QoE-Oriented Adaptive SVC Decoding in DVB-T2

Adlen Ksentini, *Member, IEEE,* and Tarik Taleb, *Senior Member, IEEE*

***Abstract*—Associating DVB-T2 and Scalable Video Coding (SVC) constitutes an efficient way for broadcasting added-value video services, such as HDTV and 3D TV, to end-users. The ultimate objective of this new approach of broadcasting video services is to ensure high Quality of Experience (QoE) for end users. Whilst Quality of Service (QoS) is the collective effect of performance that determines the degree of satisfaction of a user of a service, QoE reflects more accurately the user experience, as it is based on human perception when evaluating the video quality. Maximizing user QoE is thus becoming a crucial requirement when deploying new broadcast platforms for the provisioning of high quality video services. The contributions of this paper are two-fold. At first, we introduce a reference-less QoE measurement tool dedicated to SVC coding. Based on a learning function, this tool is able to learn the non-linear relationship between parameters affecting video quality and perceived user QoE. According to several experiments carried out using this tool, we demonstrate that decoding all SVC layers is not always efficient to achieve high user QoE, mostly when SVC enhanced layers experience packet losses. For the sake of maintaining a good QoE, it is worthwhile withdrawing enhanced layers experiencing packet losses and not displaying them to end-users. Based on this observation, we propose a QoE-Based Adaptive SVC Decoding (QoE-BASD) algorithm that assists a video receiver to select the appropriate SVC layers for video decoding in order to maximize QoE. We evaluate the performance of the proposed solution: (i) analytically, by using discrete Markov Chains to model the proposed solution; and (ii) via OPNET-based computer simulations. The obtained results are encouraging, and illustrate the gain achieved by QoE-BASD when compared to the conventional approach.**

***Index Terms*—DVB-T2, QoE, SVC, video broadcast.**

I. Introduction

RECENT trend in video distribution is the use of scalability coding to handle user heterogeneity in terms of user's terminal capacities (e.g., screen resolution and CPU) and network connection speed. Indeed, video scalability encodes the video in multiple separated layers, which enable a large number of users with heterogeneous capability to view any desired video stream from anywhere [1]. In order to achieve video encoding scalability, the H.264/MPEG-4 AVC video compression standard [2], [3] proposes the Scalable Video Coding (SVC) extension, which takes advantage of layered approach already known from previous experiences related to different video coding approaches [4]. SVC supports three fundamental types of scalabilities: spatial, temporal, and quality (Signal-to-noise ratio NR). Usually, SVC stream includes one base layer and one or several enhancement layers. The removal of an enhancement layer still lead to acceptable quality of the decoded video at reduced temporal or spatial SNR. The base layer is conforming to existing H.264/AVC profile, ensuring backward compatibility with existing receivers. SVC offers the possibility to constitute a set of layer combinations to create the video streams, which allow to target different spatial as well as temporal dimensions to be aware of user environment. Accordingly, SVC offers a flexible solution for the content provider, (such as TV broadcasters, VoD providers, and Catchup TV channels) to manage, store, and distribute several video formats toward multiple kinds and scales of terminals, and over different and transient access technologies to reach the end user.

On the other hand, DVB-T2 [5] is the newly standard for digital video broadcasting that aims at replacing the DVB-T (i.e., the first generation of the terrestrial broadcasting standard) for broadcasting terrestrial television. Based on advances made in digital signal processing, and specifically in channel coding, DVB-T2 brings a new flexibility in services' broadcasting with an increased transfer capacity of 50%, when compared to DVB-T. Besides using the Coded Orthogonal Frequency Division Multiplexing (COFDM) in order to be more robust against multipath channels [6], the DVB-T2 physical layer data channel is divided into logical entities called Physical Layer Pipes (PLP). Each PLP carries one logical data stream that could be an audio-visual multimedia stream along with the associated signaling information, or hierarchical video streams which can address at the same time different qualities. The PLP architecture is designated to be flexible so that arbitrary adjustments of robustness and capacity can be easily done. Thus, using different PLPs enables broadcasting, on a single radio channel, multiple services, or groups of services, with different channel coding and modulation settings. Broadcasting several service components over the same channel has thus become possible, with differentiated levels of robustness, which was not possible with the previous DVB-T standard or other broadcasting technologies [7]. Using this new capability allows handling users' channel diversity, where users with good channel condition can decode all PLPs and access to high quality contents, while users with poor channel conditions (such as mobile terminals) can decode only robust PLP but at least can access the lowest service quality; i.e., in case of DVB-T all services are lost. Needless to say that associating SVC with DVB-T2 can easily address the problem of broadcasting added value services to high number of users,

Manuscript received June 20, 2012; revised March 14, 2013; accepted March 18, 2013. Date of publication April 26, 2013; date of current version May 20, 2013. This paper was recommended by Associate Editor L. Zhang.

A. Ksentini is with the IRISA Laboratory, University of Rennes 1, Cedex 35042, France (e-mail: adlen.ksentini@irisa.fr).

T. Taleb is with NEC Europe, Heidelberg, Germany (e-mail: talebtarik@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2013.2254891

despite their radio environments and terminal capabilities, which represent an interesting solution for operators [8]. In fact, the hierarchical physical layer provided by DVB-T2 can be easily combined with hierarchical video coding features proposed by SVC [9]. Thus, each SVC layer is broadcasted through different PLPs. The base layer is sent through the most robust PLP, usually PLP0. The enhanced layer is sent through other PLPs, which use less robust physical modulation, but allow using higher data rates. Thus, stations with good physical channel conditions can decode all the SVC layers and benefit from high video quality, while users with poor channel conditions can at least decode the base layer and benefit from acceptable quality.

Apart of the work presented in [10], most of related research works on SVC and DVB-T2 have focused on exploiting each technology separately. In [10], the authors discuss the deployment of SVC in DVB-T2, and particularly concentrate on providing optimal usage of DVB-T2 features from SVC's point of view. In addition, they provide modifications to the error protection mechanism, at the physical layer, in order to improve users' experience. Besides considering only physical layer enhancements, the authors employed the Picture Signal Noise Ratio (PSNR) tool to calculate user perception, which is known for its lack of efficiency to reflect user's Quality of Experience (QoE).

In this paper, we further concentrate on the enhancement achieved, at the user side, when associating SVC and DVB-T2. We concentrate on user's experience, in terms of QoE, as the main criteria for evaluating this association, and demonstrate how QoE can help optimizing this association for providing context-aware video services to an end-user through the broadcast channel. The contributions of this paper are two folds: (i) introducing a new QoE measurement tool for SVC, called Pseudo-Subjective Quality Assessment (PSQA); (ii) proposing a QoE-Based Adaptive SVC Decoding (QoE-BASD) algorithm that helps a video receiver to select the appropriate SVC layer for video decoding in order to maximize QoE. In addition, we model the behavior of the proposed QoE-BASD scheme through a Time Discrete Markov Chain aiming at analytically analyzing the performance of this mechanism. It shall be noted that QoE-BASD does not need any feedback or signaling messages to work, which make it compliant with unidirectional technologies such as DVB-T2.

The remainder of this paper is organized in the following fashion. Section II presents the concept of user's QoE. Section III gives some details on the Pseudo-Subjective Quality Assessment (PSQA) tool, which is used to monitor user's QoE in case of using SVC. Section IV introduces the concept of QoE-BASD and the Markov model used to analytically evaluate QoE-BASD. Section V presents the simulation model and discusses the obtained results. The paper concludes in Section VI.

II. QoE Measurement Approaches: State of the Art

QoE is defined in [11] as the overall acceptability of an application or service, as perceived subjectively by the end use. Quality of Service (QoS) is defined in [12] as the collective effect of performance which determines the degree of satisfaction of a user of the service. In telecommunications, QoS is usually a measure of performance of the network itself. QoE instead focuses on the overall experience of the user. It depends on the global system behavior, going from the source of the services until the user, including the content itself and the network performance [13][14][15]. There are several factors that can influence QoE for video applications. Characteristics, such as frame rate of a video stream, may impact the fluidity of the video: a lower frame rate means chopiness that can degrade the perceived quality. Spatial video resolution is another significant factor: depending on the limitations of the end device, users may prefer the highest available resolution. Another important factor related to the video quality is the Quantization Parameter (QP). In fact, this parameter relates to the compression of a video stream. Thus, during compression, some amount of information is thrown away and this will introduce certain distortion in the video that may, in turn, have an impact on QoE.

In addition to the preceding remarks, the type of video content, itself, may have significant importance. For example, a video of a news reader might have low frame rate requirements, but higher quantization requirements. On the other hand, a fast moving video, such as that of Formula One racing coverage, requires higher frame rates to ensure good QoE. The network used to provide the service can significantly impact the video quality. For example, packet losses can strongly degrade the video's perceived quality. Delays and jitter in the network may incur, first, a long initial delay before a video can start to play, and then, play-out disruptions and eventual data losses because of the video packets that miss the play-out deadline. In addition, other parameters, such as network bandwidth, impose limitations on the video characteristics because some quality of the video will get downgraded, either by lowering the frame rate or by using more compression, to accommodate the video with the available bandwidth.

Fig. 1. ITU standard scales for subjective test methods.

Depending on the method used to evaluate user QoE, QoE measurement tools can be classified into two categories. The first class is based on subjective evaluation tests while the second class is based on objective evaluation (signal processing algorithm). Subjective evaluation tests are based on personal evaluation of users, where a panel of selected persons rate video sequences. The output of these tests is a Mean Opinion Score (MOS), where different scales can be used as specified by the ITU-R, as shown in Fig. 1. Indeed, the ITU-R recommendation BT.500-10 formalizes the subjective test procedure by introducing several experimental conditions, such as viewing distance and viewing conditions. Although subjective tests are highly accurate in estimating users' QoE, their preparation and execution are costly and time consuming. Furthermore, they cannot be used in real-time or automatically.

Objective quality evaluations, on the other hand, are algorithms and formulas (i.e., generally signal processing algorithms) that measure, in a certain way, the quality of a video stream. Objective video quality metrics range from very simple metrics to very complex ones [16] (e.g., metrics based on human vision systems (HVS)). Objective quality metrics can be further classified into three different categories, namely Full-Reference (FR), Reduced-Reference (RR) and No-Reference (NR), based on the amount of information available for comparison with the original content. Full- and reduced reference mechanisms are mainly used to evaluate video quality in non real-time scenarios where both the original video reference (or reduced data set) and the distorted video are available.

The PSNR metric is one of the most widely used full-reference metrics for objective video quality assessment, thanks to its simplicity and its low computational requirements. It calculates the ratio between the maximum value of a signal and the background noise. PSNR is used because of its physical significance and simplicity but the performance of this metric is quite poor, as it usually does not correlate well with subjective scores. Also, it is difficult to reliably derive MOS from this metric, despite the existence of heuristic mappings of PSNR to MOS. The Structural Similarity (SSIM) approach [17] provides an alternative and complementary way to tackle the problem of video quality assessment. It is based on a top-down assumption that HVS is highly adapted for extracting structural information from the scene, and hence a measure of structural similarity should be a good approximation of perceived image quality. Nevertheless, the SSIM index achieves the best performance when applied at an appropriate scale (i.e., viewer distance/screen height). Calibrating the parameters, such as viewing distance and picture resolution, represents an important challenge for this approach. Video Quality Metric (VQM) [18] is a standardized method for objectively measuring video quality by making a comparison between the original and the distorted video sequences based only on a set of features extracted independently from each video. The algorithm used by VQM measures the perceptual effects of several video impairments, such as blurring, jerky/unnatural motion, global noise, block distortion, and color distortion. These measurements are combined into a single metric that gives a prediction of the overall quality. For more details on FR methods reader can refer to [17]. Despite their efficiency in evaluating the video quality, FR methods are only applicable when the original video sequence is available. This constitutes a limitation when there is a need to evaluate QoE in real-time at the decoder side. To address such situation, NR methods have been proposed. The main goal of NR methods is to create an estimator based on the proposed features that would predict the MOS of human observes, without using the original image or sequence data. Furthermore, since the model does not require any comparison of signals, the calculations can be performed in near real-time. Previously introduced NR methods do not estimate the overall user quality but estimate the degree of blockiness [19], which is the most prominent artifact of block-DCT based compression methods such as H.26x, MPEG and their derivatives. To increase the estimation accuracy, work in [20] uses the bit-stream information which depends on the compression algorithm. Such method suffers from the fact that it cannot differentiate video quality degradation from features of the video itself and those introduced by the network (e.g., loss and delay). Other NR methods incorporate network information to enhance the user quality prediction. The ITU recommendation ITU G.1070 [21] (known also as opinion model) is a NR model which uses the bit rate and frame rate of the compressed video along with the expected packet loss rate of the channel to predict the quality video. Work in [22] showed that it is possible to enhance this model and make it more precise by replacing, for example, packet loss rate with packet loss event rate. A further extension of this mechanism is proposed in [23]. Pseudo Subjective Quality Assessment (PSQA) [24] is a quality assessment tool that is a hybrid between subjective and objective evaluation techniques. PSQA belongs to the category of NR approaches by allowing evaluating the video quality in real-time without requiring the original video sequence. A comparison between PSQA and other QoE metrics tools is presented in [25]. The following section gives more details on PSQA and how it can be adapted to evaluate the SVC video quality.

III. Pseudo Subjective Quality Assessment (PSQA)

A. PSQA Principle

PSQA is built on the idea of doing subjective tests for several distorted videos and uses the results of this evaluation to teach a learning function the relation between the parameters that cause the distortion and the perceived quality (see Fig. 2). The procedure consists of choosing a set of P parameters (selected a priori), which may have an impact on the perceived quality. For example, we can select the video codec used, the packet loss rate of the network, the mean loss burst size, the end-to-end delay and/or Jitter. Let this set be denoted as $P = \{\Delta_1, \Delta_2, ..., \Delta_n\}$. Once these quality-affecting parameters are defined, it is necessary to choose a set of representative values for each, together with an interval where the parameter is bounded, according to the conditions under which we expect the system to work. The number of values to choose for each parameter depends on the range of the chosen interval and on the desired precision. For instance, if we consider the packet loss rate as one of the parameters, and if we expect its values to range mainly from 0 to 5%, we could use 0, 1, 2, 5 and perhaps also 10% as the selected values. In this context, we call a configuration a set denoted as $\Omega = \{v_1, v_2, ..., v_n\}$, whereby v_i is one of the chosen values for parameter p_i. The total number of possible configurations is usually large. Therefore, the next step is to select a subset of S configurations to be subjectively evaluated. This selection may be done randomly, but it is important to cover the points near the boundaries of the configuration space. Mathematically speaking, the corresponding sampling sequence is called a low-discrepancy one. It is then not necessary to use a uniform

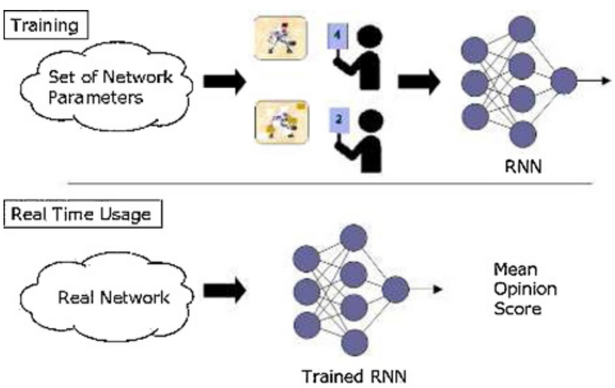

Fig. 2. PSQA methodology.

distribution. Instead, there is a need to sample more points in the regions near the configurations which are most likely to happen during normal use, or the ones considered as the most important configurations. Once the configurations have been chosen, we build a set of 'distorted samples', that is, samples resulting from the transmission of the original media over the network under the different chosen configurations. For this, we use a testbed or a network simulator, or a combination of both.

We subsequently select a set of M media samples (δ_m), $m = 1, \ldots, M$, for instance, M short pieces of video (i.e., subjective testing standards recommend the usage of sequences of an average duration of 10s). We denote the set of S configurations already sampled by $\{\Omega_1, \ldots, \Omega_S\}$ where $\Omega_S = \{v_{s1}, \ldots, v_{sp}\}$, v_{sp} being the value of parameter Δ_p in configuration Ω_S. From each sample δ_i we build a set of samples $\{\delta_{i1}, \ldots, \delta_{iS}\}$ that have encountered varied conditions when transmitted over the network in the following manner: sequence δ_{iS} is the sequence that arrived at the receiver when the sender sent δ_i through the source-network system where the P chosen parameters had the values of the configuration Ω_S. Since the distorted samples are generated, a subjective test is carried out on each received piece δ_{iS}. After statistical processing of the answers (designed for detecting and eliminating bad observers, that is, observers whose answers are not statistically coherent with the majority), the sequence δ_{iS} receives the value μ_{iS} (often, this is a MOS). The idea is then to associate each configuration Ω_S with the value:

$$\mu_s = \frac{1}{M} \sum_{m=1}^{M} \mu_{mS} \tag{1}$$

At this step, there is a quality value associated with each configuration Ω_S. We now randomly choose S_1 configurations among the S available configurations. These, together with their values, constitute the Training Database. The remaining $S_2 = S - S_1$ configurations and their respective values constitute the Validation Database, reserved for further (and critical) use in the last steps of the process.

The next phase in the process is to train the learning function in order to learn the mapping between configurations and values as defined by the Training Database. Assume that the selected parameters have values scaled into [0,1] and the same with quality. Once the learning function has captured the mapping, that is, once the learning function has been trained, we have a function $f()$ from [0,1] into [0,1], mapping any possible value of the (scaled) parameters into the (also scaled) quality metric. The last step is the validation phase: we compare the value given by $f()$ at the point corresponding to each configuration Ω_S in the validation database to its quality valueμ_{iS}; if they are close enough, the training is validated. If the validation fails, we must review the chosen architecture and configurations. Based on the trained learning function (that is, the $f()$ function), it is possible to assess the performance of all submitted sequences according to each network configuration Ω_S.

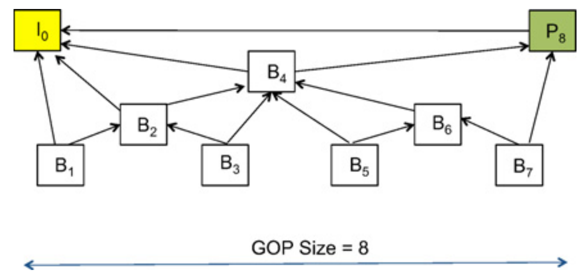

Fig. 3. GOP structure in single layer case (SVC).

B. PSQA for SVC

To Build PSQA for SVC, we need to clearly identify the parameters impacting the perceived quality of the video streams (as SVC is composed by several layers). There are different parameters that affect the quality of H.264/SVC videos. They include parameters related to the content itself (such as brightness, contrast, sharpness, color, motion), the encoding parameters (such as QP) and other parameters dependent on the transport network (such as delays, loss rate, bandwidth). To efficiently use PSQA, it is important to consider parameters that can be obtained in real-time and with low complexity. The first parameter affecting the quality of the SVC streams is the frequency of IDR (Instantaneous Decoder Refresh) at the encoding side. Unlike coding mechanisms such as MPEG, SVC uses a different structure of the Group Of Picture (GOP). In SVC, the GOP structure consists of one key frame (IDR and P) and the remaining are B frames [26]. In addition, SVC adopts a hierarchical coding structure for B frames, as illustrated in Fig. 3, to facilitate the temporal scalability implementation.

IDR frames (I frame) are special frames due to the fact that they are encoded without reference to another frame. From hereunder, we indifferently denote them as IDR or I frames. The IDR frames are periodically sent in order to refresh the decoder buffer and create a new point of reference. In fact, an increase in this IDR frequency (IDR period) means an increase in the number of IDR frames that, in turn, decreases the number of P frames and B frames. High numbers of IDR frames are beneficial to reduce error propagation during the refresh period. In Fig. 4 we illustrate the dependence of affected frames on the location of key picture within the intra refresh period.

The affected frames vary according to the lost frame type (I, P or B) as well as to the position of the lost frame in the

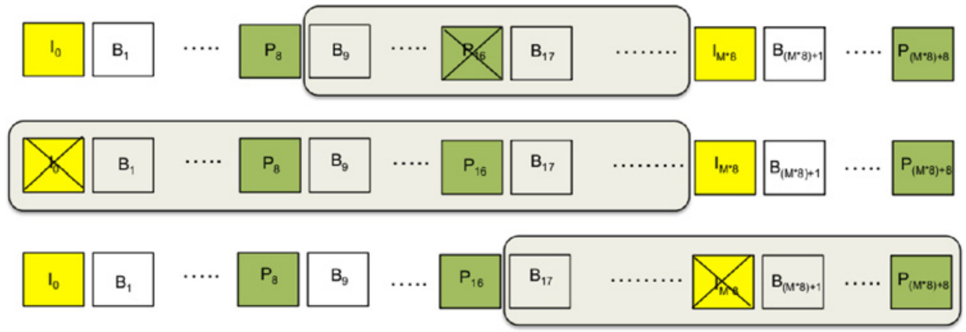

Fig. 4. Frames range affected by the loss of key frames (single layer case).

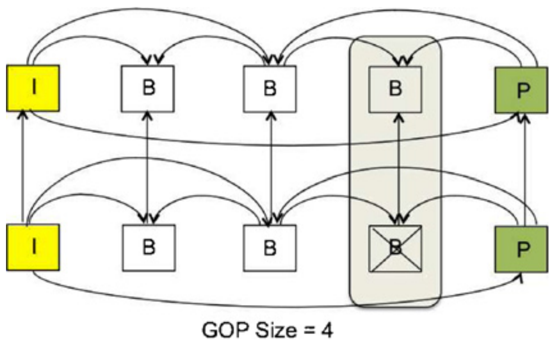

Fig. 5. Frames range affected by the loss of a B frame (base layer and one enhanced layer (CGS) case).

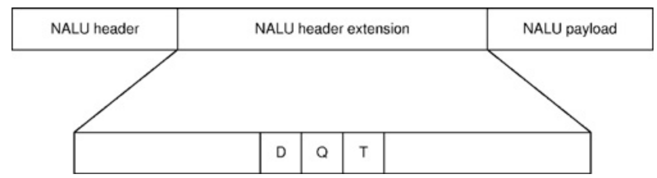

Fig. 6. Typical NALU format.

intra-refresh-period (or IDR period). Unlike MPEG, loosing a B frame has impact on other B frames, of course depending on its position in the hierarchy. For instance, in Fig. 3, loosing frame B4 impacts all other B frames in GOP. However, loosing frame B2 has no impact. Loosing a P frame has an impact on all frames (B frames) that use this frame as a reference. The worst case occurs when an IDR frame is lost. Indeed, loosing such frame causes distortion not only on the current GOP, but also to precedent GOPs. Loosing IDR frames affects both P and B frames.

As noticed in the precedent version of PSQA [27], the packet loss rate is an important parameter that affects the video quality. This is more evident in case of streaming SVC over UDP-based networks or over DVB, where the transport layer is not reliable. For SVC, we must consider the packet loss rate for each layer composing the SVC stream. The NALU (Network Abstraction Layer Unit) is the transport unit of video packet used in H.264 (SVC and AVC). In case of SVC, NALU can only carry information of one layer. The loss of a NALU affects only a single layer. However, it is worth noting that losing a NALU belonging to the base layer has more impact on the video quality, than the loss of a NALU belonging to other enhancing layers (particularly when using spatial or quality scalabilities). Besides affecting the other frames of the base layer (Fig. 4), a loss of base layer NALU impacts the other layers as all the other layers in SVC use the base layer as reference and any error in this layer propagates to other layers (see Fig. 5 for intra layer prediction in case of SNR (CGS) scalability). Hence, this situation seriously downgrades the video quality.

Usually, a NALU packet consists of one header (as AVC header), and a specific header extension (Fig. 6). This extension has particular fields D, Q and T, which are used to identify the spatial quality and temporal layers, respectively. Hereunder, we denote by $P = \{L_{BL}, L_1, L_2, ..., L_N, f_{IDR}\}$ the set of affecting parameters, where f_{IDR} represents the IDR frequency, L_{BL} and L_N denote the NALU loss rate of the base layer and layer N, respectively. It is important to note that there is another parameter that affects the quality of H.264/SVC videos, which is the error concealment procedure used by the decoder. Indeed, such mechanism can help the decoder to replace information lost due to NALU losses. In PSQA, the error concealment is implicitly taken in consideration as the distorted videos are evaluated at the subjective phase and are obtained after the decoder has applied the error concealment procedure. Therefore, in this version of PSQA, we are relying on the error concealment provided by the SVC decoder.

As stated before, PSQA needs to be able to measure those parameters automatically, above all in real-time. Usually, f_{IDR} is a static value; it can be obtained when encoding the video stream. Another way of obtaining f_{IDR} is by tracking the appearance of the IDR frames in the video stream. An IDR frame, in turn, can be identified by parsing the NALU headers that contain the information about whether its payload corresponds to an IDR frame or not. As mentioned before, higher values of the IDR frequency means higher resilience to the error propagation, but for the price of larger video sizes. On the other hand, the NALU loss rate $(L_{BL}, L_1, L_2, ..., L_N)$ for each SVC layer are obtained by relying on the Real Time Protocol (RTP) layer. Combined with RTP simple packetisation mechanism (Single NAL Unit) [28], we propose using a multi-session RTP connection for each layer. In other words, for each layer an independent RTP session is established, where one NALU is conveyed in one RTP packet. Thus, we can obtain the NALU loss rate for each SVC layer at the RTP level, either

TABLE I
VIDEOS USED FOR TRAINING

Video	NAL number
CITY	300
CREW	300
HARBOUR	300
ICE	240
SOCCER	300

TABLE II
VIDEO PARAMETERS

Fixed parameter	Value
Resolution	704x576 (4CIF)
Frame Rate	30
Layer(QP)	BL(34) - EL1(28) - EL2(22)

TABLE III
THE VALUES OF PARAMETERS AFFECTING QOE.

Parameter	Set of values
NALU loss rate for Base Layer (%) (L_{BL})	0, 0.3, 0.5, 1, 3, 5, 10
NALU loss rate for Layer 1 (%) (L_1)	0, 0.3, 0.5, 1, 3, 5, 10
NALU loss rate for Layer 2 (%) (L_2)	0, 0.3, 0.5, 1, 3, 5, 10
IDR Frequency	75, 150, 300

at the decoder side, or by enabling the RTCP protocol for a remote monitoring if there is a return channel available from the receiver to the server. IP traffic is supported in DVB-T2 by either using MPEG-TS encapsulation or through the Generic Stream Encapsulation (GSE) [29] that provides an appropriate IP encapsulation over PLPs.

C. PSQA for SVC Validation

In order to train the learning function $f()$, which monitors user's QoE, we used five different video sequences (m) (Table I) for constituting the set of the media sample (δ_m). We encoded the different videos by using the JSVM encoder [30]. For the decoder side, we used the openSVC soft [31]. The resolution is 4CIF (704 x 576), the frame rate is set to 30 frames/s, and the values of Quantization Parameter (QP) are set to 34, 28, 22, respectively for layer 0 (Base Layer), 1 and 2 as shown in Table II. The video scalability is based on the SNR (CGS) quality, by reducing the QP parameter for each enhanced layer by Delta QP (DQP). As our focus is on employing only quality-based scalability, we used only key frames (IDR and P frames) to constitute the GOP of each video sequence. In this case the GOP size is equal to 1, which means that no B frames are used and the GOPs are in the form of "PPP...PIPPP...".

Since the NALU losses (L_{BL}, L_1, L_2) have a serious impact on the final quality of a video, we proposed that each value of these parameters (NALU losses) is to be taken from the set noted $V = \{0, 0.3, 0.5, 1, 3, 5, 10\}$. The set V represents the loss rate (i.e., from 0% to 10%, quality is already very bad at 10% loss). Regarding the parameter f_{IDR}, three values were considered: 75, 150 and 300. Table III shows the set of values associated with each parameter.

TABLE IV
THE VALUES OF PARAMETERS AFFECTING QOE.

MOS	Quality	Level of Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

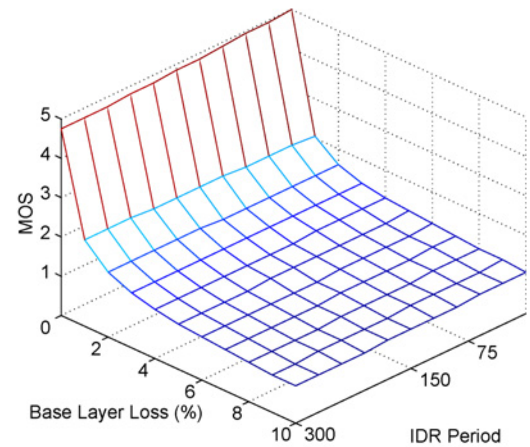

Fig. 7. PSQA scores vs. L_{BL} and f_{IDR} with L_1 = 0% and L_2= 0%.

A large number of videos were created using different combinations of the above parameters and their values given in Table III. The obtained sequences correspond to the set of configurations $\{\Omega_1, \Omega_2, ..., \Omega_n\}$. For instance, $\Omega_1 = \{0, 0, 0, 75\}$ represents a configuration, where $L_{BL} = 0\%$, $L_1 = 0\%$, $L_2 = 0\%$, $f_{IDR} = 75$. Next steps consist of: (i) reducing the set of video sequences $\{\delta_{1S}, ..., \delta_{iS}\}$ to 500 by employing uniform sampling; (ii) using manual evaluation to select around 100 among 500 for the subjective test session, such that 25 videos corresponded to a MOS score between 1 and 2, 25 videos between 2 and 3, and so on. The MOS scale, shown in Table IV, was used for all quality evaluations. These 100 video sequences were evaluated by 15 humans using DSIS (Double Stimulus Impairment Scale) methodology [32]. Indeed, the reference video is first shown to the evaluators' panel, and then followed by the distorted sequences. This method is more useful for evaluating clearly visible impairments caused by transmission. The obtained scores were then pre-processed, as in [24], to check for inconsistent scorers. Then the data was used to train the learning function.

The obtained results with PSQA are shown in Figs. 7, 8 and 9. Fig. 7 represents the relation between QoE (PSQA scores) and the two parameters L_{BL} and f_{IDR}. The obtained results show that the video quality is highly sensitive to losses in the base layer and the value of MOS quickly decreases to 1.0 along with an increase in NALU loss rate of the base layer (L_{BL}). Even a NALU loss rate of around 1% of the base layer degrades the value of MOS to a value lower than 3 out of 5. This, in turn, corresponds to impairment in the video quality as slightly annoying (Table IV). This fast degradation, due to NALU loss rate of base layer, is attributable to the fact that

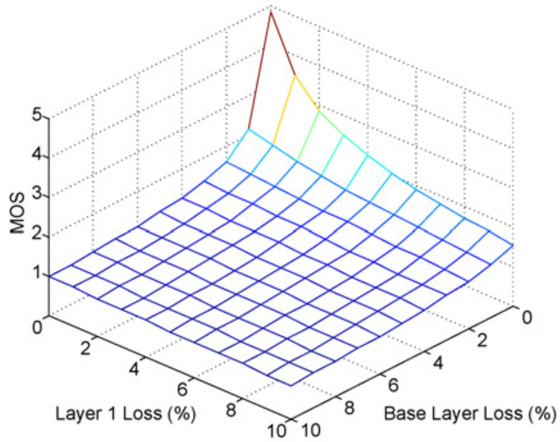

Fig. 8. PSQA scores vs. L_{BL} and L_1 with f_{IDR}=300 and L_2= 0%.

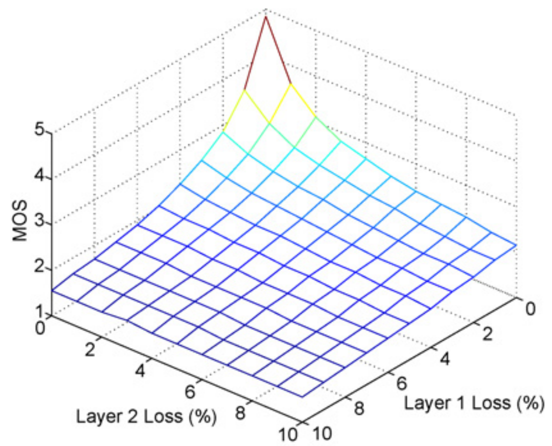

Fig. 9. PSQA scores vs. L_1 and L_2 with f_{IDR} =300 and L_{BL} = 0%.

all the other layers in SVC use the base layer as reference and any error in this layer propagates to other layers. In addition, it can be seen that MOS slightly decreases with increasing values of f_{IDR}, IDR frame frequency. This is because a high value of f_{IDR} means that a particular error has higher chances of propagating to a large number of frames until an IDR frame arrives and refreshes the decoder. For example a value of f_{IDR} = 300 means that an error occurring in the 1st frame can propagate up to a maximum of 299 frames, in other words around 10 seconds of a video being streamed at a rate of 30 frames per second.

Fig. 8 shows the value of MOS predicted by PSQA with varying L_{BL} and L_1. It can be seen that the quality is more sensitive to an increase in NALU loss rate of base layer as compared to that of layer 1. Whereas, it can be seen in Fig. 9 that QoE is, relatively, only slightly more sensitive to an increase in L_1 as compared to L_2. This is due to the fact that the base layer in SVC is the most sensitive to the losses as compared to other layers. If there is no loss in the base layer, and there is a loss in other layers, then the decoder has a better chance of error concealment as compared to the case when the data from the base layer itself are lost. From these results we can definitely conclude that the QoE has a non-linear relationship with the parameters affecting QoS.

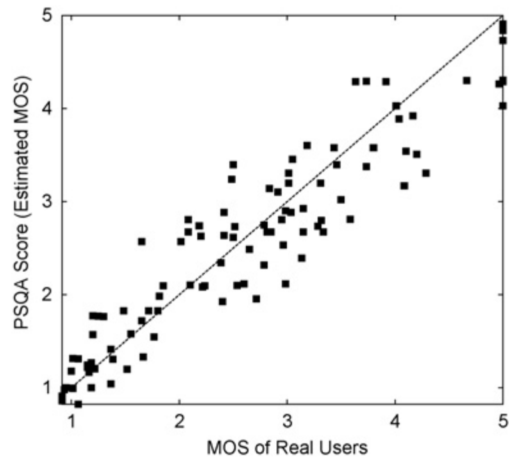

Fig. 10. PSQA scores vs. scores given by real users.

Fig. 10 shows the scatter plot with estimated MOS vs. actual MOS obtained from the subjective tests. The diagonally plotted line corresponds to the case when the estimated MOS value is equal to the actual MOS value. Thus, points lying close to the line would indicate the accuracy of the estimation tool. As it can be deducted from Fig. 10, the scatter plot shows a good accuracy of the estimation. This is reflected by the overall mean square error of about 0.1777 on the MOS scale from 0 to 5. In order to further validate the good accuracy of PSQA to estimate QoE, we use the Pearson Correlation Coefficient (PCC) as recommended by the Video Expert Group (VQEG) [33] to validate the objective models for video quality assessment. PCC is the linear correlation coefficient between the estimated MOS and the subjective MOS. It measures the prediction accuracy of a metric, i.e., the ability to predict the subjective quality ratings with low errors. For N data pairs (x_i; y_i), with $\overline{x}$ and $\overline{y}$ being the means of the respective data sets, the PCC is given by:

$$PCC = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \overline{y})^2}} \tag{2}$$

In the case of PSQA, the mean PCC for the tested video is equal to 0.9406, which represents a good result as the highest accuracy obtained with a PCC equals to one. In Fig. 11, we present the PSQA scores computed for the following three cases: (i) the three layers are decoded and displayed to users, where the enhanced layer 2 experiences NALU loss; (ii) the two lower layers are decoded; (iii) only the base layer is decoded. Although NALUs belonging to the enhanced layer 2 have less importance than those of the base layer, loosing NALUs from this layer degrades the user's QoE. Therefore, if the enhanced layer 2 (highest layer) experiences losses, for the sake of maintaining good QoE, it is worthwhile withdrawing this layer and not displaying it to the end users. Indeed, decoding only the base layer and the enhanced layer 1 achieves better MOS than decoding all layers, particularly if the L_2's NALU Loss Rate exceeds 1%. It is clear that, in some cases, withdrawing one SVC layer can maintain higher MOS than the case where this layer is decoded. Accordingly,

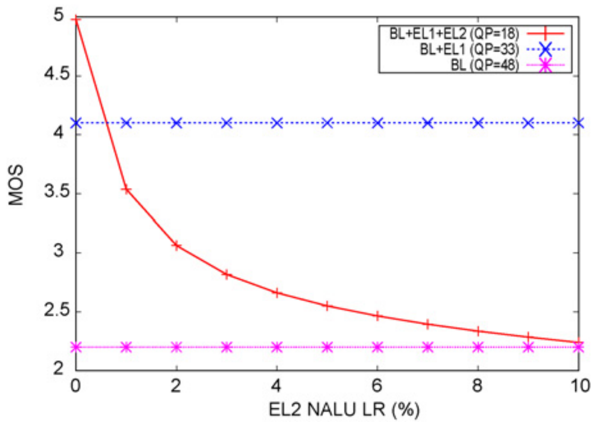

Fig. 11. MOS vs. L_2's NALU loss rate.

in the next section we propose QoE-BASD, a new scheme that dynamically selects the number of SVC layers to be decoded and shown to the end-user aiming at maximizing the user's QoE.

IV. QoE-Based Adaptive SVC Layer Decoding (QoE-BASD) in DVB-T2

A. QoE-Based Selection Algorithm

Based on the results of Fig. 12, we propose using QoE as the main criteria for selecting or withdrawing a SVC layer at the receiver side. The proposed QoE-BASD algorithm assists the decoder to decide which layers to be displayed at the end user. Furthermore, QoE-BASD is following the same principle of associating DVB-T2 and SVC as in [9]. QoE-BASD is also compatible with DVB-T2 unidirectional communication principle, since it is executed at the receiver side without the need for any feedback or signaling messages.

In the following, we assume that n layers compose the broadcasted video stream. We consider representing the initial SVC stream with a Matrix noted Mat. Indeed, the aim of this matrix is to decompose the initial SVC stream into a combination of layers composing n streams. Each line of the matrix corresponds to a possible SVC stream composed of k layers ($n \geq k \geq 1$), whereby each column represents a layer. The element $Mat(i, j)$ is equal to one if the jth layer is present in the ith SVC stream. Otherwise, the element $Mat(i, j)$ is equal to null. Usually, the received SVC stream at the decoder side represents the last line of the matrix. Using QoE-BASD, the aim is to build the matrix Mat and selects the line of the matrix (SVC stream) that maximizes the user's QoE. In contrast, the classical approach systematically decodes all the layers composing the last line of the matrix.

$$Mat_{4,4} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \tag{3}$$

For simplicity, we consider the above matrix $Mat_{4,4}$, which represents the case of an initial SVC stream composed by 4 layers. For instance, the first line shows the case of a SVC stream (noted 1) composed only by the base layer, while the 4th line represents a SVC stream (noted 4) constituted by the base layer and 3 enhanced layers. The QoE-BASD algorithm is illustrated in Algorithm 1. First, for each line of the matrix Mat, QoE-BASD computes MOS by considering the parameters $(f_{IDR}, L_{BL}, L_1, L_2)$ for line i. We denote by $\Delta = (\Delta_1, \Delta_2, \Delta_3, .., \Delta_i)$ the vector containing the i MOS values obtained for each line. After that, the maximum value of MOS (noted Δ_k) of this vector is selected. Accordingly, the kth layers composing the SVC stream are decoded and displayed to the user, while the other layers (layers higher than k) are withdrawn. The execution periodicity of QoE-BASD (the loop) is repeated every t seconds, which corresponds to the duration of a GOP.

Algorithm 1

loop
  for $i = 1$ to n **do**
    $\Delta[i] \leftarrow$ Compute_QoE(Mat_i) // Mat_i is a vector constituted by the line i of the matrix Mat.
  end for
  $\Delta[k] \leftarrow max(\Delta)$
  Decode and display the k layers
end loop

The complexity of the QoE-BASD algorithm depends on the number of layers n composing the SVC video stream. Since most of the existing implementations of SVC use three layers (i.e., one base layer and two enhancement layers), QoE-BASD can be easily implemented in most DVB-T2 products; even in those with low CPU capacities.

B. Analytical Model of QoE-BASD Algorithm

In this section, we develop an analytical model using Markov chains to evaluate the performance of the QoE-BASD mechanism. Specifically, we derive the number of times that QoE-BASD system visits a decoding state that represents the case of decoding only one, two, or three SVC layers. The elaborated model will be also used to derive performance metrics for QoE-BASD in terms of users' QoE.

Let $X(k)$ be a stochastic process $\{X(k), k \geq 1\}$, where $X(k) = (i, l, m)$ if the system decodes i SVC layers, and l and m represent the loss or successful transmissions of l packets of the enhanced layer 2, and the loss or successful transmissions of m packets of the enhanced layer 1, respectively. In fact, depending on the current number of decoded SVC layers, we need to count the number of successive successful packets of the current highest SVC layer (good quality) in order to increase the number of SVC layers to decode. Alternatively, we count the number of successive packet losses (bad quality) to decrease the number of SVC layer to decode. Intuitively, $X(k)$ is a discrete time homogeneous Markov chain model whose associated transition graph is depicted in Fig. 12 for the case of ($0 \geq l \geq 4$), ($0 \geq m \geq 4$) and a SVC stream composed of a base layer and two enhancement layers. The proposed model envisions the following assumptions:

1) The base layer does not experience any loss as it uses PLP0, which is the most robust.

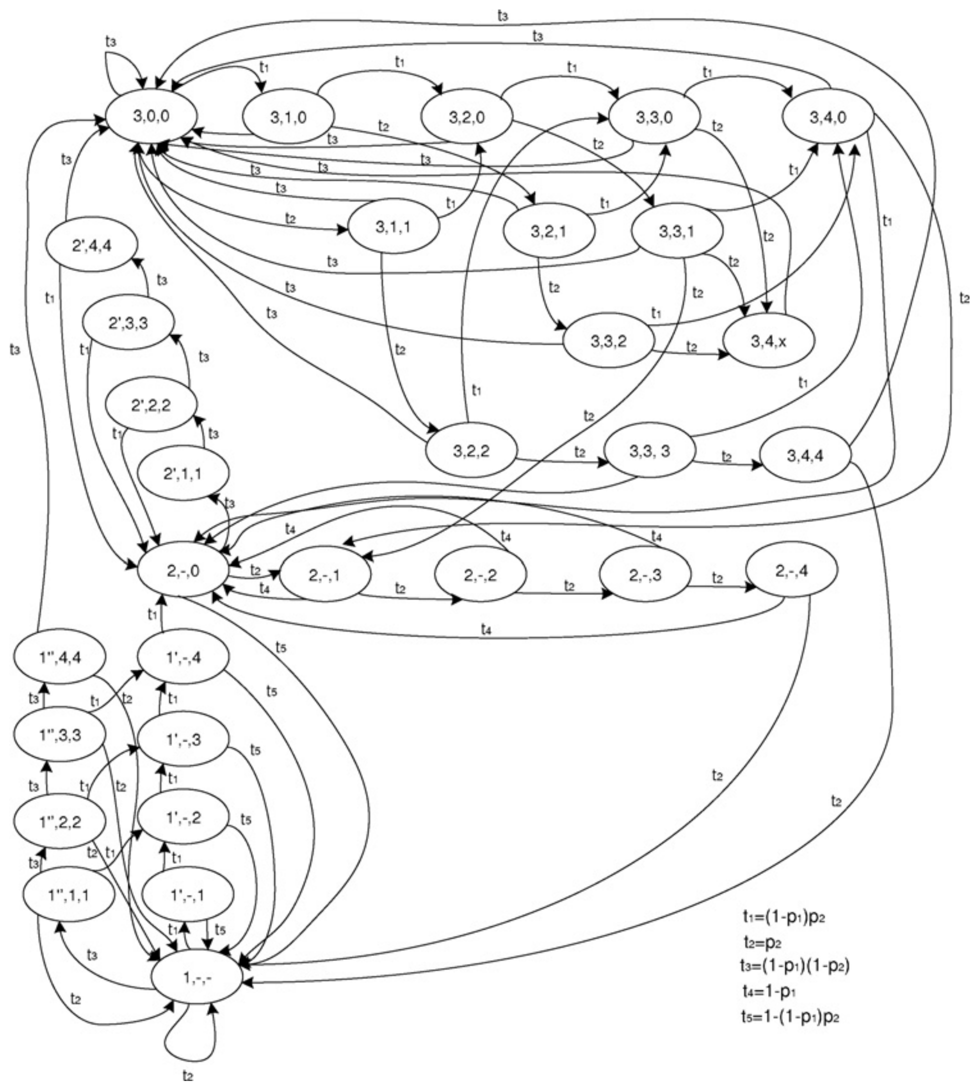

Fig. 12. Markov chain model of the proposed QoE-BASD scheme.

2) After four consecutive packet losses of an enhanced layer i, the system ignores this layer and begins decoding only the lower layers. This assumption is based on the results of Figs. 7, 8 and 9, which indicate that QoE decreases in a non-linear fashion with relation to the packet loss.
3) After the successful reception of five successive packets of an enhanced layer i, the system begins decoding layer i. This, in fact, represents the case where the quality increases as layer i is not experiencing packet losses for a certain duration (i.e., illustrated by five successive packets successfully received), which represent the period t in Algorithm 1.

The states are denoted as follows:

1) (3,l,m) represents the system when decoding three layers, and the enhanced layers 1 and 2 experiencing m and l consecutive packet losses.
2) (2,-,m) represents the system when decoding two layers and the enhanced layer 1 experiencing m consecutive packet losses.
3) (1,-,-) represents the system when decoding only the base layer.
4) (1',-,m') represents the system when decoding only the base layer, and m' consecutive packets of the enhanced layer 1 are successfully received.
5) (1",l',m') represents the system when decoding only the base layer, and l' and m' consecutive packets of the enhanced layers 2 and 1 are successfully received, respectively.
6) (2',l',m') represents the system when decoding two layers, and l' and m' consecutive packets of the enhanced layers 2 and 1 are successfully received, respectively.

Since we are interested in the stationary behavior of the proposed mechanism, we denote by X the stationary version of the Markov chain $\{X(k)\}$. This version exists since the Markov chain is finite and irreducible. The transition rules between the different states are as follows:

1) Transition between state (3,l, 0) to state (3, l+1, 0) with (l < 4) corresponds to the reception of erroneous packets of EL2 and EL1, respectively. Its probability is

$(1-p_1)p_2$, whereby p_2 and p_1 represent the probability of receiving an erroneous packet at EL2 and EL1, respectively. p_2 and p_1 are obtained according to Equation (4), which represents the relation between the probability of a packet loss, the physical Bit Error Rate (BER) and packet size. Intuitively, $p_2 \geq p_1$ ($BER_{PLP2} \geq BER_{PLP1}$ asPLP1 is using more robust physical modulation than PLP2,

2) Transition from state (3,l,m) to state (3,0,0) with ($l < 4$, $m < 4$) represents the case of successful reception of a packet belonging to EL2, so its probability is $(1-p_1)(1-p_2)$. It is worth mentioning that successfully receiving a packet belonging to EL2 implicitly means that also a packet of EL1 is successfully received (i.e., PLP1 is more robust than PLP2). Furthermore, this transition is motivated by the fact that the system resumes counting the number of successive packet losses if they did not reach the threshold for switching to a lower layer.
3) Transitions from state (3,l,m) to state (3,l+1,m+1) with ($l \leq 3$, $m \leq 3$) and from state (2,-,m) to state (2,-,m+1) correspond to the case of receiving an erroneous packet at EL1, so its probability is p_1.
4) Transition from state (3,4,0) to state (2,0,0) corresponds to the case where an erroneous packet at EL2 is received, and the system begins to decode two layers (i.e., instead of three layers). Its probability is $(1-p_1)p_2$.
5) Transition from either states (3,4,4) and (2,-,4) to (1,-,-) corresponds to the case where EL1 experiences five successive packet errors, so its probability is p_1.
6) Transition from state (1',-,m) to state (1',-,m+1) corresponds to the case where only a successful packet of EL1 is received, so its probability is $(1-p_1)p_2$.
7) Transition from state (1",l,m) to state (1",l+1,m+1) corresponds to the case where one successful packet of each enhanced layer is received, so its probability is $(1-p_1)(1-p_2)$.
8) Transition from state (1',l,m) to state (1',-,m+1) corresponds to the case where a packet of EL2 is lost and a packet of EL1 is successfully received, so its probability is $(1-p_1)p_2$. Although the system loses a packet belonging to EL2, the system continues counting the number of successfully received packets of EL1 in order to decode two layers instead of only the base layer.
9) Transition from state (1',-,4) to state (2,-,0) corresponds to the case where the system reaches the threshold (i.e., five successive successfully received packets) from where the system begins decoding two 2 layers instead of only the base layer.
10) Transition from state (2',l,m) to state (2',l+1,m+1) corresponds to the case where one successful packet of each enhanced layer is received, so its probability is $(1-p_1)(1-p_2)$.
11) Transition from states (1',4,4) and (2',4,4) to (3,0,0) corresponds to the case where five packets of EL1 and EL2 are successfully received, so its probability is $(1-p_1)(1-p_2)$. This transition also means that the system will begin decoding three layers.

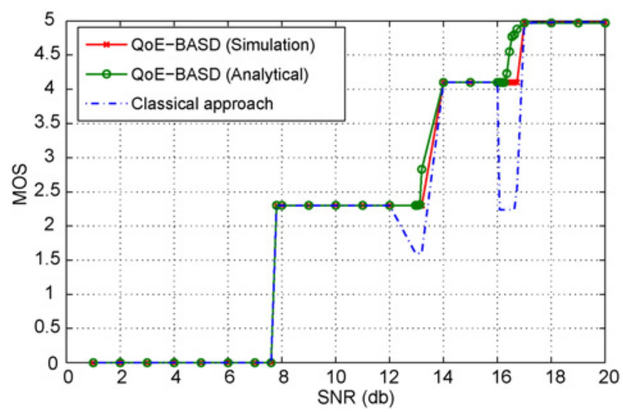

Fig. 13. MOS versus physical SNR (mobile profile).

12) The system stays in state (3,0,0), if the packets of EL2 and EL1 are successfully received. Its probability is $(1-p_1)(1-p_2)$.
13) The system stays in state (1,-,-), if packets of EL2 and EL1 are lost. Its probability is p_1.

$$p = 1 - (1 - BER)^{packet\_size} \tag{4}$$

From the above mentioned transition rules, we build the matrix of transitions P of this Markov chains. We denote by ϕ the stationary distribution of the Markov chains. We thus have

$$\begin{aligned} &\pi = \pi * P \qquad and \\ &\textstyle\sum_{i=1}^{S} \pi_i = 1 \end{aligned} \tag{5}$$

where S denotes the number of states (e.g., in case of the model of Fig. 12, S=31). To compute the average number of times the system visited each state of the chain, we need to resolve the system in (5). By computing the stationary distribution vector, we can estimate the average QoE perceived by a user in case QoE-BASD is used, as follows:

$$\begin{aligned} &E[QoE] = \pi(3,0,0)E[Q(3,0,0] \\ &+\textstyle\sum_{i=1}^{l}\sum_{j=1}^{m}\pi(3,i,j)E[Q(3,1,j)] \\ &+\pi(2,-,0)E[Q(2,-0)] \\ &+\textstyle\sum_{i=1}^{m}(\pi(2,-,i)E[Q(2,-,i)] \\ &+\pi(2',-,i)E[Q(2,-,i)] \\ &+\pi(1,-,-)E[Q(1,-,-)] \\ &+\textstyle\sum_{i=1}^{m}(\pi(1',-,i)E[Q(1',-,i)] \\ &+\pi(1",i,i)E[Q(1",i,i)]) \end{aligned} \tag{6}$$

where $E[Q(s,l,m)]$ denotes the QoE obtained in state (s,l,m); s denoting the number of layers decoded in this state, l denoting the number of erroneously or successfully received packets of EL2, m denoting the number of erroneously or successfully received packets of EL1. For instance, $E[Q(3,0,0)]$ is the maximum MOS value achievable by the system. It is equal to 4.97 over 5. Besides, $\pi(s,l,m)$ illustrates the stationary probability to have the system residing in state (s,l,m).

V. Performance Evaluation

We implemented the DVB-T2 model as in [34] and the proposed solution using the OPNET simulator. Three layers compose each SVC stream, namely one base layer and two

TABLE V
DVB-T2 PARAMETERS.

Scenario Name	PLP	Physical Rate	Physical Modulation	Min. Frequency	Trans. Power	Bandwidth
Full HD	PLP0	7.8 Mbps	16 QAM 1/2	622 Mhz	220 Watt	8 Mhz
	PLP1	34.24 Mbps	256 QAM 3/5	622 Mhz	220 Watt	8 Mhz
	PLP2	47.72 Mbps	256 QAM 5/6	622 Mhz	220 Watt	8 Mhz
Mobile Reception	PLP0	7.8 Mbps	16 QAM 1/2	622 Mhz	220 Watt	8 Mhz
	PLP1	21.1 Mbps	64 QAM 3/5	622 Mhz	220 Watt	8 Mhz
	PLP1	34.24 Mbps	64 QAM 3/5	622 Mhz	220 Watt	8 Mhz

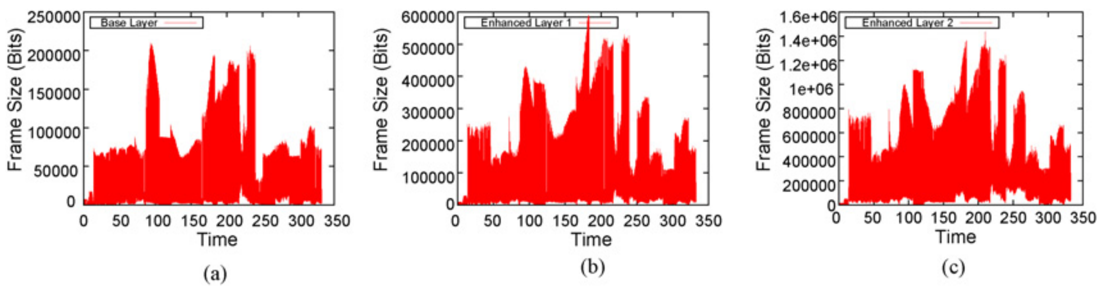

Fig. 14. Sony sequence (HD 1980x1088) characteristics. (a) Base layer. (b) Enhanced layer 1. (c) Enhanced layer 2.

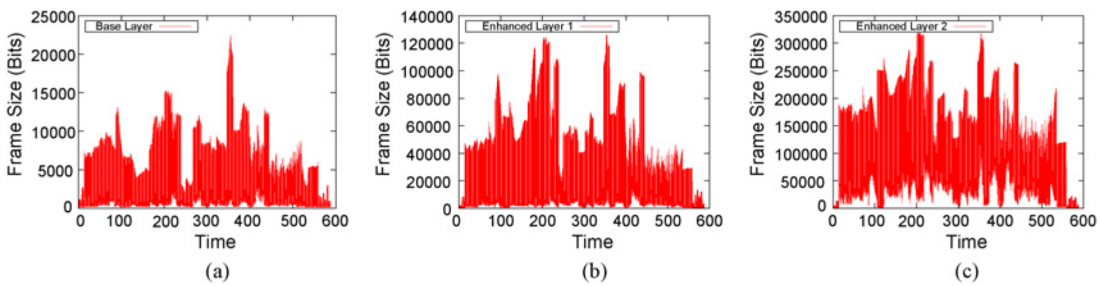

Fig. 15. Sony sequence (CIF 352x288) characteristics. (a) Base layer. (b) Enhanced layer 1. (c) Enhanced layer 2.

enhanced layers. As video streams, we used the Sony video traces available at [35]. As stated in [36], video traces do not contain the actual encoded video (bit) stream; instead, they provide a meta-characterization of the encoded video stream. A video trace provides this meta-characterization by providing the quantities that are required for simulating the transport of the actual video with a communication or networking mechanism. Basic video traces provide the time stamp, encoded size (in bytes), and PSNR quality of each encoded video frame. Video traces are employed in simulation studies of the transport of video over communication networks.

Each SVC layer is broadcasted through different DVB-T2 PLPs. The base layer is sent through the most robust PLP, namely PLP0. The enhanced layers 1 and 2 are sent through PLP1 and PLP2, respectively. Furthermore, we added the possibility to simulate fading power envelop channel by using Ricean or Rayleigh models. These models are widely used to simulate multipath models in wireless communications. It is worth mentioning that the Ricean model is used for Line-Of-Sight (LOS) communications, which depend on the signal strength of the LOS component k, and the Rayleigh model (k=0) is used for NoLOS communications. Table V shows the parameters considered for the simulated DVB-T2 network.

Fig. 13 plots the mean MOS (QoE) for different values of the physical signal noise ratio (SNR). The figure compares the results of QoE-BASD, obtained both by simulation and analytical model, to the results obtained in case of the conventional scheme whereby all layers are decoded. Before SNR reaches 8db, there is no decodable signal. From 7db, PLP0 (Base layer) begins to be decoded without errors. From 8db to 12db, only the base layer is decoded. Between 12db and 14db, packets from PLP1 begin to be decoded, but high Bit Error Rates increase the packet loss. In this situation, QoE-BASD decodes and displays only the base layer, which achieves better MOS than the case of decoding packets of the enhanced layer EL1 with errors. At 14db, BER of PLP1 decreases to zero, which implies that both schemes can decode packets of both the base layer and the enhanced layer 1 without errors. Between16db and 17db, PLP2's signal starts appearing but with high BER. In this interval, QoE-BASD decodes only the base layer and EL1, exhibiting better MOS than the case of decoding the three layers, since decoding EL2 with errors decreases users' QoE. On the other hand, this figure demonstrates that the proposed analytical model of QoE-BASD is accurate in capturing the behavior of this mechanism, and gives practically results similar to those obtained with computer simulations.

To validate further the concept of QoE-BASD, we simulated two more realistic scenarios:

1) Full HD scenario that simulates a HD video broadcast (HD version of Sony sequence) for fixed stations.

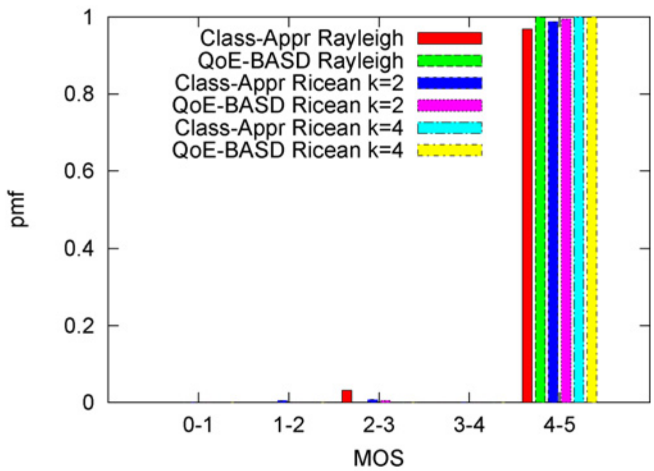

Fig. 16. PMF of MOS for the case of fixed user placed at 20 km away from the DVB-T2 gateway.

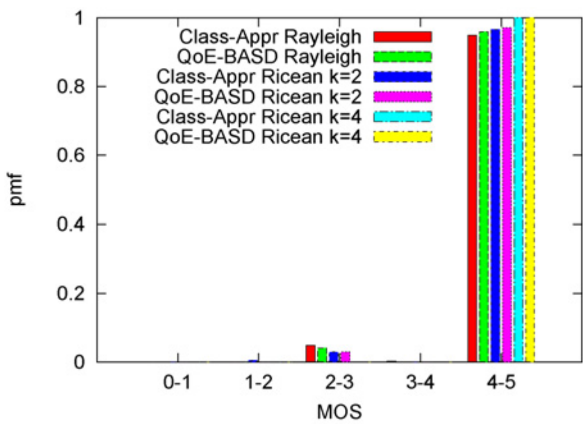

Fig. 17. PMF of MOS for the case of fixed user placed at 100 km away from the DVB-T2 gateway.

2) Mobile Reception scenario that simulates a mobile TV reception (CIF version of Sony sequence).

The characteristics of both videos are shown in Figs.14 and 15, which plot the frame size of each layer composing both sony video versions. The HD version of Sony has a spatial dimension of 1980x1088, whereby the CIF version has a spatial dimension of 352x288.The video sequences were encoded at 30 frames/sec, an IDR frequency of 16, and a GOP size of 4 and 16 in case of the HD version and the CIF version (no P frames were used), respectively. For the HD version, DQP is 10, which means that each layer reduces the overall QP about 10 (increases quality). In case of CIF version, DQP is equal to 15. For both scenarios we compared QoE-BASD to the conventional approach whereby all SVC layers are decoded. The presented results particularly focus on MOS perceived by users, since our aim it to increase users' QoE.

Full HD scenario

In the first scenario (Full HD), we varied the distance between the receiver and the DVB-T2 transmitter to investigate the impact of the distance between the transmitter and the receiver on users' QoE.

Figs. 16 and 17 plot the probability mass function (PMF) of MOS perceived by two fixed users placed at 20 km and

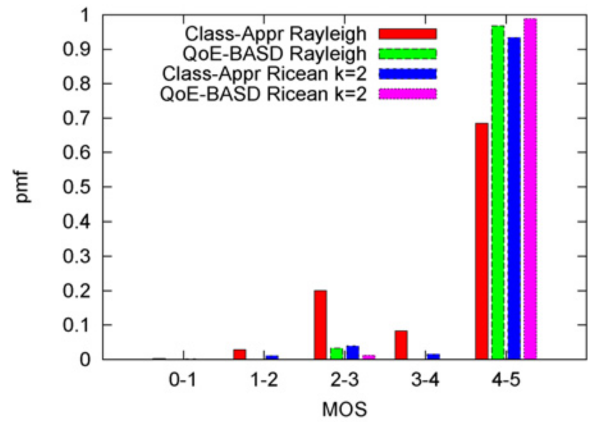

Fig. 18. PMF of MOS in case of a Pedestrian.

100 km away from the DVB-T2 transmitter, respectively. The simulation duration for this scenario is five minutes. MOS is calculated for the three fading channels: (i) Ricean with a k factor equals to 2; (ii) Ricean with a k factor equals to 4; (iii) and Rayleigh. We clearly notice that the Rayleigh channel fades more quickly than the two other Ricean-based channels, and this is attributable to the fact that Rayleigh's envelope does not contain a LOS component. The Ricean model with a high k factor fades slowlier, since the impact of the LOS component is more important on the received signal (k=4). Moreover, MOS is higher when the receiver is closer to the DVB-T2 transmitter.

From these figures, we also notice that QoE-BASD outperforms the conventional approach, mostly in the Rayleigh channel. Indeed, the Rayleigh channel is more sensitive to channel fading that increases packet losses due to high BER. Thanks to the adaptive decoding feature of the proposed scheme, QoE-BASD avoids the degradation of MOS by withdrawing SVC layers that experience packet losses. Meanwhile, the Ricean channel ensures better physical signal quality, as we note that packet losses are low in case of k=2, and almost not existent in case of k=4.

Mobile scenario

For this scenario, we simulate the movement of a mobile user through the following cases:

1) Pedestrian (1m/s): the mobile user is initially placed at 1km away from the transmitter station, and walks a distance of 5 Km further. The antenna height considered in this scenario is 1m.
2) Vehicle: the mobile is initially placed at 1km away from the transmitter, and walks a distance of 20 Km further. The antenna height is set to 2m. In this scenario, we consider two types of car movement: a car in a city driving at a speed of 50Km/h and a car on a highway driving at a speed of 110 Km/h.

The simulation duration for these cases is 15 minutes.

Fig. 18 plots the PMF of MOS obtained in the case of a pedestrian user. It compares the performance of QoE-BASD against that of the conventional approach in the case of two physical channels, namely Rayleigh and Ricean (k=2). It shall be noted that for all mobile scenarios we simulate

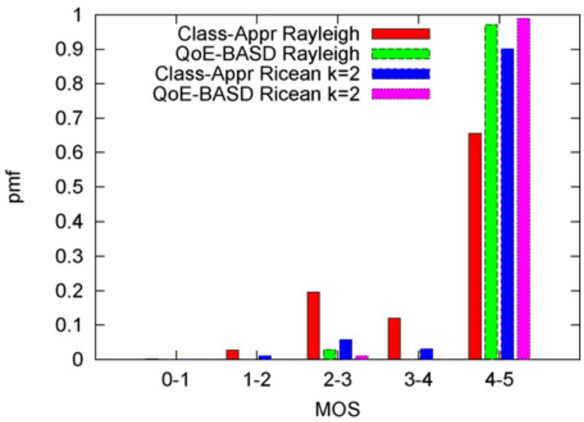

Fig. 19. PMF of MOS in case of a City car.

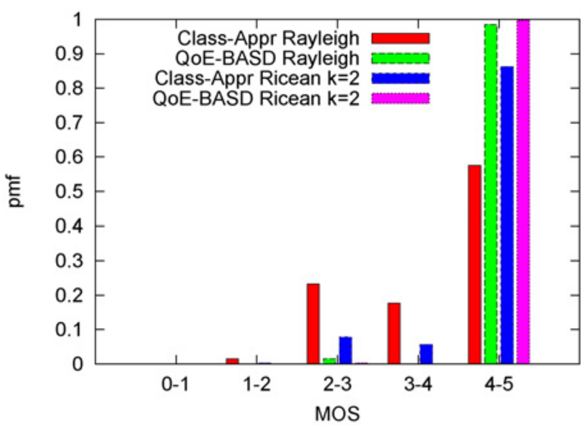

Fig. 20. PMF of MOS in case of a Highway car.

only a Ricean channel with k=2, since the case of k=4 is not realistic for simulating user mobility. From the obtained results, we clearly notice that QoE-BASD outperforms the conventional approach, regardless the underlying channel. We note that QoE-BASD maintains MOS between 4 and 5, with a probability of 0.98 and 0.94 in case of the Rayleigh and Ricean channels, respectively. High difference in MOS is achieved between QoE-BASD and the conventional approach in case of the Rayleigh channel, and that is thanks to the feature of QoE-BASD that reduces SVC layers when the channel's signaling conditions are poor.

Figs. 19 and 20 present PMF of MOS in case of higher mobility scenarios (e.g., city car and highway car). In these scenarios, the antenna's height for mobile receivers is set to 2 meters, which has an impact on the signal quality. Indeed, few db are won in comparison to the pedestrian case (1 meter). From the figures, we clearly observe the same behavior as in case of the pedestrian scenario: QoE-BASD achieves better results in case of both city and highway cars and for both simulated channels. In case of the Ricean channel, we notice that the QoE-BASD scheme and the conventional approach obtain nearly the same MOS. This is due to the fact that the Ricean channel experiences low BER. Thus, both approaches can decode the three layers. However, the difference is more evident in case of the Rayleigh channel, as PLP2 experiences high BER, impacting users' QoE. In fact, QoE-BASD adapts to the periodic degradation of the channel quality by withdrawing packets of EL2, which ultimately maintains good users' QoE. For both cases, QoE-BASD exhibits a MOS value between 4 and 5 with a probability higher than 0.95 (for both cases). However, in the conventional approach this probability decreases to 0.62 and 0.58, in case of a city car and a highway car, respectively.

VI. Conclusion

In this paper, we evaluated the concept of associating DVB-T2 with SVC in order to broadcast high added-value video services to end-user. We first, introduced a reference-less QoE measurement tool for SVC video coding. Based on this tool we conduct measurements, which showed that decoding all SVC layers is not always the best way to enhance QoE, especially when SVC enhanced layers experience packet losses. To increase user QoE, we demonstrated that, in some cases, it is better to withdraw an enhanced layer that experiences packet losses rather than decoding its content. We considered this observation when designing our proposed QoE-BASD mechanism that dynamically selects the SVC layers to be decoded and displayed to end-users in order to increase QoE. We modeled QoE-BASD using a Time Discrete Markov chain model, and validated the analytical results by computer simulations. The obtained results clearly demonstrated the enhancements achieved by QoE-BASD as it improves user QoE.

Finally, it shall be noted that the proposed solution is compatible with the DVB-T2 standard as QoE-BASD could be deployed without complexity in DVB-T2 decoders, and it does not require any feedbacks or return channel to work.

References

[1] B. Ciubotaru and G.-M. Muntean, "SASHA: A quality-oriented handover algorithm for multimedia content delivery to mobile users," *IEEE Trans. Broadcast.*, vol. 55, no. 2, p. 437, Jun. 2009.

[2] Scalable Video Coding, Joint ITU-T Rec. H.264 Ð ISO/IEC 14496-10/Amd. 3 Scalable Video Coding, Nov. 2007.

[3] T. Wiegand, "Scalable video coding for IPTV services," *IEEE Trans. Broadcast.*, vol 55, no. 2, pp. 527–538, Jun. 2009.

[4] K. Kashibuchi, T. Taleb, A. Jamalipour, Y. Nemoto, and N. Kato, "Prioritization-based layered multicast for fixed/mobile networks with fast convergence and inter-session fairness," *J. Commun. Software Syst.*, vol. 2, no. 2, pp. 89–98, Jun. 2006.

[5] *DVB-T2 the HDTV Generation of Terrestrial DTV*, DVB-TM-3997, Mar. 2008.

[6] M.-J. Rahman, W. Yiyan, T. Bin, Y. Kechu, and J.-Y. Chouinard, "A time slicing adaptive OFDM system for mobile multimedia communications," *IEEE Trans. Broadcast.*, vol. 56, no. 2, pp. 226–234, Jun. 2010.

[7] J.-Y. Chouinard, A. Semmar, W. Xianbin, and W. Yiyan, "On the channel and signal cross correlation of downlink and uplink mobile UHF DTV channels with antenna diversity," *IEEE Trans. Broadcast.*, vol. 56, no. 2, pp. 235–248, Jun. 2010.

[8] T. Taleb and K. Hashimoto, "MS2: A novel multi-source mobile-streaming architecture," *IEEE Trans. Broadcast.*, vol. 57, no. 3, pp. 662–673, Sep. 2011.

[9] A. Ksentini and Y. Hadjadj-Aoul, "On associating SVC and DVB-T2 for mobile television broadcast," in *Proc. IEEE Globecom*, Houston, TX, USA, Dec. 2011, pp. 1–5.

[10] L. Kondrad, I. Bouzizi, V. K. M. Vadakital, M. M Hannuksela, and M. Gabbouj, "Cross-layer optimized transmission of H.264/SVC streams over DVB-T2 broadcast system," in *Proc. IEEE Int. Conf. Symp. BMSB*, Bilbao, Spain, May 2009, p. 15.

[11] ITU-T SG12, *Definition of Quality of Experience*, COM12 LS 62 E, TD 109rev2 (PLEN/12), Geneva, Switzerland, Jan. 2007.

[12] ITU-T E.800, *Telephone Network and ISDN Quality of Service, Network Management and Traffic Engineering: Terms and Definitions Related to Quality of Service and Network Performance Including Dependability*, Aug. 1994.

[13] G.-M. Muntean, "Efficient delivery of multimedia streams over broadband networks using QOAS," *IEEE Trans. Broadcast.*, vol. 52, no. 2, pp. 230–235, Jun. 2006.

[14] G.-M. Muntean and N. Cranley, "Resource efficient quality-oriented wireless broadcasting of adaptive multimedia content," *IEEE Trans. Broadcast.*, vol. 53, no. 1, pp. 362–368, Mar. 2007.

[15] S.-B. Lee, A. Smeaton, and G.-M. Muntean, "Quality-oriented multiple-source multimedia delivery over heterogeneous wireless networks," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 216–230, Jun. 2011.

[16] B. Ciubotaru, G.-M. Muntean, and G. Ghinea, "Objective assessment of region of interest-aware adaptive multimedia streaming quality," *IEEE Trans. Broadcast.*, vol. 55, no. 2, pp. 202–212, Jun. 2009.

[17] S. Chikkeur, V. Sundaram, M. Reisslein, and L.-J. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 165–182, Jun. 2011.

[18] M.-H. Pinson and S. Wolf, 'A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 312–322, Sep. 2004.

[19] S. Winkler, A. Sharma, and D. McNally, "Perceptual video quality and blockiness metrics for multimedia streaming applications," in *Proc. Int. Symp. Wireless Personal Multimedia Commun.*, Aalborg, Denmark, Sep. 2001, pp. 547–552.

[20] A. Ichigay, M. Kurozumi, N. Hara, Y. Nishida, and E. Nakasu, "A method of estimating coding PSNR using quantized DCT coefficients," *IEEE Trans. Circuits Syst. Vido Technol.*, vol. 16, no. 1, pp. 251–259, Feb. 2006.

[21] ITU-T G. 1070, *Opinion Model for Video-Telephony Applications*, Apr 2007.

[22] K. Yamaghachi and T. Hayashi, "Parametric packet-layer model for monitoring video quality of IPTV services," in *Proc. IEEE ICC*, Beijing, China, Jun. 2008, pp. 110–114.

[23] F. You, W. Zhang, and J. Xiao, "Packet loss pattern and parametric video quality model for IPTV," in *Proc. IEEE Int. Conf. Comput. Inform. Sci.*, Shanghai, China, Jun. 2009, pp. 824–828.

[24] S. Mohamed and G. Rubino, "A study of real-time packet video quality using random neural networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 12, pp. 1071–1083, Dec. 2002.

[25] W. Cherif, A. Ksentini, D. Negru, and M. Sidibe, "A_PSQA: Efficient real-time video streaming QoE tool in a future media Internet context," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Barcelona, Spain, Jul. 2011, pp. 1–6.

[26] G. Auwera, P.-T. David, and M. Reisslein, "Traffic and quality characterization of single-layer video streams encoded with H.264/MPEG-4 advanced video coding standard and scalable video coding extension," *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 698–718, Sep. 2008.

[27] K. Singh, A. Ksentini, and B. Marienval, "Quality of experience measurement tool for SVC video coding," in *Proc. IEEE ICC*, 2011, pp. 1–5.

[28] S. Wenger, Y. Wang, and T. Schierl, *RTP Payload Format for SVC Video*, IETF Internet Draft, draft-ietf-avt-rtp-svc-20.txt, Dec. 2009.

[29] ETSI TS 102 606 V.1.1.1, *Digital Video Broadcasting (DVB); Generic Stream Encapsulation*.

[30] JSVM [Online]. Available: http://ip.hhi.de/imagecom_G1/savce/downloads/SVC-Reference-Software.htm

[31] OpenSVC [Online]. Available: http://sourceforge.net/projects/opensvcdecoder

[32] ITU-R BT.500-11, *Methodology for the Subjective Assessment of the Quality of Television Pictures*, Jun. 2002.

[33] Video Quality Experts Group, *Validation of Objective Models of Video Quality Assessment*, Phase II, Aug. 2003.

[34] A. Ksentini and Y. Hadjadj-Aoul, "DVB-T2 simulation model in OPNET," in *Proc. ICST SimuTools*, Desenzano, Italy, Mar. 2012, pp. 214–216.

[35] SVC Traces [Online]. Available: http://trace.eas.asu.edu/videotraces2/cgs/index.php

[36] P. Seeling and M. Reisslein, "Video transport evaluation with H.264 video traces," *IEEE Commun. Surveys Tutorials*, vol. 14, no. 4, pp. 1142–1165, Oct.–Dec. 2012.

Adlen Ksentini received the M.Sc. degree in telecommunication and multimedia networking from the University of Versailles, Versailles, France, and the Ph.D. degree in computer science from the University of Cergy-Pontoise in 2005, with a dissertation on QoS provisioning in IEEE 802.11-based networks.

He is currently an Associate Professor at the IRISA Laboratory, University of Rennes 1, Cedex, France. He is a member of the INRIA Rennes Team Dionysos. He is involved in several national and European projects on QoS and QoE support in future Internet networks. He is a co-author of over 40 technical journal and international conference papers. His other interests include future Internet networks, cellular networks, green networks, QoS, QoE, and multimedia transmission.

Dr. Ksentini has been on the technical program committees of major IEEE ComSoc conferences, ICC/Globecom, WCNC, PIMRC.

Tarik Taleb (SM'10) received the B.E. degree in information engineering with distinction, and the M.Sc. and Ph.D. degrees in information sciences from GSIS, Tohoku University, Japan, in 2001, 2003, and 2005, respectively.

He is currently a Senior Researcher and 3GPP Standards Expert at NEC Europe, Ltd., Heidelberg, Germany. Prior to his current position and until March 2009, he was an Assistant Professor at the Graduate School of Information Sciences, Tohoku University, in a lab fully funded by KDDI, the second largest network operator in Japan. From October 2005 to March 2006, he was a Research Fellow with the Intelligent Cosmos Research Institute, Sendai, Japan. His current research interests include the field of architectural enhancements to mobile core networks (particularly 3GPP's), mobile cloud networking, mobile multimedia streaming, congestion control protocols, handoff and mobility management, intervehicular communications, and social media networking.

Dr. Taleb is a Board Member of the IEEE Communications Society Standardization Program Development Board. As an attempt to bridge the gap between academia and industry, he has founded and has been the General Chair of the IEEE Workshop on Telecommunications Standards. He has also been directly engaged in the development and standardization of the evolved packet system as a member of 3GPP's System Architecture Working Group. From research to standards, a successful event that got awarded the Best Workshop Award by the IEEE Communication Society. He has been on the Editorial Board of the IEEE WIRELESS COMMUNICATIONS MAGAZINE, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, and a number of Wiley journals. He is serving as the Vice Chair of the Wireless Communications Technical Committee, the largest in IEEE ComSoC. He also served as the Secretary and then as the Vice Chair of the Satellite and Space Communications Technical Committee of IEEE ComSoc from 2006 to 2010. He has been on the technical program committees of different IEEE conferences, including Globecom, ICC, and WCNC, and has chaired some of their symposia. He was the recipient of the 2009 IEEE ComSoc Asia-Pacific Best Young Researcher Award in 2009, the 2008 TELECOM System Technology Award from the Telecommunications Advancement Foundation in 2008, the 2007 Funai Foundation Science Promotion Award in 2007, the 2006 IEEE Computer Society Japan Chapter Young Author Award in 2006, the Niwa Yasujirou Memorial Award in 2005, and the Young Researcher's Encouragement Award from the Japan Chapter of the IEEE Vehicular Technology Society in 2003. Some of his research work has been also awarded Best Paper Awards at prestigious conferences.