Using SIVA XML and SMIL for Interactive Non-linear Videos: a Comparison

Emanuel Berndl
Chair of Distributed Information Systems,
University of Passau, Germany
berndle@fim.uni-passau.de

Britta Meixner
Chair of Distributed Information Systems,
University of Passau, Germany
meixner@fim.uni-passau.de

Harald Kosch
Chair of Distributed Information Systems,
University of Passau, Germany
harald.kosch@uni-passau.de

ABSTRACT
With recent technologies, it is possible to create appealing multimedia presentations or extended videos with a high level of interactivity. Standards like SMIL provide extensive structures to describe metadata for timing and spacing of single media elements which then form a presentation. While multimedia presentations are viewed mainly in a linear manner, provide interactive and non-linear videos a much higher level of interactivity and navigational possibilities. In this work, we examine the expressiveness of SMIL for the support of interactive non-linear videos. It has to describe temporal and spatial relationships of videos and annotations, as well as interaction and navigational elements. We therefore compared SMIL with the SIVA XML. We tried to find ways to express SIVA XML structures with SMIL attributes and elements. After that, we compared the DTD/XSD of SMIL and SIVA XML using XML metrics. We thereby focus on the language implementations. We do not take their implementations in authoring tools or players into account. Concluding that SMIL has little disadvantages in terms of feasibility for interactive videos, we propose minor additions that could resolve these problems and make SMIL more appropriate for our use case.

Author Keywords
SMIL, Interactive Video, Non-linear Video, XML, Video Annotations, Multimedia Document, Metrics

ACM Classification Keywords
I.7 Document and Text Processing: Document Preparation—*Hypertext/hypermedia, Multi/mixed media, Standards*

INTRODUCTION
Nearly each web page provides multimedia contents today. These reach from animated images to sounds and embedded videos. With recent technologies and increasing Internet bandwidths, appealing combinations of different types of media and various forms of user interaction are possible. However, the contents of web pages are not temporally synchronized. Therefore, more advanced languages with spatial and temporal models are necessary. Two main fields of research can be found in this area, namely "multimedia presentations" and "hypervideos". An applicable definition of multimedia presentation is given by Nimmagadda et al. as follows: "*Multimedia presentations are collections of different media files [...] like text, images, videos, and animations with different resolutions, durations, and start-times. [...] The layout of multimedia presentations is defined by the locations and the start times of the objects*" [16]. In contrast, interactive non-linear videos are defined by us (extended from [9]) as follows: "*[...] An interactive non-linear video is a digitally enriched form of video materials arranged for an overall concept. It presents additional information beyond the original content. Furthermore, it offers new forms of influence and navigation in the video and additional contents*" [15]. They are a subset of hypervideos for which many different definitions and descriptions can be found. A summarizing definition can be given as: Hypervideo is defined as video based hypermedia that combines non-linear video structuring and dynamic information presentations. Video information is linked with different kinds of additional information (like texts, pictures, audio files, or further videos). Users can mouse-click on sensitive regions (having spatial and temporal characteristics) within the videos to access the additional information (heterogeneous hypervideo) or jump to other scenes (homogeneous hypervideo). Hyperlinks build a graph between main video scenes and additional information.

Watching multimedia presentations, the viewer is rather passive, but basic interaction and navigation may be possible. The viewer is elicited from his passivity viewing interactive non-linear videos in contrast. This form of video consists of video scenes ("main video") and additional information which enhance the scenes. Timeline and control bar are extended with additional functions. These provide control on the flow of the video and give hints on when additional information (in the remainder of this work referred to as annotations) is displayed. Decision elements in the video allow the selection of a certain branch of the video instead of watching it in a linear way. Furthermore, additional information which may be any type of medium, like text, image, or video, is added as an annotation to the main video. We proposed an XML format for this type of video in [13]. Its structure was designed for the definition of interactive non-linear videos while using SMIL [22] may be possible up to a certain point, but leads to problems and work-arounds in some areas. In this work, we try to show the advantages and disadvantages of using SMIL for the description of interactive non-

2nd International Workshop on Interactive Content Consumption at TVX14, June 25, 2014, Newcastle, UK.

linear videos compared to our XML format[1]. An overview of authoring tools and players for SMIL can be found in our previous work [14]. We furthermore tested the only working player (Ambulant Player[2]), which showed weaknesses in the display of presentations as well as stability. Thereby, this work makes the following research contributions: Requirements were identified (see section REQUIREMENTS) and both formats were checked for their suitability to implement these requirements (see section FEASIBILITY ANALYSIS). Metrics were used to compare the complexity of the SIVA XML schema and SMIL, see section COMPARISON/METRICS.

REQUIREMENTS

An analysis of usage scenarios (further described in [12] and [14]) like e-learning, virtual tours, mobile help systems, or sport events revealed several requirements according to timing and spacing of media elements in interactive non-linear videos with additional information. Needed functions and elements are as follows (see also [13]):

- *Media, main video, and annotations*: As specified for interactive non-linear videos, usually one video is displayed as the main video. Additional information may be shown with this video. Therefore, several different types of media like images, audio-files, videos, and text should be usable. It should be possible to handle them differently during playback, for example should a subtitle be positioned automatically.

- *Event-based timing model*: Main video and annotations may be time dependent or time independent. For this reason, an event-based timing model is preferred to a structured timing model due to the high level of interactivity mixed with fixed points in time were annotations are displayed or hidden. By keeping timing issues as local as possible, synchronization is realizable more easily.

- *Temporal relationships between main video and annotations*: Temporal relationships in form of start and end point or durations of display need to be defined between the main video (scene) and each of the annotations.

- *Spatial relationships between videos and annotations*: A positioning of main video and single annotations or groups of annotations needs to be defined. Annotations may be displayed statically in areas around the video or as an overlay over the video. Furthermore, dynamic annotations may move on a path on the video canvas. Automated arrangement of annotations in defined areas facilitates the authoring process.

- *Decision elements at forks in video flow*: The playback of interactive non-linear videos includes different strands of scenes. Selection elements are needed to select the next scenes which are displayed to the viewer. Selection elements may be buttons or links.

- *Table of contents*: One way of extended navigation is provided by a table of contents which has to be defined and linked with single scenes.

- *Keyword reference list*: A second way of extended navigation is implemented with a keyword search. Keywords need to be linked with scenes or annotations in order to find information more quickly.

- *Extensibility*: The structure of the XML format has to be extensible in case of new ways of interaction that should be mapped into the model. Furthermore, changes in the XML file should be kept as local as possible in the structure without changing bigger parts of the existing file. Scripting is not considered as useful with respect to the affordance of an easy to use authoring tool.

DESCRIPTION OF THE FORMATS

The SIVA XML schema and the SMIL DTD show several differences in structure and scope. While SMIL tries to cover many different areas of application, the SIVA XML schema is exactly tailored to the needs of interactive non-linear videos with additional information. We now give a short overview over the formats before we compare them based on the requirements we determined in the previous section.

SMIL DTD

SMIL stands for Synchronized Multimedia Integration Language and it is a standard for interactive multimedia presentations released by the World Wide Web Consortium (W3C). Design goals of SMIL were to define "an XML-based language that allows authors to write interactive multimedia presentations. Using SMIL 3.0, an author may describe the temporal behavior of a multimedia presentation, associate hyperlinks with media objects and describe the layout of the presentation on a screen. [Furthermore, it should allow] reusing of SMIL 3.0 syntax and semantics in other XML-based languages, in particular those who need to represent timing and synchronization" [22]. Used media files are images, text, audio files, videos, animation, and textstreams which are linked to an internal graph structure. Navigation is possible in a presentation but not in single continuous media files. Furthermore, it is possible to define hotspots for navigation or to display additional information. With the usage of the elements and attributes from the timing modules, "time can be integrated into any XML language" [6, p. 117]. It is possible to define start and end time, duration, persistence, and repetition of objects and relations between those objects [6, p. 117]. The layout of a presentation is defined by the "relative placement of (multiple) media objects", but SMIL does not concern the internal formatting media of objects [6, p. 149]. SMIL is based on CMIF [5] and the AHM [10]. The final version of this standard is the SMIL 3.0 Recommendation which was published on December 01, 2008 [22]. Previous versions of this standard were SMIL 1.0 released in 1998, SMIL 2.0 released in 2001, and SMIL 2.1 released in 2005 [6]. SMIL 3.0 consists of 12 modules of elements and attributes (Animation, Metainformation, Content Control, Structure, Layout, Timing and Synchronization, Linking, Time Manipulations, Media Objects, Transition Effects, smilState and smilText) described as a DTD [6]. Furthermore, five profiles are

[1]For a more detailed description of the SIVA XML schema see [13], SMIL is described in [6] and [22].

built which use the enlisted elements and attributes, namely the SMIL 3.0 Language Profile, the SMIL 3.0 Unified Mobile Profile, the SMIL 3.0 DAISY Profile, the SMIL 3.0 Tiny Profile, and the SMIL 3.0 smilText Profile [22]. These profiles may limit the elements and attributes of the standard or extend it with functionality from other XML languages [6].

Extensions for SMIL can be found in the work of Cazenave et al. [8], Pihkala and Vuorimaa [17], and Vaisenberg et al. [21]. These works add a table of contents, a search function, and a bookmark function [21], "location information, tactile output, forms, telephoning, and scripting" [17], and the option to publish multimedia documents on the web using HTML5, CSS, and SMIL Timesheets [8] to SMIL. We do not consider these language extensions, because they are not part of the standard. In the following sections only elements and attributes from the SMIL 3.0 specification are used.

SIVA XML Schema

The SIVA XSD[2] was designed during the projects "Interaktives Video Editierungstool zum netzwerkbasierten Wissenstransfer (ivi-Pro)"[3] and "iVi-Pro 2.0 - Interaktives Video im Zeitalter von Mobilität und Kollaboration"[4][5]. Major design goals were an easy expandability and a slim format which exactly fitted our requirements as well as existing and potential future scenarios without too many limitations. We decided to implement some logic into the player to avoid repetitive definitions in the XML file. This allows the XML files to be more flexible, easy to read and adaptable to the requirements. Besides a main video, usable media files are images, audio files, videos, rich texts, and subtitles. These can be displayed as "global annotations" during a whole video, or as "local annotations" during a single scene. It is possible to define a non-linear structure of scenes, whereat decision elements provide selection panels or quizzes to viewers. Other navigational elements are a table of contents and a keyword search. Hotspots in the video trigger the display of additional information. The timing is kept local by adding annotations to single scenes. Thereby, absolute times in the scenes are used for displaying and hiding annotations. Thus synchronization is only necessary for a single scene and not for a whole video. The SIVA XML consists of six parts represented by six main elements below the root element: `<projectInformation>`, `<sceneList>`, `<resources>`, `<actions>`, `<tableOfContents>`, and `<index>`. These elements are linked by `ID/IDREF` attributes which are checked by constraints for their consistency. A more detailed description of the XML format can be found in [13]. In contrast to SMIL, the SIVA XML schema is not an official standard.

[2]The XSD file can be downloaded from **http://siva.uni-passau.de/sites/default/files/downloads/sivaPlayer.xsd** (accessed May 12, 2014)

[3]"Interactive video editing tool for network-based knowledge transfer (iVi-Pro)"

[4]"iVi-Pro 2.0 - Interactive video in the age of mobility and collaboration"

[5]The development of the SIVA XML schema was partially funded by the European Social Fonds and the Bayrisches Staatsministerium für Wissenschaft, Forschung und Kunst (Bavarian State Ministry for Sciences, Research and the Arts)

Comparison of the SIVA XML Schema and SMIL

We are aware that in general, the SIVA XML schema is a more specific, focused, and limited approach, while SMIL is more general, flexible, and not by default made to fit our use cases of interactive non-linear videos. Both languages do not cover the same aspects, and have different focuses and levels of detail.

We do not take other models and formats like NCM/NCL [7, 19, 20], CHM [18], ZYX [4], and HTML5 [23] into account, because they are either not standardised or have other areas of application. SMIL in contrast is standardised and considered as a format capable of describing hypervideos in the multimedia community.

FEASIBILITY ANALYSIS

We already stated what an interactive non-linear video is and what requirements need to be fulfilled in order to satisfy the demands of such the videos described and analysed in the REQUIREMENTS section. The following part shows how feasible an implementation of an interactive non-linear video is with both of the given XML languages SMIL and SIVA XML. Therefore we first present the feasibility of every requirement in regard of both languages. We also propose how extensive the implementation is and what features can or can not be realised. Therefore we will especially emphasize on disadvantages in the specified requirements, which show the main points that disallow a fully satisfying solution for interactive videos. On the other hand, these points also show starting points for better adaptiveness towards this use case. Afterwards we conclude our feasibility results in the Feasibility Conclusion section. Examples used in this section are adapted from [3].

Media, Main video, and Annotations

The entire presentation of an interactive video consists of a main video with the addition of annotations. Annotations are multimedia elements, that are supposed to enhance the interactive feeling and can be used to give more information about the topic of the video. Annotations can be triggered (invoked for display) by user interaction, established by a click on certain defined portions of the video, or by reaching a specified point of time. The placement is a fixed point or a path, resulting in a moving annotation.

Both XML languages support the full variety of media annotations, but differences are met in terms of the placement. When the editing of a **SMIL** presentation is finished, the placement of all its elements is set. This can result in overlapping annotations, for example pictures, when they are placed in the same area. The **SIVA XML** is usually interpreted by a player which supports an automated placing function, that will arrange the annotations next to each other. Another weakness of SMIL is found concerning the pathing. In order to achieve the exact demanded movement, the element of the annotation would require four `<animate>` for each step of the path.

Event-based Timing Model

The **SIVA XML** is fully designed to fulfill the requested timing model that an interactive non-linear video needs. Scenes are built modularly and do not have to be processed in a linear order as in SMIL. Annotations are started by defined triggers during a scene.

In contrast, **SMIL** makes use of an interval-based timing model. Although the whole functionality of an interactive non-linear video could be implemented, there are slight disadvantages with this model. Each relation between main video and annotation is bound together as a result of the SMIL element structure. It is not as modular and as local as in the SIVA XML.

Temporal Relationships between Main Video and Annotations

Temporal relationships can be implemented well in both languages. The modularity of the **SIVA XML** makes it possible that every temporal relationship can be modeled by local XML constructs which are linked by `ID/IDREF` attributes.

SMIL on the other hand supports a broad range of elements to satisfy the temporal needs of an interactive non-linear video. By making use of the basic temporal elements `<seq>` and `<par>` combined with more complex ones like `priorityClass` and their timing attributes `start` and `end`, each relationship between and inside the parts of an interactive video can be implemented.

Spatial Relationships between Main Video and Annotations

The **SIVA XML** shows advantages in spatial relationships compared to SMIL. All media and navigational elements of the interactive non-linear video can be placed specifically where they need to be, annotations can be arranged automatically, their paths and/or positions that are defined in the SIVA Producer will be fulfilled. If elements of a displayed panel (e.g. the table of contents) cannot be shown in its full size, the player can adapt to it by using techniques like scrollbars to supply the full range of accessibility for all elements.

SMIL on the contrary has some difficulties establishing these requirements. Each element has to be aligned exactly with its `left`-, `top`-, `right`-, `bottom`-, `height`-, and `width`-attributes in order to determine its position. If the given set, for example a list of links or buttons, is too large, it cannot be displayed entirely. The portion that is too large for the displaying area will be cut out. Furthermore, links can not be sized according to the width that their text needs, so consecutively you can activate a link by clicking into the "free" area that is residing to the right side of the link, what can also result in faulty activations of links. The counterpart here would be to size the buttons to a fixed length, but if a given text exceeds that boundary, the text will be cut off as well.

Decision Elements at Forks

Enhanced navigational features of the interactive non-linear video are forks. These are usually button panels and quizzes. At a button panel the viewer can pick one option that determines the continuation of the main video. In a quiz, a row of questions with multiple choice answers is posed. Each answer will give a certain amount of points. The achieved sum will determine the continuation after the quiz. For these functionalities, different elements like the button or answer panel are needed.

In the **SIVA XML schema**, all of this is supported entirely by defined complex types. The modular structure of the XML file makes the scenes accessible by triggers which are linked to the buttons of a choice panel. The quiz functionality specifies questions together with their answers. The correct answers are marked and obtainable points per question are set. Furthermore, point ranges for a whole quiz are defined for the selection of the following scene.

All of these features can be implemented in **SMIL**. Forks and their paths are supported by button panels, that suffer the spatial disadvantage already mentioned in the Spatial Relationships section. Although complex in the XML structure, the quiz functionality is also realizable. But a problem arises from the fact, that a path after a fork may be an edge to an already played scene. Jumps inside the SMIL file are in need of a link element. In order to implement this, the viewer is confronted with a panel that has to be clicked so that the video can continue at an earlier point. This results in a break of the flow of the video. For the common consumer of videos, this is an inconvenience. Problems like this are not present in the SIVA XML.

Table of Contents

The table of contents contributes to the non-linear character of the interactive video. When displayed (after clicking a button), a panel with links in a tree structure is presented to the user. By activating one of these links, the corresponding scene will be played. Here, the **SIVA XML** allows the addition of sub entries for each item of the panel. In **SMIL**, a table of contents is constituted by a list of clickable links, so for every entry, a link will be created and then arranged in a top down manner in the specified area. This element is also suffering the spatial problems mentioned in the Spatial Relationships section.

Keyword Reference List

A keyword search could not be established for the **SMIL** export as such functionality is not supported by the language. The **SIVA XML** supports all the requirements that are needed for a search. Users have the possibility to search for strings as keywords. Keywords are linked with scenes or annotations. When the user selects a keyword of a scene, the scene starts at its beginning. Selecting the keyword of an annotation, the video starts play-back at the point where the annotation is displayed. They can be searched while the interactive non-linear video is played.

Extensibility

Regarding the possibilities to extend the given model for new features, both XML languages are capable of integrating additional sets of elements. On both sides, the DTD or XSD files have to be adapted as well as the interpretation of the resulting exported XML document at the player. But in terms of the possibilities of changing an XML document of both languages that already exists, the **SIVA XML** has slight advantages. Due to the modular structure, it is easy to add new scenes, keywords, or annotations to an interactive video. This process is more complicated in **SMIL** based on the nested composition of the elements. Especially the temporal structure needs to be kept correct. Adding one single element may have impact on different parts of the interactive video, altering it in a way that may not be intended by an author or an insertion algorithm.

Feasibility Conclusion

Table 1 represents an overview of our analysis by listing each requirement and its feasibility for both languages. The feasibility is ranged on a scale from "*very bad or not at all*" (denoted as "- -"), "*partly feasible*" (denoted as "-") over "*neutral*" (denoted as "0") to "*feasible with some drawbacks*" (denoted as "+") or "*meets all requirements*" (denoted as "++").

Table 1. Feasibility of the Requirements

Requirement	SIVA	SMIL
Media, Main video, and Annotations	++	+
Event-based Timing Model	++	+
Temporal Relationships	++	++
Spatial Relationships	++	0
Decision Elements	++	+
Table of Contents	++	0
Keyword reference list	++	- -
Extensibility	++	++

The feasibility analysis shows, that the **SIVA XML** is very well adapted to the requirements of an interactive non-linear video. **SMIL** is able to realize many of the requirements or more precisely the needed features as well, but it lacks in certain details. As Table 1 demonstrates, in terms of temporal relationships or the extensibility, both languages are suited very well. The lack of a keyword feature sets SMIL back in that requirement, while facets like the problems in the spacial relationships force an inferior evaluation compared to the SIVA XML in other categories.

COMPARISON/METRICS

In order to have a closer look and a numerical comparison of the two metadata formats, we will make use of the following XML metrics: Size, Structure Complexity, Depth, Fan-In, and Fan-Out. For a detailed description see [11].

With these five values for an XML description format you are able to make statements about the complexity, comprehensibility, reusability, and convertibility of it. The higher one of the metric count is, the more complex is the possible resulting XML file. A high Fan-Out value makes it harder to alter a format because changes in single entities or elements may have an impact on multiple locations in the file. We have manually identified the count of the results for both, the SIVA XML and SMIL, to be able to compare them. For the first one, the SIVA XSD was converted into a DTD to be able to compare it with the SMIL DTD. Knowing that such a conversion will usually have an impact on the accurateness of the file, it does not have an impact on the evaluated metrics. The DTD for SMIL can be found online as well [1]. As it contains the elements for the whole language, we have created a profile to model a DTD that only supports the modules and elements needed in the export for interactive non-linear videos. SMIL was also contemplated in two different ways: with and without the `<metadata>` element, which has a great impact on the Fan-Out value. This is caused by the fact that the element can become a child node of each SMIL 3.0 element. The use of this element is the wrapping of structured meta information. It contains an own XML tree as content and is not processed by the player at all. As we do not make use of this element in our export, it is still contained in the generated DTD of our interactive video profile. Therefore we differentiate between an analysis of SMIL with and without the `<metadata>` element. Our results are presented in Table 2. Each number depicts the count for the specific XML metric.

Table 2. Comparison of SMIL and the SIVA XML (SC = Structure Complexity, ∞ = unbounded)

	Size	SC	Depth	Fan-In	Fan-Out
SIVA XML	58	67	5	12	8
SMIL (w/o meta)	40	430	∞	21	16
SMIL (with meta)	41	507	∞	22	38

Regarding the first two entries in Table 2, one can see that SMIL gets by with less elements than the SIVA XML, but its complexity is much higher. This is caused by the fact that many SMIL elements are used recursively. The high Fan-Out value is applicable for many of the occurring elements. The potential depth for SMIL is unbounded because the temporal elements `<par>` and `<seq>` can be boxed repeatedly. In some depth analyses, the recursion is ignored to not achieve a depth that is unbounded. We do not take this into consideration because in fact there can be an apparent endless potential depth by nesting forks. The Fan-In and Fan-Out metrics also state higher values in the SMIL-DTD and therefore indicate superior complexity.

PROPOSAL FOR EXTENSION

The extensibility of SMIL allows the addition of new elements that could be used to generate a module that is more adapted and able to provide support for interactive non-linear videos with regard to the aforementioned criteria. The ideas for these elements arose from the problems that were encountered while modeling the elements which are already available in the SIVA XML schema for interactive non-linear videos in SMIL. In combination with these elements, SMIL could achieve a more dynamic structure and be more feasible referring to interactive non-linear videos. Possible useful additions might be the following.

Jumps in the XML File

The elements `<goto>` and `<end>` may change the flow of the SMIL presentation by jumping to another position of the same document when reached. While the former cause a jump to a given `ID` supported by a `to` attribute, the `<end>` element is supposed to bring the presentation to an end. They can be used like most of the timing elements in SMIL in terms of nesting as well as their attributes. Caution has to be paid, because loops and abrupt endings of the presentation can be built very easily.

Choices at Forks

To satisfy the requirement for fork elements more easily, we propose the introduction of two new elements that could facilitate the implementation of a fork: <fork> and <choice>. When a <fork> element is started, it composes a standard choice panel (that could be altered by its attributes in terms of shape etc.) which contains buttons to start one of its <choice> children nodes. One of these <choice> elements contains elements that are supposed to be played once it is activated. In combination with the preceding elements <jump> or <end>, different continuations after a path can be established as well. If this is not supported, the <choice> as well as its surrounding <fork> element will be ended and the succeeding element will be started, just like in a normal SMIL presentation. Listing 1 shows an example for the use of <fork> and <choice> elements combined with the above mentioned jump elements <goto> and <end>. The fork ranges from line 4 to 23 and contains three different choices. By selecting one of them, the code inside the corresponding <choice> element will be started. The button panel itself is designed by the attributes of the <fork> element to show round buttons with a size of 20 pixels. The panel will be displayed for 30 seconds. If no path is chosen in that timespan, according to the defaultPath attribute, the defined default path (in this example *path1*) will be played. The second path choice from line 11 to 16 contains a <goto> element in line 12 inside a sequential container with the effect, that after the other elements defined in line 13, an automated jump from line 14 to line 2 (according to the given to attribute). The third path in lines 17 to 22 consists of a parallel node containing some code and then an <end> element in line 20, which also owns a begin attribute with the value 20s. This structure of elements induces the behaviour, that no matter what the content of the <par> element is, the <end> element will be started after twenty seconds, causing the presentation to terminate.

```
1  <smil xmlns="http://www.w3.org/ns/SMIL">
2  <head>
3    <!-- Any SMIL head content -->
4  </head>
5  <body>
6    <seq xml:id="start">
7      <!-- Any SMIL content -->
8      <fork shape="circle" size="20"
9        region="main_region"
10       dur="30s" defaultPath="path1"
11       xml:id="fork">
12       <choice xml:id="path1" after="#fork">
13         <!-- Any SMIL content -->
14       </choice>
15       <choice xml:id="path2">
16         <seq>
17           <!-- Any SMIL content -->
18           <goto to="#start"/>
19         </seq>
20       </choice>
21       <choice xml:id="path3">
22         <par>
23           <!-- Any SMIL content -->
24           <end begin="20s"/>
25         </par>
26       </choice>
27     </fork>
28     <!-- Any SMIL content -->
29   </seq>
30 </body>
```

Listing 1. Body excerpt of a SMIL file showing sample code for a fork and jumping elements

CONCLUSION

In summary, the SIVA XML shows advantages regarding to the usefulness for interactive non-linear videos. As stated in section FEASIBILITY, all of the requirements that are needed to fully implement an interactive non-linear video are met by the language. SMIL can realize most of the functionality (keywords could not be established natively), but it lacks in some details like spatial problems of decision boxes, the placement of subtitles or moving annotations. A further benefit of the SIVA XML is revealed by the analysis of the underlying DTDs of both formats. SMIL is composed by a much more complex set of entities which makes the construction and the understanding of the resulting SMIL documents harder. The temporal elements in particular hamper the process of modifying a given SMIL presentation, which was stated in the Extensibility section. Many parallel, sequential, and conditional elements are stacked and interwoven, so that the addition of a simple annotation can be a very complex venture. These points state, that the SIVA XML is better suited for interactive non-linear videos. But an important detail to mention is that SMIL is not meant or designed particularly to support interactive non-linear videos. The research we did here was based on the standard SMIL 3.0 model.

A new set of extensions would make SMIL more usable for interactive non-linear videos. Therefore, SMIL could be extended by different complex structures like a decision fork, a textual link, and an element which allows to jump inside the SMIL file (without the activation of a link). In combination with these elements, SMIL could achieve a more dynamical structure and be more feasible referring to interactive non-linear videos.

The SIVA Suite[6] contains a production software (SIVA Producer) that enables the user to design and create interactive videos. As this work shows that SMIL is well suited in order to be used as metadata format for interactive non-linear videos, we implemented a SMIL exporter into the SIVA Producer. Now it is possible to export the designed interactive video into a SMIL presentation (with the limitations described in this work).

REFERENCES

1. SMIL 3.0 DTDs. Website, December 2008. http://www.w3.org/TR/smil/smil-DTD.html (accessed March 25, 2013).
2. Ambulant open smil player. Website, June 2010. http://www.ambulantplayer.org/ (accessed May 12, 2014).
3. Berndl, E. Siva-xml-schema vs. smil - strukturmapping und exportimplementierung. Mastersthesis, University

[6]The SIVA Suite can be downloaded from: **http://www.siva.uni-passau.de/?q=node/66** (accessed May 12, 2014)

of Passau, Chair of Distributed Information Systems, Passau, Germany, October 2012.

4. Boll, S., and Klas, W. Zyx - a multimedia document model for reuse and adaptation of multimedia content. *IEEE Transactions on Knowledge and Data Engineering 13*, 3 (2001), 361–382.

5. Bulterman, D., Rossum, G. V., and Liere, R. V. A structure for transportable, dynamic multimedia documents. In *In Proceedings of the Summer 1991 USENIX Conference* (1991), 137–155.

6. Bulterman, D., and Rutledge, L. *SMIL 3.0: Flexible Multimedia for Web, Mobile Devices and Daisy Talking Books*. X. media. publishing Series. Springer, 2008.

7. Casanova, M. A., Tucherman, L., Lima, M. J. D., Rangel Netto, J. L., Rodriquez, N., and Soares, L. F. G. The nested context model for hyperdocuments. In *Proceedings of the 3rd Annual ACM Conference on Hypertext. HYPERTEXT '91*, no. December, ACM (New York, NY, USA, 1991), 193–201.

8. Cazenave, F., Quint, V., and Roisin, C. Timesheets.js: when smil meets html5 and css3. In *Proceedings of the 11th ACM symposium on Document engineering*, DocEng '11, ACM (New York, NY, USA, 2011), 43–52.

9. Hammoud, R. *Interactive video: algorithms and technologies*. Signals and communication technology. Springer, 2006, ch. Introduction to interactive video, 3–25.

10. 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 (Feb. 1994), 50–62.

11. Klettke, M., Schneider, L., and Heuer, A. Metrics for xml document collections. In *XML-Based Data Management and Multimedia Engineering - EDBT 2002 Workshops*, vol. 2490 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2002, 15–28.

12. Meixner, B., Hölbling, G., Stegmaier, F., Kosch, H., Lehner, F., Schmettow, M., and Siegel, B. Siva producer - a modular authoring system for interactive videos. In *Proceedings of I-Know, 9th International Conference on Knowledge Management and Knowledge Technologies* (Graz, Austria, 2009), 215–226.

13. Meixner, B., and Kosch, H. Interactive non-linear video: definition and xml structure. In *Proceedings of the 2012 ACM symposium on Document engineering*, DocEng '12, ACM (New York, NY, USA, 2012), 49–58.

14. Meixner, B., Matusik, K., Grill, C., and Kosch, H. Towards an easy to use authoring tool for interactive non-linear video. *Multimedia Tools and Applications* (2012), 1–26.

15. Meixner, B., Siegel, B., Hölbling, G., Lehner, F., and Kosch, H. Siva suite: authoring system and player for interactive non-linear videos. In *Proceedings of the international conference on Multimedia*, MM '10, ACM (New York, NY, USA, 2010), 1563–1566.

16. Nimmagadda, Y., Kumar, K., and Lu, Y.-H. Preference-based adaptation of multimedia presentations for different display sizes. In *IEEE International Conference on Multimedia and Expo, 2009. ICME 2009.* (2009), 978–981.

17. Pihkala, K., and Vuorimaa, P. Nine methods to extend smil for multimedia applications. *Multimedia Tools and Applications 28* (2006), 51–67.

18. Sadallah, M., Aubert, O., and Prié, Y. Component-based hypervideo model: high-level operational specification of hypervideos. In *Proceedings of the 11th ACM Symposium on Document Engineering. DocEng '11*, ACM (New York, NY, USA, 2011), 53–56.

19. Soares, L. F. G., and Rodrigues, R. F. Nested context model 3.0. part 1 - ncm core. 18, Laboratorio TeleMidia DI - PUC-Rio, Rio de Janeiro, 2005. Technical Report. Website (accessed April 26, 2014).

20. TeleMidia Lab - PUC-Rio. Ncl - nested context language. Standard/Website, 2011. http://www.ncl.org.br/en/inicio (accessed Mai 12, 2014).

21. Vaisenberg, R., Jain, R., and Mehrotra, S. Smpl, a specification based framework for the semantic structure, annotation and control of smil documents. In *11th IEEE International Symposium on Multimedia, 2009. ISM '09.* (2009), 533–539.

22. W3C. Synchronized multimedia. Website, March 2012. http://www.w3.org/AudioVideo/ (accessed Mai 12, 2014).

23. W3C. Html5 - a vocabulary and associated apis for html and xhtml (w3c candidate recommendation 6 august 2013). Website, 2013. http://www.w3.org/TR/html5/Overview.html (accessed Mai 12, 2014).