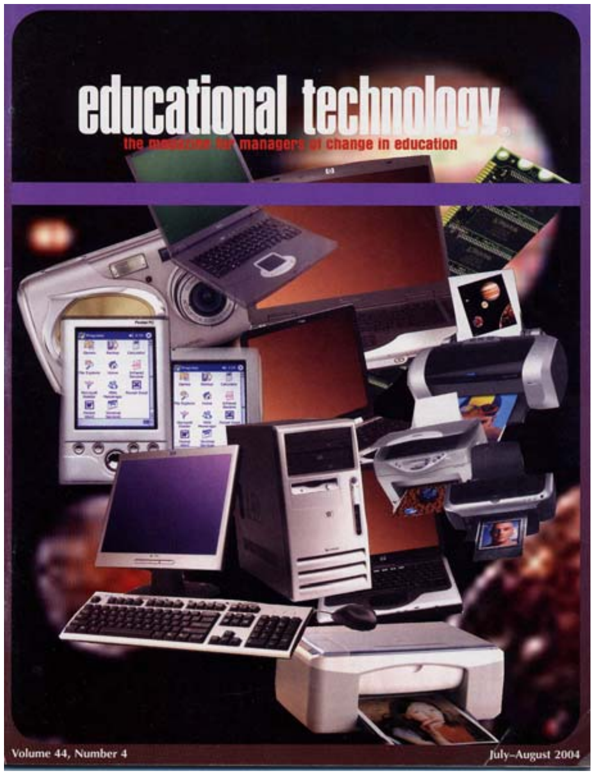
Customized Video Playback: Standards for Content Description, Customization, and Personalization

(This article was published in *Educational Technology, 44*, 4, pp, 5 – 13. Appears here by permission)

Michael D. Bush, Alan K. Melby, Thor A. Anderson, Jeremy M. Browne, Merrill Hansen, and Ryan Corradini[1]



Introduction to Customized Video Playback (CVP)

Early in the cellular telephone revolution, promoters of that industry sounded the clarion call regarding information availability that went something like "anything, anywhere, anytime." Nicolas Negroponte, director of MIT's Media Lab, later coined what can be considered to be an important slogan of the personalized content revolution, "nothing, nowhere, never unless it is timely, important, amusing, relevant, or capable of engaging my imagination" (Negroponte, 1994). Indeed, considering education and entertainment, there are hundreds of television channels available via satellite and cable, thousands of pre-recorded programs from VHS tapes and DVDs, as well as an increasing number of video files available on the Web, presenting consumers, learners, and educators alike with a variety of content that is seemingly endless. Not

---

[1] Michael D. Bush is Associate Director of the Center for Language Studies and Associate Professor of French and Instructional Psychology and Technology, Brigham Young University, Provo, Utah (e-mail: Michael_Bush@byu.edu).

Alan K. Melby is Director of the Electronic Film Review project and Professor of Linguistics, Brigham Young University, Provo, Utah (e-mail: akm@byu.edu with EFR at beginning of subject line).

Thor A. Anderson is a professor at Utah Valley State College, Orem, Utah (e-mail: thor_anderson@comcast.net).

Jeremy M. Browne is a PhD candidate in the Department of Instructional Psychology and Technology, Brigham Young University, Provo, Utah (e-mail: browne@byu.edu).

Merrill Hansen is Director of the ViewerFreedom Foundation of Cupertino, California (e-mail: merrill@viewerfreedom.org)

Ryan Corradini, a recent graduate of the Brigham Young University Linguistics program, is an independent software consultant and owner of Down East Interactive Concepts, Sea Level, North Carolina (e-mail: Ryan@intconcepts.cjb.ne t).

only do they all have access to the arrival of an incredible variety of programs via these various means, but they also can time-shift broadcasts via VCRs and now digital personal video recorders (PVRs) such as TiVo® and ReplayTV®.

These developments in video technologies help to tailor video viewing to suit consumer tastes and learner needs, expanding our options concerning ***which*** video programs we watch and ***when*** we watch them. Unfortunately, our ability to select ***how*** we watch these programs is mostly limited to our manual dexterity with the couch potato's best friend: the remote control for the TV, VCR, DVD, or PVR. Even playback software on most *computers* leaves much to be desired. Once a video program is selected from all those available, the viewer can only jump around from place to place within the chosen documentary, feature film, or perhaps recorded lecture. With respect to selectivity, however, the crudeness of this approach is apparent. Whether the viewer is an ordinary consumer, a student, or a professor as shown in our example below, the basic needed functionality is the same: To be able to select or avoid **segments** of video as desired.

Thus, two levels of selection are needed: (1) selecting a desired program, and also (2) selecting the portions of the program to be included or excluded. This second level of selection creates the possibility of true content customization for a group of viewers or even personalization for an individual viewer, which we generically call Customized Video Playback (CVP). Once a specific program has been chosen, CVP enables the viewer (or presenter in the case of an instructor) to select precisely those portions of video that correspond to predetermined objectives or choices. Customizing the playback of a video program does not necessarily mean creating a new video program with a persistent, stand-alone existence. The video program itself need not change, only the manner in which it is viewed.

In using software to achieve this customized playback, an English literature professor might choose and show several scenes from a certain rendition of Hamlet to highlight Laertes' development as a character in the play. Afterwards, to support student review in a computer lab or other student workspace, the professor could then make available the same presentation that had been made in the classroom, if the method of CVP being used supports repeated viewing of the same customized presentation. Using the same or similar software as the English literature educator, a foreign language educator might well wish to play a few lines of a conversation over and over again to focus on one communicative function, while a professor of film may wish to show only segments of a film that use certain cinematographic techniques. Supporting the idea of such an implementation, Marchionini (2003) describes video programs as a resource that "offers instructional designers a powerful medium for crafting mutually reinforcing explications of concepts while providing learners with content that engages multiple senses" (Marchionini, 2003, p. 36). He goes on to list several requirements for teachers who wish to incorporate video into their teaching, such as "tools for integrating video segments into presentations and lesson plans" (pp. 38-39).

With respect to entertainment, consumers could choose to use similar functionality to avoid content that they deem undesirable or to choose favorite segments of a film that they wish to view, all this while in the privacy of the home. A Kung-Fu fan could choose to view only the fight scenes from her preferred martial arts film, while a softhearted romantic could view the touching scenes from his favorite romantic comedy, or a watchful parent or discretionary viewer could avoid viewing specific scenes of offensive material during an otherwise appealing movie. In a similar vein, such an approach would provide a significant tool to classroom teachers who must concern themselves with the appropriateness of the films they choose for in-class viewing (Gareis, 1997).

Customized Video Playback goes beyond including or excluding selected segments and leads to the concept of annotation, an important aspect of CVP. In applications going back to the early 1980s, feature films have been annotated in such a fashion so as to open them for use for students to learn language and culture (Gale, 1989). The description of the content includes both content modeling, such as a hierarchical segmentation of a video program, and annotation of individual segments.

In all three contexts (classroom, student workspace, and home), viewers can use CVP to see what they wish to see, when they want to see it. They are able to go beyond simply watching a video program sequentially from beginning to end, and instead benefit from a level of control that better meets their needs. The increased control enables them to tailor the playback of the video program for a particular group in the case of the English professor above, or even personalize the playback in the case of the student or the home entertainment consumer.

CVP: Fine Tuning User Control and Personalization

There are at least two significantly different ways to achieve repeatable, customized viewing of a video program. One method, a file-based approach, is to encode the program digitally, place it on a computer's hard disk, and use video editing software to create a new video program that is entirely separate from the source video program. This new file can be played without access to the original video program. A second method, a selective playback approach, amounts to making selections from a video stream based on a content model of the video, where the selected segments are not separate files but rather are “virtual” clips from a single video source such as a single title on a DVD.

The file-based approach has four major drawbacks:

- It is time-consuming,
- It requires specialized skills,
- It requires specialized, expensive, software and hardware tools, and
- It can violate copyright law.

In comparison, the selective playback approach, such as is implemented in our preferred approach to CVP, involves the creation of a content model of the video program of interest. Based on viewer choices, playback software accesses that content model, and creates customized playback on-the-fly. To do this, CVP, as we envision it, depends on three coordinated technical components:

- The video program (henceforth called a video asset),
- A description of the content of the video asset, and
- A selective playback mechanism.

The video asset is any video stream that is accessible via time code. We include in the notion of video asset at least three types of video programs, from the perspective of copyright: (1) original footage whose copyright belongs to the organization doing the customized playback; (2) licensed footage for which there is a contract specifying what can and cannot be done with the video program; and (3) off-the-shelf commercial footage in a packaged medium such as a DVD that can be purchased by the public at a retail outlet. Regardless of the copyright status and regardless of the medium (e.g., optical disc or magnetic disc) and even regardless of the encoding (MPEG-2 or some other encoding of the video stream), the Video Asset Description (VAD) is represented in some file external to the asset itself and defines segments within the video asset using time codes rather than through any copying or modification of the original video material. In addition, end-user tools facilitate the creation of playback specifications from video asset descriptions. The resulting combination of software, data format, data, and digital video stream amounts to an efficient approach to a content retrieval and processing system. Preferably, the format of the video asset description is based on a suitable standard. In this paper we will focus on the MPEG-7 standard.

One advantage of using a standard content description system is that multiple authoring software tools (for segmentation and annotation) and multiple delivery software tools (for customized viewing) would be reusable in various combinations in the three different settings described above, from the classroom, to a student's workspace, to the home. This would not be the case if each authoring software tool produced only a proprietary representation of the description of the video asset.

With a standard VAD format, the same video asset description could also be used with software packages from different vendors. In order for the system to work, someone must analyze the content of a video program and publish the analysis as a VAD. An end user interested in that video program then acquires the copy of the video asset and its VAD and uses a VAD-aware tool to create a playback specification according to personal objectives. The video program is then viewed under control of that playback specification. Different end-users could potentially use the same video asset and the same VAD to produce a multitude of personalized views of the same video program, without needing to understand the inner-workings of the system and without actually copying or modifying the video program.

Within the selective playback approach to CVP, it is important at this point to differentiate between two sub-approaches to establishing the playback specification for customizing the viewing experience. Our VAD-based approach, combined with suitable tools, allows end users to create their own playback specifications. In this case, viewers begin with the video asset description and then use it to facilitate the construction of their personalized playback specification. Another approach is for viewers to purchase a pre-defined playback specification that has been created for a particular purpose, to select or avoid particular content. Such solutions employ a playback specification that is distributed to the public (gratis or for purchase), rather than beginning with a video asset description from which private, personalized video

video playback specifications are created. In both sub-approaches, a VAD could be used to create playback specifications. The crucial distinction is in whether the user creates a personal playback specification or obtains a pre-made playback specification. This distinction between publicly available and privately created playback specifications may have important intellectual property rights ramifications with respect to the third type of video program mentioned above: commercial films for which no special license has been negotiated.

Technical Description of CVP

Although various CVP applications could each be developed independently from scratch, the purpose of this article, as indicated by the title, is to argue for a standards-based approach to CVP. We will present one standard for the description of a video asset and describe several existing or evolving types of specifications for the customized playback of video assets. This section assumes a basic knowledge of XML on the part of the reader.

Every CVP application is based on one or more video assets that could exist in any of several forms. For example, a video asset could be an MPEG-2 file on a local hard disk, an MPEG-4 file on a Web server, or a commercial DVD, just to name a few possibilities. Part of the development process for a CVP application necessarily includes an analysis of the content of the relevant video asset and the identification of segments that can be played back for a particular purpose. We strongly feel that CVP developers should agree on a common Video Asset Description (VAD) format for representing the segmentation of a video asset and various types of descriptive information about the video asset as a whole, as well as about its various segments. One advantage of such a format, as previously mentioned, would be to allow various developers to share the use of the same tools not only for the creation of VADs but also for their use. The focus of this article is a description of the VAD format that we have defined within the framework of MPEG-7, in cooperation with several partners, and its connection with another standard (the IEEE LOM).

Essentially, the proposed VAD format is an XML application conformant with a subset of MPEG-7. MPEG-7 is a recently published ISO standard (ISO/IEC 15938), and the subset is being developed by the authors of the present paper, in cooperation with several partners. See Manjunath, Salembier, and Sikora (2002) for background information on MPEG-7 in general, and consult the ISO website (http://www.iso.org) for the current status of various parts of the MPEG-7 standard (MPEG-7 has ten parts in various stages of development, as of this writing). Once a VAD, that is, a VAD-compliant XML document instance, has been created for a particular video asset, it can be used as the basis for a number of CVP applications, where each application must include some particular selective playback mechanism.

As mentioned earlier, until now the most common mechanism for selective playback of a video asset has been a remote control. Unfortunately, this mechanism lacks a fundamental feature of CVP as we envision it: repeatability. There must be some way to recreate a customized visual experience without constantly having a thumb poised over the buttons of a remote control and an eye focused on a timecode display; neither should a stand-alone video program be created. In a way, the various mechanisms described below can be thought of as a fancy remote control with a memory of which buttons to "push" and when.

Suppose your video asset is a file on a Web server and suppose the client machine has a SMIL-compatible player installed. Then your CVP application could be a SMIL file (http://www.w3.org/AudioVideo/; Kennedy & Slowinski, 2002). A SMIL tag for playing a 15-second clip from a video asset, starting ten seconds into the asset, might look like this:

```
<video src="toad-movie.rm" clipBegin="10s" clipEnd="25s"/>
```

Using SMIL is only one possible mechanism for achieving customized playback. SMIL is limited in that it currently does not support playback from commercial DVDs. Another mechanism is a DVD playback object that can be invoked using ECMAScript (the standard version of JavaScript) embedded in an HTML page. This method could be called Enhanced HTML. Microsoft provides such a DVD playback object (MSVidWebDVD). MSVidWebDVD is only available for Windows XP. The previous DVD playback object (MSWebDVD) is no longer supported (Microsoft, 2002). Another HTML+ECMAscript DVD playback object is available from VisibleLight corporation (http://www.visiblelight.com/ see "onstage DVD-hybrid" products).

Furthermore, VisibleLight enables CVP of DVDs from Visual Basic™ as well as providing plug-in tools for DVD control from PowerPoint™, and Director™. Such plug-ins for existing programming/authoring languages might become more common in the future.

Another type of mechanism for selective playback besides SMIL, Enhanced HTML, and plug-ins for various presentation and authoring tools, is to use an enhanced DVD player (software or hardware) that has the ability to play a sequence of video segments specified by the user. A list of separate video assets (such as independent video files on a hard disk) is sometimes called a playlist; however, we are using the term playlist in a different sense. Within the context of CVP, we use "playlist" to refer to a sequence of segments and attributes within a single video asset. This ability to play a sequence of (virtual) clips from within a single video asset is especially important when copyright restrictions do not allow portions of the video asset to be copied into separate files. It is also important when the video asset is segmented hierarchically, making the storing of all possible segments, some of which contain other segments, rather inefficient. Various playlist-aware DVD players are under development, and the DVD Association (http://www.dvda.org/) is considering the formation of a working group to define a standard format for video-clip playlists that define customized playback of a sequence of clips from within a DVD. Given a playlist-aware DVD player and a VAD, once a playlist has been defined using that VAD, a CVP application might consist simply of a player running a particular playlist.

The DVD Forum (http://www.dvdforum.org/forum.shtml) is working on a specification for advanced consumer electronics DVD players called eNav (Enhanced Navigation). At this point it is unclear exactly what features will be included in eNav, but it will hopefully include some support for CVP.

As personal video recorders (PVR) such as TiVo and ReplayTV gain popularity, the TVAnytime Forum (http://www.tv-anytime.org) is defining a specification for a second generation personal video recorder which has a customized video playback facility. One of this group's stated four-fold objectives is to "define specifications that will enable applications to exploit local persistent storage in consumer electronics platforms" (http://www.tv-anytime.org/about/index.html). More specifically the organization's working group on "Metadata Specification" seeks to establish "Segmentation Metadata" as a means "to edit content for partial recording and non-linear viewing" and to "navigate within a piece of segmented content" (http://www.tv-anytime.org/workinggroups/wg-md.html). Non-linear viewing corresponds to what we term CVP.

Given a widely accepted VAD format at some point in the future and a few widely used selective playback mechanisms, multiple tools should arise to facilitate the creation of VADs and the building of playback specifications from VADs. In addition, there should be various CVP applications for a particular playback mechanism. Then Customized Video Playback will be able to fulfill its potential. Of course, one key to the future of CVP is an adequate VAD format that gains wide acceptance. Given that standards are important to create critical mass and a propitious environment for development (Bush, 2002) and since our proposal is based on MPEG-7, we will provide an introduction to this standard. However, since the IEEE LOM is the basis for much current work in e-learning, we will first introduce the LOM and suggest that it is insufficient as the sole basis for the kind of VAD format that is needed to support CVP.

An Introduction to the IEEE LOM and Related Activities (ARIADNE, IMS, and SCORM)

In 1996, the Learning Technology Standards Committee (LTSC) was formed within the Institute of Electrical and Electronics Engineers (IEEE). The committee is chartered to "develop accredited technical standards, recommended practices and guides for learning technology" (http://ltsc.ieee.org/). The LTSC is divided into various working groups that each hold quarterly face-to-face meetings as well as teleconferences to accomplish their standards work. The LTSC working group of most interest here is the Learning Object Metadata (LOM) group.

During the timeframe from 1995 to 1997, there were a few groups working on metadata issues. An early project that served as a foundation for work by other groups was the Dublin Core Metadata Initiative (http://www.dublincore.org). Two groups that took the Dublin Core metadata work to further expand it were the Alliance of Remote Instructional Authoring and Distribution Networks for Europe (ARIADNE) (http://www.ariadne-eu.org/) and the Instructional Management Systems (IMS) project, later renamed the IMS Global Learning Consortium (http://www.imsglobal.org/). Both groups, ARIADNE and IMS, contributed their early metadata work to the LOM working group and all three combined their efforts to settle upon the specific elements that would comprise the final LOM standard. In 2002 the LOM group finished version 1.0 of the Learning Object Metadata standard that is known as IEEE 1484.12.1 (IEEE LTSC, 2002; see also http://standards.ieee.org).

In the view of many individuals in this field, one of the benefits of having an IEEE committee working on learning technology standards is that the IEEE is an accredited standards body (in the United States, standards bodies are accredited by ANSI, the U.S. national member of ISO). This status can give IEEE standards more permanence and sometimes more credibility than specifications emerging from industry consortia such as the IMS. On the other hand, due to the careful deliberations and the need to take into account many different perspectives that are sometimes competitive in nature, accredited standards bodies are known to generally take long periods of time to produce a final standard. For example, the LOM version 1.0 standard was roughly six years in the making, a very long time in the world of learning technologies, given the number of changes that can occur in that amount of time.

Recognizing the intentionally slow and deliberate nature of accredited standards bodies, other groups in the field of learning technology began taking the preliminary metadata element set and matching it with current technologies. The IMS produced the first XML binding of an early LOM metadata set in August, 1999. Since that time the IMS has updated its binding to reflect the progressively refined versions of the LOM information model that have occurred over time. This has served IMS members well, since many wanted to produce metadata records but had no specific guidance from the LOM group on how to represent the LOM information model using technologies such as XML. The Shareable Content Object Reference Model (SCORM) coming out of the Advanced Distributed Learning (ADL) initiative (http://www.adlnet.org/) was one such group. The SCORM references the IEEE LOM information model but utilizes the IMS LOM XML binding throughout its documentation and testing software. In June of 2002, the LOM working group formed an XML binding sub-group to work on an IEEE version of an XML binding for the LOM model. As of this writing, the IEEE XML binding of the LOM (1484.12.3) is still under revision.

An examination of the data elements of the LOM reveals that although it is well suited to describe a video asset for purposes of locating and retrieving the asset as a whole, it is entirely inadequate to describe a hierarchical segmentation and segment-by-segment annotation of a video asset. Therefore, we have turned to another standard, MPEG-7, as the basis for the VAD format.

One of the explicit design decisions made by the LOM working group with their information model and the IMS with their binding was to leave LOM records as open and flexible as possible. There is always a tradeoff between flexibility and interoperability. The more flexible the schema, the less interoperable it is and vice-versa. At some point in the future, it may be possible to extend the *technical* element of the LOM schema to provide the extra detail needed for customized video playback. The additional metadata elements utilized for this aspect of the LOM would come from MPEG-7.

An Introduction to MPEG-7

An ISO/MPEG project was organized in October 1996 and began building the requirements for what is now called the “Multimedia Content Description Interface” or MPEG-7. This was created for the purpose of describing previously existing content already compressed or to be compressed using current and future standards such as MPEG 1, 2 and 4. After the call for proposals in October 1998 and going through the initial standards development process, MPEG-7 entered the final editing phase in September of 2001 and has recently been published as international standard ISO 15938. MPEG is the common name for ISO Joint Technical Committee 1, Sub-committee 29, Working Group 11.

The need for a model to describe multimedia content in this age of increased digital content production, expanded storage capabilities, and enhanced content consumption venues became readily apparent to a group of content creators, distributors, and technologists. As a result, MPEG-7 has become **the** superset of multimedia content descriptions. It thus can then be used as a resource for efforts such as ours that are working to implement content description schemas. Where MPEG 1, 2, and 4 dealt primarily with compression standards, MPEG-7 exists as a descriptive mechanism for multimedia content. As Nack and Lindsay wrote:

> MPEG-7 … focuses on the standardization of a common interface for describing multimedia materials (representing information about the content, but not the content itself—"the bits about the bits"). In this context, MPEG-7 addresses aspects such as facilitating interoperability and globalization of data resources and flexibility of data management (1999, July-September, 69)

As can be seen from the final version of the "MPEG-7 Overview (version 8)" (Martinez, 2002), the specification itself covers a comprehensive range of functionality across audio-visual media. For example, one of the playback mechanisms mentioned earlier is the TV-Anytime technology being developed for personal video recorders, a technology that is developing with some level of mutual consideration of the work underway within the MPEG community. Indeed, "TV-Anytime uses the MPEG-7 Description Definition Language (DDL) to describe metadata structure as well as the XML encoding of metadata" (Cover Pages, 2002). On the MPEG-7 side of the developments we learn that:

> MPEG-7 does not define a monolithic system for content description but rather a set of methods and tools for the different viewpoints of the description of audiovisual content. Having this in mind, MPEG-7 is designed to take into account all the viewpoints under consideration by other leading standards such as, among others, TV-Anytime, Dublin Core, SMPTE Metadata Dictionary, and EBU P/Meta (Martinez, 2002, 4).

As a member of the INCITS L3.1 (http://www.incits.org) (the ANSI member organization that provides the United States delegation to MPEG), Brigham Young University (BYU) has been involved in the development of MPEG-7 since 2002, for use in the VAD project. Since the MPEG-7 community is collecting information and sample applications showing how the standard has been implemented, we have been able to present our included sample VAD (see Appendix) at an MPEG-7 Awareness Event. Our presentation was favorably received the MPEG-7 community, and we are working on finalizing the Core Description Profile (CDP) standard, which forms the basis of the VAD format. This "profile" (an MPEG term for subset) of the MPEG-7 elements found in parts 2, 3, 4, and 5 (mostly in Part 5) will be one of the profiles in MPEG-7 Part 9.

To insure applicability across a wide variety of types of audio-visual media and content types, MPEG-7 contains a broad range of descriptive power. It implements a Description Definition Language (DDL) based on XML that contains:

- The XML Schema structural language components,
- The XML Schema datatype language components, and
- The MPEG-7 specific extensions (Martinez, 2002, 9).

In addition, the MPEG-7 specification provides for the description of visual attributes and audio attributes, as well as Multimedia Description Schemes (MDS). MPEG-7 contains descriptors for such visual information as color, texture, shape, motion, localization, and face recognition. Its audio framework "deals with low-level Descriptors, for audio features that cut across many applications (e.g., spectral, parametric, and temporal features of a signal), and high-level Description Tools that are more specific to a set of applications" (Martinez, 2002, 9). These Description Tools from the MPEG-7 audio framework work in conjunction with the Multimedia Description Schemes (DSs) to provide multiple, descriptive, data streams for the media asset:

> The DSs provide a standardized way of describing in XML the important concepts related to AV content description and content management in order to facilitate searching, indexing, filtering, and access. The DSs are defined using the MPEG-7 Description Definition Language (DDL), which is based on the XML Schema Language, and are instantiated as documents or streams. The resulting descriptions can be expressed in a textual form (i.e., human readable XML for editing, searching, filtering) or compressed binary form (i.e., for storage or transmission) (Martinez, 2002, 10).

With the possible addition of a few new metadata descriptors (called "tools") MPEG-7 seems to be adequate as the basis for the VAD format.

Defining the VAD Format and Mapping it to the LOM

In the e-learning community, most efforts to create formats for shareable and reusable content objects are based on the IEEE Learning Object Metadata (IEEE LOM) described in earlier in this paper. As previously mentioned, because the IEEE LOM does not include all the elements necessary to describe a video asset, such as a mechanism for describing a hierarchical segmentation of a film or other video asset, we had to look elsewhere. The ISO MPEG-7 standard as described briefly above seems to include all or nearly all the elements needed to describe a video asset, but MPEG-7 is not well known within the e-learning community, and thus a purely MPEG-7 approach would not fit into the larger e-learning picture of sharing descriptions of various kinds of educationally-relevant assets using a common metadata framework. Therefore, we decided to combine the strengths of the IEEE LOM and MPEG-7 standards by defining the VAD in terms of MPEG-7 and defining a mapping between the VAD format and the elements of the LOM.

From the standpoint of solid principles of metadata development, our VAD format is an "application profile" of MPEG-7, a necessary reality because "no single metadata element set will accommodate the functional requirements of all applications" with exactly the same elements (Duval, Hodgins, Sutton, and Weibel; 2002; April). The methodology we envision for coordinating the VAD format and the LOM consists of the following steps:

1. Develop a profile (subset) of MPEG-7 that is adequate for the VAD format. This profile is called CDP, and this step is substantially complete.

2. Compare the IEEE LOM elements and the VAD format to develop a map of correspondences.

3. Devise a subset of the LOM that allows a bi-directional mapping to VAD.

So far, we are still working on step 1. As of early 2004, the CDP profile is in the Committee Draft stage as part of MPEG-7 Part 9. Note that profiles of MPEG-7 must be sponsored by several MPEG member organizations. Current sponsors of the CDP profile include BYU, Motorola, Inc., the Japanese national broadcasting corporation (NHK) and others. Additional participants in the VAD project are welcome.

In order to facilitate Step 2, a formal liaison has been established between MPEG and the LTSC. The comparison of LOM and VAD is scheduled to be accomplished during 2004, now that the composition of the VAD format is fairly stable. Step 3 will follow naturally from Step 2. Once these three steps have been completed, it should be possible to begin with an MPEG-7 CDP file and generate a LOM description or to begin with a LOM description and generate a partial CDP description.

Conclusions

One way to understand the significance of this work is to understand the fundamental difference between the approaches used in IEEE LOM and MPEG-7. Our determination is that the first is well-suited to assure retrievability, while the second is useful for providing a technical description of the asset.

An analogy would be the Multiple Listing Service (MLS) widely used across the United States for cataloging available homes and other real estate. Using the MLS, it is possible to easily find a home that corresponds to a particular set of criteria. To build a copy of or to remodel that house, however, would require a lot more information, such as that contained in a set of architect’s plans. Where IEEE LOM plays the role of the MLS for educational content, MPEG-7 provides the equivalent of the architect’s plans. In other words, for certain applications, the IEEE LOM will serve the useful function of locating content, and MPEG-7 will open up the use of the content once it is located.

An obvious approach to combining VAD and LOM is to use LOM-compliant data to retrieve a video asset in an educational setting. The retrieved asset would include a VAD-compliant XML file to describe its internals.

CVP (Customized Video Playback) is important not only to education but also to home use. In either case, it will be impossible for any benefit to be derived without appropriate specifications for content modeling as well as tools for customization and personalization of the content. The development and adoption of the VAD format would facilitate CVP. Indeed, there is a good chance that intellectual property restrictions with respect to various approaches to CVP are such that distributing a VAD to facilitate personal building of playlists will be the only viable means to achieve CVP within a legal framework for content distribution when the video asset is an off-the-shelf film on DVD.

Two examples of developments in the area of CVP are (1) the Open Video Project, sponsored by the University of North Carolina at Chapel Hill and (2) MovieLearn, a division of CleverLearn.[2] Although these efforts do not address the specific issue of CVP, VAD, and playlists, they do address the need for the technical capabilities such as metadata that will make these a reality.

The Open Video Project is "an on-going effort to develop an open source digital library that can be used by researchers, teachers, students and the public" (Marchionini, p. 37). Recognizing the need for metadata for locating video segments, developers in this project also recognize the value of MPEG-7 as "a standard schema to create video metadata." (Yang, 2003, p. 1). The MovieLearn company also recognizes the need for a standard metadata format and will look into MPEG-7 (private communication at the occasion of the TESOL convention in Long Beach, California, April 2004).

Finally, although the potential significance of the VAD standard to CVP in education and the home is great, and although we have moved a long way down the road to developing a viable approach to VAD development, there is a great deal of work to be accomplished. Therefore, to insure that our work is available as widely as possible, we have set up a Web site that contains various files that document the VAD format: http://www.vad.byu.edu. In particular, this website, along with the MPEG-7 standard itself, provides information necessary to a detailed understanding of the sample VAD file provided in the appendix to this paper.

Input on the VAD approach to Customized Video Playback is solicited from any and all interested individuals.

---

[2] See the company's Web site at http://movielearn.com/en/ .

REFERENCES

Bush, M. (2002, November/December). Connecting Instructional Design to International Standards for Content Reusability. *Educational Technology,.42*(6), 5-13.

Cover Pages (2002). TV-Anytime Forum Publishes Metadata Specification for Network Broadcast Content. http://xml.coverpages.org/ni2001-10-02-c.html .

Dublin Core Metadata Initiative (2002). Web site: http://www.dublincore.org/.

Duval, E., Hodgins, W., Sutton, S. and Weibel, S. L. (2002, April). Metadata Principles and Practicalities. *D-Lib Magazine*, 8 (4). http://www.dlib.org/dlib/april02/weibel/04weibel.html

Gale, L. E. (1989). Macario, Montevideodisco and Interactive Dígame: Developing interactive video for Language instruction. In W. Flint Smith (ed.), *Modern Technology in Foreign Language Education: Applications and Projects*. pp. 235-248. Lincolnwood, IL: National Textbook Company.

Gareis, E. (1997). Movies in the Language Classroom: Dealing with Problematic Content. *TESOL Journal*, *6*(4), 20-23.

IEEE LTSC (2002). *Draft Standard for Learning Object Metadata*. IEEE Learning Technology Standards Committee (LTSC) P1484.12 Learning Object Metadata (LOM) working group. http://ltsc.ieee.org/doc/wg12/LOM_1484_12_1_v1_Final_Draft.pdf

Kennedy, T. and Slowinski M. (2002) *SMIL: Adding Multimedia to the Web*. Indianapolis, Indiana: Sams Publishing. (Synchronized Multimedia Integration Language.)

Manjunath, B., Salembier, P., and Sikora T. (2002) *Introduction to MPEG-7: Multimedia Content Description Interface*. West Sussex, England: John Wiley & Sons, Ltd.

Marchionini, G. (2003). Video and Learning REdux: New Capabilities for Practical Use. *Educational Technology, xx*(x), xx-xx.

Martínez, J. M. (Ed) (2002). *Coding of Moving Pictures and Audio, MPEG-7 Overview*. ISO/IEC JTC1/SC29/WG11N4980. MPEG Alliance Web Site. http://www.mpeg-industry.com/mp7a/w4980_mp7_Overview1.pdf.

Microsoft (2002). MSDN Web Site. http://msdn.microsoft.com/library/default.asp?url=/library/en-us/directx9_c/directx/htm/mswebdvdobject.asp (If this URL is no longer available, go to http://msdn.microsoft.com/; then search for "mswebdvd activex control".)

Nack, F. and Lindsay, A. T. (1999, July-September). Everything You Wanted to Know About MPEG-7: Part 1. *IEEE MultiMedia*, *6*(3), 65-77. http://lass.cs.umass.edu/~shenoy/courses/spring01/papers/ieeemm-mpeg7.pdf

-- (1999, October-December). Everything You Wanted to Know About MPEG-7: Part 2. *IEEE MultiMedia*, *6*(4), 64-73.

Negroponte, N. (1994, August). Prime Time Is My Time: The Blockbuster Myth. *Wired 2.08*, 2(8). http://www.wired.com/wired/archive/2.08/negroponte.html

Yang, M., Mu, X., & Marchionini, G. (2003). VIVO -A Video Indexing and Visualization Organizer. Interaction Design Laboratory, School of Information and Library Science, University of North Carolina at Chapel Hill. Retrieved April 9, 2004, from http://www.open-video.org/ovadmin/yang_jcdl_2003.pdf

Appendix (sample VAD document)

```
?xml version="1.0" encoding="iso-8859-1"?>
<Mpeg7 xmlns="urn:mpeg:mpeg7:schema:2001"
       xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
       xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001
       C:./CDP-schema_FullG_ABCD.xsd">

<!-- Mpeg7 is the root element of every MPEG-7 document instance,
       and a document instance that conforms to an MPEG-7 profile, such as CDP,
       must also validate against the master MPEG-7 schema -->

<DescriptionMetadata>
  <LastUpdate>2004-03-20</LastUpdate>
</DescriptionMetadata>

<!-- DescriptionMetada is meta-metadata, that is, data about the MPEG-7 description,
       rather than about the video asset in question -->

<Description xsi:type="ContentEntityType">
 <MultimediaContent xsi:type="AudioVisualType">
  <AudioVisual>
   <MediaInformation>
    <MediaProfile>
      <MediaFormat>
        <Content href="urn:mpeg:mpeg7:cs:ContentCS:2001:2">
          <Name>Audiovisual</Name>
        </Content>
        <Medium href="urn:mpeg:mpeg7:cs:MediumCS:2001:1.3">
          <Name>DVD</Name>
        </Medium>
      </MediaFormat>
    </MediaProfile>
   </MediaInformation>
  </AudioVisual>
 </MultimediaContent>
</Description>

<!-- After the DescriptionMetadata element, a CDP file consists of a sequence of Description elements;
      the above Description provides information about the medium of the video asset, in this case a DVD -->

<Description xsi:type="ContentEntityType">
  <MultimediaContent xsi:type="VideoType">
    <Video id="MainTitle">
      <TemporalDecomposition>

        <VideoSegment id="chapter1">

          <TemporalDecomposition>

            <VideoSegment id="chapter1scene1">
```

```
        <TextAnnotation type="description">
          <FreeTextAnnotation>opening credits; music; village aerial view</FreeTextAnnotation>
        </TextAnnotation>
        <MediaTime>
          <MediaTimePoint>T00:00:00</MediaTimePoint>
          <MediaDuration>PT1M24S</MediaDuration>
        </MediaTime>
      </VideoSegment>

      <VideoSegment id="chapter1scene2">
        <TextAnnotation type="description">
          <FreeTextAnnotation>entering church; bells; Count introduced</FreeTextAnnotation>
        </TextAnnotation>
        <MediaTime>
          <MediaTimePoint>T00:01:24</MediaTimePoint>
          <MediaDuration>PT0M20S</MediaDuration>
        </MediaTime>
      </VideoSegment>

           <!-- remaining scenes go here -->

     </TemporalDecomposition>

    </VideoSegment> <!-- Remaining chapters go here -->
   </TemporalDecomposition>
  </Video>
 </MultimediaContent>

<!-- The above MultimediaContent element defines a hierarchical segmentation of the video asset ,
     and the one  below segments the audio portion of the asset, in this case trivially into one long segment -->

 <MultimediaContent xsi:type="AudioType">
  <Audio id="english_full">
   <MediaTime>
    <MediaTimePoint>T00:00:00</MediaTimePoint>
    <MediaDuration>PT01H57M20S</MediaDuration>
   </MediaTime>
  </Audio>
 </MultimediaContent>
</Description>
/Mpeg7>
```

20 May 2004

Addendum to CVP article

Customized Video Playback: Proprietary Solutions Vs. Open Standards?

During the review and editing period for the present article, a significant new CVP-enabled product was introduced into the consumer market: a ClearPlay-enabled DVD player (model DRC232N) from RCA, available at Wal-Mart stores and from www.walmart.com. This DVD player includes non-volatile memory for storing video asset descriptions and a mechanism for defining and running playback specifications based on these descriptions. A ClearPlay video asset description, called a filter, documents segments of video on a particular DVD. Each of these segments can potentially be skipped or muted during customized video playback. Each segment has attributes that describe the content of that segment, such as "graphic violence" or "foul language". The user sets flags in the player that match with the attributes to determine which segments will actually be skipped or muted, thus creating a playback specification on the fly.

The user purchases filters on-line from the ClearPlay website and downloads them to a personal computer as a file of type CPF (presumably for ClearPlayFilters). The user then burns the CPF file to a CD and inserts that CD into the ClearPlay-enabled DVD player. At that point, the user can load selected filters from the CD into the memory of the DVD player. The player can hold approximately 300 filters at a time, each specific to one DVD. If the memory becomes full, the user can delete some filters to make room for others. Once the filter set in the player has been updated, the CD can be removed to make room for a DVD in the player's drive.

With the ClearPlay feature off, the device acts like any other commercial DVD player. With the ClearPlay feature on, when a DVD is loaded, the player searches its memory for a filter that matches it and, if found, creates a playback specification based on the player settings and customizes the playback of the DVD.

Although the primary purpose of the ClearPlay feature is to facilitate parental control by filtering out content that parents deem inappropriate for their children to see or hear, it obviously has additional applications in education. For example, could be used to create versions of documentaries that are shortened to fit into an alloted portion of a class period and tailored to include only the most relevant segments. It could also be used in a language class to show a sequence of segments that illustrate particular points of vocabularly, culture, or grammar. This could be accomplished by creating a custom ClearPlay filter and mentally redifining the meaning of some of the flags (e.g., a violence flag could be used to skip all segments not relating to a particular culture point).

Unfortunately, the potential for using ClearPlay-enabled DVD players in education is severely limited by the fact that a ClearPlay CPF file is not a text file, such as an XML file and ClearPlay corporation has opted not to release the specification for creating CPF files (source: e-mail from ClearPlay technical support dated May 19, 2004). Thus, ClearPlay becomes an negative example of the importance of using a standards-based approach to Customized Video

Playback, with video asset descriptions in XML so that anyone can create them. It may be give a commercial advantage to ClearPlay to use a proprietary format, but this approach is not ideal for RCA or the consumer. One possible compromise would be for DVD player companies to build players that accept both ClearPlay CPF files and standard playlists in an XML-format defined by the DVD Association, as suggested in the present article.

Readers who are convinced of the open standards approach described in the present article can take three kinds of action: (1) contact ClearPlay to ask that they somehow make available the format specification for CPF files (even if it requires a license); (2) contact DVD player manufacturers to explain that their market would be expanded if they included an open-standard format for playlists, in addition to the proprietary ClearPlay format, in future CVP-enabled DVD players; and (3) support a non-profit organization that promotes user rights in digital media, such as the Electronic Frontier Foundation (http://www.eff.org) or the ViewerFreedom Foundation (http://www.viewerfreedom.org).