Video workflow in the learning sciences: Prospect of Emerging technologies for argmenting work practices

Roy D. Pea, Eric Hoffert

▶ **To cite this version:**

Roy D. Pea, Eric Hoffert. Video workflow in the learning sciences: Prospect of Emerging technologies for argmenting work practices. Ricki Goldman, Poy Pea, Bridge Barron and Sharon J. Derry. Video workflow in the learning sciences, Lawrence Erlbaum Associates, pp.427-460, 2007. hal-00190028

HAL Id: hal-00190028

https://telearn.hal.science/hal-00190028

Submitted on 23 Nov 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

27

Video Workflow in the Learning Sciences: Prospects of Emerging Technologies for Augmenting Work Practices

Roy Pea
Stanford University

Eric Hoffert
Versatility Software

VIDEO WORKFLOW AND PROCESSES

The aim of our chapter is to provide our readers with a comprehensive model of the stages of video workflow and their affiliated work practices, and with a road map of present and future technologies that support these practices. We situate the contributions provided by the chapters of this section of our volume within this workflow framework. Armed with this orienteering guide for video workflow, the reader should have a sense of the sociotechnical context of digital video and its affiliated technologies that they will be able to leverage today and anticipate in the years ahead. It is important to understand not only the epistemological and representational issues involving research video, in applications of video research on peer, family, and informal learning, and on classroom and teacher learning—but to recognize and to use productively the advances that are enabling and transforming video workflow practices for the work that we do as learning scientists.

We were inspired to sketch out this framework by a talk from Carl Rosendahl, Executive Producer of *Antz* and Founder of Pacific Data Images, the company that produced *Shrek*'s digital effects for the DreamWorks studio. He highlighted how computer software and networks are transforming every stage of the filmmaking process, from development to preproduction including storyboarding, production, and postproduction, including nonlinear editing, cinematography, visual effects, and distribution. We also foresee computer software and networks transforming every stage of the video workflow for the learning sciences, and we review later with illustrative examples how these transformations are beginning to surface in core technologies for each video workflow area. These advances, exciting as they are, are nonetheless at an early stage compared to the rapid development of digital video workflow in the film industry.

The chapters of this volume together help illuminate the extraordinary workflow complexity of video research in the learning sciences, a multistage and iterative process that, as Hay and Kim argue in their chapter (this volume), is beset with too much "friction" today—in which the researcher's needs and desires to do particular things with video such as share it or open it up to collaborative commentary are slowed down with the present state-of-the-art. And as Stevens highlights in this volume, capturing ideas in digital things and structuring learning around them using new video tools is a new version of a solution to the longstanding problem of inert knowledge in education.

Figure 27.1 provides a top-level view of the video workflow framework that transitions from video capture, to analysis, sharing, and collaboration. It begins in the upper lefthand corner with strategy and planning for video record capture, and moves quickly into the tactics of preproduction: Where, when, and how will you capture the video data that you seek? Our chapter does not treat these facets of workflow as they are addressed elsewhere in this volume. Then you are on-site, capturing video records with whatever devices suit your aims. Depending on the nature of your device, encoding of your video record may happen at the same time as capture—witness the advent of consumer digital video recorders that save video to computer hard disks that are part of the recorders. The video researcher then begins the processes of pulling these records into some kind of order, from the simple act of labeling them to easily find them later to the much more intricate activities that add the value of interpretation to these records. The researcher may chunk the video record into segments defined by event boundaries, time markers, or a variety of semiotic considerations. And marking video segments of interest, creating transcripts at different levels of detail, developing and using categories that the researcher considers useful for the aims of their research works in a recursive manner with both the deepening analysis of the video records and the never-ending tracking and finding of the rapidly growing population of data through searching and browsing. The researcher marks, transcribes, and categorizes a little, analyzes and reflects a little, needs to search and find a little, and so on, in the recursive loops that define such knowledge building activities (analogously to the writing process). In essence, there are close interdependencies between the activities of video record de-composition (e.g., segmenting, naming, coding) and re-composition (e.g., making case reports, collections of instances of commonly categorized phenomena, statistical comparisons of chunked episodes). Then the workflow moves on to presenting and sharing video analyses, in a variety of formats, and such sharing may be

formative as one collaboratively develops and/or comments on a developing video analysis, or a summative account as the video analysis is published (e.g., on the web or a DVD) and commented on by others in the community. To close the loop, the substantive insights from specific video research workflow activities have the prospects of influencing the next cycles of video research workflow in the field.

VIDEO TECHNOLOGIES

Video Capture, Standards, Storage, Input/Output, and Display

Video Capture: Formats for Inputs

Video input to a computer may be in a digital or analog format. Many "legacy" video sources are still in an analog form including a wide variety of VCRs, TVs, and a previous generation of analog video camcorders; these devices use NTSC (525 lines of resolution), PAL (625 lines), or SECAM standards—each adopted in different regions of the world (e.g., North America and Japan for NTSC; Europe for PAL; Eastern Europe for SECAM). The devices all require a process of analog to digital video conversion. Classic analog video formats include VHS (250 lines of resolution), S-VHS (400 lines), Hi8, Betacam, and BetaSP. Video data on tape is in YCrCb format (one luminance/brightness channel and two chrominance/color channels), and converted to RGB (three color channels, stored as 8 bits per pixel/channel) on a computer. There are a

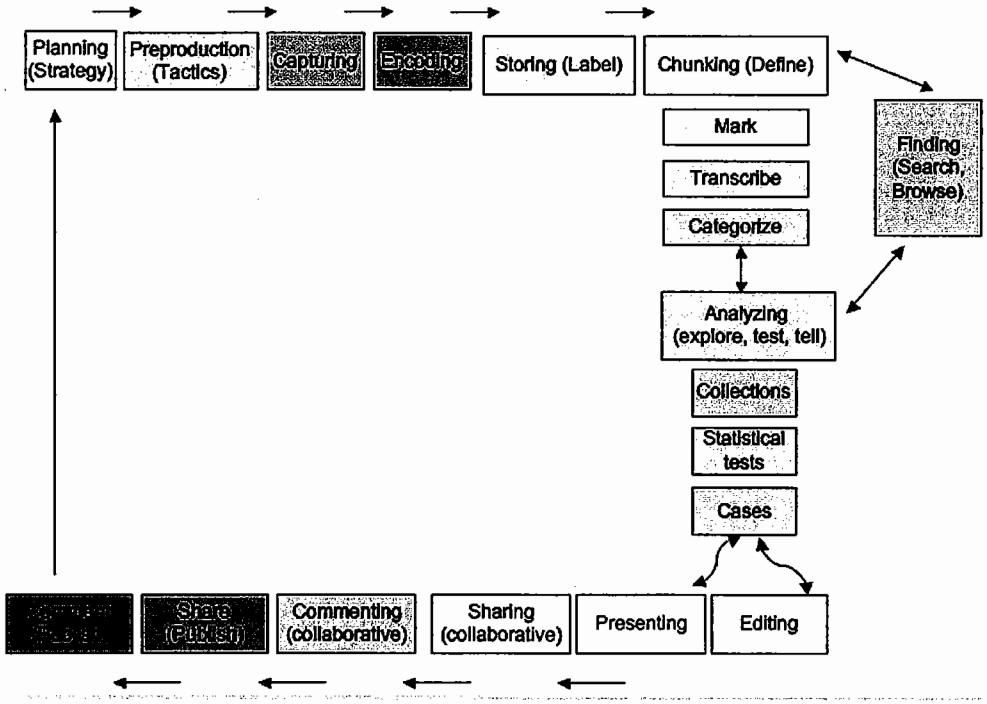

Figure 27.1. Diagram of video research workflow processes.

broad array of devices specialized in the analog to digital video conversion process. From lowest to highest quality, analog video signals span composite, S-video, and RGB component video. The serial control protocol RS-422 is often used to control a video source from a computer to allow computer-controlled video commands (i.e., rewind, fast-forward, play, jump to time code X).

Video imagery has an aspect ratio that is expressed as "x:y," where x represents its width and y its height. For digital video, pixels may be square or nonsquare; aspect ratios are typically 4:3 (1.33:1) for traditional video (the traditional TV screen, as well as IMAX), but newer widescreen formats include high definition, or HDTV (16:9, or 1.78:1), flat (1.85:1, American Theatrical Standard), and anamorphic scope (2.35:1). These widescreen video formats will gravitate from home theatre systems to the research laboratory. A proliferation of digital video formats includes D1, DV, DV25, DVCam, Digital8, DVCPro, and DVCPro50. Digital video device types include digital camcorders with video content recorded using mini-DV and other formats; direct-to-flash memory as in Nokia Series 60 3G phones; and direct to hard disk for other larger devices. PDAs, digital cameras, and cell phones now have integrated cameras with direct digital video input; video is captured using variants of digital video standards such as H.263, motion JPEG, or MPEG-4 (see later). These video-aware devices are exploding in popularity; 500 million camera phones were sold in 2005, and 63 million digital cameras were sold in 2004, with a projected 100 million digital cameras to be sold in 2008.

FireWire cables (using the IEEE 1394a standard) are rapidly becoming the prevalent high-bandwidth input mechanism that work across computer platforms to import video into a personal computer through serial bus ports. This standard is integral to a large number of popular digital camcorders for video I/O and runs at a rapid 400 megabits per sec. There are also FireWire 800 products, based on the IEEE 1394b version multimedia standard, that deliver speeds starting at 800 megabits/second, scalable to 3.2 gigabits/second. The new ultra high bit-rate standard also extends the distance that FireWire-equipped devices can send video and audio to more than 100 meters over CAT-5, plastic fiber, and other media.

The tools using these video inputs to the workflow process include digital camcorders, media enabled PDAs, and video-rich cell phones, web-cams, wireless video-on-IP (Internet Protocol) devices, and analog video cameras with video digitizers.

Video Standards

A great number of the revolutionary advances in digital video available to learning sciences researchers have developed due to world-wide MPEG standards (see http://www.mpeg.org for pointers), and because the kinds of functions available to a researcher are dependent on these standards and their evolution, we provide a brief account of them here. Whereas initially MPEG standards made for advances in video compression that were essential to reducing the costs of storing and transmitting video records, newer MPEG standards delve into the semantic content of videos, and enable new interactive capabilities with that video for authors and consumers that use such standards. In addition to the MPEG standards, several other important video standards

will be reviewed; 3GPP and 3GPP2 for mobile video, and SMIL—a World Wide Web consortium standard for describing multimedia presentations.

MPEG

MPEG (http://www.chiariglione.org/mpeg) is a series of broadly adopted worldwide audio and video coding standards. Major standards developed from MPEG include:

- MPEG-1: Coding for digital storage media (1992).
- MPEG-2: Coding for digital TV and DVD (1994).
- MPEG-4: Interactive Multimedia Audiovisual Objects (1998).
- MPEG-7: Content Description Interface (2001).
- MPEG-21: Multimedia Framework (2003).

These standards are reviewed later, with the table providing a quick summary of key aspects of the primary MPEG standards today, MPEG-1, MPEG-2, and MPEG-4.

MPEG-1 (1992): Coding for Digital Storage Media

MPEG-1 represents the first major standard from the MPEG family of video codecs, designed for coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s. The term "codec" combines the terms "compressor-decompressor" to characterize either hardware or software that can perform transformations on a data stream at both ends of its use in telecommunications. Codecs compress video for such purposes as storage, transmission, or encryption, and then decompress it for its uses including display, as in videoconferencing. In MPEG-1, each frame of video is decomposed into macroblocks—regions of 16 x 16 pixels. The macroblocks contain brightness and color samples called YUV, where U and V (chrominance information) are sampled at one quarter of the rate of Y (luminance information). Each 8 × 8 block of pixels is converted from a two-dimensional to a one-dimensional representation of 1 × 64 (in a zigzag sequence from the upper left to the lower right corner of the pixel block). A transform algorithm (discrete co-

TABLE 27.1
MPEG-x Characteristics Including Video and Audio Resolution, and Application Domain

Video Codec	*Video Resolution*	*Audio Resolution*	*Applications*
MPEG-1	352 × 240 for NTSC at 29.97 fps	224 to 384 kbps including MP3	CDROM, Internet, Video conferencing including MP3
MPEG-2	720 × 480 for NTSC at 29.97 fps	32 to 912 kbps including MP3	DVD, Digital-TV, HDTV
MPEG-4	Variable	Variable including MP3, AAC	Cell phone, PDA, PC, variable rate Internet delivery

sine transform / DCT) converts the content from the spatial to the frequency domain as a series of numeric values, where the result is further coded, with frequent values replaced by short codes and infrequent values replaced by longer codes.

"Motion compensation" is an algorithmic technique common across the MPEG-x standards, and used to predict the movement of pixel blocks (macroblocks) from frame to frame; the prediction error for a macroblock is then stored and quantized and is typically smaller than storage of the actual pixel values, increasing the compression rate further. Motion compensation works by searching for "matching" macroblocks in adjacent video frames. Compressed sequences of MPEG video are constructed from groups of pictures (GOP). There are three classes of frame types in the MPEG standard that comprise a GOP. These are:

- I-Frames (intra-coded frames) compress a single frame of video without any reference to other frames in the video sequence. For random access in an MPEG video sequence, decoding starts from an I-frame. I-frames are included every 12 to 15 frames. These frames are also used for fast forward and reverse.

- P-Frames (Predicted frames) are coded as differentials from a prior I-Frame or a prior P-Frame. The prior P-or I-frame is used to predict the values of each new pixel in order to create a new predicted P-frame. P-frames provide a compression ratio superior to the I-Frames although this is a function of the degree of motion; small amounts of motion produce better compression for P-frames.

- B-Frames (Bi-directional frames) are coded as differentials from the prior or next I-or P-frame. B-frames use prediction similar to P-frames but for each block in the image, the prior P-frame or prior I-frame is used or the next P-frame or the next I-frame is used. Because the encoder can select which I-frame or P-frame to select, the encoder can select the bi-directionally predicted frame that produces the highest possible level of compression.

A coding sequence for MPEG-1 for NTSC video is I B B P B B P B B P B B P B B I, where I-Frames may be spaced 15 video frames apart, and two B-Frames precede each P-Frame (or I-frame). MPEG-1 standards vary for NTSC video (National Television System Committee, a 525-line/60 Hz, 30 fps system, principally used in the USA/Japan) and for PAL video (phase alternation by line, a 625-line/50 Hz, 25 fps system, used principally in Europe). For PAL video, the sequence is I B B P B B P B B P B B I and is typically 12 frames long. MPEG-1 resolution is 352 × 240 for NTSC at 29.97 fps and 352 × 288 for PAL/SECAM at 25 fps. Audio bit rates are typically 224 kbps for MPEG-1 layer II audio where 384 kbps is the typical rate utilized. MPEG-1 layer III audio coding represents the well-known MP3 audio standard, with typical rates of 128kbps and 192kbps.

Importantly for research, because MPEG-1 combines intraframe and interframe encoding, the co-dependence of certain frames makes this codec inappropriate for editing and other image postproduction applications. MPEG-1 displays progressive scan images, noninterlaced frames that cannot be used for broadcast, but it can achieve three times or more the compression factors of JPEG. It is good for playback only on applications such as games, distribution, publishing, VCD, and CD-ROM although it is occasionally used for desktop-based rough cut editing applications.

MPEG-2 (1994): Generic Coding for Digital TV and DVD (Moving Pictures and Associated Audio Information)

The MPEG-2 specification was designed for broadcast television using interlaced images. It provides superior picture quality compared to MPEG-1 with a higher data rate. At lower bit rates, MPEG-1 has the advantage over MPEG-2. At bit rates greater than about 4 Mbits/s, MPEG-2 is recommended over MPEG-1. MPEG-2 includes support for high quality audio and full surround sound with 5.1 channels, representing left, center, right front, right rear, and left rear audio channels. The audio can be extended to 7.1 with left center and right center channels. Audio bit rates range from 32 kbps up to 912 kbps where 384 kbps is the typical rate utilized and the sampling rate is fixed at 48 kHz.

MPEG-2 supports variable video bit rate and broadcast applications; MPEG-2 tends to be encoded at 6 to 8 Mb/s fixed data rate. For high-end production, typically the highest bit rates are used, such as 50 Mbps. This is called master quality MPEG-2 video encoding. Component ITU-R 601 format video running at 270 Mbits/sec will run at 2–50 Mbits/sec when transcoded into MPEG-2. MPEG-2 can also support both 4:3 and 16:9 image aspect ratios. MPEG 2 is used for DVD, digital TV, and HDTV.

MPEG-2 uses a group of pictures (GOP) at 12 (PAL) or 15 (NTSC) frames in length where each frame is constructed of two interlaced fields. A coding sequence for MPEG-2 for NTSC video is I P B P B P B P B P B P B P B I, where I-Frames may be spaced 15 video frames apart, and a P-Frame precedes each B-Frame. For PAL video, the sequence is I P B P B P B P B P B P I and is typically 12 frames long. MPEG-2 resolution is 720 × 480 for NTSC at 29.97 fps and 720 × 576 for PAL/SECAM at 25 fps.

Like MPEG-1, the I-frames in MPEG-2 are encoded independently and are the only independent frames in an MPEG-2 sequence. Only the I-frames can be edited when working with MPEG-2. MPEG-2 has been proven to be a good video standard to handle the use of transcripts along with standard (and noisy) classroom interactions. MPEG-2 consisting only of I-frames at high bit rates is often used for video editing and/or production applications due to its high picture quality and flexible random access support. In the case of MPEG-2 where only I-frames are used, production quality MPEG2 at 50 Mbps is also referred to as IMX; this format is frequently utilized with equipment such as AVID editing stations and storage subsystems.

MPEG-4 (1998): Coding of Interactive Multimedia Audiovisual Objects

MPEG-4 resulted from a new international effort incorporating and extending MPEG-1 and MPEG-2 and involving hundreds of researchers and engineers. MPEG-4 builds on three fields; digital television; interactive graphics applications (synthetic content); and interactive multimedia (distribution of and access to content on the Web). MPEG-4 provides the standardized technological elements for integrating of the production, distribution, and content access paradigms for the three fields.

Unlike its predecessors, MPEG-4 is an object-based video standard. Audiovisual scenes can be composed of objects, where a compositor within a decoder places video

objects into a scene using the optimal encoding process for each object. An objective is to go beyond the typical start/stop/rewind/fast-forward level of interaction common to video content; with MPEG-4 the objective is to allow for interactivity with video objects directly embedded within a scene. Relevant to computer graphics practitioners, the standard is targeted for the combination of natural and synthetic objects in a scene. Audiovisual objects can include 2D/3D computer graphics, natural video, synthetic speech, text, synthetic audio, images, and textures. MPEG-4 streaming delivers the same quality video streaming as MPEG-2, the current industry standard, but MPEG-4 uses only one third of the MPEG-2 bit rate. This bit rate reduction at the same quality level is quite substantial and yields significant speedups in transmission time. MPEG-4 video provides very high quality across the bandwidth spectrum—from cell phones up to high bit rate broadband—that rivals the best proprietary compression algorithms available today. MPEG-4 was designed to be a scalable Codec that could support a broad array of delivery devices (PDA, PC, Set-top box, etc.) and it has delivered on that promise.

At the core of the MPEG-4 standard is the audio codec—AAC (Advanced Audio Codec). AAC offers support for multichannel audio, up to 48 channels; high resolution audio with sampling rates up to 96 KHz; decoding efficiency for faster and more efficient decoding; and compression with smaller file sizes. Multilingual support is also provided. AAC is used for audio coding at 32 kbps per channel and higher. The standard is targeted for audio coding in 3G wireless phone handsets and is used in the Apple iTunes Music Store. Apple Computer strongly supports MPEG-4 (Apple QuickTime7/MPEG, 2005). MPEG-4 is an integral element of QuickTime 7 (and beyond) and Real Networks has adopted the standard as well. However, and in notable fashion, Microsoft has yet to embrace the standard and has provided an alternative scheme in Windows Media 9 and 10. Most recently, a flavor of MPEG-4 referred to as MPEG-4 Part 10, which is also known as H.264, is rapidly coming into place as a favored standard for high-quality video compression. H.264 is being used to store video as a "broadband master" at bit rates from 3 to 6 Mbps from which the content can be further transcoded into a variety of lower bit rates for broadband distribution. Only the fastest PC and Macintosh computers can decompress H.264 at acceptable playback speeds and resolutions; as a result, this nascent format is expected to take an extended time period to come into widespread consumer usage.

3GPP and 3GPP2

Launched in 2003 as consumer services, 3GPP (the Third Generation Partnership Project) defines Mobile Video Codecs, with capability to download or stream video for mobile media devices, and often to capture video as well. The similar 3GPP and 3GPP2 are based on Mobile Video and Audio Codec Standards and are primarily targeted for ultra-low bandwidth downloadable and streaming video for cell phones and mobile devices (3GPP works for GSM networks-Global System for Mobile Communication; 3GPP2 for CDMA networks-Code Division Multiple Access). Variants of key standards are used such as H.263 and MPEG-4, with video download rates defined at 64 kbps. Streaming rates range from 25–45 kbps with AMR audio spanning 4–12 kbps and AMR-WB spanning 6–25 kbps. Image resolution and frame rates include Sub QCIF (128 × 96) and QCIF

(176 × 144 for PAL, 176 × 120 for NTSC) using 7.5, 10, and 15 frames per sec. Transport Mechanisms are designated as GPRS (General Packet Radio Service) for Internet access, WAP (wireless access protocol), MMS (multimedia messaging service) for e-mail exchange, and MMC (multimedia memory card) for memory to PC synchronization. This mobile multimedia standard resolution is 128 × 96 at 15 fps but may scale higher. Stanford's DIVER Project has been experimenting with Nokia cell phone short video capture (e.g., one min clips). A movie thus captured is sent to the DIVER software web server as an MMS e-mail attachment, and is then transcoded into Flash video format for research analyses, commentary, and remixing over standard web browsers using DIVER (see later in this chapter). As cell phone video cameras increase in resolution and storage media on phones allow capture of longer movies, this approach could enable a flexible and ever-present component of video research technology.

MPEG-7 (2001): Multimedia Content Description Interface. MPEG-7 is a standard focused on video and rich content metadata. The metadata for video includes semantic characterizations of video and interactivity. Once powerful mechanisms for video object detection and segmentation are in place and validated as a reliable capability, MPEG-7 can support these advanced functions with an end-user ability to edit out objects, people, and scenes. The main elements of the MPEG-7 standard include: (a) description tools, or descriptors (D), that define syntax and semantics of each feature (metadata element); and description schemes (DS), that specify structure and semantics of the relationships between their components, (b) a description definition language (DDL) that is used to define the syntax of the MPEG-7 description tools and allow creation of new description schemes, and (c) system tools, used to support binary representation for efficient storage and transmission, transmission mechanisms (both for text and binary formats), multiplexing of descriptions, synchronization of descriptions with content, and management and protection of intellectual property in MPEG-7 descriptions. MPEG-7 descriptions of content may include information on:

- Creation and production of the content.
  - Director, title, or short feature movie.
- Usage of the content.
  - Copyright pointers, usage history, and broadcast schedule.
- Storage features of the content.
  - Storage format, encoding.
- Spatial, temporal, or spatiotemporal components of the content.
  - Scene cuts, segmentation in regions, region motion tracking.
- Low level features in the content.
  - Colors, textures, sound timbres, and melody description.
- Reality captured by the content.
  - Objects and events, interactions among objects.
- How to browse content in an efficient way.
  - Summaries, variations, spatial, and frequency sub bands.
- Collections of objects.

- The interaction of the user with the content.
  - User preferences and usage history.

MPEG-21 (2003): Multimedia Framework

The major aim of MPEG-21 has been to establish a transparent multimedia framework for all ways in which one user interacts with another user, and the object of that interaction is a fundamental unit of distribution and transaction called the "digital item" or "resource" (where "user" has the technical sense of any entity that interacts in the MPEG-21 environment or makes use of a digital item). Digital items are the "whats" and users are the "whos" of the MPEG-21 framework. The standard defines a "resource" as an individually identifiable asset, such as video clip, audio clip, an image, and text. Interactions concerning resources include creation, production, provision, delivering, modification, archiving, rating, aggregating, syndicating, retail selling, consuming, subscribing, and facilitating as well as regulating transactions that occur from any of such kinds of interactions. The goal of MPEG-21 has been characterized as defining the technology needed to support users to access, consume, trade, and otherwise manipulate digital items in efficient, transparent, and interoperable ways. So for example, MPEG-21 includes an XML-based standard "rights expression language" for sharing digital rights, restrictions, and permissions for digital resources between creators and consumers, and for communicating ubiquitous and secure machine-readable license information (Wang, 2004).

Video Interaction: SMIL

SMIL (synchronized multimedia interaction language) is a Web Consortium standard for describing multimedia presentations (http://www.w3.org/AudioVideo). SMIL can be used to create time sequential and time parallel composited layers of image, text, and video within a single, synchronized multimedia presentation. Graphical regions of the screen are defined and temporal events can be mapped into the graphical regions. SMIL is compatible with both QuickTime (QuickTime/SMIL, 2005) and RealMedia. SMIL is based on an XML representation and allows for the integration of distributed web resources into a unified end-user experience. Examples of SMIL usage include starting one video clip after another video clip completes, or triggering a demographic trend graphic to appear beside a video news clip. In addition, a completely new user experience—such as launching a new browser window with a new user input form—can be triggered from a user mouse click in a particular graphical region or visual icon.

SMIL data files are typically comprised of links, media content, spatial and temporal layouts, semantic annotations, and alternative content (for varying bandwidths, tasks, and user characteristics). SMIL uses the concept of "layout adaptation," where SMIL documents can adapt to browsers and/or playback devices with different characteristics such as screen sizes, bit depths, language characteristics, and so forth. Adaptation can be based on environment, user, and purpose. Selected SMIL "dialects" may also be skipped using a "skip-content" flag. For example, a full color video could be

represented in black and white on a monochrome cell phone display; a text region could be shown in French rather than in English based on the location of the end-user. SMIL is rich in hierarchy—regions can be hierarchical spatially—and time-based constructs can be nested such as parallel and sequential time-based media playback (also called "temporal hierarchy"). SMIL can be adapted to the needs of the education community because of its flexibility, features for rich media and interactivity, and ability to support curriculum tool building and delivery.

Video Storage and Archives

Video storage of sufficient scale and reliability to deliver rich media to multiple users is a key requirement. Video storage capability is rising dramatically (from GB to TB) while costs are falling quickly (Langberg, 2004). In the 50 years since IBM invented computer hard disk storage, the density of information that can be recorded per square inch has increased 50 million times, from 2K bits to 100 Gigabits (Walter, 2005), with ultra-high densities achieved of 50 terabits per square inch with SeaGate's labs (McDaniel, 2005). A gigabyte of storage today costs on the order of $1 (Gilheany, 2004; Napier, 2006), with terabyte storage for $1,000; this is remarkable when contrasted with storage pricing in the year 2000, when a terabyte of industrial grade storage might cost as much as $1,000,000. Yet, there is a direct correlation between the cost of the storage and its inherent reliability. Storage with ultra high levels of redundancy and reliability can be costly, usually 10–100 times more than standard SCSI storage on a PC.

Video storage systems, when used for ongoing production and archiving, may include any of the following storage approaches; *online, nearline*, and *offline*. The access time and amount of storage utilized for each "tier" of storage increases as one progressively transitions from online to offline. Likewise, the relative expense of storage declines as one moves from online to offline. Video storage often follows a scheme similar to traditional hierarchical storage management (HSM; e.g., see Front Porch Digital, 2002; IDC, 2005).

Online Storage. Online storage is the fastest storage media and is used for all production-level work. Online provides near instantaneous access to video material and content. The access time for online storage is on the order of 10 to 15 ms measured as the time it actually takes for the disk read/write head to locate a data sector on the disk drive. Online content typically ranges from gigabytes to terabytes. Data transfer rates can range from 10 to 1,000 Mbytes/sec or more, with higher speeds using special disk arrays.

Nearline Archiving (Hierarchical Archiving). With nearline archiving, an archiving mechanism can be used for content that has not been requested for an extended period of time. Under this scenario, file data is normally stored on a server so that it can be accessed quickly as needed. When a particular event occurs, as when files are not accessed for a specified period of time, a nearline archiving system automatically transfers files to an external removable tape device, providing additional disk

space for online work. When it is necessary to open a file whose data has been transferred to a nearline archive, the data is automatically recalled from remote storage. Nearline automated tape libraries are the primary mechanism used for archival and short-term storage. Access time for nearline archives is on the order of 100s of ms or single digit seconds. Storage cost is on the order of 10¢ per GByte or less (Gray, 2004). Data transfer rates range from 10 to 250 Mbytes/sec or more. Nearline content typically ranges from gigabytes to terabytes.

Offline Archiving. With an offline archiving approach, the selected content will be moved to offline storage when it is no longer required and the content will not be available without manual content restoration. Offline storage is based on tape. Tape devices access data in sequence; this means that accessing files from tape devices can require significant time even if the tape is already on site and loaded in the tape drive. When storage costs drop, the need for offline and nearline archives is reduced. Offline archives are typically on high-density tape; for example DLT tape contains 600 GBytes of data on a single tape and it is rated to last 30 years. Access time for offline archives is on the order of 10s or 100s of seconds or more. Storage cost is on the order of 1¢ per GByte or less. Offline content can range from petabytes to exabytes, particularly for organizations with very large-scale data backup and retention requirements over large time scales. Data transfer rates range from 1 to 100 Mbytes/sec or more. Holographic memory is a new optical media contribution to superior low-cost offline archiving. In late 2006, Maxell is releasing new optical storage media with 3-D holographic recording technology so that a single 5¼" diameter optical disc has a 1.6 terabyte capacity, offering a 50-year media archive life and random data access with data rates as high as 120 mbytes/sec.

Tape archival libraries provide affordable mass storage, which often handles tape library management. An archive manager is a middleware software solution serving as an abstraction layer and bridge between an online storage system and an automated tape library's tape drive and robotics mechanisms for physical tape movements to and from use. Archive managers must be compatible with automation systems (e.g., Sundance, Encode, Probus), content management systems (e.g., Documentum, Artesia), and video servers (i.e., QuickTime, Real Networks, Quantel, Avid). Archive managers hide the complexity of interfacing to disparate tape library systems, which tend to have complex logic and proprietary interfaces. Archive operations require physical retrieval of content from remote devices, mounting of tapes and extracting of video from the tape back to an online storage facility; video files may be distributed across multiple tapes. Video is also routed to the correct video server when more than one server is utilized. Key archive manage operations include "archive," "delete," and "restore."

Online storage, especially the type that is used on high performance media servers, will often include support for RAID (Redundant Array of Independent/Inexpensive Disks), a category of disk drives that employ two or more drives in combination for fault tolerance and performance. RAID supports six tiered levels and can handle data striping (where data resides across multiple disks for faster access), and data mirroring (where data resides on multiple disks for fault tolerance). Storage Area Networks

(SANs) are becoming more widespread now, especially in networked or hosted environments. A SAN (see SAN, 2005) is a high-speed subnetwork of shared storage devices—machines that contain only disks for storing data. SAN architecture enables all storage devices to be accessible to servers on a local or wide-area network. Because data stored on disk does not reside on network servers, server power is utilized for applications only and disk servers handle data access only. Alternative and lower cost solutions are available including newer systems such as Mirra, where a specialized multimedia/data/backup server is connected to a PC and allows for external Internet access to stored content. Such storage appliance devices cost only a few hundred dollars but allow for automated backup of content, integrated with web-friendly access for storage sharing.

As storage costs drop substantially and as interest rises in the archiving of video material, a number of major video archives are rising in prominence. Notable among public archives are the Internet Moving Image Archive (Internet Archive, 2005, http://www.archive.org/movies/movies.php) and the Shoah Visual History Foundation (Shoah, 2005). The Internet Moving Image Archive is a repository of video that is searchable, indexed, and available to the public as part of the Internet Archive Project, which contains 400 terabytes of indexed content, growing at the rate of 12 terabytes per month. The video archives provide a collaborative environment (user votes, ratings, most popular videos, most popular categories, and number of downloads) and multiformat delivery including MPEG-1, MPEG-2, and MPEG-4 (standard playback and editable). Internet Archive research is underway to develop the "Petabox," a petabyte archive and processing matrix using 800 PCs to process the archived data. The Shoah Visual History Foundation has captured 120,000 hours of video testimony from holocaust survivors. This content is stored on a 400 terabyte digital library system where robotics are used to retrieve the appropriate tape matching user requests for video access; a high-speed fiber optic network connects a number of major universities to the archive and can deliver full resolution video to web browsers at these institutions. One could imagine these concepts extended to the learning sciences—for example, an archive could be developed with hundreds of thousands of hours of video captured from teachers and classroom interactions providing a valuable resource for the education community, subject to appropriate human subjects approvals from institutional research boards (IRBs) required by federally funded research in the United States.

Video Clients and Servers

Video clients are ubiquitous as digital media players, and are now used widely across the Internet for multimedia delivery. The three primary client players are Microsoft Windows Media, Real Networks, and QuickTime. These players are dominant on the Mac and PC platforms. While technically not a client video player, Macromedia Flash format streaming video files are growing in usage because Flash is installed in over 98% of the world's computers and plays through web browsers. Recent projections indicate that streaming Flash video could grow to a market share of as much as 35% or more in the next 5 years. Linux and SUN Microsystem's Solaris operating systems are not as directly connected to the media player ecosystem although there

are options available. Recent co-operations announced between SUN Microsystems and Microsoft could lead to advances for media players across the Solaris/Linux and Microsoft worlds—this would be helpful for greater standardization of media delivery, but only time will tell. The installed base of QuickTime players is now several hundred million worldwide. The media players include support for graceful degradation where CPU power, networked bandwidth limitations, or other issues that reduce throughput are adaptively handled by the players, which will reduce frame rate, picture resolution, or audio sampling rate if and where possible. Although the concept of scalable video has been highlighted for many years—where video content can be adaptively modified in real-time based on constraints of bandwidth and CPU power, alternative concepts are used for desktop delivery such as graceful degradation and multitrack reference movies (discussed later). Many video capture systems are now configured, through the use of multiple video input cards, so that video can be digitized and transcoded (converted into new media formats) in parallel into multiple downloadable and streaming media formats. Once the video content has been encoded into any of the key media formats, it can then be played back using a variety of the video server platforms. A subset of video servers is described in the next section.

Video servers store and deliver streaming and downloadable media. Streaming allows for random access at the client, no placement of the media file on the user's local hard drive, and supports delivery of live media streams. Storage for video may also be provided as a service on the Internet. Servers may be open source or proprietary. Representative video server platforms include: (a) the Real Networks Helix Server, the first major open-source streaming media server (Helix Server, 2006), which supports a large variety of video codecs (i.e., QuickTime, MPEG-2, MPEG-4, Windows Media, Real Media, etc.) as well as provides access to an open source code base for enhancing and extending the media server itself. This server is particularly useful when developing new streaming media algorithms and protocols; it can be used to build customized encoders and players for solutions requiring new codecs. A Helix web site provides source code, documentation, and useful reference information; and (b) the QuickTime Streaming Server is server technology for sending streaming QuickTime content to clients across the Internet using the industry standard RTP (Real-Time Transport Protocol) and RTSP (Real-Time Streaming Protocol) protocols. The streaming server has a number of key features including skip protection, which uses excess bandwidth to buffer ahead data faster than real time on the client machine. When packets are lost, communication between client and server results in retransmission of only the lost packets (not all of the data in a block containing lost packets, as is typically used), reducing impact to network traffic. ISO (International Standards Organization) compliant MPEG-4 files can be delivered to any ISO-compliant MPEG-4 client, including any MPEG-4 enabled device that supports playback of MPEG-4 streams over IP. QuickTime supports the concept of reference movies that allow for storing tracks of movies at alternate data rates and delivery methods. For example, you can store movies at data rates of 56 Kbits/sec, 384 kbits/sec, and 1.5 Mbits/sec all in the same QuickTime movie file and select the appropriate movie depending on connection speed. Similarly, it is feasible to include both HTTP FastStart

and Streaming media versions of the same content as separate tracks within the same QuickTime movie to allow selection based on the connection scenario for a user.

QuickTime FastStart is a delivery mechanism for QuickTime Movies that works with web browsers over HTTP, that is, after a preset amount of content has been delivered to a client over HTTP, the movie will begin to playback. Concurrent with playback, the HTTP process of downloading proceeds in parallel. A movie can thus play while the next set of content is being downloaded. The process emulates real streaming but is only an approximation. The ubiquity of HTTP, thanks to browsers, makes this a real advantage as no special server side or streaming server software is required and there are rarely any issues with traversing corporate or university firewalls or network translation tables. However, live content cannot be provided, as one must wait until an entire movie is downloaded before true random access is available, and a copy of the content is placed locally on a user's hard drive, which the content provider (commercial or researcher) may not desire.

Darwin is a related platform for the QuickTime Streaming Server. A key metric for streaming servers is the number of concurrent streams that may be delivered by a server. With Darwin streaming server, up to 4,000 simultaneous streams can be served from a single server, and resources can be scaled up to meet increased traffic by adding multiple servers. Darwin is based on the same code as Apple's QuickTime streaming server product (available for Mac OS X) and is available as open source, with support for Solaris, Windows NT/2000, Linux, and an ability to be ported to additional platforms (http://developer.apple.com/opensource/server/streaming).

Video Resolution, I/O, and Display

Megapixel image resolution and frames per second (fps) parameters continue to improve as per unit cost drops rapidly. Many digital camcorders provide much better imagery in using 520 lines of horizontal resolution versus 240 lines with VHS analog camcorders. Hi-resolution formats are becoming increasingly important due to the widespread proliferation of HDTV compatible displays and plasma screens. Based on high demand, costs are expected to drop rapidly over the next few years for HD (High-Definition) technology. Key standards for high-resolution video (SGI HDTV, 2005) include 1920 × 1080 @ 60i (an analog video interlaced display at 60 fps, i.e., SMPTE 274M), and 1280 × 720 @ 60p (a digital video progressive scan display at 60 fps, i.e., SMPTE 296M). Frame rates per second for HDTV content can include, for progressive scan, 24, 25, 30, 50, or 60 fps, and for interleaved scan, 50 or 60 fps. An important consideration for the next few years in the transition to more pervasive HDTV is the adverse effects of displaying conventional 4:3 aspect ratio interlaced video on HD displays. A standard cable feed, or video footage shot on DVD can be stretched, distorted, or aliased in appearance. This is unfortunate, because traditional analog TV displays will look better when displaying this content than would more expensive HDTV displays. Addressing these issues will be a slow process; making 4:3 aspect ratio video higher quality in appearance on 16:9 displays, and waiting for more content to be produced as original in widescreen format.

New high-resolution production systems are emerging from vendors such as Silicon Graphics, Apple, and Discreet; these environments include an ability to capture, compress, store, and manage HDTV data streams.

The Japanese broadcaster NHK has developed a possible successor to HDTV that uses the same 16:9 wide screen aspect ratio, Ultra High Definition Video (UHDV), but with an immersive field of view that is four times as wide and four times as high at HDTV, yielding a picture size of 7680 × 4320 pixels. UHDV also refreshes 60 frames per second, twice conventional video. Because HDTV took 40 years from its development as standard in 1964 to its consumer growth today, UHDV may be a long time coming.

To move from the large screen to the palm-size video display, we note that in 2003, the standard camera phone resolution was 300 kilopixels. By 2005, 2-megapixel camera phones are commonplace from Nokia and other companies, with even 7- to 8-megapixel camera phones available from Samsung. The next horizon will come from much higher resolution and much smaller form factor cameras using CMOS rather than CCD technology. CMOS (pronounced "see-moss") stands for complementary metal oxide semiconductor, and CMOS integrated circuits are very low in power consumption and heat production and thus allow for very dense packing of logic functions on a chip, resulting in greater, cheaper video functionality in a smaller package.

Video input and output capabilities span a broad range of bandwidths and form factors. Transfer mechanisms can be analog or digital. Video can be readily transferred to and from cell phones, cameras, PDAs, TVs, VTRs, and HDTV systems. For video I/O, one can transfer directly to computer from capture devices using USB or FireWire cabling, or by removable storage media like Compact Flash, Memory Stick Pro, Smart Media, Secure Digital, XD Photo Card, or tiny CD-R discs. In 2005, tape-free camcorders emerged with a tiny removable 4-gigabyte MicroDrive hard disk.

The primary worldwide standard for digital video is ITU 601, with video at a resolution of 720 × 480 pixels (SMPTE 601 @ 270 Mbps via serial digital interface, or SDI). HDTV video is typically managed using fiber-channel disk arrays with digital video transfer via the SMPTE 274M and 296M standards. However, certain classes of systems (Sony HDTV, 2005) allow HDTV signals to be encoded using the pervasive 601 standard so that they can be easily imported and then subsequently manipulated back in the HDTV format region.

The area of handheld and mobile devices continues to advance at a dramatic pace, with new models of handhelds and cell phones offering color screens, higher memory, increased network bandwidth (via WiFi and 3G) and enhanced removable storage. It is logical to consider the use of these devices as a platform on which to distribute and display rich media. For example, data storage cards such as CompactFlash, SD Memory, Memory Stick, MemPlug, and others, offer storage on the order of a gigabyte or more. This level of storage is well suited to handling compressed digital video files with duration of one hour or more. Content authors can create and store standard 4:3 or panoramic video content on this new class of data storage for mobile devices.

Kinoma (2005) has provided a strong solution for displaying high-quality digital video on handhelds. Kinoma Producer provides an authoring environment for PCs or Macs that enables the user to convert a movie into a specialized media format suitable for playback and interaction on a handheld device, using Kinoma Player software.

Kinoma video is compatible for playback on the many handhelds and cell phones running PalmOS. The Kinoma video codec supports full screen, full motion, full color, high-resolution video for Palm Powered handhelds plus VR objects, VR panoramas, animation, and still images with synchronized audio. For output, Kinoma can generate video, audio, still images, and with proper setup, can transform a PowerPoint presentation into a format for handheld display. In 2005, Apple introduced a video capable iPod which provides similar capabilities as Kinoma but for a much broader audience on the iPod, where video is more of an integral part of the product design; the video iPod is based on the MPEG-4 standard described earlier. Apple has also integrated its digital rights management technology (FairPlay) into the MPEG-4 video content used on the iPod to manage distribution of protected intellectual property including content that is for sale.

As Fishman (this volume) notes, multimedia handheld PCs and smart phones have the potential to dramatically improve teachers' ability to access multimedia records for their uses in professional development, or for making notes during instruction that can be synchronized with subsequent reflections on video for their practices, or to anchor mentoring dialogues.

Media phones, PCs, PDAs, TVs, and HDTVs can all serve now as video display devices. Users must be able to author in a multimedia multidevice world. Display technology is advancing rapidly and future advances must be anticipated now. High-resolution display can be crucial for "seeing" for analysis on larger displays like HDTV (e.g., Microsoft Windows Media 9 support for HDTV; development of Mark Cuban's HDNet HDTV channels and broadcasting) and already 13 million U.S. HDTVs compatible monitors were in the United States at the end of 2004, with projections of 74 million by 2010 (Chanko. Wiqker, & Scevak, 2005). The use of non-PC and non-TV platforms can provide nondesktop opportunities for reviewing and analyzing videos, for example, cell phones (with 320 × 240 pixel displays).

Video Editing, Indexing, and Analysis

Video Editing

Nonlinear editing is a key method to identifying and prioritizing video streams to produce final output. Nonlinear editing tools are now mainstream and available in simpler form factors than ever before, and without data loss during digital editing and copying, unlike analog videotape copying. Computer-based, nonlinear editing systems have radically changed the editing paradigm and have become the standard tools in both the film and the video industry (Hoffert & Waite, 2003), so much so that all the major television stations are dismantling their linear video editing suites and many of the youngest generation of editors have never edited linear video.

Nonlinear editing systems range from high-end, professional systems (such as the Avid Media Composer and Film Composer) to industrial grade or "prosumer" systems (such as Adobe Premiere and Apple Final Cut Pro) to the most basic consumer variants (such as Apple's iMovie or QuickTime Pro, or Pinnacle's Studio MediaSuite). All editing tools share random-access capabilities for retrieving digitized video and

sound material and most utilize the concept of the time line as a working tool. In various systems, a low-resolution and more highly compressed digital video proxy of the broadcast grade content is utilized to speed the nonlinear editing process. An edit decision list (EDL) is employed with time code pairs and pointers to the original material. EDLs are then applied to the high-resolution content to generate the final edited video content. In higher end systems (e.g., Sony HDTV, 2005), working images are often stored in uncompressed format at full resolution and with R:G:B as 4:4:4 (12 bits) where possible to avoid any degradation of image quality during multigenerational image manipulations.

The time line is a graphical representation of the edited program, which allows an overview of the linear flow of the program. It shows representations of the clips assembled to create the master edit using the length of the clips to represent their durations and vertical lines to represent the locations of edits between clips. The clip names are displayed at each edit point. Optional thumbnail image representations can be displayed, as well as symbols for segment and transition effects present in the program. The time line consists of tracks that represent separate video and audio streams. A basic master edit has three tracks—one video track and two audio tracks for stereo sound (see Fig. 27.2). More tracks are possible in the more advanced systems.

The time line allows the editor to move through the master edit without having to scroll through the program. This is achieved by providing a clear overview of the location of the various elements within the program. The position locator, represented by the long vertical line and yellow arrowhead to which the arrow is pointing in the figure is moved to any location with one click of the mouse.

Multichannel video editing is also on the horizon not only as a new art and media form, but as a legitimate tool for presentation of multiple video channels; there is a relevance to studying distributed learning with video capture of interactions at each node of collaboration. Commercial editing tools have principally been developed for creating single-channel programs, for example, a linear film or video with one image stream accompanied by simultaneous audio in mono, stereo, or multiphonic variants. Although the semiprofessional and professional systems are capable of multiple video and audio tracks, these are intended as intermediary steps in creating the single-channel master. Layered video tracks are intended for creating composites, keys, and other image effects that will ultimately be reduced to one image track via rendering. The multiple audio tracks are likewise an aid in working with different sound elements such as speech, sound effects, and music that will be mixed down to the finished stereo audio

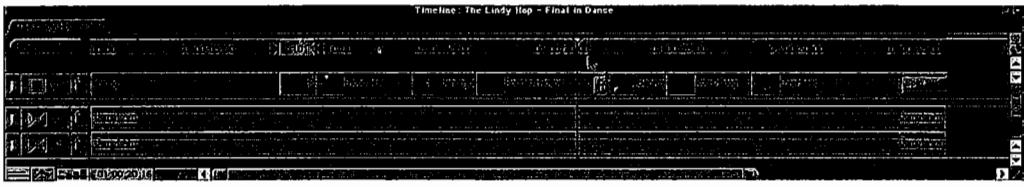

Figure 27.2. A typical time line. This example is taken from Apple Final Cut Pro. The position locator is represented by the long vertical line and yellow arrowhead near the center.

tracks. Nevertheless, these existing tools can be used to other ends for creating nonstandard, multichannel filmic environments. These include synchronized, multichannel films, nonlinear hypernarratives, and stereoscopic films.

Automated editing is a next generation area where certain decision areas can be delegated to smart video editing software. Video metadata standards such as MPEG-7 (see earlier) can be employed initially to segment video into discrete scenes and objects. Such segmentation and objectification of video can support the premise of automated editing schemes. Content can then be organized thematically and semantically based on business rules with automated editing algorithms applied. In the future, one may also employ scripting languages (a "film grammar" that allows for "when X and Y show up zoom into Z") to automate the video editing process for large corpuses of content. Davis (2003) argues that with metadata and media reuse, consumers could more readily become daily media producers through automated mass customization of media.

Video Indexing: Object and Scene Detection

Video indexing allows video to be segmented and deconstructed into component elements suitable for browsing, indexing, search, and retrieval. Tools to handle this can be manual or automated. Manual tools typically allow marking of relevant scenes, frames, or sots (sounds on tape, such as interview clips) using time codes or time code pairs. Users are also able to add annotations or links to related content. These requirements are evident in the learning sciences as many research tools employ them in the feature sets they provide to their user communities. Commercial tools such as the Virage Videologger (Virage, 2005) provide support for both manual and automated indexing of audio and video content, for both stored and lived media. Automated indexing tools seek to index media with little or no human intervention. Video object and scene detection allows for the detection of objects, scenes, key frames, and scene changes in well-understood visual domains, and is enhanced when multimodal information can be used (Snoek & Worring, 2005). Complementary algorithms for speech recognition support speaker identification, speech-to-text conversion, and transcript creation. More advanced models for automated indexing are in a research mode for event detection such as determining there is an event where two people are coming together; this work goes beyond traditional object and scene detection. Panoramic cameras are also being used to capture full 360°degree scenes (e.g., Pea et al., 2004; Sun, Foote, Kimber, & Manjunath, 2001) and the captured content may be used in conjunction with indexing to identify speakers (via audio) and to locate individuals or objects (via video).

Video Analysis

As the chapters of this volume illustrate, "video analysis" circumscribes an extremely diverse set of theoretical underpinnings, researcher objectives, and affiliated work practices concerning what is done with video when it is analyzed. Video analysis includes at least two broad and complementary categories of research; one bottom up

from observations and the second top down from theory. Both forms of analysis often rely on having ready-to-hand some form of transcripts of the talk—from levels ranging from coarse grained to phonological—and possibly annotations regarding gestures, body orientations, actions on artifacts and documents, visual regard—all depending on the purpose of the video analysis.

In the first case of bottom-up inquiries, the researcher is viewing video and building up category definitions and exemplars inductively, from watching video and noting features of activities that appear worthy of designating with a name as a category, and additional exemplars are sought out to render the utility of the category more evident. In their classic paper on interaction analysis using video recordings of human activities, Jordan and Henderson (1995) articulate an accumulating body of wisdom concerning productive interaction analysis work practices, as well as providing an exposition of a number of "analytic foci," or "ways into a tape" that are orienting strategies for the theoretical issues of special interest to interaction analysts and that help in identifying video segments for collaborative group analytic work.

In the second case, following such inductive work, video "coding" is the major video analysis activity, and it depends on having a set of categories, definitions, and exemplars of the category to guide coding practice (e.g., Barron, 2003). For the purpose of conversational analysis of video records of human interactions, researchers tend to use either commercial (e.g., Atlas/ti, HyperResearch, Qualrus), or open source software and analytic tools (e.g., CLAN, Transana, see later) as key enablers to interpret conversations in video interactions.

Current video analysis tools are strong individually in various aspects of annotation and coding of time segments of video (such as Anvil: Kipp, 2001; CAVA: Brugman & Kita, 1998, a replacement for MediaTagger: Brugman & Kita, 1995; ELAN, 2005; Silver: Myers et al., 2001; Signstream: Neidle, Sclaroff, & Athitsos, 2001), editing of video (such as Silver and Transana, 2005), or producing and analyzing transcripts (such as the CLAN programs used in the Child Language Data Exchange System/CHILDES, for studying conversational interactions: MacWhinney, this volume), and multiple points of view on video with attribute significance ratings and visualizations (Goldman's Orion, this volume). Yet none of these tools has directly tackled the core challenges of supporting the broader use, sharing, publishing, commentary, criticism, hyperlinking and XML standardized referencing of the multimedia data produced and output by the tools. This area is a key remaining challenge where support for critical collaborative commentary and cross-referencing—a process that allows researchers to make XML standardized, accessible, and direct contact with competing analyses of video and audio data—is a fundamental advance still needed for scientific disciplines that depend on video data analyses. The integration of video analysis for the work of a community of researchers and practitioners poses technical and design issues that go beyond those inherent in developing video analysis tools, such as Transana and SILVER, which are more focused on specific tasks like video editing or transcribing than on providing a generalinteroperable and global XML standards-based infrastructure for collaboration.

Our research community also faces the challenges of preserving human subject anonymity where this is required by informed consent protocols, while also desiring to develop a cumulative knowledge base where multiple perspectives and competitive

research argumentation can be brought to bear using video data of learning interactions. In order to address privacy concerns, the future of video analysis in the learning sciences may include the possibility of automatically anonymizing video using face and voice recognition (Kitahara, Kogure, & Hagita, 2004) and then transforming faces and voices. Research is also making progress in detecting emotions from facial expressions and contextual information (Picard, 2000), and mapping facial gestures onto computer-animated 3-D facelike surfaces may eventually be used to obscure identity otherwise revealed in the video source recordings.

Video Sharing

Video Asset Management

Specialized content and digital asset management systems allow video to be tagged with metadata, stored in multiple versions, transcoded into alternate formats for delivery (e.g., MPEG-2 and MPEG-4), and automated for generation of hierarchical low-bandwidth media previews and visual proxies for rapid access. E-mail notifications with hyperlinks to video and enabling video for public Web site access fit into this class of system. Metadata schema allow a variety of descriptive and rights-related parameters to be associated with the multimedia content, including but not limited to copyright information, production data, educational topic, level K–12 educational appropriateness, contractual usage restrictions, time codes, scripts or transcripts connected to the content, associations between assets, and composition structure for layered video (i.e., effects, titles, independent tracks, etc.). For example, the Corporation for Public Broadcasting has released *PBCore* (the Public Broadcasting Metadata Dictionary, http://www.utah.edu/cpbmetadata/) as a standard metadata vocabulary of 48 categories for describing and using media including video, audio, text, images, and interactive learning objects to enable content to be more easily retrieved and shared across developers, institutions, educators, and software systems. To provide a bit more detail, 13 different elements describe the intellectual content of a media resource, 7 describe the intellectual property elements that relate to the creation, creators and usage of a media resource, and 28 describe the instantiation elements that identify the nature of the media resource as it exists in some form/format in the physical/digital worlds.

Digital video asset management systems typically include a search, retrieval, and indexing engine as a core component of their design. Database indices often include indexing of free-form text as well as of structured metadata. To provide value to external systems, an increasing trend is to enable an export capability where video assets may be transmitted to external systems via an XML representation of metadata with pointers to related assets in databases. Enterprise class video asset management systems are based on multitier architecture with a canonical Web server, application server, and database server. Enterprise scale asset management systems (e.g., Artesia, Documentum, North Plains, Oracle's Intermedia) start at $50K and can range into the $500K and multimillion dollar levels when deployed for thousands or tens of thousands of users. Low-cost asset management systems as alternatives (i.e., Canto's Cumulus, Extensis, etc.) share a number of similar traits with enterprise scale systems but

with streamlined functionalities. These systems can start at much lower price points in the hundreds to low thousands of dollars for individual use, or in the $10K–$50K range, depending on the number of clients in one's workgroup. More recently, video-centric digital asset management systems such as Venaca and Ardendo have begun to generate significant interest because they embed the functionality of video logging, annotation, rough-cut video editing, transcript searching, and other video features, directly into the core digital asset management architecture.

"Web Services" have provided a new method of abstraction by establishing a globally recognized language and computer-platform independent API (application programming interface) and messaging mechanism by means of a set of definitions of the ways one piece of computer software can communicate with another. Web services are likely to become more prevalent soon for use in video development, including API access for video capture, playback, transformations and so on (http://www.w3.org/TR/ws-arch/). Standards such as XML[1] and SOAP[2] are likely to be used to create a new level of standardization for accessing rich media and video functionality across the Internet; emerging content management standards such as JSR-170 and the nascent JSR-283 should be tracked; progress on these matters also depends on resolving key issues for security, billing, and provisioning. Web services directories (e.g., GrandCentral Communications) and web services interface builders (e.g., Dreamfactory, Curl, Laszlo Systems), are expected to be integral to the advancement of Web services usage. These trends should be watched closely for learning sciences video research support infrastructure.

Video Security

Video security middleware is increasingly required to ensure the security and privacy of video content. Authentication and authorization for media access, roles, and permissions is required. Digital media files can typically be copied and distributed freely across open networks. This approach, while promoting content access and usage, provides limited protection and no direct compensation to copyright holders of media content or protection of data records required by IRB (Institutional Review Board) human subjects protocols. Digital Rights Management (DRM) systems, designed to address such issues, restrict the use of digital files in order to protect the interests of copyright holders, to monetize content delivery, and to allow consumers to legitimately access vast libraries of copyrighted multimedia material. DRM technologies control file access (number of views, length of views, timeframe during which viewing is allowable), as well as file altering, sharing, copying, printing, and saving. DRM technologies can be made available within the operating system, within dedicated software, or in the actual hardware of media capable devices. DRM systems are

---

[1]XML stands for the global standard and general purpose Extensible Markup Language, which makes it possible for groups to create markup languages for describing data (thus, metadata) to support sharing of data across Internet-connected systems.

[2]Simple Object Access Protocol is an XML messaging protocol that encodes information in Web service request and response messages before they are sent over a network. SOAP messages are independent of any operating system or protocol and can be transported using many Internet protocols, such as HTTP, MIME, and SMTP.

now widespread, with close to a billion media players in computers enabled for DRM support. Many content authors and consumers are not aware of the availability of DRM platforms relative to the ubiquity with which these systems are now utilized.

DRM solutions take two distinct approaches to securing content. The first approach is "containment," an approach where the content is encrypted in a package so that it can only be accessed by authorized users. This limits access to content where a user had a valid license to interact with the media. The second approach is "watermarking," the practice of placing a watermark on content as a signal to a device that the file is copy protected. Our focus here is on containment methods. A number of DRM systems are currently used in high-profile media on-demand commercial services to secure content and to generate content revenues for content providers. Note that DRM is not yet available as a capability across all digital media formats. Sample media on- demand services include iTunes Music Store, RealNetworks' Rhapsody Digital Music Service, and for movies—MovieLink, PressPlay, Akimbo, and LaunchMedia (part of Yahoo!).

The component of the DRM system used to package the content is often called a "License Server." DRM systems typically secure content to a server platform and require users to be authenticated for content access through use of a license key. License server platforms package media files and issue licenses. License servers encrypt a given media file, lock it with a license key, and incorporate additional information from the content provider. This results in a packaged file that can only be played by the person who has obtained a license. The license itself may be distributed together or separately from the content in a conventional or encrypted format.

When a user requests playback or access to content that is secured, they must enter the license for the content (or are redirected to a page where they can learn how to obtain a license including payment details), or there must be a communication mechanism with the server to exchange a license key with the server to enable playback. License management allows users to make a specified number of local copies of the content, and to restore media files on a secondary computer in case of a hardware failure on a primary system. Users may also transfer files to secure portable devices, to portable media, and can burn content onto CDROM; however, rules must be set by the content owner to allow each of these types of operations.

The encrypted content may be placed on a Web site, streaming media server, CD/DVD, or e-mailed. Strong encryption is used to protect the content using cryptographic and antipiracy mechanisms. A number of the algorithms are based on published ciphers that have undergone intense review from the cryptographic community. Major commercial DRM systems include Windows Media and Office DRM, RealNetworks DRM, and Apple's FairPlay system used with iTunes (Salkever, 2004). The highest revenues generated to date for digital rights management targeted at pure consumer delivery of digital music have come from Apple Computer and the iTunes Music Store. FairPlay, with many of the aforementioned security characteristics, was able to achieve critical buy-in from the content providers to enable their media for distribution and purchase. It provided a strong DRM solution, along with a networked-based metadata service that can be updated dynamically (such CDDBs, or CD Data Bases, include the open-source sites FreeDB and MusicBrainz, and the commercial encoding CDDB platform from Gracenote used by tens of millions of digital music users, Copeland, 2004).

The significance of DRM solutions for video was underscored in autumn 2005 when Apple Computer began providing major network television shows, music videos, and video podcasting capabilities for its iPod portable media players.

Video Gateways and Media Appliances

Broadband video gateways and media appliances are starting to make major headway into the home. This trend and the component technologies are powerful enough and moving to a more open architecture so that they can be considered and leveraged for research purposes. In 1999, TiVo and ReplayTV launched the first personal video recorders (PVRs), which have since then revolutionized time shifting of consumer TV viewing experiences by providing hard-disk storage for digital video recording. TiVo, the leading PVR company, has an installed base as of September 2005 of 3.6 million units, with estimates of more than 10 million PVRs across all suppliers, a trend expected to accelerate as satellite and cable companies such as DirecTV and Comcast incorporate DVR functions into their set-top boxes. In addition, TiVo is making major advances with their PC to TV connection —based on their home media option. From a price perspective, compared to a media PC—this is very low cost—because as of November 2005, a 40-hour TiVo PVR costs $50 with 12 months prepaid service at a $12.95 per month rate (or a lifetime use fee of $299). TiVo supports TV to PC linkages with TiVoToGo so video can be watched on PCs or on the road. Fishman (this volume) sees TiVo and other PVR devices' capabilities to record live content in a buffer of "constant recording" and save prior events on command as a viable direction for teachers to collect video assets as records of their practice in classrooms.

The TiVo platform today is still a relatively closed architecture, but by the time this book is published, the linkage from PC to TV is expected to strengthen with additional support for a more open programming and extension environment (as indicated by the announced partnership of TiVo and NetFlix to deliver movies-on-demand). Related to TiVo and the home video space, Happauge Digital serves as a very low-cost standalone bridge between the PC and the TV, and Digital Blue and Mattel's Vidstar each provide low-cost capture and movie-making appliances for kids.

Consumers and researchers have alternate platforms considerably more open in design yet way more costly than TiVo—since 2002, Microsoft "Home Media" PC and debuting in 2005, Apple Computer's "Front Row," each allow for direct interconnect to a TV with a specialized interface for remote control and for viewing media rich information at a distance. As prices on such products drop in the coming year, they effectively cross an "affordability chasm" for using these systems in a consumer living room context as well as in research labs. With ultra fast 64-bit chips such as Intel's "Prescott" to power the next generation of Home Media Center PCs (speeds from 2.4 to 3.4 GHz) and the new Viiv initiative, the gaps between computing and high quality home video experiences are disappearing.

Video Publishing—DVD Recorders, Video Web Sites, Hybrid Models

Digital Video Disc (DVD) has rapidly become a common distribution format for video material with low-cost authoring platforms widely available. In addition to the

digitization of video and archival protection relative to analog videotape, interactivity can be added easily with chapter markers for key scenes and shots. Typical storage is on the order of 4.5 gigabytes with discs ranging from one to 2 hours or more, depending on the video bit rate selected. Typical bit rates for MPEG-2 on DVD range from a modest picture quality video stream at 4 Mbps (1.8 GBytes per hour) up to a high quality video picture stream at 10 Mbps (4.5 GBytes per hour). An alternative to reducing the bit rate of MPEG-2 to store more video is to utilize the MPEG-4 format on DVD, which achieves a much higher level of compression (MPEG-4 compression can be 2x to 4x higher than MPEG-2). However, this would require a new generation of consumer players and as a result, is unlikely to occur soon. To further simplify the process of conversion from analog to digital video disc conversion, DVD service bureaus are now broadly available to convert archival videotape into digital video format—the typical cost per tape conversion is now in the range of $15 to $25 per hour of analog video onto DVDs.

The next generation DVD standards are also now on the horizon; key contenders for the next 10 year's of relevance in a world of high-definition video and high-quality audio include Blu-Ray and HD-DVD. A consumer electronic industry coalition led by Sony is supporting Blu-Ray (Belson, 2004)—a new intermediate format for DVDs with 8.5 gigabytes of storage and support for 4 hours or more of video at high quality, and it is called a "double layer disc" (Sharma, 2004a). Double-layer DVD discs are single sided with two data layers that can be independently recorded to and read from, where both layers can be accessed from the same side of the disc. Blu-Ray uses blue lasers instead of the red lasers typically used in optical drives to read data off discs, and supports 50 gigabytes of storage capacity with standards development backed by Dell, Hitachi, Hewlett-Packard, Sony, Samsung, Panasonic, Sharp, and so forth. The use of blue lasers allows storage of more data for the same surface area of the disc. HD-DVD uses a single-lens optical head that integrates both red and blue laser diodes, and supports 30 gigabytes of storage capacity with standards development backed by leaders Toshiba and NEC, 200 other companies in the DVD forum, and supported by Microsoft, and now Intel (in November, 2005). Although it has less storage than Blu-Ray, its backers consider HD-DVD more reliable as a storage medium (Sharma, 2004b). As of early 2006, these competing standards are playing out in a drama, with a standoff in Hollywood (Belson, 2005), as to whether consumer electronics/TV (Blu-Ray) or computer companies (HD-DVD) will rule the future of digital video disk-based technologies.

Based on the high bandwidth required for TV resolution video, DVD stands as a superior publishing medium relative to the Web. However, as bandwidth of the Internet rises overall, and broadband to the home, office and schools rise, expect that TV quality video will start to migrate to the Web. Even today, video Web sites that allow posting of indexed and searchable video with commentary can form the basis for new formats of e-publications of video material.

For example, in this volume, Beardsley, Cogan-Drew, and Olivero describe the VideoPaperBuilder system, software that enables teachers and researchers to work together to build multimedia Web-page documents called VideoPapers that closely link video, text, and still images from classroom practices. Authors may annotate segments of digital video of teaching or learners, with text comments or scanned records of student work or teacher handouts on paper or whiteboards. The completed document

uses JavaScript menus, html links and frame sets, and QuickTime image slide shows to interweave the authors' video, slides, and text into a single multimedia presentation that can be interactively experienced by users using Web browsers.

A frequently noted example of this video-sharing trend can be found in the Open Video project (Geisler, 2004). The Open Video site (http://www.open-video.org) showcases effective video sharing across the Internet, metadata use, and how to make video more broadly accessible across a range of research and public user communities. The site houses a variety of video collections comprising over 3,000 videos, such as the CMU Informedia Project, the Howard Hughes Medical Institute, and the Prelinger Archives. All video content can be easily searched and browsed; metadata (with rights information) has been tagged for all video clips, and each video segment has a short preview (7 sec), a multiframe storyboard representation, and access to the original high-resolution video source material in digital format (i.e., MPEG-x). The next stage of the project will add more video formats, genre characteristics (student television, anthropological footage, technology demonstrations), and more collections for the video community site. During 2005, a plethora of Web sites was also launched that enable the general public to upload their videos and tag them with categories and brief commentaries, including Google, YouTube, Vimeo, Clipshack, OurMedia, VideoEgg, and so forth. Secure peer-to-peer group sharing of video and audio recordings is the focus of other commercial ventures such as Grouper and Veoh.

There are also commercial Web sites, such as Teachscape, LessonLab, CaseNex, and TeachFirst, all of which incorporate high-quality digital video as an integral element of their service. In this case, the focus is on teacher professional development and/or preservice education. Some of these companies, such as Teachscape (2005), use hybrid models that utilize the connectivity and interactivity of the web in conjunction with the high-bandwidth, high-speed media delivery platforms such as DVD in combination. A video program can be authored in tandem so that users interact with a Web browser or application constructed on the Mac or PC, with high-bandwidth media accessed rapidly from a local DVD drive. Users of hybrid model Web sites are shipped physical video discs for the video. Such hybrid systems will be able to provide very rich media soon with the advent of the DVD dual layer drive, HD-DVD, and Blue-Ray standards described earlier. This model was used frequently in the past as well with CD-ROM discs; however, as video delivery over the web has improved, this mode of interaction makes less sense. The higher density DVDs make the model viable yet again, but there will always be a race between high-bandwidth optical media and high-speed Internet connectivity to the home, school, university, and office. Raul Zaritsky's stimulating chapter (this volume) works in a related vein, but with several surprises, advancing high quality video case studies as what he designates "educational research visualizations" to serve as scaffolding for teachers seeking to understand and emulate the rationale and situated practices of a reform-oriented mathematics curriculum. If effect, he argues that these visualizations serve as warrants in an argument for the appropriation of new teaching practices, and how such a "workshop in a box" as a new media form could accelerate the adoption curve for theory-driven designs into education. The results he reports are sobering for innovators seeking to advance new media grammars that exploit multiple camera angles, multiple audio tracks, 3-D graphics, and other new affordances of digital video

for educaion, as they can be perceived as complex and disorienting distractions rather than helpers without more teacher familiarity.

VIDEO COLLABORATION

Digital Video Collaboratories

Video collaboration systems will support key elements of the learning sciences video research workflow, enabling end-users to analyze, share, and collaborate around video records. Such systems will form the backbone of a video research framework. Earlier work enabled real-time and/or asynchronous text messaging among multiple participants as they are watching video broadcasts or video archives (e.g., Bargeron et al., 2002; White et al., 2000), but our work practices in the learning sciences require far more than that. Despite consumer-driven advances in video capture technology, and a sharp rise in the use of video for analysis purposes by solo researchers, video circulates sluggishly, if at all, within research communities. The same researchers who use video for analysis typically rely exclusively on text to present results. Researchers default to text because they cannot readily ensure that an audience can view video as source data, much less in a form that integrates an argument with video evidence.

This promise–reality gap for digital video has serious consequences for researchers. Connections between evidence and argument are obscured, the development of shared examples of exemplary analyses using video that can serve training and socialization functions for researchers is impeded, and sharing of video data among scientists is discouraged in favor of an isolated and inefficient approach to gathering and analyzing primary data. Research communities will not make full use of video data so long as significant obstacles remain at any of the key points of video capture, encoding, storage, retrieval, analysis, sharing, and commentary. Enabling research communities to build knowledge through sharing video data and analyses would constitute an important enhancement to the global research and education infrastructure. We see this emphasis shared throughout the chapters of this section of our volume.

In the Digital Video Collaboratory Project, where the DIVER team at Stanford has teamed with Brian MacWhinney's TalkBank team housed at CMU and the University of Pennsylvania, we have been addressing these critical issues and enabling communities of researchers and practitioners to collaborate in producing, analyzing, and commenting on an evolving corpus of video records in diverse disciplines studying learning and human interactions. Our project is centered on creating highly accessible tools for video analysis, sharing, and collaboration. We seek to establish a strong basis for broader impact across multiple disciplines and applications with our focus on accessibility, ease of use, core technical advances, and metadata/API standards. Achieving these goals requires leveraging information technology advances and innovations in Web-based computing, video analysis and collaboration tools, and video compression and streaming. To achieve the primary goal and validate our tools, we have been conducting our research initially as a multi-institution collaboration between Stanford University and Carnegie Mellon University, but we plan to develop and use our enabling infrastructure as a unified Digital Video Collaboratory for broad accessibility to

researchers in a range of disciplines studying interactions in classrooms and in other contexts of human activity.

Enabling the free flow of video data and analyses within research communities requires three capacities that are currently lacking, each of them addressed by our Digital Video Collaboratory Project. First, video data has to be universally accessible without regard to its physical location. Currently, video data is available, if at all, in heterogeneous repositories with idiosyncratic access control, search and retrieval interfaces, and metadata structures. We are addressing this obstacle by developing a virtual video data repository and video analysis community portal, implementing a metadata scheme designed to support research use of video-as-data.

Second, research communities require video analysis tools that support the full range of scientific activities from inductive development of categories for interpretation to coding analysis and through collaboration, critique, and publishing. Currently, video analysis tools maroon data on islands of incompatible file formats, making it difficult to share data among applications, much less among other researchers in a free flow of data and argumentation. We address this obstacle by developing both generic and discipline-specific XML-based schema for video analysis to facilitate application interoperability, and flexible desktop and Web-based video analysis tools that directly support sharing, critique, and output of video analyses. The use of XML will facilitate development of specialized XML extensions to represent discipline-specific metadata for use by such components, and also make possible data exchange with other video analysis tools using XML, such as Atlas.ti and SignStream tools.

Finally, if video is to be a primary communications medium, other researchers must be able to respond to a video analysis using the medium of video itself. Our DIVER team at Stanford in this project has developed an approach for enabling distributed video analysis that allows random space-time access into compressed video streams, while not requiring the downloading of video into local computer storage for authoring new video clips (see Pea, 2006).

The DIVER system uniquely enables "point of view" authoring of video analyses in a manner that supports sharing, collaboration, and knowledge building around specific references into video records. We do this by enabling users to easily create an infinite variety of new digital video clips from a video record. This process is called "diving," and the author a "diver," because the DIVER user "dives" into a video record by controlling—with a mouse, joystick, or other input device—a virtual camera viewfinder (see yellow rectangle, Fig. 27.3) used to mark snapshots of specific moments, or to record multiframe video "pathways" through a video to create their "dive." The use of DIVER to focus the attention of an observer of one's dive on a video resource is what we call "guided noticing." Guided noticing is a two-part act for a visual scene that has been a vital part of cultural learning episodes, long before computers existed: First, a person points to, marks out, or otherwise highlights specific aspects of that scene. Second, a person names, categorizes, comments upon, or otherwise provides a cultural interpretation of the topical aspects of the scene upon which attention is focused. In the case of DIVER, such guided noticing is time-shifted and shareable by means of recording and display technologies. Diving creates a persistent act of reference with dynamic media—which can then be experienced by others remote in time and space, and which can additionally serve as a focus of commentary and re-interpretation. Why is

guided noticing important? Because achieving "common ground" (e.g., Clark, 1996) in referential practices can be difficult to achieve, and yet is instrumental to the acquisition of cultural categories generally, and for making sense of novel experiences in the context of learning and instruction especially.

As illustrated in Figure 27.3, a dive is made up of a collection of "panels" on the right side of the web page, each containing a small video key frame representing a mark or video clip, as well as a text field containing an accompanying annotation, code, or other interpretation. Both the annotations and the space–time coordinates of a user's dive on video records are represented by the DIVER software system as XML metadata, so that one is not literally creating new video clips, but simply views into parts of one or more video files through a dive.

DIVER is designed to serve the purposes of both the video researcher who captured the video records, and his or her research collaborator or colleague who desires to have conversational exchanges anchored in specific moments that matter to them in the video segments. First, the video researcher uploads video data in any one of a typical range of formats to the DIVER server. Once DIVER software services automatically transcode the video into a streaming format, the researcher then may use a client-side Web browser to mark and record space–time segments of videos with the virtual camera, and to make text annotations about them as they build up their "dive" for analyzing the video record. (To provide security to video records, streaming video files in the Macromedia Flash, or .flv format, are made accessible through a Web server over the Internet, so that video files will not be downloaded to personal computers.)

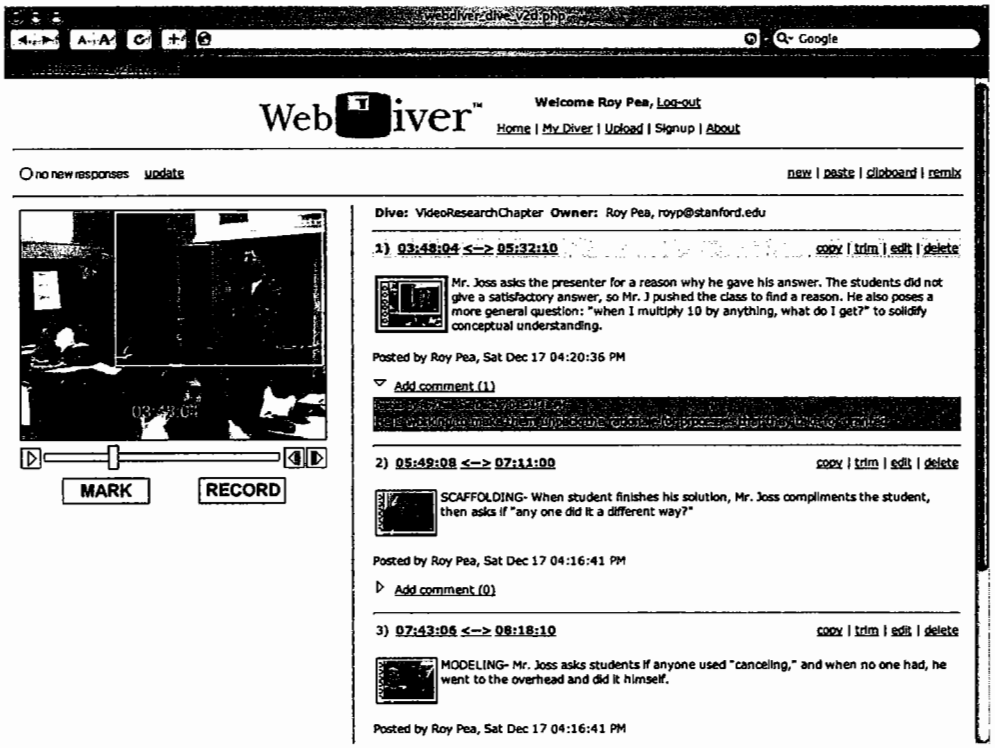

Figure 27.3. WebDIVER: Streaming media interface for Web-based diving.

The researcher's dive may then serve as the multimedia base medium for processes of scientific interchange, supporting collaboration and elaboration, as well as a critique of scientific argument and research evidence. The databases of primary video records and secondary analyses, or "dives" then become available to approved users through a browsable, text-searchable community-based Web site. Other researchers viewing the originating researcher's dive can respond to that diver's annotations by posting their own textual comments linked to the video in question (as in the first panel of Fig. 27.3), which could then be viewed by the diver and by other researchers, and be developed further in a Web-enabled video-anchored dialog. Dive respondents may also make their own dives on video records, referencing segments from any of the terabytes of video files available through the DIVER servers. All of these activities generate searchable metadata, and support finding analysis-relevant clips and analyses in a video research community of practice.

Our DIVER team is extending these developments to realize our vision of the Digital Video Collaboration (DVC) for robust access control, group formation, e-mail notifications of changes in dives one has authored or subscribed to, and so on. We have also integrated WebDIVER with the DVC virtual data repository concept, so WebDIVER users can store and retrieve video data and analyses without regard to their underlying physical storage locations. WebDIVER users can make dives into videos stored and served from distributed web servers or content delivery networks (CDNs), and play back dives as 'remixes' that reference only the spatio-temporal segments pointed to within the dive.

MacWhinney (this volume) characterizes progress toward building tools to facilitate this new process, which he calls "collaborative commentary," and defines as the involvement of a research community in the interpretive annotation of electronic records, with the goal of evaluating competing theoretical claims. The collaborative commentary process involves linking comments and related evidence to specific segments of digital video, transcripts, or other media. He describes seven spoken language database projects that have reached a level of Web-based publication that makes them good candidates as targets and beneficiaries of collaborative commentary technology.

Goldman (this volume) has been pioneering for many years in her video ethnography software systems the importance of "points of viewing theory," most recently software in her work on the Web-based Orion video analysis system, in which the insights generated by diverse participants on a given video record are valued as ethnographic contributions. Reed Stevens' desktop VideoTraces software (this volume) is oriented to reflection and presentation—enabling users to lay down a "trace" on top of a "base" video record (playable at variable speeds). The trace consists of voice annotation and a gesture depicted as a pointed hand cursor. When a VideoTraces file is played, one hears the audio trace overlay and sees its gestural focus. Stevens and colleagues have used this system in science education museums and in higher education courses such as rowing and dance composition. VideoTraces' uses of virtual pointing and voice-recorded commenting on video provide a complementary mechanism to our use in DIVER of guided noticing for achieving common ground in a referring act in the complexity of a video record.

We hope these new capabilities to establish digital video collaboratories will accelerate scientific advances across a range of disciplines. Beyond the bounds of acade-

mia, the availability of a fluid and reliable mechanism for publishing video data and analyses can support research dissemination by providing a means for public access to research results and enabling community commentary (see Pea, 1999).

MacWhinney (this volume) reviews a variety of approaches for enabling collaborative commentary among research video user communities—collaboration around the core elements of video research including video clips, video, transcripts, coding schemes, and annotations. Such community environments allow users to view and annotate material from one another, and include tools such as Zope, Zannot, Annotea, B2Evolution, Blogger.com, Blogging.com, and RSS (Really Simple Syndication).

Baecker, Fono, and Wolf (this volume; also Baecker, Moore & Zijdemans, 2003) have been developing a system called ePresence, designed to enable global broadcasts over the Web—of video and slides and presentations and live software demos, real-time interactive access to broadcasts by remote viewers who can have public or private text-based chats and submit questions to presenters, and postevent access to presentation archives. Although not initially conceived as a collaboratory for video data sharing in the learning sciences, Baecker et al. highlight how the video channel in ePresence can be used not only for video of lecturers, but for collaborating researchers to share video data access and to have text-based chats and threaded discussions about such resources. Because these discussions can incorporate Web links to other video sources and documents, ePresence provides a potentially powerful infrastructure for digital video collaboratory activities among learning science communities.

Communities of Interest Networks (COIN)

Communities of interest networks have emerged in recent years, thanks to blogging, RSS Web feeds (e.g., pubsub, newstrove, rocketinfo), and Web-based community platforms that enable participants to specify topics of interest so that they are regularly notified of results of searches (e.g., Google Alert), other news streams (RSS) culled from millions of Web sites, or new citations of articles published (e.g., Ingenta provides 20 million online articles from nearly 30,000 publications). We believe that as video resources become more widely available on the Web and in communities of practice, such as learning sciences research, and interest "grows" around them, that COIN infrastructure services will become available and widely used. One attractor to such COIN services is that they come to provide a form of "social information filtering" in which highly used, highly rated, or highly cited resources bubble to salience through patterns and levels of use of these resources by participants in the networks who are using these resources. In this way, video collaboratories can come to leverage the network effects (Katz & Shapiro, 1985; Rohls, 1974) seen in other Web spheres from e-commerce to communications, in which the value of a network grows exponentially with the number of nodes attached to it (see Barabasi, 2002).

CONCLUSIONS

It is noteworthy that we have observed a great proliferation of genres in the past few years that incorporate interactive multimedia and differing levels and kinds of af-

fordances for collaboration. These include systems such as Orion's "constellations," DIVER's "dives," VideoPapers, VideoTraces documents, ePresence "archives," TalkBank video transcripts, and so on. The multiplicity of such systems raises a number of important issues, and dimensions for comparative analysis as researchers seek to find the best fit to their desired work practices. These issues include ease of use, embeddability in Web-commentary layers, access/security/IP regarding content, search capabilities, and virtual access to video stored across multiple distrivuted servers. As we have indicated, video researchers studying learning and teaching have a great deal to look forward to as the converging advances of computing and media communication technologies make formerly advanced technologies into everyday consumer and research tools.

ACKNOWLEDGMENTS

The National Science Foundation has provided support at Stanford University that we gratefully acknowledge for work reported in this chapter (#0216334, #0234456, #0326497, #0354453). The opinions and findings expressed are our own. We'd like to thank Joe Rosen for his many contributions as DIVER software architect and engineer for advancing the agenda of open-standard digital video collaboratories.

REFERENCES

Apple QuickTime 7/MPEG. (2005). Retrieved October 30, 2005 from http://www.apple.com/mpeg4

Baecker, R. M., Moore, G., & Zijdemans, A. (2003). Reinventing the lecture: Webcasting made interactive. *Proceeding of HCI International 2003* (Vol 1, pp. 896–900). Mahwah, NJ: Lawrence Erlbaum Associates.

Barabasi, A.-L. (2002). *Linked: The new science of networks*. Cambridge, MA: Perseus.

Bargeron, D., Grudin, J., Gupta, A., Sanocki, E., Li, F., & LeeTiernan, S. (2002). Asynchronous collaboration around multimedia applied to on-demand education. *Journal of Management Information Systems. 18*(4), 117–145.

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences, 12*(3), 307–359.

Belson, K. (2004, September 20). New economy: Format wars, Part 3—Sony and its allies battle 200 companies over the next generation of digital videodiscs. *New York Times, Section C, p. 3.*

Belson, K. (2005, July 11). A DVD standoff in Hollywood. *New York Times, Section C, p. 1.*

Brugman, H., & Kita, S. (1998, February). CAVA: Using a relational database system for a fully multimedial gesture corpus. *Workshop: Constructing and Accessing Multi-media Corpora: Developments in and around the Netherlands*. Nijmegen, The Netherlands.

Brugman, H., & Kita, S. (1995). Impact of digital video technology on transcription: A case of spontaneous gesture transcription. *KODIKA/CODE Ars Semeiotica, An international journal of semiotics, 18,* 95–112.

Chanko, T., Wiqder, Z. D., & Scevak, N. (2005, July 19). *US DTV and iTV Forecast, 2005 to 2010*. Report may be purchased from http://www.jupiterresearch.com/bin/item.pl/research:vision/1211/id=96505/

Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.

Copeland, M. (2004, March). The magic behind the music. *Business 2.0, 5*(2), 40.

Davis, M. (2003). Editing out video editing. *IEEE Multimedia, 10*(2), 2–12.

ELAN. (2005). *EUDICO linguistic annotator.* Software downloaded October 30, 2005 from http://www.mpi.nl/tools/elan.html

Front Porch Digital. (2002). *An overview of digital video archives in broadcast: A white paper for the media and entertainment industries*. Retrieved October 30, 2005 from http://www.fpdigital.com/uploads/1115225972.pdf

Geisler, G. (2004, February). *The open video project: Redesigning a digital video digital library.* Paper presented to the American Society for Information Science and Technology Information Architecture Summit, Austin, TX.

Gilheany, S. (2004). *Projecting the cost of magnetic disk storage over ten years.* Retrieved November 5, 2005 from http://www.aiim.org/documents/costmagstorage.pdf

Gray, J. (2004). *The five minute rule.* Microsoft Research Presentation on PetaByte Server Infrastructure. Retrieved October 30, 2005 from http://research.microsoft.com/~Gray/talks/FiveMinuteRule.ppt

Helix Server. (2006). Retrieved December 28, 2006 from http://www.realnetworks.com/products/media-delivery.html ; http://www.helixcommunity.org

Helix Community. (2006). Retrieved December 28, 2006 from http://www.helixcommunity.org

Hoffert, E. M., & Waite, C. (2003, August). *Post-Linear Video: Editing, Transcoding, and Distribution.* Paper presented at the Conference Proceedings of the ACM SIGGRAPH 2003, San Diego, CA.

IDC. (2005, January). *WGBH digital asset management prototype lays foundation for lower costs, increased efficiencies, and enhanced services: An IDC eBusiness case study*. [IDC Report FE2016-0]. Retrieved October 30, 2005 from http://www.artesia.com/pdf/IDC_CaseStudy_WGBH.pdf

IEEE 1394b. (2005). *About IEEE 1394b technology*. Retrieved from http://www.1394ta.org/Technology/About/1394b.htm

Internet Archive. (2005). *Internet moving image archive.* Retrieved October 30, 2005 from http://www.archive.org/movies/movies.php

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *The Journal of the Learning Sciences, 4*(1), 39–103.

Katz, M. L. & Shapiro, C. (1985). Network externalities, competition, and compatibility. *The American Economic Review, 75*(3), 424–440.

Kinoma. (2005). Kinoma player, Kinoma producer. Retrieved October 30, 2005 from http://www.kinoma.com

Kipp, M. (2001). ANVIL: A Generic Annotation Tool for Multimodal Dialogue, *Proceedings of Eurospeech* (pp. 1367–1370). Aalborg, Denmark, September, 2001.

Kitahara, I., Kogure, K., & Hagita, N. (2004, August). Stealth vision for protecting privacy. *Proceedings of the 17th International Conference on Pattern Recognition, 4*, 404–407.

Langberg, M. (2004, March 26). New hard drives may turn handhelds into tiny TiVos. *Forbes*. Retrieved October 30, 2005 from http://www.forbes.com/technology/2004/03/0326harddrivespinnacor_ii.html

McDaniel, T. W. (2005). Ultimate limits to thermally assisted magnetic recording. *Journal of Physics: Condensed Matter, 17*, R315–R332.

Myers, B., Casares, J. P., Stevens, S., Dabbish, L., Yocum, D., & Corbett, A. (2001). A multi-view intelligent editor for digital video libraries. *Proceedings of the 1st ACM/IEEECS Joint Conference on Digital Libraries* (pp. 106–115). Roanoke, VA.

Napier, D. (2006, January). Build a home terabyte backup system using Linux. *Linux Journal, 141*, p. 3.

Neidle, C., Sclaroff, S., & Athitsos, V. (2001). A tool for linguistic and computer vision research on visual–gestural language data. *Behavior Research Methods, Instruments, and Computers, 33*(3), 311–320.

Pea, R. D. (2006). Video-as-data and digital video manipulation techniques for transforming learning sciences research, education and other cultural practices. In J. Weiss, J. Nolan, & P. Trifonas (Eds.), *International handbook of virtual learning environments* (pp. 1321–1393). Dordrecht: Kluwer.

Pea, R. D. (1999). New media communication forums for improving education research and practice. In E. C. Lagemann & L. S. Shulman (Eds.), *Issues in education research: Problems and possibilities* (pp. 336–370). San Francisco: Jossey Bass.

Pea, R., Mills, M., Rosen, J., Dauber, K., Effelsberg, W., & Hoffert, E. (2004, January/March). The DIVER™ Project: Interactive digital video repurposing. *IEEE Multimedia, 11*(1), 54–61.

Picard, R. W. (2000). Toward computers that recognize and respond to user emotion. *IBM Systems Journal, 39*(3&4), 705–719.

QuickTime/SMIL. (2005). *Usage of the synchronized multimedia interface language in the QuickTime multimedia standard.* Cupertino, CA: Apple Developer Technical Specification.

Rohlfs, J. (1974). A theory of interdependent demand for a communications service. *The Bell Journal of Economics and Management Science, 5*(1), 16–37.

Salkever, A. (2004, March 24). Digital music: Apple shouldn't sing solo. *Business Week Online.*

SAN. (2005). Retrieved from November 2, 2006 from http://www.webopedia.com/TERM/S/SAN.html

SGI/HDTV. (2004). Real-Time, Full-Bandwidth HDTV I/O with SGI HD XIO. Discreet / Silicon Graphics Technical Specification.

Sharma, D. (2004a, March 17). SONY debuts double-layer DVD drive. *News.com Article.* Retrieved October 30, 2005 from http://news.com.com/2100–1041–3–5174122.html

Sharma, D. (2004b, January 7). Toshiba spotlights high-definition DVD player, *News.com Article.* Retrieved October 30, 2005 from http://news.com.com/2100–1041–5136601.html

Shoah. (2005). *Survivors of the Shoah Visual History Foundation* Retrieved October 30, 2005 from http:// www.vhf.org

Snoek, C. G. M., & Worring, M. (2005). Multimodal video indexing: A review of the state-of-the-art. *Multimedia Tools and Applications, 25,* 5–35.

SONY/HDTV. (2005). *Real-time, HDTV I/O with the SONY HDCAM Software Codec. Discreet / SONY Technical Specification, 2005.*

Sun, X., Foote, J., Kimber, D., & Manjunath, B. S. (2001). Panoramic video capturing and compressed domain virtual camera control. *Proceedings of the 9th ACM International Conference on Multimedia,* pp. 329–338.

Teachscape. (2005). Teachscape web site and service. Retrieved October 25, 2005 from http:// www.teachscape.com

Transana. (2005). *Qualitative analysis software for video and audio data.* Retrieved from http://www.transana.org

Virage. (2005). Retrieved October 25, 2005 from http://www.virage.com/

Walter, C. (2005, August). Kryder's law. *Scientific American, 293*(2), 32–33.

Wang, X. (2004). MPEG-21 rights expression language: Enabling interoperable digital rights management. *IEEE MultiMedia, 11*(4), 84–87.

White, S. A., Gupta, A., Grudin, J., Chesley, H., Kimberly, G., & Sanocki, E. (2000). Evolving use of a system for education at a distance. *Proceedings of the 33rd Hawaii International Conference on System Sciences* (Vol. 3, p. 3047).