

GRiNS: A GRaphical INterface for Creating and Playing SMIL Documents.

Dick Bulterman¹, Lynda Hardman¹, Sjoerd Mullender¹, Lloyd Rutledge¹, Martin Prime²,
and Michael Wilson²

¹*CWI, Kruislaan 413, PO Box 94079, 1090 GB Amsterdam, Netherlands*

Email: dick.bulterman@cwi.nl

²*CLRC Rutherford Appleton Laboratory, Chilton, Oxon, OX11 0QX, United Kingdom*

Email: M.D.Wilson@rl.ac.uk

Abstract. W3C has developed a language for Web-Based Multimedia Presentation called SMIL: the Synchronized Multimedia Integration Language (pronounced 'smile'). SMIL provides a simple generic declarative language for synchronizing multimedia which could lead to a change in the appearance of WWW presentations away from looking like document pages towards the look of television programmes.

The Chameleon Esprit project has developed a browser to play documents written in SMIL in order to demonstrate the language and encourage its adoption, and an authoring environment GRiNS, that can be used to create SMIL-compliant documents. The GRiNS authoring environment allows the original media assets to be allocated to screen locations in a view of the document to be presented, and have their presentations synchronized on a time line view. A third view presents a hierarchy of the node structure of the multimedia document to promote the re-use of its components, while a fourth view presents the hyperlink structure within the document and pointing to the outside web.

1. Introduction

If you wish to publish multimedia presentations on CD-ROM and over the World Wide Web (WWW) there are a variety of solutions today, each of which has substantial limitations:

HTML documents cannot express the synchronisation primitives required. Video and audio files can be downloaded and played locally, although no synchronisation is supported outside a single file, the download time is usually unacceptable.

Dynamic HTML using Java and Cascading Style Sheets (CSS) requires programming language skills and still only supports the motion and resizing of objects on a page.

CD-ROM authoring packages provide the synchronisation primitives which may require programming skills in languages such as Lingo [1], but when the product is distributed over the WWW layout is limited to a single small object embedded in a page, and the synchronisation primitives, and linking do not operate over the rest of the published page.

Streamed video services provided by Microsoft NetShow [2], or Real Networks' Real Media [3] allow the product to be presented with internal synchronisation, and linking to other WWW objects from maps on the video stream, but the whole product must be transformed into a video stream using

considerable bandwidth, and synchronisation is still limited to the video presentation window.

A standard solution is required which overcomes these limitations in order to facilitate a change in the appearance of the web from resembling a magazine, to look more like television. This change may further facilitate the development of low cost digital television over the Internet, for use by educational institutions, small companies, and community groups.

The declarative multimedia format CMIF [4],[5], based on the Amsterdam Hypermedia Model [6], allows the control of interactions required for multimedia applications to be declaratively encoded in a text file as a structured set of object relations. A declarative specification is often easier to edit and maintain than a program-based specification, and it can potentially provide a greater degree of accessibility to the network infrastructure by reducing the amount of programming required for producing any particular presentation.

In early 1997 W3C established a working group on Synchronized MultiMedia (SYMM) to define a declarative multimedia format for the WWW. Based on CMIF, they have produced the Synchronized Multimedia Integration Language (SMIL - pronounced *smile*) as a W3C recommendation[7]. The SYMM group has restricted its attention to the development of the base language, without specifying any particular playback or authoring environment functionality. A brief summary of SMIL is presented in the next section.

The Esprit Chameleon project has developed a viewer and editor for CMIF which has been modified to use SMIL called GRiNS a GRaphical INTERface for SMIL. Section three provides an overview of the GRiNS authoring and viewer interfaces. Section four will provide an overview of the link editor for video supporting SMIL. Section five will then discuss the availability of the GRiNS environment as part of the Chameleon suite of authoring and presentation tools.

2. Synchronised Multimedia Integration Language - SMIL.

A typical application for which SMIL can be used is that of an electronic newscast. An example newscast is used here to illustrate SMIL's functionalities. The plot of the newscast is that the newscast titles, logo and theme music are presented, followed by an anchor's introduction and the first story about the growth of the WWW; during this section a graph appears when growth statistics are mentioned and then disappears; the anchor turns to a remote reporter, who is initially introduced by a textual title which disappears during his report; the reporter returns to the anchor, who wraps up the broadcast, closing with music. At any point in this example, the user can select objects in the presentation which will activate links to other WWW resources.

SMIL documents describe the temporal and spatial layout details of a multimedia presentation along with links from those media, and details of alternatives for communication and presentation. The SMIL language is intended to use a variety of communication protocols including HTTP which is normally used on the WWW for HTML pages, and RTSP which supports streaming of continuous media such as video and audio. The choice of communication protocol will depend on the available networks. Individual assets in many media languages (e.g. HTML, GIF, etc.) can be included within a SMIL document, and the choice of asset may depend on a combination of the available communication network bandwidth, the resources available at the display station, and the end user's abilities and preferences. The following sub-sections describe these features of the language in more detail.

2.1 Spatial Layout in SMIL

SMIL documents, like HTML documents are divided into a head and a body. The head is used to define the spatial layout of the media assets to be presented as channels on the screen.

```
<smil>
  <head>
    <layout type="text/smil-basic">
      <tuner id="left-video" left="20" top="50" z="2"/>
      <tuner id="left-text" left="20" top="120" z="2"/>
      <tuner id="right-video" left="150" top="50" z="2"/>
      <tuner id="right-text" left="150" top="120" z="2"/>
    </layout>
  </head>
  <body>
    XML comments & XML processing instructions
  </body>
</smil>
```

The example above shows four channels, two for each of video and text, positioned using the smil-basic layout language. Other layout languages are possible, and alternative layouts can be used with various conditions using the switch statement explained in section 2.4.

2.2 Temporal Synchronization in SMIL

The document body is used to synchronize the presentation of different media assets in the layout channels. This is controlled through the use of `<PAR>`, `</PAR>` and `<SEQ>`, `</SEQ>` statements to enclose assets which should be presented either in parallel or sequentially.

```
<body>
  <seq>
    <par>
      <img href="graph" loc="left-video" dur="45s"/>
      <text href="graph-text" loc="left-text"/>
    </par>
    <par>
      <video href="tim-video" loc="left-video"/>
      <text href="tim-text" loc="left-text"/>
    </par>
  </seq>
</body>
```

The example above shows that a sequential presentation is made, consisting of two presentations, the first of which contains an image and some text presented in parallel, the second of which contains a video and some text presented together in parallel.

The attribute `dur="45s"` defines the duration of the first presentation to be forty five seconds while the attribute `loc="left-text"` selects which of the channels whose location was specified in the header should be used to present the information.

2.3 Linking in SMIL

Conventional HTML links can be used from complete assets, for example :

```
<a href="http://www.w3.org/People/Berners-Lee">
  <video href="tim-video" loc="left-video"/>
```

states that if the user clicks anywhere on the video "tim-video" a link will be traversed to the location "http://www.w3.org/People/Berners-Lee", which provides further information on the person feature in the video.

In addition, the `<anchor>` statement may be used to define links attached to anchors at locations specified by the `coords="0%,0%,50%,50%"` attribute in a running video, at times specified by the `begin="0s"` and `end="5s"` attributes. The following example shows how two anchors are assigned to portions of the video `http://www.w3.org/CoolStuff` pointing to a WWW page `"http://www.w3.org/AudioVideo"`. More information on writing anchor statements is provided in section four.

```
<video src="http://www.w3.org/CoolStuff">
  <anchor href="http://www.w3.org/AudioVideo" begin="0s"
    end="5s" coords="0%,0%,50%,50%"/>
  <anchor href="http://www.w3.org/AudioVideo" begin="5s"
    end="10s" coords="50%,50%,100%,100%"/>
</video>
```

2.4 Alternatives and Tailoring Presentations in SMIL

The fourth major construct in SMIL is to manage the use of alternatives. This uses a `<switch>` statement dependent on attributes provided by the viewer tool. This construct allows presentations to be tailored to the bandwidth of the communications network by such alternatives as using still images instead of video; to allow audio or text to be presented in the preferred language of the user in order to facilitate internationalization of information; to select the graphics and sound content which best suited to the presentation equipment available to the end user; and to allow the presentation of text instead of audio, or sound instead of visual information to make the presentation accessible to users with disabilities or strong preferences for a presentation medium because of their environment or personal taste. Each of these uses of the alternatives construct increases the size market which the presentation can reach and therefore increases both the impact of the information and its potential profitability.

The example below shows how the language of an audio channel will be selected to match the preferred language of the user:

```
<switch>
  <audio src="french_audio.wav" language="fr"/>
  <audio src="english_audio.wav" language="en"/>
</switch>
```

2.5 The potential of SMIL

SMIL has been designed to support each of these four constructs of temporal and spatial specification, and the use of links and alternatives, in order to allow multimedia presentations to be optimised for both the user and the available technology. It allows each of these constructs to apply throughout a presentation, and can be used with either download or streaming communication protocols to minimise delays due to downloading

assets. Since SMIL does not require that the whole presentation be converted to video and streamed, so that for the newscast example it requires only a sixth of the bandwidth that tools using that approach do. Finally, SMIL has been defined in XML[8], which is a subset of the ISO standard SGML, and it has been demonstrated [9] that because of this it can be easily converted to and from other standard languages for storing text and multimedia in databases such as HyTime[10]. In summary, SMIL is intended to overcome the limitations found in the alternative means of presenting multimedia over the WWW.

Although SMIL has been designed to use a declarative language with attribute/value pairs in order to be easy to write by hand without the need for programming skills following the example of HTML in the hope that it will be as widely received. Like HTML, many users will prefer to use an authoring environment to produce SMIL documents. The next section describes the GRiNS authoring environment for SMIL.

3. GRiNS Authoring and Viewer Interface.

The GRiNS authoring environment supports creation of a SMIL document in terms of four views shown in the left pane of Figure 2: the end user view shown on the lower and upper left, the hierarchical structure of the assets used shown in the lower centre, the hyperlinks and anchors shown in the lower right, and the timeline of the presentation shown in the upper right. In the upper centre are shown the main control menu of the editor, and the control panel for the end user view.

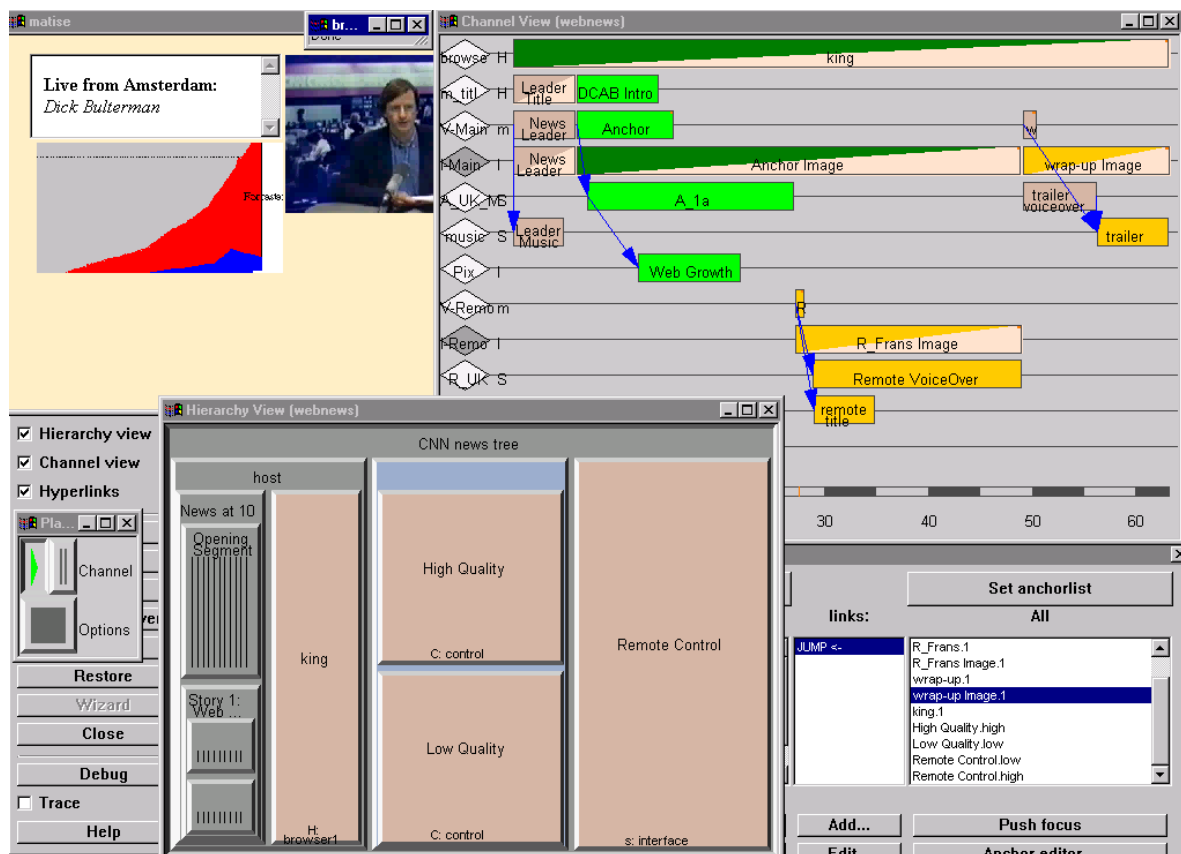


Figure 1: The user interface to the GRiNS authoring environment showing the end user view (top left), the channel view (top right), the hyperlink view (bottom right) and the hierarchy view (bottom center); the control panel for the overall environment is shown at the bottom left.

3.1 The end user view

The end user view provides a WYSIWYG view of the presentation under development, as well as a mechanism to interactively lay out the spatial position of the layout channels associated with the presentation.

The end user view activates the controls of the viewer which the end user will use.

3.2 The Hierarchical Structure View

The newscast example can be viewed as a hierarchy of three main segments, with the story breaking down into the titles and introduction, the remote segment, and the anchor wrap up & close. In turn each of these segments can be further decomposed to produce a hierarchical structure terminating with the individual media assets to be presented. Such hierarchies can generalize across a several stories, and the hierarchy view allows this form of decomposition to be viewed explicitly, and such structures to be re-used as templates.

3.3 The Timeline View

The hierarchical structure is intended to represent the coarse timing relationships reflected in the `<par>` and `<seq>` constructs. While the attributes associate with an element can be used to define more fine grained relationships which are often difficult to visualize with the hierarchical view alone. For this reason GRiNS also supports a timeline projection of a presentation. Unlike some systems which use the timeline as the initial view of the application, the GRiNS timeline displays relationships derived from the hierarchical structure. This means that the user is not tied to a particular clock or frame rate since the actual timing relationships will only be known at execution time.

When working with the timeline view the user can define exact start and end offsets within `<par>` groups by using a synchronisation arc shown as an arrow.

3.4 The Hyperlink View

The Hyperlink view presents the anchors and links in the presentation and allows their creation and modification. Since the `<anchor>` statement separates anchors from links as is done in sophisticated hypermedia languages such as HyTime, but not in simple ones such as HTML, a separate view is required to explicitly enter the link after the anchors are created. SMIL supports the use of different classes of anchor, such as FORK which must be coded in JavaScript for HTML pages, but have been introduced into the XML language [8] in which SMIL is defined. A fork link is used in the example to show the home page of the remote reporter in a separate window to the main SMIL presentation.

4. Video Link Editor.

One novelty in SMIL is the introduction of selectable anchor maps on video presentations. Such maps were considered for audio too, but no common agreement on how to provide the user with a selection mechanism could be reached so they were not included. Video anchor maps present a particular problem in production since the object to be selected moves around the screen within a shot from frame to frame. To hand code such maps is very time consuming so a special editor called MOVie has been developed to encode video maps. This allows the user to mark areas of a screen at the start of a shot with the identity of an object, and mark the position at the end of a shot, and linearly interprets the position of the object on the intervening frames. This allows the fast production of `<anchor>` statements

for a video clip, although for long video segments which cut between many shots this is still a considerable effort.

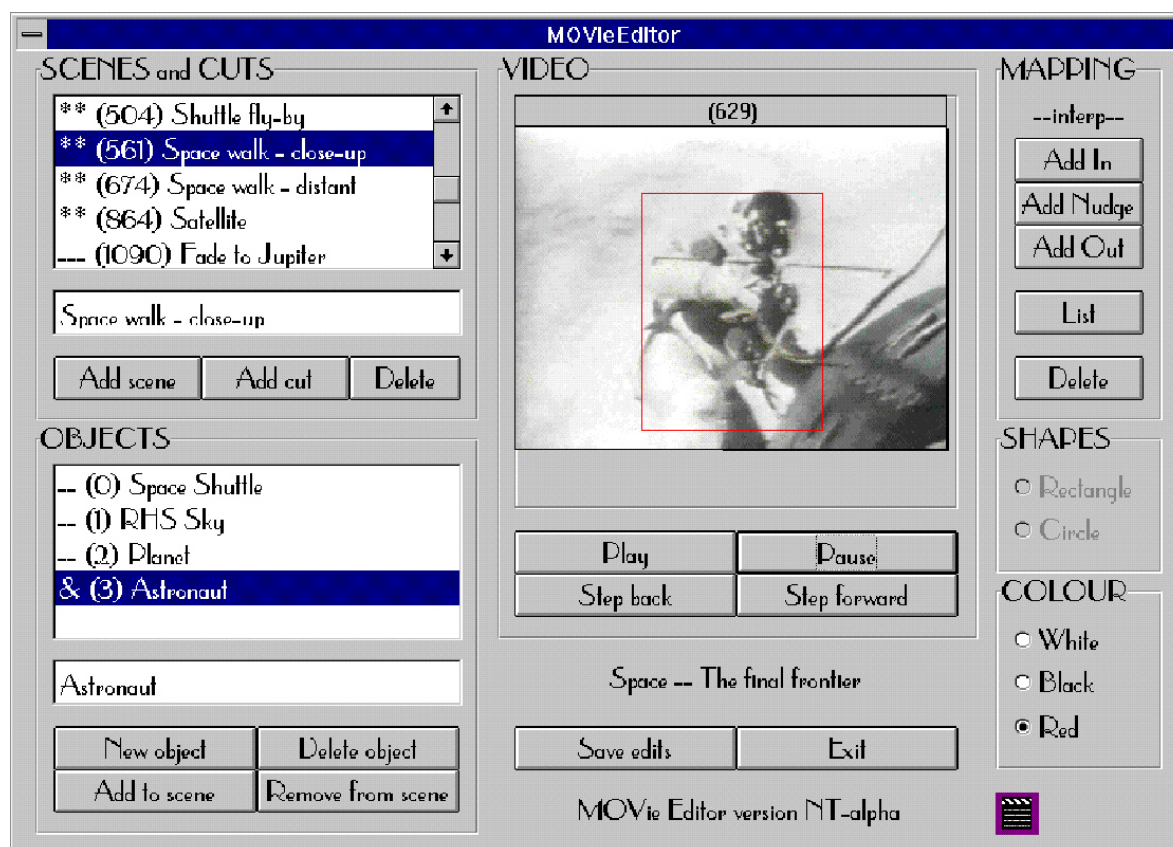


Figure 2: The user interface to the MOVie editor for producing anchor maps for video is shown in the left pane, and the place of the editor as tool support for the selectable video production process is shown in the right pane.

Figure 2 shows the user interface to the MOVie editor and a schematic of the role this has in the production process. The output of MOVie must still be passed to the GRiNS editor so that the anchors specified can have links associated with them. Similar editors are commercially available for other video formats [12] so we expect that they will soon be available for SMIL too, although without an overall SMIL editor to establish the links with the rest of the document, they may have more limited functionality than MOVie.

5. Availability of GRiNS & competitors

The GRiNS authoring environment and viewer have been implemented on all major presentation platforms (Windows-95/NT, Macintosh and UNIX). They are intended to provide a reference architecture for playing out SMIL documents and to integrate structured authoring concepts in SMIL document creation.

A SMIL player is currently available from DIGITAL Inc. as a plug-in for Netscape Explorer [13] while Real Networks have announced that their next streamed media player and server products will be built around SMIL[16]. Further players are expected soon from most of the major browser manufacturers. The GRiNS SMIL viewer is currently available to W3C members [14], and the whole GRiNS environment will soon be widely available from the company established in Amsterdam to exploit it [15].

It is expected that a wide range of applications will be developed in SMIL, and GRiNS is intended to be the first authoring environment available for it. Since RealNetworks have announced that their product range will conform to SMIL, a number of companies which currently produce HTML, or web site editors have announced the intention of producing SMIL editors, although there are currently no details of release dates [17]. The market in this area will obviously develop quickly, and it is hoped that GRiNS will establish an early market position which it can maintain.

Several organisations have so far expressed interest in using SMIL to develop low quality television programmes for Internet delivery, particularly for applications in remote learning.

5. Conclusions

SMIL has great potential for defining multimedia presentations which could move the look and feel of the web from a model of paper documents towards one of television programmes. W3C expect it to be adopted as a recommendation this year, and various WWW browser manufacturers are expected to produce viewers for it soon. GRiNS is intended to be the first authoring environment for this language, and given the long history of academic research which has led to it, it is hoped that it will have a realistic chance even when the market for SMIL authoring matures.

Acknowledgements

The work of the W3C's SYMM working group was coordinated by Philipp Hoschka of W3C/INRIA. A list of contributors is available at [7]. The research reported in this paper was partly funded by CEC Esprit IV grant 20597 to the CHAMELEON project.

References

- [1] Macromedia Inc.: Director 6.0 <http://www.macromedia.com/software/director/>
- [2] Microsoft Inc. Netshow <http://www.microsoft.com/netshow/>
- [3] Real Networks Inc. Real Media <http://www.real.com/>
- [4] D.C.A. Bulterman, R. Van Liere and G. Van Rossum. A Structure for transportable, Dynamic Multimedia Documents. In Proceedings of Usenix Spring Conference on Multimedia Systems, Nashville, TN, 1991, 137-155.
- [5] D.C.A. Bulterman Synchronisation of Multi-Sourced Multimedia Data for Heterogeneous Target Systems. In P. Venkat Rangan (Ed.) Network and Operating System Support for Digital Audio and Video, LNCS-712, Springer Verlag, 1993, 119-129.
- [6] L. Hardman, D.C.A. Bulterman and G van Rossum. The Amsterdam Hypermedia Model: Adding Time and Context to the Dexter Model. CACM 37(2), Feb 1994, 50-62.
- [7] W3C, SMIL Proposed Recommendation <http://www.w3c.org/TR/1998/PR-smil-19980409/>
- [8] W3C, XML Recommendation, <http://www.w3c.org/TR/1998/REC-xml-19980210>
- [9] L. Rutledge, J. Van Ossenbruggen, L. Hardman and D.A. Bulterman, Practical Application of Existing Hypermedia Standards and Tools. In Proceedings of Digital Libraries '98., in press.
- [10] S. DeRose and D Durand. Making Hypermedia Work: A User's Guide to HyTime. Kluwer Press, Boston 1994.
- [11] V A Burrill, T Kirste, JM Weiß Time-varying sensitive regions in dynamic multimedia objects: a pragmatic approach to content-based retrieval from video. Information and Software Technology 36 (4) p.213-224, April 1994.
- [12] Veon Inc. V.Active Video Link Editor. <http://www.ephyx.com/>
- [13] Digital Inc, HPAS SMIL Applet for Netscape Navigator, <http://www.research.digital.com/SRC/HPAS/>
- [14] GRiNS SMIL Player, <http://www.cwi.nl/SMIL/grins.html>
- [15] GRiNS documentation, <http://www.cwi.nl/Chameleon/GRiNS>
- [16] RealNetworks, SMIL Compliant Player, <http://www.real.com/technology/smil/index.html>
- [17] RealNetworks, Third Party Authoring Tools, <http://www.real.com/g2/developer/tools.html>