Intelligent Multimedia Presentation in Ubiquitous Multidevice Scenarios

Ralf Kernchen, Stefan Meissner, and Klaus Moessner
Centre for Communication Systems Research, University of Surrey, UK

Pablo Cesar and Ishan Vaishnavi
Centrum Wiskunde & Informatica (CWI), Netherlands

Matthieu Boussard
Alcaltel-Lucent Bell Labs, France

Cristian Hesselman
Novay, Netherlands

This intelligent multimedia adaptation and delivery framework tailors to ubiquitous environments, so that users can experience multimedia content using multiple devices in various mobility situations.

Traditionally, multimedia consumption has been a static and monolithic activity with users consuming content on a single device. Consumption is increasingly movable, because of the advent of multimedia-capable mobile devices and truly networked environments.[1] It makes sense, then, that users desire (and increasingly require) better collaboration between different devices.[2]

We can identify two major trends on movable and context-aware multimedia research: dynamic generation of multimedia presentations[3-5] and network-based session mobility.[6-8] The first body of research arises from the multimedia community and focuses on the delivery of rich multimedia content in a specific context. Nevertheless, these solutions don't consider the problem of dynamic continuity across devices and they don't use advanced device-description standards. The second body of research originates from the network community, often focusing on telephony and video conferencing. They provide solutions for device discovery, dynamic session transfer, and video adaptation based on transcoding. However, their approaches often have a restrictive view on multimedia content as a simple or multiplexed continuous stream.

Keeping this mind, we created an intelligent multimedia adaptation and delivery framework that combines the best of both worlds. An implementation of the framework is the open source initiative named Multimodal Delivery and Control System (MDCS; see http://sourceforge.net/projects/mdcs/). The goal of MDCS is to permit dynamic and seamless adaptation of interactive multimedia according to the user's context, resulting in a richer experience that leverages the powerful constructs provided by both the structured multimedia research and network communities. For example, when users are on the move, they can order movies or TV shows on their mobile phones, but when they reach home, MDCS can adapt the same content and render it to an available high-definition television (HDTV) in the living room while the movie-controlling interface remains on the mobile phone. We came up with this approach in part after reading O'Hara et al.'s[9] report that users find it frustrating to view video material on mobile devices and how they would like to see better device integration.

Our system realizes a generalized adaptation and delivery process backed by the functionalities defined in its multimedia presentation adaptation and delivery framework. The process scores and ranks all feasible presentation options and device combinations depending on user preferences and context information. Finally, the system renders the device-tailored presentations in a distributed way to the multidevice environment so that the user can experience better device integration.

Ubiquitous multimedia consumption

To illustrate the relevance and potential of our research in future multimedia adaptation

1070-986X/10/$26.00 © 2010 IEEE

Published by the IEEE Computer Society

systems, we chose a representative mobility scenario that shows the general features for ubiquitous multimedia presentation consumption in multidevice environments. The scenario revolves around Maria, who is on a business trip and staying in a hotel.

Figure 1 shows an overview of the scenario. Through MDCS, Maria receives a personalized recommendation for a documentary about Africa (her next holiday destination) on her mobile phone. She selects it using her personalized content guide (see Figure 1a), which allows her to access rich multimedia presentations.

Her mobile phone then notifies Maria of possible active devices in her surroundings—such as an HDTV and high-fidelity (hi-fi) audio equipment in her hotel room. Selecting the documentary triggers the intelligent multimedia presentation delivery process. The presentation contains video and scene information (images and text). Based on the available devices and Maria's personal preferences, the system decides to play the video content on the TV, while making image scene information available on the mobile device (see Figure 1b). The MDCS transforms as well the accompanying text to speech—and plays it in the hi-fi system synchronized with the video rendered.

When Maria decides to leave the room (which the device reads as a context change) a dynamic evaluation of available devices will take place, resulting in a new delivery decision while keeping media continuity (see Figure 1c). The system then transfers media content from the TV to the mobile device, which involves transforming the content to scale for the smaller device.

With this goal of providing continuous, while dynamic, multimedia experiences that use the most appropriate devices surrounding the user, we can identify a number of fundamental challenges:

- We need to define a suitable adaptation and delivery process that incorporates presentation-handling functions—namely presentation scheduling, media item transformations, device and presentation matching, contextual personalization, and device-tailored content delivery.

- The system must offer high-quality user experiences in every presentation and device configuration. It must accurately discover devices, offer an appropriate description of the devices' characteristics, match media item properties, and provide multimedia continuity in an integrated fashion.

- Personalization and user preferences are important factors for the usability and acceptance of such complex systems. Usage must be efficient in different contexts and personalization should be automatic—yet controllable—for the user. We propose applying learning techniques based on context parameters, to minimize configuration efforts.

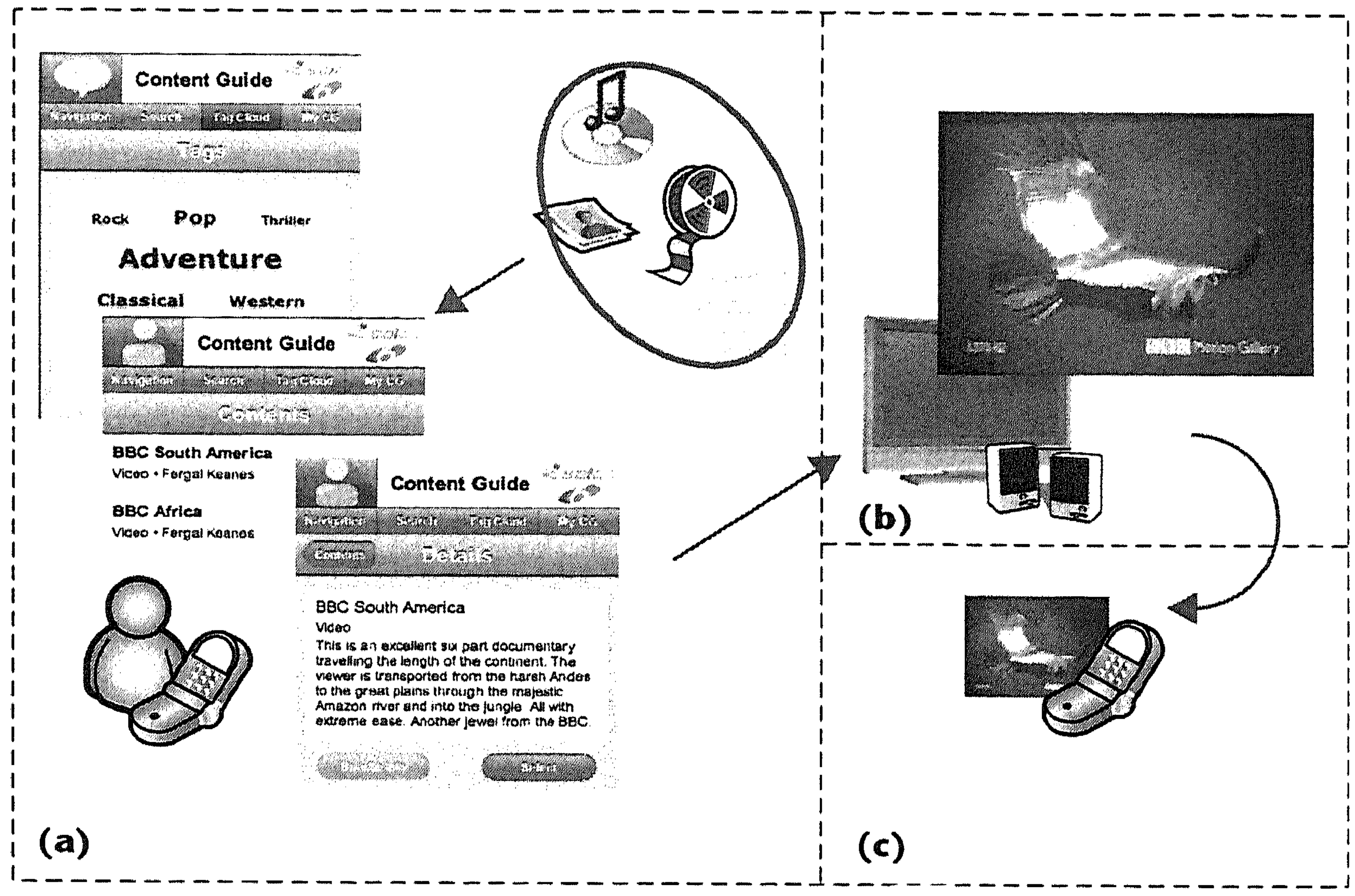

Figure 1. An intelligent presentation delivery scenario for a multimedia content guide: (a) content guide selection, (b) synchronized content delivery (for high-quality video on a TV), and (c) session transfer when Maria leaves the hotel room (to low-quality video on the mobile device).

Overall, to overcome these challenges, our novel intelligent multimedia delivery system integrates mobility characteristics for the optimization of multimedia presentation adaptation. It also integrates session continuity, taking different device configuration and personalization features into account.

Intelligence design

As we explain the steps of our multimedia adaptation and delivery process—along with the framework's related functionalities—we should note that we kept the process generic so that it isn't tied to any specific device or multimedia description language, multimedia encoding mechanism, or network infrastructure. We show the six steps that encompass the process in Figure 2 (next page).

Following our hypothetical scenario, we assume that users can select their preferred rich multimedia presentation, in which a description document is normally available. After presentation selection, the multimedia service

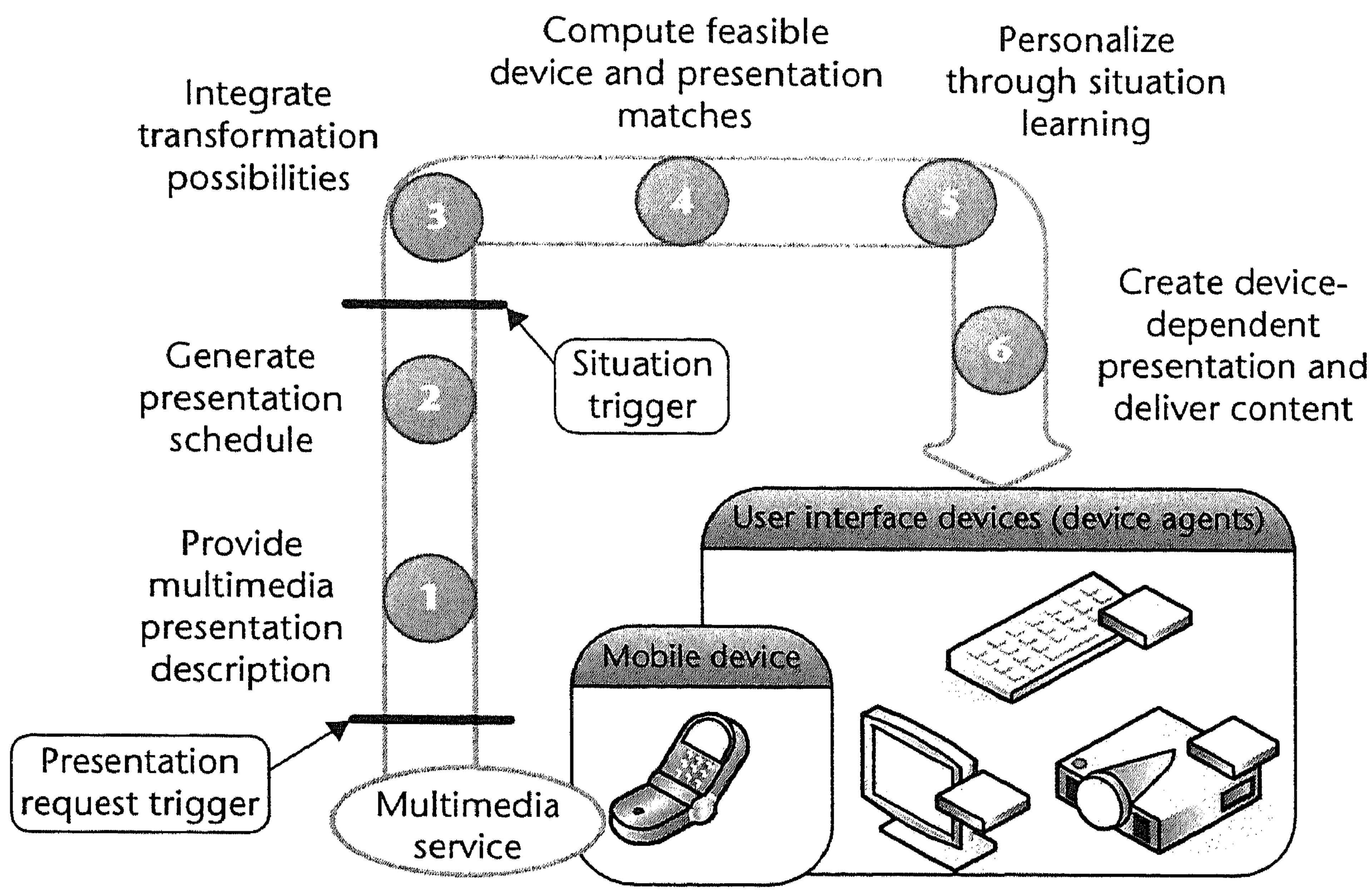

Figure 2. Six steps involved in the multimedia adaptation and delivery process. We kept the process generic so that it will easily work with more devices, languages, encoders, and network infrastructures.

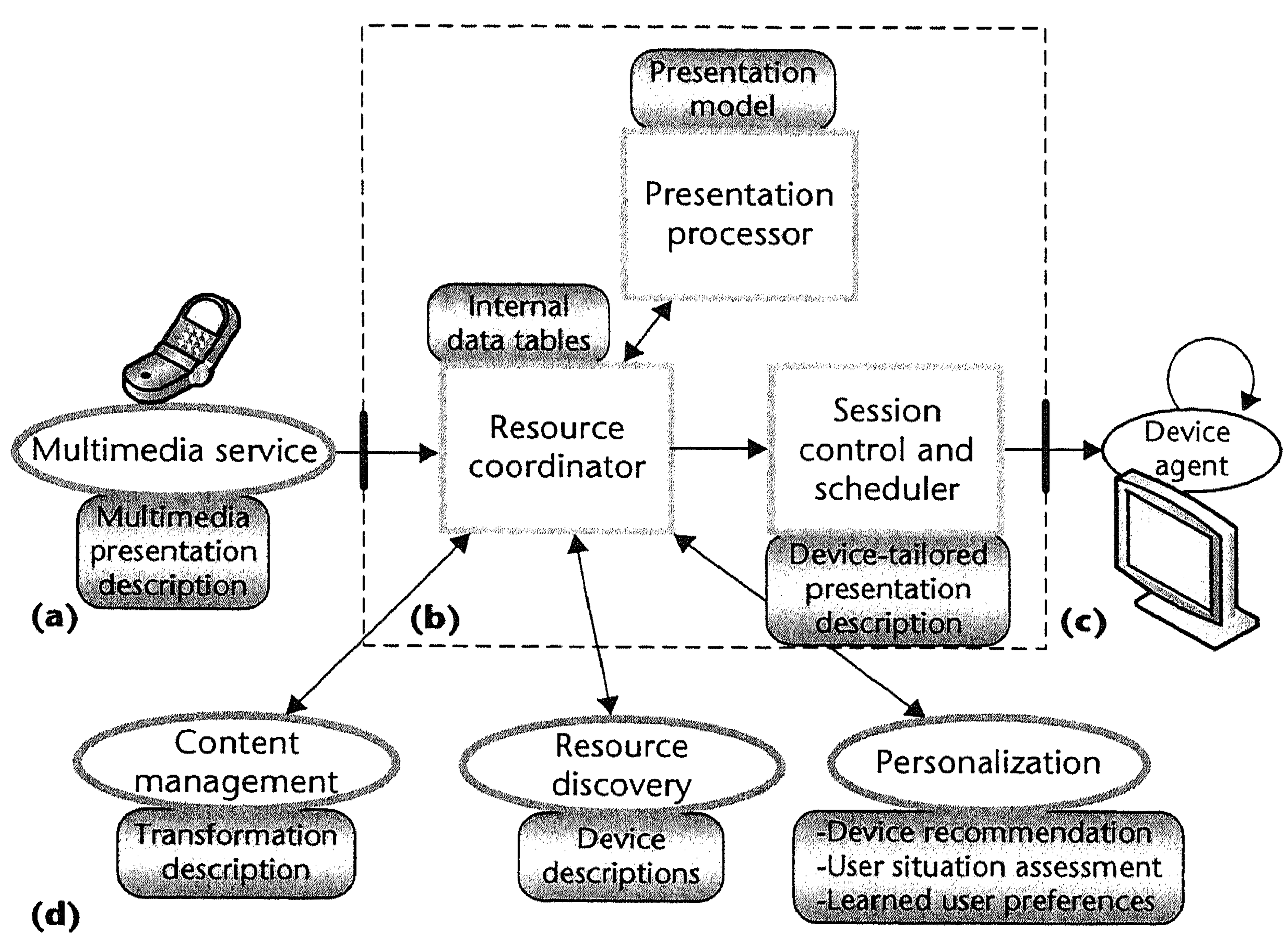

Figure 3. Adaptation functionalities and involved models for our system's framework. (a) A multimedia service requests presentation delivery from (b) the framework's core presentation adaptation components for optimal delivery in (c) a multidevice environment. (d) The delivery process is supported through content management, resource discovery, and personalization.

provides the multimedia description document to the system (step 1). From the description document, the system derives a presentation schedule (step 2). It then identifies available media content transformations (such as media encoding, resizing, or other transformations) and links them to the presentation schedule (step 3), thus increasing the number of devices that can render the content. Description of the active devices and the resulting presentation schedule serve as the basis for computing all feasible presentation and device combinations (step 4).

To achieve an optimal adaptation for a user in different contexts, we propose integrating a personalization mechanism that matches the presentation and device combinations with the user's actual situation (step 5). When the system finds optimal combinations, it creates device-dependent multimedia presentations with the right timing information and delivers them to the selected devices' agents (step 6). The process takes mobility aspects into account, as indicated by the situation trigger. Whenever a user's environment changes, the system reevaluates any active multimedia presentations. It then stores the presentation's session progress and takes this into account for seamlessly continuing the presentation within a different device configuration.

To enable this process, we identified fundamental adaptation and delivery functionalities and mapped them into our framework's design (see Figure 3). In this article, we focus mostly on the framework, but we provide more details about the MDCS implementation elsewhere.[10] In Figure 3a, we can see that the multimedia service accesses the framework via the resource coordinator (see Figure 3b), which is also the presentation request trigger. The resource coordinator governs internal data tables for each step of the process. The other components in Figure 3b provide state-of-the-art technology required to deliver and render multimedia content. A presentation processor is responsible for constructing the internal presentation model and the session controller or scheduler controls the timely progress of the presentation during the actual playing of the presentation.

However, because we intend to provide intelligent multimedia presentation delivery in ubiquitous multidevice scenarios, the system requires additional functionalities (see Figure 3d). According to the process, the content-management component provides a list of content transformation possibilities based on its available transformers and their descriptions. The user's mobile device gathers knowledge about the environment through a resource-discovery component that

Related Web References

For further information on some of the standards, ontologies, and protocols we mention in this article, please visit the following Web sites.

- *W3C's SMIL*: http://www.w3.org/AudioVideo/
- *UPnP Forum's UPnP 1.0*: http://www.upnp.org/specs/arch/UPnP-arch-DeviceArchitecture-v1.0.pdf
- *Open Mobile Alliance's User Agent Profile 2.0*: http://www.openmobilealliance.net/Technical/release_program/docs/UAProf/V2_0-20060206-A/OMA-TS-UAProf-V2_0-20060206-A.pdf)
- *W3C's Composite Capability/Preference Profiles Structure and Vocabularies 2.0 profile for mobile handsets*: http://www.w3.org/TR/2007/WD-CCPP-struct-vocab2-20070430/
- *EU-IST Spice project's mobile ontology description*: http://ontology.ist-spice.org/spice_ontologies_files.htm
- *IETF's SIP*: see http://www.ietf.org/rfc/rfc3261.txt

continuously scans the user's environment using the Session Initiation Protocol (SIP) or other close-range network discovery protocols for available multimedia resources and captures their device capability descriptions. The system later uses the device descriptions for discerning feasible presentation and device combinations and for personalized content delivery. The system assesses the user's situation and presentation and device recommendations before deciding how to deliver the content. User situations can include the user's activity, location, and social situation (for example, whether other people are around).

The system produces device recommendations for specific content based on user preferences (for example, high-quality video at home should be played on an HDTV display instead of a personal device's small screen). Both resource discovery and personalization provide dynamic aspects for the adaptation and delivery process indicated earlier as a situation trigger. In our opinion, these functionalities are mandatory to guarantee the best possible user experience and an optimal multimedia presentation delivery. To simplify the process of defining user preferences and to avoid a manual procedure, we apply learning mechanisms, which we detail later.

Rich multimedia presentation description

Referring back to step 1 of the adaptation and delivery process, we should highlight that there are a number of multimedia presentation formats available and they all inherit different characteristics because of their application area. But what we need is a flexible description approach providing the potential to accommodate most common format characteristics. By looking at temporal models of presentation description formats, we can identify the best candidate. Rogge et al.[11] provide a good evaluation of these formats by introducing their temporal model that includes 29 multimedia interval relationships and making an exhaustive comparison with relevant multimedia description formats. They concluded that only the World Wide Web Consortium's Synchronized Multimedia Integration Language (W3C's SMIL)—a well-known standard for presentation modeling—can represent all 29 multimedia interval relationships sufficiently. For this reason, along with its proven characteristics and flexibility, we chose SMIL as our reference document model in the design and MDCS implementation. (Our "Related Web References" sidebar lists the Web site for SMIL, as well as other standards, ontologies, and protocols mentioned in this article.)

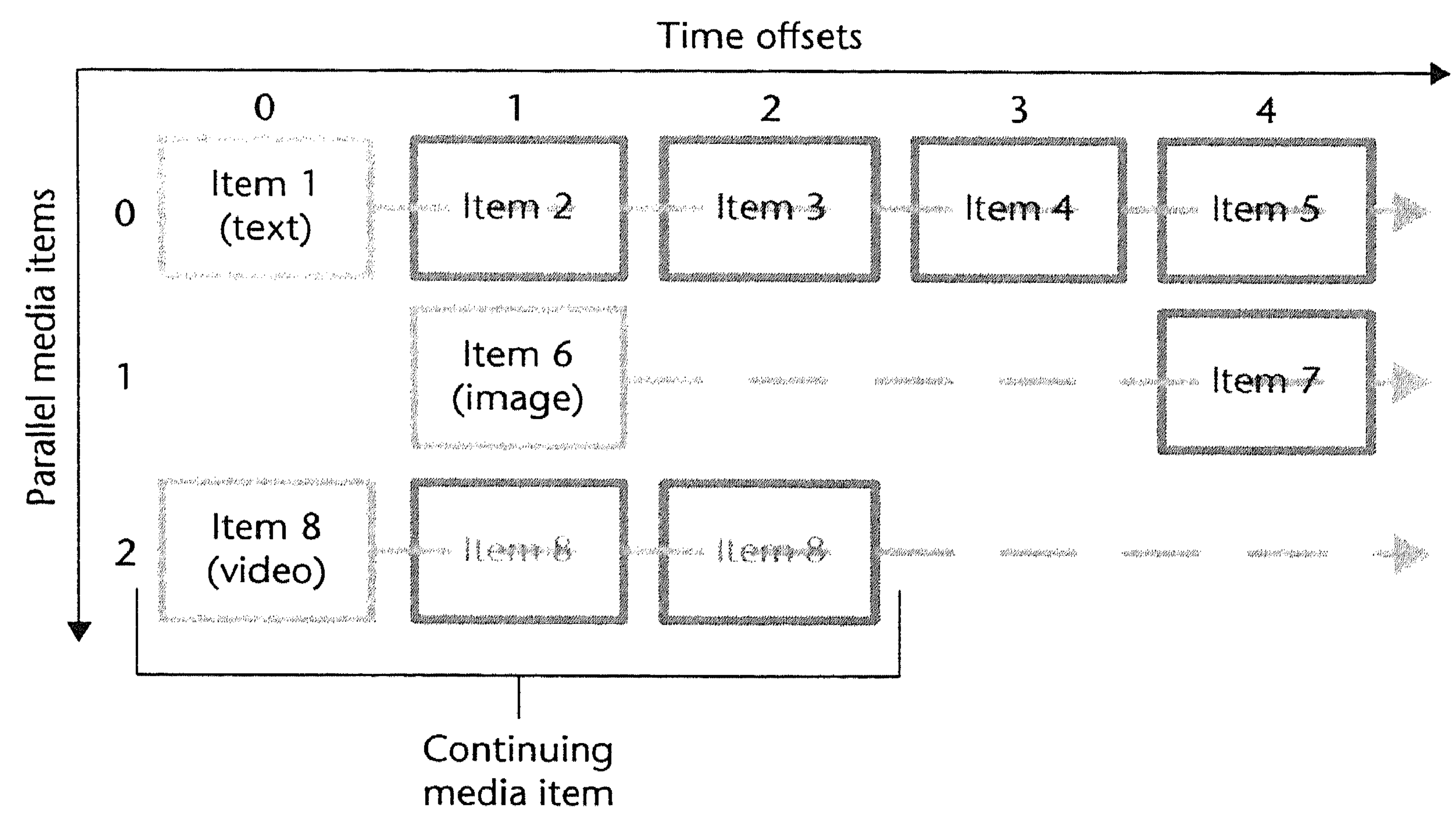

Figure 4. Presentation schedule with media items.

Generating the presentation schedule

For each rich multimedia presentation, the system generates an internal presentation schedule representing temporal and spatial dependencies between contained media items (step 2 of Figure 2). Figure 4 shows some

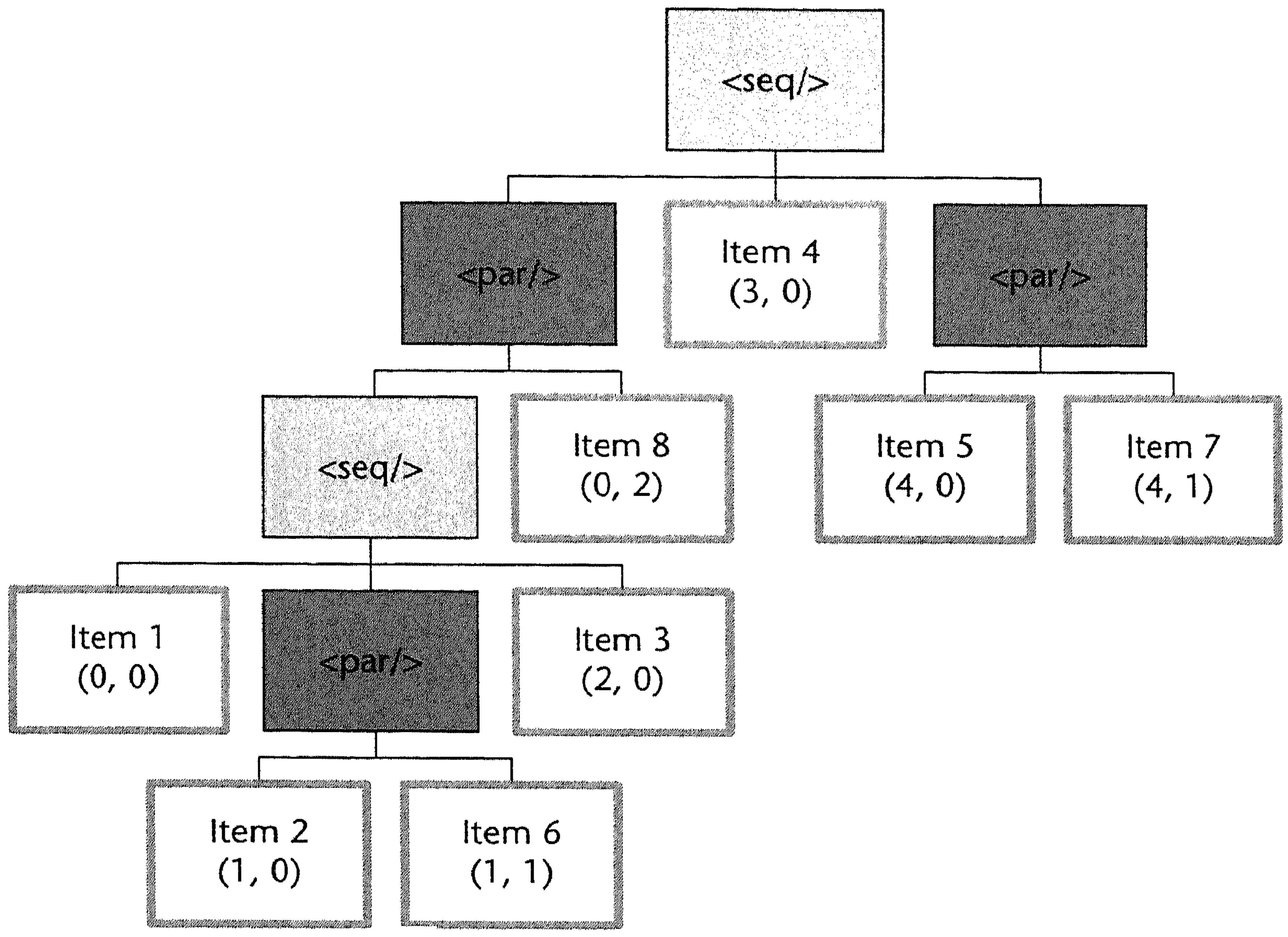

Figure 5. Presentation description tree example. Child elements of <seq/> are time dependent, whereas child elements of <par/> show spatial dependencies.

of the assumptions we made that simplified the process while keeping it general and adequate.

The illustration in Figure 4 shows a number of media items structured within a table, which on the *x*-axis defines the temporal dependencies and on the *y*-axis defines the spatial (media items to be rendered in parallel) dependencies. For example, item 1 must be displayed before item 2, item 2 is parallel to item 6, and item 8 continues during items 1, 2, and 3. As Figure 4 shows, the temporal relationships are based on time offsets and indicated in absolute terms (for example, one media item has to finish before the other starts), which is adequate for representing a movie that's composed of several media clips. In this article, we only consider linear presentations, which don't allow for interaction and therefore won't change the spatial relationships during the presentation time. Nevertheless, we believe that our approach can facilitate nonlinear presentations by breaking more complex presentations down to linear cases if necessary.

Further, we define a base media item (highlighted in green in Figure 4) for each parallel set of items. A base item characterizes the remaining sequence of items, and thus we use it for computing the transformation calculations and the adaptation decision. Media items 1, 6, and 8 (text, image, and video, respectively) are the selected base items in this example.

The system determines the presentation schedule from the presentation description. As we mentioned, we base our work on the SMIL standard. We therefore explain how a SMIL document would be mapped to the presentation schedule. In SMIL, time-dependent media items are so-called sequences and are grouped by the parent element <seq/>. Spatial-dependent media items that should be rendered in parallel are grouped by the parent element <par/>. The combinations of these elements are provided in a presentation description tree. Figure 5 shows an example of a presentation description tree. The numbers in brackets depict the (*x*, *y*) position of the media item in the presentation schedule table shown in Figure 4. Every application output request results in such a tree, with the assumption that the presentation always starts with a <seq/> element.

In step 1 of the process, the multimedia service provides a multimedia presentation (that is, the SMIL presentation description) to the resource coordinator. Then the service generates the presentation tree and prepares to generate the presentation schedule in step 2. In our system, the presentation processor component calculates the schedule table from a SMIL presentation tree using our own recursive scheduling algorithm.

Integrating transformation possibilities

The objective in step 3 is to extend the presentation schedule with media item alternatives. Therefore, the system analyzes the base media items' format information and integrates new format proposals (based on available transformation) into the presentation schedule. In some cases it's beneficial to transform media items from one format to another (for example, video encoding) for the same modality, but in other situations it's better to transform the modality (for example, text-to-speech conversions).

As Figure 3 shows, the content management component receives the transformation request from the resource coordinator together with the base media item descriptions and returns a list of feasible transformations. Therefore, in the case of a transformation request, transformers get a media item and certain conversion parameters as input. If conversion is possible, the system returns a reference to a new media element that fits all the conversion parameters. The system can do the actual conversion on-the-fly or using existing media items from a multimedia database (such as a content repository).

After it calculates the set of alternative media items, the system updates the presentation schedule table depicted in Figure 4 accordingly.

Computing feasible presentation and device matches

In step 4, after the system generates the complete schedule table—including the possible media item transformations—it then computes all feasible presentation and device matches. This means that media items capable of composing the multimedia presentation are compared with the characteristics of available rendering devices.

A prerequisite for this step is finding suitable device descriptions provided by the resource discovery component. There are many device description standards—such as the UPnP Forum's UPnP 1.0, the Open Mobile Alliance's User Agent Profile 2.0, and the W3C's Composite Capability/Preference Profiles Structure and Vocabularies 2.0 profile for mobile handsets. But none can be directly applied to the requirements imposed by multidevice ubiquitous multimedia environments, since each of them addresses one specific environment. Therefore, we defined a technology-independent layer that lets us define ternary relations between media items and devices with their multimedia capabilities. This work was motivated by earlier findings on a device description approach for mobile multimodal interfaces,[12] and later refined and expressed into a complete mobile ontology description within the European Information Service Technologies' Service Platform for Innovative Communication Environment project (EU-IST Spice; see the "Related Web References" sidebar). This in turn allows the integration of different device description models into one processable model.

The inputs for the matching steps are the presentation schedule table—which contains the base and alternative media items—and the device descriptions. For each media item and device, media comparators perform a match.

As Figure 6a shows, a media comparator gets a device description and a media item description as input. It then performs the comparison based on its predefined computation mechanism. This results in the matching-quality coefficient (MC), which on a scale from 0 to 1 provides a measure of how well the media item matches the device (0 = lowest match, 1 = highest match).

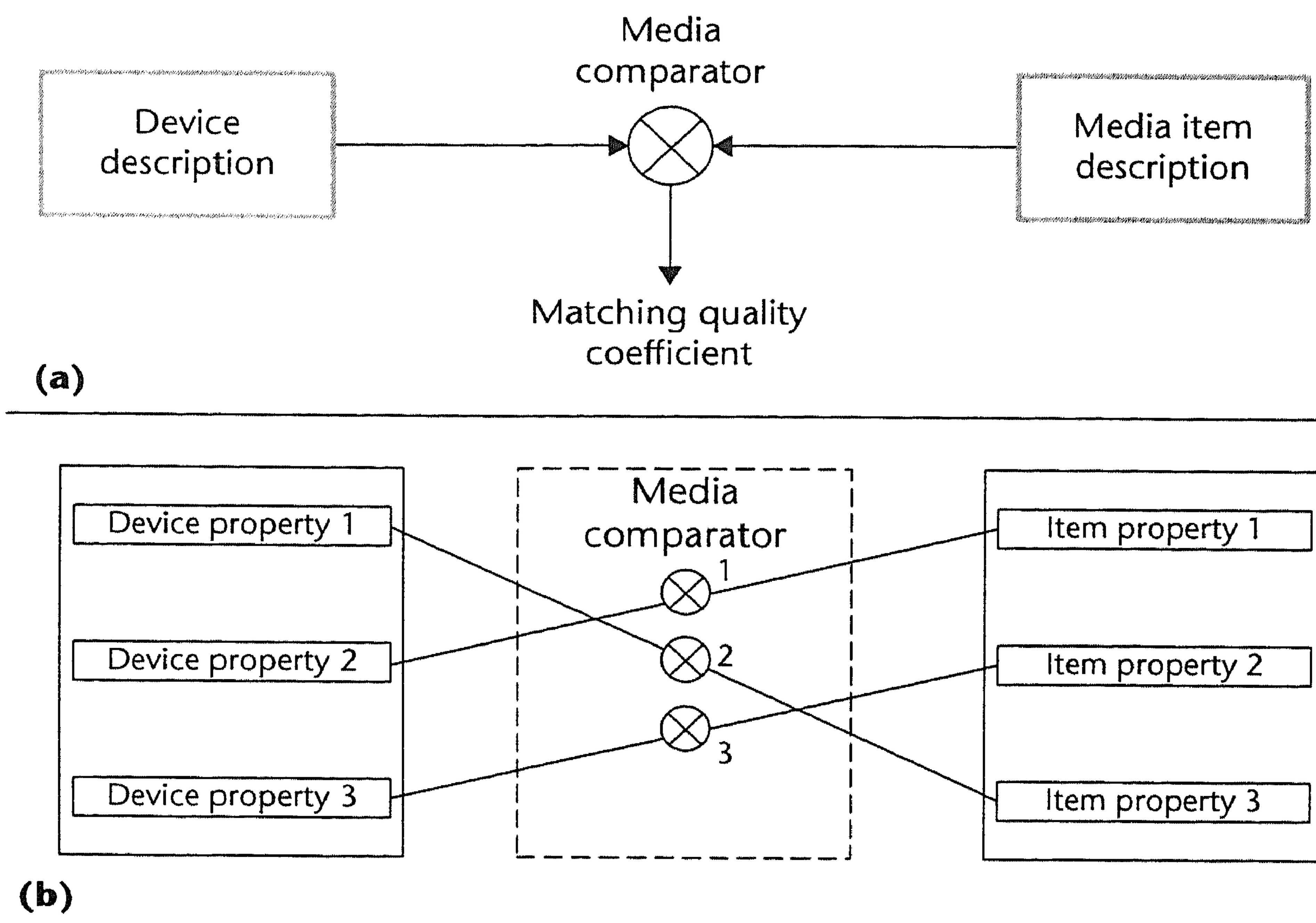

Figure 6. Media item and device comparison mechanism. (a) Overview and (b) property matching.

Figure 6b shows the media comparator's internal behavior. Each media comparator inherits the knowledge of how to weight media item and device property matching. Then it calculates the overall MC. We formalize the compare function in Equation 1:

$$MC(dd, mid) = \frac{1}{n}\sum_{i=1}^{n} dd.property_i \otimes mid.property_i \quad (1)$$

We define the MC by the average values of the compared device description (*dd*) and media item description (*mid*) property matches. In this general approach, we define the comparison of the parameters by the $\otimes$ operator, which stands for the flexibility of the approach to encompass different comparison methods. For example, the operator could be "=" or a defined value range. Equality is important when comparing the device's general modality capability with the media item modality type (for example, a video-capable device can't necessarily display a text or HTML media item). Nevertheless, the final values are recorded within the system's internal data tables. Table 1 (next page) illustrates example results, referring to the presentation schedule shown in Figure 4.

The system assigns every feasible media item and device match an MC. During the presentation transformation in step 3, the content management component provides alternative media items. We grouped media items by base items in Table 1, to highlight their relations. Media items in this example include text, image, and video with different quality values,

Table 1. Example content and device matching.

Base items	Media items for matching	Devices	Matching-quality coefficient
Item 1 (text)	Item 1 (text)	Device 1 (mobile)	0.6
	Item 1 (text)	Device 3 (TV set)	0.8
	Alternative content 1 (text-to-speech)	Device 1 (mobile)	0.8
	Alternative content 1 (text-to-speech)	Device 2 (hi-fi)	0.4
Item 6 (image)	Item 6 (high-resolution image)	Device 3 (TV set)	0.9
	Alternative content 6 (low-resolution image)	Device 1 (mobile)	0.8
Item 8 (video)	Item 8 (high-resolution video)	Device 3 (TV set)	0.7
	Alternative content 8 (low-resolution video)	Device 1 (mobile)	0.5

but they could include 3D graphics or any other rich multimedia items. Devices in this example are a mobile phone, a TV, and a hi-fi audio system.

Next, the system calculates all feasible device–presentation combinations with a combination-quality coefficient (CC). Derived from the example in Table 1, this step will calculate the two combinations (as Table 2 shows). In this example, we only allowed one media item to be played on one device to reduce complexity.

The system derives the CC by calculating the average score of each MC for each possible combination. The results are rated as presentation and device matches. In Table 2 the combination in the first row has a coefficient of 0.6 and the combination in the second row has a coefficient of 0.63, which indicates that the second row provides a slight advantage over the first in terms of user experience quality. Using the CC, the system produces all the feasible presentation and device matches. At that point, the system will include the user's personal preferences, taking the users' current context into account.

Personalizing via situation learning

The system personalizes the adaptation decision in step 5 by applying the user's personal preferences and contextual profile (according to the user's situation) to feasible presentation and device matches. In our approach, user preferences relate to learned rules about the media item, device, and context configurations (the context can be the location, user's activity, physical limitations, and so on). We use an association rule learning approach that we described elsewhere[13] to derive rules that the system can apply to a user's situation. The basis for this is snapshot data, referring to the configurations we mentioned and taken when users confirm a presentation delivery scenario. Later, the system applies the rules and derives the configuration recommendations for a given situation. Learning is a flexible approach that reacts to the user's behavior and changes that occur over time, relieving users of an explicit configuration step. Figure 7 shows an overview of the personalization system's internal components and interactions.

During the personalization process, the resource coordinator requests a recommendation providing the current presentation and device combinations (step 1). The recommender applies the learned rules stored in the personalization system's internal profile to the received request, incorporating current context knowledge (step 2). In step 3, the values returned to the resource coordinator are presentation and device recommendations with a situation confidence coefficient (SC). Using a weighting function, as Equation 2 shows, the system calculates the final confidence coefficient (which for our

Table 2. Feasible presentation and device combinations.

Item 1 (text)	Matching-quality coefficient	Item 6 (image)	Matching-quality coefficient	Item 8 (video)	Matching-quality coefficient	Combination-quality coefficient
Device 2 (hi-fi)	0.4	Device 3 (TV set)	0.9	Device 1 (mobile)	0.5	0.6
Device 2 (hi-fi)	0.4	Device 1 (mobile)	0.8	Device 3 (TV set)	0.7	0.63

purposes is also the final device and presentation combination, or FC) from the SC and CC.

$$FC = w_1 CC + w_2 SC \quad (2)$$

We can alter weights for fine tuning between better presentation quality or user situation adaptation, but the sum of w_1 and w_2 has a maximum of 1. For our example, we applied an equal distribution of 0.5 for w_1 and w_2. Table 3 shows the final presentation and device combinations, applying the calculation presented in Equation 2 to the results introduced in Table 2.

For the first row of Table 3, the SC is 0.3 and for the second 0.7. Using the previous calculated CC of both cases, the FC is slightly altered, producing a higher match for the current feasible presentation quality and situation.

This calculation result concludes the presentation and device matching score calculation. From here, the system either can take an automatic decision—which in this instance is the second row—or the system can present the user with a list of the two choices on her portal device, highlighting the second presentation and device match.

After making the final decision, the resource coordinator sends a snapshot to the situation learning system (step 4). The learning system then adds the snapshot to its internal database (step 5) and calculates new confidence rules available for the next request (step 6).

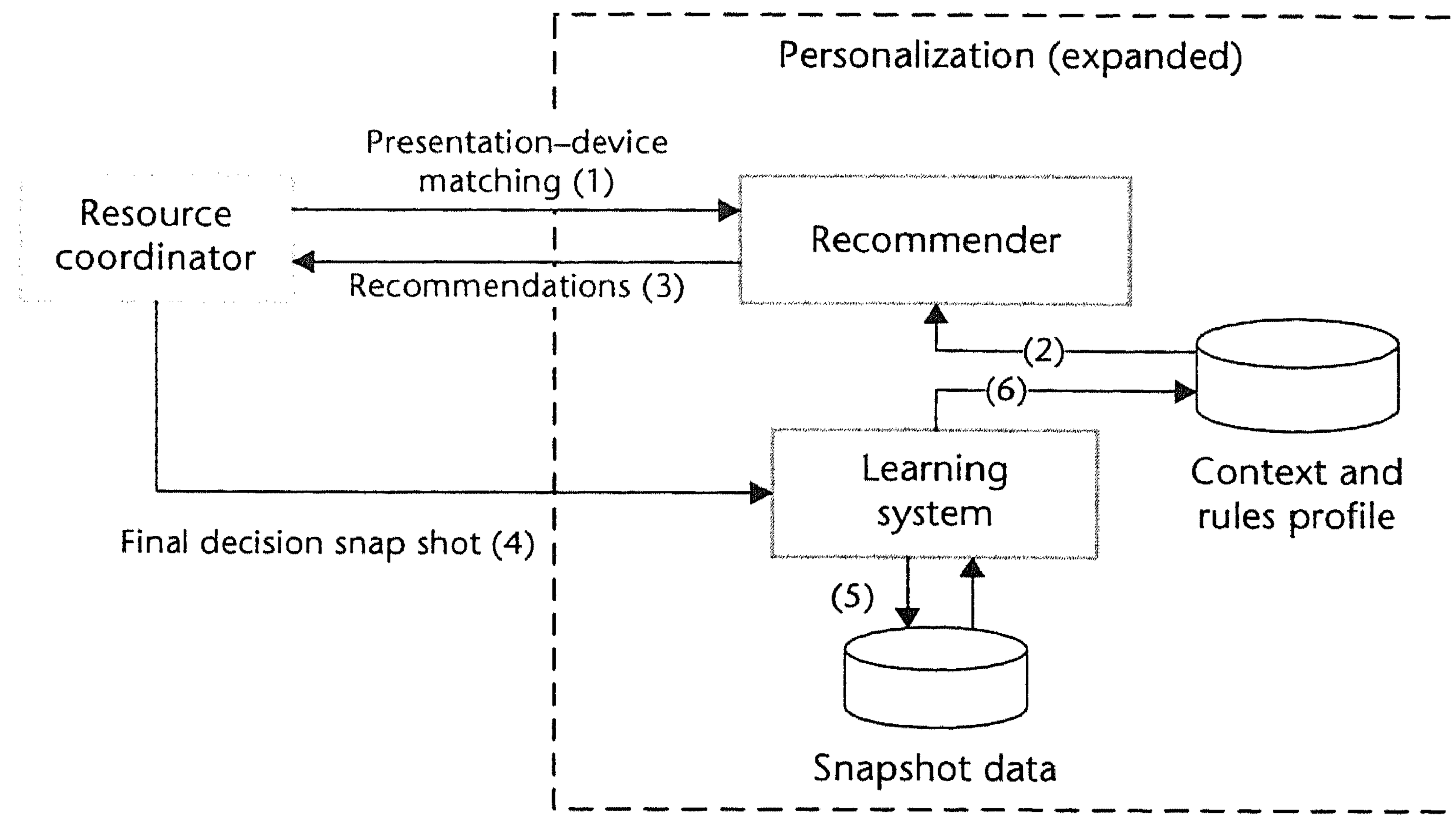

Figure 7. Six steps in the interaction and learning process for system personalization.

Dynamically creating presentations and delivering content

Based on the decision reached in the previous steps, the resulting multimedia presentations are constructed and delivered to the most suitable devices. We use SMIL as the presentation delivery format of choice to assure ubiquitous content delivery. The biggest benefit of using SMIL is its structured multimedia document nature.[14] Since the system is dealing with documents, and not with encoded content, session continuity, dynamic adaptation, and distributed rendering are easier to implement and deploy.

The presentation's timeline is composed of temporal constructs such as parallels and sequences, as Figure 8a (next page) shows. The media items that compose the presentation follow the temporal structure imposed by the multimedia service. Normally layout information of the presentation and media items (that is, the regions) provide information for deciding where the items should be rendered. This is in line with our base media item approach that we explained earlier. The framework allows for layout transformations to adjust spatial arrangements and resize media items, temporal transformations to assure session continuity, and media asset transformations for alternative media items that a specific device can use.

Transformations are dynamic and dependent on the situation and multimedia devices in the user's environment. In the earlier scenario, Maria has a TV set and a mobile phone at her disposal. In this case, the decision is to render the video and its captions on the TV set while the commercial images (the ads) are played on the mobile phone. The final presentation will look like the one represented in Figure 8b, where `customTest` attributes indicate whether the system shows an element. Device agents are in charge of checking this

Table 3. Final presentation and device combinations.

Item 1 (Text)	Item 6 (image)	Item 8 (video)	Combined confidence coefficient	Situation confidence coefficient	Final presentation and device combinations
Device 2 (hi-fi)	Device 3 (TV set)	Device 1 (mobile)	0.6	0.3	0.45
Device 2 (hi-fi)	Device 1 (mobile)	Device 3 (TV set)	0.63	0.7	0.665

Figure 8. Comparison between (a) the original SMIL document and (b) the annotated one for device-tailored delivery.

```
<layout>
 <region id="video" width="..." height="..."/>
 <region id="ad" width="..." height="..."/>
 <region id="caption" width="..." height="..."/>
</layout>
...
<seq>
   <par>
     <video region="video" src=".../video.flv"/>
     <text region="caption" src=".../info.rt"/>
     <img region="ad" src=".../ad.mov"/>
   </par>
 </seq>
```

(a)

```
Document sent to the TV
 <video region="video" src=".../video.mov"
  customTest="show"/>
<text region="caption" src=".../info.rt"
 customTest="show"/>
  <img region="ad" src=".../ad.mov"/>
Document sent to the mobile phone
 <video region="video" src=".../video.mov"/>
 <text region="caption" src=".../info.rt"/>
 <img region="ad" src=".../ad.3gp"
  customTest="show"/>
```

(b)

attribute for each media item. We detail other transformation mechanisms elsewhere.[14]

After producing the adequate structured presentation for each of the rendering device agents, the system delivers the modified presentation document to the selected device agents. The delivery phase takes advantage of the resulting format being a textual descriptive format and transmits it using the Internet Engineering Task Force's SIP mechanism. We included the SIP mechanism in an IP Multimedia Subsystem (IMS) communication infrastructure that we detail elsewhere,[14] but the system could also work with any other ubiquitous communication standard. As we previously validated,[15] such an infrastructure assures acceptable transmission times and session continuity when the system needs to reevaluate the user's environment (for example, when the system can use a new device for rendering).

Evaluation

We tested MDCS's performance within the EU-IST Spice project. We obtained useful feedback by having end users evaluate it quantitatively (via focus group interviews) and also through expert workshops (via individual in-depth interviews).

We showed the system to eight groups of approximately 12 persons each (four groups in Spain and four in Poland) and asked them to give their opinions related to general understanding, evaluation of chosen demo aspects (relevance, comfortable to use, usefulness, and novelty), market potential, and areas of improvement. We provided an explanatory video depicting the possibilities of the solution from a user's perspective, together with a slide presentation given by a moderator. Half of the groups had access to a prototype implementation, and thus received firsthand experience on how the system works. In total, more than 100 people took part in the test. The participants' ages ranged from 27 to 40, with a balanced representation in terms of gender.

In Italy, eight experts in pervasive-related projects participated in the evaluation. We showed the experts the same material as the user groups (video, presentation, and comments) adapted to their background, and they had access to the prototype implementation.

In general, participants didn't find it difficult to understand the purposes and benefits resulting from multimedia content adaptation in ubiquitous scenarios. Participants indicated the usefulness and convenience of such a solution, highlighting the mobility aspects. A key advantage, according to the participants, is the benefit of achieving a high-level of interoperability among the various device and media components, as well as the convenience of being able to use the most adequate devices in their surroundings at any specific moment. Nevertheless, one of the main concerns raised by the professional testers was the support of interoperability aspects between different device manufacturers, which we would need to overcome to make the system work in all situations. This is a concern that we share.

We detail further information on the evaluation process and recruitment, evaluation methodology, and a full set of results elsewhere.[16]

Conclusion

Ubiquitous multimedia research should consider scenarios that go beyond mobile consumption. Multidevice environments offer the potential to enhance the user experience in terms of flexibility and interactivity and will enable novel applications in education, entertainment, collaboration, and communication. In order to realize truly ubiquitous multimedia applications, suitable content adaptation methods need to be defined and deployed.

The contribution of this article is a novel multimedia adaptation and delivery process and its underlying framework that focuses on mobile users in daily and dynamic situations. We analyzed the different processing steps and defined related framework functionalities such as the generation of the presentation schedule, the computation of the presentation–environment matches, personalization through situation learning, and device(s)-tailored presentation delivery. The novelty of this solution resides in its dynamism, where mobility aspects such as location and device changes are taken into account, while assuring optimal session continuity across different situation configurations. Initial results based on user testing of a working system indicate clear advances over today's multimedia consumption landscape.

The impact of this article reaches beyond the process and architectural framework. It spans through a number of standardization activities (for example, SMIL and device ontologies) and it's supported by a number of industrial partners (including Alcatel-Lucent, France Telecom, and Telefonica). Moreover, the results from evaluating the system with more than 100 participants indicate that end users and professionals found it easy and convenient for accessing a broad range of multimedia content, resulting in an entertaining, new, and distinctive experience.

Nevertheless, ubiquitous multimedia adaptation and delivery in multidevice scenarios still must overcome a number of issues. This includes the promotion of further standardization for homogeneous device and content descriptions, identification of relevant context parameters for multimedia adaptation beyond location (such as network parameters and user activity), concurrent access to devices when more than one person is in the same vicinity, and cross-device synchronization mechanisms in heterogeneous networks. **MM**

References

1. S. Mate, U. Chandra, and I.D.D. Curcio, "Movable-Multimedia: Session Mobility in Ubiquitous Computing Ecosystem," *Proc. 5th Int'l Conf. Mobile and Ubiquitous Multimedia,* ACM Press, 2006; http://doi.acm.org/10.1145/1186655.1186663.
2. D. Dearman and J.S. Pierce, "It's on My Other Computer!: Computing with Multiple Devices," *Proc. Special Interest Group on Computer–Human Interaction Conf. Human Factors in Computing Systems,* ACM Press, 2008, pp. 767-776.
3. A. Sherp and S. Boll, "MM4U—A Framework for Creating Personalized Multimedia Content," *Managing Multimedia Semantics,* U. Srinivasan and S. Nepal, eds., IRM Press, 2005, pp. 246-287.
4. J.V. Ossenbruggen et al., "Towards a Multimedia Formatting Vocabulary," *Proc. 12th Int'l Conf. World Wide Web,* ACM Press, 2003, pp. 381-393.
5. A. Scherp and S. Boll, "Paving the Last Mile for Multi-Channel Multimedia Presentation Generation," *Proc. 11th Int'l Multimedia Modelling Conf.,* IEEE CS Press, 2005, pp. 190-197.
6. S. Berger et al., "Ubiquitous Computing Using SIP," *Proc. 13th Int'l Workshop on Network and Operating Systems Support for Digital Audio and Video,* ACM Press, 2003, pp. 82-89.
7. K. Kaneko, H. Morikawa, and T. Aoyama, "Session Layer Mobility Support for 3C Everywhere Environments," *Proc. 6th Int'l Symp. Wireless Personal Multimedia Comm.,* IEEE Press, 2003, pp. 347-351.
8. R. Schacham et al., "Ubiquitous Device Personalization and Use: The Next Generation of IP Multimedia Communications," *ACM Trans. Multimedia Computing, Comm., and Applications,* vol. 3, no. 2, 2007; http://doi.acm.org/10.1145/1230812.1230818.
9. K. O'Hara, A.S. Mitchell, and A. Vorbau, "Consuming Video on Mobile Devices," *Proc. Special Interest Group on Computer–Human Interaction Conf. Human Factors in Computing Systems,* ACM Press, 2007, pp. 857-866.
10. C. Hesselman et al., "Delivering Interactive Multimedia Services in Dynamic Pervasive Computing Environments," *Proc. 1st Int'l Conf. Ambient Media and Systems,* The Institute for Computer Sciences, Social Informatics, and Telecommunications Engineering (ICST), 2008, pp. 1-8; http://www.lab.telin.nl/~hesselma/publications/AmbiSys08_1.60-clean-web.pdf.
11. B. Rogge, J. Bekaert, and R. Van de Walle, "Timing Issues in Multimedia Formats: Review of the Principles and Comparison of Existing Formats," *IEEE Trans. Multimedia,* vol. 6, no. 6, 2004, pp. 910-924.
12. R. Kernchen et al., "Device Description for Mobile Multimodal Interfaces," *Proc. 15th Information Soc. Technologies' Mobile and Wireless Comm. Summit,* 2006; http://personal.ee.surrey.ac.uk/Personal/R.Kernchen/Publications/DevDescr_IST_Summit_2006.pdf.
13. Y. Du et al., "Context-Aware Learning for Intelligent Mobile Multimodal User Interfaces,"

Proc. 18th Ann. IEEE Int'l Symp. Personal, Indoor, and Mobile Radio Comm., IEEE Press, 2007, pp. 1-5.
14. P. Cesar et al., "Multimedia Adaptation in Ubiquitous Environments: Benefits of Structured Multimedia Documents," *Proc. ACM Symp. Document Eng.,* ACM Press, 2008, pp. 275-284.
15. I. Vaishnavi, P. Cesar, and D.C.A. Bulterman, "A Presentation Layer Mechanism for Multimedia Playback Mobility in Service Oriented Architectures," *Proc. Int'l ACM Conf. Mobile and Ubiquitious Multimedia,* ACM Press, 2008, pp. 205-212.
16. P. Miziołowski et al., "Deliverable 8.3—Pioneering SPICE Project, a View by Focus Groups from the Fields of Dynamic Desktop and Terminal Manager, Intelligent Service, End-User Studio, Split Terminal GBA Authentication, Content Guide Demos," tech. report, EU-IST FP6 Spice project, 2008; http://www.ist-spice.org/documents/SPICE_deliverable_8.3_final_version_clean.pdf.

Ralf Kernchen is a researcher and doctoral student at the Centre for Communication and Systems Research, University of Surrey, UK. His research interests include mobile multimodal user interfaces, context awareness, service platforms, and C2X applications. Kernchen has a Dipl.-Inf. in computer science from the Technical University of Berlin. Contact him at r.kernchen@surrey.ac.uk.

Stefan Meissner is a researcher and doctoral student at the Centre for Communication and Systems Research, University of Surrey, UK. His research interests include knowledge modeling, multimodal user interfaces, context awareness, and service-oriented architecture platforms for Internet of Things (IoT). Meissner has a Magister Artium in computer science and communications research from the Technical University of Berlin. Contact him at s.meissner@surrey.ac.uk.

Klaus Moessner is a professorial research fellow in the Centre for Communication and Systems Research, University of Surrey, UK. His research interests include reconfiguration management, service platforms, and adaptability of multimodal user interfaces. Moessner has a PhD in reconfigurable mobile communication networks from the University of Surrey. Contact him at k.moessner@surrey.ac.uk.

Pablo Cesar is a researcher at Centrum Wiskunde en Informatica (CWI), Netherlands. His research interests include multimedia systems and infrastructures, social multimedia, interactive media, multimedia content modeling, and user interaction. Cesar has a PhD from the Helsinki University of Technology. Contact him at p.s.cesar@cwi.nl.

Ishan Vaishnavi is pursuing a doctoral degree in real-time multimedia transmission mechanisms at Centrum Wiskunde en Informatica (CWI), Netherlands. His research interests include operating systems, multimedia networking, and real-time systems. Vaishnavi has an MTech in mathematics and computing from the Indian Institute of Technology, Delhi. Contact him at i.vaishnavi@cwi.nl.

Matthieu Boussard is a research project leader at Alcatel-Lucent Bell Labs, France. His research interests include mobile multimodal interfaces, context awareness, semantics, service composition, and mobile applications. Boussard has an ingénieur diplômé in telecommunication engineering from the French Institut National des Télécommunications. Contact him at mathieu.boussard@alcatel-lucent.com.

Cristian Hesselman is a senior researcher at Novay, Netherlands. His research interests include converged multimedia systems, adaptive multimodal user interfaces, and the integration of both with the real world. Hesselman has a PhD in computer science from the University of Twente, the Netherlands. Contact him at Cristian.Hesselman@novay.nl.