THE USE OF SMIL: MULTIMEDIA RESEARCH CURRENTLY APPLIED ON A GLOBAL SCALE

LLOYD RUTLEDGE, LYNDA HARDMAN AND JACCO VAN OSSENBRUGGEN
CWI (Centrum voor Wiskunde en Informatica), P.O. Box 94079, NL-1090 GB Amsterdam, The Netherlands
E-Mail: {Lloyd.Rutledge, Lynda.Hardman, Jacco.van.Ossenbruggen}@cwi.nl

This paper describes the current use of the multimedia standard SMIL. SMIL features that relate to active areas of multimedia research are discussed. SMIL current implementation in existing browsers is described. Examples from the Web of SMIL applications representing different types of multimedia are presented. These discussions together provide an overview of how SMIL currently addresses the needs of multimedia distributed on the Web.

1. Introduction

SMIL (Synchronized Multimedia Integration Language, pronounced "SMIL") is the W3C format for multimedia on the Web [9]. Its HTML-like syntax encodes the screen layout, interaction, adaptivity and timing of multimedia presentations. With its W3C status and several free browsers currently available, SMIL is gaining a substantial presence on the Web. Various concepts from multimedia research have a presence in the SMIL format. This gives SMIL the potential to apply multimedia research on a global scale — putting multimedia into common use, just as HTML put hypertext into common use.

This paper discusses SMIL as a result of multimedia research. First, SMIL is discussed in terms of ongoing multimedia research, showing the relationship between this research and its existing, and potential, impact on large-scale multimedia use. Then, existing software for processing SMIL is described, specifying what SMIL is capable of today. Finally, some current applications of SMIL on the Web are presented, showing the manner with which SMIL is — and can be — used.

2. Multimedia Research Applied to SMIL

SMIL, as a format for multimedia, also serves as a model for multimedia. Its constructs define a distinct set of multimedia concepts. The structure within which these constructs fit in is one model for how these concepts can be composed into multimedia presentations. These concepts and this model are related with concepts and models that come from multimedia research. This shows the impact research has had on the SMIL standard, and it demonstrates the potential impact ongoing research can still have on SMIL and Web-based multimedia. Some research has started as a result of the SMIL standard, such as the adaptation of existing models and systems to incorporate SMIL, and the comparison of the format with other models to suggest extensions to it. This section presents some examples of multimedia research that relates to the SMIL format and its implementation.

2.1 Top-level Document Structuring

SMIL is an XML-defined format. Its XML structure is time-based — the primary semantic impact of the syntactic composition and progression the XML code is on the timing of the document. This contrasts with typical XML documents and HTML documents. Their structure is textflow-based, with its primary semantic impact being on how far along the wrapped-around single flow of text a part of the document appears. In other words, a component's position in SMIL structure says *when* it appears, and a component's position in HTML and typical XML structure says *where* it appears. This distinction and its ramifications have been investigated in recent research [7]. These ramifications include how temporal information can be applied to mainly textual documents, and how text-flow information can be applied to mainly temporal documents. An example of where these issues applies with HTML+Time, which uses SMIL-based temporal constructs in HTML documents. With HTML+Time, the XML structure can be either textflow-based or time-based or a mixture of both.

2.2 User Adaptation

There is much research on adapting presentations to individual users. SMIL has constructs that provide some user-centered adaptation. Research with the CMIFed system has investigated means of encoding user-centered adaptation in SMIL and also in non-standardized extensions to SMIL [3].

The primary SMIL constructs for user-centered adaptation are the switch element and the test attributes. These specify that one or zero selections are to be made among a collections of alternatives for inclusion at individual locations in the temporal hierarchy of a SMIL document. These alternatives can be, for example, equivalent media objects in different formats. They can also be different subtrees of SMIL temporal composition and media objects.

CMIFed's approach is to put adaptation not in the temporal/media structure but on the devices for presented individual media objects. The CMIFed construct for such media devices is the *channel*. The CMIFed channel corresponds to the SMIL region element, except that the region applies only to visual media, while the channel can apply to any medium.

Channel-based adaptation consists of having channels be turned on or off. For example, in CMIFed, subtitling for the hearing impaired would consist of audio media objects throughout the temporal/media hierarchy that output to one subtitle channel. This subtitle channel would be turned on for the hearing impaired and turned off for other users. To encode this in SMIL, each piece of subtitling in the hierarchy would have to be have test attributes assigned to it an be put in a switch element. With channels, neither additional attributes nor the additional layers of the temporal/media hierarchy are needed. Since SMIL subtitles would be assigned a region, which is the equivalent of a channel, assigning subtitles in CMIFed a channel involves the use of no addition constructs over SMIL.

The use of channels also facilitates the enabling of different combinations of adaptation. For example, a presentation can adapt by providing subtitles and also by being in different languages. Both the main spoken soundtrack and the subtitles can be in these different languages. Adaptation can consist of providing just sound, just subtitles, sound and subtitles in the same language, or sound and subtitles in the different languages. With SMIL, each different combination would need to be explicitly accounted for in the temporal/media hierarchy, its switch elements, and the test attributes. With the use of CMIFed channels, the temporal/media hierarchy is simpler because it encodes little of the adaptive information. The hierarchy is structured as if both soundtracks and subtitles in all languages are played simultaneously. In CMIFed, the selection is made with the channels, not with the media objects and temporal hierarchy. The different combinations are specified by the turning on and off of different channels, where there is one channel for each language in sound and subtitles.

2.3 Authoring Multimedia

Much research has taken place regarding the authoring of multimedia [14,18,24]. Some of this has been applied to authoring SMIL [4,12,15]. The primary authoring issue that applies to SMIL is the author's specification of temporal constraints. The research authoring system CMIFed [18], which was later developed into the commercial authoring system GR*i*NS [4], used a timing model that is similar to that used in SMIL. When CMIFed was extended to output SMIL as well as its own format, not many changes were required.

Other systems have used a different model for temporal constraints than that which SMIL uses and have thus needed to perform a transformation between the models to output SMIL code. Mikado [12], for example, primarily generates multimedia presentations in the format Madeus [13], whose model is based on sets of temporal constraints, not parallel and sequential composites. Mikado has a component named SMILY [15] that generates a subset of SMIL. SMILY has an algorithm for translating the temporal constraints inherent in Mikado and Madeus into SMIL temporal composites along with SMIL delay and event synchronization attributes. This algorithm describes the relationship between the two timing models.

2.4 Dynamic Quality-of-Service

Recent research with the REMDOR architecture has explored means of extending SMIL to support more dynamic graceful degradation of quality-of-service [26].

These extensions would provide the author more power to specify what media to start downloading and when to adapt the presentation to varying bandwidth and processing constraints. REMDOR is an architecture using a transport level partial-order/partial-reliability model for selecting media components for multimedia integration. This model has been compared with the SMIL timing model to investigate how dynamic QoS control could be incorporated into SMIL processing and the SMIL syntax.

This research is producing suggestions for the extension of the SMIL standard to handle more dynamic control of graceful degradation. The strong and weak values of their proposed start attribute state for a SMIL par (parallel) element whether the elements the par states should be played in parallel must all start at the same time, thus waiting for all the elements' media objects to be loaded before any start playing, or if each element can start playing as soon at it's media objects are downloaded. The proposed reliability attribute states whether a media object in a presentation is essential or expendable. From this attribute, when it is time to play a media object that has not yet been downloaded, the browser can determine whether to wait for the download and then play the object before continuing, or to continue the presentation without playing the object. The proposal also includes adding new event types to SMIL synchronization, such as when a media object has been downloaded.

2.5 The Automatic Generation of SMIL

Research with the Berlage architecture has investigated the potential automatic generation of SMIL encoded multimedia presentations [22]. The presentations are generated from media archives and metadata describing that media. Berlage is structured in terms of the Standard Reference Model for Intelligent Multimedia Presentation Systems (SRM-IMMPSs), an established model for generating multimedia in general [2]. The SRM-IMMPSs accounts for having different multimedia output formats generated by the same system. This way, the same code that generates SMIL could also be used to generate other formats as well. Such a multi-format system would need a base multimedia presentation model from which presentations in the particular formats would be generated. A transform would have to occur from this model to an output format such as SMIL.

3. SMIL Implementations

The extent to which an idea can be exercised is the extent to which tools exist for applying it. Breaking down the idea into intellectually manageable components, developing models from these for expressing the idea in its various forms and then designing architectures for processing the idea are crucial first steps. Toward the end of the research life cycle comes the making of running tools. The nature of these tools can have a large impact on how the community in general applies an idea on a larger scale. While an idea or model may be influenced strongly by the research community, industry has naturally a stronger impact on final implementations and thus to how an idea or model is realized. The impact a particular implementor has on the general, long-term execution of an idea can be influenced by that implementor's other work, or by the nature of an existing implementation that was adapted for incorporating the new idea. This section describes the existing tools for processing SMIL, and how these tools shape, and have the potential to shape, the use of SMIL.

3.1 G2

RealPlayer G2 is a SMIL player made by RealNetworks [17]. RealPlayer G2 is free-of-charge and has a commercially available version called RealPlayer Plus G2. G2 is a player for the media formats developed by RealNetworks, such as RealAudio and RealVideo. These formats include much timing information, and G2 can play multiple formats simultaneously, making G2 effectively a player for synchronized multimedia. One G2 focus is the ability to play a lot of media efficiently without noticeable delays in presentation. Another focus has been to function with RealNetworks streaming servers.

With the release of SMIL last year, G2 was extended to play SMIL presentations as well as Real media files. With tens of millions of G2 players having been downloaded since, there is an enormous amount of SMIL browsers hooked up with the Web. While not having implemented all of the SMIL constructs, the SMIL-related features of G2 are regularly enhanced and extended. G2 has recently been made to play many more media formats and to function as a plug-in within HTML players. The impact RealNetworks and G2 are having on SMIL is in terms of well-processed timing and of the integration and large-scale distribution of high-volume media in multimedia presentations.

The community authors that write for the G2 player sometimes produce SMIL code that relies on external media file for specifying timing rather than putting the synchronization data in the SMIL file itself. Before SMIL was developed, G2 was a player primarily for specialized mono-media formats. The formats that RealNetworks developed and that G2 played were typically mono-medium with timing information for that medium embedded in the format. The arrival and implementation of SMIL has enabled G2 presentations to have a more intricate spatial layout by encoding it in SMIL. However, authors that have been familiar primarily with the G2 player often create SMIL presentations with much of the timing information not in SMIL but in the integrated media objects encoded with RealNetworks media formats, even though G2 properly plays the equivalent SMIL code. This results in SMIL presentations with timing information spread between multiple files in multiple formats, rather than in one, integrating file. It also results in presentations that are difficult to adapt for players other than G2.

3.2 GRiNS

GR*i*NS (GRaphical *i*Nterface to SMIL) is an authoring and player system developed by CWI and Oratrix Development [16]. The GR*i*NS player is available free-of-charge and has had several thousand downloads. Compared to other players, GR*i*NS has a focus on the SMIL standard itself. It has implemented almost all of the SMIL constructs, more than the other players. But it does not have some of the browser environment integration features of other players, such as the ability to operate as a plug-in.

The GR*i*NS player was made as a companion to the GR*i*NS authoring environment (see Figure 1). This provides a tool for large-scale authoring of SMIL presentations. Graphical interfaces are provided for direct manipulation of SMIL temporal compositional, fine-tuned synchronization, spatial layout and hyperlinking.

GR*i*NS was originally the research multimedia authoring system CMIFed [18]. Multimedia presentations created with CMIFed were stored in the format CMIF. When SMIL was developed, CMIFed was adapted to output SMIL as well as CMIF. CMIF encodes some behavior that SMIL does not encode. Such behavior can be edited with GR*i*NS. When it is, this behavior is encoded in SMIL as an extension of SMIL. The SMIL specifications specifies some techniques and constructs for output such extensions. GR*i*NS uses these constructs and techniques in its SMIL output of non-SMIL behavior. Still, even with the proper use of extension constructs, GR*i*NS may make presentations that do not play well on other browsers.

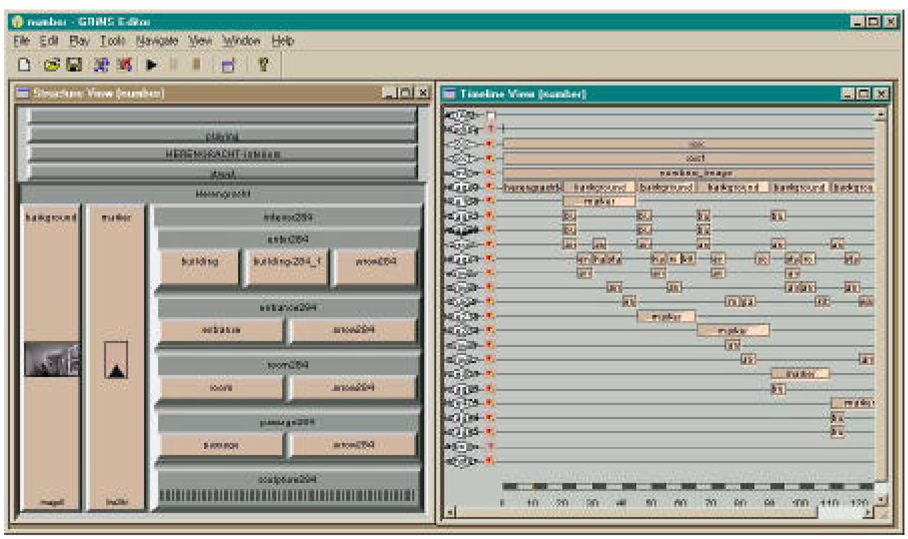

Figure 1. Screen Display from the GR*i*NS Authoring System

3.3 SOJA

The Helio organization has developed a SMIL browser called SOJA [8]. It is available free-of-charge and has had several thousand downloads. As a Java-based player, it integrates well within HTML presentations on HTML browsers. A typical presentation of a SMIL document on SOJA involves downloading not just the SMIL and media files but also the SOJA applet itself. The applet itself takes up relatively small system space. In relation to the other players, SOJA's focus is on having SMIL integrate well with the general Web browsing environment in terms of user interface, data transfer and software processing. Since SOJA was created only for SMIL presentations, and since work on it began only after SMIL became a standard, its use is less likely to develop SMIL dialects than the other players.

4. Some SMIL Applications

Once tools are made for expressing and processing an research result, a community can use them on a large enough scale to develop conventions for their use and for the application of the research result. These conventions consist of established patterns of how features in a tool, or constructs in a format, are applied. Conventions then effectively become tools and languages themselves.

This section presents some examples of the application of SMIL. These examples of the early use of SMIL provide insight into how the standard can be used, and into what conventions for its use may form. The multimedia applications discussed in this section fall in three different areas: infotainment, accessibility, and conceptual multimedia art. These analyses not only describe how multimedia research is applied to SMIL as it exists today, but also how in can continue to contribute the development of future versions of SMIL and further integration of it with other Web standards [20].

4.1 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 [11,23] (see Figure 2). 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 [23]. Experience with Fiets as infotainment provides insight into how SMIL encodes a primary genre of multimedia, while focusing on the three main constructs of multimedia: 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 [20], 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 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

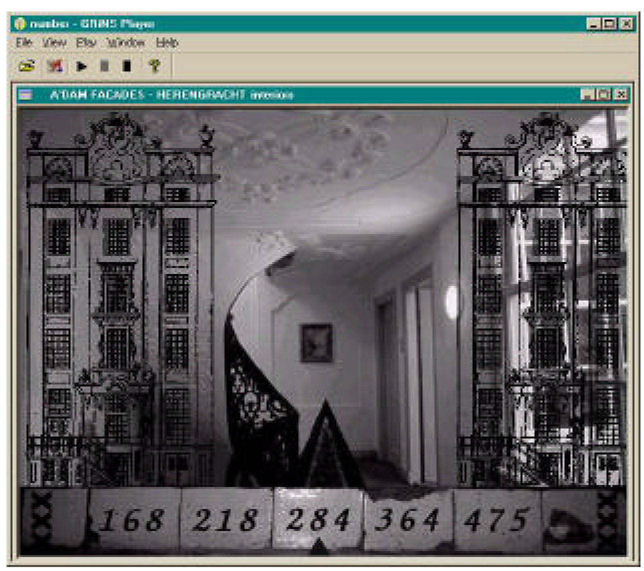

Figure 2. Screen Display from *Fiets*

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 [6,20], 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.

4.2 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 3). It is being developed by The CPB/WGBH

Figure 3. Screen Display from *PIVOT*

National Center for Accessible Media (NCAM) [5] 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.

4.2.1 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.

4.2.2 Tucked Audio Descriptions

Some presentations require audio descriptions, which describe visual events in the video. For example, when a video shows a ball bouncing as part of a physics experiment, visually impaired users will not perceive it. An audio description such as "the ball bounces on the table" would provide these users with information essential to the presentation that they would otherwise miss. Such audio descriptions would be in addition to the original audio of the presentation, which both sighted and visually impaired users would hear.

One technique for putting audio descriptions in a presentation is to "tuck" them into 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 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.

4.2.3 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.

4.3 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.

One example of conceptual multimedia is *GoTo0*, an interactive story by Maja Kuzmanovic [10] (see Figure 4). It's SMIL-encoding is a multimedia storyboard for a planned virtual reality presentation. In the final virtual reality, the user's interface with the story will be more physically immersive. The multimedia version emulates the virtual reality with a point-and-click interface with which the user navigates through the story. The challenge with writing the SMIL version what choosing what media objects (sound, images and video), screen display layouts, timing relationships and navigational interface will best convey the final virtual reality.

Another example of 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 an outfit made of loose-fitting yellow industrial material [22] (see Figure 5). 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 [22].

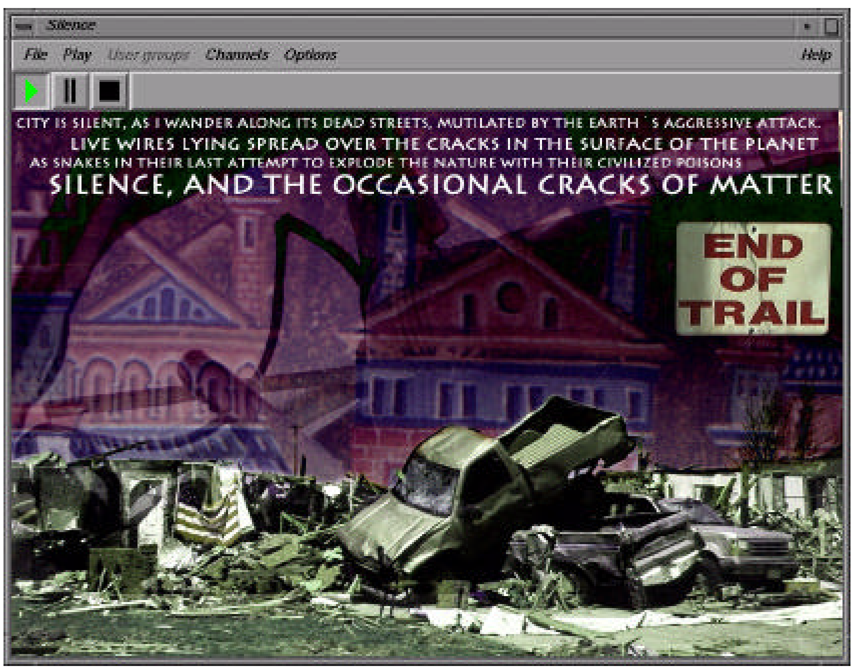

Figure 4. Screen Display from *GoTo0*, by Maja Kuzmanovic

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

4.3.1 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.

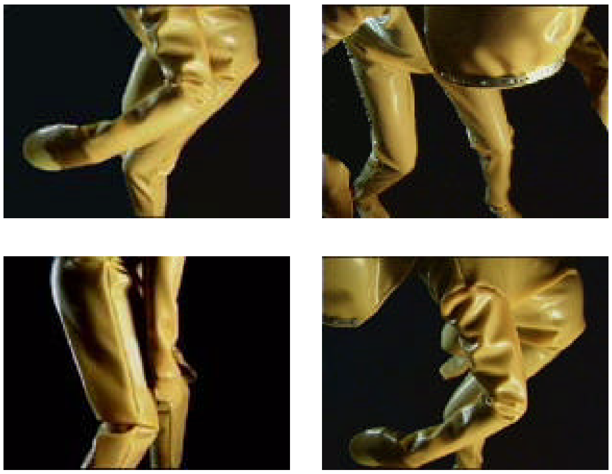

Figure 5. Screen Displays from *Off the Wall*, by Margret Wibmer and Günther Zechberger

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 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.

4.3.2 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, it is easier to encode that there is a perceived change of the object during a period of time.

4.3.3 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.

5. Summary

This paper presented SMIL as applied multimedia research. Examples of established and ongoing research were presented, along with their existing and potential impact on SMIL. Existing software was described, showing the tools that are available for demonstrating research concepts with SMIL. Examples of SMIL applications were given, illustrating some potential conventions for use of the standard, and illustrating how the standard can be used to demonstrate multimedia research ideas. Issues that arose include the need to transform some multimedia models into SMIL.

6. Acknowledgments

The Fiets artwork and graphic design was made by Maja Kuzmanovic. Many images used in Fiets come from the Amsterdam Heritage Website [1]. 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).

7. References

[1] City of Amsterdam Municipal Department for Preservation and Restoration of Historic Buildings and Sites. *Amsterdam Heritage*, http://www.amsterdam.nl/bmz/adam/adam_e.html.

[2] Bordegoni, M., Faconti, G., Feiner S., Maybury, M.T., Rist, T., Ruggieri, S., Trahanias, P. and Wilson, M. Standard Reference Model for Intelligent Multimedia Presentation Systems." *Computer Standards and Interfaces*, **18**(6,7) (December 1997), pp. 477-496.

[3] Bulterman, D.C.A. User-Centered Abstractions for Adaptive Hypermedia Presentations, *Proceedings of ACM Multimedia 98* (Bristol, England, September, 1998) pp. 247-256.

[4] Bulterman, D.C.A, Hardman, L., Jansen, J. Mullender, K.S. and Rutledge, L. GRiNS: A GRaphical INterface for Creating and Playing SMIL Documents, *Proceedings of the Seventh International World Wide Web Conference (WWW7)* (Melbourne, Australia, April 1998).

[5] Freed, G. *NCAM Web Access Project*, http://www.wgbh.org/wgbh/pages/ncam/webaccess/aboutproject.html.

[6] Hardman, L., Bulterman, D.C.A. and van Rossum, G. The Amsterdam Hypermedia Model: Adding Time and Context to the Dexter Model. *Communications of the ACM*, **37**, 2 (ACM, 1994) pp. 50-62.

[7] Hardman, L., van Ossenbruggen, J., Rutledge, L. and Bulterman, D.C.A. Do You Have the Time? Composition and Linking in Time-based Hypermedia, *Proceedings of ACM Hypertext 99* (Darmstadt, Germany, February, 1999) pp. 189-196.

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

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

[10] Kuzmanovic, M., *GoTo0*, http://www.cwi.nl/SMIL/GoTo0/.

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

[12] Jourdan, M., Layaïda, N. and. Roisin, C., A survey on authoring techniques for temporal scenarios of multimedia documents, *Handbook of Internet and Multimedia Systems and Applications, part 1: Tools and Standards* (CRC Press, April 1998).

[13] Jourdan, M., Layaïda, N., Roisin, C., Sabry-Ismail, L. and Tardif, L., "Madeus, an Authoring Environment for Interactive Multimedia Documents", *Proceedings of ACM Multimedia 98* (Bristol, England, September, 1998) pp. 267-272.

[14] Jourdan, M., Tardif, L. and Villard, L., Multiviews Interfaces for Multimedia Authoring, *Proceedings of Multimedia Modeling 98 (MMM98)* (Lausanne, Switzerland, October 1998) pp. 72-79.

[15] Jourdan, M., Tardif, L. and Villard, L., SMILY, a Smil Authoring environment, Technical Demonstration, *Proceedings of ACM Multimedia 99* (Orlando, Florida, USA, November 1999).

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

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

[18] van Rossum, G., Jansen, J., Mullender, M.S., and Bulterman, D.C.A. CMIFed: a presentation environment for portable hypermedia documents, *Proceedings of ACM Multimedia 93* (Anaheim, California, USA, August 1993) pp. 183-188.

[19] Rutledge, L. Hardman, L. and van Ossenbruggen, J. Evaluating SMIL: Three User Case Studies, *Proceedings of ACM Multimedia 99* (Orlando, Florida, USA, November 1999).

[20] 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).

[21] Rutledge, L. Hardman, L. and van Ossenbruggen, J. Evaluating SMIL: Three User Case Studies, Poster Paper, *Proceedings of ACM Multimedia 99* (Orlando, Florida, USA, November 1999)

[22] Rutledge, L., Hardman, L., van Ossenbruggen, J. and Bulterman, D.C.A. Implementing Adaptability in the Standard Reference Model for Intelligent Multimedia Presentation Systems, *Proceedings of Multimedia Modeling 98 (MMM98)* (Lausanne, Switzerland, October, 1998) pp. 12-20.

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

[24] Saade, D.C.M., Soeares, L.F.G, Costa, F.R. and Souza Filho G.L., Graphical Structured-Editing of Multimedia Documents with Temporal and Spatial Constraint, *Proceedings of Multimedia Modeling 97 (MMM97)* (Singapore, October 1997) pp. 279-295.

[25] Schmitz, P., Yu, J. and Santangeli, P. *Timed Interactive Multimedia Extensions for HTML (HTML+TIME): Extending SMIL into the Web Browser*, Note submitted to the Wide Web Consortium. September 1998. http://www.w3.org/TR/NOTE-HTMLplusTIME/. (note for discussion)

[26] Urbano, M. J., Mendes, A.J., Monteiro, E.H. and Amer, P.D. Specification of Order and Reliability in SMIL Documents, *Proceedings of the IEEE Workshop on Internet Applications (WIAPP'99)* (San Jose, California, USA, July 1999).

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