Dynamic and Distributed Adaptation of Scalable Multimedia Content in a Context-Aware Environment

Michael Ransburg, Hermann Hellwagner
Department of Information Technology
Klagenfurt University
Klagenfurt, Austria
{michael.ransburg,hermann.hellwagner}@itec.uni-klu.ac.at

Renaud Cazoulat
France Télécom R&D
Cesson Sévigné, France
renaud.cazoulat@orange-ft.com

Benoit Pellan, Cyril Concolato
Multimedia Group - Signal and Image Processing Dept.
GET - ENST
Paris, France
{benoit.pellan,cyril.concolato}@enst.fr

Saar De Zutter, Chris Poppe, Rik Van de Walle
Multimedia Lab
Ghent University - IBBT - IMEC
Ghent, Belgium
{saar.dezutter, chris.poppe, rik.vandewalle}@ugent.be

Andreas Hutter
CT IC 2 MC
Siemens AG
Munich, Germany
andreas.hutter@siemens.com

***Abstract*— The seamless access to rich multimedia content on any device and over any network, usually known as Universal Multimedia Access, requires interoperable description tools and adaptation techniques to be developed. To address this, MPEG-21 introduces an adaptation framework, which provides several mechanisms for making adaptation decisions according to usage environment and adapting multimedia contents in a coding format independent way. This paper gives an overview of the European FP6 project DANAE which not only implements and extends the existing MPEG-21 adaptation mechanisms but also kicked off several new standardization activities in the area of dynamic and distributed adaptation and resource conversion. MPEG-21 DIP enables static stream selection which is a first step in a series of adaptations. The existing BSD-based adaptation mechanisms enable the efficient implementation of generic adaptation engines, which can be used for existing and future coding formats. These mechanisms were extended to enable dynamic and distributed adaptation. Alternatively to the BSD-based adaptation, resource conversion was investigated which does not rely on scalable media and allows adaptations at the scene level.**

MPEG-21, Digital Item Adaptation, Resource Conversion, Digital Item Processing, Distributed Adaptation, QoS, QoE, Scalable coding, Efficient use of network/radio

I. INTRODUCTION

More and more multimedia contents are becoming available on the Internet, end user devices are becoming more and more diverse, and users with different preferences access the multimedia contents through heterogeneous networks. Therefore a need exists for multiple versions of the same content in order to meet all these different usage environments. This burden on the content and service providers can be removed by applying BSD-based adaptation which enables scaling a media content in a codec-agnostic way [1]. However, BSD-based adaptation relies on scalable media and is therefore constrained by the scalability options offered by a particular media codec. Therefore, the DANAE project uses BSD-based adaptation along with other adaptation mechanisms in order to provide the best quality of experience to the end-user.

Indeed, static stream selection enriches the set of possible adaptations by allowing an end user or an intelligent software to make a decision at session setup e.g., on the type of content based on the usage environment.

Furthermore, if the usage environment changes considerably during session lifetime (e.g., the bandwidth drops to a value where scaling of the video is no longer possible or efficient) then e.g., the conversion of the initial video to a slide show might be needed.

The combination of these adaptation strategies allows for a very complete set of adaptation possibilities which fit many usage environments. Moreover, the extension of the BSD-based adaptation to distributed scenarios benefits the service provider by making bandwidth-efficient usage of his network architecture.

Section II briefly introduces relevant parts of the MPEG-21 multimedia framework. Section III provides a use case which motivates our research of the presented adaptation mechanisms. Section IV provides a more detailed description of the DANAE adaptation framework. Finally, Section V concludes this paper.

II. MPEG-21 - The Multimedia Framework

The aim of the MPEG-21 standard, the so-called Multimedia Framework, is to enable transparent and augmented use of multimedia resources across a wide range of networks, devices, user preferences, and communities. MPEG-21 provides the next step in MPEG's standards evolution, i.e,. the transaction of Digital Items (DIs) among users. The Digital Item aggregates multimedia resources together with content-related metadata, licenses, identifiers, intellectual property management and protection (IPMP) information, and methods within a standardized structure and is therefore the fundamental unit of transaction and distribution within the MPEG-21 Multimedia Framework.

The MPEG-21 standard currently comprises 18 parts which can be clustered into five major categories each dealing with different aspects of the Digital Items: declaration (and identification), rights management, adaptation, processing, and systems aspects. For this paper, the parts related to the declaration, adaptation, and processing of Digital Items are relevant and are briefly reviewed in the following. For further information the reader is referred to [1].

A. Digital Item Declaration

The standard representation of Digital Items is defined by a model which describes a set of abstract terms and concepts and is expressed by the XML Schema-based Digital Item Declaration Language (DIDL) [2]. The resulting XML document, conformant to DIDL, is called Digital Item Declaration (DID). The DID may contain several building blocks (e.g., Resource, Descriptor and Choice/Selection) as defined in the DIDL which basically depends on the application and domain requirements. A resource usually references the multimedia content by using a Universal Resource Identifier (URI). A descriptor may provide additional information both at the semantic and syntactic levels by using any kind of (content-related) metadata.

B. Digital Item Processing

Digital Item Processing (DIP) [3] provides ways to interact with Digital Items. It allows specifying Digital Item Methods (DIMs) which can be provided to the User, who can then decide how to interact with the DI. An example is a configureChoice DIM which uses the DIDL choice/selection mechanism to enable a conditional configuration of the Digital Item in a dynamic way. As further detailed in Section IV.A, this approach allows for the so-called adaptation by selection, i.e., pre-computed variations of the multimedia resources can be selected by means of the available context information. This results in a DI with an updated DID where the choice in question is resolved. There is a set of DIMs which are implemented by normative Digital Item Base Operations (DIBOs). Additionally there is also the possibility to extend this by DIMs which are implemented by custom Digital Item eXtended Operations (DIXOs).

C. Digital Item Adaptation

Digital Item Adaptation (DIA) specifies normative XML-based description tools to assist with the adaptation of Digital Items, in particular the multimedia content referenced by a DID. As such tools used to guide the adaptation engine are specified by DIA, the implementation of the adaptation engines themselves are left open to industry competition. In particular, the DIA standard specifies means enabling the construction of device and coding format independent adaptation engines [4]. These tools provide means for adaptation decision-taking (i.e., the Adaptation Quality of Service tool, refered to as AQoS) and scalable multimedia content adaptation independent of the actual coding format (i.e., the generic Bitstream Syntax Description, refered to as gBSD). The AQoS describes all adaptation possibilities for a specific content, including the objective quality one can expect as a result from a particular adaptation. The gBSD provides a generic high-level syntax description of the content. This description enables generic adaptation engines to be implemented which can adapt any content which is described by such a gBSD. An additional set of DIA tools is referred to as Usage Environment Description (UED) tools which provide a comprehensive and extensible vocabulary for describing the context information pertaining to the environment where Digital Items are used. Such descriptions provide a fundamental input to any adaptation engine because "if you can describe it, you can adapt to it[1]".

III. Use Case

This section sketches a use case that motivates MPEG-21-based adaptation, as discussed in Section IV

An Internet Service Provider (ISP) provides a new feature which offers customers live streams of events and two subscribers use this service to watch a Formula 1 race. They are both in the same room and use their PC with a 30” LCD to watch this program. After some time, one of the subscribers has to leave the room. However, she wants to continue watching the race, therefore she picks her tablet PC and instructs it to duplicate the session from the PC. The Formula 1 program appears on the tablet PC and she can continue watching the race in another room. Consequently, the other subscriber has to go to work. The race is almost over and he does not want to miss the finish. He therefore picks his UMTS-enabled PDA and instructs the PC to transfer the session so that he can watch the finish of the race on his way to work. At some point in time during his travel, the network available bandwidth is significantly reduced so that it prevents from enjoying the race with a satisfactory quality level: the video stream is replaced by a slide show, in order to retain a good audio quality and watch key moments of the race.

The architecture, which the service provider relies on in order to fulfill this scenario, is explained below and illustrated in Figure 1. The set-top box needs to feature adaptation functionalities and is therefore context-aware. When the subscribers start watching the Formula 1 race on their PC, the set-top box needs to be aware of the terminal capabilities context of the PC. This includes the display resolution of the LCD and the decoding capabilities of the PC. If the requested stream is not already made available to the set-top box, it forwards the request, including the related terminal capabilities,

[1] W3C Device Description Working Group (DDWG); http://www.w3.org/2005/MWI/DDWG/.

to the service provider. The service provider uses the terminal capabilities to perform DIA-based adaptation as introduced in Section II.C and starts to stream the channel (the adapted media stream) to the set-top box. The set-top box forwards this adapted stream to the PC, where it is displayed on the LCD.

Since the network available bandwidth on the UMTS-enabled PDA depends on various factors such as the location and speed of the subscriber, the live stream also needs to be adapted on-the-fly against network conditions in order to always provide the best possible quality. The service provider performs DIA-based adaptation, according to user network characteristics, which can consist in degrading the audio and video quality or even dynamically switching media types (from video to slideshow for instance).

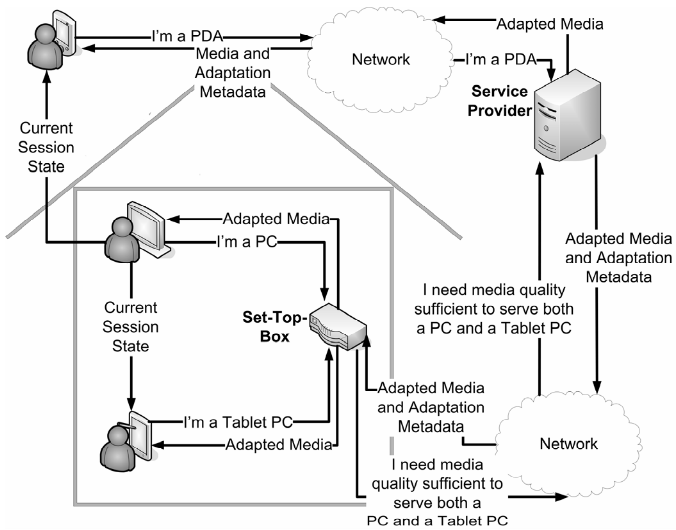

Figure 1: Dynamic and Distributed Adaptation Use Case

At some time, as described above, one of the subscribers requests to continue watching the channel on the tablet PC. This means, that the tablet PC requests the same channel from the set-top box, and provides its terminal capabilities with the request. When the set-top box receives this request, it analyzes the provided terminal capabilities in order to find out whether the quality of the channel, which it receives from the media provider, is sufficient, or if it needs to request a better quality. Actually, the quality is not appropriate - it is in fact too high given the display resolution and the decoding capabilities of the tablet PC. Therefore, the set-top box replicates the stream which it already receives from the Service Provider and adapts this replicated stream according to the terminal capabilities of the tablet PC.

By facilitating this architecture, the Service Provider saves bandwidth: Actually, multicast-like delivery is achieved by replicated unicast. However, the replication and, in addition, adaptation processes are being performed at the customers' premises without putting load on the ISP's equipment. Additionally, the architecture also enables the subscribers to transfer or duplicate the session to any MPEG-21 enabled device in their vicinity. Furthermore, the adaptation techniques employed in this architecture enable the service provider to provide its content to its subscribers "anywhere, anytime".

IV. Adaptation Framework

In the course of the DANAE project three different MPEG-21 based approaches to adaptation where researched, contributed to and implemented. The DANAE work in the context of resource conversion and distributed adaptation led to new standardization activities at MPEG.

Digital Item Processing (DIP) enables *static stream selection* at the start of a session and in session mobility scenarios. *Resource conversion* enables dynamic stream selection at any time in a stream. Alternatively, *BSD-based adaptation* enables fine-grained, dynamic scalability of multimedia content in a generic way. Finally, *distributed adaptation* extends the BSD-based adaptation approach in order to enable this adaptation mechanism anywhere along the delivery chain.

In the following, the different adaptation approaches are detailed using the architecture overview in Figure 2.

A. Static Stream Selection

The session is started by the Multimedia Player, which requests content to the Digital Item Processing (DIP) engine on the server. The DIP engine fetches the Digital Item (DI) which includes the requested content from the Content Database and performs a first adaptation step which is called static stream selection. The Digital Item Declaration (DID) is used to describe a set of alternatives for multimedia content which differ in, e.g., bit rate, resolution, encoding or detail level of the content. DIP can then be used to reconfigure the DID to only include those alternatives which fit the usage environment context, which is made available from the Terminal by the Context Aggregation Tool (CAT). One example of this is a DIM which facilitates the expertise level of the user with regards to a specific topic. This expertise level is then used to reconfigure the DID to contain only those content alternatives fitting the user's expertise level. In the next step, the remaining alternatives might be further reduced by a DIM which automatically retrieves the screen size of the end user device and uses it to exclude alternatives not fitting the provided screen size. Continuation of this static stream selection process results in a set of alternatives (at least one) which equally well satisfy the end-users usage environment.

This static stream selection process, which only happens once at session setup, is a first step in a series of adaptations which ultimately lead to the best available content for a given usage environment. Normative DIA descriptors are utilized to provide a description of the end-user's usage environment (such as the screen size of his terminal) to the DIP engine which reconfigures the DID appropriately.

The resulting reconfigured DI is then made available to the Fragmenter which facilitates dynamic and distributed adaptation.

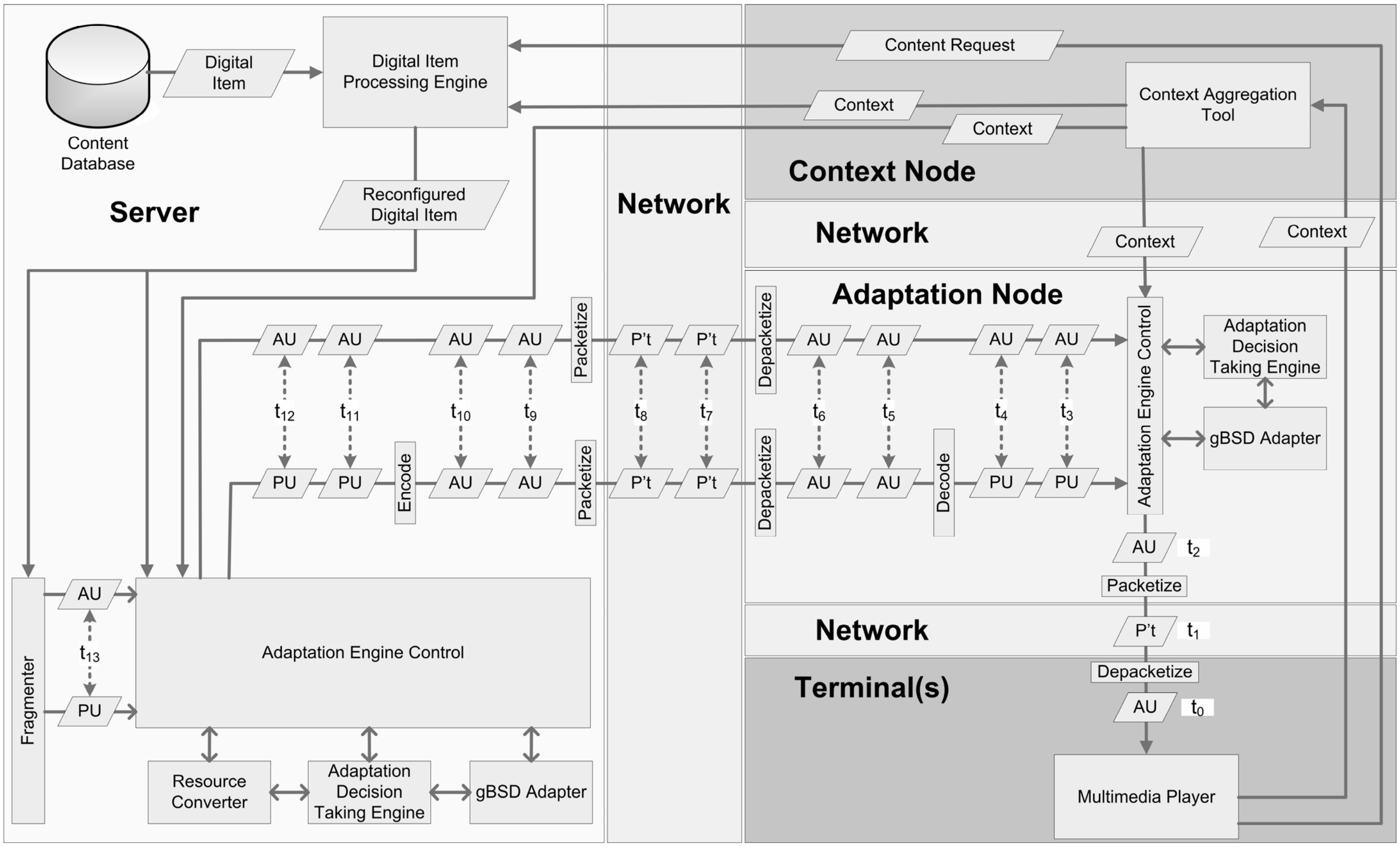

Figure 2: DANAE Architecture

B. Dynamic Adaptation

Dynamic adaptation refers to piece wise adaptation of media content while it is delivered to the Terminal. In particular, an adaptation decision is taken for each Access Unit (AU) of the content. In order to enable this, the Fragmenter fragments the media content into its AUs and similarly fragments the adaptation metadata (i.e., the gBSD and the AQoS) into so-called Process Units (PUs) by facilitating XML and media streaming instructions [5]. This fragmentation process is straight-forward for media content, since the content is already represented as AUs. For the adaptation metadata, the composition of the corresponding PUs is not straight forward, e.g., because there is redundant data (e.g., the ancestor XML nodes) needed in each PU in order to guarantee that it is a valid XML description. The XML streaming instructions enable to signal this composition. Each gBSD PU describes the syntax of the related AU. Similarly, each AQoS PU describes the adaptation options for each media segment, which usually consists of multiple AUs. For the sake of clarity, only one of the both metadata streams, i.e., the gBSD PUs, is depicted in Figure 2.

The Adaptation Engine Control gets the current AU and the corresponding PUs (gBSD and AQoS) from the Fragmenter. It is also constantly updated with the latest usage environment context from the CAT. The AQoS PU and the current context are provided to the Adaptation Decision Taking Engine (ADTE) which can select the appropriate adaptation by interpreting these normative descriptors. As a result, the decision is to either apply BSD-based adaptation or resource conversion, using a specific set of adaptation parameters.

1) BSD-based Adaptation

The characteristics of scalable coding formats enable the generation of a degraded version of the original bitstream by means of simple remove operations followed by minor update operations, e.g., removal of spatial layers and updates of certain header information comprising the horizontal and vertical resolution [6]. The gBSD PU, which is an XML document which describes an AU of a scalable bitstream, enables the adaptation thereof in a codec-agnostic way. The level of detail of this description depends on the bitstream characteristics and the application requirements. The adaptation comprises the transformation of the gBSD PU and the generation of the adapted bitstream AU, both processes are performed in the gBSD Adapter.

The gBSD transformation process transforms the gBSD PU according to the ADTE decision by using XSLT. The bitstream generation process parses the transformed gBSD PU and generates the adapted bitstream AU by only copying the parts of the AU which are described by the transformed gBSD. Furthermore, the transformed gBSD is updated to correctly reference the adapted AU with regards to start and length.

The resulting adapted AU and the corresponding updated gBSD PU are then further processed for delivery, as explained in Section C.

2) *Resource Conversion*

The resource converter is a media adapter that is able to adapt the media characteristics (e.g., modality, coding format or resolution) according to the ADTE decision by partially decoding and re-encoding the initial media. This kind of adaptation is much more CPU-demanding than BSD-based adaptations and will only be decided upon when scalable adaptations cannot be performed (either the resource is not scalable, or scalable adaptations are not appropriate).

Alternatively, the resource conversion can be done off-line. Hereby, the resource converter simply switches from the initial content to a pre-encoded content which has been generated during the content generation phase and is available in the (reconfigured) DID. DANAE concentrated on a common and interoperable ADTE for resource conversion and BSD-based adaptation and provides a dynamic alternative to the static stream selection, which is described in Section A.

In order to signal the incoming converted data to the Terminal, the MPEG-4 Binary Format for Scenes (BIFS) [8] update mechanism has been used in DANAE. This presentation format not only achieves seamless transitions between two media streams but also allows adaptation at the scene level. For instance, the adaptation of a video through resizing might free enough space for a short text explanation.

The overall architecture of Resource Converter is based on several conversion tools that are identified by conversion act URIs (registered ISO/IEC 21000-5 terms) [9]. Each conversion act has specific semantics (and conversion parameters) and is referenced in DIA descriptors[7]. The conversions are based on dynamic constraints that may vary during the session lifetime and therefore require opening a RTP channel for each adaptation option.

C. *Distributed Adaptation*

In order to enable distributed BSD-based adaptation, not only the AUs need to be streamed but also the adaptation metadata, i.e., the gBSD and AQoS PUs. While the encoding and transport of media AUs is a well known concept, it is novel for XML-based adaptation metadata.

In a first step, the PUs, which are still represented in the text domain, are encoded into AUs using a proper encoding mechanism by the Encoder. After evaluating different means to do this [10] the MPEG-7 Binary XML codec (BiM) [11] has been chosen as the most efficient compression method. BiM is a schema-aware encoding mechanism which, if properly configured, removes any redundancy existing between consecutive PUs. The redundancy, resulting from the requirement that PUs need to be able to be processed independently, is removed and only the new information is encoded into Access Units.

After encoding the PUs into AUs, the media, gBSD and AQoS AUs are packetized for transport by the Packetizer. In this step, the timing information provided by media and XML streaming instructions is mapped onto the transport layer, by including it into the packet header. Both the media and gBSD AUs are then streamed into the network, where (an) adaptation node(s) can perform additional BSD-based adaptation steps, as described in the use case in Section III.

Details on the encoding and transport of adaptation metadata have been published in [10].

Other content-related metadata, which do not have fragmentation or timing requirements are not streamed, but provided using other out-of-band mechanisms, i.e., as attributes in the Session Description Protocol (SDP, RFC2327).

V. CONCLUSION

In this paper, we presented the adaptation framework which was developed in the course of the European FP6 project DANAE. By facilitating interoperable MPEG-21 mechanisms, a high quality of experience for the end-user can be retained, independent from his usage environment. Three different adaptation mechanisms were introduced, DIP-based static stream selection, resource conversion and BSD-based adaptation. Additionally, it was shown how DANAE extends the BSD-based adaptation to distributed scenarios where the adaptation can happen anywhere along the delivery chain.

ACKNOWLEDGMENT

Part of this work was supported by the European Commission in the context of the DANAE project (IST-1-507113); http://danae.rd.francetelecom.com.

REFERENCES

[1] Ian Burnett, Fernando Pereira, Rik Van de Walle, and Rob Koenen (eds.), “The MPEG-21 Book”, John Wiley and Sons Ltd, 2006.

[2] Frederik De Keukelaere and Rik Van de Walle, “Digital Item Declaration and Identification”, in: Ian Burnett, Fernando Pereira, Rik Van de Walle, and Rob Koenen (eds.), The MPEG-21 Book, John Wiley and Sons Ltd, pp. 69-117, 2006.

[3] Frederik De Keukelaere and Gerrard Drury, “Digital Item Processing”, in: Ian Burnett, Fernando Pereira, Rik Van de Walle, and Rob Koenen (eds.), The MPEG-21 Book, John Wiley and Sons Ltd, pp. 333-399, 2006.

[4] Christian Timmerer, Sylvain Devillers, and Anthony Vetro, “Digital Item Adaptation - Coding Format Independence”, in: Ian Burnett, Fernando Pereira, Rik Van de Walle, and Rob Koenen (eds.), The MPEG-21 Book, John Wiley and Sons Ltd, pp. 282-331, 2006.

[5] ISO/IEC 21000-7:2006 FPDAM/2: “Dynamic and Distributed Adaptation”.

[6] Sylvain Devillers, Christian Timmerer, Jörg Heuer, and Hermann Hellwagner, “Bitstream Syntax Description-Based Adaptation in Streaming and Constrained Environments”, IEEE Transactions on Multimedia, vol. 7, no. 3, pp. 463- 470, June 2005.

[7] ISO/IEC 21000-7:2006 FPDAM/1: “Conversions and Permissions”.

[8] ISO/IEC 14496-11:2005: “Scene description and application engine”.

[9] Thomas DeMartini, Jill Kalter, Mai Nguyen, Edgar Valenzuela, and Xing Wang, “Rights Expression Language”, in: Ian Burnett, Fernando Pereira, Rik Van de Walle, and Rob Koenen (eds.), The MPEG-21 Book, John Wiley and Sons Ltd, pp. 139-213, 2006.

[10] Michael Ransburg, Christian Timmerer, and Hermann Hellwagner, “Transport Mechanisms for Metadata-driven Distributed Multimedia Adaptation”, First International Conference on Multimedia Access Networks (MSAN 2005), July 2005.

[11] Niedermeier U. et al., "An MPEG-7 tool for compression and streaming of XML data", IEEE International Conference on Multimedia and Expo (ICME 02), Lausanne, Switzerland, August 2002.