Measuring Quality of Experience for MPEG-21-based Cross-Layer Multimedia Content Adaptation*

Christian Timmerer[1], Víctor H. Ortega[2], José M. González[3], and Alberto León[3]
[1]Klagenfurt University, Dep. of Information Technology (ITEC), Austria, [2]Tecsidel, Spain, [3]Telefónica I+D, Spain
[1]christian.timmerer@itec.uni-klu.ac.at, [2]victor.ortega@tecsidel.es, [3]josemgc@tid.es, [3]alm@tid.es

Abstract

The aim of this paper is to describe a Quality of Service (QoS) model enabling to measure the perceptual quality of video transmissions by exploiting metrics from different layers (service, application, network) in an interoperable way. As such we are able to keep the quality experienced by the end user at a satisfactory level without cost-intensive subjective tests. Therefore, we propose a detailed QoS model for video transmission following the philosophy of the ITU-T's E-model for audio and show how this can be translated into interoperable description formats offered by the MPEG-21 Multimedia Framework.

1 Introduction

The requirement to access multimedia content such as video and audio streams during everyday's life is omnipresent. Research and standardization efforts around to what is commonly known as Universal Multimedia Access (UMA) has gained momentum and offer a rich set of tools enabling such an access from a technological point of view. However, most of these techniques exclude the human end user who is actually the source of the above mentioned requirements and ultimately wants to consume multimedia content independent of his/her context. The issues resulting from a more user-centric perspective are collectively referred to as Universal Multimedia Experience (UME) [1] where the user takes a center stage.

An important aspect with regard to UME is to measure the quality experienced by the user in an objective way and to signal the required quality metrics by standardized, i.e., interoperable, description formats. As the objective measures may require quality metrics coming from various layers (i.e., service, application, and network) we propose to adopt cross-layer interactions, especially when transmitting multimedia content over wireless channels [2].

In this paper we propose a QoS model enabling to measure the perceptual quality of video transmissions taking into account quality metrics coming from different layers and following the philosophy of the ITU-T's E-model for audio. In order to enable interoperability among the involved parties – mainly the service provider and the content consumer – we propose to adopt description formats (i.e., tools) from the MPEG-21 Digital Item Adaptation (DIA) standard. In particular, we demonstrate how our QoS model can be instantiated using MPEG-21 DIA tools enabling a generic metadata-driven decision-taking component to determine which parameters of the content needs to be adjusted and how in order to provide a satisfactory quality experienced by the end user.

The remainder of this paper is organized as follows. Section 2 describes the background in terms of probing QoS and MPEG-21. Our proposed QoS model including how it is mapped to interoperable description formats offered by MPEG-21 is described in Section 3. Conclusions and future work items are outlined in Section 4.

2 Background

2.1 Probing Quality of Service

For audio streams, in 1996 the European Telecommunications Standards Institute (ETSI) and the International Telecommunications Union (ITU-T) publish a model that estimates the perceived quality experimented by users in phone calls. This model (E-model) is based on the premise that: "Psychological factors on the psychological scale are additive". The E-model takes into account factors such as packet loss, delay, and others like the equipment impairment and the packet loss robustness factors that depends on the codec used in the connection as is described by the recommendation ITU-T G.113.

When we are dealing with video streams there are multiple parameters that can change between two videos even if they are coded with the same codec.

* This work is supported by the European Commission in the context of the ENTHRONE project (IST-1-507637). Further information is available at http://www.ist-enthrone.org.

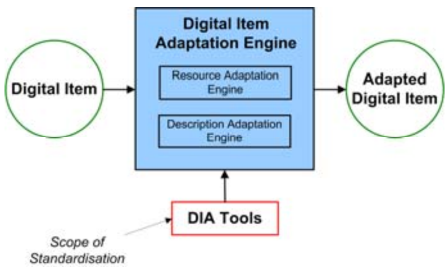

Figure 1. Concept of MPEG-21 Digital Item Adaptation [8].

Most of up-to-date codecs define different profiles with several resolutions, frames per second, bit rates, etc. Some approaches have used automatic measures of the video quality based on the Peak Signal to Noise Ratio (PSNR) [3]. But because it is impossible to have the original picture at the destination, the PSNR can only be used to extract some parameters of the Mean Opinion Square (MOS) in function of the packet loss.

If we analyze the PSNR picture by picture, it does not take into account the transition between pictures, and the human perception is very sensitive to this transitions. Thus, this method does not give a good perceptual approach of a video sequence. Other solutions such as the Video Quality Metric (VQM) offer an objective quality approach to help in the design and control of digital transmission systems [4]. In [5] appears a study which takes into account the perceptual impression of packet loss, variation in the frame rate and synchronization with the audio signal. Finally, a more sophisticated study considering the opinion of the users is also explained in [3].

2.2 MPEG-21 Multimedia Framework

The MPEG-21 Multimedia Framework comprises a set of standards (17 parts) with the aim to enable transparent and augmented use of multimedia resources across a wide range of networks, devices, user preferences, and communities, notably for trading (of bits) [7]. In particular, it shall facilitate the transaction of *Digital Items* (DIs), i.e., media resources, metadata, and structural information among *Users* of the entire digital media distribution chain. Note that Users within this framework include also communities and organizations and are not restricted to the single human end user. The standard can be clustered into six categories each dealing with different aspects: *declaration and identification* of Digital Items, *digital rights management*, *adaptation* of DIs regarding different usage contexts, *processing* of DIs, various *systems* aspects (e.g., a file format), and *miscellaneous* (e.g., reference software, conformance, etc.).

A vital part with respect to UMA/UME as introduced in Section 1 is part 7 of MPEG-21 entitled Digital Item Adaptation (DIA) [8]. This part of MPEG-21 specifies normative description tools to assist with the adaptation of Digital Items as depicted in Figure 1. That is, only the information required by media resource and description adaptation engines is standardized while the actual adaptation engines are not normatively defined. Furthermore, such an open architecture allows for various QoS management techniques to be applied including means for cross-layer multimedia content adaptation as proposed in this paper. Therefore, the relevant MPEG-21 DIA description tools are briefly described in the following subsections:

- Usage Environment Description (UED)
- Universal Constraints Description (UCD)
- Adaptation Quality of Service (AQoS)

2.2.1 Usage Environment Description

The Usage Environment Description (UED) tool allows one to instantiate the fundamental input to any kind of adaptation engine in an interoperable way. The UED includes various properties of the context in which a Digital Item is consumed (or, more generally, processed) and comprises four categories.

First, *terminal* capabilities provide means for describing both receiving and transmitting capabilities in terms of supported en-/decoders, display and audio output properties, user interaction support, and power and storage characteristics, etc.

Second, *network* characteristics and conditions describe the static and dynamic properties of the network enabling multimedia adaptation for improved transmission efficiency. The static properties include information about the maximum capacity of the network and the minimum guaranteed bandwidth a network can provide among others. The dynamic properties deal with available bandwidth, error and delay.

Third, properties related to the *user* characteristic may include information about the user, usage preferences and history, presentation preferences, accessibility characteristics, and location where the user wants to consume the Digital Item.

Finally, *natural environment* characteristics are related to the physical environment conditions around a user such as lightning condition, noise level, or time and location. For further reading the reader is referred to [8][9].

2.2.2 Universal Constraints Description

The UED as described in the previous section mainly describe static properties of the context in which a Digital Item is processed. However, for

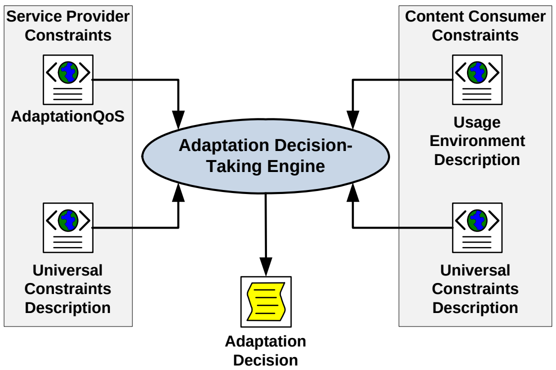

Figure 2. High-Level View of an Adaptation Decision-Taking Engine.

adaptation purposes it is desired to further constrain UEDs by explicitly signaling the adaptation engine a range of adaptation possibilities. For example, a terminal may have a resolution of 480×320 pixels but one could also adapt to a resolution below that. Therefore, the Universal Constraints Description (UCD) tool provides means for describing limitation and optimization constraints which may be used to further constrain the usage of Digital Items or the usage environment with respect to the possible adaptation space.

2.2.3 AdaptationQoS

The AdaptationQoS (AQoS) tool provides means to specify the relationship between constraints (i.e., the UED), feasible adaptation operations (e.g., transcoding, scaling, etc.) satisfying these constraints, and associated utilities (i.e., qualities, for example, in peak signal-to-noise ratio). Together with the UCD (and also UED) it allows one to formulate a mathematical optimization problem for which a variety of algorithms can be applied [10]. In the worst case, the problem to be solved is a mixed-variable multi-criteria optimization problem with general constraints as it is discussed in [11].

A high-level architecture of an adaptation decision-taking engine based on the above introduced description tools is depicted in Figure 2.

3 An Interoperable QoS Model for Video Transmission Exploiting Cross-Layer Interactions

3.1 QoS Probes and Mapping

In our study we made an analysis of the perceptual quality for video transmission with different ranges of parameters. We used two different video sequences codified with AVC/H.264 and a frame rate from 6.25 to 25 frames per second. The video bandwidth used is between 150 and 1500 kbps, and finally, we introduced simulated random packet loss up to 10%. With all this ranges we build a huge repository of videos with some of their parameters modified. In this way, we can observe not only the effect on the subjective quality of the video when varying only one parameter, but also can simultaneously study the cumulative effect on the quality of several of them.

A public survey[2] has been distributed in order to include as wide and heterogeneous audience as possible in both, internal and external approach, in a national and European environment. Each person watched a minimum of 10 videos randomly selected from the repository and rated the quality of the video between a value of 1 for a bad quality and 5 for a perfect quality. From this evaluation and the corresponding content parameters we were able to derive the following formulas (bottom-up) to be used in a video transmission system.

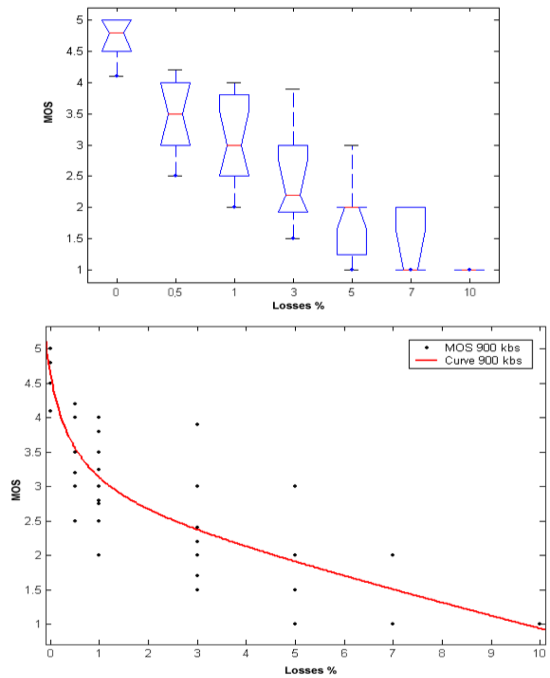

Figure 3. MOS vs. Packet Loss for 900 kbps.

3.1.1 Impact of Packet Loss

For loss distribution we have used a Bernoulli model. All the packet loss introduced in the videos were made assuming random losses distributed over an uniform probability density function which means that all the packets have the same probability to be dropped.

[2] http://www.enthrone2.tid.es/ (last accessed Jan. 2008)

$$f_{150}(l)=\frac{-0.06914\cdot l^2+1.545\cdot l+1.719}{l+0.447}$$

$$f_{300}(l)=\frac{-0.1343\cdot l^2+2.061\cdot l+2.041}{l+0.4721}$$

$$f_{600}(l)=\frac{-0.1591\cdot l^2+2.333\cdot l+2.342}{l+0.5274}$$

$$f_{900}(l)=\frac{-0.1757\cdot l^2+2.496\cdot l+2.5}{l+0.5391}$$

$$f_{1500}(l)=\frac{-0.21\cdot l^2+2.73\cdot l+3.012}{l+0.6336}$$

Equation 1. MOS vs. Packet Loss equations by bit rate.

$$f_L(l)=\frac{P_1\cdot l^2+P_2\cdot l+P_3}{l+Q_1}$$
$$\forall l\in[0,10]\%$$

Equation 2. Generic MOS vs. Packet Loss equation.

In the real world, packet dropping used to appear in the form of bursts of random length. A burst is a period with a high density of losses with independent probability which produces a larger distortion than isolated losses. Other models like Gilbert and Markov describe state models that transition between gap (good) states and bad (burst) states which have high or low density of independent losses. For this reason when we calculate the quality in short intervals, the packet loss density distribution can be considered uniform even if we are inside a burst.

The first step to process all the gathered data was to remove the atypical qualities, i.e., those values that are too different from the majority. These atypical data can be explained due to mistakes made during the input of the data, the videos were watched in non-optimum conditions, the user had a bad day, etc. We consider atypical (outliers values) all the data that were deviated from the mean more than 3/2 of the standard deviation. Then we found the equations that fit the curves described by the data clouds as shown in Figure 3.

The equations that minimizes the mean error among all the opinions (once we discard all the atypical data) for the different values of the analyzed bandwidth is shown in Equation 1.

3.1.2 Impact of Bandwidth

Another factor that can influence in the video stream quality is the bandwidth, for this reason, many bandwidth rates has been studied. In order to obtain a model that fits all the curves generated (one for each bandwidth) we calculated the equations that describe

$$P_1(br)=-0.1387\cdot e^{\frac{2.721\cdot br}{10000}}+0.2823\cdot e^{\frac{-8.885\cdot br}{1000}}$$

$$P_2(br)=2.154\cdot e^{\frac{1.584\cdot br}{10000}}-2.125\cdot e^{\frac{-7.8\cdot br}{1000}}$$

$$P_3(br)=1.95\cdot e^{\frac{2.887\cdot br}{10000}}-1.307\cdot e^{\frac{-9.414\cdot br}{1000}}$$

$$Q_1(br)=\frac{1.75\cdot br^3}{10^{10}}-\frac{4.327\cdot br^2}{10^7}+\frac{4.19\cdot br}{10^4}+0.3876$$

Equation 3. Bandwidth Model.

$$f_L(l,br)=\frac{P_1(br)\cdot l^2+P_2(br)\cdot l+P_3(br)}{l+Q_1(br)}$$

Equation 4. MOS vs. Packet Loss and bit rate equation.

the different constant parameters from Equation 1 as shown in Equation 2.

The values of the constants P_1, P_2, P_3, and Q_1 for the different bit rates analyzed when represented describe increasing/decreasing curves that can be easily approximated (Equation 3). This way we can describe the relationship between the bit rate and the packet loss and how they influence in the obtaining of the quality perceived by the users as shown in Equation 4.

3.1.3 Impact of Frame Rate

An interesting study in how the quality perception changes as a function of the frame rate is shown in [6]. This study categorizes the media streams using three parameters: *temporal nature* of the data (i.e., soccer match vs. interview), *audio* (auditory) and *visual content*. Based on this categorization the *watchability* of the media streams is analyzed for all the possible combinations and making a classification based on the perception of the users between 1 and 7 (where 7 is the best quality).

We extrapolated this study to the case of a video on demand scenario where the video sequences have a high temporal nature. Considering that 30 fps has the best quality (i.e., 7) and normalizing the values, we can see the degradation factor that suffers the media rating in function of its frame rate in Figure 4. The equation that describes the curve of degradation of quality in function of the frame rate and minimizes the mean quadratic error is shown in Equation 5 where *fps* is the number of frames per second.

The curves that describe the different frame rates studied were calculated using an initial curve obtained – as explained in Section 3.1.1 – and the degradation factor of Equation 5 were applied. In order to check the accuracy of the proposed equation, the different curves have been compared with the data obtained from the real users as is shown in Figure 5.

$$f_R(fps) = \frac{-0.00102 \cdot fps^2 + 1.164 \cdot fps + 1.704}{fps + 5.714}$$
$$\forall \; fps \in [5, 30]$$

Equation 5. Frame Rate Model.

$$MOS(l, br, fps) = f_L(l, br) \cdot f_R(fps)$$
$$\forall \; l \in [0, 10]\%$$
$$\forall \; br \in [150, 1500] \text{kbs}$$
$$\forall \; fps \in [5, 30]$$

Equation 6. Our proposed QoS Model.

3.1.4 Impact of Delay

The delay is important when a high interactivity is required, e.g., video conference, whereas in scenarios like video streaming where a buffer space is created and as long as the data can be downloaded as fast as it is used up in playback, the influence of this parameter will practically disappear. Therefore, we have not further considered this factor.

3.1.5 Model proposed

The quality perceived by the user will be obtained mainly from the values of the bit rate and of the packet loss within the network, this maximum value will be influenced by the frame rate of the video proportionally. Therefore, the final formula used to obtain the perceived quality of a video streaming is shown in Equation 6.

3.2 Adding MPEG-21 Support Enabling Interoperable Cross-Layer Interactions

In order to provide interoperability when well established layers are broken up, we propose to utilize MPEG-21 DIA tools for describing functional dependencies between the layers. In particular, QoS mappings as described in Section 3.1 – possibly ranging across well-defined network layers – are instantiated with the MPEG-21 DIA tools introduced in Section 2.2. In this section we will show how the proposed QoS model as defined in Equation 6 can be instantiated using interoperable description formats according to the three-step approach described in [12].

The functional dependencies of the MOS function are described using MPEG-21 DIA AdaptationQoS' stack functions, i.e., XML-based reverse polish notation (RPN), given a range of possible content frame rate and bit-rate combinations (expressed via a so-called look-up table) with the objective to maximize the MOS. The context in which the multimedia content is consumed is characterized through the packet loss measured at the receiving terminal and communicated

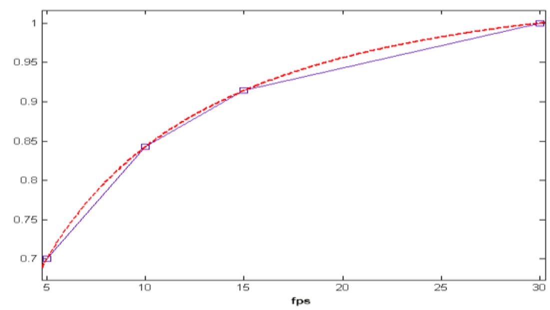

Figure 4. Degradation of Quality based in Frame Rate.

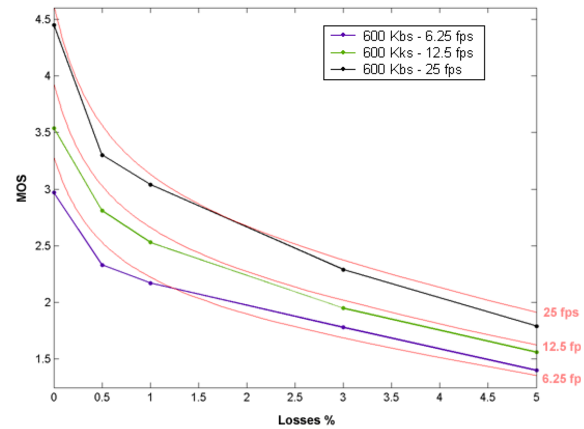

Figure 5. Frame Rate Model Comparison.

towards the multimedia content providing entity using an MPEG-21 DIA Usage Environment Description (UED). In addition to the UED, the constraints of the probe – as indicated by Equation 6 – are expressed by an MPEG-21 DIA Universal Constraints Description (UCD) utilizing limitation constraints which are attached to the UED.

Excerpts of the aforementioned interoperable descriptions are shown in Figure 6. Note that the complete descriptions can be found elsewhere[3]. The AdaptationQoS excerpt provides the multiplication of the functions representing the MOS vs. Packet Loss and bit rate equation (Equation 4) and the frame rate model (cf. Equation 5) respectively. Both functions are also represented as stack functions and so on demonstrating the modular usage of this tool. The maximization of the MOS is indicated by an optimization constraint referencing the MOS IOPin of the AdaptationQoS description. The UED excerpt describes a network with 1500 kbps available bandwidth and three percent packet loss. Finally, the probes' UCD excerpt defines that only packet losses of smaller than 10 percent are allowed.

[3] http://www-itec.uni-klu.ac.at/~timse/WISe08/wise08.zip

AdaptationQoS Stack Function for MOS (aqos.xml):

```
<!-- Stack Function for MOS calculation -->
<Module xsi:type="StackFunctionType" iOPinRef="MOS">
  <StackFunction>
    <Argument xsi:type="InternalIOPinRefType"
       iOPinRef="F_FRAMERATE"/>
    <Argument xsi:type="InternalIOPinRefType"
       iOPinRef="F_PACKETLOSS"/>
    <!-- multiply -->
    <Operation operator=":SFO:18"/>
  </StackFunction>
</Module>
```

UCD maximizing the MOS (ucd_provider.xml):

```
<OptimizationConstraint optimize="maximize">
  <Argument xsi:type="ExternalIOPinRefType"
     iOPinRef="aqos.xml#MOS"/>
</OptimizationConstraint>
```

UED (ued.xml):

```
<Network xsi:type="NetworkType">
  <NetworkCharacteristic
       xsi:type="NetworkConditionType">
    <AvailableBandwidth average="1500000"/>
    <Error packetLossRate="0.03"/>
  </NetworkCharacteristic>
</Network>
```

UCD for probe constraints (ucd_probe.xml):

```
<!-- packet loss <= 0.1 (10%) -->
<LimitConstraint>
  <Argument xsi:type="SemanticalRefType"
      semantics=":AQoS:6.6.5.8"/>
  <Argument xsi:type="ConstantDataType">
    <Constant xsi:type="FloatType">
      <Value>0.1</Value>
    </Constant>
  </Argument>
  <Operation operator=":SFO:38"/>
</LimitConstraint>
```

Figure 6. MPEG-21 DIA Description Excerpts.

Once these descriptions are available, they can be fed into an adaptation decision-taking engine (ADTE) as shown in Figure 2. In our work we rely on an existing implementation [10]. In particular, the implementation is generic in a sense that the core is independent of the actual description format and solves a mathematical optimization problem by restricting the solution space in order that the limitation and optimization constraints are fulfilled. In our example, the ADTE will assign values to the AdaptationQoS' IOPins which can be used to adjust the bit-rate and frame rate according to the measured packet loss while maximizing the MOS.

4 Conclusion and Future Work

When delivering multimedia data through heterogeneous networks and in order to guarantee a good QoS it is necessary to know what is the quality of multimedia traffic experimented by the user. This requires a reliable model able to translate the problems encountered in the network (packet loss, etc.) into a measure of the quality perceived by the final user.

In that sense, this paper proposes, following the philosophy of the E-model for audio, a new model for evaluating the quality of video streams in a non-intrusive way. This new model depends on network factors easily extracted from the RTP traffic data.

Furthermore, this paper shows how such a model can be translated into interoperable description formats offered by the MPEG-21 Multimedia Framework enabling optimization strategies applied across layers ranging from the service layer to the network layer.

Future work will include the evaluation of this model in large-scale pilots featuring interconnected test-beds across Europe during the course of the IST-ENTHRONE project [13].

5 References

[1] F. Pereira, I. Burnett, "Universal multimedia experiences for tomorrow", *IEEE Signal Processing Magazine*, vol. 20, no. 2, Mar. 2003, pp. 63 – 73.

[2] M. van der Schaar, N. Sai Shankar, "Cross-Layer Wireless Multimedia Transmission: Challenges, Principles, and New Paradigms", *IEEE Wireless Communications*, vol. 3, no. 4, Aug. 2005, pp. 55-58.

[3] J. Klaue, B. Rathke, A. Wolisz, "EvalVid - A Framework for Video Transmission and Quality Evaluation", *Proc. 13th International Conference on Modeling Techniques and Tools for Computer Performance Evaluation*, Urbana, 2003. pp. 233-237.

[4] M.H. Brill, J. Lubin, P. Costa, J. Pearson, "Accuracy and cross-calibration of video-quality metrics: new methods from ATIS/T1A1", *Proc. of the International Conference on Image Processing*, Rochester, 2002. pp. 37-40.

[5] P. Ruiz, "Seamless Multimedia Communications in Heterogeneous Mobile Access Network", *Proc. of the Terena Networking Conference*, Rhodes, 2004.

[6] R.T. Apteker, J.A. Fisher, V.S. Kisimov, H. Neishlos, "Video acceptability and frame rate", *IEEE MultiMedia*, vol. 2, no. 3, Fall 1995.

[7] F. Pereira, J.R. Smith, A. Vetro, (eds.), "Special Section on MPEG-21", *IEEE Trans. on Multimedia*, vol. 7, no. 3, June 2005.

[8] A. Vetro, C. Timmerer, "Digital Item Adaptation: Overview of Standardization and Research Activities", *IEEE Trans. on Multimedia*, vol. 7, no. 3, June 2005, pp. 418-426.

[9] A. Vetro, C. Timmerer, S. Devillers, "Digital Item Adaptation - Tools for Universal Multimedia Access", in: I. Burnett, F. Pereira, R. Van de Walle, and R.Koenen (eds.), *The MPEG-21 Book*, John Wiley and Sons Ltd, 2006, pp. 243-281.

[10] I. Kofler, C. Timmerer, Hermann Hellwagner, A. Hutter, F. Sanahuja, "Efficient MPEG-21-based Adaptation Decision-Taking for Scalable Multimedia Content", *Proc. of the 14th Multimedia Computing and Networking Conference (MMCN'07)*, San Jose, USA, Jan./Feb. 2007.

[11] D. Mukherjee, E. Delfosse, J.-G. Kim, Y. Wang, "Optimal Adaptation Decision-Taking for Terminal and Network Quality of Service", *IEEE Trans. on Multimedia*, vol. 7, no. 3, Jun. 2005, pp. 454-462.

[12] I. Kofler, C. Timmerer, H. Hellwagner, T. Ahmed, "Towards MPEG-21-based Cross-layer Multimedia Content Adaptation", *Proc. 2nd International Workshop on Semantic Media Adaptation and Personalization (SMAP 2007)*, London, UK, Dec. 2007.

[13] T. Ahmed (eds.), "Pilot and services integration and tests", *ENTHRONE Deliverable D27*, http://www.ist-enthrone.org, under preparation, due in Mar. 2008.