Hypervideos and Interactive Multimedia Presentations

BRITTA MEIXNER, University of Passau, Germany

Hypervideos and interactive multimedia presentations allow the creation of fully interactive and enriched video. It is possible to organize video scenes in a nonlinear way. Additional information can be added to the video ranging from short descriptions to images and more videos. Hypervideos are video-based but also provide navigation between video scenes and additional multimedia elements. Interactive multimedia presentations consist of different media with a temporal and spatial synchronization that can be navigated via hyperlinks. Their creation and description requires description formats, multimedia models, and standards—as well as players. Specialized authoring tools with advanced editing functions allow authors to manage all media files, link and arrange them to an overall presentation, and keep an overview during the whole process. They considerably simplify the creation process compared to writing and editing description documents in simple text editors. Data formats need features that describe interactivity and nonlinear navigation while maintaining temporal and spatial synchronization. Players should be easy to use with extended feature sets keeping elements synchronized. In this article, we analyzed more than 400 papers for relevant work in this field. From the findings we discovered a set of trends and unsolved problems, and propose directions for future research.

Categories and Subject Descriptors: H.1.2 [**Information Systems**]: Models and Principles—*User/machine systems*; H.5.1 [**Information Systems**]: Information Interfaces and Presentation—*Multimedia information systems*; H.5.4 [**Information Systems**]: Information Interfaces and Presentation—*Hypertext/hypermedia*; I.7.2 [**Computing Methodologies**]: Document and Text Processing—*Hypertext/hypermedia*

General Terms: Human Factors

Additional Key Words and Phrases: Hypervideo, interactive multimedia presentation, interactivity, nonlinearity, player, authoring tools, data formats, standards

ACM Reference Format:
Britta Meixner. 2017. Hypervideos and interactive multimedia presentations. ACM Comput. Surv. 50, 1, Article 9 (March 2017), 34 pages.
DOI: http://dx.doi.org/10.1145/3038925

1. INTRODUCTION

The history of moving pictures goes back to 1891 when the first film camera was invented by Thomas Edison [Monaco 2009, p. 641]. But it took nearly 100 years until films or videos became interactive in 1989 (first hyperlinked videos “Video Finger” [Watlington 1987] and “Elastic Charles” [Brondmo and Davenport 1989]). One of the earliest research papers on hypervideo was published by Sawhney et al. [1996] about

This work is supported by the European Social Fonds and the Bayrisches Staatsministerium für Wissenschaft, Forschung und Kunst (Bavarian State Ministry for Sciences, Research and the Arts) and the Bundesministerium für Bildung und Forschung (German Federal Ministry of Education and Research) (BMBF) under project number 03V0633. The research was done at the University of Passau, the article was written at FX Palo Alto Laboratory, Inc., 3174 Porter Drive, Palo Alto, CA 94304, USA.
Author’s address: Centrum Wiskunde & Informatica, Science Park 123, 1098 XG Amsterdam, The Netherlands; email: meixner.britta@gmail.com.

DOI: http://dx.doi.org/10.1145/3038925

a hypervideo prototype called HyperCafe. “HyperCafe is an experimental hypermedia project we developed to illustrate general hypervideo concepts. HyperCafe places the user in a virtual cafe [...]” [Sawhney et al. 1997].

Since then, the number of mobile devices with integrated cameras providing sufficient qualities has grown. This led to a growth of the number of videos in social media websites like Facebook and on video platforms like YouTube and Vimeo. According to the Cisco Visual Networking Index (VNI), the usage of online video will grow further in the coming years [Cisco 2014]. “Because mobile video content has much higher bit rates than other mobile content types, mobile video will generate much of the mobile traffic growth through 2018. [...] Of the 15.9 exabytes per month crossing the mobile network by 2018, 11 exabytes will be due to video. Mobile video represented more than half of global mobile data traffic beginning in 2012, indicating that it is having an immediate impact on traffic today, not just in the future” [Cisco 2014]. According to Shipman et al. [2003] “[...] the growing use of digital cameras (video and “still”) to capture short video snippets makes home authoring of interactive video an application that is likely to emerge.” With new technologies and improvements in standards, transmission hardware, processors, internal storage, new methods of programming, and reliable software libraries, it is possible to provide various different ways of interaction with multimedia contents nowadays. Besides requesting additional information, nonlinear navigation in a video via a table of contents or keyword search is possible. This allows one to create meaningful and appealing video presentations. Use cases are virtual tours through houses or cities where the viewer can decide where to go [Meixner et al. 2010], training videos for physiotherapy [Meixner et al. 2014, 2015], interactive video stories with possible different endings, e-learning [imc AG 2010], or mobile help systems that guide one through a troubleshooting process [Tonndorf et al. 2012].

However, compared to a traditional linear video, the editing and postproduction process of hypervideos or interactive multimedia presentations is more difficult because of the various media types that have to be arranged in space and time while providing a link structure to navigate them. Authoring tools are needed for creating these presentations that support authors in media management and media linking. Finished projects are then provided in players with an extended range of navigational and interactive functions compared to traditional video players. Depending on the framework, a defined data exchange format may be necessary. Authoring tool, player, and data format have to provide features to seamlessly integrate different types of media and synchronize them in space and time. In addition, navigational and interactive features require further attention. Not only do they have to be synchronized with the presentation in time and space, they also have to be kept consistent and updated during the lifecycle of a project.

This article is structured as follows: We first outline the survey scope in Section 2. We then give an overview of description formats, models, and standards in Section 3 and of authoring tools and players in Section 4. In Section 5, observations and trends that can be drawn from the survey are outlined and discussed. The article ends with a conclusion and an outlook on future work in Section 6. Research in the proposed areas may help to make hypervideos and interactive multimedia presentations more popular in the future.

2. SURVEY SCOPE

In this article, we analyze and compare description formats, models, and standards as well as authoring tools and players for hypervideos and interactive multimedia presentations. The search for literature for this survey was driven by the following research questions:

—What different types of videos providing interactive and nonlinear functions do exist and how can different forms be delimited from one another?
—What description formats, models, and standards exist that are capable of linking between and within media files and describing interactivity and nonlinearity?
—Which authoring tools and players do exist in these areas?

2.1. Methodology

The data collection and analysis process of this article closely follows the instructions for systematic literature reviews [Brereton et al. 2007; Kitchenham and Charters 2007]. Data collections and analysis were performed by the author and checked for completeness and correctness by two other reviewers. Initially we conducted a literature search in the following databases: dblp computer science bibliography (http://dblp.org/search/), Google Scholar (https://scholar.google.com/), the ACM Digital Library (http://dl.acm.org/), the IEEE Xplore Digital Library (http://ieeexplore.ieee.org/Xplore/home.jsp), and ScienceDirect (http://www.sciencedirect.com/). In a second step we checked the references of papers found in the initial search for further relevant work. In addition, we searched the Web for available (nonscientific) tools.

Used search terms were "interactive video, non-linear video, nonlinear video, hypervideo, hyper-video, clickable video, multimedia presentation, video hotspot, annotated video," and variations thereof. We did not search for the terms "multimedia document, hypermedia document, adaptive hypermedia," and combinations. Altogether, we found more than 400 papers[1] for further analysis, which were then evaluated regarding the following inclusion and exclusion criteria for this work:

—**Inclusion criteria:**
  —Terms "hypervideo," "hyper-video," or "multimedia presentation" in title, abstract, or body.
  —Terms "interactive video," "non-linear video," "nonlinear video," "clickable video," "video hotspot," or "annotated video" in title, abstract, or body AND features described in the paper fit definitions of "hypervideo" or "interactive multimedia presentation."
  —Publications in conference or workshop proceedings, journal articles, technical reports, online and downloadable software from the Web.
  —Language: English.
  —Any publication date.
  —(Additional for Formats/Models/Standards): Any description format, but at least one diagram describing the model or pseudocode/code of the format is given.
  —(Additional for Software): At least one screenshot of the software.
  —(Additional for Software): Tools published after 1998 (publication of SMIL 1.0).

—**Exclusion criteria:**
  —Work dealing with simpler video types like "interactive video," "non-linear video," "clickable video," "video hotspot," or "annotated video."
  —Work from the areas of "multimedia documents," "hypermedia documents," and "adaptive hypermedia."
  —Work from the areas of "video browsing," "video search," "multi-view video," "video annotation using video analysis," and "social video/social TV."
  —Work about multimedia presentations without interactivity exceeding VCR actions (only linear playback), so-called basic multimedia presentations.

[1]Due to the iterative process, the precise number of papers that were analyzed cannot be stated.

We analyzed work with the terms "interactive video," "non-linear video," "nonlinear video," "clickable video," "video hotspot," or "annotated video" in title, abstract, or body, because there is no clear understanding of how the different terms are delimited. Authors may use one term, but actually describe another type of video. We analyzed all papers and extracted key features to find and describe differences between the different types of presentations. Finding that there is no common understanding of the terms, we provide working definitions for the terms "hypervideo" and "multimedia presentation" in the following section.

2.2. Definition of "Hypervideo" and "Interactive Multimedia Presentation"

With different usages of the terms "hypervideo" and "interactive multimedia presentation" in the literature, we first give working term definitions. To find these definitions, we analyzed works for their definitions and searched for similarity throughout them to find a common denominator.[2] Both concepts are subsets of hypermedia (which again is a subset of multimedia), meaning that different types of media are linked via hyperlinks. No further limitation due to structure or layout applies.

—**Interactive Multimedia Presentation:** Most of the term definitions for "multimedia presentation" have three elements in common: static and continuous media, temporal relationships (sometimes leading to synchronization issues), and spatial relationships. It can be noted that videos, audio files, text, and images are part of many multimedia presentations (e.g., Adjeroh and Lee [1995] and Nimmagadda et al. [2009]). We define the term "interactive multimedia presentation" following Nimmagadda et al. as follows:

Definition 2.1 (*Interactive Multimedia Presentation*). "Multimedia presentations are collections of different media files [...] like text, images, videos, and animations with different resolutions, durations, and start-times. [...] The layout of multimedia presentations is defined by the locations and the start times of the objects" [Nimmagadda et al. 2009]. Prerendered static multimedia presentations allow VCR actions while interactive multimedia presentations feature additional navigational structures.

The most important and distinctive features and research areas dealing with (interactive) multimedia presentations are as follows:

—*Synchronization issues and temporal relationships:* The definition of temporal relationships (parallel, sequential, time-shifted) between media elements may lead to conflicts during playback. These have to be resolved (e.g., Blakowski et al. [1992] and Abdelli [2008]).
—*Spatial relationships and layout:* Fixed or varying spatial layouts with overlapping media elements need to be defined, which may also change over time (e.g., Chung and Shih [1997] and Nimmagadda et al. [2009]).
—*User interactions:* "A multimedia presentation whose content is predefined before rendering is called a static multimedia presentation. A multimedia presentation whose content is composed dynamically during runtime is called a dynamic multimedia presentation" [Kleinberger et al. 2008]. Dynamic composition may be based on user interaction or previously defined conditions that are evaluated during playback. Static multimedia presentations allow only VCR actions, while dynamic multimedia presentations enable the user to choose certain contents directly in an overlying navigational structure (e.g., Chung et al. [1995] and Abdelli [2008]).

[2]The term *element* is used to describe an arbitrary object or a person in the video.

—**Hypervideo:** Hypervideos are found in different forms in the literature. They either provide nonlinear navigation between video scenes (homogeneous hypervideo) or they consist of linear videos with additional information (heterogeneous hypervideo). Hybrid forms having nonlinear navigation between scenes and at the same time providing additional information also exist. Closely following the definition of Stahl et al. [2005], we define the term "hypervideo" as follows:

Definition 2.2 (*Hypervideo*). Hypervideo is defined as video-based hypermedia that combines nonlinear video structuring and dynamic information presentations. Video information is linked with different kinds of additional information (like texts, pictures, audio files, or further videos). Users can interact with sensitive regions having spatial and temporal characteristics. These are shown in video scenes or separately but are synchronized with the video. They provide access to additional information (heterogeneous hypervideo) or allow jumps to other scenes (homogeneous hypervideo). Hyperlinks build a graph between main video scenes and additional information, if available.

The most important and distinctive features and research areas dealing with hypervideos are as follows:

—*Hyperlinks (to scenes or additional information):* Hyperlinks are usually represented by hotspots or sensitive regions that depend on space and time in the main video (e.g., Sawhney et al. [1997] and Mujacic and Debevc [2007]).
—*(Linear) heterogeneous hypervideos:* Video and additional information like text, images, audio files, animations, and other videos (e.g., Correia and Chambel [1999] and Bochicchio and Fiore [2005]).
—*Homogeneous hypervideos:* Links between videos (e.g., Sawhney et al. [1996] and Seidel [2011]), illustrated graph structure [Sawhney et al. 1996], detail-on-demand hypervideo (one link at a time to jump to additional (explanatory) video, returns to main video automatically) (e.g., Doherty et al. [2003] and Shipman et al. [2005]).
—*Hybrid hypervideos:* Hybrid hypervideos have a graph structure linking videos as well as links from videos to additional information (e.g., Chambel and Guimarães [1999] and Sadallah et al. [2011]).

Comparing hypervideos and interactive multimedia presentations, it can be stated that hypervideos are more interactive regarding navigation, but interactive multimedia presentations provide more possibilities to combine and arrange media. The only video-centric form is homogeneous hypervideo. All other forms combine different media like videos, images, audio, text, and other multimedia files. Spatial alignment and synchronization play a big role in multimedia presentations, while it is less considered for hypervideos. Table I shows a comparison of the different aspects.

2.3. Fundamentals and Terms

Before we start with our literature review, we first explain the terms that we frequently use to describe authoring tools, players, standards, and formats:

—**Used Medium:** It describes which kinds of media are used and if one of these media are used primarily during playback.
  —*Video:* The only used medium is video. When we use the term "linear video" in the remainder of this work, we want to emphasize that one single video file is used during playback (resulting in the same playback length for each playback without user interaction). When we use the term "video scenes," more than one video file is used. Video scenes are linked in an underlying graph or tree structure.

Table I. Comparison of Hypervideos and Multimedia Presentations

Type	Media	Interactivity	Temporal synchronization	Spatial layout
Homogeneous hypervideo	video	navigate in graph	no	no
Heterogeneous hypervideo	linear video and media	jump to/interact with media	between main video and media	yes
Hybrid hypervideo	videos and media	navigate in graph and jump to/interact with media	between main video and media	yes
Static multimedia presentation	media	VCR actions	yes	yes
Dynamic / interactive multimedia presentation	media	VCR actions and content selection	yes	yes

—*Media Files:* No main medium exists using media files. Used file types are text, image, audio, and video. The media files may be shown in sequence or parallel.

—**Further Characteristics:**
- —*Basic VCR Actions:* Basic VCR actions are play and pause, as well as sound volume control. Fast forward and fast backward are also considered basic VCR actions in most papers.
- —*Hyperlinks:* Hyperlinks are references between elements that are in a direct relation to each other. A user selection of the hyperlink in the source document loads another document or invokes a certain action.
  - —*Hotspots:* Hotspots are a limited form of hyperlinks. They are interactive areas (e.g., buttons or marked objects) in a video that invoke an action (e.g., open additional information, navigate in the structure) after a user interaction.
  - —*Choice Elements:* Choice elements are panels with buttons or menu-like structures that are displayed at a certain time or at the end of a scene. Viewers have to choose one option to proceed with the video. They are a special form of hyperlinks following the main narrative during playback.
- —*(Underlying) Nonlinear Structure:* The video/presentation has an underlying nonlinear structure like a graph or a tree. Scenes can be watched only according to paths along the edges in the underlying structure, assuming that media elements are represented as nodes. This results in *alternative playback paths* where the viewer selects a path through the structure that depends on her/his decisions and may vary from playback to playback.
- —*Overlying Navigational Structure:* Additional navigational structures allowing jumps to certain elements in the underlying nonlinear structure can be provided for further navigation. Possible are, for example, a table of contents, a keyword search, or other features providing overview.
- —*Additional Information:* Additional information are media objects that provide further information about topics in a main medium. They may either be invoked by the viewer or appear in predefined areas at certain points in time. An *annotation* is additional information displayed with a main medium. It consists of an anchor attaching it to the main medium and a body. The body of an annotation may be a text (plain or formatted), a picture, a graphic, a video, an audio file, an animation, or any other kind of medium that can be shown in a player.
- —*Spatial Layout:* The spatial layout characteristic describes how single media elements are arranged. They may either be placed in fixed areas or variable on a canvas. The spatial layout may change over time.

2.4. Related Fields

As already noticed by Schöffmann et al. [2015], there is confusion about the categorization of work in different fields related to video interaction, because not all researchers use the same classification of their works. The types of video described hereafter have certain interactive or nonlinear navigation features. They are, however, usually less complex than hypervideos or interactive multimedia presentations but may be seen as simplified versions thereof in literature.

—**Clickable Videos** are the simplest type of enriched videos. They consist of a single linear main video and hotspots [Clickable Video 2012]. Interaction with one of these hotspots makes available additional information, mainly text, images, and links [Incisive Interactive Marketing LLC and Wegert 2006], as an overlay in the video area or on an additional side region [VideoClix Technologies Inc. 2012; Clear-Media 2012]. Hotspots can have a different appearance ranging from small icons to outlines of an object in the video. They may move as the element in the video moves [ReelSEO 2011]. Extended timelines may allow the viewer to jump to a point in time where a hotspot is displayed [VideoClix Technologies Inc. 2012; WireWax ltd. 2012]. Clickable videos are mainly used for monetizing products or services in the Internet.

—**Interactive Video** is mainly based on linear videos and rarely considers other types of media. Basic interactive functions are play, pause, stop, fast-forward, and rewind (e.g., Zheng and Atiquzzaman [2005] and Leung and Chan [2003])—possibly at different speeds (e.g., [Parsloe 1983; Fei et al. 2005])—as well as jumps forwards and backwards [Fei et al. 1999, 2005; Liao and Li 1998]. These are extended by more complex functions changing either the main video presentation or the scene sequence based on user interaction. Interactive videos may be extended with additional information. "The basic idea of interactive video is to provide more complex operations and feedback to users" [Chang et al. 2008]. The main video is altered, that is, "different view angles, or different zoomed-in and slow-motion factors" [Fernandez et al. 2012], "zooming into individual frames" [Naman and Taubman 2011], and "resolution scalability, progressive refinement (or quality scalability), spatial random access, and highly efficient compression" [Naman and Taubman 2011] are provided. Furthermore, "the user sets the pace, skimming through familiar material and pausing to review new or difficult ideas" [Parsloe 1983]. A main component of interactive videos is a browsing functionality that enables a user to access a linear video in a nonlinear way. After partitioning a video into smaller segments [Mackay and Davenport 1989; Chang et al. 2007], single scenes can be omitted [Chang et al. 2007] or jumps to specific video parts are possible [Zhang et al. 2006]. Zhang et al. "allow proactive and random access to video content based on queries or search targets" [Zhang et al. 2006] in interactive videos. As a result, "users can select or play a segment with minimal search time" [Zhang et al. 2006]. Mackay and Davenport state that it is possible to increase the level of interactivity "by allowing users to build, annotate, and modify their own environments" [Mackay and Davenport 1989].

—**Nonlinear Videos** consist of a set of scenes or video sequences, which are arranged in a library or as a walk-through in a graph structure [Kozuch et al. 2000]. Selections in video libraries are classified as "non-real-time, non-linear video applications" and walk-throughs are classified as "real-time, non-linear application" [Kozuch et al. 2000]. Selection elements are provided to either create a video selection from the library or to select a path in the graph, leading to an individualized presentation flow (e.g., Gotz [2006] and Hausenblas [2008]). Jumps from one presentation scene/part to another may be possible [Carlsson et al. 2008; Yeung et al. 1996]. Video walk-throughs are based on a tree or graph pattern and are thus more structured than video libraries. These structures allow the implementation of "different endings depending on the user interactions taking place during story consumption" [Spaniol et al. 2006]. Parallel sequences of frames allow dynamic selection of a branch during playback [Zhao et al. 2007]. Additional information and functions to control the playback reception are not part of nonlinear videos.

—**Basic Multimedia Presentations:** Nearly all tools for basic multimedia presentations allow the combination of text, images, audio files, and videos—often described as media files or media elements. Navigation in the presentations can be performed using VCR actions or jumps on a timeline. Basic multimedia presentations can be created with SMIL Builder [Bouyakoub and Belkhir 2011], GRiNS [Bulterman et al. 1998], SMILAuthor [Yang and Yang 2003]/SMILAuthor2 [Yang et al. 2008], the Synchronization Editor [Blakowski et al. 1992], TYRO [MacNeil 1991], MPRES Author [Wong et al. 1997], Java-Assisted SMIL (JAS) [Dai et al. 2006], SIMPLE [Murthy et al. 2006], Popcorn Maker [Mozilla Corporation 2013], and the tools described by Sung and Lee [2005], Villard [2001], Deng et al. [2002b], Jokela et al. [2008], and Shih et al. [1998]. These tools allow the definition of alternative playback paths (jumps) and the use of navigational elements only to a limited extent. Prior tools that were developed before SMIL became W3C Recommendation in 1998, like the tools described by Shih et al. and Deng et al., in the Synchronization Editor, TYRO, SIMPLE, and MPRES Author use self-defined models or XML formats. More recent tools like SMIL Builder, GRiNS, SMILAuthor, Java-Assisted SMIL (JAS), and the tools described by Sung and Lee and Jokela et al. use SMIL as description language for the interactivity.

—Related work and tools from the areas of **video browsing**, **video search** like YOVISTO [Sack and Waitelonis 2010; Waitelonis et al. 2011], **multiview video** [Kelly et al. 1995; Katkere et al. 1996; Maugey and Frossard 2011; Miller et al. 2011; Xiu et al. 2012], tools for **video annotation using video analysis**, m-ontomat-annotizer [Petridis et al. 2006], or iVAT [Bianco et al. 2013], and **social video/social TV** [Wang et al. 2012; Shin 2013] are not taken into account in this survey. They provide certain interactive or nonlinear features, but they are not closely related to the types of extended videos described in this section.

Definitions and descriptions of work from the areas of clickable videos, interactive videos, and nonlinear videos can be found in Meixner [2014].

3. DESCRIPTION FORMATS, MODELS, AND STANDARDS

Hypervideos and interactive multimedia presentations require a definition of their internal structures and interaction possibilities. Usually, these formats are used to represent information and settings that are created and saved in an authoring tool and then interpreted, edited, and rendered by a player. These descriptions are mainly file-based, thus readable by the player, and require an underlying model.

Requirements for a data structure are the possibility to define temporal and spatial relationships between videos and annotations. It should furthermore provide hyperlinks between elements defining nonlinearity. Some data structures provide additional overlying navigational layers. Interactive elements may be defined in the data structure as well. The data structure should be extensible in case of new ways of interaction that may be mapped into the model in the future. The following subsections give a detailed description of models for interactive multimedia presentations and hypervideos.

Figure 1 shows the publication/standardization chronology of description formats, standards, and models (whereby the first publication is marked). First publications are from 1991. The AMF (Amsterdam Multimedia Framework) "provides an explicit partitioning of control concerns across components in a network infrastructure" [Bulterman 1993]. This concept was implemented in many models in the following

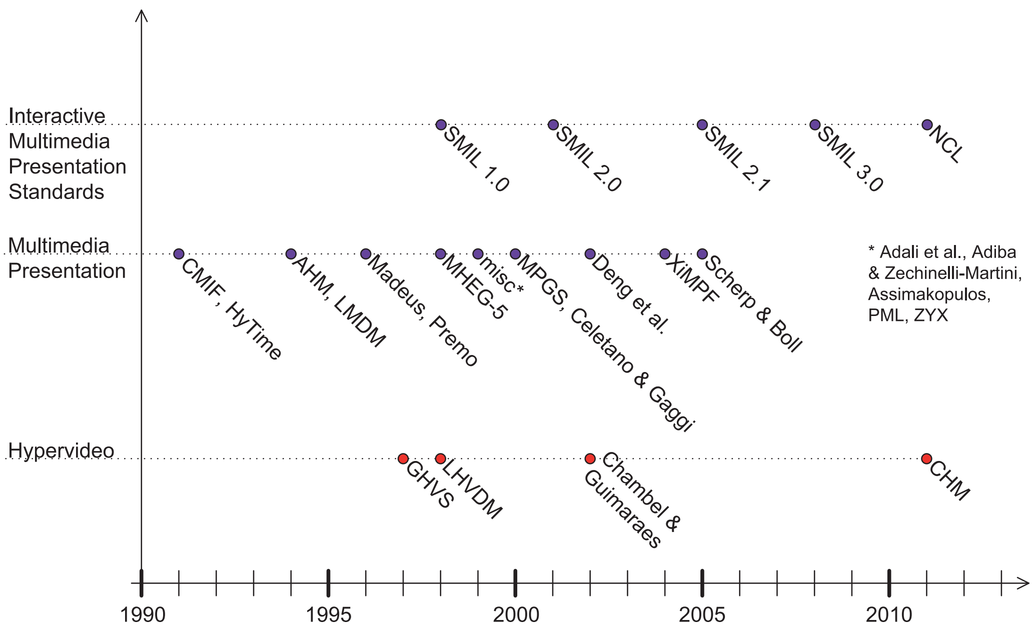

Fig. 1. Publication/standardization chronology of description formats, standards, and models.

years. Especially CMIF [Bulterman et al. 1991; van Rossum et al. 1993], AHM [Hardman et al. 1994, 1997; Hardman and Bulterman 1997], and Madeus [Jourdan et al. 1998] form the basis for later works and the Synchronized Multimedia Integration Language (SMIL) standard. It should be noted, that from 2006 on only a few new developments could be found.

3.1. Standards, Models, and Languages for Interactive Multimedia Presentations

The two most important languages in the area of interactive multimedia standards are the SMIL [W3C 2012] and the Nested Context Language (NCL) [ITU 2011]. Another combination of standards—HTML5 [W3C 2013b], CSS [W3C 2013a], and SVG [W3C 2013c]—can be used to write documents or websites, which are then displayed in web browsers. The elements of HTML allow the definition of metadata, to enable scripting (e.g., with JavaScript [MDN 2013]), create hyperlinks to other websites, and to define section and group content. Content like images, videos, and audio files can be embedded in the document. It is possible to create tables and implement forms for data submission. These elements can be used to create parts of multimedia presentations for the Web that may be hyperlinked with each other. However, due to the declarative nature of HTML5, some documents may contain continuous media, but they are overall mainly used in a static and text-based way to describe the content and structure of a website. Hereafter, we only describe formats that were designed for interactive multimedia presentations, namely, SMIL and NCL.

—**SMIL:** SMIL is a standard for interactive multimedia presentations released by the World Wide Web Consortium (W3C). Design goals of SMIL were to define “an XML-based language that allows authors to write interactive multimedia presentations. Using SMIL 3.0, an author may describe the temporal behavior of a multimedia presentation, associate hyperlinks with media objects, and describe the presentation layout on a screen. [It should allow] reusing of SMIL 3.0 syntax and semantics in other XML-based languages, in particular those who need to represent timing and synchronization" [W3C 2012]. Used media files are images, text, audio files, video, animation, and text streams that are linked to an internal graph/tree structure. Navigation is possible in a presentation, but not in single continuous media files. It is possible to define hotspots for navigation or to display additional information. Using elements and attributes from the timing modules, "time can be integrated into any XML language" [Bulterman and Rutledge 2008, p. 117]. It is possible to define start and end time, duration, persistence, repetition, and accuracy of objects and relation between those objects [Bulterman and Rutledge 2008, p. 117]. The spatial layout of a presentation is defined by the "relative placement of (multiple) media objects," but SMIL does not involve the internal formatting of media objects [Bulterman and Rutledge 2008, p. 149]. SMIL is based on CMIF [Bulterman et al. 1991] and the AHM [Hardman et al. 1994]. SMIL 3.0 consists of 12 major functional areas of elements and attributes (Animation, Metainformation, Content Control, Structure, Layout, Timing and Synchronization, Linking, Time Manipulations, Media Objects, Transition Effects, smilState, and smilText) described as a DTD. The "Timing and Synchronization" part is the most important [Bulterman and Rutledge 2008]. Five profiles are built that use the enlisted elements and attributes, namely, the SMIL 3.0 Language Profile, the SMIL 3.0 Unified Mobile Profile, the SMIL 3.0 DAISY Profile, the SMIL 3.0 Tiny Profile, and the SMIL 3.0 smilText Profile [W3C 2012]. These profiles may limit the standard's elements and attributes or extend it with functionality from other XML languages [Bulterman and Rutledge 2008]. The final version of this standard is the SMIL 3.0 Recommendation, which was published on December 01, 2008 [W3C 2012]. Previous versions of this standard were SMIL 1.0 released in 1998, SMIL 2.0 released in 2001, and SMIL 2.1 released in 2005 [Bulterman and Rutledge 2008].

Extensions for SMIL can be found in different areas but are mainly based on SMIL 2.0. Hu and Feijs describe "IPML, a markup language that extends SMIL for distributed settings" [Hu and Feijs 2006]. SMIL concepts are brought into HTML and web browsers by HTML+TIME [Schmitz et al. 1998]. Hereupon is XHTML+SMIL based. It "defines a set of XHTML abstract modules that support a SMIL 2.0 specification subset. It includes functionality from SMIL 2.0 modules providing support for animation, content control, media objects, timing and synchronization, and transition effects. The profile also integrates SMIL 2.0 features directly with XHTML and CSS, describing how SMIL can be used to manipulate XHTML and CSS features. Additional semantics are defined for some XHTML elements and CSS properties" [W3C 2002]. Limsee3 tries to simplify the authoring process of SMIL files by providing templates for certain purposes. Thereby it integrates "homogeneously logical, time and spatial structures. Templates are defined as constraints on these structures" [Deltour and Roisin 2006; Mikác et al. 2008].

Vaisenberg et al. [2009] introduce the SMPL framework, which is able to add a table of contents, a search function, and a bookmark function to SMIL presentations. Thereby, a semantic layer is added to SMIL presentations. Pihkala and Vuorimaa describe "nine methods to extend SMIL for multimedia applications" (like, e.g., multimedia consoles) in Pihkala and Vuorimaa [2006]. Thereby, SMIL 2.0 is extended with "location information, tactile output, forms, telephoning, and scripting" [Pihkala and Vuorimaa 2006]. A generic, document-oriented way to publish multimedia documents on the Web using HTML5, CSS, and SMIL Timesheets is called Timesheets.js and presented by Cazenave et al. [2011]. Combining different standards allows one to merge logical and temporal structures. Additional libraries provide a table of contents and other forms of navigation.

The Narrative Structure Language (NSL), which is used together with SMIL, is proposed by Ursu et al. [2008]. NSL can be used to achieve a variation in prerecorded materials “by selecting and rearranging atomic elements of content into individual narrations.” The basic elements in this language are “Atomic Narrative Objects (ANO)” [Ursu et al. 2008]. Interactions for ANOs and links between them can be defined building a scene graph. Different types of so-called “selection groups” (comparable to our selection control element) can be defined. Selection criteria for ANOs (or paths in the graph) can be specified with Boolean expressions. Different types of variables are stored. These can be accessed by the language. NSL uses its own computational language syntax that makes a direct translation into XML impossible.

Several other extensions for different versions of SMIL exist. Some extensions of one version of SMIL became part of the subsequent version of the standard. Bulterman examines SMIL 2.0 for document-related requirements of interactive peer-level annotations in Bulterman [2003]. An extension to XLink 1.0 called XConnector is proposed by Muchaluat-Saade et al. [2002]. Reaction to user inputs of different forms is integrated into XML documents and evaluated with real-time programming by King et al. [2004]. Both extensions are applicable to SMIL 2.0 documents. An extension for SMIL 2.1 called SMIL State is proposed by Jansen and Bulterman [2008, 2009]. It allows one to add variables to a multimedia presentation enabling dynamic adaptation to user interactions. SMIL State became part of SMIL 3.0. A temporal editing model for SMIL 3.0 is described by Jansen et al. [2010]. Thereby, different forms of document transformations are analyzed.

—**NCL:** The Nested Context Language (NCL) is a declarative XML-based language for hypermedia document authoring designed at the “TeleMidia Lab - PUC-Rio” [2011]. It is standardized as “H.761: Nested context language (NCL) and Ginga-NCL” [ITU 2011]. Being designed as a hypermedia document specification for the Web, its main field of application is digital television (DTV) systems [TeleMidia Lab - PUC-Rio 2011]. “As NCL has a stricter separation between content and structure, NCL does not define any media itself. Instead, it defines the glue that holds media together in multimedia presentations. [A] NCL document only defines how media objects are structured and related, in time and space” [TeleMidia Lab - PUC-Rio 2011]. Variable and state handling in NCL is described and discussed by Soares et al. [2010]. It describes the temporal behavior of a multimedia presentation and the spatial layout of elements on different end-user devices. User interaction with single objects can be defined as well as the activation of alternative parts of a presentation [ITU 2011]. Media files that can be linked with each other are images, video, audio files, and text. Objects with imperative code content (LUA code, etc.) and objects with declarative code content (HTML, LIME, SVG, MHEG, nested NCL applications, etc.), including other NCL embedded objects [TeleMidia Lab - PUC-Rio 2011; ITU 2011] can be added. NCL is based on the Nested Context Model (NCM) [Casanova et al. 1991; Soares and Rodrigues 2005] and inherits modules from SMIL [Silva et al. 2004].
The current version of this standard is version 3.0. Previous versions of this language are NCL 1.0, which was defined as a DTD; the second version, NCL 2.0, was defined in a modular way using XML Schema. According to that, a combination of single modules in language profiles was possible [TeleMidia Lab - PUC-Rio 2011]. NCL 2.0 contained 21 modules from 11 functional areas [Silva et al. 2004]. Versions 2.1, 2.2, 2.3, and 2.4 refined previous versions and introduced new modules [TeleMidia Lab - PUC-Rio 2011]. NCL 3.0 specifies attribute values and introduces new functions named “Key Navigation” and “Animation.” “NCL 3.0 made [in-]depth modifications on the Composite-Node Template functionality. NCL 3.0 also reviewed the hypermedia

connector specification in order to have a more concise notation. Relationships among NCL imperative objects and other NCL objects are also refined in NCL 3.0, as well as the behavior of NCL imperative object players" [TeleMidia Lab - PUC-Rio 2011]. NCL 3.0 contains 29 modules and four different predefined profiles. NCL 4.0 is work in progress.[3]

—**Further Multimedia Presentation Models and Languages:** Further multimedia presentation/document models and languages are described by Adali et al. [1999, 2000], Adiba and Zechinelli-Martini [1999], Assimakopoulos [1999], Deng et al. [2002a], and Scherp and Boll [2005]. Further models are ZYX [Boll et al. 1999, 2000; Boll and Klas 2001], the Layered Multimedia Data Model (LMDM) [Schloss and Wynblatt 1994]; Madeus [Layaida and Sabry-Ismail 1996], and MPGS [Bertino et al. 2000]. Interchange formats are the CWI Multimedia Interchange Format (CMIF) [Bulterman et al. 1991] and the Procedural Markup Language (PML) [Ram et al. 1999]. Both, models and formats are described for PREMO (Presentation Environment for Multimedia Objects) [Herman et al. 1996a, 1996b] and XiMPF: eXtensible Interactive Multimedia Presentation Format [Van Assche et al. 2004]. These models and formats commonly consist of a temporal and a spatial model/description defining when and where media elements are displayed. Media elements are in general videos, audio files, images, and texts. PREMO and PML allow the use of animated graphics. Jumps on the timeline can be specified in LMDM and in the approach described by Scherp and Boll. These models provide some interactive and navigational features, but are mainly defined for basic multimedia presentations and not discussed in more detail hereafter.

Models and languages for multimedia presentations focus on the temporal and spatial arrangement of media objects, which can be navigated on a timeline. Interaction with these objects is possible depending on the document format. Additional navigational elements are not provided in the standards and require extensions of the format for an efficient implementation. The SMPL framework for SMIL may, for example, provide ideas for the implementation of a table of contents and a search function. Similar to SMIL, NCL does not provide native structures to define additional navigational structures. A common problem in many of the description languages and models for multimedia presentations (especially in interval-based models) is the temporal element synchronization, which may occur in different temporal relationships as described by Allen [1983]. Different approaches that try to overcome this problem are published. Used techniques are, among others, dynamic extended finite-state machine (DEFSM) models [Huang and Wang 1998], the Z notation [Shih et al. 1999], Timed Petri Nets and the logic of temporal intervals [Little and Ghafoor 1990], Dynamic Petri Nets (DPN) [Tan and Guan 2005], and collections of Petri nets [Shih 1998]. Synchronization models and languages are, for example, proposed by Meira and Moura [1994], Schnepf et al. [1996], Bailey et al. [1998], Aygün and Zhang [2002], Presti et al. [2002], Buchanan and Zellweger [2005], and Hakkoymaz [2009]. A detailed overview of this problem and a comparison of possible solutions are described by Blakowski and Steinmetz [1996]. Further discussion can be found in Meixner [2014]. Synchronization issues increase in complexity, when user interaction is allowed. Wahl et al. [1995] name temporal interaction forms and their temporal dependencies and suggest an integrated model for time and interaction. Existing languages can be extended by the properties described by Bes and Roisin [2002], namely, "priorities, more abstract properties, and fallback positions." Keramane and Duda extend basic media segments with "executable code, live feeds, and links." They take user interactions, content-sensitivity, and new

[3]http://www.telemidia.puc-rio.br/?q=pt-br/projetoNCL40 (accessed July 28, 2016).

sources of multimedia data into account while providing support for sharing and reuse [Keramane and Duda 1997].

3.2. Models and Languages for Hypervideos

Four different models were identified for the description of hypervideos. All models have video as a main medium. Annotations are mainly images, text, audio files, and videos. The video scenes are linked to a graph structure by the definition of hyperlinks.

—**Generalized HyperVideo System (GHVS) Model:** GHVS can be used to specify hyperlinks between frames. It meets “basic goals like physical data independence, the ability to compose arbitrarily complex presentations of hypervideos, completeness in expressibility, and simplicity” [Hung 1997]. A graph consisting of video scenes is defined by video-to-video hyperlinks. Rectangled hotspots allow the definition of jumps to other frames, between scenes, and to audio files and images. The defined language in this work is called GHVS language and it is based on the “Generalized HyperVideo System (GHVS) model,” which in turn is based on the PRES model [Wong et al. 1996].

—**Logical Hypervideo Data Model (LHVDM):** “In addition to multilevel video abstractions, the model is capable of representing video entities that users are interested in (defined as hot objects) and their semantic associations with other logical video abstractions, including hot objects themselves” [Jiang and Elmagarmid 1998]. Links between videos define a graph structure. It is possible to define links in videos, to jump to certain frames. Contents shown with the videos, like images and audio files are extracted from the main video as images out of frames or as audio files by saving the soundtrack. Temporal information describes the time intervals during which an object is activated as a link (hot object). The object has a certain spatial information during that time. Spatial relations between hot objects exist. A video query language is defined for the LHVDM.

—**Chambel and Guimarães:** Chambel and Guimarães [2002] describe a “hypervideo model [which] is based on the hypermedia model of the Web, extended with additional concepts required to support the temporal dimension of hyperlinking in dynamic media, such as video” [Chambel and Guimarães 2002]. The main media in this model are videos that are enriched with images and text. The media are linked to a graph structure by hyperlinks. Different types of links like “multilinks, dynamic links, synchronized links” as well as a table of contents and various maps are used to navigate in the hypervideo. “Link anchors can be spatially scattered in the pages and images, allowing for a more fine grained addressing of links origin and destination” [Chambel and Guimarães 2002]. Jumps to points on the timeline can be defined in a video index. Temporal links that are established for a time interval are dependent on time conditions. Spatial links depend on space conditions and make it possible to establish links from certain video regions. The language used for hypervideo construction is called HTIMEL.

—**Component-Based Hypervideo Model (CHM):** The CHM is a “high level representation of hypervideos that intends to provide a general and dedicated hypervideo data model” [Sadallah et al. 2011]. This model consists of a spatial, a temporal, and an event-based model. The main medium is video. Videos are linked and extended with text, video, audio files, and rich text. Jumps to points on a timeline, in a map, in a history, or to links associated with a table of contents are possible. The model provides “high level components such as data readers, enrichment content viewers, video players, hotspots, timelines, maps and tables of contents” [Sadallah et al. 2011]. A timeline-based model with virtual time references is used. The spatial model is derived from SMIL.

No standard specifically designed for hypervideos was found. The models found during the review do not provide a machine-readable structure. A second layer of navigation implemented as a table of contents or a keyword search can rarely be found. The models described in this subsection provide timing and spacial settings in the following way: Basic constructs to describe the position of an object are available, but it is not possible to define interactive moving areas (e.g., hotspots) in some models. The timing models vary in their way of synchronizing single elements. Two of the models provide a table of contents as an additional navigational structure. The model described by Chambel and Guimarães does not provide the impression of an overall video, because the linking is realized between websites with embedded videos and not in a single video player that loads different video files. The Generalized HyperVideo System (GHVS) model is a basic and relatively static model without additional navigational structures. No annotations are used in the LHVDM which provides hotspots and linking. None of the models was transferred into a usable (XML-based) language.

4. AUTHORING TOOLS AND PLAYERS

The authoring process of interactive and nonlinear media is more complicated than the process for traditional linear video. Writing control files for players by hand is a very tedious job. Back in 1989, Fox demanded that "efficient tools and environments for authoring and editing of interactive multimedia programs must be developed" [Fox 1989]. Bulterman and Hardman [2005] describe "issues that need to be addressed by an authoring environment" for multimedia presentations. They identify "seven classes of authoring problems," namely, the definition and realization of media assets, synchronization composition, spatial layout, asynchronous events, adjunct/replacement content, performance analysis, and publishing formats. They describe and explain "four different authoring paradigms," which are (please refer to Bulterman and Hardman [2005] for further reading) structure-based, timeline-based, graph-based, and script-based. Our literature review showed that not all tools deal with all seven issues. Most of the tools use a combination of the authoring paradigms, which are enhanced by other GUI elements and paradigms to provide the full functionality needed in an authoring tool (also depending on the underlying model).

Contrary to authoring tools, which are usually used by a small number of authors, players are used by a much wider range of viewers. Regarding the fact that videos with interactivity and nonlinearity can be used in lots of different scenarios (like e-learning, medical or sports training, or guided tours), potential users may be of every age group and skill level in using a playback device. While standard controls like play, pause, stop, fast-forward, fast rewind, volume control, or a timeline are sufficient for traditional linear videos, players for videos with extended functions require more advanced concepts for display and interaction. Jain and Wakimoto [1995] claim that "in the field of entertainment and training, where interactive video is expected to be useful, much more friendly interface is desired [sic]." This requires an intuitive arrangement and labeling of buttons as well as interactive elements like hotspots depending on the type of video. Many authoring tools offer some kind of own player implementation. These optimize the content output created in the authoring tool and provide functions suited for the desired use cases.

The remainder of this section gives an overview of authoring tools for interactive multimedia presentations and hypervideos. Tools described in scientific work were not tested for the use with current operating systems.

4.1. Authoring Tools and Players for Interactive Multimedia Presentations

Tools for the authoring and playback of interactive multimedia presentations all provide some basic features. Nearly all tools allow the combination of text, images, audio files, and videos—often described as media files or media elements. The tools either use self-defined models or XML formats or they use SMIL as description language for the interactivity.

—**Madeus:** Madeus provides "various kinds of context-dependent navigation: step by step navigation [...], structural navigation [...], and user defined navigation [...]" [Jourdan et al. 1998] in an interactive multimedia presentation. Two forms of temporal navigation, "Context dependent navigation" and "Context independent navigation" are possible. It allows "efficient support for the specification of temporal scenarios and this in an architecture that allows the integration of both authoring and presentation phases of multimedia documents." It uses audio, video, different image files, and formatted text. The authoring tool provides timelines, graphs, and multiple views, which are not further specified in the paper. Presentations are exported in the Madeus language (XML). The player is implemented as a stand-alone desktop player, the so-called "Madeus presentation engine;" no detailed description of the player is available.
—**GRiNS:** GRiNS [Bulterman et al. 1998] "allows the original media assets to be allocated to screen locations [...], and have their presentations synchronized." It "presents a hierarchy of the node structure of the multimedia document to promote the re-use of its components." The authoring tool presents the structure of hyperlinks within the document and to the Web. Different types of media can be linked in the authoring tool providing a WYSIWYG end-user view, a hierarchical structure view, a timeline view (channel view), and a hyperlink view. A SMIL file is exported. The GRiNS player is implemented as a preview in the authoring tool. No stand-alone desktop player is available.
—**MEMORY:** MEMORY is an "integrated approach for adaptive multimedia presentations enabling universal access for situational learning" [Kleinberger et al. 2008]. MEMORY allows the definition of search queries for media documents and navigation in the search results. Various media (audio file, video file, XML file, PDF file, DOC file) can be linked. "Navigation possibilities for jumping to different media documents or fragments [are] presented in a hit list" [Kleinberger et al. 2008]. The projects are exported in LOM [LTSC 2002] format. The MEMORY player is a web-based presentation tool. It provides a list of jump destinations. Controls for single media or channels are implemented. Annotations are presented in a fixed arrangement around the main medium. Interaction with the video is play, pause, and stop. Forward, Rewind, and other interaction is possible with additional information.
—**MediaTouch:** MediaTouch is a "visual-based authoring tool [...]. It's based on the native approach, which lets the author operate at the level of MHEG-5 objects" [Echiffre et al. 1998]. MediaTouch enables authors to edit MHEG-5 objects providing editors for hierarchy, properties, spatial layout, and links [Echiffre et al. 1998]. Scenes and media elements are arranged in a tree structure. It is possible to create links between elements/scenes, hotspots, and hyperlinks. No player is described in this work.
—**LECTURNITY 4:** LECTURNITY 4 allows authors to create "screen recordings for software training and e-learning content for company training, [and] e-lectures for teaching and sales training productions" [imc AG 2010]. Used media are Powerpoint presentations, audio files, video, and images. The authoring tool has views with parallel timelines, a preview area, and a toolbar. These tools allow authors to create buttons, transparent areas, a directory, thumbnails, a timeline, a title, searches, and hotspots for navigation and further information in the resulting presentation. No specification of an export format could be found. The LECTURNITY 4 Player is implemented as a stand-alone desktop player. Besides standard controls

like play/pause, fast-forward, rewind, stop, and restart for a whole presentation, it provides markers on a timeline, hyperlinks, buttons/hotspots in presentation, and a list of jump destinations. Annotations are presented in a fixed arrangement around the main medium.

—**NextSlidePlease:** NextSlidePlease uses "a directed graph structure approach for authoring and delivering multimedia presentations" [Spicer et al. 2012] that are mainly based on presentation slides as media objects. It can be used for "authoring and delivering agile multimedia presentations." The authoring tool provides tools called "Overview Inset, Time Cost and Priority Controls, Presentation Graph Editor Plane, Zoom Slider, [and] Graph View" [Spicer et al. 2012]. These allow the creation of alternative playback paths, jumps between related/linked slides, a graph structure, and a next slide definition. No data format is described. The NextSlidePlease player is implemented as a preview in the authoring tool. No stand-alone player is available. It provides a graph structure that illustrates the current position, hyperlinks, and a list of jump destinations.

—**Matchware Mediator 9:** Matchware Mediator 9 can be used to "create interactive CD-ROM presentations, dynamic HTML pages and Flash projects [. . .] [providing] icon-based editing, [. . .] without requiring any coding or scripting" [MatchWare A/S 2012]. The authoring tool links different media files and provides functions to create an extended navigation by using hotspots, buttons, and links. A nonlinear structure with choice elements (e.g., links, menus) and hotspots for "hyperlinking" to websites or other sections is created. It provides a WYSIWYG editor, a hotspot-editing function, parallel timelines, a toolbar, input fields for annotation content, an event editor, and an animation editor. The project is exported in HTML [W3C 2013b] and JavaScript [MDN 2013] code making it playable/presentable in web browsers.

—**Cutts et al.:** Cutts et al. [2009] describe the "use of a video segmentation process that provides contextual supplementary updates produced by users. Supplements consisting of tailored segments are dynamically inserted into previously stored material in response to questions from users." The authoring tools allow the definition of a table of contents, a search function, and markers on the timeline. Used media are multimedia documents, supporting text (with links), and frequently asked questions. The projects are exported as XML files; no standard is used. The player is implemented as a stand-alone desktop player with editing functions. It has standard controls like play/pause, fast-forward, rewind, stop, and restart for the whole presentation, as well as a table of contents and markers on a timeline. Annotations are presented in a fixed arrangement around the main medium.

—**Gaggi and Celentano:** Interactive multimedia presentations with parallel and sequential execution and presentation of media, and hyperlinks for navigation can be created with the tool described by Gaggi and Celentano [2002]. Used media are video, audio clips, images, and text pages annotated with various media. The authoring tool offers editors for a timeline represented as a tree, a graph view, and a spatial layout view. The output is saved to an XML file; no standard is used. No player interface was implemented in the work of Gaggi and Celentano. They propose an execution simulator with special controls for the simulator (start, pause, stop, end, reset, import, close). Media are represented by placeholders.

—**AMBULANT SMIL player:** The AMBULANT SMIL (2.0/2.1/3.0) player [CWI 2010; Bulterman et al. 2004] has standard controls like play/pause, fast-forward, rewind, stop, and restart for a whole presentation. All media files are positioned independently on the canvas; no fixed areas are given by the AMBULANT SMIL player. A browser plug-in is available. While the player can interpret the whole SMIL standard, interaction and navigation are strongly dependent from the functionality provided by the authoring tool that produced a SMIL file.

Summarizing the analysis of tools for interactive multimedia presentations, the following statements can be made: Authoring tools usually provide more than one view. Used patterns and editors vary greatly depending on the complexity of the resulting presentation and the level of interactivity, or the extent of additional functions. The user can choose between timeline-based, graph-based, and structure-based editors in one authoring tool depending on which editor supports the current step in the authoring process best. Often, the same multimedia document is presented in different ways. Many of the tools do not provide an overview, be it for the whole video or for single elements. This may make it hard for end users to keep track of their elements in a large project. Most of the described tools for interactive multimedia presentations either do not describe the description language or do not use self-defined formats and models. Players for interactive multimedia presentations mainly have standard controls for navigation in the whole presentation depending on the underlying synchronization model. Some players provide navigation for single media or media channels and hyperlinks to other presentation parts. Dealing with interactive multimedia presentations, players provide navigational and interactive features like hotspots. The outcome of authoring tools without players is either presented with SMIL players like AMBULANT SMIL (2.0/2.1/3.0) player [CWI 2010; Bulterman et al. 2004], or with standard web browsers. A feature overview of different authoring tools and players for interactive multimedia presentations can be found in Table II (sorted by year of first publication). The table shows the most distinctive features between the tools.

4.2. Authoring Tools and Players for Hypervideos

Most of the tools found for the creation of hypervideos are described in scientific papers; only few web or commercial tools could be found. Navigation and player controls vary widely in player implementations. Most players (except those described otherwise in the following) were implemented as stand-alone players for desktops:

—**Chang et al.:** Chang et al. [2004] present an "object-based hypervideo authoring system. Video objects can be described by semantic annotation and multistory movies can be produced." The projects are based on one linear video that is enhanced with "multimedia descriptions." Accordingly, if different videos are needed, they have to be merged to one single video before editing, which decreases usability and requires that the user merges the parts in another tool. Hotspots are defined to jump to points on the timeline. "Additional data can be a text, a video clip, a URL link, or a still image" [Chang et al. 2004]. Hotspots are used as choice elements in the video to jump to other scenes, which creates a nonlinear link structure. Annotated regions in a segment are chosen to be "branch points" (forks). The GUI provides a graph view, a video preview, and an overview for defined video parts. No information about the description language is given. The player is "developed for the video viewer to view the annotated film efficiently" [Chang et al. 2004]. Implemented player controls cannot be determined from this work. Additional information are "multimedia descriptions," or more precisely "a text, a video clip, a URL link, or a still image" [Chang et al. 2004]. The snapshot in this work shows the additional information in a two-part GUI in the right area. The video is played in the first and annotations are presented in the second area.

—**Finke and Balfanz:** Finke and Balfanz [2004] (partially based on Balfanz et al. [2001]) describe "basic functional building blocks" of a "generic hypervideo concept." The main media are videos that can be linked with other videos. Rectangled hotspots, which show additional information as HTML page after user interaction, are defined. The implementation does not allow the editing of annotations; these can only be added as already edited files. This requires the usage of other tools even for smaller

Table II. Authoring Tools and Players for Interactive Multimedia Presentations

Source	Main medium/video scenes + additional information	Playback paths (jumps), non-linearity, interaction	Patterns/editors	Type of player	Player controls	Language
Madeus [Jourdan et al. 1998]	"Mpeg audio and video, different image formats and formatted text"	"Temporal navigation [. . .]: Context dependent navigation [. . .] and Context independent navigation"	timelines, graphs, multiple views (no screenshots available)	"Madeus presentation engine"	no detailed description of player available	Madeus language (XML)
MediaTouch [Echiffre et al. 1998]	scenes, media elements in a tree structure	links between elements/scenes, hotspots, hyperlinks	MHEG-5	"Hierarchy Editor, Properties Editor, Layout Editor, Links Editor"	–	–
GRiNS [Bulterman et al. 1998]	media files	mainly linear, navigation interaction via hyperlinks	end-user view, hierarchical structure view, timeline view (channel view), hyperlink view	player preview in authoring tool	WYSIWYG view for development, no stand-alone player	SMIL
Gaggi and Celentano [Gaggi and Celentano 2002]	video, audio clips, images, text pages annotated with various media	mainly linear, parallel and sequential execution, navigation by hyperlink activation	timeline represented as a tree, graph view, spatial layout view	execution simulator	simulator: start, pause, end, reset, import, close	XML file, no standard
AMBULANT SMIL (2.0, 2.1, 3.0) player [Bulterman et al. 2004; CWI 2010]	–	–	–	stand-alone player, browser plug-in	open file, play, pause, stop, view source	SMIL
MEMORY [Kleinberger et al. 2008]	various media (audio file, video file, XML file, PDF file, DOC file)	"Navigation possibilities for jumping to different media documents or fragments presented in a hit list"	–	web-based presentation tool	video: play, pause, stop; add. info.: forward, rewind, other	LOM
Cutts et al. [Cutts et al. 2009]	multimedia documents, supporting text (with links), freq. asked questions	navigation in table of contents, search, or with marker on timeline	–	stand-alone player with editing function	forward, rewind, play, timeline	XML files, no standard
LECTURNITY 4 [imc AG 2010]	Powerpoint-presentation, audio files, video, images	buttons, transparent areas, directory, thumbnails, timeline, title, searches, hotspots	parallel timelines, preview-area, toolbar	stand-alone player	standard video controls, buttons/ hotspots in presentation	–
Matchware Mediator 9 [MatchWare A/S 2012]	media files	nonlinear structure with choice elements (links, menus), hotspots for hyperlinking to website or other section	WYSIWYG editor, hotspot-editing function, parallel timelines, toolbar, input fields for annotation content, event editor, animation editor	–	–	HTML (+JS)
NextSlide-Please [Spicer et al. 2012]	presentation slides	alternative playback paths jumps between related/linked slides, graph structure, next slide definition	"Overview Inset, Time Cost and Priority Controls, Presentation Graph Editor Plane, Zoom Slider, Graph View"	–	–	–

editing tasks. The system is web-based. The GUI depicted in Balfanz et al. [2001] shows a tree view with keyframes and an editor to place rectangled clickable areas on a frame. The description format for the annotation metadata and the format to describe the links between single video nodes is not described in any paper. The web video player consists of three areas for display. A video area, a navigation view, and an information and communication view are available. Video scenes in a linear order are used. Player controls are play/pause, jump forward/backward between scenes, and a timeline. The navigation view provides links to multimedia annotations.

—**Hyper-Hitchcock:** Hyper-Hitchcock is an authoring tool for the creation of detail-on-demand video. "Detail-on-demand video is a form of hypervideo that supports one hyperlink at a time for navigating between video sequences" [Shipman et al. 2008]. The main medium is video, but additional information is also provided as video. A nonlinear structure is defined by several types of links (detail links, prerequisite links, related information links, alternate view links, action choice links) defining playback paths [Shipman et al. 2008]. Choice elements for navigation between the video scenes are the linked video key frames. A timeline, a clip selection panel, a tree view, and a workspace area are GUI parts. It is not possible to define hotspots for navigation. The internal link structure format is not described in any paper. The player is implemented as a stand-alone player, which was iteratively developed over several user studies. "All keyframes are clickable, thus enabling the user to return several link levels at once" [Shipman et al. 2008]. The Hyper-Hitchcock player has one single area, where the main video and additional information (keyframes as links to other videos) are shown. Controls of the player GUI are buttons for play, stop, and navigation as well as a timeline with a keyframe preview. All videos are displayed in the main video area.

—**Advene:** Advene [Aubert and Prié 2005; Aubert et al. 2012] is a tool for active reading in videos and the generation of hypervideos based on rules. "One of the results of active reading applied to audiovisual material can be hypervideos, that we define as views on audiovisual documents associated with an annotation structure" [Aubert and Prié 2005]. One linear video is used as a main medium; annotations are rendered to different views. Alternative playback paths (jumps) in the audio-visual document are definable by the annotation layer. Navigation depends on the annotations defined for the video. The definition of hotspots is not possible. GUI elements are a stream-based view, a view for note taking, a tree view, parallel time lines, a description area, and a video area. A self-defined model and description format for the projects is used. Advene provides two implementations/views for playback: a static view and a dynamic view. One video can be navigated via timeline or URLs. Neither choice elements that have influence on the video structure nor hotspots provide interactivity to the viewer. The player GUI offers standard controls, hyperlinks, an URL stack, navigation links, and a position indicator for navigation in the video. Annotations are shown around the video and as overlay over video. They are mainly text-based.

—**Hsu et al.:** Hsu et al. [2005] describe a tool for "hyper-interactive video browsing by a remote controller and hand gestures." Video scenes are arranged in a graph structure. This nonlinear structure can be navigated with "hyperlink[s] in a specified temporal-spatial domain" [Hsu et al. 2005]. Additional information are "text descriptions, existing image files, web page files, or URLs on the Internet" [Hsu et al. 2005]. Alternative playback paths (jumps) are not possible. It is not possible to create hotspots in a scene. The tool contains a video preview, an annotation area, and a graph view. For playback, they describe "hyper-interactive video browsing by a remote controller and hand gestures." Nonlinear video with a graph structure offers hyperlinks "in a specified temporal-spatial domain" [Hsu et al. 2005]. These links are navigated by gesture controls. Additional information like text descriptions, existing image files, web page files, or URLs on the Web can be displayed with the video. They are either shown as smaller overlays or they replace the video.

—**Zhou et al.:** Zhou et al. [2005] present a system for "automatic generation of additional information and the integration of the additional information to its corresponding selectable video object." The outcome is a limited form of hypervideo, called detail-on-demand video. The main medium is a video that is annotated with video frame images and HTML files. Latter ones may contain links to further information. No choice elements or hotspots can be created with the described tool. The GUI is implemented as a converter view with two tree structures. It is not an authoring tool in the traditional sense of the term, because it is not possible to compose different media files to an overall presentation. The GUI does not provide functions to link different media files in the GUI manually. Structure and relations between the elements are described in MPEG-7 [ISO/IEC 2009], which is converted to SMIL [W3C 2012]. The generated files can be played with SMIL players.

—**HyPE and Jeherazade:** HyPE and Jeherazade are combined to implement "narrative intelligence in hypervideo" [Hoffmann and Herczeg 2006]. Jumps in linear videos are triggered by hotspots. Hotspots are used to trigger the display of additional information like video, audio files, text, and images. A self-defined XML file is used to describe the hypervideo structure. A video view and a list with hotspots (polygons) are GUI parts. HyPE provides a GUI that gives no structured overview of the whole video, making authoring of larger projects difficult due to reduced usability. Based on a linear video, nonlinearity is implemented by jumps that are triggered by hotspots. Hotspots are used to display additional information. The player GUI is implemented as a two-part window with a video or audio player on the left side, and a text or an image viewer on the right side. No other controls are offered by the GUI.

—**SIVA Suite:** The SIVA Producer "provides all functions to manage the creation of an interactive nonlinear video. [...] It provides functions for the enrichment of the video content with additional multimedia annotations. [...] A non-linear flow is implemented by separate video scenes instead of jumps in one linear video. Thereby buttons are used for the selection of the follow up scene. The tagging of annotations and scenes with keywords enables a search function in the player. A table of contents can be added by building up a tree structure and assigning scenes from the scene graph" [Meixner et al. 2012]. The projects are saved as XML/JSON files with the structure defined in Meixner and Kosch [2012]. "In addition to standard player functions, [the SIVA] player provides extended navigation (button panels, table of contents, keyword search) in the video, areas for additional information, opportunities for collaboration, and a logging function" [Meixner et al. 2013]. The latter allows user behavior analysis. The player GUI is configured in the authoring tool; it provides extended controls for intra- and inter-scene navigation.

—**Klynt:** Klynt is a web platform for visual storytellers [Honkytonk Films 2013]. The desktop editor has a visual storyboard to create a scene graph consisting of video scenes or multimedia pages. Multiple media formats can be added to the videos, which are then played in an HTML5 player. Buttons are added to scenes that show additional information mainly consisting of text and images or they load another scene. The integration of Facebook, Twitter, and Google maps is possible. It is possible to add a Google maps menu consisting of a map with markers, which are then linked to video scenes. Other navigational elements are presentation-like screens with buttons to other screens or to video segments. These elements allow the creation of hypervideos with focus on different media types. Klynt provides a GUI similar to those known from Adobe products like Adobe Premiere [Adobe Systems Incorporated 2017]. It provides a WYSIWYG editor for links, a scene graph editor, annotation editors, and a timeline. No description format could be found. The Klynt player is implemented as a web player in HTML5. It provides different customized buttons as overlays on the video for navigation between video scenes and for the display of annotations. These may contain more buttons, images, text, videos, or other web-based contents. The Klynt player has one single area, where the main video and additional information are shown. The latter are either shown as smaller overlays or they replace the video. Rectangular hyperlink areas are implemented and menu-like structures can be displayed.

—**LinkedTV/VideoHypE:** LinkedTV, respectively VideoHypE [Redondo-Garcia and Troncy 2013; Baltussen et al. 2013], is a tool for "supervised automatic video hyperlinking" mainly focusing on the annotation and the hyperlinking of video segments. A video can be selected. For this video shots are defined. These are arranged in an overview. Chapters can be defined from the shots. It is possible to select, name, and categorize entities. Hyperlinks can be specified, which link to websites. The tool provides a timeline view. The player from the LinkedTV project is designed as a second screen application for desktops, smart phones, and tablets. The first screen is used to play the video while the second screen can be used to control the main screen. It is possible to navigate to another chapter in the presentation. The second screen has an interface that shows "detected entities of the video grouped by persons, objects and locations" [Ockeloen and van Leeuwen 2013]. Different external control interfaces are available or under development.

Most of the tools described in this section were implemented to show new annotation principles or to combine editing principles. Usability was rarely taken into account. The tools are capable of producing hypervideo with an additional navigation layer to a certain extent. All four authoring paradigms from Bulterman and Hardman [2005] can be found in the described tools and systems. The range of functions varies between the tools. Accordingly, no general structure of an editor can be stated for hypervideo authoring tools. Many tools use self-defined languages (often XML format) or the authors do not describe the description language in their work. SMIL, a standard for multimedia presentations, is used by some tools from this area. The players for hypervideos show differences in presenting the additional information and in provided controls. Standard controls like play and pause are implemented in most of the players. Some players provide timelines (sometimes with key-frame preview). Players are implemented as web players, web browser plug-ins, or stand-alone desktop players. A feature overview of different authoring tools and players for hypervideos enlisting the most distinctive features can be found in Table III (sorted by year of first published paper).

5. OBSERVATIONS AND TRENDS

In the previous sections, we analyzed description formats, models, standards, authoring tools, and players for hypervideos, and interactive multimedia presentations. In this section, we describe the observations and findings that emerged during this work.

5.1. Description Formats, Models, and Standards

In our literature search, we found two standards for multimedia presentations: SMIL and NCL. While no standards were found for hypervideos, both SMIL and NCL can be used for hypervideos as well (maybe with extensions). The literature analysis revealed that these standards are rarely used when authoring tools and players are implemented; however, we found no explicit reasons for that. One possible reason may be the complexity of SMIL and NCL. Because of this complexity, many research prototypes were implemented using formats that exactly fit the needs of the tools rather than write exports for SMIL or NCL. While both SMIL and NCL offer many features, some

Table III. Authoring Tools and Players for Hypervideos

Authoring Tools and Players for Hypervideos						
Source	*Main medium/video scenes + additional information*	*Playback paths (jumps), nonlinearity, interaction*	*Patterns/editors*	*Type of player*	*Player controls*	*Language*
Chang et al. [Chang et al. 2004]	one linear video + "multimedia descriptions," text, video clip, URL, image	based on annotations, choices by hotspots to jump to other scenes	graph view, video preview, overview for defined video pieces	stand-alone player	–	–
Finke and Balfanz [Finke and Balfanz 2004]	list of video scenes + "any form of information media," HTML	"navigation engine" for orientation, jumps between scenes (previous, next), jumps on timeline, choices not described, rectangled hotspots track objects	–	web player	play/pause, jump forward/ rewind, navigation view with hyperlinks	hyper-video metadata model, data model, data repository
Hyper-Hitchcock [Shipman et al. 2005, 2008]	video + videos	defined by several types of links (keyframes) and user behavior, no hotspots	timeline, clip selection panel, tree view, workspace	stand-alone player	play, stop, navigation buttons, timeline, keyframes	–
Zhou et al. [Zhou et al. 2005]	video + video frame images and HTML files	alternative paths, no choices or hotspots in videos, but links in annotations	converter view with two tree structures	–	–	MPEG-7, SMIL
Advene [Aubert and Prié 2005; Aubert et al. 2012]	one audio-visual document + annotations rendered to different views	jumps defined by annotation layer, no choice elements or hotspots	stream-time-based view, view for note taking, tree view, parallel timelines, description area, video area	stand-alone player, interactive homepage	standard controls, hyperlinks, URL stack, navigation links, position indicator	own model
Hsu et al. [Hsu et al. 2005]	video scenes in graph structure + text, images, web page files, URLs	graph, choices by "hyperlink in a specified temporal-spatial domain," no hotspots	video preview, annotation area, graph view	stand-alone player	gesture controls	–
HyPE and Jeherazade [Hoffmann and Herczeg 2006]	linear video + video or audio player, text or image window	choices and jumps triggered by hotspots, hotspots also display additional information	video view, list with hotspots (polygon)	stand-alone player (HyPE stand-alone player)	none	XML file
SIVA Suite [Meixner et al. 2010; Meixner 2014]	video scenes in graph structure, images, text, audio, other videos	choices of next scene in button panel, hotspots display additional information, table of contents, keyword search	scene graph view, graph overview, media repository, scene repository, annotation editor, toolbar, editor for table of contents	web player (HTML5)	play/pause, timeline, full screen, volume control, search, table of contents, prev/next scene, settings	SIVA format (XML file) [Meixner and Kosch 2012]
Klynt [Honkytonk Films 2013]	video scenes and media in graph structure + text, graphic, audio files, video, hyperlinks	visual storyboard to create a graph, links between sequences, choices by buttons on the video, several menus, no hotspots	graph view, WYSIWYG editor, timeline view	web player	play/pause, timeline, full screen, volume control, social media, (menus)	–
LinkedTV/ VideoHypE [LinkedTV Consortium 2013; Redondo-Garcia and Troncy 2013]	video scenes + multimedia content (on second screen)	list of video scenes, hyperlinks between video segments, no choices, no hotspots	chapter editor, timeline view, link editor	player with second screen	controls on second screen, external control interfaces	LinkedTV ontology, Media Fragments URI, RDF

concepts—like a table of contents, a keyword search, or other overlying navigational structures—are hard to implement with these standards. This makes it necessary to either extend, or makes one's own model fit into SMIL or NCL.

We also discovered that considerable research dealt with the temporal synchronization and validity of SMIL documents. Several approaches were proposed and are described in this work. However, newer standards like HTML5 rarely reuse the findings of these works. Precise synchronization between media elements is hardly possible in HTML5 [Meixner and Einsiedler 2016]. This makes it difficult to transfer ideas from earlier research into appealing web presentations, which is necessary to make this form of multimedia document widely accepted. While many of the early research prototypes were stand-alone desktop players, newer implementations are mainly web-based, providing access to more viewers. However, media delivery over networks may be delayed due to latency and packet loss during transmission. Fine grained control of media elements in the browser cache is not possible due to missing mechanisms in HTML5 and different implementations from browser to browser. This makes it nearly impossible to create a temporally precise synchronized presentation of multimedia elements, leading to a reduced user experience.

5.2. Authoring Tools and Players

Summarizing the findings of this survey, it can be stated that the earlier implementations of authoring tools and players were research prototypes. These implementations were used to prove concepts that were possible with new developments like available programming frameworks and a growing use of video on personal computers. Different paradigms were proposed in the authoring tools described in this article, but studies on usability of the resulting prototypes are hard to find. Consequently, no knowledge about best practices for designing authoring tool and player GUIs exists. There may be a tendency towards web-based authoring, but no clear trend can be noticed. It is furthermore hard to establish one authoring tool in this area due to the wide gap between existing user groups. Hobby users, or nonprofessionals who capture videos with their camcorders or smartphones, have different needs than professionals who want to create larger projects. This fact is rarely taken into account in the works analyzed in this survey. Specialization of tools for certain scenarios (like e-learning, house or building walk-throughs, or physiotherapy trainings) can help establish this type of multimedia document for certain user groups, providing better value for their customers. This, however, requires customized and easy to use authoring tools.

While earlier players were implemented as stand-alone desktop players, the introduction of HTML5 and JavaScript allowed the implementation of web players that can be used by a wide range of users without the need to install software and download videos in advance. However, no trend can be noticed regarding a higher presence of these presentations on the Web. This may have different causes; for example, the lack of authoring tools or knowledge about available technologies. With the trend away from large screens (laptops and desktop computers) to smaller screens (smartphones and tablets), new GUI concepts are necessary to make hypervideos and interactive multimedia presentations available on all end-user devices. However, almost no research or concepts can be found in this area.

Another observation that can be made is that most tools are described in scientific papers; only a few web or commercial tools could be found, both for hypervideos and interactive multimedia presentations. While in many cases concepts are first shown in scientific papers and then adopted to web or commercial tools, this did not yet happen in this area. Possible reasons may be the lack of commonly used paradigms that are well known to producers and consumers, missing user studies that prove the importance and advantages of these videos, or an absent integration of new technologies like sensors into these types of video.

6. CONCLUSION AND OUTLOOK

In this article, we analyzed related work in the areas of "hypervideo" and "interactive multimedia presentation" and provided working definitions for each of the terms. Interactive multimedia presentations consist of different types of media that are arranged spatially and temporally. Besides VCR actions, they have additional navigational structures that add interactivity. Hypervideos are video-based hypermedia with a nonlinear structure and additional information.

Interactive multimedia presentations can be defined using the standards SMIL or NCL, but the research in these areas mainly focuses on temporal synchronization. Four different models could be found for hypervideo, but no standard or commonly used data format exists. Depending on the resulting type of video, authoring tools have different features. Authoring tools for interactive multimedia presentations provide different views that can be used in parallel depending on the editing task. The views are mainly timeline-based, graph-based, or structure-based. A whole project overview is often missing, because each view can only show a certain aspect. Authoring tools for hypervideos have the same editors and views as authoring tools for interactive multimedia presentations. They may have additional functions depending on the tool's scope. Players are mainly implemented as stand-alone desktop players or web players. Players for interactive multimedia presentations are mainly implemented as stand-alone desktop players or players that are integrated into authoring tools. They provide standard VCR controls and additional controls depending on the authoring tool/player and scope. Older hypervideo players were mainly implemented as stand-alone desktop players; newer players are implemented as Web players mainly using HTML5 and JavaScript.

Different forms of advanced videos can be found throughout the Web nowadays, but the more features and interaction they provide, the more complicated and expensive is the creation process. This suggests some topics for future research:

—**Usability:** An important area of research is the usability of authoring tools and players. Despite the fact that users mostly interact with the players, authoring tools should also be easy to access.
—**Specialized (authoring-)tools:** Nonspecialized tools require lots of configuration to create a video for a certain use case, often requiring that settings be configured repeatedly. Easy configuration of fixed settings for a video could speed up the authoring process of hypervideos and interactive multimedia presentations and lead to wider use as a consequence.
—**Authoring on mobile devices:** While recording media on mobile devices is done every day, the media often remain unconnected. Authoring tools for mobile devices that allow linking of media could make recorded media more useful.
—**Automation in the creation process:** Finding media to link and linking the media in a hypervideo or interactive multimedia presentation can be a tedious job. Algorithms for image and video search may help to find good materials and produce better videos. Algorithms that link media of given media collections can speed up the authoring process and guide authors.
—**Movement towards user generated content:** Web players are especially suitable for making the recipient of video also a video author. Adding new content may improve quality and informative value of a hypervideo. This requires tools in the players that enable all users to easily add content.

—**Re-use of existing content/copyright issues:** While content is created and uploaded to the Web every day, copyright laws vary from country to country and are often hard to understand. Uniform rules and licenses would help to simplify this process.
—**Web tools and mobility:** While older players suffered from missing web frameworks for video playback, HTML5 and JavaScript offer the opportunity to reach a wider range of users on different end-user devices (like smartphones, tablets, laptops, desktop computers with large monitors, or large multitouch screens). This should be taken into account when new players are implemented.
—**Sensor input and location-aware mobile players:** Location awareness and sensor input may create new usage scenarios for mobile players. For example, players for physical training settings can be extended with correction and guidance features to avoid overexertion. Depending on the training progress, new exercises may be activated.
—**Integration of 3D, panoramic, and 360° videos:** These types of multidimensional video require additional editors in authoring tools as well as control and navigation elements in players. A special focus has to lie on usability, ensuring that users are not overstrained while using the software.
—**Individualization for end users:** Content presentation should be adaptable to the end user. Machine learning can help preselect relevant presentation content for the viewer.
—**Prefetching and synchronization with current Web technologies:** HTML5, JavaScript, and CSS3 allow the creation of appealing players and presentations. However, it is difficult to ensure that media are available for display at a certain point in time, because of less support for media synchronization (e.g., compared to SMIL). Fine grained presentations where synchronous playback is key are hard to create.

As this survey shows, a lot of work has already been done in the area of hypervideos and interactive multimedia presentations. However, improvements and extensions as indicated in the topics for future research could lead to a wider use of these types of video making them better accessible and more diverse to use.

ACKNOWLEDGMENTS

I want to thank my PhD thesis advisor, Professor Dr. Harald Kosch, and the external examiner of my PhD thesis, Professor Dr. Maximilian Eibl, who provided insight and expertise that greatly assisted this survey. I thank John Doherty and Jan Willem Kleinrouweler for comments that greatly improved the readability of the manuscript.

REFERENCES

A. Abdelli. 2008. Extending the verification of multimedia presentation consistency to resource requirements. In *Proceedings of the 2nd International Conference on Future Generation Communication and Networking (FGCN'08)*, Vol. 1. IEEE, 216–219. DOI:http://dx.doi.org/10.1109/FGCN.2008.125

S. Adali, M. L. Sapino, and V. S. Subrahmanian. 1999. A multimedia presentation algebra. *SIGMOD Rec.* 28, 2 (1999), 121–132. DOI:http://dx.doi.org/10.1145/304181.304193

S. Adali, M. L. Sapino, and V. S. Subrahmanian. 2000. An algebra for creating and querying multimedia presentations. *Multimedia Syst.* 8, 3 (2000), 212–230. DOI:http://dx.doi.org/10.1007/s005300000046

M. Adiba and J.-L. Zechinelli-Martini. 1999. Spatio-temporal multimedia presentations as database objects. In *Database and Expert Systems Applications*, T. J. M. Bench-Capon, G. Soda, and A M. Tjoa (Eds.). Lecture Notes in Computer Science, Vol. 1677. Springer, Berlin, 974–985. DOI:http://dx.doi.org/10.1007/3-540-48309-8_92

D. A. Adjeroh and M. C. Lee. 1995. Synchronization mechanisms for distributed multimedia presentation systems. In *Proceedings of the International Workshop on Multi-Media Database Management Systems, 1995*. IEEE, 30–37. DOI:http://dx.doi.org/10.1109/MMDBMS.1995.520420

Adobe Systems Incorporated. 2017. Adobe Premiere Pro CC. (2017). Website (accessed January 20, 2017). http://www.adobe.com/products/premiere.html.

J. F. Allen. 1983. Maintaining knowledge about temporal intervals. *Commun. ACM* 26, 11 (1983), 832–843. DOI:http://dx.doi.org/10.1145/182.358434

N. A. Assimakopoulos. 1999. A generic spatial temporal computation model for systemic multimedia presentations. *Cybern. Syst.* 30, 6 (1999), 509–531. DOI:http://dx.doi.org/10.1080/019697299125082

O. Aubert and Y. Prié. 2005. Advene: Active reading through hypervideo. In *Proceedings of the 16th ACM Conference on Hypertext and Hypermedia (HYPERTEXT'05)*. ACM, New York, NY, 235–244. DOI:http://dx.doi.org/10.1145/1083356.1083405

O. Aubert, Y. Prié, and D. Schmitt. 2012. Advene as a tailorable hypervideo authoring tool: A case study. In *Proceedings of the 2012 ACM Symposium on Document Engineering (DocEng'12)*. ACM, New York, NY, 79–82. DOI:http://dx.doi.org/10.1145/2361354.2361370

R. S. Aygün and A. Zhang. 2002. Modeling and verification of interactive flexible multimedia presentations using PROMELA/SPIN. In *Model Checking Software*, D. Bosnacki and S. Leue (Eds.). Lecture Notes in Computer Science, Vol. 2318. Springer, Berlin, 205–212. DOI:http://dx.doi.org/10.1007/3-540-46017-9_15

B. Bailey, J. A. Konstan, R. Cooley, and M. Dejong. 1998. Nsync—A toolkit for building interactive multimedia presentations. In *Proceedings of the 6th ACM International Conference on Multimedia (MULTIMEDIA'98)*. ACM, New York, NY, 257–266. DOI:http://dx.doi.org/10.1145/290747.290779

D. Balfanz, M. Finke, C. Jung, and R. Wichert. 2001. An interactive video system supporting e-commerce product placement. In *Proceedings of the World Scientific and Engineering Society Conferences, 2001*. Paper 255. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.67.3305&rep=rep1&type=pdf.

L. B. Baltussen, J. Blom, and R. Ordelman. 2013. VideoHypE: An editor tool for supervised automatic video hyperlinking. In *Intelligent Technologies for Interactive Entertainment*, M. Mancas, N. d'Alessandro, X. Siebert, B. Gosselin, C. Valderrama, and T. Dutoit (Eds.). Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Vol. 124. Springer International Publishing, 43–48. DOI:http://dx.doi.org/10.1007/978-3-319-03892-6_5

E. Bertino, E. Ferrari, and M. Stolf. 2000. MPGS: An interactive tool for the specification and generation of multimedia presentations. *IEEE Trans. Knowl. Data Eng.* 12, 1 (2000), 102–125. DOI:http://dx.doi.org/10.1109/69.842254

F. Bes and C. Roisin. 2002. A presentation language for controlling the formatting process in multimedia presentations. In *Proceedings of the 2002 ACM Symposium on Document Engineering (DocEng'02)*. ACM, New York, NY, 2–9. DOI:http://dx.doi.org/10.1145/585058.585061

S. Bianco, G. Ciocca, P. Napoletano, R. Schettini, R. Margherita, G. Marini, and G. Pantaleo. 2013. Cooking action recognition with iVAT: An interactive video annotation tool. In *Proceedings of Image Analysis and Processing (ICIAP'13)*, A. Petrosino (Ed.). Lecture Notes in Computer Science, Vol. 8157. Springer, Berlin, 631–641. DOI:http://dx.doi.org/10.1007/978-3-642-41184-7_64

G. Blakowski, J. Hübel, and U. Langrehr. 1992. Tools for specifying and executing synchronized multimedia presentations. In *Network and Operating System Support for Digital Audio and Video*, R. G. Herrtwich (Ed.). Lecture Notes in Computer Science, Vol. 614. Springer, Berlin, 271–282. DOI:http://dx.doi.org/10.1007/3-540-55639-7_24

G. Blakowski and R. Steinmetz. 1996. A media synchronization survey: Reference model, specification, and case studies. *IEEE J. Select. Areas Commun.* 14, 1 (1996), 5–35. DOI:http://dx.doi.org/10.1109/49.481691

M. A. Bochicchio and N. Fiore. 2005. Teacher-centered production of hypervideo for distance learning. *Int. J. Dist. Ed. Technol. (IJDET)* 3, 4 (2005), 19–34. DOI:http://dx.doi.org/10.4018/jdet.2005100103

S. Boll and W. Klas. 2001. ZYX—A multimedia document model for reuse and adaptation of multimedia content. *IEEE Trans. Knowl. Data Eng.* 13, 3 (2001), 361–382. DOI:http://dx.doi.org/10.1109/69.929895

S. Boll, W. Klas, and U. Westermann. 1999. Exploiting ORDBMS technology to implement the ZYX data model for multimedia documents and presentations. In *Datenbanksysteme in Büro, Technik und Wissenschaft*, A. P. Buchmann (Ed.). Springer, Berlin, 232–250. DOI:http://dx.doi.org/10.1007/978-3-642-60119-4_14

S. Boll, W. Klas, and U. Westermann. 2000. Multimedia document models: Sealed fate or setting out for new shores? *Multimedia Tools Appl.* 11, 3 (2000), 267–279. DOI:http://dx.doi.org/10.1023/A:1009606112260

S. Bouyakoub and A. Belkhir. 2011. SMIL Builder: An incremental authoring tool for SMIL documents. *ACM Trans. Multimedia Comput., Commun., Appl.* 7, 1 (2011), 2:1–2:30. DOI:http://dx.doi.org/10.1145/1870121.1870123

P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil. 2007. Lessons from applying the systematic literature review process within the software engineering domain. *J. Syst. Softw.* 80, 4 (April 2007), 571–583. DOI:http://dx.doi.org/10.1016/j.jss.2006.07.009

H. P. Brondmo and G. Davenport. 1989. Creating and viewing the Elastic Charles—A hypermedia journal. In *Hypertext: State of the ART, Papers Presented at the Hypertext II Conference*, R. McAlesse and C. Green (Eds.). Intellect, York, England, 43–51. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.85.5291&rep=rep1&type=pdf.

M. C. Buchanan and P. T. Zellweger. 2005. Automatic temporal layout mechanisms revisited. *ACM Trans. Multimedia Comput., Commun., Applications* 1, 1 (2005), 60–88. DOI:http://dx.doi.org/10.1145/1047936.1047942

D. C. A. Bulterman. 1993. Specification and support of adaptable networked multimedia. *Multimedia Syst.* 1, 2 (1993), 68–76. DOI:http://dx.doi.org/10.1007/BF01213485

D. C. A. Bulterman. 2003. Using SMIL to encode interactive, peer-level multimedia annotations. In *Proceedings of the 2003 ACM Symposium on Document Engineering (DocEng'03)*. ACM, New York, NY, 32–41. DOI:http://dx.doi.org/10.1145/958227.958228

D. C. A. Bulterman and L. Hardman. 2005. Structured multimedia authoring. *ACM Trans. Multimedia Comput., Commun., and Appl.* 1, 1 (2005), 89–109. DOI:http://dx.doi.org/10.1145/1047936.1047943

D. C. A. Bulterman, L. Hardman, J. Jansen, K. S. Mullender, and L. Rutledge. 1998. GRiNS: A GRaphical INterface for creating and playing SMIL documents. *Comput. Netw. ISDN Syst.* 30, 1–7 (1998), 519–529. DOI:http://dx.doi.org/10.1016/S0169-7552(98)00128-7

D. C. A. Bulterman, J. Jansen, K. Kleanthous, K. Blom, and D. Benden. 2004. Ambulant: A fast, multi-platform open source SMIL player. In *Proceedings of the 12th Annual ACM International Conference on Multimedia (MULTIMEDIA'04)*. ACM, New York, NY, 492–495. DOI:http://dx.doi.org/10.1145/1027527.1027646

D. C. A. Bulterman, G. V. Rossum, and R. V. Liere. 1991. A structure for transportable, dynamic multimedia documents. In *Proceedings of the Summer 1991 USENIX Conference*. 137–156. Retrieved from http://homepages.cwi.nl/ robertl/papers/1991/usenix/paper.pdf.

D. C. A. Bulterman and L. W. Rutledge. 2008. *SMIL 3.0: Flexible Multimedia for Web, Mobile Devices and Daisy Talking Books* (2nd, illustrated ed.). Springer.

N. Carlsson, A. Mahanti, L. Zongpeng, and D. Eager. 2008. Optimized periodic broadcast of nonlinear media. *IEEE Trans. Multimedia* 10, 5 (2008), 871–884. DOI:http://dx.doi.org/10.1109/TMM.2008.922847

M. A. Casanova, L. Tucherman, M. J. D. Lima, J. L. Rangel Netto, N. Rodriquez, and L. F. G. Soares. 1991. The nested context model for hyperdocuments. In *Proceedings of the 3rd Annual ACM Conference on Hypertext (HYPERTEXT'91)*. ACM, New York, NY, 193–201. DOI:http://dx.doi.org/10.1145/122974.122993

F. Cazenave, V. Quint, and C. Roisin. 2011. Timesheets.js: When SMIL meets HTML5 and CSS3. In *Proceedings of the 11th ACM Symposium on Document Engineering (DocEng'11)*. ACM, New York, NY, 43–52. DOI:http://dx.doi.org/10.1145/2034691.2034700

T. Chambel and N. Guimarães. 1999. The role of hypervideo in learning environments. In *Proceedings of WebNet World Conference on the WWW and Internet 1999*. AACE, 1672. Retrieved from http://www.editlib.org/p/7587.

T. Chambel and N. Guimarães. 2002. Context perception in video-based hypermedia spaces. In *Proceedings of the 13th ACM Conference on Hypertext and Hypermedia (HYPERTEXT'02)*. ACM, New York, NY, 85–94. DOI:http://dx.doi.org/10.1145/513338.513365

H. Chang, H. H. Hsu, Y.-C. Liao, T. K. Shih, and C.-T. Tang. 2004. An object-based hypervideo authoring system. In *IEEE International Conference on Multimedia and Expo (ICME'04)*, Vol. 3. IEEE, 2219–2222, Vol. 3. DOI:http://dx.doi.org/10.1109/ICME.2004.1394711

H.-B. Chang, H.-H. Hsu, and L. R. Chao. 2008. Interactive video game platform for game-based learning. In *Advances in Web Based Learning—ICWL 2008*, F. Li, J. Zhao, T. Shih, R. Lau, Q. Li, and D. McLeod (Eds.). Lecture Notes in Computer Science, Vol. 5145. Springer, Berlin, 232–240. DOI:http://dx.doi.org/10.1007/978-3-540-85033-5_23

H.-B. Chang, H.-H. Hsu, and T. K. Shih. 2007. Using interactive video technology for the development of game-based learning. In *Proceedings of the International Conference on Parallel Processing Workshops (ICPPW 2007)*. IEEE, 24–28. DOI:http://dx.doi.org/10.1109/ICPPW.2007.81

C.-M. Chung and T. K. Shih. 1997. On automatic generation of multimedia presentations. *Inf. Sci.* 97, 3–4 (1997), 293–321. DOI:http://dx.doi.org/10.1016/S0020-0255(96)00195-8

C.-M. Chung, T. K. Shih, J.-Y. Huang, Y.-H. Wang, and T.-F. Kuo. 1995. An object-oriented approach and system for intelligent multimedia presentation designs. In *Proceedings of the International Conference on Multimedia Computing and Systems, 1995*. IEEE, 278–281. DOI:http://dx.doi.org/10.1109/MMCS.1995.484934

Cisco. 2014. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018. White Paper. Retrieved from http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf.

Clear-Media. 2012. ConciseClick: clickable video made easy. Retrieved from http://conciseclick.com/.

Clickable Video. 2012. CLICKABLE VIDEO. Add a call to action to your video. Retrieved from http://www.clickablevideo.eu/.

N. Correia and T. Chambel. 1999. Active video watching using annotation. In *Proceedings of the 7th ACM International Conference on Multimedia (Part 2) (MULTIMEDIA'99)*, Vol. 99. ACM, New York, 151–154. DOI:http://dx.doi.org/10.1145/319878.319919

S. Cutts, P. Davies, D. Newell, and N. Rowe. 2009. Requirements for an adaptive multimedia presentation system with contextual supplemental support media. In *Proceedings of the 1st International Conference on Advances in Multimedia (MMEDIA'09)*. IEEE, 62–67. DOI:http://dx.doi.org/10.1109/MMEDIA.2009.19

CWI. 2010. AMBULANT open SMIL player (2.4). Retrieved from http://www.ambulantplayer.org/.

X. Dai, S. Tabirca, and E. Lenihan. 2006. JAS—An e-learning tool for building multimedia presentations. In *Proceedings of the 1st International Multi-Symposiums on Computer and Computational Sciences (IMSCCS'06)*, Vol. 1. IEEE, 743–746. DOI:http://dx.doi.org/10.1109/IMSCCS.2006.81

R. Deltour and C. Roisin. 2006. The Limsee3 multimedia authoring model. In *Proceedings of the 2006 ACM Symposium on Document Engineering (DocEng'06)*. ACM, New York, 173–175. DOI:http://dx.doi.org/10.1145/1166160.1166203

L. Y. Deng, R.-X. Chen, R.-C. Chang, and T.-S. Huang. 2002a. Adaptive content model for multimedia presentation. In *Proceedings of the 1st International Symposium on Cyber Worlds, 2002*. IEEE, 209–216. DOI:http://dx.doi.org/10.1109/CW.2002.1180881

L. Y. Deng, T. K. Shih, S.-H. Shiau, W.-C. Chang, and Y.-J. Liu. 2002b. Implementing a distributed lecture-on-demand multimedia presentation system. In *Proceedings of the 22nd International Conference on Distributed Computing Systems Workshops, 2002*. IEEE, 111–115. DOI:http://dx.doi.org/10.1109/ICDCSW.2002.1030756

J. Doherty, A. Girgensohn, J. Helfman, F. Shipman, and L. Wilcox. 2003. Detail-on-demand hypervideo. In *Proceedings of the 11th ACM International Conference on Multimedia (MULTIMEDIA'03)*. ACM, New York, 600–601. DOI:http://dx.doi.org/10.1145/957135.957140

M. Echiffre, C. Marchisio, P. Marchisio, P. Panicciari, and S. Del Rossi. 1998. MHEG-5—Aims, concepts, and implementation issues. *IEEE Multimedia* 5, 1 (1998), 84–91. DOI:http://dx.doi.org/10.1109/93.664745

Z. Fei, M. H. Ammar, I. Kamel, and S. Mukherjee. 1999. Providing interactive functions through active client-buffer management in partitioned video multicast VoD systems. In *Networked Group Communication*, L. Rizzo and S. Fdida (Eds.). Lecture Notes in Computer Science, Vol. 1736. Springer, Berlin, 152–169. DOI:http://dx.doi.org/10.1007/978-3-540-46703-8_10

Z. Fei, M. H. Ammar, I. Kamel, and S. Mukherjee. 2005. An active buffer management technique for providing interactive functions in broadcast video-on-demand systems. *IEEE Trans. Multimedia* 7, 5 (2005), 942–950. DOI:http://dx.doi.org/10.1109/TMM.2005.854403

I. A. Fernandez, C. De Vleeschouwer, G. Toma, and L. Schumacher. 2012. An interactive video streaming architecture featuring bitrate adaptation. *J. Commun.* 7, 4 (2012), 265–280. DOI:http://dx.doi.org/10.4304/jcm.7.4.265-280

M. Finke and D. Balfanz. 2004. A reference architecture supporting hypervideo content for ITV and the internet domain. *Comput. Graph.* 28, 2 (2004), 179–191. DOI:http://dx.doi.org/10.1016/j.cag.2003.12.005

E. A. Fox. 1989. The coming revolution in interactive digital video. *Commun. ACM* 32, 7 (1989), 794–801. DOI:http://dx.doi.org/10.1145/65445.65446

O. Gaggi and A. Celentano. 2002. A visual authoring environment for prototyping multimedia presentations. In *Proceedings of the 4th International Symposium on Multimedia Software Engineering, 2002*. IEEE, 206–213. DOI:http://dx.doi.org/10.1109/MMSE.2002.1181614

D. Gotz. 2006. Scalable and adaptive streaming for non-linear media. In *Proceedings of the 14th Annual ACM International Conference on Multimedia (MULTIMEDIA'06)*. ACM, New York, 357–366. DOI:http://dx.doi.org/10.1145/1180639.1180717

V. Hakkoymaz. 2009. A specification model for temporal events in multimedia presentations. In *1st International Conference on Networked Digital Technologies (NDT'09)*. IEEE, 417–422. DOI:http://dx.doi.org/10.1109/NDT.2009.5272109

L. Hardman and D. C. A. Bulterman. 1997. Document model issues for hypermedia. In *Handbook of Multimedia Information Management*, W. I. Grosky, R. Jain, and R. Mehrotra (Eds.). Prentice-Hall, Inc., Upper Saddle River, NJ, 39–68.

L. Hardman, D. C. A. Bulterman, and G. van Rossum. 1994. The Amsterdam hypermedia model: Adding time and context to the dexter model. *Commun. ACM* 37, 2 (1994), 50–62. DOI:http://dx.doi.org/10.1145/175235.175239

L. Hardman, M. Worring, and D. C. A. Bulterman. 1997. Integrating the Amsterdam hypermedia model with the standard reference model for intelligent multimedia presentation systems. *Comput. Stand. Interfaces* 18, 6–7 (1997), 497–507. DOI:http://dx.doi.org/10.1016/S0920-5489(97)00014-7

M. Hausenblas. 2008. Non-linear interactive media productions. *Multimedia Syst.* 14, 6 (2008), 405–413. DOI:http://dx.doi.org/10.1007/s00530-008-0131-3

I. Herman, G. J. Reynolds, and J. van Loo. 1996a. Premo: An emerging standard for multimedia presentation part I: Overview and framework. *IEEE Multimedia* 3, 3 (1996), 83–89. DOI:http://dx.doi.org/10.1109/MMUL.1996.556543

I. Herman, G. J. Reynolds, and J. van Loo. 1996b. Premo: An emerging standard for multimedia presentation. Part II: Specification and applications. *IEEE Multimedia* 3, 4 (1996), 72–75. DOI:http://dx.doi.org/10.1109/93.556462

P. Hoffmann and M. Herczeg. 2006. Hypervideo vs. storytelling integrating narrative intelligence into hypervideo. In *Technologies for Interactive Digital Storytelling and Entertainment*, S. Göbel, R. Malkewitz, and I. Iurgel (Eds.). Lecture Notes in Computer Science, Vol. 4326. Springer, Berlin, 37–48. DOI:http://dx.doi.org/10.1007/11944577_4

Honkytonk Films. 2013. Klynt. Retrieved from http://www.klynt.net/.

H.-H. Hsu, T. Shih, H.-B. Chang, Y.-C. Liao, and C.-T. Tang. 2005. Hyper-interactive video browsing by a remote controller and hand gestures. In *Proceedings of the Iternational Conference on Embedded and Ubiquitous Computing (EUC 2005)*, T. Enokido, L. Yan, B. Xiao, D. Kim, Y. Dai, and L. T. Yang (Eds.). Lecture Notes in Computer Science, Vol. 3823. Springer, Berlin, 547–555. DOI:http://dx.doi.org/10.1007/11596042_57

J. Hu and L. Feijs. 2006. IPML: Extending SMIL for distributed multimedia presentations. In *Interactive Technologies and Sociotechnical Systems*, H. Zha, Z. Pan, H. Thwaites, A. C. Addison, and M. Forte (Eds.). Lecture Notes in Computer Science, Vol. 4270. Springer, Berlin, 60–70. DOI:http://dx.doi.org/10.1007/11890881_8

C.-M. Huang and C. Wang. 1998. Synchronization for interactive multimedia presentations. *IEEE Multimedia* 5, 4 (1998), 44–62. DOI:http://dx.doi.org/10.1109/93.735868

Y.-C. Hung. 1997. A hypervideo system generator. *Softw.: Prac. Exper.* 27, 11 (1997), 1263–1281. DOI:http://dx.doi.org/10.1002/(SICI)1097-024X(199711)27:11<1263::AID-SPE129>3.0.CO;2-3

imc AG. 2010. LECTURNITY 4. (2010). Website (accessed April 26, 2014). http://www.lecturnity.com/en/lecturnity/features/.

Incisive Interactive Marketing LLC and T. Wegert. 2006. What ever happened to clickable video? Retrieved from http://www.clickz.com/clickz/column/1704276/what-ever-happened-clickable- video.

ISO/IEC. 2009. MPEG-7. ISO/IEC 15938. Multimedia content description interface. Retrieved from http://mpeg.chiariglione.org/standards/mpeg-7.

ITU. International Telecommunication Union. 2011. H.761: Nested context language (NCL) and Ginga-NCL. Retrieved from http://www.itu.int/rec/T-REC-H.761-201106-I/en.

R. Jain and K. Wakimoto. 1995. Multiple perspective interactive video. In *Proceedings of the International Conference on Multimedia Computing and Systems, 1995*. 202–211. DOI:http://dx.doi.org/10.1109/MMCS.1995.484925

J. Jansen and D. C. A. Bulterman. 2008. Enabling adaptive time-based web applications with SMIL state. In *Proceeding of the 8th ACM Symposium on Document Engineering (DocEng'08)*. ACM, New York, 18–27. DOI:http://dx.doi.org/10.1145/1410140.1410146

J. Jansen and D. C. A. Bulterman. 2009. SMIL State: An architecture and implementation for adaptive time-based web applications. *Multimedia Tools Appl.* 43, 3 (2009), 203–224. DOI:http://dx.doi.org/10.1007/s11042-009-0270-3

J. Jansen, P. Cesar, and D. C. A. Bulterman. 2010. A model for editing operations on active temporal multimedia documents. In *Proceedings of the 10th ACM Symposium on Document Engineering (DocEng'10)*. ACM, New York, 87–96. DOI:http://dx.doi.org/10.1145/1860559.1860579

H. Jiang and A. K. Elmagarmid. 1998. Spatial and temporal content-based access to hypervideo databases. *VLDB J.* 7, 4 (1998), 226–238. DOI:http://dx.doi.org/10.1007/s007780050066

T. Jokela, J. T. Lehikoinen, and H. Korhonen. 2008. Mobile multimedia presentation editor: Enabling creation of audio-visual stories on mobile devices. In *Proceedings of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems (CHI'08)*. ACM, New York, 63–72. DOI:http://dx.doi.org/10.1145/1357054.1357066

M. Jourdan, N. Layaïda, C. Roisin, L. Sabry-Ismaïl, and L. Tardif. 1998. Madeus, an authoring environment for interactive multimedia documents. In *Proceedings of the 6th ACM International Conference on Multimedia (MULTIMEDIA'98)*. ACM, New York, 267–272. DOI:http://dx.doi.org/10.1145/290747.290780

A. Katkere, J. Schlenzig, A. Gupta, and R. Jain. 1996. Interactive video on WWW: Beyond VCR-like interfaces. *Comput. Network. ISDN Syst.* 28, 7–11 (1996), 1559–1572. DOI:http://dx.doi.org/10.1016/0169-7552(96)00025-6

P. H. Kelly, A. Katkere, D. Y. Kuramura, S. Moezzi, and S. Chatterjee. 1995. An architecture for multiple perspective interactive video. In *Proceedings of the 3rd ACM International Conference on Multimedia (MULTIMEDIA'95)*. ACM, New York, 201–212. DOI:http://dx.doi.org/10.1145/217279.215267

C. Keramane and A. Duda. 1997. Operator based composition of structured multimedia presentations. *From Multimedia Services to Network Services* 1356 (1997), 1–32. Website (accessed April 26, 2014). http://www.springerlink.com/index/D5568R436290594M.pdf.

P. King, P. Schmitz, and S. Thompson. 2004. Behavioral reactivity and real time programming in XML: Functional programming meets SMIL animation. In *Proceedings of the 2004 ACM Symposium on Document Engineering (DocEng'04)*. ACM, New York, 57–66. DOI:http://dx.doi.org/10.1145/1030397.1030411

B. Kitchenham and S. Charters. 2007. *Guidelines for Performing Systematic Literature Reviews in Software Engineering—Version 2.3*. EBSE Technical Report. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.117.471.

T. Kleinberger, A. Holzinger, and P. Müller. 2008. Adaptive multimedia presentations enabling universal access in technology enhanced situational learning. *Univ. Access Inf. Soc.* 7, 4 (2008), 223–245. DOI:http://dx.doi.org/10.1007/s10209-008-0122-3

M. Kozuch, W. Wolf, and A. Wolfe. 2000. An experimental analysis of digital video library servers. *Multimedia Syst.* 8, 2 (2000), 135–145. DOI:http://dx.doi.org/10.1007/s005300050156

N. Layaida and L. Sabry-Ismail. 1996. Maintaining temporal consistency of multimedia documents using constraint networks. In *Proc. SPIE 2667, Multimedia Computing and Networking 1996*. SPIE Digital Library, 124–135. DOI:http://dx.doi.org/10.1117/12.235866

Y.-W. Leung and T. K. C. Chan. 2003. Design of an interactive video-on-demand system. *IEEE Trans. Multimedia* 5, 1 (2003), 130–140. DOI:http://dx.doi.org/10.1109/TMM.2003.808818

W. Liao and V. O. K. Li. 1998. Synchronization of distributed multimedia systems with user interactions. *Multimedia Syst.* 6, 3 (1998), 196–205. DOI:http://dx.doi.org/10.1007/s005300050088

LinkedTV Consortium. 2013. LinkedTV. Retrieved from http://www.linkedtv.eu/.

T. D. C. Little and A. Ghafoor. 1990. Synchronization and storage models for multimedia objects. *IEEE J. Sel. Areas Commun.* 8, 3 (1990), 413–427. DOI:http://dx.doi.org/10.1109/49.53017

LTSC. IEEE Learning Technology Standards Committee. 2002. Learning Object Metadata (LOM), (1484.12.1-2002 IEEE Standard for Learning Object Metadata). Retrieved from http://ltsc.ieee.org/wg12/.

W. E. Mackay and G. Davenport. 1989. Virtual video editing in interactive multimedia applications. *Commun. ACM* 32, 7 (1989), 802–810. DOI:http://dx.doi.org/10.1145/65445.65447

R. MacNeil. 1991. Generating multimedia presentations automatically using TYRO, the constraint, case-based designer's apprentice. In *Proceedings of the 1991 IEEE Workshop on Visual Languages*. IEEE, 74–79. DOI:http://dx.doi.org/10.1109/WVL.1991.238847

MatchWare A/S. 2012. Mediator 9—Multimedia authoring software. Retrieved from http://www.matchware.com/en/products/mediator/default.htm.

T. Maugey and P. Frossard. 2011. Interactive multiview video system with low decoding complexity. In *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP'11)*. IEEE, 589–592. DOI:http://dx.doi.org/10.1109/ICIP.2011.6116618

MDN. Mozilla Developer Network and individual contributors. 2013. JavaScript. Retrieved from https://developer.mozilla.org/en-US/docs/JavaScript.

S. R. L. Meira and A. E. L. Moura. 1994. A scripting language for multimedia presentations. In *Proceedings of the IEEE International Conference on Multimedia Computing and Systems, 1994*. IEEE, 484–489. DOI:http://dx.doi.org/10.1109/MMCS.1994.292493

B. Meixner. 2014. *Annotated Interactive Non-linear Video—Software Suite, Download and Cache Management*. PhD thesis. Universität Passau. Retrieved from https://opus4.kobv.de/opus4-uni-passau/files/222/Meixner_Britta.pdf.

B. Meixner and C. Einsiedler. 2016. Download and cache management for HTML5 hypervideo players. In *Proceedings of the 27th ACM Conference on Hypertext and Social Media (HT'16)*. ACM, New York, 125–136. DOI:http://dx.doi.org/10.1145/2914586.2914587

B. Meixner, S. John, and C. Handschigl. 2015. SIVA suite: Framework for hypervideo creation, playback and management. In *Proceedings of the 23rd ACM International Conference on Multimedia (MM'15)*. ACM, New York, 713–716. DOI:http://dx.doi.org/10.1145/2733373.2807413

B. Meixner and H. Kosch. 2012. Interactive non-linear video: Definition and XML structure. In *Proceedings of the 2012 ACM Symposium on Document Engineering (DocEng'12)*. ACM, New York, 49–58. DOI:http://dx.doi.org/10.1145/2361354.2361367

B. Meixner, K. Matusik, C. Grill, and H. Kosch. 2012. Towards an easy to use authoring tool for interactive non-linear video. *Multimedia Tools Appl.* (2012), 1–26. DOI:http://dx.doi.org/10.1007/s11042-012-1218-6

B. Meixner, B. Siegel, G. Hölbling, F. Lehner, and H. Kosch. 2010. SIVA Suite—Authoring system and player for interactive non-linear videos. In *Proceedings of the International Conference on Multimedia (MM'10)*. ACM, New York, 1563–1566. DOI:http://dx.doi.org/10.1145/1873951.1874287

B. Meixner, B. Siegel, P. Schultes, F. Lehner, and H. Kosch. 2013. An HTML5 player for interactive non-linear video with time-based collaborative annotations. In *Proceedings of the International Conference on Advances in Mobile Computing & Multimedia (MoMM'13)*, R. Mayrhofer, L. Chen, M. Steinbauer, I. Khalil, and G. Kotsis (Eds.). ACM, New York, 490–498. DOI:http://dx.doi.org/10.1145/2536853.2536868

B. Meixner, K. Tonndorf, S. John, C. Handschigl, K. Hofmann, and M. Granitzer. 2014. A multimedia help system for a medical scenario in a rehabilitation clinic. In *Proceedings of the 14th International Conference on Knowledge Technologies and Data-Driven Business (i-KNOW'14)*. ACM, New York, Article 25, 8 pages. DOI:http://dx.doi.org/10.1145/2637748.2638429

J. Mikác, C. Roisin, and B. Le Duc. 2008. An export architecture for a multimedia authoring environment. In *Proceedings of the 8th ACM Symposium on Document Engineering (DocEng'08)*. ACM, New York, 28–31. DOI:http://dx.doi.org/10.1145/1410140.1410147

G. Miller, S. Fels, M. Ilich, M. M. Finke, T. Bauer, K. Wong, and S. Mueller. 2011. An end-to-end framework for multi-view video content: Creating multiple-perspective hypervideo to view on mobile platforms. In *Proceedings of Entertainment Computing (ICEC'11)*, J. C. Anacleto, S. Fels, N. Graham, B. Kapralos, M. Saif El-Nasr, and K. Stanley (Eds.). Lecture Notes in Computer Science, Vol. 6972. Springer, Berlin, 337–342. DOI:http://dx.doi.org/10.1007/978-3-642-24500-8_37

J. Monaco. 2009. *How to Read a Film: Movies, Media, and Beyond* (4th, illustrated ed.). Oxford University Press.

Mozilla Corporation. 2013. Popcorn Maker. Retrieved from https://popcorn.webmaker.org.

D. C. Muchaluat-Saade, R. F. Rodrigues, and L. F. G. Soares. 2002. XConnector: Extending XLink to provide multimedia synchronization. In *Proceedings of the 2002 ACM Symposium on Document Engineering (DocEng'02)*. ACM, New York, 49–56. DOI:http://dx.doi.org/10.1145/585058.585069

S. Mujacic and M. Debevc. 2007. A formal approach to hypervideo design. In *Proceedings of the 14th International Workshop on Systems, Signals and Image Processing, 2007 and 6th EURASIP Conference Focused on Speech and Image Processing, Multimedia Communications and Services*. IEEE, 189–192. DOI:http://dx.doi.org/10.1109/IWSSIP.2007.4381185

U. Murthy, K. Ahuja, S. Murthy, and E. A. Fox. 2006. SIMPEL: A superimposed multimedia presentation editor and player. In *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'06)*. IEEE, 377–377. DOI:http://dx.doi.org/10.1145/1141753.1141873

A. T. Naman and D. Taubman. 2011. JPEG2000-based scalable interactive video (JSIV). *IEEE Trans. Image Process.* 20, 5 (2011), 1435–1449. DOI:http://dx.doi.org/10.1109/TIP.2010.2093905

Y. Nimmagadda, K. Kumar, and Y.-H. Lu. 2009. Preference-based adaptation of multimedia presentations for different display sizes. In *Proceedings of the 2009 IEEE International Conference on Multimedia and Expo (ICME'09)*. IEEE, 978–981. DOI:http://dx.doi.org/10.1109/ICME.2009.5202660

D. Ockeloen and P. van Leeuwen. 2013. Deliverable 3.6: Interface and presentation engine version 2. Retrieved from http://www.linkedtv.eu/wp/wp-content/uploads/2013/12/LinkedTV_D3.6.pdf.

E. Parsloe. 1983. Interactive video. *J. Eur. Indust. Training* 7, 3 (1983), 28–32. DOI:http://dx.doi.org/10.1108/eb002148

K. Petridis, D. Anastasopoulos, C. Saathoff, N. Timmermann, Y. Kompatsiaris, and S. Staab. 2006. M-OntoMat-Annotizer: Image annotation linking ontologies and multimedia low-level features. In *Knowledge-Based Intelligent Information and Engineering Systems*, Lecture Notes in Computer Science, Vol. 4253, B. Gabrys, R. J. Howlett, and L. C. Jain (Eds.). Springer, Berlin, 633–640. DOI:http://dx.doi.org/10.1007/11893011_80

K. Pihkala and P. Vuorimaa. 2006. Nine methods to extend SMIL for multimedia applications. *Multimedia Tools Appl.* 28, 1 (2006), 51–67. DOI:http://dx.doi.org/10.1007/s11042-006-5120-y

S. L. Presti, D. Bert, and A. Duda. 2002. TAO: Temporal algebraic operators for modeling multimedia presentations. *J. Netw. Comput. Appl.* 25, 4 (2002), 319–342. DOI:http://dx.doi.org/10.1006/jnca.2002.0135

A. Ram, R. Catrambone, M. J. Guzdial, C. M. Kehoe, D. S. McCrickard, and J. T. Stasko. 1999. PML: Adding flexibility to multimedia presentations. *IEEE MultiMedia* 6, 2 (1999), 40–52. DOI:http://dx.doi.org/10.1109/93.771372

L. J. Redondo-Garcia and R. Troncy. 2013. Television meets the web: A multimedia hypervideo experience. In *Proceedings of the Doctoral Consortium co-located with 12th International Semantic Web Conference (ISWC 2013)*, L. Aroyo and N. Noy (Eds.). 48–55. Retrieved from http://ceur-ws.org/Vol-1045/paper-07.pdf.

ReelSEO. 2011. How to overcome the pitfalls of interactive, clickable video. Retrieved from http://www.reelseo.com/pitfalls-clickable-video.

H. Sack and J. Waitelonis. 2010. Exploratory semantic video search with yovisto. In *Proceedings of the 2010 IEEE 4th International Conference on Semantic Computing (ICSC)*. IEEE, 446–447. DOI:http://dx.doi.org/10.1109/ICSC.2010.98

M. Sadallah, O. Aubert, and Y. Prié. 2011. Component-based hypervideo model: High-level operational specification of hypervideos. In *Proceedings of the 11th ACM Symposium on Document Engineering (DocEng'11)*. ACM, New York, 53–56. DOI:http://dx.doi.org/10.1145/2034691.2034701

N. Sawhney, D. Balcom, and I. Smith. 1996. HyperCafe: Narrative and aesthetic properties of hypervideo. In *Proceedings of the 7th ACM Conference on Hypertext (HYPERTEXT'96)*. ACM, New York, 1–10. DOI:http://dx.doi.org/10.1145/234828.234829

N. Sawhney, D. Balcom, and I. Smith. 1997. Authoring and navigating video in space and time. *IEEE Multimedia* 4, 4 (1997), 30–39. DOI:http://dx.doi.org/10.1109/93.641877

A. Scherp and S. Boll. 2005. Paving the last mile for multi-channel multimedia presentation generation. In *Proceedings of the 11th International Multimedia Modelling Conference (MMM'05)*. IEEE, 190–197. DOI:http://dx.doi.org/10.1109/MMMC.2005.58

G. Schloss and M. Wynblatt. 1994. Building temporal structures in a layered multimedia data model. In *Proceedings of the 2nd ACM International Conference on Multimedia (MULTIMEDIA'94)*. ACM, New York, 271–278. DOI:http://dx.doi.org/10.1145/192593.192674

P. Schmitz, J. Yu, and P. Santangeli. 1998. Timed interactive multimedia extensions for HTML (HTML+TIME): Extending SMIL into the Web browser. Retrieved from http://www.w3.org/TR/NOTE-HTMLplusTIME.

J. Schnepf, J. A. Konstan, and D. H.-C. Du. 1996. Doing FLIPS: Flexible interactive presentation synchronization. *IEEE J. Select, Areas Commun.* 14, 1 (1996), 114–125. DOI:http://dx.doi.org/10.1109/49.481698

K. Schoeffmann, M. A. Hudelist, and J. Huber. 2015. Video interaction tools: A survey of recent work. *ACM Comput. Surv.* 48, 1, Article 14 (Sept. 2015), 34 pages. DOI:http://dx.doi.org/10.1145/2808796

N. Seidel. 2011. Enable wikis for seamless hypervideo integration. In *Proceedings of the 29th Annual European Conference on Cognitive Ergonomics (ECCE'11)*. ACM, New York, 251–252. DOI:http://dx.doi.org/10.1145/2074712.2074765

T. K. Shih. 1998. Participator dependent multimedia presentation. *Inf. Sci.* 107, 1–4 (1998), 85–105. DOI:http://dx.doi.org/10.1016/S0020-0255(97)10043-3

T. K. Shih, H.-C. Keh, Y.-H. Wang, and Y.-F. Kuo. 1999. Temporal properties underlying multimedia presentations with Z notations. *J. Inf. Sci. Eng.* 15, 1 (1999), 107–129. Retrieved from http://www.iis.sinica.edu.tw/page/jise/1999/199901_08.pdf.

D.-H. Shin. 2013. Defining sociability and social presence in social TV. *Comput. Human Behav.* 29, 3 (2013), 939–947. DOI:http://dx.doi.org/10.1016/j.chb.2012.07.006

F. Shipman, A. Girgensohn, and L. Wilcox. 2003. Hyper-hitchcock: Towards the easy authoring of interactive video. In *Proceedings of the Conference on Human-Computer Interaction (INTERACT'03): IFIP TC13*, M. Rauterberg, M. Menozzi, and J. Wesson (Eds.). IOS Press. Retrieved from http://www.fxpal.com/publications/FXPAL-PR-03-219.pdf.

F. Shipman, A. Girgensohn, and L. Wilcox. 2005. Hypervideo expression: Experiences with Hyper-Hitchcock. In *Proceedings of the 16th ACM Conference on Hypertext and Hypermedia (HYPERTEXT'05)*. ACM, New York, 217–226. DOI:http://dx.doi.org/10.1145/1083356.1083401

F. Shipman, A. Girgensohn, and L. Wilcox. 2008. Authoring, viewing, and generating hypervideo: An overview of Hyper-Hitchcock. *ACM Trans. Multimedia Comput., Commun., Appl.* 5, 2 (2008), 15:1–15:19. DOI:http://dx.doi.org/10.1145/1413862.1413868

H. V. O. Silva, R. F. Rodrigues, L. F. G. Soares, and D. C. Muchaluat Saade. 2004. NCL 2.0: Integrating new concepts to XML modular languages. In *Proceedings of the 2004 ACM Symposium on Document Engineering (DocEng'04)*. ACM, New York, 188–197. DOI:http://dx.doi.org/10.1145/1030397.1030433

L. F. G. Soares and R. F. Rodrigues. 2005. *Nested Context Model 3.0. Part 1—NCM Core*. Technical Report No. 18. Laboratorio TeleMidia DI - PUC-Rio, Rio de Janeiro. Retrieved from http://www.telemidia.puc-rio.br/sites/telemidia.puc-rio.br/files/Part1-NC M3.0.pdf.

L. F. G. Soares, R. F. Rodrigues, R. Cerqueira, and S. D. J. Barbosa. 2010. Variable and state handling in NCL. *Multimedia Tools Appl.* 50, 3 (2010), 465–489. DOI:http://dx.doi.org/10.1007/s11042-010-0478-2

M. Spaniol, R. Klamma, N. Sharda, and M. Jarke. 2006. Web-based learning with non-linear multimedia stories. In *Proceedings of Advances in Web Based Learning (ICWL 2006)*, W. Liu, Q. Li, and R. W.H. Lau (Eds.). Lecture Notes in Computer Science, Vol. 4181. Springer, Berlin, 249–263. DOI:http://dx.doi.org/10.1007/11925293_23

R. Spicer, Y.-R. Lin, A. Kelliher, and H. Sundaram. 2012. NextSlidePlease: Authoring and delivering agile multimedia presentations. *ACM Trans. Multimedia Comput., Communic., Appl.* 8, 4 (2012), 53:1–53:20. DOI:http://dx.doi.org/10.1145/2379790.2379795

E. Stahl, C. Zahn, and M. Finke. 2005. How can we use hypervideo design projects to construct knowledge in university courses? In *Proceedings of the 2005 Conference on Computer Support for Collaborative Learning 2005: The Next 10 Years! (CSCL'05)*. International Society of the Learning Sciences, Morristown, NJ, 641–646. DOI:http://dx.doi.org/10.3115/1149293.1149377

M. Y. Sung and D. H. Lee. 2005. A Java-based collaborative authoring system for multimedia presentation. In *Proceedings of Advances in Multimedia Information Processing (PCM 2004)*, K. Aizawa, Y. Nakamura, and S. Satoh (Eds.). Lecture Notes in Computer Science, Vol. 3332. Springer, Berlin, 96–103. DOI:http://dx.doi.org/10.1007/978-3-540-30542-2_13

R. Tan and S.-U. Guan. 2005. A dynamic Petri net model for iterative and interactive distributed multimedia presentation. *IEEE Trans. Multimedia* 7, 5 (2005), 869–879. DOI:http://dx.doi.org/10.1109/TMM.2005.854377

TeleMidia Lab - PUC-Rio. 2011. NCL—Nested Context Language. Retrieved from http://www.ncl.org.br/en/inicio.

K. Tonndorf, T. Knieper, B. Meixner, H. Kosch, and F. Lehner. 2012. Challenges in creating multimedia instructions for support systems and dynamic problem-solving. In *Proceedings the 12th International Conference on Knowledge Management and Knowledge Technologies (i-KNOW'12)*. ACM, New York, 33:1–33:4. DOI:http://dx.doi.org/10.1145/2362456.2362497

M. F. Ursu, I. C. Kegel, D. Williams, M. Thomas, H. Mayer, V. Zsombori, M. L. Tuomola, H. Larsson, and J. Wyver. 2008. ShapeShifting TV: Interactive screen media narratives. *Multimedia Syst.* 14, 2 (2008), 115–132. DOI:http://dx.doi.org/10.1007/s00530-008-0119-z

R. Vaisenberg, R. Jain, and S. Mehrotra. 2009. SMPL, a specification based framework for the semantic structure, annotation and control of SMIL documents. In *Proceedings of the 11th IEEE International Symposium on Multimedia (ISM'09)*. IEEE, 533–539. DOI:http://dx.doi.org/10.1109/ISM.2009.114

S. Van Assche, F. Hendrickx, N. Oorts, and L. Nachtergaele. 2004. Multi-channel publishing of interactive multimedia presentations. *Comput. Graphics* 28, 2 (2004), 193–206. DOI:http://dx.doi.org/10.1016/j.cag.2003.12.006

G. van Rossum, J. Jansen, K. S. Mullender, and D. C. A. Bulterman. 1993. CMIFed: A presentation environment for portable hypermedia documents. In *Proceedings of the 1st ACM International Conference on Multimedia (MULTIMEDIA'93)*. ACM, New York, 183–188. DOI:http://dx.doi.org/10.1145/166266.166287

VideoClix Technologies Inc. 2012. VideoClix. Turning viewers into customers. Retrieved from http://www.videoclix.tv.

L. Villard. 2001. Authoring transformations by direct manipulation for adaptable multimedia presentations. In *Proceedings of the 2001 ACM Symposium on Document Engineering (DocEng'01)*. ACM, New York, 125–134. DOI:http://dx.doi.org/10.1145/502204.502206

W3C. 2002. XHTML+SMIL profile. Retrieved from http://www.w3.org/TR/XHTMLplusSMIL/.

W3C. 2012. Synchronized Multimedia (SMIL). Retrieved from http://www.w3.org/AudioVideo/.

W3C. 2013a. Cascading Style Sheets home page. Retrieved from http://www.w3.org/Style/CSS/.

W3C. 2013b. HTML5—A vocabulary and associated APIs for HTML and XHTML (W3C candidate recommendation 6 August 2013). Retrieved from http://www.w3.org/TR/html5/Overview.html.

W3C. 2013c. Scalable Vector Graphics (SVG). Retrieved from http://www.w3.org/Graphics/SVG/.

T. Wahl, S. Wirag, and K. Rothermel. 1995. TIEMPO: Temporal modeling and authoring of interactive multimedia. In *Proceedings of the International Conference on Multimedia Computing and Systems, 1995*. IEEE, 274–277. DOI:http://dx.doi.org/10.1109/MMCS.1995.484933

J. Waitelonis, N. Ludwig, and H. Sack. 2011. Use what you have: Yovisto video search engine takes a semantic turn. In *Semantic Multimedia*, T. Declerck, M. Granitzer, M. Grzegorzek, M. Romanelli, S. Rüger, and M. Sintek (Eds.). Lecture Notes in Computer Science, Vol. 6725. Springer, Berlin, 173–185. DOI:http://dx.doi.org/10.1007/978-3-642-23017-2_12

Z. Wang, L. Sun, X. Chen, W. Zhu, J. Liu, M. Chen, and S. Yang. 2012. Propagation-based social-aware replication for social video contents. In *Proceedings of the 20th ACM International Conference on Multimedia (MM'12)*. ACM, New York, 29–38. DOI:http://dx.doi.org/10.1145/2393347.2393359

J. A. Watlington. 1987. *Synthetic Movies*. Master's thesis. Massachusetts Institute of Technology. Retrieved from http://alumni.media.mit.edu/~wad/ms-thesis/ms-thesis.pdf.

WireWax ltd. 2012. wireWAX. Retrieved from http://www.wirewax.com.

J. Wong, S. Kini, and K. Doobagunta. 1996. Synchronization in specification-based multimedia presentations. *Software: Practice and Experience* 26, 1 (1996), 71–81. DOI:http://dx.doi.org/10.1002/(SICI)1097-024X(199601)26:1<71::AID-SPE997>3.0.CO;2-R

J. Wong, S. Rao, and N. Ramaiah. 1997. A multimedia presentation toolkit for the World Wide Web. *Softw.: Pract. Exper.* 27, 4 (1997), 425–446. DOI:http://dx.doi.org/10.1002/(SICI)1097-024X(199704)27:4<425::AID-SPE92>3.0.CO;2-2

X. Xiu, G. Cheung, and J. Liang. 2012. Delay-cognizant interactive streaming of multiview video with free viewpoint synthesis. *IEEE Trans. Multimedia* 14, 4 (2012), 1109–1126. DOI:http://dx.doi.org/10.1109/TMM.2012.2191267

C.-C. Yang, C.-K. Chu, and Y.-C. Wang. 2008. Extension of timeline-based editing for non-deterministic temporal behavior in SMIL2.0 authoring. *J. Inf. Sci. Eng.* 24, 5 (2008), 1377–1395. Retrieved from http://erdos.csie.ncnu.edu.tw/ccyang/Publication/JISE2008-3.pdf.

C.-C. Yang and Y.-Z. Yang. 2003. SMILAuthor: An authoring system for SMIL-based multimedia presentations. *Multimedia Tools Appl.* 21, 3 (2003), 243–260. DOI:http://dx.doi.org/10.1023/A:1025770817293

M. Yeung, B.-L. Yeo, and B. Liu. 1996. Extracting story units from long programs for video browsing and navigation. In *Proceedings of the 3rd IEEE International Conference on Multimedia Computing and Systems, 1996*. IEEE, 296–305. DOI:http://dx.doi.org/10.1109/MMCS.1996.534991

D. Zhang, L. Zhou, R. O. Briggs, and J. F. Nunamaker. 2006. Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Inf. Manag.* 43, 1 (2006), 15–27. DOI:http://dx.doi.org/10.1016/j.im.2005.01.004

Y. Zhao, D. L. Eager, and M. K. Vernon. 2007. Scalable on-demand streaming of nonlinear media. *IEEE/ACM Trans. Netw.* 15, 5 (2007), 1149–1162. DOI:http://dx.doi.org/10.1109/TNET.2007.896534

B. Zheng and M. Atiquzzaman. 2005. System design and network requirements for interactive multimedia. *IEEE Trans. Circ. Syst. Video Technol.* 15, 1 (2005), 145–153. DOI:http://dx.doi.org/10.1109/TCSVT.2004.839982

T. T. Zhou, T. Gedeon, and J. S. Jin. 2005. Automatic generating detail-on-demand hypervideo using MPEG-7 and SMIL. In *Proceedings of the 13th Annual ACM International Conference on Multimedia (MULTIMEDIA'05)*. ACM, New York, 379–382. DOI:http://dx.doi.org/10.1145/1101149.1101230

Received November 2015; revised January 2017; accepted January 2017