

# Open Standard { and Open Sourced SMIL for Interactivity

by Daniel F. Zucker | Wake3, Inc. | [zucker@stanfordalumni.org](mailto:zucker@stanfordalumni.org)

by Dick Bulterman | CWI | [dick.bulterman@cwi.nl](mailto:dick.bulterman@cwi.nl)

**SMIL, SYNCHRONIZED MULTIMEDIA INTEGRATION LANGUAGE**, was the first member in the family of open, XML-based standards developed and supported by the World Wide Web Consortium (W3C). It is an important resource in the interaction designer's toolbox. It can be used not only to develop time-based multimedia presentations, but also to implement media-rich interfaces for PC or embedded applications and devices. SMIL is supported by a number of open source players [1, 2], as well as by Microsoft's Internet Explorer, Apple's QuickTime, and RealNetwork's RealPlayer.

SMIL 1.0 was first published as a W3C recommendation in June 1998. The SYMM Workgroup [3], responsible for the SMIL specification, is planning for a new release of SMIL with updated and enhanced functionality as SMIL 3.0 [4]. SMIL 3.0 is expected to become a recommendation by early 2008. The main new features of SMIL 3.0 will be discussed later in this article.

SMIL is used to specify the time-based interactions between different multimedia objects. SMIL does not encode the multimedia objects themselves, but references Web addresses where the media can be found. In creating a multimedia presentation consisting of video, still images, audio, and captions, for example, the SMIL document would specify when each of these various media

elements is activated, as well as where they are rendered. Activation is triggered either by time-based rules or user interaction.

SMIL can be used to structure passive multimedia presentations that play without user input like a TV show. More interestingly, SMIL can be used to react to user input to build a more interactive presentation such as the news [5] shown in Figure 1. In this example, SMIL is used to create a control panel in which clicking still-picture key frames causes the audio-video program to start in the main window.

SMIL can be used for any application requiring synchronization and presentation of media in time. The DAISY Consortium [6] uses SMIL, for example, to control the playback of talking books for the visually impaired. SMIL is useful because it provides a standards-based structure to DAISY's time-based audio presentation. Using SMIL, the user can easily navigate throughout the document, rapidly moving to the next word, paragraph, or chapter. SMIL is also useful because it enforces the separation between presentation content and structure. This means that a common talking-book presentation structure can be used for many different individual talking books that differ only in media content.

By virtue of being an open standard, SMIL is well supported by the open source community. SMIL can be used license and royalty free. Combined



Figure 1: An interactive SMIL presentation  
Web Edition

with similar royalty-free media codecs, SMIL offers a complete multimedia platform comparable to other closed proprietary standards requiring hefty license fees.

**SMIL AND INTERACTION DESIGN.** For the interaction designer, SMIL is useful for any presentation or interface requiring time-based interactions with media. It is especially useful for media that can be repurposed for playback on different form factor devices.

SMIL has several advantages. SMIL enforces the separation of structure and media in a presentation or interface. By design, the media objects themselves are kept separate from the SMIL document describing the presentation. This greatly simplifies maintenance and eases the task of, for example, using a single SMIL document to describe a presentation format while updating the media to produce a new presentation. It also allows the media objects to change over time, even though the basic presentation logic remains the same.

SMIL uses a declarative format rather than a procedural scripting language such as EcmaScript. Advanced effects such as transitions and animations are supported implicitly in the language and require no heavy-duty scripting. The declarative nature of SMIL also allows it to be automatically generated or to provide a basis for further transformation and customization using standard Web tools.

SMIL is structured to allow efficient repurposing of content for adaptation to different types of display devices. SMIL has a content-control architecture that allows a designer to specify content alternatives for different devices or for differing user preferences. The run-time binding of objects to the presentation also allows a presentation's content model to change without having to redefine the SMIL container.

**SMIL STRUCTURE.** Compared with a temporally static markup language such as HTML or SVG (SVG without SMIL, that is), SMIL adds the concept of time. HTML and SVG describe how something looks at a single point in time—animated images, EcmaScript, and SMIL encapsulated in these formats are exceptions to this. Let's focus on the core language for this analogy; SMIL provides a framework to allow such presentations to vary over time.

To be clear, SVG does incorporate the concept of time for animations. It does this by including SMIL in a modular fashion using the SMIL animation module. In fact, SMIL's uses for SVG animation are perhaps one of the better-known uses of SMIL. This allows SVG elements to be animated over time. For example, in this way a clock can be constructed entirely in SVG without using EcmaScript.

MMS (multimedia messaging) is another well-known use of SMIL. Cell phone users can create, send, and receive slideshows. For MMS, the OMA (owners of the MMS specification) define SMIL as the presentation markup [7]. In this case, SMIL is the vehicle that allows the stringing together of single images into a time-based slideshow with optional transitions between the media. What is less known is that SMIL can do much more than this.

**Par, Seq, and Excl.** SMIL's declarative syntax to describe the interaction between time-based media relies on three primary constructs, or containers: par, seq, and excl. Par is for parallel, seq for sequential, and excl for exclusive. Two media objects placed in a par container play in parallel, meaning they both play at the same time. (Layout description likely has them located in two different places so that both media objects are visible in a meaningful way.) Two media objects placed in a seq play sequentially. That is, they play one after another. Excl was added for SMIL 2.0 and means that only one media element in the group of elements can play at a time. Usually, some sort of event logic is used to determine which media object is playing.

Each of SMIL's time containers define a local timeline, in which a group of related media objects can be managed. The nature of the time container provides a basic set of activation constraints that eases the designer's task of creating a presentation. A seq container imposes general slideshow-like temporal constraints among the objects: Adding new slides with default timing is easy, since only a new media reference needs to be added to the

SMIL file. None of the timings on individual objects needs to be changed. In a parallel container, a common multitrack timeline is defined that provides a common reference time base for the activation of multiple objects. The par and seq containers can be nested, allowing an audio track to accompany a slideshow, or to provide several logical collections of media objects within the same presentation context.

SMIL is used to determine the interaction between media elements. It does not define the media itself. Media elements can be audio, video, text, decorated text such as HTML or time-based text. The last is useful for subtitles.

**SMIL Timing.** The par/seq/excl time-container structure of SMIL provides a default set of timing relations among objects. In many cases, all timing within a presentation can be determined by simply placing objects in an appropriate time container. The default timing can be further refined by using a collection of timing attributes: attributes that add a specific begin offset, an end time, or a duration to a media object.

The timing attributes in SMIL override any timing within a particular media object. For example, a 15-second video clip can be trimmed or extended using SMIL timing attributes. The attributes clip-Begin/clipEnd also allow a fragment of a larger video to be played. Note that unlike formats such as MPEG4, which define a timeline based on the encoding of a particular video object, SMIL provides a flexible timeline that abstracts timing away from the media and into the overall presentation.

In addition to timing attribute control, SMIL also provides time navigation within a presentation via a temporal hyperlink architecture. Jumping from one object to another in a presentation has the effect of adjusting the presentation timeline to the context of the link's destination anchor. This allows all of the related content that would otherwise be active when the destination had been reached normally to also be active once the link is followed. It is the SMIL scheduler that determines the temporal relationship among elements, meaning that each of the individual media objects does not need to be aware of the presence of other media in the presentation—a major benefit of SMIL.

**SMIL Events.** Interaction in SMIL is provided by a declarative event-based architecture that distinguishes between internal player events (such

as a media object's beginning or terminating) and external user events (such as an object within the presentation—a next button or a navigation arrow—being selected interactively by a user).

Nearly all SMIL timing-related actions define a begin or end event. This allows companion media objects to be scheduled interactively, based on the duration or (conditional) activation of related content. No scripts are required to control this interactivity; instead, a begin (or end) condition is set on the companion object.

For user-centered interaction, SMIL provides a mechanism for associating user events such as mouse clicks or hovering activity with the start or end of either individual media objects or with SMIL timing containers. This allows basic interaction within a presentation with the need to invoke a scripting architecture. SMIL also provides basic support for interaction using a DOM interface, although many SMIL players do not yet support this feature.

## Multimedia Messaging:

MMS is a short text message optionally including media; the entire message can be played as a slideshow. A user can start with a short message using SMS on a cell phone, and add elements; some clients start with MMS. Adding specific fields changes the SMS to MMS, which is tarified at a higher rate. For example, adding a subject line or a CC to an SMS makes it MMS. MMS can also have media attachments—a photo, a video, or an audio clip—but only one attachment can be dynamic (audio or video). The user can add and rearrange pages. When the message is received, it can be played as a slideshow or viewed as a series of pages.

**REFERENCES** 1. Ambulant Open SMIL Player. <http://www.cwi.nl/projects/Ambulant/distPlayer.html> 2. "RealNetworks Releases Industry Standard SMIL Source Code to Helix Community" 2006 Press Releases. 7 July 2003. <<http://www.realnetworks.com/company/press/releases/2003/smilsources.html>> 3. W3C Synchronized Multimedia Home page. <<http://www.w3.org/AudioVideo>> 4. Synchronized Multimedia Integration Language. <<http://www.w3.org/TR/SMIL3>> 5. Ambulant Open SMIL Player Demos <<http://www.cwi.nl/projects/Ambulant/Demos/News.zip>> 6. Digital Access Information SYstem (DAISY) Consortium. <<http://www.daisy.org>> 7. Open Mobile Alliance Release Programs and Specifications <[http://www.openmobilealliance.org/release\\_program/docs/MMS/V1\\_2-20030923-C/OMA-MMS-CONF-V1\\_2-20030929-C.pdf](http://www.openmobilealliance.org/release_program/docs/MMS/V1_2-20030923-C/OMA-MMS-CONF-V1_2-20030929-C.pdf)> 8. Freed, N. and N. Borenstein "RFC 2046: Multipurpose Internet Mail Extensions (MIME) Part Two: Media Types," November 1996, <<ftp://ftp.isi.edu/in-notes/rfc2046.txt>> 9. "Ogg Vorbis audio format" from Xiph.Org Foundation. A fully open, non-proprietary, patent-and-royalty-free, general-purpose compressed audio format. The Ogg Vorbis specification is available at <http://www.xiph.org/vorbis/doc/> 10. Randers-Pehrson, Glenn and Thomas Boutell. "Registration of new Media Type image/png," 27 July 1996. <<http://www.iana.org/>

In SMIL 3.0, a state mechanism is expected to be added to SMIL in which a collection of variables can be defined to further control presentation interaction. These variables can be used to implement counters, to store dynamic content or to interact with the content-control mechanism in SMIL to influence the selection of individual media objects.

**MEDIA IN SMIL.** Prior to SMIL 3.0, W3C SMIL profiles did not require particular media types or codecs. Unfortunately, this has led to interoperability problems in that compliant SMIL players can play the same SMIL document, but they may not be able to play the same SMIL media. This means that a SMIL presentation authored for one player did not necessarily play for another player. This is perhaps one factor that has dampened SMIL's popularity.

SMIL 3.0 attempts to remedy this situation by including a common set of popular, royalty-free formats that are required for the Mobile, Extended Mobile, and Language Profiles. This means that a presentation authored in SMIL using these supported media types is guaranteed to be playable on any compliant player. Additionally, it enhances interoperability between content intended for both desktop and mobile devices. Future set-top box profiles should require the same content set to continue this interoperability and leverage one of SMIL's key advantages for cross-device interoperability.

The required formats from the SMIL Mobile, Extended Mobile, and Language Profiles are defined as follows:

- audio/basic [8]
- Ogg Vorbis [9]
- image/png [10,11]
- image/jpeg [8, 12]

#### COMMERCIAL EXAMPLES OF SMIL

**ACCESS/KDDI deployment.** One example in which SMIL is successfully used in a commercial service is KDDI's EZ Channel Plus Service [13]. It is based on ACCESS's Netfront SMIL player. EZ Channel was first released in a suite of new applications as part of KDDI's 3G Service, based on CDMA2000 1x EVDO technology. Third Generation Wireless (3G) provides high bandwidth (2.4 Mbps maximum) data communications. EZ Channel, based on SMIL, is intended to provide added value to the user while taking advantage of the new high bandwidth availability.

EZ Channel downloads content packages of SMIL-based multimedia content that is played for the user like an interactive television show. The content is downloaded during off-peak hours, stored on the handset, and is then played on demand for the user. Today, there are many channels available showcasing a variety of content. The KDDI EZ Channel service has been perhaps one of the most successful services based on SMIL to date.

**ENHANCEMENTS FOR SMIL 3.0.** SMIL 3.0 includes several significant enhancements from SMIL 2.1. It includes several new modules—SMILtext, State, External Timing, and DOM—as well as revisions to many existing modules.

The bulk of SMIL structure and syntax remains unchanged from SMIL 2.1, while several new additions add functionality often requested from previous versions or were implemented as private extensions by individual players (such as the RealPlayer from RealNetworks).

**SMILtext.** SMILtext is a media type defined for use within SMIL. Prior to SMIL 2.1, there was no convenient way to include inline text within a SMIL presentation. The text needed to be referenced from an external file. Thus, a new file often had to be created to include even a single sentence of text.

SMILtext defines an implicit SMIL media type to allow the inclusion of simple text. Common types of text formatting and decoration, such as the ability to specify colors, fonts, and text styles, are supported. SMILtext is meant to be lightweight. If a heavier solution is needed, then "Distribution Format Exchange Profile" (DFXP) [14] is recommended.

The addition of SMILtext is extremely useful to include simple inline text without the overhead of creating a new external media object.

**DOM.** For the first time SMIL 3.0 explicitly defines a DOM to allow interaction with scripting languages such as EcmaScript. The DOM functionality is divided into primarily two components. The first has the ability to start and stop a presentation during playback using scripting. The second has the ability to dynamically change attributes. This ability is similar to the ability to change attributes using animation. The key differentiator is that animation is described in a declarative manner, while the ability to modify attributes using the DOM is achieved in a procedural manner. An effort has been made to completely align the DOM model in SMIL with

the model earlier provided for SMIL Animation within SVG.

**External Timing, or Timesheets.** Timesheets are an exciting new addition to SMIL 3.0. The goal of timesheets is to allow the inclusion of SMIL information into another XML-based markup language such as HTML. In this way, timing information is added to a presentation format that does not include timing information. The original document can still be played using a legacy host player since the addition of SMIL external timing information does not modify the host-language syntax. Of course, in this case the timing information is lost, but the presentation is still playable.

External timing uses an external file to specify timing information. In this way, it is syntactically similar to CSS [15]. The functionality provided in the ability to append time-based information to an existing XML document can be considered analogous to the ability to append style information using CSS. Similarly, a single timesheet document, like a style sheet, can be defined once and reused in multiple documents to provide a common temporal framework for a multimedia presentation.

**State.** SMIL 3.0 adds the ability to test state variables from within SMIL. The intent of this is to allow more explicit control and visibility for the state of the SMIL presentation without requiring the use of DOM and an external scripting language. Applications for state could include quizzes and computer-aided instruction or interactive adaptation of presentations to user preferences.

**SMIL 3.0 PROFILES.** SMIL is designed to be modular. There are 63 modules grouped into 13 functional areas. Individual modules of SMIL are grouped to create a profile. A profile is designed to fit a particular application's needs in terms of balancing functionality and complexity. When a given player supports a profile, then it is guaranteed that any content authored for that profile will play in that player.

SMIL 3.0 adds two new profiles, the Daisy Profile and Tiny Profile, and keeps four profiles—Mobile, Extended Mobile, Scalability Framework, and Language—from SMIL 2.1.

**Mobile Profiles.** The SMIL 3.0 Mobile Profile and Extended Mobile Profile are mostly unchanged from their introduction for SMIL 2.1. A key motivator for the release of SMIL 2.1 was to define

profiles to meet the needs of the mobile industry for cellular handsets such as used in the KDDI SMIL service described earlier. The Mobile Profile is based on 3gpp's SMIL profile, while the Extended Mobile Profile is based on 3gpp2's SMIL profile. In SMIL 3.0, required content types have been added to improve interoperability.

**Tiny Profile.** The Tiny Profile replaces SMIL Basic from earlier versions of SMIL. This is meant to be the lightest feature set that can be implemented while still supporting the core elements of SMIL. It is designed for low-complexity devices such as digital cameras or music players. It can also be used to support playlist functionality as would be used from a media server.

**Daisy Profile.** The Daisy Consortium defines a format for digital talking books for the print impaired, including those with blindness, low vision, and dyslexia. Daisy is a recognized worldwide standard for talking books and has been involved with SMIL for some time. SMIL 3.0 includes a formal Daisy profile for the first time. Profiles such as the Mobile Profile are too heavy, since Daisy does not have needs for visual aspects such as transitions and multiarc timing, and SMIL Tiny is too lightweight for Daisy.

**USING SMIL.** An example better helps illustrate SMIL and SMIL's benefits to the designer. Along with the Ambulant open source SMIL player, a number of examples are provided for download [16]. In this section we discuss the Flashlight example. It is difficult to appreciate all the aspects of a multimedia presentation in print format, so the reader is encouraged to download and play this example on a SMIL player.

The Flashlight example illustrates several key points about SMIL. First, it shows how to structure a SMIL presentation for basic navigation. If the user does nothing, then the presentation plays automatically from start to finish in sequence. However, the user has the option of clicking on any of the major topic buttons, and that part of the presentation begins playing immediately.

Second, this presentation shows how two SMIL presentations can be created referring to the same content library in order to develop presentations targeted for different devices. In this case, there is a presentation for a desktop player and a presentation for a handheld player. The handheld presenta-

assignments/media-types/image/png.> **11.** "Portable Network Graphics (PNG) Specification (Second Edition)" Information technology—Computer graphics and image processing—Portable Network Graphics (PNG): Functional specification. ISO/IEC 15948:2003 (E). Thomas Boutell (Ed.), World Wide Web Consortium 01 October 1996. This W3C Recommendation is available at <http://www.w3.org/TR/REC-png>. **12.** "JPEG File Interchange Format, Version 1.02"; Eric Hamilton, World Wide Web Consortium September 1, 1992. This specification is available at <http://www.w3.org/Graphics/JPEG/jfif.txt>. **13.** "NetFront(tm) SMIL Player Adopted as Multimedia Player for KDDI's New 'EZ Channel Plus' Service" ACCESS Press Release. 21 September 2006. <[http://www.access-company.com/news/press/PalmSource/2006/092106\\_kddi\\_smil.html](http://www.access-company.com/news/press/PalmSource/2006/092106_kddi_smil.html)> **14.** Timed Text (TT) Authoring Format 1.0—Distribution Format Exchange Profile (DFXP) <<http://www.w3.org/TR/taf1-dfxp/>> **15.** Cascading Style Sheets Home Page <<http://www.w3.org/Style/CSS/>> **16.** Ambulant Demos <<http://www.cwi.nl/projects/Ambulant/Demos.html>>

tion is targeted for a low-bandwidth device and uses still pictures with an audio voiceover, while the desktop presentation uses video. The handheld presentation also makes more use of animation, which tends to be more efficient in bandwidth usage. The layouts are also adjusted to accommodate different-size displays. Figure 2 compares the two presentations.

Last, the presentation takes advantage of the `systemTest` element, allowing a single presentation to be dynamically modified based on application requirements. In this case, audio voiceovers are supplied using both British and American English. The attribute `systemLanguage` is used to globally select, at runtime, the proper audio files based on intended audience.

The code for this is shown here:

```
<switch>
<audio id="InBattIn-US-0" region="content-
text" src="LITE\dataCE/InBattIn-US.mp3"
dur="8.8s" systemLanguage="en-us"/>
<audio id="InBattIn-EU-0" region="content-
text" src="LITE\dataCE/InBattIn-EU.mp3"
dur="2.1s"/>
</switch>
```

**SUMMARY.** Both open source SMIL players as well as many SMIL content examples are available for downloading from the Web. The next time you need to create a multimedia presentation, try the open source approach. SMIL may be the technol-

ogy you are looking for; it is already well known for its role in SVG and MMS. SMIL 3.0 will add missing features and functionalities to further broaden SMIL's appeal.



**ABOUT THE AUTHORS** *Dr. Daniel Zucker is an independent consultant specializing in the mobile Web. He was co-chair for the W3C SMIL Workgroup, and most recently was senior director of technology and FAE for ACCESS, a leading supplier of non-PC Web browsers. Prior to ACCESS, Zucker was VP of engineering for Mobilearia and CTO for ePocrates.*



*Dr. Dick Bulterman is a senior researcher at CWI in Amsterdam, where he has headed the distributed multimedia languages and interfaces theme since 2004. Prior to joining CWI, he was on the faculty of the division of engineering at Brown University, where he was part of the Laboratory for Engineering Man/Machine Systems. Other academic appointments include visiting professorships in computer science at Brown (1993-94) and in the information theory group at TU Delft (1985), and a part-time appointment in computer science at the University of Utrecht (1989-1991). Dr. Bulterman received a Ph.D. in computer science from Brown University in 1982. He is on the editorial board of the ACM/Springer Multimedia Systems Journal and Multimedia Tools and Applications. He is a member of Sigma Xi, ACM, and IEEE.*



Figure 2: Presentations with Mobile Layout (left) and Desktop Layout (right) referencing the same content library.