SMIL 2.0

— repurposing broadcast content for the Web

Dick C.A. Bulterman
Oratrix Development, Amsterdam

As end-user bandwidth increases to a level where the (re)distribution of audio/video material via the Internet becomes attractive, XML-based standards that help broadcasters migrate their existing content to the Web are becoming richer and more powerful. SMIL 2.0 – developed by the World Wide Web Consortium (W3C) – is the newest version of the Web's most popular multimedia format.

This article provides an introduction to the concepts and facilities of the SMIL 2.0 language, in the context of the work flow requirements for taking existing broadcast content and making it available for a Web-centric audience.

Since the introduction of streaming audio in the mid-1990s, the Web has been seen by many to be a natural extension of over-the-air and cable-based content distribution models. Content producers would have an open arena for their wares, and users would have unprecedented freedom to select which programmes they wanted to see (and when!).

Where some prophets predicted that the Web would create an on-demand replacement to conventional broadcasting, the reality has been that the impact of the Internet on rich media distribution has been limited. While the Web has provided a compelling environment for the timely and on-demand distribution of text documents that are augmented with images and graphics, and while the redistribution of audio – and to a lesser degree, video – has made some important in-roads, the use of the Web for wide-scale distribution of conventional broadcast content has been minimal.

There are two reasons for the lack of migration of broadcast content to the Web. First, the limited and often unpredictable bandwidth available between the content on the server and the user at the client have resulted in a viewing/listening experience that has fallen far short of what most consumers have come to associate with broadcast content. Second, the costs in production time and server resources to host migrated content have exceeded any real or perceived benefit from making the content available on-line, especially given the lack of a clear business model for recovering these costs.

While the lack of bandwidth and a negative revenue model have combined to provide a chicken-and-egg stagnation of Web multimedia, there is reason to believe that the fundamental advantage of Web-based distribution remains compelling enough to overcome the initial performance and cost problems. This advantage is the Web's ability to reach a differentiated audience that can be more selectively targeted than conventional broadcast distribution.

Content differentiation: the key to the Web

The key to understanding multimedia and the Web is not to focus on the Web as a replacement for broadcast distribution, but to view it as a vehicle to build user communities that augment the broadcast audiences. In order to capitalize on the inherent diversity of the on-line community, broadcasters (and other Web content providers) need to be able to add value easily to existing assets. This diversity is reflected in:

- **Connection speed/type** — unlike the uniformity of the broadcast world, some Web users have high-speed ADSL/Cable connections (and expect a correspondingly high user experience), others use low-cost conventional infrastructures (and are willing to trade-off quality for cost), while others will use mobile devices such as telephones and PDAs (and be willing to pay a premium for anytime/anywhere access). Focusing on only one group at the expense of the others will lead to user alienation rather than acceptance.
- **Device characteristics** — where TV set characteristics are uniform and well-known, the physical characteristics of the target of a Web presentation will vary widely in terms of resolution, screen size, pixel depth, and available storage/memory. Rather than providing convergence over time, the Web is becoming increasingly diverse in the devices available to the end users.
- **User community profile and make-up** — in an increasingly inter-connected world, the national and cultural composition of cross-border audiences require content tailored to multiple languages. Even within a single country's borders, the multi-cultural society will provide both challenges and opportunities for targeting information appropriately. At the same time, society will demand – through government regulation – that equal access to information be provided to all Web users, including those of varying physical (dis)abilities.
- **Diverse semantics, not only syntax** — where the mass audience of broadcast content encourages a common homogeneity, the targeted composition of the Web inspires heterogeneity. Some users will want to be able to get in-depth information out of a programme (ranging from additional information on the place of a disaster to the on-line address at which the remote reporter's coat can be purchased), while others will want to selectively skim a programme's content. If the first generation of Web content has taught us anything, it is that content selectivity is far more important to users than design quality or even performance.

If managing diversity in a cost-effective manner is the key to both controlling the cost of broadcast content repurposing and maximizing its impact, only one Web multimedia standard really makes sense for wide-scale broadcast-to-Web deployment: SMIL. The W3C's Synchronized Multimedia Integration Language, first released in 1998 and now about to go into version 2.0, has been specifically designed to meet the needs of multimedia from a Web perspective. Already installed on over 200,000,000 desktops worldwide, SMIL provides an XML framework for identifying, scheduling and activating media objects that can be combined dynamically to offer a tailored presentation. Perhaps more importantly, SMIL can also be used to distinguish what is and what isn't relevant to an individual user: using content selection mechanisms built into the language, a presentation can be designed to be used with several target audiences from a single source document.

Abbreviations

3GPP	3rd Generation Wireless Protocol
ADSL	Asynchronous digital subscriber line
GRPS	2nd Generation Wireless Protocol
HTML	Hypertext markup language
MPEG	Moving Picture Experts Group
PDA	Personal digital assistant
SMIL	Synchronized multimedia integration language
SVG	Scalable vector graphics
SYMM	(W3C) Synchronized Multimedia Working Group
W3C	World Wide Web Consortium
XML	Extensible markup language

SMIL: Synchronized Multimedia Integration Language

The first thing that most broadcasters focus on when migrating content to the Web is the low-level encoding of the media objects being sent. In this respect, it is important to note that SMIL is not a *media encoding format,* but instead a *presentation description language*. A SMIL presentation contains references to media objects (along with timing and layout information), but not the media objects themselves. SMIL does not impose any restrictions on the types of media objects used in a presentation, or their encodings. Of course, some SMIL players may not support all available formats, or they may favour their vendor's proprietary media encodings.

The general structure of a SMIL presentation is shown in the yellow text panel on the next page. A SMIL file starts with the *<smil>* tag, followed by an XML *namespace identifier* which defines the version of the language being used. (A particular SMIL player may also identify any extensions it supports using another namespace ID.) An individual SMIL file may also identify any extensions it requires, by using additional namespace prefixes and the *systemRequired* attribute. The SMIL file further contains *<head>* and *<body>* sections: the *head* section contains meta information, custom control, layout and transition definitions, and the *body* contains a collection of scheduling elements/attributes and media object references.

Orchestrating the content activation, placement and delivery

The example SMIL file illustrates some of the main facilities available with version 2.0 of the language:

- **Layout control provides predictable object placement** — whereas object placement is often a by-product of document style in some Web languages, the SMIL 2.0 Language Profile has a full set of layout elements and attributes that give full graphic design control of the presentation to the designer.
- **Timing is structured via nested collections of *par* and *seq* containers** — rather than using a single fixed notion of time, SMIL structures the objects within parallel (*par*) and sequential (*seq*) containers. Objects inside a *seq* container are activated in monotonic order, while objects in a *par* container can be scheduled flexibly.
- **Conditional content is supported via the *switch* container** — given the diversity of the Web, SMIL provides a *switch* container that allows alternatives for a media object to be defined. The first object whose *test attribute* predicate resolves to *true* is used.
- **Transitions and animations can be attached to media object references** — various forms of visual effects can be added to objects. The key aspect of these transitions and animations are that they occur locally in the player, and thus require no extra network bandwidth.

Figure 1
Two views of the document shown on the following page, as determined by the *switch* setting.

Fig. 1 shows two views of a SMIL file under different presentation conditions: on the left is the view provided when using a connection speed of greater than 56 kbit/s, and on the right is the view of the same presentation when using less than 56 kbit/s. Note that the SMIL language does not define how the available bandwidth is measured. Some players will use a static estimate, stored in a configuration file, while other players will use actual measured bandwidth.

A-temporal composition and event timing

In the normal course of processing a SMIL file, the activation of individual objects depends on the (nested) position of the object in the *par*/*seq* container hierarchy. Although there may be slight variations from the modelled time during the actual presentation (due to network characteristics or player performance issues), the time structure of the document will be translated directly into the runtime behaviour of the presentation.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE smil PUBLIC "-//W3C//DTD SMIL 2.0//EN"
                        "http://www.w3.org/2001/SMIL20/PR/SMIL20.dtd">
<smil xmlns="http://www.w3.org/2001/SMIL20/PR/Language">
 <head>
   <meta name="title" content="Oratrix New Year's Promo"/>
   <meta name="generator" content="GRiNS Pro Composer for SMIL, v2.0 build a58"/>
<!--
    This example defines one audio region and two internal drawing regions.
-->
   <layout>
     <topLayout id="Window" backgroundColor="black" width="528" height="395">
       <region id="Backgrd" width="528" height="395"/>
       <region id="Audio"/>
       <region id="Main" left="62" width="420" top="42" height="314" z-index="1"/>
     </topLayout>
   </layout>
   <transition id="w-0" type="pushWipe" dur="1.25" subtype="fromLeft"/>
   <transition id="f-0" type="fade" dur="0.5"/>
   <transition id="f-1" type="fade" dur="2.5"/>
 </head>
<!--
  The <head> section contains the main structure of the SMIL presentation.
  <par> defines a general timing container, while <seq> constrains its
  children to execute in sequence.
-->
 <body>
   <par id="NewYear" dur="59s">
     <seq id="background" fill="remove" fillDefault="hold">
       <img id="btrade" region="Backgrd" dur="1" src="btrade.JPG" transIn="f-1"
                        left="186" width="130" top="3" height="38" transIn="f-0"/>
       ...
       <img id="trailer" region="Backgrd" begin="7.25" src="tr.jpg" transIn="f-1">
         <area id="trailer-0" show="new" href="http://www.oratrix.com/GRiNS/"/>
       </img>
     </seq>
     <audio id="AudioMix" region="Audio" src="Noel1.wav"/>
<!--
  The following switch selects the first element that meets the
  switch criteria. (If the bandwidth detected by the player is at least
  34.4 kbps, then the first <seq> is selected. Otherwise the second is used.
-->
     <switch id="pictures">
       <seq id="ImgColor" title="Color Images" systemBitrate="34400">
         <img id="fill" region="Main" dur="3" src="fill.gif" transIn="f-0"/>
         ...
         <img id="in2" region="Main" dur="1" src="in2.jpg" transIn="f-1"/>
       </seq>
       <seq id="ImgBW" title="BW Images" fill="remove" fillDefault="hold">
         <img id="fill-3" region="Main" dur="0" src="fill.gif" transIn="f-0"/>
         ...
         <img id="in2-0" region="Main" dur="1" src="in2.jpg" transIn="f-0"/>
       </seq>
     </switch>
   </par>
 </body>
</smil>
```

One exception to this are objects that are contained inside a SMIL 2.0 <*excl*> container. At most, one of the content objects of the exclusive container can be active at any one time. When a new element becomes active, based on timing and activation attributes, it either pauses or replaces a previously activated object (if any). Rather than defining purely temporal composition, the exclusive container defines a composition not directly time based (i.e. *a-temporal*). For example, consider the following SMIL fragment:

```
<par id="Slideshow" endsync="Bkgd-Clip">
  <excl id="OnDemandSequence">
    <par id="Slide_1"  begin="0s; next_3.activateEvent" dur="10s">
      <img id="Image_1" region="M" src="img1.jpg"/>
      <img id="next_1"  region="N" src="next.gif" begin="5s"/>
    </par>
    <par id="Slide_2"  begin="next_1.activateEvent" dur="10s">
      <img id="Image_2" region="M" src="img2.jpg"/>
      <img id="next_2"  region="N" src="next.gif" begin="5s"/>
    </par>
    <par id="Slide_3" begin="next_2.activateEvent" dur="10s">
      <img id="Image_3" region="M" src="img3.jpg"/>
      <img id="next_3"  region="N" src="next.gif" begin="5s"/>
    </par>
  </excl>
  <audio id="Bkgd-Clip" region="audio" src="tune.MP3"/>
</par>
```

In this fragment, we see a new SMIL 2.0 timing construct: the <*excl*>, or exclusive. The <*excl*> has three children: Slide_1, Slide_2 and Slide_3. (Each of these are <*par*>s containing an image and a "next" button. When each <*par*> starts, the corresponding image is displayed and 5 seconds later, the "next" button appears.) All of this is accompanied by a music clip that plays throughout the presentation.

The first child of the exclusive (Slide_1) starts immediately: when begin="0". Each following <*par*> starts when its predecessor's "next" button has been clicked. The entire substructure ends when the background audio ends.

In addition to the *a-temporal* uses of the exclusive container, SMIL 2.0 also supports a general event-based activation model: individual objects can have event names associated with them, and then only start – or end – when that event occurs. SMIL also supports the notion of timed hyperlinks: when a link leads to an object, it starts in its "temporal context". That is, all of the other objects in the presentation that would have been active if the linked-to object had started as the result of the presentation's normal execution, are also started when a timed link is followed.

As we will show later in this article, the combination of SMIL's constructs can lead to compelling Web content that can easily be tailored to the needs and environment of the user.

Export to multiple formats and profiles

One of the key features of SMIL is that it is purely *declarative*, not algorithmic: a SMIL file defines how a presentation should behave, rather than defining a particular algorithm to implement that behaviour. By not using a scripting language to define the presentation, the SMIL file can be used as a single source that can be transformed into many different formats. Some of these formats may be defined on code fragments generated automatically from the SMIL file, such as transforming SMIL into MPEG-4 or Quicktime formats.

Another example of transforming SMIL is to move from one SMIL 2.0 *profile* to another. Using essentially automatic transformation techniques, a presentation created for the SMIL 2.0 *Language Profile* can be mapped to the SMIL 2.0 *XHTML+SMIL Profile*. It can also be mapped to the SMIL 2.0 *Basic Profile* for use in reduced complexity devices. At present, the Language Profile – which is an extension to the existing SMIL 1.0 language – is the most prevalent, but various vendors are working on a range of SMIL integration efforts.

Using SMIL to reuse content

The SMIL 2.0 specification is a large and imposing document. Although its 500 pages is of moderate size when compared with many broadcast standards, one can easily lose the intent of the language without a more concrete example. In this section, we attempt to provide such an example in terms of the migration of an Evening News programme to the Web.

Using a combination of current HTML and multimedia player technologies, it is already fairly easy to take broadcast content and migrate it to the Web. One typical production sequence might be to take a video tape version of, say, a 40-minute broadcast, convert the entire file to a streaming format (such as RealVideo or QuickTime's video format), place the converted file on a streaming server and then place a link to the server on an HTML page. Among the production decisions that need to be answered are defining a target bandwidth for the encoded news, as well as a target screen size for the encoded video. In order to comply with government directives to make the video accessible, the closed captions may need to be extracted from the source, displayed in a version with a separate subtitle area.

The goal

Once the process described above is complete, a user would have access to the entire 40 minute newscast as an integral media object. Currently, the quality of the video would be marginal: a postage-stamp, common denominator presentation lasting (in Web terms) an eternity, with a user-defined starting time being the only advantage over broadcast distribution.

The goal of our design example is to identify ways of giving added value to the Web content. Rather than simply duplicating over-the-air uses, our Web version uses SMIL 2.0 to offer selectivity, presentation tailoring and device customization. The following set of design goals define the presentation's scope:

- **Delivery device** — the presentation will be viewable on a conventional PC or a next-generation SMIL-powered telephone.
- **Connection bandwidth** — rather than using a "one size fits all" model, the presentation will offer content differentiated by available bandwidth: users on high-speed connections will receive video, while users on (relatively) slow speed connections will receive a sequences of images.
- **Display size** — users with a PC screen will receive image-driven menus, while users on small devices (such as a telephone or PDA) will be given text menus. The graphic size of the presentation will also vary by device and connection speed.
- **Captions/subtitles** — depending on an end-user's needs and wants, captions will be provided using a timed-text format.
- **Multiple languages** — the document will be structured so that the presentation's labels and captions will be provided in a base language (in this case, English) and in at least one other language, perhaps provided by a local redistributor.
- **Selective activation** — rather than forcing a user to watch the entire newscast, viewers will be able to select stories based on a menu of story choices.
- **Additional content** — in addition to the basic news story, viewers will be able to access additional features on an on-demand basis. Once the on-demand story ends, the newscast should continue from the point of diversion.

While presentation tailoring is currently seen as a production luxury that few broadcasters can afford, such customization will become increasingly important as the Web becomes a more diverse collection of high/low speed, large/small display, and fixed/wireless devices. While customization may seem to bring with it an explosion of costs, it may in fact bring with it a basis for revenue: by offering customized differentiated content, users groups can be identified and marketed. While a basic version may always be "free" (except for device and connect charges), augmented versions for high-end or speciality groups can be offered at premium prices, either by subscription or on an individual pay-per-view basis.

Work flow issues

Fig. 2 shows the basic workflow required to meet the design described above.

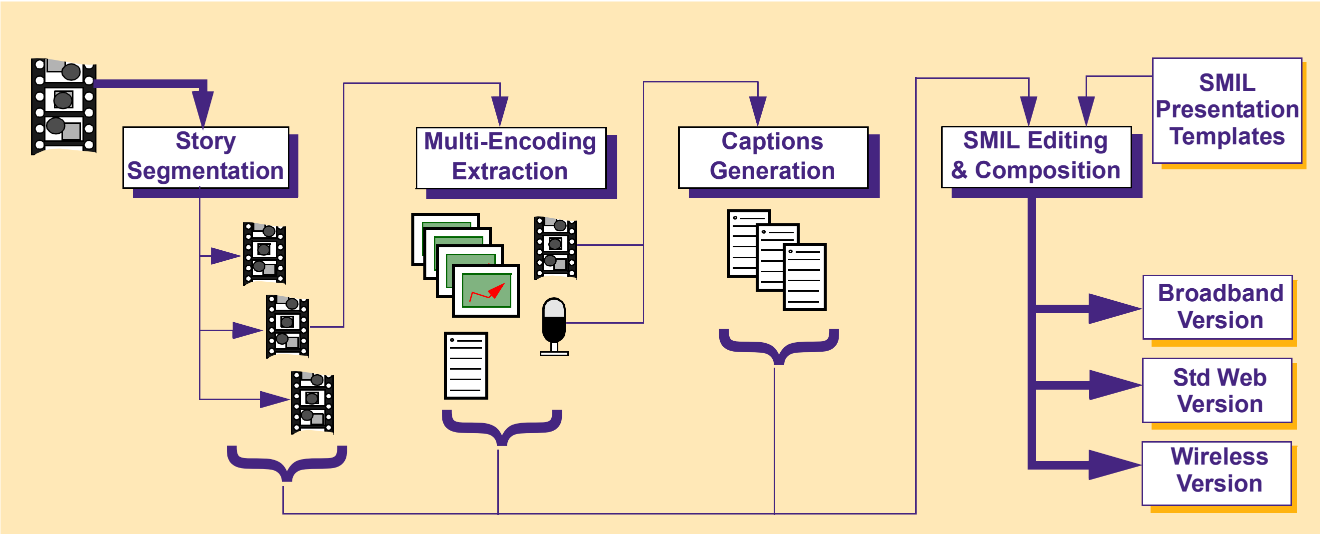

Figure 2
The work flow required to give added value to Web content.

The process starts by segmenting the newscast into a collection of stories. This segmentation can be derived using the over-the-air video, or by using post-production copies of individual stories and adding the news reader introductions from the broadcast. Many commercial software packages exist for story segmentation and meta-information labelling.

The media type for each story can then be multi-sourced. In our example, we will make a slide-show reduction of each story by splitting the audio from the video, and extracting video frames into slides. Depending on the desired level of customization, multiple versions of the audio, video and image files (in various sizes or formats) could be generated. While much of this process can be semi-automated, the actual frame selection of images is best done by a production staffer to avoid generating slides that are inappropriate to a story's content.

The next step is to either recapture or generate captions for each story. This can be done by extracting closed captions from the video source, by reusing scripted material or by generating new captions during Web production using a captioning tool. If desired, the captions can also be translated into multiple languages.

Once the composite audio/video object and the individual slides, audio and text objects are completed, logical groups of assets are available for packaging into individual story groups. These story groups can be inserted into the SMIL file either by using a template-based presentation-editing system, or by a text editor. An advantage of an editing system is that incremental preview and bandwidth simulation may be available. Good authoring tools will also support multi-format output generation.

Incremental development

One of the advantages of SMIL is that, as a text format, the presentation can be created incrementally. An initial version of the presentation can be created for the widest market, and then fine-tuned for the needs of each target environment. The entire design process can be simplified by using a set of SMIL templates. The designing of compelling templates which make full use of each environment is a specialist task but, once this is done, the template can often be reused with little or no changes for every evening's broadcast.

Examples

Fig. 3 shows various aspects of the presentation just discussed. In the *Broadband* and *Web* editions, the user is given an image-based menu of stories. Using the SMIL *exclusive* element, only one of the available stories is actually displayed. (Automatic – lazy – playback is also available.) The user can also navigate from one story to another using the navigation bar. Note that in this illustration, the "Southeast Newsroom" is made available: this can easily be regionalized using SMIL's *custom test* elements. The image area used in the *Broadband* edition is larger than its *Web* counterpart and the *Broadband* audience will get to see video in a pseudo 16x9 format while the *Web* edition contains slides with transitions in 4x3 format.

(a) Broadband version (b) Web version (c) Mobile version

Figure 3 — The BBC 10 O'clock News
This demonstration illustrates how a single broadcast news production can be repurposed for use in broadband Internet, conventional Web, and mobile environments. (The broadcast version, available to ADSL/cable customers, contains video content: the Web and wireless versions contain images.) Note the ability to support multi-lingual captions.

The *Wireless* edition of the presentation has a different look and feel (because of the small screen size), but the functionality is identical. The *Wireless* edition uses a slideshow (without transitions, since these aren't supported by the particular player). Also, the menu just shows the main categories rather than individual stories.

The use of captions is controlled by the preference setting of the user. Of course, you don't need to be deaf to need captions – if you are watching the news during a staff meeting, it is typically wise to turn off the volume and use the captions instead! (Note that the images of the *Wireless* edition shows captions available in Dutch. These may have been added in the version of the SMIL file made available by the Dutch distributor.)

One other interesting aspect: the "What's On" buttons at lower right of the *Broadband* and *Web* editions allow a viewer to selectively activate the content outside the newscast. (In this case, they get a promo of a programme to be broadcast later that day on each channel.) When the promo ends, the news continues where it was paused.

During production, the entire presentation can be saved as a single SMIL file, or the master SMIL file can be partitioned into separate files when the presentation is published. Even if separate files are generated, the media assets themselves only need to be stored once, since they are included in the presentations by reference.

The Structure of the SMIL 2.0 standard

Unlike SMIL 1.0, which was a monolithic standard, SMIL 2.0 has been defined as a collection of reusable modules. There are twelve major functional areas, including *timing/synchronization*, *layout*, *transitions*, *animation*, *content control*, *metainformation*, *linking* and *media objects*. Many of the major functional areas are further subdivided into module sets that define *basic*, *enhanced* and *advanced* elements and attributes. The reason for modularizing SMIL is to provide a set of reusable components. At present, for example, SMIL animation has been integrated into the SVG standard, and other XML languages will reuse SMIL components in the future.

Obtaining SMIL players and tools

Table 1 (on the next page) gives a partial list of vendors who support commercial implementations of SMIL. Mass-market SMIL players (such as the *RealPlayer*) are available as part of the standard software available on most user environments (e.g. PCs). Typical costs for SMIL authoring tools (other than text editors) range from $50 - $2,500 for a single-user licence, depending on features and work flow integration.

Closing Comments

This article has presented a basic overview of the facilities available in SMIL 2.0 for supporting content repurposing. Readers who wish to obtain more details or who would like copies of the SMIL files discussed in this article should feel free to contact: **Dick.Bulterman@oratrix.com**.

The SMIL 2.0 specification can be found via **http://www.w3.org/AudioVideo/**.

The author would like to thank the BBC and Ericsson for the use of their copyrighted materials in the illustrations, and reminds readers that the use of these illustrations does not imply commercial availability of the functionality described.

Dick Bulterman is managing director of Oratrix Development in Amsterdam, NL. Prior to founding Oratrix, he was head of multimedia and human-computer interaction at CWI, the Dutch national centre for computer science and mathematics in Amsterdam, where he had worked in various capacities since 1988. In 1991, he started a research project on networked multimedia authoring and delivery systems (the CMIF project), many of whose concepts were used by the SMIL 1.0 and 2.0 languages. Prior to joining CWI, Dr Bulterman was a professor of computer engineering at Brown University in Providence, RI (USA). He holds undergraduate and Masters degrees in Economics and Applied Mathematics, and a Ph. D. in Computer Science.

Dick Bulterman was a founding member of the W3C SYMM working group and has been active in the development of both SMIL 1 and SMIL 2. His group at CWI released the first SMIL 1 player in mid 1998, and his company Oratrix released the first publicly-available SMIL 2 player in September 2000. His principal interest is in providing authoring environments and tools that help to make rich presentations available to the widest possible audiences, regardless of national language usage, device/bandwidth characteristics or user-accessibility constraints.

Table 1
Vendors of SMIL-based products.

Product	Vendor	When
GR*i*NS SMIL 1.0 Player GR*i*NS SMIL 2.0 Player GR*i*NS SMIL 1.0 Editor/Publisher GR*i*NS SMIL 2.0 Editor/Publisher	Oratrix [a]	Available now Available now Available now Q3, 2001
XHTML+SMIL in IE-5.5 XHTML+SMIL in IE-6.0	Microsoft [b]	Available now Expected Q4, 2001
Proprietary GPRS Wireless (SMIL 1.0)	Ericsson [c]	Expected Q3, 2001
RealPlayer-6/7/8 (SMIL1.0) RealPlayer-*x* (SMIL 2.0)	Real Networks [d]	Available now Expected Q3/Q4, 2001
QuickTime 4.1 (SMIL 1.0)	Apple [e]	Available now
3GPP Wireless (SMIL 2.0 Basic)	3GPP Consortium [f]	Proposal available now

a. http://www.oratrix.com/GRiNS/.
b. http://msdn.microsoft.com/workshop/author/behaviors/time.asp.
c. http://www.ericsson.com/ia/Offerings/Interactivemedia/index.shtml.
d. http://www.real.com/.
e. http://www.apple.com/quicktime/download/.
f. http://www.3gpp.org/ftp/tsg_sa/TSG_SA/TSGS_11/Docs/PDF/SP-010079.pdf.