MPEG-21-based Cross-Resource Adaptation Decision-Taking

Ingo Kofler and Hermann Hellwagner
Department of Information Technology (ITEC), Klagenfurt University, Klagenfurt, Austria
{ingo.kofler, hermann.hellwagner}@itec.uni-klu.ac.at

Abstract

The adaptation of multimedia resources is a common method to enable the transport and consumption of audio-visual content in constrained environments. An important aspect in this field is adaptation decision-taking, which aims to find adaptation parameters that maximize the quality for the consumer while considering the constraints of the networks and terminals involved. In this paper we focus on improving the adaptation of audio-visual content by maximizing the perceived quality. This can be realized by using a multimedia quality model and content-related metadata. We present an approach to derive this content-related metadata from subjective tests and use it for adaptation decision-taking within the MPEG-21 multimedia framework.

1. Introduction

The actual vision of the multimedia community is to enable a uniform delivery of multimedia resources to a growing variety of end terminals. The ultimate goal is to support the transport of multimedia content over a heterogeneous infrastructure of delivery networks and the consumption of the content at an arbitrary end device. This idea is commonly known as Universal Multimedia Access (UMA) [1] and is one of the recent research topics in the field of multimedia communication.

A key technology to overcome the individual limitations of the heterogeneous infrastructure is the adaptation of the multimedia resource. Examples for such limitations are a limited network bandwidth or bit rate restrictions of the terminal's decoder. The adaptation of the content can be performed along the adaptation dimensions that are determined by the type of the resource (image, video, audio) and the codec used. In general, the adapted resource has fewer requirements concerning the network capacity and the resources at the terminal but also comes along with a decreased perceived quality at the consumer's end device. Therefore the selection of appropriate adaptation parameters, which is also known as adaptation decision-taking, is a very important task for achieving the vision of UMA. The task of decision-taking is to find a way to adapt the resource that maximizes the perceived quality for a given set of network and terminal limitations. Different approaches to fulfill this task can be found in the current literature, ranging from a simple look-up mechanism [2] to a knowledge-based planning scheme [3].

Adaptation of multimedia resources plays also a vital role within the MPEG-21 multimedia framework. In particular part 7 of the MPEG-21 standard – Digital Item Adaptation (DIA) [4] – deals with the adaptation of Digital Items, which are multimedia resources and associated metadata in the terminology of MPEG-21. The approach of adaptation decision-taking that is introduced in the standard is based on solving a mathematical optimization problem, which is described by metadata. Adaptation parameters and the resulting characteristics are expressed as mathematical variables. Dependencies between the variables are modeled using functions that can either be defined by explicitly listing all tuples or by defining an algebraic expression. E.g., the dependency between the frame rate and the resulting bitrate can be expressed by a look-up table. In addition to variables and functional dependencies, the optimization problem consists of a set of limitation and optimization constraints. While the limitation constraints can be used to restrict the value of a variable, e.g., to force the bitrate of a video to be lower than the available network bandwidth, the optimization constraints can be used to steer the selection of the adaptation decision. An optimization constraint consists of an objective function that has to be maximized or minimized. In most cases the optimization constraint is used to maximize the resulting quality of the adapted resource, which can either be assessed using a subjective measure, e.g., Mean Opinion Score (MOS), or an objective measure, e.g., Peak Signal-to-Noise Ratio (PSNR).

In our paper we focus on cross-resource decision-taking. In contrast to the ordinary decision-taking that only considers the adaptation of a single resource, the cross-resource variant deals with the adaptation of an audio-visual multimedia resource. The handling of both the audio and visual part of the resource results in some additional challenges. As both parts offer independent adaptation dimensions there exists a higher number of available adaptation possibilities that have to be investigated. However, the more complex task is the selection of the final adaptation decision from the set of available adaptation possibilities. Again, this selection is steered by an optimization constraint that tries to maximize the perceived quality. Unfortunately there exists no objective measure that can be used to measure the global quality of an audio-visual resource. Instead of that we used a multimedia quality model [5] that approximates the perceived quality. As the parameters of the model are specific to a certain kind of content we performed subjective quality assessments to determine these parameters for different types of audio-visual content. Based on the results we formulated an optimization problem using MPEG-21 DIA metadata, enabling MPEG-21 compliant cross-resource adaptation decision-taking.

The rest of the paper is organized as follows. In Section 2 the global quality model that was taken as a basis for adaptation decision-taking, is introduced in more detail. The subjective quality assessment that was performed to derive the content-specific parameters for the quality model is described in Section 3 while Section 4 discusses the results. The translation of the quality model into an optimization problem that is expressed using MPEG-21 DIA metadata is covered in Section 6. The paper is concluded in Section 8.

2. Multimedia Quality Model

Most of the research concerning quality models is focused on understanding the human perception of unimodal content, e.g., audio only perception. But at the moment there exists no objective quality model that enables the prediction of the perceived global quality that results from a given audio and video quality.

Approaches towards measuring the global multimedia quality were introduced in [5]. The results of subjective experiments showed that predicting the perceived quality of an audio-visual clip is more complicated than simply adding or averaging the audio and video quality. Furthermore, the impact of the individual qualities on the human perception is highly depending from the type of multimedia content. It is somehow obvious but also validated by subjective tests [14], that the human perception is more sensitive to audio degradation when watching head and shoulder sequences. On the other hand, in the case of content with a high amount of motion, e.g., an action movie, the perceived global quality suffers more from a degraded video quality than from a bad audio quality. The assumption for the multimedia quality model was that there already exist quality models for both video and audio. Therefore, the multimodal quality should be predictable by applying multimodal combination rules to the individual qualities. The proposed combination rule for the global multimedia quality (MMQ) [5] is

$$MMQ = a \cdot AQ + b \cdot VQ + c \cdot AQ \cdot VQ + d$$

The rule is mainly a weighted sum of both the audio quality (AQ) and video quality (VQ) with an additional multiplicative term. It was derived based on subjective tests and statistical analysis. As the contribution of audio and video quality to the perceived quality depends on the type of content, the coefficients *a, b, c, d* have to be determined for the actual content. This requires subjective tests which are expensive, time consuming and in some cases (live broadcast) even impossible. Therefore, a practical approach is to derive coefficients for certain genres, e.g., news, sports, action, etc., which can be used as an approximation for a concrete instance of a given genre.

In the case of adaptation decision-taking, the multimedia quality model can be utilized to improve the quality of the adaptation decision. Instead of trying to maximize both audio and video quality, the knowledge of the perceived multimedia quality can be used for guiding the adaptation. That means that the formula introduced above is used as objective function in the optimization problem.

3. Subjective Quality Assessment

In order to implement an MPEG-21 compliant cross-resource adaptation decision-taking we first determined the coefficients for a given set of audio-visual content. In this section the process of deriving a global multimedia quality function for our content is discussed. We performed subjective tests to derive coefficients for different types of content. The approach is similar to the one in [1] and [6], respectively. First, subjective tests were made to determine the perceived quality of the audio and the video streams separately. In a second step the combinations of audio and video streams in different qualities were assessed. The subjective ratings of the perceived quality are then input to a statistical analysis which results in corresponding coefficients for the multimedia quality model.

3.1. Test material

Representative clips taken from the television programme were used for the experiments. The content was captured from the normal television broadcast and cut to a length of about 10 seconds which is sufficient for the subjective quality assessment.

The monaural audio stream was encoded by using the MPEG-4 Audio BSAC (bit sliced arithmetic coding) codec [9]. The encoder was based on the MPEG-4 Audio reference software. The stream consists of one base layer and 48 enhancement layers, which allows a fine granular scalability of the bitrate between 16 and 64 kbit/s. The adaptation of the bitstream can be accomplished by truncating enhancement layers which results in a lower bitrate and a lower quality. The degradation of the quality was quantified by using the Objective Difference Grade (ODG) [12] as objective quality measure.

The scalable video codec (SVC) [10] was chosen for encoding the video stream. The encoder was based on the JSVM 3.3 reference software. The selected frame size was QCIF (176 x 144 pixels) and the frame rate was 25 fps. It was not possible to use larger frame sizes at this frame rate because of performance lacks of the SVC decoder. Quality scalability was used as the only possible adaptation dimension. The resulting video quality was measured in terms of PSNR.

For the experiment three different clips were used, which were considered to be representative for different genres of content. A description of the clips is given in Table 1. In the following the clips are denoted as *action*, *docu* and *cartoon*. For both the audio and the video stream of the clips, different instances of quality were generated. For each clip 5 different quality levels of video and 5 different levels of audio were generated. The levels were selected in a way that they cover the adaptation possibilities quite uniformly. For the audio-visual content all possible combinations of audio and video quality levels were combined. However, the worst quality of the audio and the video was omitted in order to get a total of 16 different audio-visual degradations (4 audio x 4 video) for each clip.

Table 1. Description of the test content

Name	Description
action (12 sec)	A scene from the science-fiction series "Stargate". It shows an alien that is screaming, surrounded by moving infantrymen. The audio consists of the alien's scream and the shouting of one infantryman.
cartoon (9.5 sec)	A cartoon clip taken out of the cartoon series "Batman". First it shows two persons that are talking and then fades to a photograph of two basketball players.
docu (11 sec)	A scene taken from a documentation, showing a diver with an octopus. The underwater video recording is very greenish with less motion. The audio stream contains the voice of the speaker that explains where the name of the octopus is originated.

3.2. Test methodology

The assessment methodology follows the ITU-T Recommendations P.910 [7] and P.911 [8] for subjective multimedia quality assessment. The Absolute Category Rating (ACR), which is a single stimulus method, was selected for the assessment. Following the ACR method the content, that has to be assessed, is presented for about 10 seconds and afterwards the subject has to rate the quality on a discrete scale. We decided to use the 11-grade scale that is ranging from 0 (worst) to 10 (best). For a better guidance of the subjects, some of the levels had labels that described the quality level (e.g. excellent, poor).

3.3. Experimental environment

The experiment was performed in a dark, sound-insulated room. The video clips were presented on a 17" CRT monitor running at a resolution of 640 x 480 pixels. This relatively low resolution was chosen to display the QCIF-sized video clips in an appropriate size concerning the display's dimensions. For the audio clips two speakers were positioned to the left and the right of the monitor. The subject was sitting on a chair in front of the monitor. The interaction with the application that controlled the experiment was done by a mouse that was positioned on the right side of the monitor.

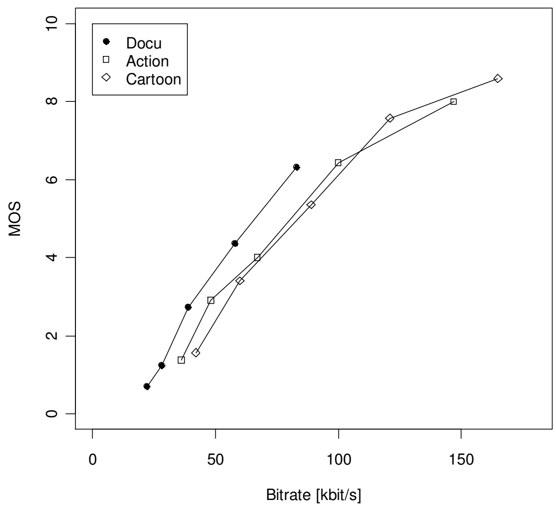

Figure 1. Video bitrate and perceived quality

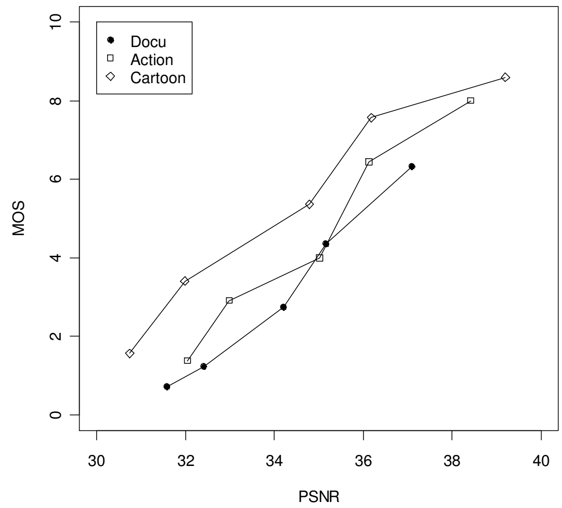

Figure 2. Objective and perceived video quality

3.4. Test procedure

At the beginning of the test each subject was briefed using written instructions to ensure equal conditions and information for each participant. After the briefing the subjects had the possibility to ask questions in case of any obscurities. Then the subjects moved into the room where the experimental environment was set up. The experiment was controlled by an application that managed the playback of the clips, the logging of the subject's voting and the guidance of the subject through the experiment.

The experiment itself consisted of three different parts. In Part 1 the video test material was presented and assessed by the subject. Part 2 consisted of the audio material and finally, in Part 3, the audio-visual test material was assessed by the subject. The procedure within each part was always the same. First the clips with the highest and the lowest quality were presented to the subject. These references should enable the subject to get an idea of the minimum and maximum of the quality spectrum. The subject could decide to see the sequence of references again and again, until it was ready to perform the actual assessment. During the assessment the test material was presented in a random order. In Part 1 and 3 the videos were presented in the centre of the screen and were surrounded by a grey background according to [7]. After the playback of each clip the subject had to rate the quality by using a vertical slider, that was presented in the middle of the screen. The initial position of the slider was at a rating of 5. The subject was told to assess the clip by moving the slider to a position that expressed the perceived quality. After a subject rated a clip no modification or retraction of the rating was possible. Three subsequent clips were used as replication and were presented twice and in the same order. These replications were inserted to test intra-subject reliability and to remove the voting of unreliable subjects afterwards.

3.5. Subjects

22 subjects were involved in the experiment. All of them were students at the University of Klagenfurt. Their age ranged from 19 to 35 with an average at 25 years. The subjects were selected in a way that the same number of male and female students participated in the experiment. All of the subjects reported normal or corrected vision and normal hearing. None of the participants was working or had knowledge in the field of audio or video encoding or multimedia quality. All of them were non-technicians and the majority of them were students of business administration, psychology and journalism.

4. Results

The ratings of the clips were logged in text files by the application, which build the basis for the statistical analysis of the data. For the statistical analysis the GNU R package [11] was used. At the beginning the ratings of the unreliable subjects were removed by using the replications. This removal of ratings was performed separately for each of the three parts of the

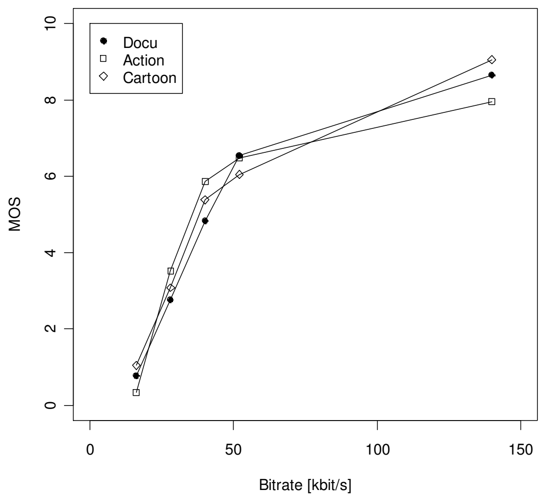

Figure 3. Audio bitrate and perceived quality

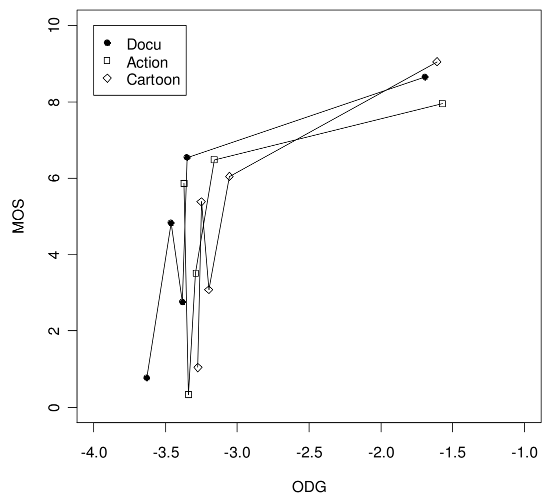

Figure 4. Objective and perceived audio quality

experiment. Afterwards, the Mean Opinion Score (MOS) for each clip and quality level was calculated.

First of all the perceived quality of the audio and video resource was examined individually. The figures show that perceived quality of the video is decreasing at lower bitrates which can be seen as chart in Figure 1. Furthermore, a good correlation between the objective quality (PSNR) and the subjective quality (MOS) of the video clips can be observed (Figure 2).

The same trend can be observed for the audio clips in Figure 3. The perceived quality is decreasing drastically at low bitrates. However, the difference between the MOS for the audio bitrates of 52 kbit/s and 40 kbit/s seems to be very low for the *action* and the *cartoon* clip. Another interesting observation is that the perceived difference between the highest possible quality at 140 kbit/s and the quality level at 52 kbit/s is not that high as suggested for two of the three clips. When comparing the perceived quality and the objective quality in terms of the Objective Difference Grade (ODG) no clear correlation can be found for all quality levels (Figure 4). This is due to the fact that the ODG was designed as an objective measure for high quality audio and is therefore not suitable for measuring low quality audio clips.

In order to investigate the contribution of the audio and video quality to the perceived global quality, a regression analysis was performed for each of the three clips. Therefore, a linear model that predicts the global quality as function of the subjective audio and video quality was generated. Interestingly, for all three clips the model achieved a better precision when omitting the multiplicative term, which means that the coefficient *c* is zero. In the following the statistical analysis for each of the clips is discussed in detail.

4.1. Content *action*

The influence of both audio and video quality on the perceived global quality of the *action* clip is illustrated in Figure 5. Statistical analysis showed that the video quality alone predicts the global quality with $R^2 = 0.647$. The contribution of the audio quality to the global quality is significant lower with $R^2 = 0.192$. The linear model that combines both audio and video quality achieves a very good prediction ($R^2 = 0.904$). The resulting model is as follows

$$MMQ = 0.488 \cdot AQ + 0.644 \cdot VQ - 1.7$$

The coefficients show that the influence of the video quality is higher than those of the audio. Expressed in percentages of the global quality the contribution of audio quality and video quality is 43 and 57, respectively.

4.2. Content *cartoon*

When analyzing the *cartoon* clip the data show that the video quality is a predictor for the global quality with $R^2 = 0.589$. The audio quality alone predicts the global quality with $R^2 = 0.237$. The linear model that includes both audio and video quality achieves a better prediction, with an $R^2 = 0.890$. The model for the cartoon clip is as follows

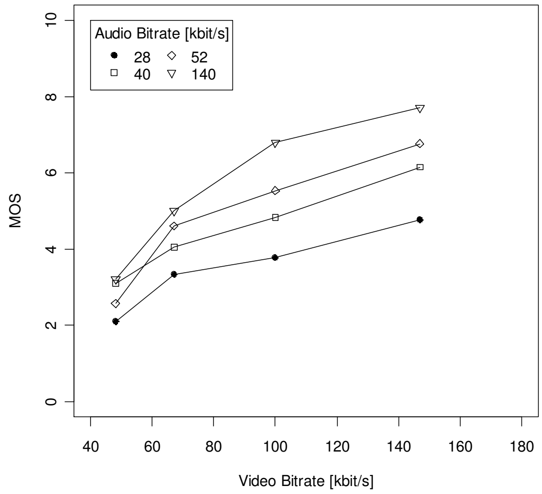

Figure 5. Cross-resource interaction - content *action*

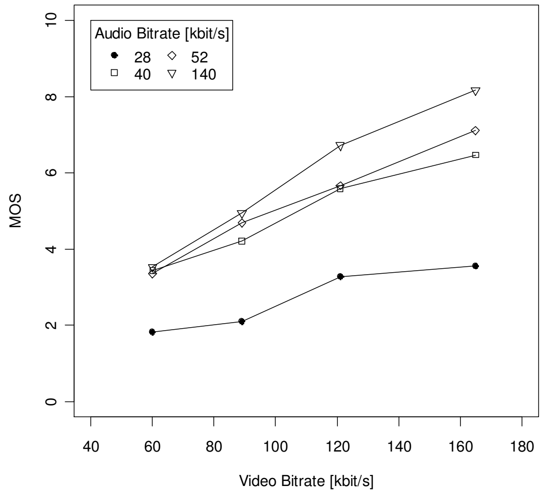

Figure 6. Cross-resource interaction - content cartoon

$$MMQ = 0.415 \cdot AQ + 0.7 \cdot VQ - 0.75$$

One can see that the dominance of the video quality is higher than for the action clip. The weightings of audio and video can be expressed as 37 and 63 percentages, respectively. Figure 6 shows the influence of the video quality and the audio quality at different quality levels. One can see that the impact of the audio quality on the global quality is comparatively low, especially at low video bitrates. However, when using a poor audio quality it significantly decreases the perceived global quality.

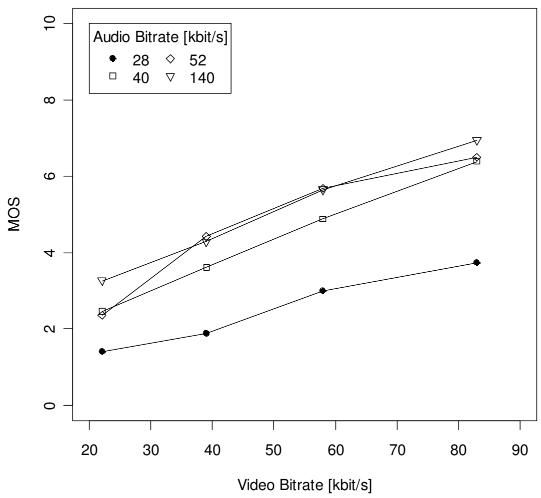

Figure 7. Cross-resource interaction - content *docu*

4.3. Content *docu*

The last clip under investigation was the *docu* clip. By applying statistical analysis it can be shown that the video quality alone predicts the global quality with $R^2 = 0.589$. The audio quality alone as predictor for the global quality results in an $R^2 = 0.237$. The combination of both audio and video quality results in the following linear model with an $R^2 = 0.890$.

$$MMQ = 0.415 \cdot AQ + 0.7 \cdot VQ - 0.75$$

Again, the video quality is dominating the global quality. This behavior can also be seen in the diagram in Figure 7. As with the cartoon clip, the impact of the audio quality seems to be low at an average to high audio quality level. Nevertheless a bad audio quality drastically reduces the global quality.

5. MPEG-21 implementation

The adaptation decision-taking within MPEG-21 Digital Item Adaptation is based on a mathematical optimization problem. Therefore, we translated the results from the subjective tests and the adaptation possibilities of the resource into a corresponding optimization problem and described it using normative MPEG-21 metadata. However, in this paper we forbear from using the normative XML syntax for lack of space and use an equivalent mathematical notation.

The adaptation possibilities and the resulting properties of a resource can be described by using the Adaptation QoS (AQoS) tool [4]. The description is based on variables and their functional dependencies, which are called *IOPins* and *Modules* in the MPEG-21 terminology. For the AQoS description of the video content we introduce the variables FGS, VIDBITRATE, PSNR, and VIDMOS. As the quality scalability is the only adaptation dimension we consider, the only parameter for the adaptation is the amount of the fine-grained scalability layer to use (FGS). When adapting the resource with a given value for FGS it results in a certain average bitrate of the video (VIDBITRATE), a lower PSNR value (PSNR) and in a subjective quality (VIDMOS). The adaptation possibilities of the audio resource are quite similar. The audio stream encoded with BSAC offers the truncation of up to 48 enhancement layers to adapt the bitrate. The parameter for the adaptation is the number of layers to be removed, expressed by the variable NUMLAYERS. The resulting properties of the resource are the bitrate (AUDBITRATE) and both a subjective and objective measure of the quality (AUDMOS and ODG, respectively). These dependencies are realized as look-up tables in the Adaptation QoS descriptions and can be stated formally as follows:

$$
\begin{aligned}
&FGS \rightarrow VIDBITRATE \\
&FGS \rightarrow VIDMOS \\
&FGS \rightarrow PSNR \\
&NUMLAYERS \rightarrow AUDBITRATE \\
&NUMLAYERS \rightarrow AUDMOS \\
&NUMLAYERS \rightarrow ODG
\end{aligned}
$$

Another important part of information for the adaptation decision-taking is the usage context of the consumer. It covers both preferences of the user and technical aspects like capabilities and limitations of the end-device and the delivery networks. MPEG-21 introduces the Usage Environment Description tool (UED) [4] to specify the usage context in a normative way. For the cross-resource use case there are some parts of the UED that are of particular interest: the average available bandwidth of the network (in kbit/s), the maximum bitrate of the terminal's SVC decoder, and the maximum bitrate of the terminal's BSAC decoder (both in kbit/s). It is obvious that the adaptation has to be performed in a way that neither the audio nor the video bitrate exceeds the capabilities of the corresponding decoder, i.e., its maximum bitrate. Furthermore, to enable a transport of the media via the network the sum of the audio and video bitrate has to be lower than the available network bandwidth.

These restrictions can also be expressed by using MPEG-21 DIA metadata. The Universal Constraint Description tool (UCD) enables the formulation of expressions that constrain the values of variables (=IOPins) by using values taken from the usage context (UED). In MPEG-21 this kind of constraints is called limitation constraint because it limits the solution space of the optimization problem. An adaptation decision, which is actually an assignment of a value for each variable, has to satisfy each of the constraints to enable an adequate adaptation. The notation used for the limitation constraints is XML-based and uses the reverse polish notation, which is also known as postfix notation. These expressions are called stack functions within MPEG-21 because this notation allows a simple evaluation by using a stack. For the cross-resource adaptation decision-taking the following three limitation constraints can be stated

$$
\begin{aligned}
&VIDBITRATE \le svcdecbitrate \\
&AUDBITRATE \le bsacdecbitrate \\
&AUDBITRATE + VIDBITRATE \le nwbwidth
\end{aligned}
$$

In the mathematical notation used above, VIDBITRATE and AUDBITRATE are references to the IOPins within the Adaptation QoS description. The arguments on the right side of the constraints, which are used as upper limits, are references to values taken from the UED.

In addition to the limitation constraints the UCD tool offers the specification of optimization constraints. They guide the selection of the adaptation parameters from the set of variable assignments that satisfy all of the limitation constraints. Optimization constraints are also expressed as stack functions, with the additional information whether the function value has to be maximized or minimized. In most cases the optimization constraint aims at achieving the best quality within the possible adaptation parameters. In the case of the cross-resource adaptation decision-taking the formula derived for the global multimedia quality is used. In the case of the *docu* content the following optimization constraint is used:

$$
\text{maximize} \quad AUDMOS \times 0.415 + VIDMOS \times 0.7 - 0.75
$$

The limitation and optimization constraints in combination with the variables and their functional dependencies finally represent a mathematical optimization problem. The optimal adaptation parameters can be calculated by solving the optimization problem [13], which is mostly performed by a generic software component called Adaptation Decision-Taking Engine (ADTE) [15]. As this component operates on normative metadata, cross-resource decision-taking can be achieved by exploiting components that already exist in an MPEG-21 infrastructure.

6. Conclusions

In this paper we discussed the adaptation decision-taking process for audio-visual content referred to as cross-resource decision taking. We used a simple multimedia model that predicts the perceived global quality of an audio-visual clip using the audio and video quality. Based on this model we performed subjective tests to derive content specific coefficients for different audio-visual content. The results from the tests were then translated into MPEG-21 DIA metadata that describes a mathematical optimization problem. Adaptation decision-taking within MPEG-21 is done by solving this optimization problem. As the metadata is normative, a generic software component called Adaptation Decision-Taking Engine can be used for this purpose. The cross-resource decision-taking can therefore be realized without implementing or introducing new software components. It can rather be seen as an application of the powerful MPEG-21 Digital Item Adaptation framework. Besides, the metadata approach is not limited to this specific additive multimedia model. As all dependencies and adaptation possibilities are expressed using metadata, this approach is open and flexible to support more sophisticated multimedia models that may emerge in the future.

7. Acknowledgement

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

8. References

[1] R. Mohan, J. R. Smith, and C.-S. Li, "Adapting Multimedia Internet Content for Universal Access", IEEE Transactions on Multimedia, Volume 1, Issue 1, pp. 104-114, March 1999.

[2] L. Böszörményi, H. Hellwagner, H. Kosch, M. Libsie, S. Podlipnig, "Metadata Driven Adaptation in the ADMITS Project", EURASIP Signal Processing: Image Communication Journal, Vol. 18, No. 8, pp. 749-766, September 2003.

[3] D. Jannach, K. Leopold, C. Timmerer, and H. Hellwagner, "A Knowledge-based Framework for Multimedia Adaptation", Applied Intelligence – Special Issue: Innovations in Applied Artificial Intelligence, Vol. 24, No. 2, pp. 109-125, 2006.

[4] ISO/IEC 21000-7:2004, Information Technology – Multimedia Framework – Part 7: Digital Item Adaptation, 2004.

[5] D.S. Hands, "A Basic Multimedia Quality Model", IEEE Transactions on Multimedia, Volume 6, Issue 6, pp. 806 – 816, December 2004.

[6] S. Winkler, C. Faller, "Audiovisual Quality Evaluation of Low-Bitrate Video", Proceedings of SPIE/IS&T Human Vision and Electronic Imaging, Volume 5666, pp. 139 – 148, January 2005.

[7] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications", International Telecommunication Union, Geneva, Switzerland, 1999.

[8] ITU-T Recommendation P.911, "Subjective audiovisual quality assessment methods for multimedia applications", International Telecommunication Union, Geneva, Switzerland, 1998.

[9] ISO/IEC 14496-3:2005/Amd.2:2006, Information Technology – Coding of audio-visual objects – Part 3: Audio, Amendment 2: Audio Lossless Coding (ALS), new audio profiles and BSAC extensions, 2006.

[10] ISO/IEC 14496-10:2005/Amd.3, Information Technology – Coding of audio-visual objects – Part 10: Advanced Video Coding, Amendment 3: Scalable Video Coding, 2006.

[11] The R Project for Statistical Computing, URL: www.r-project.org.

[12] ITU-R Recommendation BS. 1387-1, "Method for objective measurements of perceived audio quality", International Telecommunication Union, Geneva, Switzerland, 2001.

[13] D. Mukherjee, E. Delfosse, J.G. Kim, and Y. Wang, „Optimal adaptation decision-taking for terminal and network quality-of-service", IEEE Transactions on Multimedia, Volume 7, Issue 3, pp. 452-462, June 2005.

[14] D. Wijesekera, J. Srivastava, A. Nerode, and M. Foresti, "Experimental evaluation of loss perception in continuous media", Multimedia Systems, Vol. 7. pp. 486-499, 1999.

[15] I. Kofler, C. Timmerer, H. Hellwagner, A. Hutter, and F. Sanahuja, "Efficient MPEG-21-based Adaptation Decision-Taking for Scalable Multimedia Content", Proceedings of the SPIE/ACM Multimedia Computing and Networking Conference (MMCN 2007), San Jose, California, to appear.