audio track, and the MRAS client uses to support audio annotations. Based on results from pilot tests, we decided to automatically pause the video in both the audio and text cases to make the user experience more consistent.

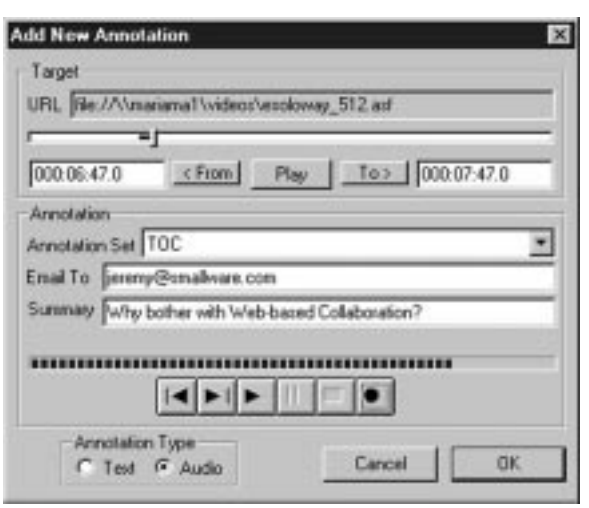

Fig. 4. MRAS 'Add New Annotation' dialog box. Here, an audio annotation is being recorded. It will be added to the 'TOC' annotation set and mailed to `Jeremy@smallware.com`. The user has specified that the annotation is contextualized by the target video from 6 min and 47 s to 7 min and 45 s.

2.3.2. Positioning annotations

Just as an annotation on a text-based document can correspond to a text span, so an MRAS annotation can correspond to a time range within a target video. Each annotation maintains its own range positioning data. Annotation ranges may overlap. Controls at the top of the 'Add' dialog box (Fig. 4) allow users to refine or change the range beginning and end so that their annotations can be positioned more precisely 'over' the appropriate segment of the target video.

2.3.3. Organizing annotations

After the annotation has been positioned, a user must choose from a drop-down list the Annotation Set to which the annotation will be added. Annotation Sets are the fundamental organizational mechanism in MRAS. An Annotation Set can be thought of as any collection of comments, questions, or notes. For example, it may correspond to an individual user's personal notebook, or to the transcript of a meeting. As discussed later, annotation sets also form the basis for access control and sharing. For now, Annotation Sets are implemented as a single-level flat organizational scheme, and an individual annotation can belong to one and only one Annotation Set. This relatively limited implementation will be relaxed in the future, however, so that Annotation Sets can be organized in arbitrary user-defined hierarchies, and so that individual annotations and sets may belong to more than one set.

2.3.4. Sending annotations in e-mail

After an Annotation Set has been specified, a user can enter SMTP e-mail addresses to which the annotation will be sent after being validated by the server. The server copies contextual information and annotation data into the e-mail message, allowing the message recipient(s) to quickly and easily navigate to the appropriate place in the target video. Replies to annotation e-mail messages — as well as user-created e-mail messages which contain the appropriate meta data tags — are handled by the MRAS E-mail Reply Server (ERS). The ERS receives, validates, and processes e-mail messages as if they were annotations. By fully integrating MRAS with SMTP, we take advantage of the power and flexibility of existing e-mail applications as collaborative tools. Users who do not have access to the MRAS client can still participate in the shared discussions supported by the MRAS framework.

2.3.5. Retrieving annotations

Fig. 5 shows the MRAS 'Query Annotations' dialog box, used to retrieve existing annotations from the MRAS server. This dialog box is accessed via the 'Query' button on the MRAS toolbar (Fig. 3).

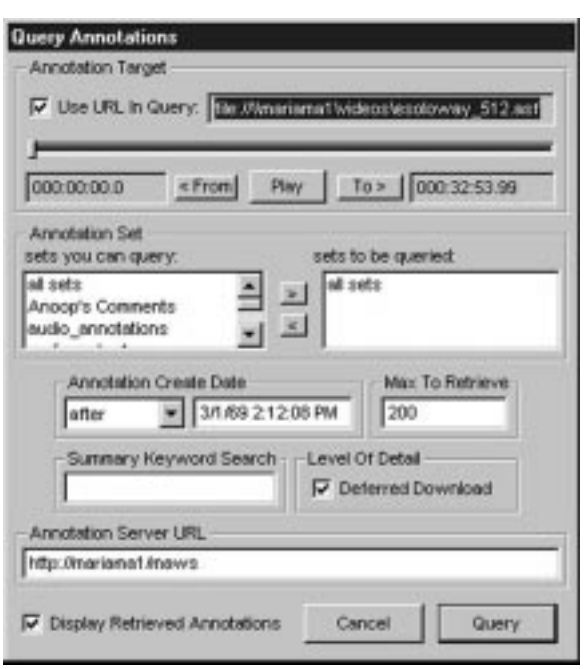

Fig. 5. Query Annotations dialog box.

The controls at the top of the dialog are similar to those in the 'Add New Annotation' dialog box (Fig. 4); however, here they describe the time range in the target video for which annotations are to be retrieved. This, along with the 'Max To Retrieve' and 'Annotation Create Date' boxes in the dialog box, narrow the range of annotations to be retrieved.

The 'Level of Detail' box provides users with control of bandwidth usage when downloading annotations, by allowing them to download various amounts of information for the annotations being retrieved. For instance, if a user selects the 'deferred download' checkbox, then only enough meta data to identify and position each annotation is downloaded. Later, if the user wishes to see the full content of a particular annotation, she can download the data for only the one she is interested in, without having to download the content of other annotations in the query result set.

As in the 'Add New Annotation' dialog box, users are presented with a list of annotation sets in the Query dialog. However, here the list shows sets to which a user has read access, and which contain annotations on the current target video (all other annotation sets are either off-limits to the user or are empty, so it does not make sense to query from them). Users can choose any combination of annotation sets to constrain their query. When the query is launched, only annotations from the specified sets will be retrieved.

The 'Use URL in Query' checkbox at the very top of the dialog box allows a user to specify whether the current video target content URL will be used as a constraint in the annotation query operation. If it is selected, only the annotations on the current target (the video being viewed in the Web browser's current page, for instance) will be retrieved from the server. If it is deselected, the URL will not be used, and *all* annotations in the specified set(s) will be retrieved (regardless of what video they were created on). This opens the possibility for annotation 'playlists', wherein a list of annotations can be used to define a composite view of information across a number of targets. We have prototyped several composite video presentations involving different Web pages using this feature.

The 'Summary Keyword Search' box allows a user to retrieve only those annotations which meet

all of the other query criteria and which also contain the specified keywords in their summary strings. This feature is particularly useful if users know what they are looking for in a large set of annotations.

2.3.6. Access control and sharing

The Annotation Set list presented in the 'Query Annotations' dialog differs slightly from the list in the 'Add' dialog. Here, it lists sets to which a user has read access *and* which contain annotations on the current target video, whereas in the 'Add' dialog it is simply a list of sets to which the user has write access. In addition to being the fundamental organizational mechanism, Annotation Sets serve as the basic entities against which access rights are established and controlled. Users are identified and authenticated in MRAS by their network user accounts and user group affiliations. A table in the Annotation Meta Data Store (Fig. 2) stores a permissions map that relates user groups to Annotation Sets. In this way, MRAS affords a fine-grained mechanism for controlling access rights between groups of users and sets of annotations.

Another important use of Annotation Sets is to support collaboration through fine-grained, structured sharing of annotations. In a college course, for instance, one set could be created for each student called 'student X's notebook', to which only the student has read/write access and the professor has read access. Another set entitled 'class discussion' could grant read/write access to all class members. And a set called 'comments on homework' could be created, to which TAs have read/write access and everyone else has read access.

2.3.7. Viewing and using annotations

The MRAS 'View Annotations' window (Fig. 6) displays annotations retrieved from the server in a tree structure. Several interesting MRAS features are introduced here, including annotation reply, seek, tracking, and preview.

Reply and seek functionality are accessed through a menu that appears when a user clicks over an annotation in the view window. 'Reply' allows a user to elaborate on a previously created annotation by creating a child annotation. The ability to create child annotations forms the basis for support of threaded discussions in MRAS. Child annotations

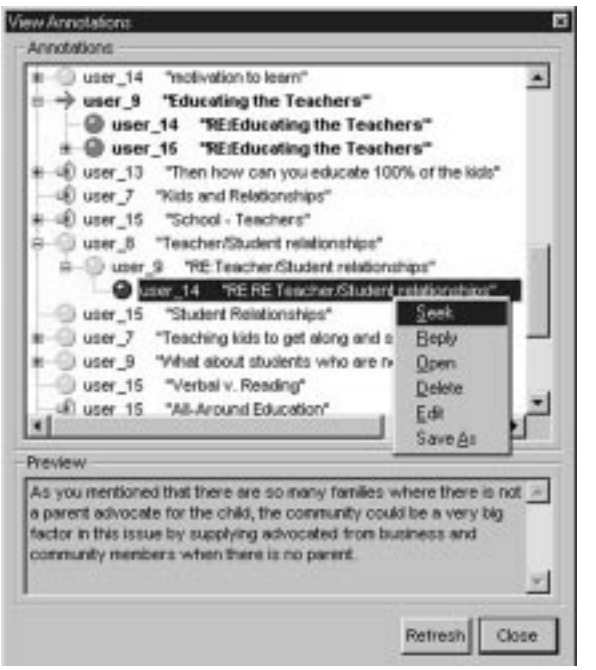

Fig. 6. MRAS 'View Annotations' window. The annotation entitled 'Educating the Teachers' and its children are being tracked. A reply to the annotation entitled 'Teacher/Student relationships' has been selected by the user. The preview pane shows the contents of the user-selected annotation in this example, but it can also show the contents of the tracked annotations. The mouse-click menu for manipulating the selected annotation is also displayed.

inherit all of their contextual meta data from their parent, so that a threaded discussion is consistently contextualized by the target video at every point in the discussion, and the context can be recovered from any part of the discussion.

'Seek' changes the current video position to the beginning of the range over which an annotation was created. Using seek, users can easily see the part of the video that corresponds to the annotation. This also allows annotations to be used to create personal or shared table-of-contents.

The tracking feature polls the target video at regular intervals for its current position and updates the 'View' window. Annotations *close* to the current target video position (within some epsilon) are highlighted with red icons and bold typeface, and the *closest* annotation gets an arrow icon. Tracking allows users to see where they are in a list of annotations as the target video plays. As annotations

are tracked, the preview feature displays the content of the closest annotation, without the user having to manually expand it. This was a popular feature in our usage studies, especially in the 'shared annotations' condition in which users added comments and questions to a set containing annotations added by previous users.

3. Personal notes usage study

We conducted two studies of annotation creation on streaming video over the Web. The first concerned personal note-taking using MRAS. Since there is little if any past experience with the use of such systems, our primary motivation was to gain intuition about how people do use it — how many annotations they create, how they position annotations within the lecture video, what their reaction is in contrast to pen and paper (with which they have decades of experience).

3.1. Experimental procedure

3.1.1. Participants

Six people participated in the 'Personal Notes' usage study. They were intermediate to advanced Microsoft Windows users with no involvement with the research, and were given software products for their participation.

3.1.2. Methods

Participants were asked to assume that they were students in a course. As preparation for a discussion planned for the next class meeting, they needed to watch a video presentation of Elliot Soloway's ACM '97 conference talk "Educating the Barney Generation's Grandchildren". Their goal was to generate questions or comments from the video for the class discussion. Each was given as much time as needed to complete the task.

During the video, subjects made annotations in two ways: using handwritten notes and using text-based MRAS notes. Half of the subjects (three) took handwritten notes during the first half of the video and switched to MRAS for the second half of the video (paper-first condition). The other three used MRAS for text-based note taking during the first half, and pen and paper for the second half (MRAS-first condition).

Subjects began the study session by completing a background questionnaire. For the MRAS-first condition, subjects watched the first half of the video after familiarizing themselves with the MRAS user interface. For the paper-first condition, subjects were told to take handwritten notes in the way they normally would in a 'live' classroom situation. Halfway through the talk, the video stopped and subjects switched to the other note-taking method. Those in the paper-first condition familiarized themselves with taking notes using MRAS before continuing with the video, and those in the MRAS-first condition were told to take handwritten notes in the way they were accustomed to.

From behind a one-way mirror we observed behaviors such as pausing or rewinding the video. Task time and annotation number and content were also recorded (see Table 1). After the study, sub-

Table 1
Personal notes study results

Cond.	Subject	MRAS Notes	Paper Notes	Total Notes	MRAS Time	Paper Time	Total Time
MRAS First	1	15	16	31	46.58	28.05	74.63
	2	19	13	32	40.00	28.00	68.00
	3	39	21	60	38.59	28.29	66.70
	Avg (SEM)	24.33 (7.42)	16.67 (2.33)	41.00 (9.50)	41.69 (2.48)	28.08 (0.06)	69.78 (2.46)
Paper First	4	7	14	21	31.28	30.00	61.28
	5	13	14	27	34.00	17.07	51.07
	6	8	19	27	25.97	16.50	42.47
	Avg (SEM)	9.33 (1.86)	15.67 (1.67)	25.00 (2.00)	30.41 (2.36)	21.19 (4.41)	51.61 (5.44)
Total Avg(SEM)		16.83 (4.79)	16.17 (1.30)	33.00 (5.63)	36.06 (2.95)	24.64 (2.50)	60.69 (4.86)

'SEM' is the standard error of the mean.

jects completed a post-study questionnaire and were debriefed.

3.1.3. Number of notes and time spent taking notes

Although the video was just 33 min, the participants using MRAS took an average of 16.8 notes; subjects using pen and paper took a statistically indistinguishable average of 16.2 notes.

Interestingly, taking notes using MRAS took 32% longer than on paper (this difference was significant at probability $p = 0.01$, based on a repeated measures analysis of variance (ANOVA), the test used below unless indicated otherwise). With paper, users often took notes while the video continued to play, thus overlapping listening and note-taking. For MRAS, the system pauses the video when adding, to avoid audio interference. One might reconsider the design decision to force pausing when conditions allow (e.g., for text annotations). However, this difference did not seem to negatively affect the subjects' perception of MRAS — as noted below, all six subjects reported that the benefits of taking notes with MRAS outweighed the costs.

Another surprising discovery was how MRAS-first subjects took paper notes differently from paper-first subjects. *All* three MRAS-first subjects paused the video while taking paper notes; *none* of the paper-first subjects paused. This suggests that people modified their note-taking styles as a result of using MRAS.

3.1.4. Contextualization and positioning personal notes

An important aspect of MRAS is that annotations are automatically linked (contextualized) to the portion of video being watched. We were curious how subjects associated their paper notes with portions of lecture, and their reactions to MRAS.

Examining the handwritten paper notes of subjects, we found little contextualization. It was difficult or impossible to match subjects' handwritten notes with the video. The exception was one subject in the MRAS-first condition who added timestamps to his handwritten notes. On the other hand, it was easy to match subjects' notes taken with MRAS with the video, using the seek feature. Subjects' comments reinforced this. One subject in the paper-first case told us that MRAS "...allows me to jump to areas in the presentation pertaining to my comments." We elaborate on this in the shared-notes study described below.

3.1.5. User experience

In comparing note-taking with MRAS to using paper, all six subjects expressed preference for MRAS to paper. Comments emphasized organization, readability, and contextualization as reasons for preferring MRAS annotations. One subject stated that the notes she took with MRAS were "...much more legible, and easier to access at a later time to see exactly what the notes were in reference to".

Subjects also felt that the notes taken with MRAS were more useful. When asked which method resulted in notes that would be more useful six months down the road (for instance for studying for an exam) *all* six subjects chose MRAS. They again cited issues such as better organization, increased readability, and the fact that MRAS automatically positions notes within the lecture video. In this regard a subject said of her notes "the end product is more organized... The outline feature ... allows for a quick review, with the option to view detail when time permits." Subjects noted that the seek and tracking features of the MRAS user interface were particularly useful for relating their notes to the lecture video, and this added to the overall usefulness of their notes. While the subject population was too small to make any broad generalizations, these results are extremely encouraging.

4. Shared-notes study

Our second study sought to assess the benefit of sharing notes, and to using a particular medium for making notes.

When notes are taken with pen and paper, the possibilities for collaboration and sharing are limited at best, and people generally are restricted to making text annotations, incidental markings, and drawings. With a Web-based system such as MRAS, users can share notes with anyone else on the Web. Users can group and control access to annotations. The potential exists for annotations using virtually any media supported by a PC. In our second study, we explore these possibilities.

4.1. Experimental procedure

The design for the 'Shared Notes' study was similar to that of the 'Personal Notes' study. The same video and similar instructions were used. The difference was that all subjects used MRAS to take notes for the entire duration of the video, and they were told that they were adding their comments and questions to a shared set of notes.

4.1.1. Subjects

Eighteen new participants were recruited from the subject population used in the 'Personal Notes' study. Each participated in one of three conditions: a Text-Only condition where they only added text annotations, an Audio-Only condition, and a Text-and-Audio condition, where both text and audio annotations were allowed.

4.1.2. Procedure

Again, subjects were told to assume that they were creating discussion questions and comments for participation in a class discussion. For each condition, subjects participated in sequence and notes were saved to a common annotation set, so that each subsequent subject could see, review, and respond to the annotations created by previous subjects. The first subject in each condition started with a set of 'seed' annotations, which had been created by a subject in an earlier pilot study and adapted to the appropriate annotation medium for each condition. Subjects were asked to look at existing annotations before adding their own to avoid redundant annotations.

4.2. Results

4.2.1. Number and nature of annotations

One goal of the 'Shared Notes' study was to evaluate the impact of different annotation media on how users annotate. Several previous studies have compared the use of text and audio for annotation of text [4,14]; however, little if any work has been done to investigate the effect of the annotation medium on Web-based video annotations. Also, whereas these previous studies have focused on a qualitative analysis of annotation content, we focus on quantitative issues.

There was a fairly high rate of participation in all three conditions: A total of 50 annotations were added in Text-Only, 41 in Audio-Only, and 76 in the Text and Audio condition (Table 2). However, the number per subject was not significantly different for adding new annotations (one-way ANOVA, $p = 0.45$), replying to existing ones ($p = 0.35$), or for the combination of the two ($p = 0.37$). Neither medium nor the choice of medium had a significant effect on the number of annotations added.

Table 2 shows the results of the 'Shared Notes' usage study, in which subjects were divided into three conditions (Text Only, Audio Only, and Text + Audio). The number of annotations added by each subject was analyzed for type (new vs. reply). In the

Table 2
Shared-notes study results

	Text Only Condition			Audio Only Condition			Text+Audio Condition					Time		
Subject	New	Reply	Total	New	Reply	Total	New	Reply	Audio	Text	Total	Text	Audio	Text+Audio
1st	7	0	7	6	2	8	4	2	2	4	6	55.30	118.85	55.45
2nd	0	1	1	2	1	3	3	3	1	5	6	52.63	71.40	62.32
3rd	7	2	9	5	4	9	5	3	3	5	8	61.58	87.06	45.63
4th	9	6	15	1	0	1	6	6	7	5	12	52.00	44.33	77.05
5th	4	8	12	4	7	11	8	4	1	9	10	75.00	82.48	55.07
6th	5	1	6	6	3	9	14	20	12	22	34	87.07	59.70	66.53
Avg (SEM)	5.33 (1.20)	3.00 (1.32)	8.33 (1.99)	4.00 (0.86)	2.83 (1.01)	6.83 (1.60)	6.33 (1.61)	6.33 (2.79)	4.33 (1.76)	8.33 (2.82)	12.67 (4.37)	63.99 (5.79)	77.14 (10.51)	66.18 (7.42)

Subjects were divided into three conditions (Text-Only, Audio-Only, and Text + Audio). The number of annotations added by each subject was analyzed for type (new vs. reply). In the Text + Audio condition, it was also analyzed for medium (text vs. audio). Subjects tended to use text as an annotation medium slightly more than audio for both new and reply annotations; subjects did not take significantly longer to complete the task in any of the three conditions, and on average the number of replies went up per subject in sequence.

Table 3
Effect of medium on reply annotation counts

Original Medium	Reply Medium	Text-Only Cond. Total	Audio-Only Cond. Total	Text+Audio Cond. Total	All Cond. Total
Audio	Audio	0 (0%)	24 (100%)	6 (16%)	30 (37%)
Audio	Text	0 (0%)	0 (0%)	7 (18%)	7 (9%)
Text	Audio	0 (0%)	0 (0%)	3 (8%)	3 (4%)
Text	Text	19 (100%)	0 (0%)	22 (58%)	41 (50%)

The effect of annotation medium on the number of replies was substantial. Subjects in the Audio-Only and Text-Only conditions were limited to a single medium, so all replies were in that medium. Subjects in the Text + Audio condition were more likely to reply to text annotations using text, and were more likely to reply to text annotations overall.

Text + Audio condition, it was also analyzed for medium (text vs. audio). Subjects tended to use text as an annotation medium slightly more than audio for both new and reply annotations, subjects did not take significantly longer to complete the task in any of the three conditions, and on average the number of replies went up per subject in sequence.

Furthermore, subjects in the Text-Only condition took an average of 64.0 min (SEM or standard error of the mean ± 5.8) to complete watching the video and taking notes, whereas Audio-Only subjects took 77.1 min (± 10.5), and Text + Audio subjects took 65.2 min (± 7.4). Again, the differences were not significant ($p = 0.47$), indicating that neither medium nor the choice of medium affected the time to take notes using MRAS.

Subjects in the Text + Audio condition tended to use text more than audio for creating annotations (both new and reply, see Table 3) ($p = 0.06$), and most expressed a preference for text. When asked which medium they found easier for adding annotations, 4 out of 6 chose text. One said typing text "...gives me time to think of what I want to put down. With the audio, I could have paused, gathered my thoughts, then spoke. But, it is easier to erase a couple of letters... than to try to figure out where in the audio to erase, especially if I talked for a while."

Subjects also favored text for creating reply annotations in particular. This is shown in Table 3. When we looked at replies in the Text + Audio condition, we discovered that subjects were as likely to use text to reply to audio annotations as they were to use audio. Furthermore, subjects were more likely ($p = 0.03$) to use text when replying to text.

Interestingly, subjects in the Text + Audio condition were much more likely to reply to text annotations in the first place ($p = 0.01$). User feedback from both the Text + Audio and Audio-Only conditions explains why: subjects generally felt it took more effort to listen to audio than to read text. One subject in the Text + Audio condition was frustrated with the speed of audio annotations, saying that "I read much faster than I or others talk. I wanted to expedite a few of the slow talkers." Another subject pointed out that "it was easy to read the text [in the preview pane] as the video was running. By the time I stopped to listen to the audio (annotations) I lost the flow of information from the video."

Medium therefore does have a quantitative effect on the creation of annotations in a Web-based system, albeit a subtle one. We anticipated some of these difficulties in our design phase (e.g., as pointed by [14]), and that was our reason for providing a text-based summary line for all annotations, regardless of media type. Several subjects in the Audio-Only and Text + Audio conditions suggested a speech-to-text feature that would allow audio annotations to be stored as both audio and text. This would make it easy to scan audio annotation content in the preview window. We plan to explore this in the near future.

4.2.2. Annotation sharing

Web-based annotations can be shared easily. This opens the door for rich asynchronous collaboration models. We explored several dimensions of annotation sharing, including the impact on the number of annotations over time and the distribution of new and 'reply' annotations added over time. Neither of these dimensions has been explored in depth in the literature.

With subjects participating sequentially, we expected to see the number of new annotations drop off

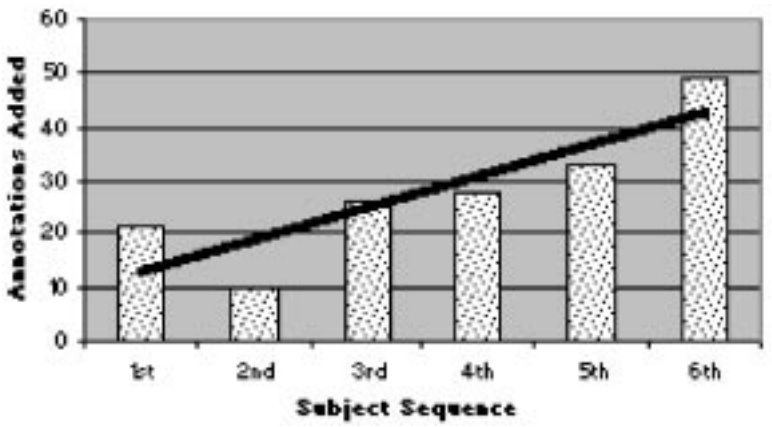

Fig. 7. Annotations added over time. The number of annotations added over time increased (Pearson $r = 0.491$, $p = 0.039$). Here, the numbers for all three conditions (Text-Only, Audio-Only, and Text + Audio) have been combined. A linear least-squares trend line is superimposed in black.

as the study progressed. We thought earlier subjects would make notes on the most interesting parts of the lecture video. Later subjects would have fewer opportunities for adding new and different annotations, and so would add fewer overall. In fact, something very different happened. We looked at total annotations added in sequence across all three conditions, and found that the number increased significantly (Pearson $r = 0.49$, $p = 0.04$). This is illustrated in Fig. 7. Furthermore, this increase in the total number of annotations arose from a significant increase in replies alone (Pearson $r = 0.52$, $p = 0.02$), suggesting that more sharing occurred as the study progressed. This suggests than an easy means for supporting moderation could be good to add to the system.

4.2.3. Positioning shared notes

Positioning annotations in this study was particularly important because the notes were shared. Subjects who participated later in the study relied on the positioning of earlier subjects' annotations to contextualize the annotation contents. Our hypothesis was that if annotations were not accurately positioned in the video, they would be confusing. We expected an 'accurately' positioned annotation to be placed directly over the place in the video that contextualized it.

As in the 'Personal Notes' study, users mostly positioned their annotations 10 to 15 s after the relevant part of the video. (This was not a conscious decision of the subjects, but just that they pressed the 'add' button after they had formulated a response to something in the video.) Contrary to our expectations, subsequent subjects were not confused or distracted by the lack of accuracy. When they watched the lecture video from beginning to end, most found it natural to see annotations display in the preview window a few seconds after the relevant part in the video.

4.2.4. User experience

Subjects generally liked using MRAS for taking shared notes on streaming video over the Web. As already noted, their feedback reinforced many of our observations, especially with respect to their choice of medium for creating annotations. Moreover, their feedback lends fresh insight into the process of creating and sharing Web-based annotations.

Asked whether they made more comments and questions using MRAS than they would have in a live lecture, 14 out of 18 said yes. However, this was due in part to the fact that they were watching a video that could be paused and replayed. One participant in the Text-Only condition told us he "...would not have had the opportunity to make as many comments or ask as many questions" in a live lecture. Another subject in the Audio-Only condition commented that she "...would not have had as much to say since... I'd be taking up everyone's time [and] 'hogging the stage'." Regardless of the reason, this is good news for asynchronous collaboration in education

contexts: Web-based support for making comments and asking questions asynchronously may increase class participation.

The fact that Web-based annotations can be shared also had a big impact on users. Asked whether the other annotations helped or distracted them, some reported the additional information confusing because it was not perfectly synchronized with the activity in the lecture video. The majority, however, found others' annotations useful in guiding their own thinking. This indicates that the positive results of [12] for annotation of text carry over to Web-based annotations. One Text + Audio subject said "it was thought-provoking to read the comments of others." Another remarked "it was more of an interactive experience than a normal lecture. I found myself reading the other notes and assimilating the lecture and comments together." These comments were echoed by subjects in the other conditions. An Audio-Only subject said that it was interesting and useful to see "...what other people find important enough to comment on." And a Text-Only participant pointed out that MRAS "...gives you the opportunity to see what others have thought, thus giving you the chance to prevent repetition or [to] thoughtfully add on to other's comments." Note-sharing in a Web-based annotation system could significantly enhance collaboration in an asynchronous environment.

Finally, subjects were asked whether they found using MRAS to be an effective way of preparing for a class discussion. A strong majority across all conditions agreed. The average response of subjects in the Text + Audio condition was 6.0 out of 7 (with 7 'strongly agree' and 1 'strongly disagree'). The average in both the Audio-Only and Text-Only conditions was 5.0 out of 7. One Text-Only subject was particularly enthusiastic: "I think that this software could really get more students involved, to see different viewpoints before classroom discussions, and could increase students' participation who would otherwise not get too involved."

5. Related work

Annotations for personal and collaborative use have been widely studied in several domains. Annotation systems have been built and studied in educational contexts. CoNotes [5] and Animal Landlord [18] support guided pedagogical annotation experiences. Both require preparation of annotation targets (for instance, by an instructor). MRAS can support a guided annotation model through annotation sets, but does not require it, and requires no special preparation of target media. The Classroom 2000 project [1] is centered on capturing all aspects of a live classroom experience, but in contrast to MRAS, facilities for asynchronous interaction are lacking. Studies of handwritten annotations in the educational sphere [12] have shown that annotations made in books are valuable to subsequent users. Deployment of MRAS-like systems will allow similar value to be added to video content.

The MRAS system architecture is related to several other designs. OSF [17] and NCSA [9] have proposed scalable Web-based architectures for sharing annotations on Web pages. These are similar in principle to MRAS, but neither supports fine-grained access control, annotation grouping, video annotations, or rich annotation positioning. Knowledge Weasel [10] is Web-based. It offers a common annotation record format, annotation grouping, and fine-grained annotation retrieval, but does not support access control and stores meta data in a distributed file system, not in a relational database as does MRAS. The ComMentor architecture [16] is similar to MRAS, but access control is weak and annotations of video are not supported.

Several projects have concentrated on composing multimedia documents from a base document and its annotations. These include Open Architecture Multimedia Documents [6], DynaText [19], Relativity Controller [7], Multivalent Annotations [15], DIANE [2], and MediaWeaver [22]. These systems either restrict target location and storage, or alter the original target in the process of annotation. MRAS does neither, yet it can still support document segmentation and composition with the rich positioning information it stores.

The use of text and audio as annotation media has been compared in [4] and [14]. These studies focused on using annotations for feedback rather than collaboration. Also, the annotations were taken on text documents using pen and paper or a tape recorder. Our studies targeted an instructional video and were conducted online. Finally, these studies re-

ported qualitative analyses of subjects' annotations. We concentrated on a quantitative comparison and have not yet completed a qualitative analysis of our subjects' notes.

Research at Xerox PARC has examined the use of software tools for 'salvaging' content from multimedia recordings [13]. This work concentrated primarily on individuals transcribing or translating multimedia content such as audio recordings into useable summaries and highlights. Our work focuses not only on empowering the individual to extract meaningful content from multimedia presentations, but also on providing tools for groups to share and discuss multimedia content in meaningful ways.

Considerable work on video annotation has focused on indexing video for video databases. Examples include Lee's hybrid approach [11], Marquee [21], VIRON [8], and VANE [3], and they run the gamut from fully manual to fully automated systems. In contrast to MRAS, they are not designed as collaborative tools for learning and communication.

6. Concluding remarks

The ability to annotate video over the Web provides a powerful means for supporting 'in-context' user notes and asynchronous collaboration. We have presented MRAS, a prototype client–server system for annotating video. MRAS interfaces to other server components (database, video, Web, e-mail), employs protocols that provide access across firewalls, and provides control over bandwidth use in annotation retrieval. MRAS supports asynchronous collaboration via annotation sharing, fine-grained access control to annotation sets, threaded replies, and an e-mail interface. Unique interface design issues arose, including annotation tracking while video is playing, seek capability, and annotation positioning.

We report very encouraging results from two studies of video annotation. In a study of personal note-taking, all subjects preferred MRAS over pen-and-paper (despite spending more time with MRAS). In a study of shared-note taking, 14 of 18 subjects said they made more comments and questions on the lecture than they would have in a live lecture. Precise positioning of annotations was less of an issue than we expected, and annotation tracking was heavily used by all subjects when annotations were shared.

Although these preliminary results are exciting, we are only scratching the surface of possibilities that a system like MRAS provides. We are now experimentally deploying MRAS in a number of workplace and educational environments to evaluate its effectiveness in larger-scale and more realistic collaborative communities. Users in these environments access video content on the Web and annotate it with questions, comments, and references. For these sessions, face-to-face time in meetings or class can be reduced or used for in-depth discussion, and users are able to organize their participation in the community more effectively.

Acknowledgements

Thanks to the Microsoft Usability Labs for use of their lab facilities. Mary Czerwinski assisted in our study designs. Scott Cottrille, Brian Christian, Nosa Omoigui, and Mike Morton contributed valuable suggestions for the architecture and implementation of MRAS.

References

[1] G. Abowd, C.G. Atkeson, A. Feinstein, C. Hmelo, R. Kooper, S. Long, N. Sawhney and M. Tani, Teaching and learning as multimedia authoring: the Classroom 2000 Project, in: Proc. of Multimedia '96, Boston, MA, Nov. 1996, pp. 187–198, 1996.

[2] S. Bessler, M. Hager, H. Benz, R. Mecklenburg and S. Fischer, DIANE: a multimedia annotation system, in : Proc. of ECMAST '97, Milan, May 1997.

[3] M. Carrer, L. Ligresti, G. Ahanger and T.D.C. Little, An annotation engine for supporting video database population, in: Multimedia Tools and Applications 5, Kluwer, Rotterdam, 1997, pp. 233–258.

[4] B.L. Chalfonte, R.S. Fish and R.E. Kraut, Expressive richness: a comparison of speech and text as media for revision, in: Proc. of CHI '91, pp. 21–26, 1991.

[5] Davis and Huttonlocker, CoNote System Overview, 1995, available at http://www.cs.cornell.edu/home/dph/annotation/annotations.html

[6] B.R. Gaines and M.L.G. Shaw, Open architecture multimedia documents, in: Proc. of Multimedia '93, Anaheim, CA, pp. 137–146, Aug. 1993.

[7] E.J. Gould, Relativity controller: reflecting user perspective in document spaces, in: Adjunct Proc. of INTERCHI '93, pp. 125–126, 1993.

[8] K.W. Kim, K.B. Kim, H.J. Kim, VIRON: an annotation-based video information retrieval system, in: Proc. of COMPSAC '96, Seoul, pp. 298–303, 1996.
[9] D. Laliberte and A. Braverman, A protocol for scalable group and public annotations, NCSA Technical Proposal, 1997, available at http://union.ncsa.uiuc.edu/~iberte/www/scalable-annotations.html
[10] D.T. Lawton and I.E. Smith, The Knowledge Weasel hypermedia annotation system, in: Proc. of HyperText '93, pp. 106–117, 1993.
[11] S.Y. Lee and H.M. Kao, Video indexing — an approach based on moving object and track, in: Proc. of the SPIE, Vol. 1908, pp. 25–36, 1993.
[12] C.C. Marshall, Toward an ecology of hypertext annotation, in: Proc. of HyperText '98, Pittsburgh, PA, pp. 40–48, 1998.
[13] T.P. Moran, L. Palen, S. Harrison, P. Chiu, D. Kimber, S. Minneman, W. van Melle and P. Zellweger, 'I'll get that off the audio': a case study of salvaging multimedia meeting records, Online Proc. of CHI '97, Atlanta, GA, March 1997, http://www.acm.org/sigchi/chi97/proceedings/paper/tpm.htm
[14] C.M. Neuwirth, R. Chandhok, D. Charney, P. Wojahn and L. Kim, Distributed collaborative writing: a comparison of spoken and written modalities for reviewing and revising documents, in: Proc. of CHI '94, Boston, MA, pp. 51–57, 1994.
[15] T.A. Phelps and R. Wilensky, Multivalent annotations, in: Proc. of the 1st European Conference on Research and Advanced Technology for Digital Libraries, Pisa, Sept 1997.
[16] M. Roscheisen, C. Mogensen and T. Winograd, Shared Web annotations as a platform for third-party value-added, information providers: architecture, protocols, and usage examples, Technical Report CSDTR/DLTR, 1997, Stanford University, available at http://www-diglib.stanford.edu/rmr/TR/TR.html
[17] M.A. Schickler, M.S. Mazer and C., Brooks, Pan-browser support for annotations and other meta information on the World Wide Web, in: Proc. of the Fifth International World Wide Web Conference, Paris, May 1996, available at http://www5conf.inria.fr/fich_html/papers/P15/Overview.html
[18] B.K. Smith and B.J. Reiser, What should a wildebeest say? Interactive nature films for high school classrooms, in: Proc. of ACM Multimedia '97, Seattle, WA, pp. 193–201, 1997.
[19] M. Smith, DynaText: an electronic publishing system, Computers and the Humanities 27, 415–420, 1993.
[20] Stanford Online, Masters in Electrical Engineering, http://scpd.stanford.edu/cee/telecom/onlinedegree.html
[21] K. Weber and A. Poon, Marquee: a tool for real-time video logging, in: Proc. of CHI '94, Boston, MA, 58–64, 1994.
[22] S.X. Wei, MediaWeaver — a distributed media authoring system for networked scholarly workspaces, Multimedia Tools and Applications 6, 97–111, 1998.

David Bargeron is a Research Software Design Engineer at Microsoft Research in the Collaboration and Education Group. Before joining Microsoft, from 1992 to 1994, Mr. Bargeron taught Mathematics in Gambia, West Africa. He received his B.A. in Mathematics and Philosophy from Boston College, Boston, MA, in 1992 and is currently completing his M.S. in Computer Science at the University of Washington, Seattle, WA.

Anoop Gupta is a Senior Researcher at Microsoft Research where he leads the Collaboration and Education Group. Prior to that, from 1987 to 1998, he was an Associate Professor of Computer Science and Electrical Engineering at Stanford University. Dr. Gupta received his B. Tech. from the Indian Institute of Technology, Delhi, India in 1980 and Ph.D. in Computer Science from Carnegie Mellon University, Pittsburgh, PA in 1986.

Jonathan Grudin is a Senior Researcher at Microsoft Research in the Collaboration and Education Group. He is Editor-in-Chief of *ACM Transactions on Computer–Human Interaction* and was Co-chair of the CSCW'98 Conference on Computer Supported Cooperative Work. He received his Ph.D. from the University of California, San Diego in Cognitive Psychology and has taught at Aarhus University, Keio University, University of Oslo, and University of California, Irvine, where he is Professor of Information and Computer Science.

Elizabeth Sanocki is currently a consultant in Experimental Psychology and Human Computer Interaction at Microsoft. Dr. Sanocki was a Senior Research Fellow at the University of Chicago and University of Washington before joining Microsoft as a consultant. She received her B.S. in Psychology from Michigan State University in 1986, and her Ph.D. in Experimental Psychology and Visual Perception from the University of Washington in 1994.