# A Cross-Media Adaptation Strategy for Multimedia Presentations

Susanne Boll, Wolfgang Klas, Jochen Wandel
Databases and Information Systems (DBIS)
Computer Science Department, University of Ulm, Germany
{boll, klas, wandel}@informatik.uni-ulm.de

## Abstract

Adaptation techniques for multimedia presentations are mainly concerned with switching between different qualities of single media elements to reduce the data volume and by this to adapt to limited presentation resources. This kind of adaptation, however, is limited to an inherent lower bound, i.e., the lowest acceptable technical quality of the respective media type. To overcome this limitation, we propose *cross-media* adaptation in which the presentation alternatives can be media elements of different media type, even different fragments. Thereby, the alternatives can extremely vary in media type and data volume and this enormously widens the possibilities to efficiently adapt to the current presentation resources. However, the adapted presentation must still convey the same content as the original one, hence, the substitution of media elements and fragments must preserve the presentation semantics. Therefore, our cross-media adaptation strategy provides models for the automatic *augmentation* of multimedia documents by semantically equivalent presentation alternatives. Additionally, during presentation, *substitution* models enforce a semantically correct information flow in case of dynamic adaptation to varying presentation resources. The cross-media adaptation strategy allows for flexible reuse of multimedia content in many different environments and, at the same time, maintains a semantically correct information flow of the presentation.

**Keywords:** adaptation, quality of information, multimedia authoring, multimedia presentation

## 1 Introduction

Timely delivery of multimedia content to the users of a multimedia information system as well as timely rendering of multimedia presentations are essential in order to provide a smooth and continuous presentation service. Hereby, the multimedia information system must cope with the *user context*, i.e., the user's preferences, the targeted user system environment, and varying resources like available network bandwidth and CPU time. Consequently, trying best to deliver a continuous, synchronized multimedia presentation, the multimedia information system must *adapt* the delivery and the rendering of the presentation to the current user context.

User context-specific adaptation is one of the challenges we obtain from our research project "Gallery of Cardiac Surgery" (Cardio-OP[1]) [10] that aims at the development of an Internet-based, database-driven multimedia information system in the domain of cardiac surgery common to its different user groups, namely physicians, medical lecturers, students, and patients. The different users will access the multimedia repository from different locations ranging from a low-end PC with a low bandwidth connection (ISDN) at home to high-end PCs with high bandwidth connections on campus. As *reuse* of multimedia information is of high importance and an important cost factor in such a system, we aim at modeling multimedia information such that it can be reused in and adapted to different user contexts. As we cannot assume network connections that allow resource reservation for the project, we are concerned with adaptation techniques that follow a best effort approach.

A lot of current best-effort adaptation techniques are considered with switching between different qualities of single continuous media elements, i.e., adapting/smoothing the data rate. For the adaptation of video, we find, e.g., approaches that drop single frames [8] (MJPEG, MPEG), allow for dropping enhanced layers of MPEG-2 (e.g., [13]), skip group of pictures [7], or assign priorities to frames [6, 12]. Some approaches try to adapt not only in one quality dimension but also in two or more, e.g., [15]. [14] presents an overall adaptation scheme for handling multiple presentations in multimedia databases but also restricts adaptation to the quality of single media streams. All these techniques focus on low-level quality-of-service adaptation and, in addition, usually take alternative media quality as granted or inherent with the media. As long as the adaptation stays with the media type, the limitation of the adaptation is the minimal technical quality of the respective media type which is still acceptable for a user. For example, an MPEG video stream, already maximally adapted, still needs a certain network bandwidth and sufficient CPU time for decod-

[1]Cardio-OP - Gallery of Cardiac Surgery - is partially funded by the German Ministry of Research and Education, grant number 08C58456. Our project partners are the University Hospital of Ulm, Dept. of Cardiac Surgery and Dept. of Cardiology, the University Hospital of Heidelberg, Dept. of Cardiac Surgery, an associated Rehabilitation Hospital, the publisher Hüthig-Verlagsgruppe, Heidelberg, FAW Ulm, and ENTEC GmbH, St. Augustin. For details see also URL www.informatik.uni-ulm.de/dbis/Cardio-OP/

ing to be reasonably presented. From our point of view, these approaches lack the usage of alternative media elements of different media type for adaptation and, hence, leave out more powerful adaptation possibilities.

In the context of our work, we widen the term of adaptation to *cross-media* adaptation which means that media elements and also entire fragments of a multimedia presentation can be replaced by different media elements/fragments which can be of different media type. We find the following related approaches that address some aspects of cross-media adaptation. [17] proposes a cross-media adaptation technique focusing on the scheduling of alternatives aiming at best effort assignment of currently available resources. The approach presented in [11] proposes cross-media adaptation of Web pages to client resources in the Internet before the actual delivery to the respective client environment. SMIL [9] allows for cross-media adaptation, however, the number of parameters the presentation can be adapted to is quite limited. The approach discussed in [4, 5], mainly driven by research in authoring environments, addresses support of quality adaptation on the user level and look at abstractions for handling presentation adaptation on a much higher level. However, the related approaches presented above do not necessarily allow for continuous/dynamic adaptation and, even more important, do not ensure the semantic equivalence of the used presentation alternatives but rather concentrate on the scheduling and selection of alternatives.

From our perspective, cross-media adaptation means to replace/adapt a media element not only by the same media element of different quality but also by a media element of different type. Additionally, we want to replace/adapt not only single media elements but also *fragments* of a multimedia presentation, i.e., coherent, logical parts of multimedia documents. For example, on low network bandwidth, we would like to replace an audio containing a talk about bypass surgery by its transcript or a video-audio-presentation, that cannot satisfyingly be presented to a user, by a sequence of pictures with corresponding captions that has a much lower data volume.

Changing the media type during adaptation instantly raises the question of how to preserve the presentation semantics as the adapted presentation still must convey the same content to the user as the original one. Therefore, both semantically equivalent alternatives must be proposed for an adaptation and the dynamic change between alternatives during presentation must not disrupt the "story", i.e., the *information flow*.

In this paper, we propose formal models that ensure that the presentation alternatives preserve the presentation semantics and a correct flow of information. The *augmentation* models can be used to verify semantic equivalence of automatically generated presentation alternatives. Hereby, we can relieve the author of a multimedia presentation from specifying different presentations for different user contexts and save an enormous amount of effort and costs. *Substitution* models control the adaptation of the actual presentation. They give the consumer of a presentation the possibility to select a specific degree of adaptation, i.e., how "close" the adapted presentation must stay to the original presentation. We developed these models in the framework of the adaptive document model ZYX [1, 2, 3], and a first version of an adaptive presentation engine has been implemented within the context of the Cardio-OP project.

The remainder of this paper is organized as follows: Section 2 gives the reader an overview of the adaptation strategy we propose. Sections 3 and 4 present the *augmentation models* that ensure the correctness of different presentation alternatives and the *substitution models* that ensure the correct switching between alternatives in a dynamic adaptation during the presentation. Section 5 summarizes our work and points out the implications of the models.

## 2 The Cross-Media Adaptation Strategy

For the underlying technical infrastructure, we assume a multimedia database server that manages multimedia documents, media data, and meta data, and that is connected with heterogeneous multimedia clients via different network connections and protocols. To introduce adaptivity in this system, we model it *within* the multimedia documents. This means the multimedia document must model so called presentation alternatives representing the adaptation possibilities, i.e., alternative media elements and fragments that can be exploited for adaptation. Having such multimedia documents, adaptation to the targeted presentation environment can be realized by the selection of the most appropriate alternatives out of those specified. The particular interest of this work, however, is not the blindfolded substitution of media elements and fragments, but their substitution by *semantic equivalent alternatives*. The provision and usage of such presentation alternatives for adaptation maintaining a semantically correct information flow is the goal of our cross-media adaptation strategy.

In our strategy, the verification and selection of potential alternatives within the cross-media adaptation process takes place in two phases: First, the cross-media adaptation strategy must verify that the alternatives of the multimedia document preserve the semantics of the presentation. Second, during the actual presentation playout, the adaptation strategy has to assure that switches between the available alternatives maintain the information flow. In the following, we present the single steps of the proposed cross-media adaptation strategy which are illustrated in Figure 1 in more detail.

For a fine grained adaptation to many different user contexts, it is mandatory that a high number of alternatives is available. To relieve the author from the time consuming burden of specifying all possible alternatives, we propose to automatize the specification of the alternatives. We call this step *augmentation* of the multimedia document which takes place after the document has

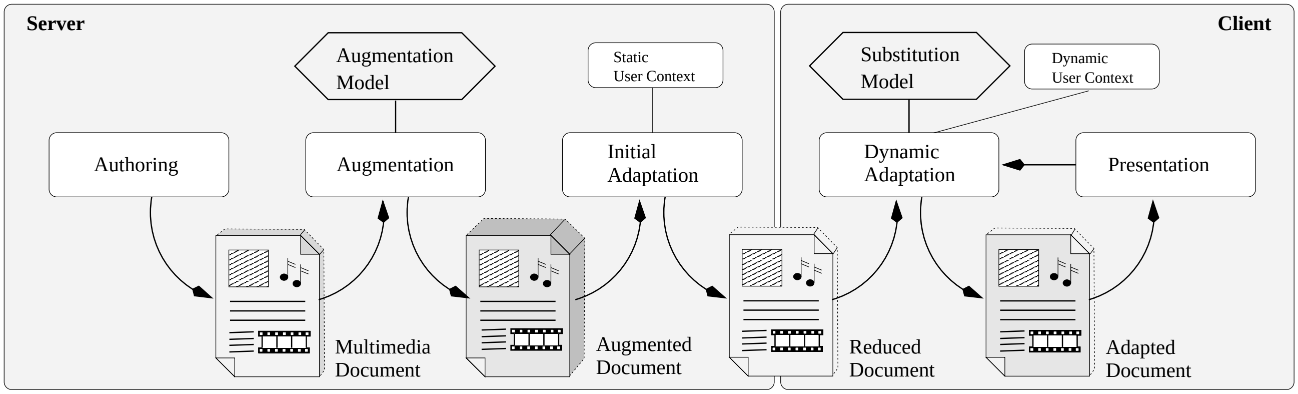


Figure 1: The steps of the cross-media adaptation strategy

been composed by the author. To find possible presentation alternatives, the augmentation process queries the underlying database exploiting inherent technical data and the meta data the media elements have been annotated with to receive potential presentation alternatives. To preserve the semantics of the presentation intended by the author, the suggested alternatives cannot simply be inserted into the document but undergo a verification to assure that the augmented document is still valid with regard to the presentation semantics. We have formalized this verification in the so called *augmentation models*. The result of the augmentation is an augmented multimedia document that covers the heterogeneous universe of different renderings of the presentation.

The augmented document then undergoes an *initial adaptation* that, given a targeted (static) user context, removes those alternatives that are not relevant for that context and, hence, forms the context specific "view" on the augmented document. For example, in the case that the presentation environment has no sound device installed, the alternatives for audios never come into question for presentation and are removed from the document. The initial adaptation has two objectives: First, it reduces the number of alternatives and consequently the data volume of the document. Second, it reduces the calculation effort of the adaptation algorithm during playout because less alternatives must be considered. After this step, the document is ready for presentation in the targeted user context.

When presenting the multimedia document, clients prefetch raw data for only those media elements specified in the document that are likely for rendering. During the presentation, due to unpredictable changes in the (dynamic) user context, i.e., fluctuations in the available system resources, *dynamic adaptation* may be necessary. This means on-the-fly switching to those alternatives that are more appropriate to a changed user context and by this continuously adjusting the presentation to the current user context. Thereby, the maintenance of the *information flow* of the presentation must be assured. For example, if a video is subject to substitution by a sequence of images, the change between the alternatives must not disrupt the presentation flow, i.e., the presentation of the image sequence has to begin with that image that correlates to the current scene in the video. Similarly to the augmentation models, we, hence, provide *substitution models* that check if a dynamic adaptation maintains a correct information flow.

## 3 Augmentation — Enhancing Multimedia Documents with Alternatives

The augmentation process identifies presentation alternatives used to enhance a pre-orchestrated multimedia document for adaptation. In the following, we present augmentation models that assure the semantic correctness of the alternatives added to the document. For each alternative suggested, the rules of the augmentation models check and ensure if the alternative conveys the same content as the original, i.e., if it is semantically equivalent and can be used for adaptation.

As the notion of *equivalence of contents* can be interpreted in different degrees of strictness, we have elaborated and formalized different augmentation models that vary in their definition of strictness. Before we introduce the augmentation models in Section 3.3, we make some assumptions on the document model and give basic definitions in Section 3.1 and semantic relationships that are fundamental for the formal models in Section 3.2.

### 3.1 Assumptions on the Document Model and Basic Definitions

For the underlying multimedia document model, we make the following assumptions: The document model must offer the possibility to model alternatives and to explicitly model the structural composition of a presentation in terms of media elements. It is required that the structural elements of alternatives can be accessed and exploited by the adaptation strategy. Furthermore, the document model must be able to model alternatives not only on the level of single media elements but also on the level of fragments. We consider fragments to be encapsulated by *complex media elements* and distinguish them from *atomic media elements* that represent single media elements. This distinction is necessary as complex and atomic media elements have different characteristics and, depending on these, the adaptation technique is of different complexity, which is reflected in the following formal models.

The features as described above can be found in different document models to a different extent and different peculiarity [3]. For example, in SMIL [9], being XML based, the structural elements can be accessed and the *switch* operator serves to specify presentation alternatives. The ZYX [1] model uses a similar *switch* operator to model alternatives and *complex media elements* encapsulate substructures of ZYX documents. CMIF [4] allows nested presentations and offers *channels* that can be used to model alternative presentation parts. TIEMPO [17] is another example of a document model that supports the modeling of presentation alternatives by means of selection groups that can be nested.

Obviously, the cross-media adaptation strategy we propose can be applied in the context of all these models. We have applied the model exemplarily in the ZYX context [16].

The symbols introduced in Definition 1 are used in the formal definitions to follow.

**Definition 1 — Symbols:** $\mathcal{AM}$,$\mathcal{CM}$,$\mathcal{M}$,$\mathcal{SUBJ}$,$\mathcal{DOC}$

Let $\mathcal{AM}$ denote the set of all *atomic media elements*, $\mathcal{CM}$ denote the set of all *complex media elements*, and $\mathcal{M} = \mathcal{AM} \cup \mathcal{CM}$ denote the set of all media elements.

Let $\mathcal{SUBJ}$ denote a set of terms that represent subjects of interest from the application domain.

Let $\mathcal{DOC}$ denote the set of all multimedia documents.

Each atomic and complex media element has associated a certain amount of meta data which is exploited for our adaptation strategy. With basic functions as exemplarily given in Definition 2, we access some of the basic features of media elements like the associated subjects.

**Definition 2 — Basic Functions:** *subjects*, *duration*

$subjects : \mathcal{M} \rightarrow 2^{\mathcal{SUBJ}}$

The function *subjects* returns the set of those subjects that are associated with an atomic or complex medium $m \in \mathcal{M}$.

$duration : \mathcal{M} \rightarrow \mathbb{N} \cup \{\infty\}$

The function *duration* returns the duration of a medium $m \in \mathcal{M}$. If $m$ is an atomic discrete medium, its duration is set to $\infty$ by definition. If $m$ is an atomic continuous medium, it has an inherent duration $dur \in \mathbb{N}$. For complex media elements, the *duration* can be determined taking into account the media elements and their temporal composition within the complex media element.

As these are only samples of basic functions and the set of functions extracting the media elements' properties is much more comprehensive, let $\mathcal{PROPFUNC}$ denote the complete set of basic functions.

### 3.2 Semantic Relationships

As we want to augment atomic and complex media elements with cross-media alternatives, we need to define the notion of equivalence between media elements. Media-equivalence (Definition 3) defines equivalence only with respect to certain aspects of media elements. This restriction is reasonable, as it might be sufficient for an adaptation strategy that two media elements correspond only with respect to *some* features but not necessarily with respect to all features.

**Definition 3 — Media-Equivalence $\simeq_D$**

For $m_1, m_2 \in \mathcal{M}$ and $D \subseteq \mathcal{PROPFUNC}$ the predicate $m_1 \simeq_D m_2$ is defined as *media-equivalence of $m_1$ and $m_2$ with respect to $D$* iff the two media elements are equal in each discriminating aspect $d \in D$.

Figure 2 illustrates media-equivalence for the discriminating aspects *duration* and *subjects*. Given this general definition of media-equivalence, one can introduce specific relationships, e.g., a type-equivalence for atomic media elements that compares the media elements' media type. It may be also of interest to compare the temporal course, the spatial arrangement, or the entire composition of media elements, especially for complex media elements. Such equivalence relationships can easily be defined by specifying appropriate sets $\mathcal{D}$ for $\simeq_{\mathcal{D}}$. Hence, Definition 3 establishes a family of media-equivalence relationships. In the definition of the augmentation models and the substitution models, we will see when and how more or less strict media-equivalences are of advantage.

The definition of media-equivalence covers the comparison of features associated with entire media elements. However, the subject of continuous atomic and complex media elements may vary over time. Therefore, to express the semantics of these media in more detail, the subjects are associated to temporal intervals of the media elements. In consequence, we need a function *subjRel* (Definition 4) that maps the subjects of atomic or complex media elements to the temporal intervals in which they are relevant. As the duration of discrete media is set to $\infty$, a subject can be associated with the temporal interval 0 to $\infty$.

**Definition 4 — Subject-Relevant Temporal Intervals** *subjRel*

$subjRel : \mathcal{M} \times \mathcal{SUBJ} \rightarrow 2^{\mathbb{N} \times \mathbb{N} \cup \{\infty\}}$

For a medium $m \in \mathcal{M}$ and a given subject $s \in \mathcal{SUBJ}$ the function *subjRel* returns the set of temporal intervals of $m$ that have been annotated with $s$.

The association of subjects to temporal intervals of a media element gives a more fine-grained description of the presentation semantics and, therefore, more detailed information on how a semantically equivalent alternative should look like.

Given the subject-relevant temporal intervals, we now can define a predicate *subjectSync* (Definition 5) that uses subject-relevant temporal intervals to express a more

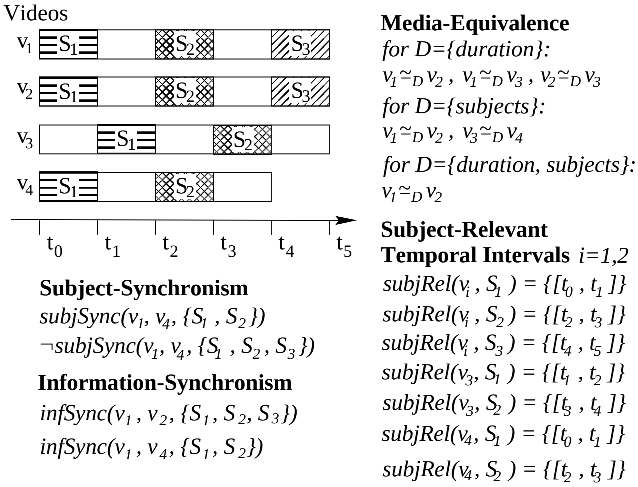


Figure 2: An example for the basic functions and semantic relationships

fine-grained equivalence relationship, especially for continuous media elements.

**Definition 5 — Subject-Synchronism Predicate** $subjSync$
Let $m_1, m_2 \in \mathcal{M}, S \subseteq \mathcal{SUBJ}$.

$$subjSync(m_1, m_2, S) \Leftrightarrow \forall s \in S : subjRel(m_1, s) = subjRel(m_2, s)$$

If two media elements $m_1, m_2$ are *subject-synchron*, all temporal intervals of $m_1$ and $m_2$ that are relevant with respect to the subjects $s \in \mathcal{S}$ are identical. If the $subjSync$ predicate holds for a media element and a proposed alternative, we can assume that the alternative still conveys a content close to the original media element. Thus, the $subjSync$ predicate can be employed to assure that presentation alternatives keep the semantics of the original media element. Figure 2 illustrates the application of $subjRel$ and $subjSync$ to the four sample videos.

For an equivalence relationship between two media elements, one could require additionally that, if a switch is performed, not only the subject stays the same but also the information flow. For example, if in one audio a sentence is commenced one may require that after a switch the sentence is correctly resumed in a second one. For rather strict adaptation approaches a predicate that ensures the maintenance of a correct information flow may be of interest.

Therefore, we define, in addition to subject-synchronism, *information-synchronism*. First, we give an auxiliary definition $syncInt$ to specify information-synchronism between two media elements during a given single interval (Definition 6), and second, the definition of information-synchronism $infSync$ between two entire media elements regarding a set of subjects (Definition 7).

**Definition 6 — Information-Synchronous Interval** $syncInt$
Let $m_1, m_2 \in \mathcal{M}, interval \in \mathbb{N} \times \mathbb{N}, s \in \mathcal{SUBJ}$.
$syncInt(m_1, m_2, interval, s)$ is a predicate that holds, if both media elements $m_1, m_2$ communicate the same temporal sequence of information within the given interval $interval$ with regard to subject $s$.

**Definition 7 — Information-Synchronism Predicate** $infSync$
Let $m_1, m_2 \in \mathcal{M}, S \subseteq \mathcal{SUBJ}$.

$$infSync(m_1, m_2, S) \Leftrightarrow subjSync(m_1, m_2, S) \wedge \forall s \in S, \forall interval \in subjRel(m_1, s) : syncInt(m_1, m_2, interval, s)$$

Two media elements are information-synchron regarding a set of subjects $\mathcal{S}$ if they are subject-synchron and information-synchron within all subject-relevant intervals. If the $infSync$ predicate holds, switches between media elements during their presentation can be performed without disrupting the information flow. However, what exactly "information-synchron" means is very much dependend on the application. We assume information-synchronism $infSync$ between some of the sample videos as given in Figure 2.

The information-synchronism always holds for the different technical qualities of a medium. As these different qualities can simply be produced automatically, an enormous amount of information-synchron alternatives can be offered. The information-synchronism feature for cross-media elements, however, must be annotated by means of an intellectual task and cannot be derived from the media automatically so far. The more fine-grained the subjects of continuous media elements are annotated to intervals, the more precise is a comparison of media elements. The size of the intervals directly influences the efficiency and accuracy of the predicates.

Obviously, the effect of both the $subjSync$ and $infSync$ predicates very much depends on the quality of metadata. Only if the annotated subject really meets the content, these predicates work properly. Hence, the quality of the metadata determines if the automatically computed and selected alternatives really fit with the content of the original document. Even though comprehensive annotation seems to be costly at first sight, frequent reuse of the material in different contexts charges off the effort.

### 3.3 Augmentation Models

An *augmentation model* is used to control the augmentation process. It comprises a set of constraints that determine which potential alternatives are actually admitted for augmentation. As mentioned before, different degrees of strictness may be of interest for the augmentation. Therefore, augmentation models cover different levels of strictness each of which stands for a certain quality of adaptation. Let $\mathcal{AUG}$ denote the set of the different augmentation models. The selected augmentation model serves as input to the augmentation process defined by the function $aug$ as follows.

**Definition 8 — Augmentation** $aug$

$aug : \mathcal{DOC} \times \mathcal{AUG} \to \mathcal{DOC}$ augments the atomic and complex media elements of a multimedia document $doc \in \mathcal{DOC}$ with alternatives. Media elements are inserted as alternatives into the document if they comply with the constraints of the given augmentation model $ag \in \mathcal{AUG}$.

Before we introduce the different augmentation models, we define an auxiliary function $M$ which extracts all media elements of a given document (Definition 9) and a function $alts$ that determines the set of augmented alternatives for a given media element (Definition 10).

**Definition 9 — Document Media** $M$

$M : \mathcal{DOC} \to 2^{\mathcal{M}}$ is a function, that returns the set of those media elements $m \in \mathcal{M}$ that are contained in a given document $doc \in \mathcal{DOC}$.

**Definition 10 — Augmenting Media** $alts$

Let $doc, doc' \in \mathcal{DOC}$ and $ag \in \mathcal{AUG}$, with $doc' = aug(doc, ag)$.

$$alts(doc, doc', m) = \{m'_1, \ldots, m'_n \mid m'_i \in M(doc') \wedge$$
$$m'_i \text{ is a presentation alternative for } m \in M(doc)\}$$

The function $alts$ returns the set of those alternatives $m'_i$ that have been added to a media element $m$ by the augmentation of $doc$ resulting in $doc'$.

Now we can specify the semantics of our first example of an augmentation model, called *strict augmentation model*.

**Definition 11 — Strict Augmentation Model**

Let $doc, doc' \in \mathcal{DOC}$. Let $strict \in \mathcal{AUG}$ denote the strict augmentation model. Let $doc' = aug(doc, strict)$.

Then $doc'$ satisfies the following condition:

$$\forall m \in M(doc) : \forall m' \in alts(doc, doc', m) :$$
$$m' \simeq_{\{subjects, duration\}} m \wedge$$
$$infSync(m, m', subjects(m))$$

The semantics of the strict augmentation model is that each alternative $m'$ in the augmented document $doc'$ covers the same subjects and is of the same length as the original element $m$ in document $doc$. Additionally, the media elements $m$ and their corresponding alternatives $m'$ need to be information-synchron with respect to all associated subjects. In consequence, this augmentation model can lead to augmented documents that contain atomic media elements of different technical quality as alternatives, e.g., a talk alternatively encoded as 8-bit audio and 16-bit audio. The augmented document can also contain, e.g., two videos covering the same content but captured from different camera perspectives. The document may also contain alternative media elements of different type, e.g., a video and a slide show. All such alternatives are available in the document $doc'$ only if $infSync$ holds, i.e., each $m'$ covers the same subjects as $m$, and $m'$ and $m$ are of the same length. Note that the strict augmentation model always considers *all* subjects assigned to individual media elements, although some subjects may not be of any relevance in the context of a particular presentation. Figure 3 shows a small part of a tree-structured multimedia document. Video $v_1$ is arranged in parallel ($par$) with other media elements. When augmented under the strict augmentation model, this part of the document is extended by a kind of $switch$ element that inserts video $v_2$ as an alternative for $v_1$. The available alternatives are those presented in Figure 2.

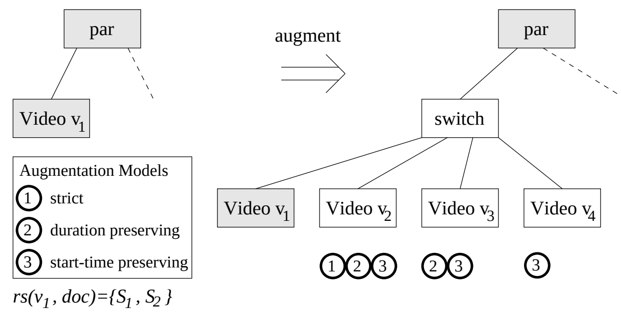


Figure 3: An augmentation example

As the strict augmentation model is a very restrictive model for later adaptation, one might wish to relax the constraints of this model. As mentioned above, not all subjects associated to individual media elements used by an author are necessarily relevant in the context of a presentation. Therefore, one could give the author the possibility to identify those subjects that are relevant to this special presentation (Definition 12). If an author adds this kind of relevance information to media elements, this annotation can be exploited by less restrictive augmentation models. For example, for the augmentation one could require media-equivalence regarding only the subjects relevant in a particular presentation context instead of media-equivalence considering all subjects. As a consequence, more media elements will be considered to become alternatives. It is also possible to relax the $infSync$ condition and demand this constraint only for media elements that have been specified to be played out in lip-synchronization mode (Definition 13). Taking these relaxations into account, we can define a more relaxed augmentation model, the *duration-preserving augmentation model* (Definition 14).

**Definition 12 — Relevant Subjects** $rs$

$rs : \mathcal{M} \times \mathcal{DOC} \to 2^{\mathcal{SUBJ}}$ returns the set of subjects that is relevant to a given media element $m \in \mathcal{M}$ in the context of a multimedia document $doc \in \mathcal{DOC}$. The set of relevant subjects $rs(m, doc)$ is a subset of $subjects(m)$ (see Definition 2).

**Definition 13 — Lip-Synchronization Predicate** *lipsync*

Let $doc \in \mathcal{DOC}, m \in M(doc)$.

*lipsync*$(m, doc)$ is a predicate that holds if $m$ has to be played out in lip-synchronization with another media element in a given document $doc \in \mathcal{DOC}$.

**Definition 14 — Duration-Preserving Augmentation Model**

Let $doc, doc' \in \mathcal{DOC}$. Let *duration-preserving* $\in \mathcal{AUG}$ denote the duration-preserving augmentation model. Let $doc' = aug(doc,$ *duration-preserving*$)$.

Then $doc'$ satisfies the following conditions:

$$\begin{aligned}&\forall m \in M(doc) : \forall m' \in alts(doc, doc', m) :\\ &\quad m' \simeq_{\{duration\}} m \wedge\\ &\quad subjects(m') \supseteq rs(m, doc) \wedge\\ &\quad (lipsync(m, doc) \Rightarrow\\ &\qquad infSync(m, m', rs(m, doc)))\end{aligned}$$

With the duration-preserving augmentation model, the alternatives for a media element $m$ must cover only the relevant subjects of $m$. Additionally, the information-synchronism condition has been relaxed. The consequence of this model for the resulting documents is that cross-media adaptation is possible in many places as the $infSync$ condition applies only for media elements that have been specified to be presented in lip-synchronization with another media element. With this condition, it is ensured that if an author has explicitly modeled two media elements, e.g., a video and an audio, to be presented in lip-synchronization, this condition still holds after the augmentation with alternatives. Figure 3 shows, that under the duration-preserving augmentation model the document is augmented by the videos $v_2$ and $v_3$.

The duration-preserving augmentation model still requires the same duration of alternative media elements. This condition can be relaxed, but reasonably only so far that the starting point in time of each media element in the presentation still stays the same. By this, the adapted presentation keeps the general temporal course intended by the author. Again, this relaxation increases the number of media elements that are suitable alternatives during the augmentation. To allow for this relaxation, it is not sufficient to consider the single media elements when looking for alternatives but also their temporal "neighbours". Therefore, we introduce the triggering relation $\prec$ that relates a media element with those that are temporally adjacent, and the definition of the set of triggering media $M_{trig}$ that collects those media elements that temporally precede the media elements of a document (Definition 15).

**Definition 15 — Triggering Relation $\prec$, Triggering Media $M_{trig}$**

Let $doc \in \mathcal{DOC}, m_1, m_2 \in M(doc), m_1 \neq m_2$.

$m_1 \prec m_2$ holds if and only if the end of the presentation of media element $m_1$ triggers the start of the presentation of media element $m_2$.

$M_{trig}(doc) = \{m \in M(doc) \mid \exists m' \in M(doc) \text{ with } m \prec m'\}$

$M_{trig}(doc)$ describes the set of presentation elements that occur on the left hand-side of the *Triggering Relation* and that are part of the document *doc*.

With the notion of triggering media, we can now define a third augmentation model, the *start time-preserving augmentation model* (Definition 16). This model relaxes the condition that the presentation alternatives need to have the same duration as the original media elements. With this third model, only the start times of each media element must stay the same, i.e., whatever alternative is selected, its starting time is the same as that of the original media element. As this is valid for each media element, the general temporal course of the original presentation is maintained. The start of each media element depends on the end of the presentation of its triggering media element. Therefore, the alternative must adopt the role of the triggering media element and keep the presentation duration of the original media element. Figure 3 shows, that under the start time-preserving augmentation model the document is augmented by the videos $v_2, v_3$, and $v_4$.

**Definition 16 — Start Time-Preserving Augmentation Model**

Let $doc, doc' \in \mathcal{DOC}$. Let *st-preserving* $\in \mathcal{AUG}$ denote the start time-preserving augmentation model. Let $doc' = aug(doc,$*st-preserving*$)$.

Then $doc'$ satisfies the following conditions:

$$\begin{aligned}&\forall m \in M(doc) : \forall m' \in alts(doc, doc', m) :\\ &\quad subjects(m') \supseteq rs(m, doc) \wedge\\ &\quad (lipsync(m, doc) \Rightarrow infSync(m, m', rs(m, doc)))\\ &\forall m \in M_{trig}(doc) : \forall m' \in alts(doc, doc', m) :\\ &\quad m \simeq_{\{duration\}} m' \ \wedge m' \in M_{trig}(doc')\end{aligned}$$

Obviously, one can think of many different augmentation models comprising a specific set of constraints that formally define a specific quality of adaptation. Of course, it is necessary that the constraints founding an augmentation model are very well chosen in order to get exactly the notion for adaptation an application needs.

## 4 Dynamic Adaptation — Substitution of Media Elements

The initial adaptation reduces the augmented document by those presentation alternatives that would never be selected for presentation in the current user context. Note that the original media elements of the document are not removed by the initial adaptation. The augmented and initially adapted document is subject to dynamic adaptation and presentation. The first time the dynamic adaptation is carried out, it operates on the initially adapted document. In this, per definition and to stay close to the original, all original media elements are selected for presentation. Undergoing the dynamic adaptation, alternatives for those media elements are selected for presentation that, given the current user context, cannot be presented. Then the rendering of the presentation starts. Whenever the user context changes, e.g., the available resources decrease, a further dynamic adaptation is initiated which again substitutes those media elements currently selected by those alternatives better matching the newly changed user context. This presentation-adaptation-loop continues as long as the presentation is rendered and there are changes in the user context detected (Figure 1). The dynamic adaptation selects those out of the available alternatives that both best match the current user context and fulfil the constraints of the respective substitution model.

A *substitution model* is used to control the dynamic adaptation process. It comprises a set of constraints that determine which potential alternatives are actually admitted for presentation. Similar to the augmentation models, we introduce substitution models that cover different levels of strictness. By selecting a substitution model, a user can determine how strict he wants the substitution of media elements to be carried out.

Before we introduce the substitution models, we give some basic definitions.

### 4.1 Basic Definitions

During the presentation of a document, additional information must be maintained to reflect the current presentation state. This includes information about the current presentation time, the selected alternatives, and the like. The association of the presentation of the document with the presentation state is called a *presentation plan*. Each plan is an instance or snapshot of the presentation at a specific point in time.

The symbols introduced in Definition 17 are used in the formal definitions to follow.

**Definition 17 — Symbols:** $\mathcal{PSTATE}, \mathcal{PLAN}, \mathcal{UC}$

Let $\mathcal{PSTATE}$ denote the set of all possible states of a running presentation. An element of this set contains all information that is related to the current presentation state of a document.

Let $\mathcal{PLAN} = \mathcal{DOC} \times \mathcal{PSTATE}$ denote the set of all presentation plans. A plan associates a multimedia document with its current presentation state.

Let $\mathcal{UC}$ denote the set of all possible user contexts. An element of this set contains all necessary information to represent a user's preferences, system environment, and current system state.

For the substitution models, we distinguish the substitution of those media elements that are currently presented (*active*) from those that are presented in the future or already have been presented in the past (*inactive*). We need this distinction as for the substitution of active media elements stricter conditions must hold than for the inactive elements. Therefore, we introduce the following definition.

**Definition 18 — Active Media:** $isActive$, $M_{active}$

Let $doc \in \mathcal{DOC}$, $m \in M(doc)$, $pstate \in \mathcal{PSTATE}$, and $pl = [doc, pstate] \in \mathcal{PLAN}$.

$isActive(m, pl)$ is a predicate that holds if the media element $m$ is currently rendered within the presentation of $doc$.

$M_{active}(pl) = \{m \in M(doc) \mid isActive(m, pl)\}$
$M_{active}(pl)$ denotes the set of media elements $m \in M(doc)$ that are currently rendered within the presentation of $doc$.

### 4.2 Substitution Models

A substitution model is used to control the switching of currently selected media elements to one of their alternatives. It comprises a set of constraints that determines whether a proposed substitute can actually be chosen or not. The substitution models cover different levels of strictness each of which stands for a distinctive quality of dynamic adaptation. Let $\mathcal{SUBST}$ denote the set of the different substitution models. The selected substitution model serves as input to the dynamic adaptation process defined by the function *adapt* given in Definition 19.

**Definition 19 — Adaptation** $adapt$

$adapt : \mathcal{PLAN} \times \mathcal{UC} \times \mathcal{SUBST} \rightarrow \mathcal{PLAN}$ selects those media elements $m \in M(doc)$ in a current presentation $pl = [doc, pstate]$ that are not appropriate anymore to the current user context $uc \in \mathcal{UC}$ and substitutes each of those with that alternative that best matches $uc$ under the control of a given substitution model $sm \in \mathcal{SUBST}$.

The function *adapt* changes the presentation state of a current presentation plan $pl$ which maintains the currently selected media elements in $pstate$. The only requirement to an implementation of *adapt* is, that the given substitution model is not violated. How good the function *adapt* calculates the best matching alternative is out of the scope of this paper.

For the distinction of media elements and their substitutes we introduce the following definition.

**Definition 20 — Substitute** $subst$

Let $doc \in \mathcal{DOC}$, $pstate, pstate' \in \mathcal{PSTATE}$, $pl, pl' \in \mathcal{PLAN}$, $pl = [doc, pstate]$, and $pl' = [doc, pstate']$. Let $uc \in \mathcal{UC}, sm \in \mathcal{SUBST}$. Let $pl' = adapt(pl, uc, sm)$.

$$subst : \mathcal{M} \times \mathcal{PLAN} \times \mathcal{PLAN} \rightarrow \mathcal{M}$$

$subst$ is a function that returns that media element $m' = subst(m, pl, pl')$ that has substituted the formerly active media element $m \in M(doc)$ after adapting $pl$ to $pl'$. If no substitution has taken place, $m$ itself is returned.

Now we can specify the semantics of our first example of a substitution model, called *strict substitution model*.

**Definition 21 — Strict Substitution Model**
Let $doc \in \mathcal{DOC}$, $pstate, pstate' \in \mathcal{PSTATE}$, $pl, pl' \in \mathcal{PLAN}$, $pl = [doc, pstate]$, and $pl' = [doc, pstate']$, $uc \in \mathcal{UC}$. Let $strict \in \mathcal{SUBST}$ denote the strict substitution model. Let $pl' = adapt(pl, uc, strict)$.

Then $pl'$ satisfies the following conditions:

$$\forall m \in M(doc) \text{ with } subst(m, pl, pl') \neq m :$$
$$subst(m, pl, pl') \simeq_{\{subjects, duration\}} m \wedge$$
$$infSync(m, subst(m, pl, pl'), subjects(m))$$

If a user selects this substitution model for the dynamic adaptation, it assures that only those alternatives are selected for a substitution that preserve the information flow. This is a very restrictive substitution model but defines that the adapted presentation stays very close to the original. If the strict substitution model is employed, the dynamic adaptation ensures that the substitute $m'$ for a media element $m$ is information-synchron with $m$. Here, one of the differences between augmentation models and substitution models becomes clear. Whereas the augmentation places the information-synchronization relationship on the *original* media element and their alternatives, for the substitution this relationship must hold between the *selected* media element and potential substitutes.

Figure 4 gives an example of the substitution of media elements of an augmented multimedia document. The participating videos are those introduced in Figure 2, and video $v_1$ is currently selected for presentation. On dynamic adaptation under the strict substitution model only video $v_2$ is a valid substitute for $v_1$. With this model substitution of active and inactive media elements are treated the same way.

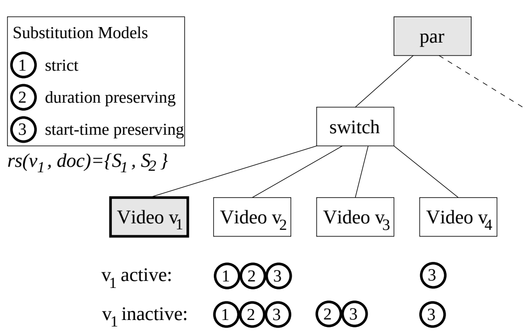


Figure 4: Sample candidates for substitution

However, to allow for a less strict adaptation the *duration-preserving substitution model* relaxes the information-synchronism condition (Definition 22).

**Definition 22 — Duration-Preserving Substitution Model**
Let $doc \in \mathcal{DOC}$, $pstate, pstate' \in \mathcal{PSTATE}$, $pl, pl' \in \mathcal{PLAN}$, $pl = [doc, pstate]$, and $pl' = [doc, pstate']$, $uc \in \mathcal{UC}$. Let $duration\text{-}preserving \in \mathcal{SUBST}$ denote the duration-preserving substitution model.
Let $pl' = adapt(pl, uc, duration\text{-}preserving)$.

Then $pl'$ satisfies the following conditions:

$$\forall m \in M(doc) \text{ with } subst(m, pl, pl') \neq m :$$
$$subst(m, pl, pl') \simeq_{\{duration\}} m \wedge$$
$$subjects(subst(m, pl, pl')) \supseteq rs(m) \wedge$$
$$((lipsync(m, pl) \vee isActive(m, pl))$$
$$\Rightarrow infSync(m, subst(m, pl, pl'), rs(m)))$$

With this substitution model, another difference between augmentation and substitution models becomes obvious. For the substituted media elements, the model distinguishes between active and inactive media elements. Whereas the conditions for potential alternatives for the substitution have been relaxed for inactive media elements, the information-synchronism relationship still must hold for the active media elements and potential substitutes. This gives the user a less restrictive model for the substitution of media elements as more alternatives can be considered. But the model still ensures smooth changes for all active media elements involved in the substitution.

Figure 4 illustrates the potential substitutes for video $v_1$ under the duration-preserving substitution model. If $v_1$ is currently active, only video $v_2$ is a relevant substitute. In the case that $v_1$ is currently inactive, also video $v_3$ is a candidate for dynamic adaptation. Though video $v_4$ is $infSync$ with $v_1$, it is not a candidate as it is not duration-preserving.

To relax the duration-preserving substitution model but still to preserve the starting point in time of each media element, we define the *start time-preserving substitution model*.

**Definition 23 — Start Time-Preserving Substitution Model**
Let $doc \in \mathcal{DOC}$, $pstate, pstate' \in \mathcal{PSTATE}$, $pl, pl' \in \mathcal{PLAN}$, $pl = [doc, pstate]$, and $pl' = [doc, pstate']$, $uc \in \mathcal{UC}$. Let $st\text{-}preserving \in \mathcal{SUBST}$ denote the start time-preserving substitution model.
Let $pl' = adapt(pl, uc, st\text{-}preserving)$.

Then $pl'$ satisfies the following conditions:

$$\forall m \in M(doc) \text{ with } subst(m, pl, pl') \neq m :$$
$$subjects(subst(m, pl, pl')) \supseteq rs(m) \wedge$$
$$((lipsync(m, pl) \vee isActive(m, pl))$$
$$\Rightarrow infSync(m, subst(m, pl, pl'), rs(m)))$$
$$\forall m \in M_{trig}(doc) \text{ with } subst(m, pl, pl') \neq m :$$
$$m \simeq_{\{duration\}} subst(m, pl, pl') \wedge$$
$$subst(m, pl, pl') \in M_{trig}(doc)$$

Figure 4 illustrates especially that the start time-preserving substitution model admits video $v_4$ for substitution as it is $infSync$ with video $v_1$ with respect to the relevant subjects $S_1$ and $S_2$ and its duration does not violate the triggering media relationship. If $v_1$ is currently active, only video $v_2$ is an additional relevant substitute. $v_3$ becomes a candidate only if $v_1$ is inactive.

## 5 Conclusion

In this paper, we have presented a cross-media adaptation scheme that serves as the basis for implementing adaptation of multimedia presentations offered by the online multimedia information system of the Cardio-OP project. The paper focuses on the modular design of the adaptation scheme in terms of augmentation and substitution models that describe the identification and selection of media of different type and even different fragments of a presentation as presentation alternatives for a given user context. This approach goes far beyond those techniques that focus on the provision of alternative media qualities depending on actual system performance. The criteria that can be captured by the augmentation and substitution models allow for various dimensions concerning the quality of adaptation.

The impact of our approach is manifold: Explicit modeling of the structural composition of multimedia presentations is a prerequisite in order to allow for the reuse of media elements and fragments [3], but this becomes available in the context of document models like ZYX [1], CMIF [4], TIEMPO [17], SMIL [9], and, e.g., other XML-based models.

For the authoring process automatic, constraint checked augmentation relieves the author from explicitly, comprehensively specifying *all* alternative content for different environments. This obviously has an impact on the costs of authoring multimedia presentations for different environments. The specification of constraints in terms of different augmentation models allows for the determination of different "degrees" of adaptability. By selecting a particular augmentation model one can determine how "far away" a presentation potentially adapted later on can be from the original presentation.

Users can determine the kind of dynamic adaptation by selecting a particular substitution model: The less restrictive a substitution model, the more powerful the adaptation to the current user context. However, a user still could select a very strict substitution model, forcing the presentation to stay very close to the original one and by this possibly abandon a smooth presentation flow.

Adaptation techniques might be of importance for future billing models: Augmentation as well as substitution models can be used either by content providers or by consumers of multimedia presentations to select a pricing category the system is operating in. In this paper the models define the upper limit of the adaptability of the document. However, one could also realize these models in terms of the level of desired adaptability. Then, a user could select a very "cheap" augmentation or substitution model that allow for adaptation of all high-quality media elements by low quality and low data volume media elements and, as a consequence, being charged based on lower price.

### Acknowledgments

We would like to thank Christian Heinlein and Utz Westermann for their valuable comments on the paper.

## References

[1] S. Boll and W. Klas. ZYX — A Semantic Model for Multimedia Documents and Presentations. In *Proc. Data Semantics (DS-8): "Semantic Issues in Multimedia Systems"*. Kluwer Academic Publishers, Rotorua, NZ, Jan 1999.

[2] S. Boll, W. Klas, and U. Westermann. Exploiting OR-DBMS Technology to Implement the ZYX Data Model for Multimedia Documents and Presentations. In *Proc. of Datenbanksysteme in Büro, Technik und Wissenschaft (BTW99), GI-Fachtagung*, Freiburg, Germany, March 1999. Springer.

[3] S. Boll, W. Klas, and U. Westermann. Multimedia Document Formats — Sealed Fate or Setting Out for New Shores? In *Proc. ICMCS'99*, volume 1, pages 604–610, Florence, Italy, June 1999.

[4] D. C. A. Bulterman. User-centered Abstractions for Adaptive Hypermedia Presentations. In *Proc. ACM Multimedia'98*, Bristol, UK, Sep 1998.

[5] D.C.A. Bulterman, L. Rutledge, L. Hardman, and J. van Ossenbruggen. Adaptive and Adaptable Presentation Semantics. In *Proc. Data Semantics (DS-8): "Semantic Issues in Multimedia Systems"*. Kluwer Academic Publishers, Rotorua, NZ, Jan 1999.

[6] H. Chen, A. Krishnamurthy, T.D.C. Little, and D. Venkatesch. A Scalable Video-on-Demand Service for the Provision of VCR-like Functions. In *Proc. ICMCS'95*, Washington DC, USA, May 1995.

[7] M. Chen, D.D. Kandlur, and P.S. Yu. Support for Fully Interactive Playout in a Disk-Array-based Video Server. In *Proc. ACM Multimedia'94*, San Francisco, USA, Oct 1994.

[8] S. Hollfelder, A. Kraiss, and T.C. Rakow. A Client-Controlled Adaptation Framework for Multimedia Database Systems. In R. Steinmetz and L.C. Wolf, editors, *Proc. IDMS'97*, pages 397–409, Darmstadt, Germany, September 10–12, 1997. Springer.

[9] P. Hoschka, S. Bugaj, D. Bulterman, et al. *Synchronized Multimedia Integration Language – W3C Working Draft 2-February-98*. W3C, URL: http://www.w3.org/TR/1998/WD-smil-0202, Feb 1998.

[10] W. Klas, C. Greiner, and R. Friedl. Cardio-OP — Gallery of Cardiac Surgery. In *Proc. ICMCS'99*, volume 2, pages 1092–1095, Florence, Italy, June 1999.

[11] Rakesh Mohan, John R. Smith, and Chung-Shen Li. Adapting Content to Client Resources in the Internet. In *Proc. ICMCS'99*, volume 1, pages 302–307, Florence, Italy, June 1999.

[12] F. Moser, A. Kraiß, and W. Klas. L/MRP: A Buffer Management Strategy for Interactive Continuous Data Flows in a Multimedia DBMS. In *Proc. VLDB 1995*, USA, 1995. Morgan Kaufmann.

[13] P.J. Shenoy and H.M. Vin. Efficient support for scan operations in video servers. In *Proc. ACM Multimedia'95*, San Francisco, Nov 1995.

[14] H. Thimm, W. Klas, J. Walpole, C. Pu, and C. Cowan. Managing Adaptive Presentation Executions in Distributed Multimedia Database Systems. In *Proc. Int. Workshop on Multi-Media Database Management Systems*. Blue Mountain Lake, NY, USA, IEEE Computer Society Press, Aug 1996.

[15] J. Walpole, C. Krasic, L. Liu, D. Maier, C. Pu, D. McNamee, and D. Steere. Quality of Service for Multimedia Database Systems. In *Proc. Data Semantics (DS-8): "Semantic Issues in Multimedia Systems"*, Rotorua, NZ, Jan 1999. Kluwer Academic Publishers.

[16] J. Wandel. A Framework for an Adaptive Presentation Environment in the Internet Based on ZYX. Diploma Thesis, in German, University of Ulm, Computer Science Department, Databases and Information Systems, Germany, 1998.

[17] S. Wirag. Specification and Scheduling of Adaptive Multimedia Documents. Technical Report 1999/04, University of Stuttgart, Computer Science Department, Germany, 1999.