Evaluating SMIL: Three User Case Studies

Lloyd Rutledge, Lynda Hardman and Jacco van Ossenbruggen
CWI (Centrum voor Wiskunde en Informatica)
P.O. Box 94079
NL-1090 GB Amsterdam
The Netherlands
+31 20 592 41 27
{Lloyd.Rutledge, Lynda.Hardman, Jacco.van.Ossenbruggen}@cwi.nl

1. ABSTRACT

This paper presents three user case studies of the multimedia standard SMIL. Each conveys a different kind of typical multimedia. The studies illustrate how SMIL can be used for these forms of multimedia. Analysis of these studies also show potential areas for extension of the language to better suit the needs of Web-based multimedia.

1.1 Keywords

SMIL, hypermedia, multimedia, World Wide Web.

2. INTRODUCTION

SMIL (Synchronized Multimedia Integration Language, pronounced "SMIL") is the W3C format for multimedia on the Web [4]. Its HTML-like syntax encodes the screen layout, interaction, adaptivity and timing of multimedia presentations. With its W3C status and at least three free browsers [3][6][7] currently available, SMIL is gaining an increasingly larger presence on the Web. This paper provides insight into the utility of SMIL in practice by presenting cases of the application of SMIL for multimedia on the Web that the authors have been involved with. These multimedia applications fall in three different areas: infotainment, accessibility, and conceptual multimedia art. These case studies also provide possibilities for independent extensions to SMIL or for future versions of SMIL itself.

3. INFOTAINMENT MULTIMEDIA

Perhaps the most typical type of multimedia presentations is "infotainment", in which the presented information is made more entertaining and engaging through increased use of audio and video media and more interaction with the user. One example of infotainment multimedia is *Fiets* (Dutch for "bicycle", pronounced "feets"), a collection of SMIL presentations about Amsterdam [5][9] (see Figure 1). Fiets consists of 9 different presentations on Amsterdam. Each conveys either the spatial, temporal, or relational information inherent to Amsterdam itself by using either the spatial, temporal or navigational aspects of the multimedia presentation [9]. Experience with Fiets as infotainment provides insight into how SMIL encodes a primary genre of multimedia, while focussing on the three main constructs of hypermedia: screen display, timeline and navigational links. This section presents Fiets' encoding in SMIL of these three areas and how the coding for each could be facilitated with improvements to SMIL itself or its use with other Web formats. Some of these changes have been proposed in earlier work [8], but not in the context of Fiets.

One measurement of SMIL is its ability to provide the basic components for multimedia upon which the bulk of typical multimedia presentations can be made. For Fiets, SMIL did provide the basic components needed for defining the spatial layout, timeline and hyperlinking desired. However, the means needed to encode some behaviors with SMIL was sometimes bulky and inefficient.

The SMIL `region` element and its attributes for two-dimensional placement and sizing provide the basics for screen placement of visual media objects. While this was able to represent all the placement in Fiets, the coding would have been more efficient if a mechanism for centering visual media in their assigned regions was available to SMIL

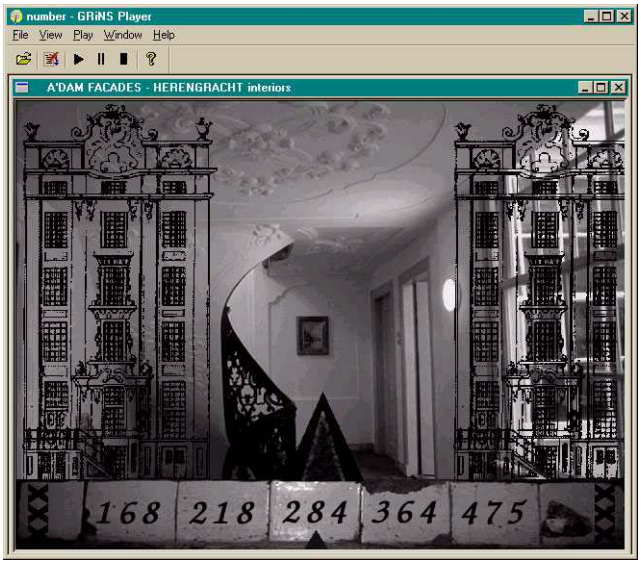

Figure 1. Screen Display from *Fiets*

processing. Without it, any centering would have to be explicitly calculated for each media object of different size. Grouping regions within other regions and relative positioning of regions would also facilitate positioning calculation by the author and make the code more efficient. The means of implementing these improvements could be put in a future SMIL version.

SMIL's hierarchical temporal composition and pair-wise synchronization successfully encoded all the timing desired in Fiets. However, the requirement in SMIL that synchronization be only between siblings in the temporal hierarchy necessitated either cumbersome rearranging of the hierarchy or chaining multiple synchronizations together. Allowing broader synchronization in SMIL would remove the need for these inefficiencies.

Finally, the SMIL-encoded linking in Fiets would have required less repeated code if a constructs such as the *choice node*, discussed in earlier work [2][8], were introduced into SMIL. This would enable behavior similar to that provided by frames in HTML: having part of the presentation stay the same while the rest changes. This is only possible in SMIL by repeating the code for the static part of the presentation.

4. ACCESSIBLE MULTIMEDIA

While infotainment multimedia makes the presentation of information more engaging, accessible multimedia serves to adapt the presentation of information for users who may otherwise not be able to perceive it. Sight- and hearing-impaired may need additional visual or audio descriptions of information they cannot perceive. Media also needs to be adapted for users under special circumstances, such as driving a car, or users with certain systems, such as portable, hand-held devices.

The "Physics Interactive Video Tutor" (PIVOT) is a Web-based multimedia physics curriculum (see Figure 2). It is being developed by The CPB/WGBH National Center for Accessible Media (NCAM) [1] and the Center for Advanced Educational Services at the Massachusetts Institute of Technology (MIT). PIVOT is built around MIT Professor Walter Lewin's class on Classical Newtonian Mechanics.

A primary goal of PIVOT is to have it be accessible to students who are deaf or blind. The source presentations, without any accessibility features used typically have one audio and one video, where the video shifts from face-and-gesticulating-hand shots of the professor to diagrams being drawn by the professor. There are three basic accessibility components for PIVOT that can be added to these source presentations: closed captions, tucked audio descriptions, and pausing audio descriptions. For each of these three, the use of SMIL for them, and possibilities of SMIL extensions for them, are discussed below.

4.1 Closed Captions

If closed captions are to be used, then pieces of text must be shown in a portion of the display, and their display must be timed with the audio. In PIVOT, closed captioning is represented with a SMIL `switch` element containing a `text` element with the captioned text. The `text` element would have its `system-captions` attribute set to `"on"`. This would cause a browser to recognize that `text` element as appropriate to select for playing, as a child of the `switch` element, only if captions are desired for the presentation. The `text` element would assign a screen display region on which to display the captions. All of this behavior can be encoded in SMIL.

Desired behavior for PIVOT that cannot be encoded in SMIL is the altering of the layout if closed captions are used. This would enable screen display to be rearranged to make room for the captions in a graceful way. A possible extension to SMIL that would enable this behavior would be to allow `layout` elements to have test attributes like `system-captions` so that alternative layouts could be specified and at runtime selected based on the browser's setting for the use of captions.

4.2 Tucked Audio Descriptions

Some presentations require audio descriptions, which describe visual events in the video. One technique is to "tuck" these in natural gaps in the original audio. This maintains the original timing of the original audio and video. This is important if the presentation is a scheduled broadcast to be adapted on an individual basis to each member of the audience watching it simultaneously.

The SMIL encoding for this is similar to the encoding for closed captions. A `switch` element contains a single `audio` element contain a clip of audio description. If this audio clip is considered appropriate for playing, then it is played —otherwise, it is not. SMIL-defined synchronization between the audio clips and the original audio and video put

Figure 2. Screen Display from *PIVOT*

the descriptive audio in the natural gaps in the original audio track. A useful addition to SMIL for specifically defining this behavior would be a `system-audio-desc` test attribute, which works for audio descriptions in the same manner that the `system-captions` attribute works for captions.

4.3 Pausing Audio Descriptions

Instead of tucking, audio descriptions can be played during pauses imposed on the original audio and video. The advantage of pausing the original video and audio is that more elaborate and informative audio descriptions can be used. SMIL can define this behavior by cutting the audio and video into clips that begin and end with the pausing points. The clips would be played sequentially if there is no audio description. The video clip would have its `fill` attribute set to `"freeze"`, and the audio descriptions would push the timing of the `video` elements to be longer than the clips themselves. This causes the pausing behavior when the audio descriptions are played and thus changes the timeline.

Currently, no SMIL browser performs this behavior without error or without visible distortion in the video progress with audio description turned off. Improving the performance of this behavior is necessary to make pausing audio descriptions effective. A possible extension to SMIL that may encode this behavior better is a "pause" command that can work at multiple places on a single clip of video. This might enable browsers to more seamlessly play the video when audio descriptions are not used.

5. CONCEPTUAL MULTIMEDIA ART

Infotainment and accessible multimedia each typically uses the same basic model and behavior for most of their presentations. On the other hand, artists that use multimedia to express certain concepts often make presentations that have models of interaction and display that vary widely. Since it is impossible to make a model that applies to all possible conceptual multimedia art, artists usually must struggle to find a mapping between their concept and the multimedia presentation models that exist, often with the need to make compromises on the original vision.

An example conceptual multimedia comes from the artwork *Off the Wall*[1], by Margret Wibmer and Günther Zechberger. *Off the Wall* involves multiple photographs from different perspectives of one object: a person entirely encased in a loose-fitting industrial-looking yellow rubber outfit [10] (see Figure 3). A CD-ROM is being made for *Off the Wall* containing different QuickTime VRs of this object. With the object being represented in QuickTime VR, the user can spin the object around in all different orientations, and zoom in and out: like a 3-D digital image. As a companion to the project and its QuickTime VR CD-ROM, a SMIL presentation is being made for installation on the Web for emulating the behavior of the QuickTime VR [10].

5.1 Encoding the Manipulating of a Object

An important aspect of the concept being conveyed in *Off the Wall* is that of an object being manipulated by the user. The QuickTime VR provides the user with the interface to move the object around in a virtual space. The primary design for representing this type of interaction in SMIL is to replace the QuickTime VR objects with "linear" videos and integrate them in SMIL. Each video would represent the object being moved in one particular way.

For example, there could be a separate video for the object being rotated along each of the three axes of rotation. For each rotation, there could be a separate video for each different distance the view could be from the object as it rotates. There could also be video clips of the object being moved toward/away from the user, with each clip having a particular orientation of the object along the three axes. Different videos could display varying speeds of each movement. Of course, there are infinite possibilities for such movement, and each video can only capture one. For an effective presentation with these videos, they should represent a broad sample of the types of motion that the artist desires to convey.

Links can be established for specific points in the movement that trigger other video clips moving the object from that position and orientation to another. The video clips loaded as a result of these triggers can be portions of video files. In SMIL, these would be defined with begin clip attributes. The reason for using clips is to have the object in the same position and orientation for starting the next movement as it was when ending the previous.

With QuickTime VR, you have virtually infinite possibilities for interaction with the grabbing of the object to rotate it and the motion of the mouse to move it in and away. With the multimedia presentation described here, all

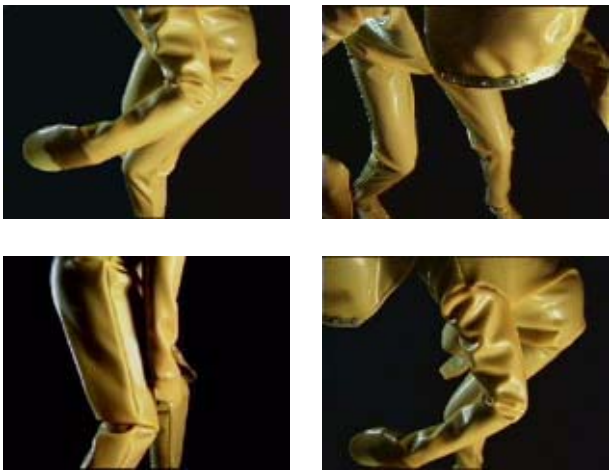

Figure 3. Screen Displays from *Off the Wall*, by Margret Wibmer and Günther Zechberger

[1] An exhibition of this work premiered in July 1999 at the Museum Ferdinandeum in Innsbruck, Austria.

movement is along the linear paths of rotation and motion established by the videos. Also, the number of potential interactions is finitely set with the discrete number of hotspots for the links.

5.2 Encoding a Dynamic Object

This SMIL-based approach can be used to convey the object changing form. Described above is a collection of videos conveying the motion of one static object. An equivalent (in terms of motion along the axes) collection can be made for the object in a different shape. Transition videos showing the object change from one shape another can also be used. The transition videos could also convey movement during the transformation. SMIL links can trigger shape changes in the same manner as they convey movement of the object.

This requires having anchors within individual media objects that effect the entire integrated presentation. Because QuickTime VR is multi-dimensional, it is harder to have portions of its display trigger events. With SMIL and the "linear" video approached described, a period of time can much more easily be made to lead to the perceived change of the object.

5.3 Reducing Storage Space And Bandwidth

A concern that arises is the space required. More videos mean more freedom of control. Increasing the amount of video also greatly increases the space required. QuickTime VR has an implicit compression over this multi-video model in that one collection of 3D pixels can be used for all movement at all speeds of a static object. With multiple videos, there is much repeated visual information stored.

A further compromise to save space would be to use all images and no videos. Each video would be replaced with a sequence of images, set up as a sequence in SMIL code. The exact same user interaction model would apply, with the same buttons and the same number of state changes. There would be much reuse of images. For example, conveying the same movement at different speeds could use the exact same images, whereas in the video model different videos would have to be used. One image could also be a common intersection between movements along different axes. The disadvantage would be the visual chunkiness of the progression. This could be lessened with more images shown at shorter durations, with a cost in space and possibly also in processing-incurred delays.

6. SUMMARY AND CONCLUSION

This paper presents three user cases studies of the application of SMIL in different areas. The case study of SMIL for infotainment provided insight into one of the most common types of multimedia. The discussion of the PIVOT project described the use of SMIL for making information on the Web more accessible and adaptive for a wider variety of users. The case study of *Off the Wall* provided an example of SMIL's use in conveying particular artistic concepts. These analyses not only help in determining how to use SMIL today, but also how to guide the development of future versions of SMIL and further integration of it with other Web standards [8].

7. ACKNOWLEDGMENTS

The Fiets artwork and graphic design was made by Maja Kuzmanovic. The creation of Fiets was funded by the European Union ESPRIT Chameleon project. Funding for the PIVOT project comes from Mitsubishi Electric America Foundation (MEAF).

8. REFERENCES

[1] Freed, G. *NCAM Web Access Project*, URL: http://www.wgbh.org/wgbh/pages/ncam/webaccess/about-project.html.

[2] Hardman, L., Bulterman, D.C.A. and van Rossum, G. The Amsterdam Hypermedia Model: Adding Time and Context to the Dexter Model. Commun. ACM 37, 2, 50-62.

[3] Helio, *SOJA "Cherbourg 2"*, URL: http://www.helio.org/products/smil/.

[4] Hoschka, P. (ed.). *Synchronized Multimedia Integration Language*, World Wide Web Consortium Recommendation. June 1998. URL: http://www.w3.org/TR/REC-smil.

[5] Kuzmanovic, M., and Rutledge, L. *Fiets: A Tour of Historic Amsterdam Buildings*, URL: http://www.cwi.nl/SMIL/fiets.html.

[6] Oratrix Development BV. *GRiNS*, URL: http://www.oratrix.com/GRiNS/.

[7] RealNetworks, Inc. *RealPlayer G2*, URL: http://www.real.com/products/player/.

[8] Rutledge, L., van Ossenbruggen, J., Hardman, L. and Bulterman D.C.A. Anticipating SMIL 2.0: The Developing Cooperative Infrastructure for Multimedia on the Web, in *Proc. The Eighth International World Wide Web Conference (WWW8)*, (Toronto, Canada, May 1999).

[9] Rutledge, L. Hardman, L., van Ossenbruggen, J. and Bulterman, D.C.A. Structural Distinctions Between Hypermedia Storage and Presentation, in *Proc. ACM Multimedia 98*, (Bristol, England, September 1998), 145-150.

[10] Wibmer, M. and Zechberger, G. *Off the Wall*, URL: http://www.off-the-wall.at/.