Overview of the SVC4QoE Project

Marcus Barkowsy[5], Médéric Blestel[3], Mathieu Carnec[1], Adlen Ksentini[4], Patrick Le Callet[5], Gérard Madec[8], Raoul Monnier[2], Jean-François Nezan[3], Romuald Pepion[5], Yohann Pitrey[5], Jean-François Travers[6], Mickaël Raulet[3], and Alain Untersee[7]

[1] AccepTV, [2] Grass Valley France, [3] IETR-INSA Rennes, [4] INRIA, [5] IRCCYN, [6] TDF, [7] Teamcast, [8] Telecom Bretagne

alain.untersee@teamcast.com, yohann.pitrey@univ-nantes.fr, mblestel@insa-rennes.fr

Abstract. The aim of this paper is to give an overview of the SVC4QoE project which purpose is to use Scalable Video Coding (SVC) to optimize video transmission in terms of Quality of Experience over DVB-T2 channels. The originality of this project is to consider the influence of the whole chain of processing and delivery performed on the video data in terms of user-perceived quality. The encoding process, as well as transmission, decoding and display are included in the optimization process. Particularly, the multi-layer structure of SVC is to be exploited to circumvent alterations of the video related to transmission on error-prone networks. Combining SVC with QoE thus provides a way to reduce network operating costs, while increasing the quality from the user's point of view. Several innovations are also to be mentioned, such as the use of a real-time open-source SVC decoder, evaluation of visual quality through subjective quality assessment tests, transmission and synchronization of SVC layers on the receiver's side, and development and integration of an end-to-end DVB-T2 chain which implements the multi-PLP (Physical Layer Pipes) functionality in an SVC context.

Keywords: Scalable Video Coding, DVB-T2, Quality of Experience.

1 Introduction

The SVC4QoE project aims at associating Scalable Video Coding (SVC) and Quality of Experience (QoE) evaluation techniques in order to maximize the user experience when receiving audiovisual content over broadcast networks. The purpose of the project is to explore and demonstrate how to optimize the broadcast network infrastructure in various receiving conditions by lowering the necessary bandwidth, while providing the best quality of the signal to a wide range of portable and hand-held receivers. To this end, new strategies have to be studied to address two conflicting problems: lowering the acquisition and operational cost of the network infrastructure, while improving the end-user experience.

The originality of the SVC4QoE project is not just to use the QoE as a diagnostic tool to evaluate the visual quality *a posteriori*, but as a way to define optimal

J. Rodriguez, R. Tafazolli, C. Verikoukis (Eds.): MOBIMEDIA 2010, LNICST 77, pp. 560–570, 2012.

encoding configurations, transmission schemes, decoding strategies and layer selection techniques.

In October 2007, the SVC standard was finalized in ISO/IEC 14496-10 Amd 3 [1], [4], as an amendment of the ITU-T Rec. H.264 | ISO/IEC 14996-10 Advanced Video Coding (AVC) standard. A video sequence encoded with SVC is a single hierarchical bitstream structured in several layers with specific spatial and temporal features. A base layer is first encoded, providing a basic quality, compatible with the non scalable AVC standard. Enhancement layers are added in the same access unit of the base layer to make increasing levels of quality available, at increased computational costs. Three types of scalability are specified by the standard: spatial, temporal and fidelity scalabilities. Using these three types of scalability together with inter-layer predictive coding, the SVC standard provides a flexible way to adapt a video sequence to the target needs and requirements, while minimizing the impact on the network in terms of bit-rate increase and deployment costs.

As a part of the standardization effort, the Joint Scalable Video Model (JSVM) reference software was developed, providing not only a reference SVC encoder and decoder, but also a set of useful tools to manipulate and analyze scalable video streams [2]. As this reference implementation is not intended to perform real-time encoding and decoding, various contributions have been proposed during the standardization process. In the SVC4QoE project, an open-source real-time solution is used to decode SVC streams. This decoder is one of the strengths of the project, as its performance both in terms of decoding time and of implemented SVC features have been constantly praised by the community.

The recently started DVB-T2 broadcasting standard represents a significant breakthrough when compared to the previous transmission standards. The permitted transfer rates and signal reception ranges make it a strong candidate for future digital terrestrial transmissions. As one of the goals of the SVC4QoE project is to study the performance of video transmission in a mobile environment with varying reception conditions, DVB-T2 was selected for the wide range of possible targets this standard is designed to address.

Quality of Experience (QoE) is related to the visual quality of a video sequence from the end-user point of view. It is different from the well-known Quality of Service for it takes into account the actual displayed video sequence, instead of a set of network-based parameters. The evaluation of QoE is built on subjective experiments, during which human viewers are asked to give their opinion about the feeling of quality they experience. Perceptual models can further be extracted from the experimental data, in order to replace the viewers' opinions with objective quality metrics.

As already mentioned, Scalable Video Coding provides three types of stream adaptation, leading to a possibly very large set of configurations. The spatial and temporal characteristics, as well as the parameters used to encode each layer and the differences between layers have an influence on both the quality of each layer and the quality of the whole stream. Therefore, the number of parameters to take into account in order to evaluate the visual quality of a scalable stream is quite high and represents a first challenge to address. Moreover, the evaluation of QoE in a scalable context raises new problems such as comparing various spatial and temporal resolutions in terms of visual quality.

The layered structure of SVC provides independent versions of the video sequence with intermediate spatial and temporal resolutions. In a transmission error-prone environment such as DVB-T2 networks, these intermediate versions can be exploited to minimize the impact of packet loss on the visual quality of the decoded video. To this end, a QoE-based layer selector must be inserted at the decoder's side, in order to display the optimal layer of the video in terms of visual quality (see Figure 1).The combination of the QoE module with the SVC decoder represents one of the strongest points of the SVC4QoE project.

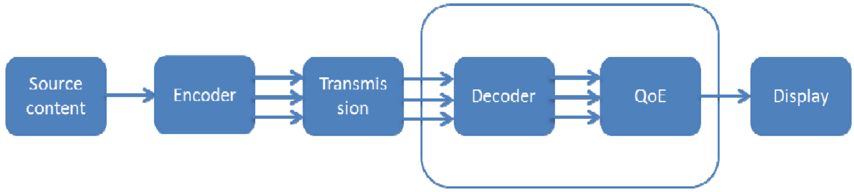

Fig. 1. Considered workflow for distribution of video content in the SVC4QoE project

2 Scalable Video Coding and Open SVC Decoder

The IETR lab of INSA Rennes has developed a flexible SVC decoder named Open SVC Decoder [3] supporting all tools to deal with spatial, temporal and fidelity scalabilities. It is based on a fully compliant H.264 baseline decoder, implementing most tools from the Main profile. The Open SVC Decoder has been developed in a modular way [5] which provides a convenient design framework, easing both conception and multi-partner collaborations. During the Scalim@ges project [6], this decoder was used to promote the SVC standard in production and distribution contexts. Both its time performance and compliance with the latest versions of the standard were remarked in the final review of the project.

As an illustration, Table 1 presents the time-improvement made possible by the Open SVC Decoder when compared to the JSVM on several conformance sequences. The Open SVC Decoder is up to 50 times faster than the JSVM reference decoder version 9.16, making it quite an interesting candidate for real-time decoding. Moreover, the Open SVC Decoder has been ported and tested on different platforms such as x86 platforms, Personal Digital Assistants (PDA), PlayStation 3 and Digital Signal Processors (DSP) using a validated algorithm-architecture matching methodology.

Its performance in terms of decoding time and compliance, as well as its portability make the Open SVC Decoder particularly suited to the requirements of the SVC4QoE project. Additionally, the Open SVC Decoder offers to decode a specific layer in the scalable structure. This feature, which is not implemented in the JSVM reference software, is particularly useful in a broadcast environment as the layer selection can be done during the decoding process. When transmission errors cause some part of an enhancement layer to be missing, the decoder can switch to the base layer immediately, compensating thus transmission errors to optimize the visual quality of the displayed video sequence.

Table 1. Comparison of the JSVM reference decoder 9.16 and the Open SVC Decoder in terms of decoding times

Sequence	Decoding time (s)		Speed up
	JSVM	Open SVC Decoder	
SVCBST-1	31.2	0.87	35 times faster
SVCBST-2	23.3	0.87	26 times faster
SVCBST-14	137	2.69	50 times faster
SVCBST-15	50	2.11	23 times faster

3 Quality of Experience and Scalable Video Coding

A growing interest has been devoted to the notion of Quality of Experience (QoE) in the last years, as attested by the increasing audience to the Video Quality Experts Group (VQEG) [11] meetings. Quality of Experience is defined as the visual quality perceived by the user of a service. The interest for QoE comes from the observation that simply evaluating the quality of transmission in terms of network-based indicators is not sufficient to get a precise idea about the visual quality of a received video sequence. The broadly used Quality of Service (QoS) exploits a set of network parameters to evaluate the quality [13]. QoS parameters are relatively easy to retrieve as they are often measured within the network itself. Unfortunately, it does not include any consideration about the Human Visual System (HVS) [14] and its complex behavior towards distortions in the displayed sequence.

3.1 QoE Evaluation

Various techniques exist to evaluate the quality of a video sequence.

To measure the Quality of Experience, two approaches exist: subjective quality measurement and objective quality measurement. In the first case, human observers are asked to evaluate the quality of video sequences displayed in a controlled environment. This approach ensures that the evaluated visual quality is closely related to the viewers' feeling, as long as the test conditions respect some restrictions.

The second approach to evaluate perceptual quality consists in using objective metrics [12]. Usually, such metrics use the original and/or the distorted video, and/or parameters extracted from the video, and provide an estimation of its quality. There are four types of objective metrics, namely Full Reference (FR), Reduced Reference (RR), No Reference (NR), and hybrid metrics. Quality metrics can be quite simplistic such as the Peak-Signal-to-Noise Ratio (PSNR), or they can be more complex and try to simulate the behavior of the HVS. Such objective models are often bound to evaluate the impact of a very specific kind of distortion, as they require a complex learning process. Therefore, a model designed to evaluate the quality of videos distorted induced by a H.264 encoding might fail to predict the impact of MPEG-2 or JPEG2000 distortions.

So far, no objective metric dedicated to SVC has been validated yet. Simplistic metrics like PSNR can be used but since they require the use of reference videos, they can only be applied during encoding and can't enable QoE monitoring on the end-user's side.

Therefore the need for a new metric is quite obvious. The choice has been made to develop a hybrid metric which will jointly use parameters from the bitstream and features extracted from the decoded frames. In our context, full reference metric cannot be used since the reference video is not accessible at the end-user side. Among possible metrics, hybrid metrics are the ones which enable to get the most precise quality scores. Several parameters will be considered, such as bit rate, quantization parameter (QP) of each macroblock, motion vector of each macroblock and type of frame (I, B, P). Additionally, the QoE metric will use a decoding-map generated by the decoder, to identify macroblocks not decoded due to packet loss (or packet corruption, packet jitter, etc.).

3.2 New Challenges for QoE Evaluation in a Scalable Context

The current recommendations for subjective evaluation do not include any tool to evaluate several spatial and temporal resolutions during a single subjective test. To this end, new protocols for the evaluation of this type of configuration must be designed. As the process of designing and validating new methodologies is both long and expensive, other approaches using existing metrics should also be investigated.

Currently, the most appropriate approach is to use spatial and temporal up-scaling methods on the tested video sequences, so that the resolutions of all the displayed video sequences are identical. The existing methodologies can then be used to evaluate the subjective quality of the videos. However, the up-scaling step introduces a new type of artifacts. As a result, the up-scaling method as well as the differences in spatial and temporal resolution and quality between layers have to be taken into consideration in the QoE model.

The main interest of SVC regarding the enhancement of the QoE is its ability to switch to a lower resolution layer when the distortions in the highest layer are considered too disturbing. However, the HVS is particularly responsive to spatial and temporal discontinuities. The impact of the number and duration of each switch from one layer to another must therefore be evaluated and included as well in the QoE model.

Because of the predictive structure on which the whole H.264 and SVC coding standard rely, data-loss in a video sequence might lead to error propagation. Whereas the impact of error propagation has been investigated for H.264 single layer coding, it has not been studied for SVC. The scalable features such as inter-layer prediction in SVC introduce a new dimension in error propagation, which has to be characterized. Determining the impact of the enumerated factors in a scalable context is essential for the evaluation of QoE.

On the decoder side, several factors can affect the visual quality. First, as decoding an SVC stream requires a lot of computations, the time-efficiency of the decoder might not be sufficient to process the data in real time. Frame skipping and freezing effects might appear which have been shown to have a dramatic effect on the visual quality. Evaluating the influence of the parameters enumerated in this section is a critical part of the SVC4QoE project. Subjective tests will be performed to extract objective models capable of exploiting the structure of SVC and enhance the end-user experience.

3.3 Scalable Video Coding for QoE Enhancement

In the case of single layer coding, transmission errors lead to a total loss of some parts of the data. Some error concealment techniques exist based on in-painting from the previous frames or from the frame itself. Using scalable video coding, a lower resolution of the video data might still be available. As the lower quality layers represent less data, one can assume that better protection can be afforded, in order to make sure there is no error in this part of the stream. Combined with up-scaling techniques, this lower-resolution version can be exploited to perform error concealment in the high-resolution video.

As an illustration, Figure 2 presents the results of a subjective pre-test performed by the IRCCyN-IVC group. This test compares four scenarios in which the displayed video is of size 640x480 pixels at 30 frames per second (i.e. VGA@30). The first scenario is the non-coded video, used as a reference quality. In the second scenario, the enhancement layer of an SVC stream containing two layers is used. The base layer is of size 320x240 pixels at 30 frames per second (i.e. QVGA@30) encoded at 200 kbits/s, while the enhancement layer is VGA@30, encoded at 600 kbits/s using inter-layer prediction from the base layer. The third scenario depicts the same video encoded with AVC at 600 kbits/s, on which data loss was simulated by removing 2 seconds of video out of 10. A typical error concealment method (see Figure 3) is used to compensate this loss. It makes use of data from the previous frame to reduce the impact of the lost data. The last scenario represents the base layer from the SVC stream, displayed in VGA@30 format after spatial up-scaling. The Mean Opinion Scores (MOS) from a set of 19 naive viewers are displayed for the four presented scenarios, acquired using the common Absolute Category Rating (ACR) methodology [10]. An example of typical distortions produced by the error concealment in AVC and the up-scaling in SVC is presented in Figure 3.

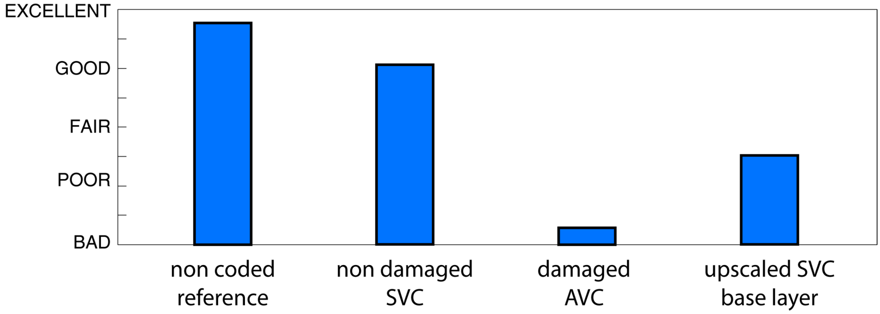

Fig. 2. Mean Opinion Scores of the presented AVC and SVC scenarios

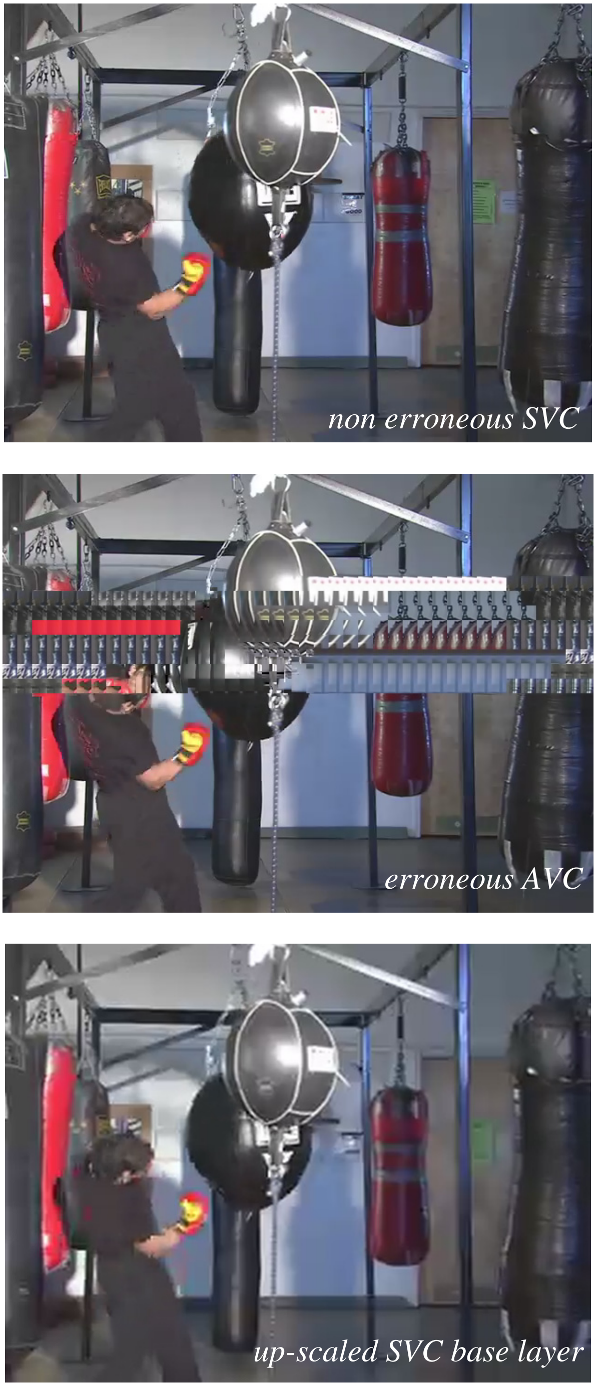

Fig. 3. Overview of the distortions introduced by coding artifacts, conventional error concealment techniques and spatial up-scaling

It can be observed that the score of the up-scaled version of the base layer from the SVC stream is higher than the score of the distorted AVC stream. This illustrates the interest of displaying a video up-scaled from a lower-resolution instead of using a single layer stream combined with conventional error concealment. As a result, switching to the up-scaled version provides a way to enhance the QoE using SVC when compared to a scenario where only AVC is used.

The selection of the layer to display is handled by the QoE model at the decoder's side. In order to optimize the QoE, the model needs to reflect as closely as possible the behavior of the HVS and include the parameters mentioned in section 3.2, such as the spatial and temporal difference between the layers and the number of temporal switches.

Several policies can be considered to minimize the spatial and temporal distortions produced by the transmission errors. A first strategy is to replace the whole image with the image from the base layer as soon as the measured quality is below a given threshold. In this case, the layer selector simply acts as a switch at frame level. A second strategy is to replace only the areas affected by the transmission errors, using up-scaled parts from the base layer. The impact of the spatial and temporal discontinuities produced by both strategies will have to be evaluated in order to determine which policy is the best in terms of QoE.

4 Transmission of SVC Signals over DVB-T2 Networks

Distribution of video content on heterogeneous networks is one of the most significant breakthroughs made possible by the development of new transmission media and techniques. DVB-T2 represents the upcoming candidate for future over-the-air transmissions, providing a flexible framework to address a wide range of services and users (ranging from mobile services for small screens to high quality services for large wide flat screens).

The extended flexibility provided by the DVB-T2 standard justifies its choice as a starting point for the SVC4QoE project. DVB-T2 allows multiple service components to be transmitted over a single RF channel through the "multi-PLP" technique (Multiple Physical Layer Pipes). These components can use different coding rates and modulation schemes. As a result, they can provide different tradeoffs in terms of capacity and robustness. It becomes then possible to have components transmitted with a very high robustness and a limited bandwidth (a few hundreds of kbps per service), while other components can be transmitted with a higher bandwidth but less robust (1 to 2 Mbps for instance per service).

DVB-T2 is therefore well suited to transmit audiovisual signals encoded with SVC. Indeed, the base layer is encoded with a reduced bit rate so that it can be received by as many terminals as possible, even in very challenging receiving situations, while the enhancement layers are encoded using higher bit rates and can be received when transmission conditions are more comfortable or better quality receiver and antenna. As an example, Table 2 presents a scenario for transmitting the SVC base layer and two enhancement layers using 3 different PLPs. The flexibility of DVB-T2 allows many variants of this scenario to be implemented and evaluated in order to match the bit rate requirements of the 3 layers.

Table 2. SVC stream transmission scenario within three DVB-T2 profiles

Layer	PLP Modulation scheme	Transmission capacity of a complete 8MHz channel	Required C/N
Base	QPSK - 2/3	10 Mbits / s	3.1 dB
Enhancement 1	16 QAM - 2/3	24 Mbits / s	9.9 dB
Enhancement 2	256 QAM - 2/3	42 Mbits / s	17.8 dB

Before transmission, the video and audio data need to be split into network-adapted packets in order to maintain the synchronization among the SVC layers at the receiver's side. In the SVC4QoE project, the splitting and transmission step is assumed by a Multimedia Aware Network Element (MANE) [7]. At the encoder's side, each SVC enhancement layer is differentiated and is encapsulated into Real-time Transport Protocol (RTP) packets by the MANE using the RTP Payload Format for SVC. Audio and base layer share the same RTP session [8]. Note that each Network Abstraction Layer Unit (NALU) is encapsulated into a single RTP packet and the SVC layers share one RTP session. By using a single RTP session, it is easy to synchronize the SVC layers thanks to the RTP timestamp.

DVB-T2 provides two main encapsulation protocols: MPEG2 Transport Streams (MPEG-2 TS), Generic Stream Encapsulation (GSE). The MPEG-2 TS packetization has been the classical encapsulation scheme for DVB services, while the Generic Stream Encapsulation was designed to provide appropriate encapsulation for IP traffic with the following pros and cons:

- The standard ways to carry IP datagrams over MPEG2-TS are the Multiprotocol Encapsulation (MPE) and the Unidirectional Lightweight Encapsulation (ULE). However, MPE/ULE IP encapsulation uses additional overhead, thus reducing the channel bandwidth in a significant way.
- An alternative for encapsulating IP packets is GSE, which reduces the overhead by a factor of 2 to 3 compared to MPEG-TS. However, GSE is currently not so much deployed in DVB products.

Accordingly, the SVC4QoE bitstreams are encapsulated into MPEG2-TS (MPEG2 Transport Streams defined in ITU-T H.222.0 | ISO/IEC 13818-1[9]) for DVB-T2 transmission. TS packets are then associated to each DVB-T2 PLP according to the conveyed SVC layers. Figure 4 shows the global architecture adopted by the SVC4QoE project. At the receiver's side, the DVB-T2 decoder sends the DVB-T2 PLPs to the MANE merger, which reconstructs the RTP stream according to the RTP sequence numbers. Thus, it ensures that the SVC decoder receives the NALUs in the decoding order.

The choice of carrying NALU over RTP/MPEG-TS, instead of using directly MPEG-TS, is motivated by:

- Using the same encoding flow for both broadcast and broadband (for ADSL or 3G users);
- Conforming to others broadcast technologies such as DVB-H;
- Simplifying the synchronization of the different SVC layers received at the terminal side. In fact, as each layer is associated with a different DVB-T2 PLP.

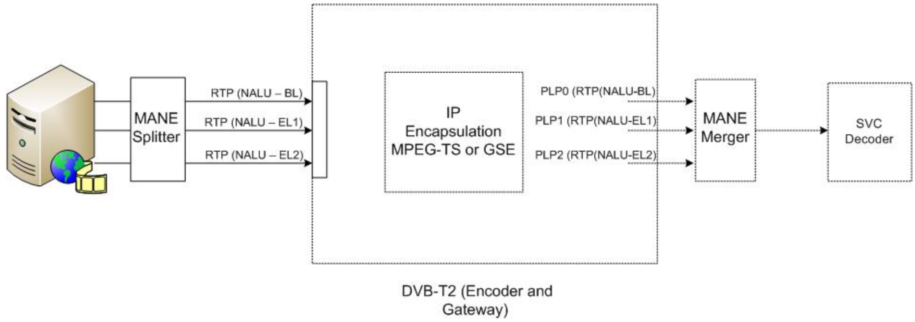

Fig. 4. Architecture of transmission of SVC streams over DVB-T2

5 Demonstration Platform

To show the service improvement provided by the association of SVC and QoE for a DVB-T2 transmission, a demonstrating platform is being developed by the partners of the project. Two main different use cases are defined to show the usefulness of the developed architecture. The first use case involves television broadcasting towards a pedestrian, moving in various environments. The adaptability of the system to limited changes in reception conditions is to be tested through this use case. The second use case exhibits television broadcasting towards an end-user in a moving car with an embedded receiver and a car-roof antenna. The reception conditions might change in a faster way than with the first use case. It will provide a way to test the system in extreme environments.

The main objective of the SVC4QoE project is to ensure that the end-user is able to receive an acceptable version of the video in all these conditions. If the transmission conditions are good enough, a high-quality layer is decoded. If not, the QoE module comes into action and decides which layers to be displayed.

6 Conclusion

This paper presents an overview of the SVC4QoE project which aims at combining Scalable Video Coding with Quality of Experience considerations and DVB-T2 transmission. The flexibility of SVC combined with DVB-T2 transmission provides new possibilities to broadcast layered streams with different levels of protection. The quality of these streams will be evaluated in terms of QoE, in order to provide the end-user with the best visual experience. This paper presents the technological aspects, as well as the new issues to address in order to develop the full chain between the source and the displayed video sequence. Some early ideas about the strategies to combine the innovative points were also mentioned in order to show the current investigations carried out by the partners of the project.

At the moment, the project steering committee is involved in a process aiming at refining the theoretical concepts and practical test protocols to be developed in order to orient and validate further work. About one year after this publication, the project will reach its end, and results from both computer simulation and laboratory trials (possibly field trials as well) will be available for presentation.

Acknowledgements. The SVC4QoE project started in October 2009 for a duration of 24 months. It was labelled by the "Images & Réseaux" Cluster. Nine partners are involved in the project:

- Academics : IETR-INSA Rennes, INRIA, IRCCYN, Telecom Bretagne;
- Industrials: AccepTV, Grass Valley France, Neotilus, TDF, TeamCast.

The authors would like to thank the DGCIS (French Ministry for Industry), Région "Bretagne" and Région "Pays de la Loire" (regional councils) for partially funding this project.

References

[1] Joint ITU-T Rec. H.264 | ISO/IEC 14496-10/Amd.3 Scalable Video Coding (November 2007)
[2] Joint Scalable Video Model JSVM-19.9, Available in CVS repository at Rheinisch-Westfälische Technische Hochschule (RWTH) Aachen, http://ip.hhi.de/imagecom_G1/savce/downloads/SVC-Reference-Software.html
[3] Blestel, M.: Open SVC Decoder, http://opensvcdecoder.sourceforge.net/
[4] Schwarz, H., Marpe, D., Wiegand, T.: Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. IEEE Trans. on Circuits and Systems for Video Tech. 17(9), 1103–1120 (2007)
[5] Pelcat, M., Blestel, M., Raulet, M.: From AVC decoder to SVC: Minor impact on a data flow graph description. In: PCS 2007 (June 2007)
[6] Scalim@ges project, http://www.images-et-reseaux.com/en/les-projets/fiche-projets-finances.php?id=125
[7] RFC 3984: RTP Payload Format for H.264 Video
[8] RTP Payload Format for SVC Video: draft-ietf-avt-rtp-svc-21
[9] ITU-T H.222.0 | ISO/IEC 13818-1:2007/Amd 3:2009
[10] Subjective video quality assessment methods for multimedia applications, ITU-T P.910 Recommendation (1996)
[11] Video Quality Experts Group (VQEG), http://www.vqeg.org
[12] Final report from the video quality experts group on the validation of objective models of video quality assessment, Tech. Rep., VQEG (2003)
[13] Wang, Z.: Internet Quality of Service. Elsevier publications (2000)
[14] Voulos, T., Pappas, N.: Perceptual Criteria for Image Quality Evaluation. In: Bovik, A. (ed.) Handbook of Image & Video Processing (2000)