

Adaptation of Scalable Multimedia Documents

Benoît Pellan and Cyril Concolato

TELECOM ParisTech, CNRS LTCI

TSI Department, Multimedia Group

Paris, 75634 Cedex 13, France

{Benoit.Pellan, Cyril.Concolato}@telecom-paristech.fr

ABSTRACT

Several scalable media codecs have been standardized in recent years to cope with heterogeneous usage conditions and to aim at always providing audio, video and image content in the best possible quality. Today, interactive multimedia presentations are becoming accessible on handheld terminals and face the same adaptation challenges as the media elements they present: quite diversified screen, memory and processing power capabilities. In this paper, we address the adaptation of multimedia documents by applying the concept of scalability to their presentation.

The *Scalable MSTI* document model introduced in this paper has been designed with two main requirements in mind. First, the adaptation process must be simple to execute because it may be performed on limited terminals in broadcast scenarios. Second, the adaptation process must be simple to describe so that authored adaptation directives can be transported along with the document with a limited bandwidth overhead. The *Scalable MSTI* model achieves both objectives by specifying *Spatial*, *Temporal* and *Interactive* scalability axes on which incremental authoring can be performed to create progressive presentation layers.

Our experiments are conducted on scalable multimedia documents designed for Digital Radio services on DMB channels using MPEG-4 BIFS and also for web services using XHTML, SVG, SMIL and Flash. A scalable image gallery is described throughout this article and illustrates the features offered by our document model in a rich multimedia example.

Categories and Subject Descriptors

I.7.2 [Document and Text Processing]: Document Preparation – Languages and systems, Multi/mixed media; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems.

General Terms

Algorithms, Design, Languages.

Keywords

Document model, document adaptation, multimedia scalability.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

DocEng'08, September 16-19, 2008, Sao Paulo, Brazil.

Copyright 2008 ACM 978-1-60558-081-4/08/09...\$5.00.

1. INTRODUCTION

The ongoing research in the audiovisual field has become mature enough to provide ubiquitous multimedia services [1]. Indeed, the same multimedia document can be retrieved anywhere by using different transport technologies (DVB, DMB, Wi-Fi or 3G) and viewed at anytime on appropriate multimedia terminals: HD displays (set-top boxes), media centers (PCs), portable multimedia devices (PMP) or handheld terminals (mobile phones). The diversity of multimedia receivers raises severe issues as far as content production is concerned. Indeed, multimedia services must be created to cope with the lowest handset capabilities while providing enhanced visual quality to advanced multimedia terminals. So far, much research has been done on the adaptation of traditional media (video, images and audio). For instance, video adaptation to terminal constraints can be achieved by using Fine-Granular Scalability in MPEG-4 [2] and efficient video stream switching can be performed using specially-coded slices (SP/SI) in MPEG-4 AVC. The MPEG-4 SVC standard [3] is even tackling the core of this problem by focusing on the scalability features of efficient video coding by providing spatial, temporal and quality enhancement layers.

Multimedia documents are composed of numerous media elements (audio sequences, video clips, images, graphics, text...) and the adaptation of multimedia documents cannot be performed by only adapting individual media since the semantics of the whole presentation must be preserved [4]. In this paper, we propose addressing the challenges of document adaptation by applying the concept of scalability to multimedia presentations. This paper presents our document model in two progressive phases: first the *MSTI* model and then the *Scalable MSTI* model.

The *MSTI* model (*Media Spatial Temporal* and *Interactive*) is based on state-of-the-art multimedia document models [5][6][7] that clearly separate media elements and logical structures from the multimedia presentation. Our *MSTI* model specifically extends the XML-based multimedia-processing model described in [5] by defining a multimedia document as a logical structure referencing several media elements on which additional transformations, similar to *spatialSheet* or *temporalSheet*, can be applied for presentation. Hence, we do not propose full specifications of the spatial, temporal and interactive properties of multimedia scenes like Madeus [5] or ZyX [6] but we choose to rely on standardized scene formats such as XHTML, SVG [8], SMIL [9] or MPEG-4 XMT [10]. The *Spatial*, *Temporal* and *Interactive* descriptions of the *MSTI* model reference elements in the *Media* description to simply apply their presentation properties. In this paper, the advantages of the *MSTI* model are shown on a choice-based adaptation scheme and illustrated in an interactive and timed slideshow of an SVG image gallery.

The *Scalable MSTI* model proposed in this paper extends the *MSTI* model by leveraging its *Spatial*, *Temporal* and *Interactive* components to provide efficient and author-friendly multimedia document adaptation functionalities. The *Scalable MSTI* model further divides *Spatial*, *Temporal* and *Interactive* descriptions into progressive layers in order to provide multimedia document scalability. Our model is author-friendly since the progressive layers of *MSTI* scalable documents follow a natural incremental authoring process when producing adaptable services [11]. It is also efficient since it reduces the redundancies between alternative documents by providing a common description framework. However, the authoring of scalable multimedia documents presents several limitations since some adaptation scenarios may not be compatible with the progressive nature of multimedia scalability. The *Scalable MSTI* model addresses these limitations through the definition of layer parameters that can be used to define flexible adaptation paths.

The remainder of this paper is organized as follows. In Section 2, we describe the *MSTI* model and syntax. We also illustrate its usage in adaptation scenarios in an image gallery example. In Section 3, we define the principles of multimedia scalability and describe the generic adaptation properties that the *Scalable MSTI* model offers. Finally, Section 4 concludes this paper and introduces perspectives for future work.

2. THE MSTI DOCUMENT MODEL

The principle of the *MSTI* document model is to decompose a multimedia document into several components so that a simple adaptation process can be applied to each of them. The *MSTI* model is based on state-of-the-art approaches: it relies on standardized scene description formats and defines a simple syntax to apply presentation properties to multimedia documents. An adaptation framework based on the *MSTI* model is given in Section 2.3 and illustrates how the selection of *Spatial*, *Temporal* and/or *Interactive* descriptions can be used to generate documents adapted to the specific characteristics of a multimedia terminal. Additionally, adaptation alternatives for an *MSTI* image gallery are described in Section 2.4.

2.1 Multimedia document adaptation

The adaptation of multimedia documents is a complex task because many scene properties rely on multiple technical capabilities of a terminal. We can illustrate this issue with the following three examples. First, animations sometimes require significant processing power when moving media elements of a multimedia scene for instance. But, animations also impose a screen size big enough for animations to be noticed. Second, the number of visible media elements may greatly impact the memory consumed by the scene but also require a minimal resolution to actually display all these items. Clickable interactive areas (e.g. buttons) are a final example. They clearly depend on the availability of pointing devices on the terminals but also require that the size of sensitive areas is sufficient to be accessible by the user. Therefore, it is difficult to associate properties of multimedia documents with the characteristics of multimedia devices since there is no one-to-one mapping as illustrated in Figure 1.

Moreover, multimedia documents can be accessed by a large variety of multimedia terminals. Therefore, even if document alternatives can be dynamically generated at publication time to cope with some typical classes of devices (HD TV, PMP, mobile phones...) as in the *MM4U* framework [12], it is difficult for an adaptation engine to compute an appropriate presentation for every configuration. For

instance, a mobile phone usually has a small screen, limited memory, limited processing power and keypad navigation but some of them possibly feature a touch screen or a graphics accelerator for which the appropriate presentation design may be very specific. One option consists in focusing on some specific characteristics of multimedia receivers (e.g. screen size and aspect ratio) to provide efficient and adaptive capabilities through designed constraints [13]. We try to address this challenge in a more generic way by taking into account the full range of scene properties.

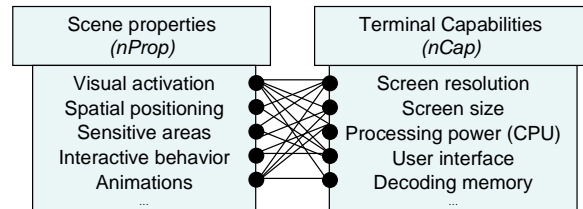


Figure 1. Mapping scene properties to terminal capabilities

The *MSTI* approach proposes transforming the complex document adaptation problem, which consists in matching the *nProp* multimedia properties with the *nCap* terminal capabilities, into two simpler operations. First, we focus on grouping presentation properties into three different components (*Spatial*, *Temporal* and *Interactive* descriptions) and designing alternative descriptions relevant to the document semantics during the authoring phase. Then, the spatial, temporal and interactive properties are analyzed and matched with the *nCap* terminal capabilities. In our approach, service providers may have little knowledge (if any) about the capabilities of receiving terminals when generating a multimedia document and should only worry about designing alternatives that are semantically significant for typical application scenarios. As a consequence, the *MSTI* document model aims at structuring and guiding the adaptation of multimedia documents by assembling possible presentation properties according to author's intentions. This approach simplifies the mapping of scene properties to terminal capabilities but does not claim to provide a solution to all adaptation scenarios since it is based on pre-designed alternatives. Some of these limitations are overcome by the *Scalable MSTI* model that extends the adaptation capabilities of the *MSTI* model through fine-grained scalability. However, it is clear that the *MSTI* model contributes to the modification of document engineering lifecycle from post-production document analysis and transformation towards adaptation-oriented multimedia authoring as initiated in [13][14][15].

2.2 MSTI model and syntax

The *MSTI* document model does not provide primitives to represent all aspects of a multimedia document as is the case in *Madeus* [5] and in *ZyX* [6], nor does it propose extensions to existing presentation formats such as adding *one-way constraints* to SVG [13]. Instead, it takes advantage of existing multimedia document formats to structure their functionalities for adaptation purposes. Therefore, the *MSTI* document model does not try to tackle multimedia adaptation complexity through a new authoring paradigm [16] but rather relies on functionalities of existing multimedia standards and on state-of-the-art models "to provide comprehensive support for authoring needs".

The underlying principle of the *MSTI* model is the use of style-agnostic media structures that are transformed by additional descriptions to generate final presentations. This approach is

inspired from common practices: e.g. by the layout of web pages using W3C Cascading Style Sheets (CSS); or by the use of SMIL Time Sheets [17] to extend the temporal properties of a document. In our approach, three different descriptions can be associated with the style-agnostic *Media* description to form the presentation of the multimedia document: *Spatial*, *Temporal* and *Interactive* descriptions.

2.2.1 Media description

The *Media* description references all the media elements that compose the multimedia document: not only audio sequences, video clips, or images, but also text, graphic elements... It also includes structures and application logic that define relationships between media elements. For instance, if some media elements have been included in the same document to be shown as a slideshow, the display order and media priorities should be described in the *Media* description since they do not depend on presentation choices (e.g. timed and/or interactive slideshow). An example of *Media* description (using SVG and JavaScript) is given in Figure 2 for an image gallery composed of 6 images. Note that we use JavaScript code in our SVG example because our slideshow requires it, but our approach does not mandate the use of JavaScript code.

```

<svg id="root" version="1.1">
  image0 = document.getElementById("img_0");
  ...
  image5 = document.getElementById("img_5");
  rect0 = document.getElementById("rect_0");
  ...
  rect5 = document.getElementById("rect_5");
  slideshow_token = 0;

  function activateNextSlide(){
    slideshow_token = (slideshow_token + 1) % 6;
    updateSlideshow(slideshow_token);
  }
  function updateSlideshow(token) {
    switch (token) {
      case 0: <!-- Apply presentation for image0 --> break;
      ...
      case 5: <!-- Apply presentation for image5 --> break;
    }
  }
  <g>
    <rect id="rect_0" visibility="hidden"/>
    <image id="img_0" visibility="hidden" xlink:href="pic0.png"/>
    ...
    <rect id="rect_5" visibility="hidden"/>
    <image id="img_5" visibility="hidden" xlink:href="pic5.png"/>
  </g>
</svg>

```

Figure 2. Media description of an image gallery in SVG

The *Media* description of the *MSTI* model could be compared to the *Logical* model of *Madeus* [5] with two main differences. First, the language used to define the *Media* description is the same as the final presentation language. Therefore, no specific syntax needs to be defined at the *Media* level. Second, the *Media* description does not contain the intrinsic properties of media elements (i.e. duration, size) but rather provides information related to the context of use (e.g. image gallery). The *Media* description of the *MSTI* model can also be seen as a multimedia document that does not include any presentation means.

Compared to the work of S. Boll [6], the *MSTI* model does not comply with the given “presentation-neutral” definition because in our approach the multimedia document is expressed in the

“presentation-specific format used for playout of the multimedia materials.” As a consequence, our approach cannot be used to enable publication of content in multiple formats. However, the *MSTI* model succeeds in completely separating the multimedia presentation properties from the semantics of the multimedia document without requiring a publication phase, in which an internal document model needs to be converted into a standardized multimedia format. This is an important choice when authoring Rich-Media documents because “an internal document model that abstracts from the different characteristics of today’s multimedia presentation formats and, hence, forms the greatest common denominator of these formats” [12] greatly reduces the available functionalities and sometimes proves unsatisfactory from an editorial point of view. The *MSTI* approach can take advantage of all the richness, including specific low-level features, of any of the multimedia standards such as XHTML, SVG [8], SMIL [9] or MPEG-4 BIFS [10].

2.2.2 Presentation of a Media description

To present a multimedia document, several media elements must be styled and composed spatially, temporally and interactively on the user’s display. The composition properties of a presentation are defined in *Spatial*, *Temporal* and *Interactive* descriptions in the *MSTI* model. The style of a multimedia presentation is defined by assigning decorative properties to its media elements (e.g. color, border width or transparency values of graphics elements). Decorative properties do not relate to the composition of media elements and are not specifically targeted by the *MSTI* model. However, they constitute presentation properties that can be defined as part of the spatial layout of a multimedia document (*Spatial*). They can also be applied automatically during document lifetime (*Temporal*) or by user action (*Interactive*). As a consequence, the *Spatial*, *Temporal* and *Interactive* descriptions of the *MSTI* model are closely related to a *Media* description.

In order to apply a presentation to a multimedia document, several requirements must be taken into account. First, the presentation transformation may happen on the receiving terminal (e.g. in broadcast or multicast mode) and this terminal may have limited processing power. Hence, our first requirement is to have a mechanism which can be implemented efficiently. Then, applying presentations may require modifying the structure of the multimedia document. Our second requirement is therefore to have a flexible mechanism to handle multimedia scene trees. Several options have been considered that did not fulfil our requirements. First, Cascading Style Sheets (CSS), which are widely used to apply styles to documents. However, they do not allow the modification of the structure of an XML document. Second, the Extensible Stylesheet Language (XSL), which is quite powerful for modifying the structure of a document, but which we deemed too complex as a presentation description for our needs. Therefore, the *MSTI* model proposes a straightforward syntax, inspired by MPEG-4 BIFS updates, that simply allows inserting new XML elements with a defined identifier (*id*) into an existing document (*insert*) and to replace existing XML elements or attributes (*replace*) through the use of reference identifiers (*ref_id*). This simple syntax is used by the *MSTI* model to describe *Spatial*, *Temporal* and *Interactive* properties. The *MSTI* model currently relies on the identifier mechanism of existing multimedia standards to link presentation properties with media parameters. However, XML identifiers [18] or XPath expressions [19] could be used for that purpose.

The transformation that consists in applying the *Spatial*, *Temporal* and *Interactive* descriptions to a *Media* description to create

presentable multimedia documents is called the *STI* composition. According to application scenarios, the *STI* composition can be either done directly (using DOM APIs), or each description can be translated into W3C Remote Event XML (REX) [20], or transformed into an XSL document in order to actually perform the document transformation with an XSLT engine. It should be noted that the *STI* composition does not rely on any ordering of *STI* transformations since these components are designed to be independent from each other as explained in Section 2.2.6.

2.2.3 Spatial description

The *Spatial* description of the *MSTI* model defines the layout for all the media elements that are part of the document. This spatial description relies entirely on the spatial model of the targeted presentation format. It includes the position of the media elements, the media size, and the visual activation of elements. An example of a *Spatial* description is provided in Figure 3 and can be applied to the *Media* description illustrated in Figure 2 to create a one-image presentation.

```
<spatial>
  <replace ref_id="root" attribute="viewBox">0 0 250 200</replace>
  <replace ref_id="img_0" attribute="x">45</replace>
  <replace ref_id="img_0" attribute="y">20</replace>
  <replace ref_id="img_0" attribute="width">160</replace>
  <replace ref_id="img_0" attribute="height">160</replace>
  <replace ref_id="img_0" attribute="visibility">visible</replace>
</spatial>
```

Figure 3. Spatial description of a one-image gallery in SVG

2.2.4 Temporal description

The *Temporal* description of the *MSTI* model defines the timing of the multimedia presentation as a whole but also the timing of each individual media element. The *Temporal* description only exploits the functionalities of the targeted presentation format. Such a format usually includes local timer definitions, sequential or parallel behavior and the configuration of timed animators. An example of a *Temporal* description is provided in Figure 4 and can be applied to the *Media* description illustrated in Figure 2 to create a timed image slideshow.

```
<temporal>
  <insert ref_id="root">
    <script id="script_timedUpdate">
      function timedSlideshowUpdate() {
        activateNextSlide();
        setTimeout("timedSlideshowUpdate()", 5000); }
    </script>
  </insert>
  <insert ref_id="root"> timedSlideshowUpdate(); </insert>
</temporal>
```

Figure 4. Temporal description of a timed slideshow in SVG

The *MSTI* model basically assumes that the structure of *Media* descriptions is only logical or usually groups media elements to define common presentation parameters such as position, visibility or sensitive areas. However, some multimedia documents in formats such as SMIL [9] may be structured on a temporal basis (*seq/par* elements). In such cases, these temporal structures are defined in the *Media* description and timing parameters (*dur*) are defined in the *Temporal* description.

2.2.5 Interactive description

The *Interactive* description of the *MSTI* model adds interactive aspects to the multimedia document and defines the behavior associated with interactions. It may specify navigation schemes but also keyboard events and mouse actions such as *isOver* and *onClick* events. An example of an *Interactive* description is provided in Figure 5 and can be applied to the *Media* description illustrated in Figure 2 to create a shadow when moving the mouse over an image of the slideshow.

```
<interactive>
  <replace ref_id="img_0" attribute="onmousemove">
    mouseOverImage0(1);
  </replace>
  <replace ref_id="img_0" attribute="onmouseout">
    mouseOverImage0(0);
  </replace>
  <insert ref_id="root">
    <script id="script_is_over_img_0">
      function mouseOverImage0 (isOver) {
        if ( isOver ) { rect0.setAttribute("visibility", "visible"); }
        else { rect0.setAttribute("visibility", "hidden"); }
      }
    </script>
  </insert>
</interactive>
```

Figure 5. Description of an interactive slideshow in SVG

2.2.6 Relationships between STI components

The *MSTI* model relies on a complete separation of multimedia presentations into three different components: *Spatial*, *Temporal* and *Interactive (STI)* descriptions. However, these components might be closely related in some cases. For instance, the animation of a media element requires a timer (*Temporal*) that will drive the movements of the media along a path (*Spatial*). Similarly, an interactive anchor applied to the surface of a media element (*Interactive*) requires that this media element is correctly positioned in the presentation (*Spatial*). Of course, the links between *STI* components can be handled through the definition of specific elements that combine them as is the case of the *S-Relation* element in the *Madeus* model [5] for “spatial animations”. However, this approach is not extensible to all *STI* dependencies and standardized presentation formats sometimes rely on low-level generic tools to cope with such authoring cases (e.g. absence of a specific animation node in MPEG-4 XMT).

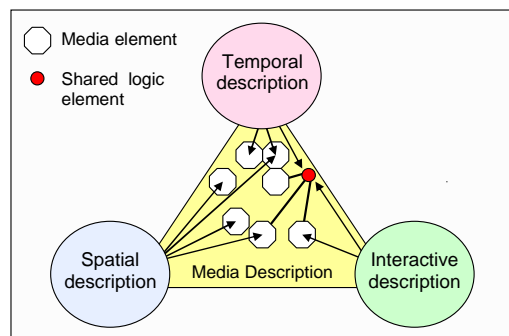


Figure 6. Multimedia document in the MSTI model

In the *MSTI* model, all dependent presentation properties are separated into *STI* components by referencing common media elements or logical structures in the *Media* description. Hence, *STI* components do not depend upon each other but are all linked to the same *Media* description. For instance, an interactive button may

animate slightly by changing its size when the mouse is over it. In that case, the shape of the ‘button’ is part of the *Media* description; the position, size and extended size of this shape is specified in the *Spatial* description; the duration and a method to trigger the animation of this shape is defined in the *Temporal* description and finally the interactive behavior triggered by the mouse is associated to the ‘button’ in the *Interactive* description. Furthermore, complex *STI* dependencies can also be handled through the definition of shared presentation elements such as variables, presentation states or methods. As depicted in Figure 6, *STI* descriptions can be closely related through references to the same elements in the *Media* description but the *STI* descriptions never directly depend on each other.

The *MSTI* model relies greatly on the capabilities of targeted formats to decompose their primitives into separated *STI* properties. However, some presentation fragments, such as the SVG *animate* element, cannot be decomposed. In such a case, the content can be reorganized as follows: the SVG *animate* element is part of the *Media* description, *begin*, *dur* and *fill* attributes are part of the *Temporal* description while *from* and *to* attributes are defined in the *Spatial* description. This kind of hybrid element can be found in existing scene description formats and define the limits of our approach since the *Media* description may contain presentation-oriented elements. As a consequence, the complete separation of document structures and semantics from presentation properties as explained in Section 2.2.2 remains a general authoring principle that may suffer from some exceptions depending on the targeted scene description format.

2.3 The *MSTI* model in the adaptation context

The *MSTI* model can be used to produce multimedia documents adapted to terminal capabilities. The adaptation process described in Figure 7 illustrates how *Spatial*, *Temporal* and *Interactive* (*STI*) descriptions can be selected for adaptation purposes using separate algorithms depending on the usage context. This decision-making can be done client-side or server-side depending on application scenarios. For instance, a client-side algorithm could use CSS Media Queries [21]. In practice, *STI* component selection can be guided through metadata associated to *STI* descriptions such as CC/PP [22].

The adaptation process is performed by the *STI* composition and consists in applying pre-authored *STI* components to a *Media* description to generate an adapted presentation. A server-side adaptation process could be based on XSL transformations whereas client-side processing could be performed by Remote Events for XML [20]. The *MSTI* model does not mandate any particular technology as long as the *insert* and *replace* functionalities described in Section 2.2.2 can be expressed.

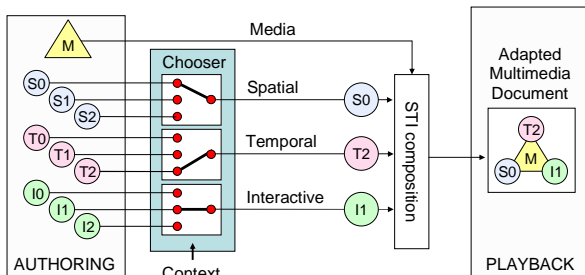


Figure 7. Choice-based *MSTI* adaptation framework

As shown in Figure 7, the *Media* description can, but does not have to, be adapted along with its presentation. In fact, the document adaptation is completely driven by *STI* descriptions and selected *STI* components will be applied to the *Media* description to create the adapted multimedia document. As a consequence, a server-side adaptation engine might implement a media filter mechanism to optimize the transmitting time of multimedia documents by not sending the unused media elements of the adapted presentation.

The decision-making process (*Chooser*) illustrated in Figure 7 consists in individually selecting a *Spatial*, a *Temporal* and an *Interactive* description based on the environment context. This approach is flexible, since it allows n descriptions per *STI* component leading to $n_s \times n_t \times n_i$ presentation alternatives for the same multimedia document, and offers a wide range of adaptation scenarios. For instance, a well-equipped web browser on a PC will select a *Spatial* description corresponding to its large screen size and resolution, an advanced *Temporal* description with appealing animations and a complete *Interactive* description for *STI* composition (S_2, T_2, I_2). A PDA has a smaller display (S_1) and cannot handle complex animations as smoothly as a PC (T_1). Additionally, a PDA usually has a sensitive screen but no visual pointer and a limited keypad (I_1). Therefore, such a PDA would select (S_1, T_1, I_1) authored descriptions to compose a multimedia document adapted to its constraints. Similarly, an advanced smart phone has a small display (S_0) but possibly features significant processing power and full interactivity (tiny keyboard, sensitive screen, and track ball). Therefore, this smart phone would compose a (S_0, T_1, I_2) presentation whereas a simpler mobile phone may select a (S_0, T_0, I_0) configuration with only a few animations and simple touchpad navigation.

2.4 Example of an *MSTI* image gallery

The example described in this section is an image gallery that can be presented as an advanced image slideshow with the *MSTI* syntax (M, S_i, T_i, I_i). This image gallery has been described using SVG and JavaScript and some parts of its code are provided in Figures 2, 3, 4 and 5. A simplistic presentation (S_0, T_0, I_0) has also been created in order to have a clearer view of individual *STI* components of the *MSTI* model. Possible adaptation configurations are illustrated in Table 1.

The *Media* description is composed of 6 images, 6 rectangles that create a ‘shadow’ effect behind the images and 4 ‘arrows’ that can be enabled for forward, fast-forward, backward and fast-backward navigation. Two variables are also defined in the *Media*: *slideshowTempLock* and *slideshowPermLock* which can be used to temporarily or indefinitely lock the slideshow when both the *Temporal* and *Interactive* components are jointly applied to the image gallery.

Table 1. *MSTI* alternatives for an image gallery

<i>STI</i>	Description
$S_0 \ T_0 \ I_0$	S_0 provides the position for a single image: the first of the gallery as described in Figure 3. T_0 and I_0 are empty descriptions and do not grant the user access to other images.
$S_1 \ T_0 \ I_0$	S_1 provides the position for 3 images that are aligned horizontally. This layout can be activated when a higher resolution than S_0 is available and if the capabilities of the terminal are sufficient (e.g. display size, decoding memory...). This document configuration grants the user access to other images.

S_0	T_1	I_0	T_1 animates a one-pane image slideshow (S_0) on a periodic time basis of 5 seconds. Transitions from one slide to the next are performed by sliding the current image to the left and inserting the next from the right. The user can therefore have access to the slideshow without any interactivity but this multimedia document requires the support of a timing function and sufficient processing power to move the images.
S_0	T_0	I_1	I_1 enables the four graphic ‘arrows’ of the <i>Media</i> description that have been positioned in S_0 . I_1 appends an event listener to these four “button” shapes and describes the mechanisms which allow the user to replace the current image by the next (or the previous) one according to interactions on an available user interface (pointing device, keypad). A step of 3 images is defined for fast-forward and fast-backward buttons. Furthermore, I_1 also triggers a ‘shadow’ effect when focusing on an image with a pointing device (e.g. mouse).
S_1	T_1	I_0	The combination of S_1 and T_1 generates a 3-pane image slideshow. The spatial activation of two additional images that are not visible in S_0 enables animations that are part of T_1 but that were not executed in the (S_0, T_1, I_0) configuration. The transition between slides is performed by removing the last image on the left, sliding all the remaining images to the left and inserting a new image on the right.
S_1	T_0	I_1	The combination of S_1 and I_1 generates a 3-pane interactive image gallery. S_1 provides a <i>Spatial</i> description of the slideshow for every state illustrated in Figure 2 and I_1 only triggers the slideshow state corresponding to a user’s requests.
S_0	T_1	I_1	The combination of T_1 and I_1 illustrates the flexibility of the <i>MSTI</i> model and achieves a 1-pane (S_0) interactive slideshow. In practice, the coexistence of the automated navigation paradigm and the user-centric navigation is managed by sharing common variables in the <i>Media</i> description. For instance, if the user is interacting with the slideshow, I_1 enables the <i>slideshowTempLock</i> variable of the <i>Media</i> description that will disable the slideshow scheduler defined in T_1 . Indeed, T_1 checks this variable before automatically animating images and releases this lock after 2 seconds when triggered. Similarly the <i>slideshowPermLock</i> variable of the <i>Media</i> description is used to lock the slideshow when an image remains in selected mode (mouse over an image). Moreover, interactive features of the slideshow are enhanced when T_1 is applied since animations used as automated transitions between images are triggered when interacting with the multimedia document.
S_1	T_1	I_1	The combination of all <i>STI</i> components results in a complete interactive slideshow with a 3-image display.

This *MSTI* document was generated by transforming the image gallery *Media* description with *STI* descriptions (dynamically transformed into XSL format) using a Saxon XSLT processor¹. Materials and generated documents are available at: <http://www.tsi.enst.fr/mm/MSTI/ImageGallery.html>. Additional tests designed for validation of the model and for demonstration purposes are available at <http://www.tsi.enst.fr/mm/MSTI> and cover several multimedia formats (including SVG without JavaScript).

¹ <http://saxon.sourceforge.net>

3. MULTIMEDIA DOCUMENT SCALABILITY

In previous sections, we have presented the *MSTI* model which separates the presentation properties of a multimedia document into *Spatial*, *Temporal* and *Interactive* descriptions by centralizing media elements, structures and semantics of a multimedia document in a *Media* description. In this model, the *STI* composition applies the *Spatial*, *Temporal* and *Interactive* descriptions to the *Media* description to generate a multimedia document. Document adaptation can be performed by selecting and combining appropriate *STI* descriptions for composition as described in Section 2.3. However, one limitation of adaptation frameworks based on the *MSTI* model lies in the balance between authoring efforts required to generate n different *STI* descriptions compared to the adaptation flexibility it offers (n^3 presentation options). In fact, many alternative *STI* descriptions have to be designed to address a wide range of devices. In the following, we present an extension of the *MSTI* model which enables the authoring of adaptable multimedia documents where *STI* descriptions are not alternatives but complements to each other. This extension of the *MSTI* model is called the *Scalable MSTI* model.

The terms “scalable” or “scalability” should be understood in this paper as in the SVC standard [3] and “refer to the removal of [video] bit-stream in order to adapt it to the various needs or preferences of end users as well as to varying terminal capabilities or network conditions.” In the following, we consider that the same definition can apply to multimedia documents and not only to video, audio or image content.

3.1 Scalable *MSTI* layers

Scalability, as defined above, imposes two related notions: axis and order relation. These notions can be clarified by analogy with the adaptation of video content. Indeed, for video, the resolution, the frame rate or the quality (PSNR), can be independently reduced on from another. In that case, the video is said to have three adaptation axes: spatial, temporal and quality and is scalable along an axis if an order relation can be defined to organize video data in a progressive manner. The scalability along an axis can be fine-grained or coarse-grained according to the length of adaptation steps. Each grain is called a layer. In the context of scalability applied to multimedia documents, we propose considering the *Spatial*, *Temporal* and *Interactive* components of an *MSTI* model as scalability axes.

3.1.1 *Spatial* axis

The *Spatial* component of the *MSTI* model is, by analogy to a video or image content, a good candidate for a scalability axis of a multimedia document. However, spatial properties being bi-dimensional (and sometimes tri-dimensional), it is difficult to define a total order relation in the spatial domain. To circumvent this issue, we restrict targeted spatial dimensions to incremental values as for the generalized spatial scalability of SVC [3]. As a consequence, we divide the *Spatial* axis into *Spatial* layers with the condition that “neither the horizontal nor the vertical resolution can decrease from one layer to the next.”

A layered description of the spatial properties of multimedia documents can address many screen resolutions through the use of fine-grained scalable descriptions but possibly implies overwriting spatial parameters from one layer to the next. Performance evaluations that have been carried out in [23] show that spatial adaptation through presentation updates is very efficient on constrained multimedia terminals and has significant advantages

compared to switching approaches in terms of adaptation capabilities (granularity) and memory requirements. Furthermore, presentation updates can be combined with one-way constraints [13] or linear interpolations [14] in some application scenarios.

3.1.2 Temporal axis

The *Temporal* component of the *MSTI* model can be considered as a scalability axis if the description of the timed properties of the presentation is given in a progressive manner. In the *Scalable MSTI* model, the specification of *Temporal* layers is left to the author but should mainly rely on progressive requirements in terms of processing complexity and memory requirements. For instance, a sequence of images may be proposed as a slow slideshow, then as a faster slideshow presenting key frames and finally as a low-motion video featuring key frames and intermediate images. Another example is the 3-pane image slideshow described in Section 2.4 that proposes moving to the next image by replacing the displayed images, then by proposing an animation that translates images from right to left to introduce a new image or finally by activating a 3D animation similar to the *CoverFlow* mode of iTunes.

The layers of the *Temporal* axis should be as progressive as possible from an authoring point of view. However, there might be good reasons to completely change the *Temporal* description from one layer to the next. Such cases might be triggered by the authoring need to deeply modify the temporal properties of a layer towards an enhanced timed multimedia presentation requiring advanced capabilities for playback. Thus, this kind of layer implies overheads in terms of bandwidth and may even require removing some description parts of the previous layers (see Section 3.3.4).

3.1.3 Interactive axis

In the *Scalable MSTI* model, we consider that the *Interactive* component of the *MSTI* model is also a scalability axis that can be divided into layers by progressively providing interactivity functions to the user. The ordering of the interactive features of the multimedia presentation is left to the author but should mainly rely on memory requirements and media complexity. For instance, a multimedia document may propose optional media elements that are progressively accessible by the user depending on the capabilities of his/her multimedia terminal. The *Interactive* scalability axis can also be followed to improve the ergonomic aspects of the multimedia document by providing navigation shortcuts or highlights on the current interaction status (e.g. focus on media elements).

The layers of the *Interactive* axis should be as progressive as possible from an authoring point of view. However, it might be difficult, in some cases, to order various types of media elements on a single axis. Therefore, some interactive layers may be considered optional if they do not impact the enhanced interactive layers of the same axis (see Section 3.3.1).

3.2 Cascading STI Composition

The division of *STI* components into scalable layers requires an incremental transformation of the *Media* description to progressively compose the adapted multimedia document. Therefore, the *STI* composition defined in Section 2.2.2 has been extended to allow cascading *STI* compositions in the *Scalable MSTI* model. Indeed, the *STI* composition outputs a multimedia document that can be used as '*Media*' input for a new *STI* composition as long as new *STI* descriptions update previous ones. This flexibility of the *STI* composition comes from the fact that a *Media* description of the *MSTI* model can be considered a *MSTI* document since they are both described in the same format and contain the same identifiers

referenced by *STI* components. This extended *STI* composition is depicted in Figure 8 and shows how *STI* compositions can be cascaded to generate an adapted multimedia document with two *STI* layers.

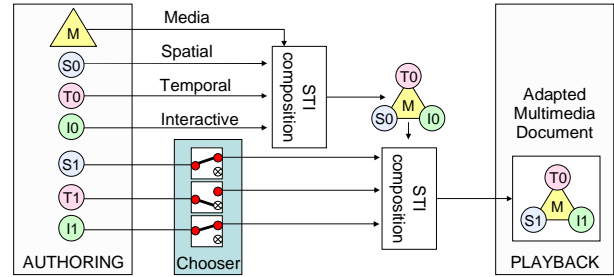


Figure 8. Cascading STI compositions

The cascading property of the *STI* composition process can be used to progressively provide spatial, temporal and/or interactive improvements to a base multimedia document. Of course, one main objective of multimedia scalability is to minimize authoring efforts by only providing the descriptions that are required from one layer to the next. As a consequence, a *Spatial* layer does not intend to completely update the previous *Spatial* layer but rather provides an update of the presentation properties that are different from the previous layer. This approach is efficient compared to the generation of separate presentation alternatives because differential updates can achieve compact document representation and low-memory consumption [23].

In the *Scalable MSTI* model, a multimedia document is the result of successive *STI* compositions for each layer on each scalability axis. In practice, the multimedia document does not necessarily have to be processed for each layer if a progressive document rendering is not required. Indeed, *STI* descriptions can be aggregated prior to the composition process. In our experiments, this optimisation has been performed by generating Extensible Style Sheets (XSL) that overwrite *replace MSTI* commands when applying several layers on the same axis by tracking *ref_id* identifiers.

Finally, a scalable multimedia document relies on a base layer (S_0, T_0, I_0) that determines the minimal presentation of the multimedia document. This base layer contains spatial, temporal and interactive properties that define a simple presentation of the media elements of the multimedia document. It may also contain *STI* properties required for some enhancement layers as explained in Section 3.3.2 and decorative properties introduced in Section 2.2.2.

3.3 The Scalable MSTI model

The *Spatial*, *Temporal* and *Interactive* (*STI*) scalability axes described in Section 3.1 make it possible to further decompose the *STI* components of the *MSTI* model into layers. As a result, the *MSTI* document syntax described in Section 2.2 has evolved to a scalable *MSTI* document syntax.

The *Scalable MSTI* document syntax is only used for the *Spatial*, *Temporal* and *Interactive* descriptions since the *Media* description is provided in the targeted multimedia format. *STI* descriptions are divided into *layer* elements to which a number is associated (*value* attribute) and which may provide informative parameters (*width*, *height*...) to guide the adaptation decision-making process. A *layer* element is composed of *insert*, *replace* and *delete* elements that can be used to modify the *Media* description by inserting, replacing or deleting presentation fragments. Compared to the *MSTI* syntax, a *delete* element has been introduced in the *Scalable MSTI* syntax in

order to remove an XML fragment by referencing an identifier (*ref_id*) that is part of the *Media* description or that has been previously inserted on the same axis. A scalable *Spatial* description is illustrated in Figure 9 and can be applied to the *Media* description illustrated in Figure 2.

```

<spatial>
  <layer value="0" width="250" height="200">
    <replace ref_id="root" attribute="viewBox">0 0 250 200</replace>
    <replace ref_id="img_0" attribute="x">45</replace>
    <replace ref_id="img_0" attribute="y">20</replace>
    <replace ref_id="img_0" attribute="width">160</replace>
    <replace ref_id="img_0" attribute="height">160</replace>
    <replace ref_id="img_0" attribute="visibility">visible</replace>
  </layer>
  <layer value="1" width="435" height="200">
    <replace ref_id="root" attribute="viewBox">0 0 435 200</replace>
    <replace ref_id="img_0" attribute="x">50</replace>
    <replace ref_id="img_1" attribute="x">225</replace>
    <replace ref_id="img_1" attribute="y">20</replace>
    <replace ref_id="img_1" attribute="width">160</replace>
    <replace ref_id="img_1" attribute="height">160</replace>
    <replace ref_id="img_1" attribute="visibility">visible</replace>
  </layer>
</spatial>

```

Figure 9. A scalable spatial description

Additionally, eleven *layer* attributes have been defined in the *Scalable MSTI* model to specify adaptation paths in three-axis adaptation graphs: *skippable*, *dependsOn{S/T/I}LayerValue*, *requiredFor{S/T/I}LayerValue*, *blocks{S/T/I}LayerValue*, and *randomAccessLayer*. They are described in the following sections.

3.3.1 One-dimension adaptation path

When considering scalability along a single axis, the adaptation path is straightforward since it simply follows the layer order and only ends when it reaches a layer containing data that does not fit the usage scenario or terminal capabilities.

The *Scalable MSTI* approach is a bit more flexible. Indeed, the layers of *Scalable MSTI* model provide progressive enhancements to the multimedia document. However, some of these enhancements may be considered optional as explained in Section 3.1.3. Therefore, the *Scalable MSTI* model introduces the concept of *skippable* layer to allow skipping layers that are optional and do not impact the enhanced layers of the same scalability axis. Such an adaptation path is described in Figure 10.

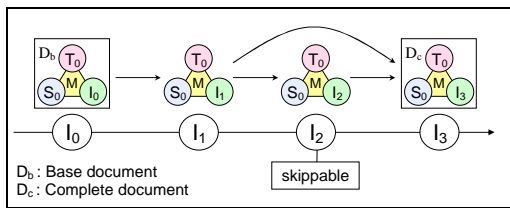


Figure 10. Adaptation path with skippable layers

3.3.2 Three-dimension constrained adaptation paths

As explained in Section 2.2.6, *STI* components are separated and do not refer to each other but share common structures and variables in the *Media* description. However, *STI* layers are not independent since the spatial, temporal and interactive properties of a multimedia presentation can be closely coupled. Hence, for a typical presentation fragment describing an animation triggered by a user action, one option is to describe all the dependent properties of the multimedia presentation in the base layer of the *STI* scalability axis. This authoring approach is advantageous since all adaptation paths are possible and will lead to the complete presentation. However, it

may not be efficient since the base layers of *STI* scalability axes may include many unused descriptions. For instance, the position of all buttons defined in the *Interactive* description will be given in the S_0 *Spatial* base layer whereas the actual use of one of these buttons may be defined in the I_2 *Interactive* description. This is not optimal since a multimedia document in configuration (S_0, T_0, I_0) will perform the positioning of buttons that will never be visible to the user.

In order to cope with such dependencies, the *Scalable MSTI* model defines layer dependency using symmetric attributes: *dependsOn{S/T/I}LayerValue* and *requiredFor{S/T/I}LayerValue*. When used, the *Scalable MSTI* model does not grant the user complete freedom in the *MSTI* adaptation graph but proposes a set of adaptation paths to the complete presentation defined during the authoring phase. The layer dependency mechanism is illustrated in Figure 11.

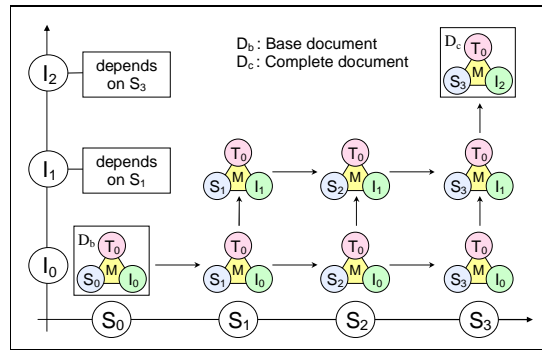


Figure 11. Constrained adaptation paths

3.3.3 Dead-end adaptation paths

A single multimedia document may be used in several usage cases for which adaptation paths may diverge. These specific versions do not fit well with the *STI* progressive approach because they do not aim to converge toward the same complete presentation. Such a typical application scenario is illustrated by a print view of a multimedia document. Indeed, a printable multimedia document does not require any temporal and interactive presentation properties and supports greater resolutions than usual multimedia terminals. In that case, the *Spatial* description can be built progressively by applying *Spatial* layers to reach a suitable printable size but this layout will probably be incompatible with the temporal and navigation scheme defined by the content creator to grant the user media element access.

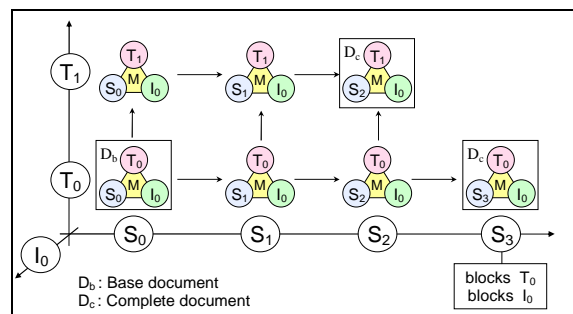


Figure 12. Adaptation path dead-end

The *Scalable MSTI* model copes with such specific adaptation needs by defining the *blocks{S/T/I}LayerValue* attribute on an *STI* layer. When applied to a layer, this attribute specifies the highest layer from the *STI* component it is compatible with. For instance, the print

layer of the *Spatial* component may block the adaptation path on the *Temporal* and the *Interactive* axes to the base layer of the scalable document (*blocksSTLayerValue*='0' and *blocksLLayerValue*='0'). As a consequence, this mechanism introduces dead-ends on purpose in the adaptation path as illustrated in Figure 12.

3.3.4 Random Access Layers

The main drawback of the *Scalable MSTI* model in terms of playback is that it necessitates processing overheads to enable advanced multimedia presentations (the highest *STI* layers) when fine-grained multimedia documents are displayed [23]. The *scalable MSTI* model also assumes that a reliable mapping between *STI* layers and terminal capabilities is always available when the adaptation is processed. This mapping is feasible since it can be completed by evaluating all presentation options of the adaptation graph as is done in the MPEG-21 adaptation framework [24]. However, the data needed to describe this mapping when fine-granular multimedia documents are created might be too large in some application scenarios such as broadcast services. These two limitations are handled in the *Scalable MSTI* model by the definition of Random Access Layers.

A Random Access Layer (RAL) flag is assigned to a *STI* layer by defining a *randomAccessLayer* attribute. This attribute can be used for fast access to an advanced presentation and partly disables progressive features of the *Scalable MSTI* model in order to minimize processing overheads for some application scenarios. A RAL flag can also be used to signal a valid fallback solution (without reloading all the content) when a player blindly tries to reach an optimum presentation by successively applying *STI* layers without relying on metadata describing a mapping to terminal capabilities. Therefore, RAL provides a backward and forward refresh of *STI* layers on a single axis. For instance, a S_k RAL will reset all the modifications incurred by $[S_{k+1}, S_n]$ *Spatial* layers (forward refresh), it will collapse all *Spatial* descriptions from the S_0 to the S_{k-1} layer (backward refresh) and it will include the *Spatial* description specifically corresponding to layer k . Therefore, despite the fact that the definition of Random Access Layers introduces description overheads, it can provide alternative entries to (S_0, T_0, I_0) in the adaptation graph and backward adaptation paths as illustrated in Figure 13.

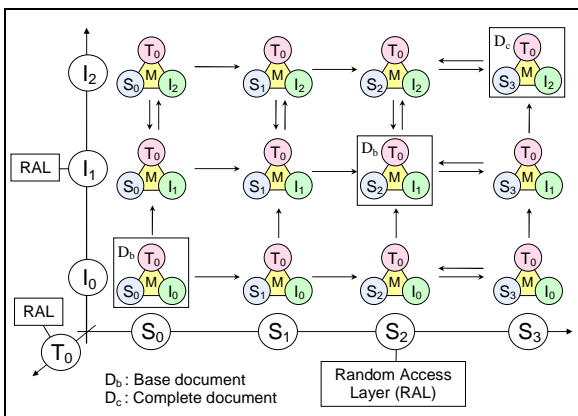


Figure 13. Adaptation path with Random Access Layers

3.4 Example of a scalable image gallery

The example described in this section is a scalable version of the image gallery document described in Section 2.4. In this section, *STI* components are described separately for a clearer explanation but they can be combined as illustrated in Table 1. The scalable image

gallery example described in this paper is composed of 5 *Spatial* layers (S_0, S_1, S_2, S_3, S_4), 3 *Temporal* layers (T_0, T_1, T_2) and 4 *Interactive* layers (I_0, I_1, I_2, I_3). This scalable example is available at <http://www.tsi.enst.fr/mm/MSTI/ImageGallery.html> and more examples can be found at <http://www.tsi.enst.fr/mm/MSTI>.

3.4.1 Spatial Layers

The *Spatial* description of the image gallery is responsible for the layout of all media elements: 6 images, 6 'shadow' rectangles and 4 'arrows'. During authoring, 5 document layouts have been designed as illustrated in Figure 14.

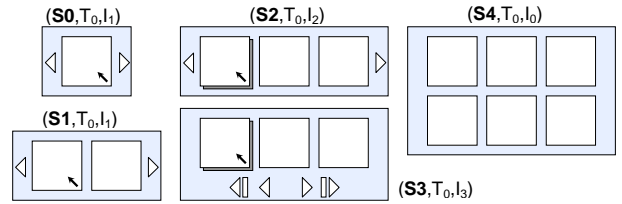


Figure 14. Various spatial configurations of an image gallery

This scalable image gallery document has been designed to provide flexible adaptation paths. However, the S_4 layer blocks the T_0 and the I_0 layers since all the navigation schemes are disabled when all the images are displayed.

3.4.2 Temporal Layers

The *Temporal* scalability layers of the image gallery can be used to create two types of slideshows since the T_0 layer does not provide any temporal description. First, the T_1 layer is based on a global timer that automatically updates images every 5 seconds. Second, the T_2 layer enhances the T_1 layer by triggering image animations on global timer ticks as described in Section 2.4 and illustrated in Figure 15.

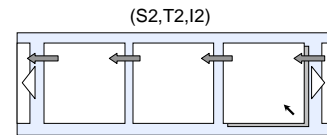


Figure 15. Interactive timed slideshow presentation

The activation of the T_2 layer of the image gallery document can be decided from the capabilities of the targeted terminal. In our case, this authoring has mainly been driven by the fact that Mozilla Firefox SVG viewer cannot handle the T_2 layer because it does not support SVG *animate* elements while Opera does.

3.4.3 Interactive Layers

The *Interactive* description of the image gallery document provides several means of navigation to the user (I_0 is empty). First, the I_1 layer activates forward and backward buttons that give access to images through interactivity. Additionally, the I_2 layer enables sensitive areas over images so that a selection mode (display of a shadow rectangle behind the image) is visible when focusing on an image. This layer can be considered cosmetic in some cases and is therefore defined as skippable. However, this image selection mode is designed to give the user a chance to pause an automated slideshow that can be defined in the *Temporal* component of the multimedia document as explained in Section 2.4. Finally, the I_3 layer activates fast-forward and fast-backward buttons that can be used to navigate in the gallery with a step of three images. Therefore, the I_3 layer depends on the S_3 layer since these buttons are only accessible from this layout.

4. CONCLUSION

The *Scalable MSTI* document model provides a generic and flexible framework for multimedia document production that efficiently minimizes adaptation complexity by relying on the authors wishes to guarantee the quality of their presentation. Our experiments show that the decomposition of multimedia documents into *Media*, *Spatial*, *Temporal* and *Interactive* descriptions makes it possible to define scalability axes that are relevant for fundamental authoring needs. Indeed, the progressive layers of *scalable MSTI* documents leverage the usual incremental authoring process of creative tasks which consists in producing something simple and then improving upon it. Additionally, the *Scalable MSTI* model provides a simple toolbox to overcome the adaptation limitations of progressive authoring by defining specific exceptions in the adaptation graph of scalable documents.

We are currently finalizing our performance measurements on the adaptation of *MSTI Scalable* multimedia documents on mobile devices that will complement our studies on spatial scalability in broadcast scenarios [23]. As a next step, we plan to focus on authoring difficulties when editing scalable multimedia documents. In particular, the definition of scalable multimedia templates and the specification of a file format for the exchange of *Scalable MSTI* multimedia documents are pertinent topics.

5. ACKNOWLEDGMENTS

This work is partially supported by RTL (EDIRADIO) and the NoE INTERMEDIA funded by the European Commission (NoE 038419). The authors would like to thank Marc Brelot and James Benenson for their review of this paper and their helpful comments.

6. REFERENCES

- [1] A. Vetro, C. Christopoulos, T. Ebrahimi, eds., Special Issue on Universal Multimedia Access, IEEE Signal Processing Magazine, vol. 20, no. 2, March 2003.
- [2] W. Li, Overview of Fine Granularity Scalability in MPEG-4 Video Standard, IEEE Trans. on Circuit Systems for Video Technology, vol. 11, no. 3, pp. 301-317, March 2001.
- [3] H. Schwarz, D. Marpe, T. Wiegand, Overview of the Scalable Video Coding Extension of the H.264/AVC Standard, IEEE Trans. on Circuit Systems for Video Technology, vol. 17, no. 9, pp. 1103-1120, Sept. 2007.
- [4] B. Pellan, C. Concolato, Media-Driven Dynamic Scene Adaptation, Proc. of the 8th Workshop on Image Analysis for Multimedia Interactive Services, pp.67-71, June 2007.
- [5] L. Villard, C. Roisin, N. Layaida, A XML-based Multimedia Document Processing Model for Content Adaptation, Digital Documents and Electronic Publishing, LNCS 2023, Springer Verlag, pp.104-119, Sept. 2000.
- [6] S. Boll, W. Klas, ZYX-A Multimedia Document Model for Reuse and Adaptation of Multimedia Content, IEEE Trans. On Knowledge and Data Engineering, vol. 13, no. 3, pp. 361-382, May 2001.
- [7] M. Bilasco, J. Gensel, M. Villanova-Oliver, STAMP: A Model for Generating Adaptable Multimedia Presentations, Multimedia Tools and Applications, Springer Netherlands, vol. 25, no. 3, pp. 361-375, March 2005.
- [8] O. Andersson, R. Berjon, E. Dahlström et al., Scalable Vector Graphics (SVG) Tiny 1.2 Specification, W3C, <http://www.w3.org/TR/SVGMobile12>, Aug. 2006.
- [9] D. Bulterman, G. Grassel, J. Jansen et al., Synchronized Multimedia Intergration Language (SMIL 2.1), W3C, <http://www.w3.org/TR/SMIL>, Dec 2005.
- [10] ISO/IEC 14496-11:2005, Information technology - Coding of audio-visual objects - Part 11: Scene description and application engine, 2005.
- [11] M. Nanard, J. Nanard, P.R. King, L. Gaillard, Genre Driven Multimedia Document Production by Means of Incremental Transformation, Proc. of ACM Symposium on Document Engineering, pp. 111-120, Winnipeg, Canada, Aug. 2007.
- [12] A. Scherp, S. Boll, MM4U: A Framework for Creating Personalized Multimedia Content, Managing Multimedia Semantics, IRM Press, pp. 246-287, Feb. 2005.
- [13] K. Marriott, B. Meyer, L. Tardif, Fast and Efficient Client-Side Adaptivity for SVG, Proc. of the 11th International Conference on World Wide Web, pp. 496-507, Honolulu, USA, May 2002.
- [14] P. Dragicevic, S. Chatty, D. Thevenin, J.L. Vinot, Artistic Resizing: A Technique for Rich Scale-Sensitive Vector Graphics, Proc. of the 18th annual ACM Symposium on User Interface Software and Technology, pp. 201-210, Seattle, USA, Oct. 2005.
- [15] G. Durand, G. Kazai, M. Lalmas, U. Rauschenbach, P. Wolf, A metadata Model Supporting Scalable Interactive TV Services, Proc. of the 11th International Multimedia Modeling Conference, pp. 386-391, Jan. 2005.
- [16] D.C.A. Bulterman, L. Hardman, Structured Multimedia Authoring, ACM Trans. On Multimedia Computing, Communications, and Applications, vol. 1, no. 1, pp. 89-109, Feb. 2005.
- [17] P. Vuorimaa, D. Bulterman, P. Cesar, SMIL Timesheets 1.0, W3C, <http://www.w3.org/TR/timesheets>, Jan. 2008.
- [18] J. Marsh, D. Veillard, N. Walsh, xml:id Version 1.0, W3C, <http://www.w3.org/TR/xml-id>, Sept. 2005.
- [19] J. Clark, S. DeRose, XML Path Language (XPath), W3C, <http://www.w3.org/TR/xpath>, Nov. 1999.
- [20] R. Berjon, Remote Events for XML (REX) 1.0, W3C, <http://www.w3.org/TR/rex>, Oct. 2006.
- [21] H.W. Lie, T. Celik, D. Glazman, Media Queries, W3C, <http://www.w3.org/TR/css3-mediaqueries>, June 2007.
- [22] C. Kiss, Composite Capability/Preference Profiles (CC/PP): Structure and Vocabularies 2.0, W3C, <http://www.w3.org/TR/CCPP-struct-vocab2>, April 2007.
- [23] B. Pellan and C. Concolato, Spatial Scene Adaptation in Broadcast Environment, Proc. of the International Conference on Multimedia & Expo (ICME), pp. 389-392, Hannover, Germany, June 2008.
- [24] M. Ransburg, R. Cazoulat, B. Pellan, C. Concolato et al., Dynamic and Distributed Adaptation of Scalable Multimedia Content in a Context-Aware Environment, Proc. of EuMob, Alghero, Italy, Sept. 2006.