Perceptual Quality Metric as a Performance Tool for ATM Adaptation of MPEG-2 based Multimedia Applications

Olivier Verscheure and Xavier Garcia Adanez
Telecommunications Services Group, TCOM Laboratory
Swiss Federal Institute of Technology
CH-1015 Lausanne, Switzerland
E-Mail : {verscheure,garcia}@tcom.epfl.ch
URL : http://tcomwww.epfl.ch/{~verscheu, ~garcia}

Abstract

In this paper we study the perceptual impact of data loss on MPEG-2 video coded streams transmitted over an ATM network. This impact is measured using a perceptual quality metric based on a spatio-temporal model of the human visual system. Video streams have been transmitted on top of both new network and ATM adaptation layers which provide a robust transmission by applying per-cell sequence numbering combined with a selective Forward Error Correction (FEC) mechanism. We compare their performance against a transmission over AAL5. Results show that the proposed AAL behaves better in terms of both network performance and perceived quality of the MPEG-2 decoded sequence.

1 Introduction

Nowadays, quality assessment of image sequences becomes a very important issue and receives a large amount of attention from the multimedia communications community. For the past several years, a tremendous number of audiovisual services has been emerging (e.g. Video on Demand (VOD), Interactive Distance Learning (IDL), home shopping, etc). Today, ATM technology, efficient compression techniques, and other developments in telecommunications make possible the offer of such services. These services should be optimized to be proposed at very attractive prices. One of the major issues in order to achieve this optimization is the utilization of a perceptual video quality metric. Such quality metrics that have the advantage of being coherent with quality as perceived by human observers are only recently beginning to emerge.

This work focuses on the study of the perceptual impact of data loss on MPEG-2 coded streams transmitted over an ATM network. The paper is divided as follows : a perceptual quality metric is presented in Sec. 2. The sensitivity to data loss of MPEG-2 video applications is studied in Sec. 3. Section 4 gives some insights on the transmission of multimedia streams over ATM networks. Section 5 describes the simulation setup and shows some results. Finally, Sec. 6 concludes the paper.

2 A Perceptual Quality Metric

Several studies have shown that a correct estimation of subjective quality has to incorporate some modeling of the Human Visual System [1]. A spatio-temporal model of human vision has been developed for the assessment of video coding quality [2, 3]. The model is based on the following properties of human vision:

- The responses of the neurons in the primary visual cortex are band limited. The human visual system has a collection of mechanisms or detectors (termed "channels") that mediate perception. A channel is characterized by a localization in spatial frequency, spatial orientation and temporal frequency. The responses of the channels are simulated by a three-dimensional filter bank.
- In a first approximation, the channels can be considered to be independent. Perception can thus be predicted channel by channel without interaction.
- Human sensitivity to contrast is a function of both frequency and orientation. The *contrast sensitivity function* (CSF) quantizes this phenomenon by specifying the detection threshold for a stimulus as a function of frequency.
- Visual masking accounts for inter-stimuli interferences. The presence of a background stimulus modifies the perception of a foreground stimulus. Masking corresponds to a modification of the detection threshold of the foreground according to the local contrast of the background (see Sec. 3.2).

The vision model described in [2] has been used to build a computational quality metric for moving pictures [3] which proved to behave consistently with human judgements. Basically, the metric, termed Moving Pictures Quality Metric (MPQM), first decomposes an original sequence and a distorted version of it into perceptual components by a Gabor filter bank. Indeed, the profile of the channels is very close to Gabor functions. A channel-based distortion measure is then computed accounting for contrast sensitivity and masking. Finally, the data is pooled over the channels to compute the quality rating which is then scaled from 1 to 5 as described in Tbl. 1 [4] (see Fig. 1).

Rating	Impairment	Quality
5	Imperceptible	Excellent
4	Perceptible, not annoying	Good
3	Slightly annoying	Fair
2	Annoying	Poor
1	Very annoying	Bad

Table 1: Quality scale that is often used for subjective testing in the engineering community

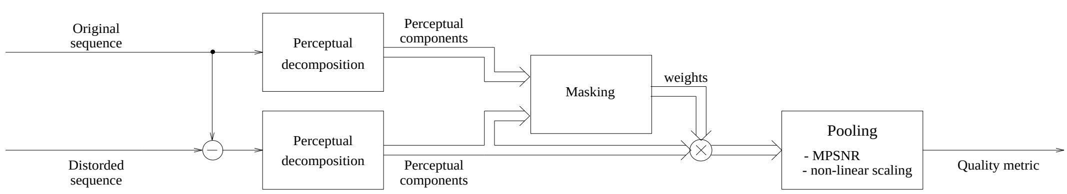

Figure 1: *Moving Pictures Quality Metric (MPQM) block diagram.*

Recently, an improved metric, termed Normalized Video Fidelity Metric (NVFM), is based on a finer modeling of vision and has been introduced in [5]. This new metric adds a modeling of the saturation characteristic of the cortical cells' responses and a modeling of inter-channel masking. It is an extension of a still-picture model developed by Teo&Heeger [6].

Figure 2 presents the MPQM quality assessment of MPEG-2 video for the Mobile & Calendar sequence as a function of the bit rate. This sequence has been encoded, with a software simulator of the Test Model v5 (TM-5) of MPEG-2 [7], as interlaced video with a constant Group of Pictures (GOP) of 12 frames and 2 B-pictures between every reference frame. An important result that can be extracted from the graph is that the perceptual quality saturates at high bit rates. Increasing the bit rate may thus result, at some point, in a waste of bandwidth since the end user does not perceive an improvement in quality anymore.

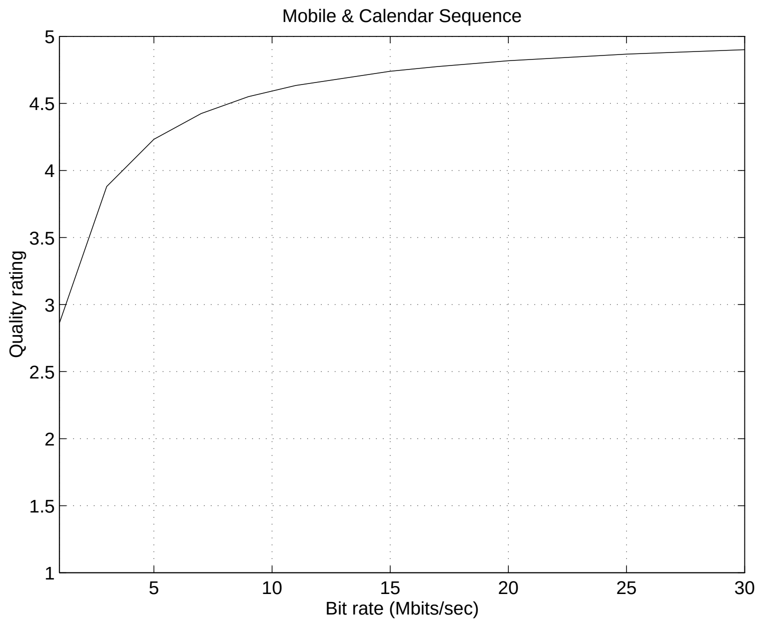

Figure 2: *MPQM quality assessment for the Mobile & Calendar sequence as a function of the bit rate.*

Additional compressed video sequences show results with the same behaviour in [3].

3 MPEG-2 Video Applications Sensitivity to Data Loss

3.1 Network Requirements

In the last years, multimedia applications have developed at a very fast pace. The principal characteristic of multimedia data is that it is of a continuous nature. As such, it has very stringent constraints in terms of delay and delay jitter. Multimedia applications are *delay sensitive*. Data arriving beyond a certain point in time is considered as lost or meaningless by the application. This has also been referred to as "Timely information" [8]. Conversely, traditional data transfer applications are considered as *loss sensitive*. They generally rely on robust transport protocols that guarantee no loss to the application. This has been achieved by using retransmission schemes to correct data loss. In return, low delays are not guaranteed. The deployment of broadband networks such as ATM have reduced to some extent the delay problem and are able to guarantee certain delay bounds when needed. The utilization of compression techniques such as MPEG-2, considered as the standard for video applications, considerably reduces the requirements of video applications in terms of bandwidth. However, compression techniques basically reduce the redundancy in the data to be transmitted thus making multimedia applications also sensitive to data loss. So basically, the transport of real-time multimedia data has to be reliable and timely. ATM networks fulfill both conditions but in some cases may fail to guarantee loss ratios.

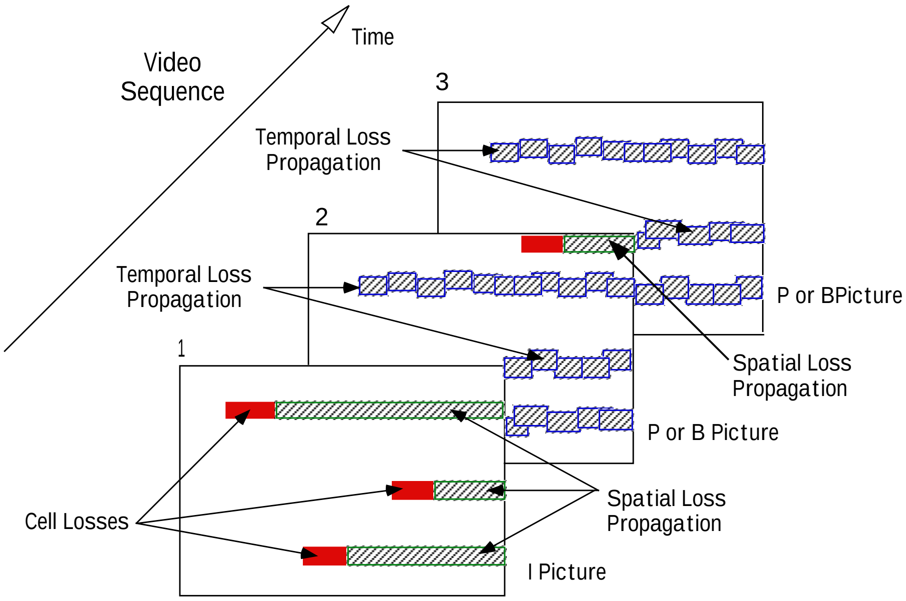

Figure 3: *Data Loss Propagation.*

3.2 Perceptual Impact of Cell Losses

In an MPEG-2 video stream, data loss reduces quality depending strongly on the type of the lost information. Losses in syntactic data, such as headers and system information, affect the quality differently than losses of semantic data such as pure video information. Furthermore, the quality reduction depends also on the location of the lost semantic data due to the predictive structure of an MPEG-2 video coded stream.

Let us consider Fig. 3 showing how network losses map onto visual information losses in different types of MPEG-2 pictures. Indeed, data loss spreads within a single picture up to the next resynchronization point (e.g. slice headers) mainly due to the use of variable length coding, run length coding and differential coding. This is referred to as spatial propagation and may damage any type of picture. When loss occurs in a reference picture (intra-coded or predictive frame), it will remain until the next intra-coded picture is received. This causes the errors to propagate across several non intra-coded pictures until the end of the current GOP. This is known as temporal propagation and is due to inter-frame predictions.

The impact that the loss of syntactic data may have is in general more important and more difficult to recover than the loss of semantic information. This data loss may induce frame loss in the decoded sequence. Indeed, when a frame header (a few syntactic bytes before each frame in the bitstream) is lost, the entire corresponding frame is skipped because the decoder is not able to detect the beginning of the frame. If the skipped frame corresponds to a predictive picture (I or P), it may strongly reduce the perceptual quality due to the predictive structure of the MPEG-2 video stream.

The problem is actually that when a header is lost, in general, the whole information it carries is skipped. Some headers are thus more crucial than others. For instance, sequence headers, predictive (I or P) picture headers, PES headers, slice headers in intra-coded pictures can be considered as essential in comparison to slice headers in B pictures.

Error concealment algorithms have already shown that it is possible to reduce the impact of data loss on the visual information [9, 10, 11]. These error concealment algorithms include, for example, spatial interpolation, temporal interpolation and early resynchronization techniques. Early resynchronization decoding techniques limit the spatial propagation of errors by decoding some semantic information that is normally discarded from the damaged MPEG-2 video streams. In other words, it helps the decoder to resynchronize quickly (at least before the following header). This method is based on the identification of allowed codewords as proposed in [10] and, unfortunately, works only with intra-coded frames. These techniques are thus, by far, not perfect. Data loss may still involve very annoying perturbation in the decoded video, especially when frames are lost.

4 Transmission over ATM

ATM Adaptation Layers (AAL) have been designed to provide specific sets of functions adapted to different identified classes of service. Among these AALs, AAL5 is able to cope with any ATM transfer capability and therefore can handle any kind of application. The main reason for this is its simplicity and small overhead. AAL5 receives SDUs from the upper layer, appends an 8 byte trailer and sends the AAL5-PDU to the ATM layer (see Fig 4). The information contained in the trailer consists of a length indicator, a CRC-32 check and padding information for boundary alignment. The reason for this lack of sophisticated functions is that AAL5 was basically designed for *loss sensitive* applications that make use of reliable transport protocols that handle error correction with retransmission mechanisms based on some kind of feedback scheme. Consequently, because upper layers are packet based, when cell losses are detected by the AAL, the full AAL5-SDU is discarded.

As already stated in Sec. 3.1, multimedia applications are delay and loss sensitive. Therefore, they cannot rely on retransmission mechanisms for error correction. The current specifications to transmit MPEG-2 based applications do not make use of any transport protocol albeit they use AAL5. This means that in case of cell loss, no action could be taken to correct the error or even to inform the receiver that data loss occurred. In addition, the packet discard mechanism of AAL5 amplifies this problem.

In order to meet the requirements of multimedia applications, some mechanisms need to be included in the AAL. We proposed in [12] an AAL mechanism that improves the cell loss detection mechanism, giving better results in terms of data loss seen by the application. Basically, we increased the resolution of the cell loss detection to the cell level. To achieve this, we introduced a cell sequence numbering. To make this possible, we use 47-byte cell payloads

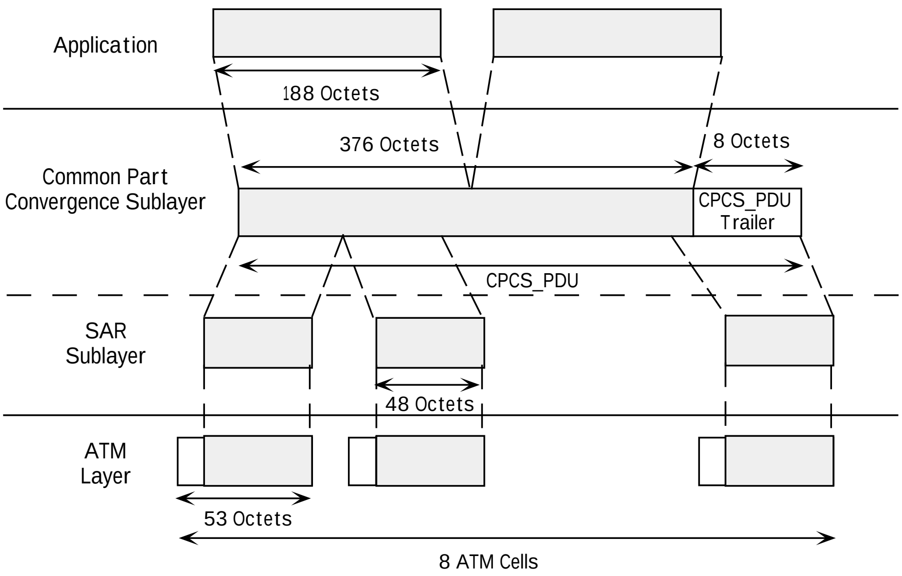

Figure 4: *AAL5 Segmentation Mechanism.*

which frees an octet per ATM cell to insert the sequence number. This packetization process, however, does not provide any kind of PDU delineation, as AAL5 does with its trailer. Our approach assumes that the PDU size is negotiated at connection setup and that it will remain constant for the duration of the connection. Therefore, the receiver does not need any extra information to delineate the packets. This scheme has proven to reduce the data loss seen by the receiver because it increases the resolution of the data loss detection algorithm avoiding the packet discard. The receiver uses the sequence numbers to detect the number and position of the lost cell in a packet. This information can thus be passed to the upper layers that can take necessary action to conceal data loss. In our proposal, we use a dummy cell insertion mechanism, when cell losses are detected (see Fig 5).

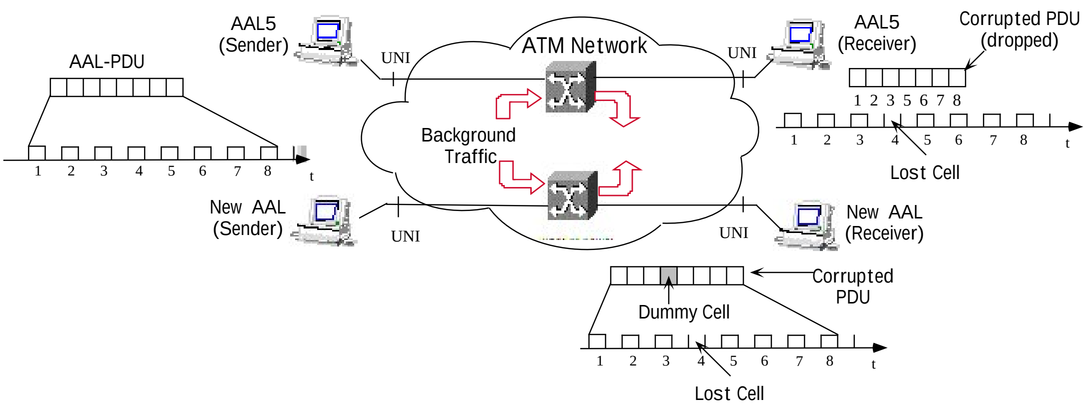

Figure 5: *Dummy Cell Insertion Mechanism.*

If we consider the transport of MPEG-2 TS packets, the proposed mechanism has several advantages. First, it always gives an integer number of cells, since the length of the TS packets is 188 bytes which gives exactly 4×47 byte payloads. Second, if dummy cell insertion is applied, it can be used as an error message since sequences of 47 zero-octets are not allowed by the MPEG-2 standard [13]. This mechanism can thus be exploited by the decoder as an error indicator.

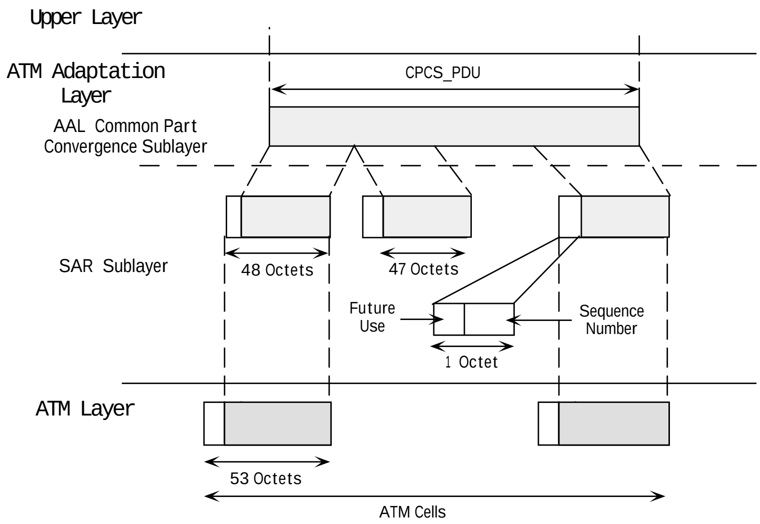

Figure 6: *Proposed ATM Adaptation Layer for Multimedia.*

To further improve the reliability of the AAL, a selective FEC mechanism has been proposed in [14]. It is based on burst erasure codes (RSE) [15] which takes into account the specifics of ATM. It relies on the fact that erasures are limited to fixed boundaries (cells) to correct cell losses only. Since this mechanism does not rely on interleaving, it introduces low delay. The advantage of such a selective method is twofold. First, it reduces the overhead, and second, it adapts to highly structured information such as compressed video. Other advantages of such a mechanism are that it is well suited for point-to-multipoint configurations that eventually should be widely used by such multimedia applications and also that it can be used to protect audio separately from video since losses in audio are much more noticeable and annoying for the user than video losses. Besides, timing information can be selectively protected to reduce the probability of losing synchronization.

However, to achieve this selective FEC mechanism, a knowledge of the data to be transmitted is necessary. Since AALs are generic, another layer has to be able to detect which information has to be protected. This is the task of the proposed Network Adaptation Layer (see Fig. 7). It is an application specific layer since it has to identify information elements in the bitstream. In our example, we have used MPEG-2 based applications so we have developed for our experiments a network adaptation layer specific to this standard. The network adaptation main functionalities are: delivery of fixed-size PDUs (in the case of MPEG-2 it is straightforward) which includes packet segmentation/reassembly and alignment to boundaries, and detection of loss sensitive data combined with the generation of FEC_request messages. The second functionality in the case of MPEG-2 applications consists of detecting a set of headers and generating FEC_request messages accordingly. Based on these messages passed with the PDU, the AAL will or will not generate the FEC data.

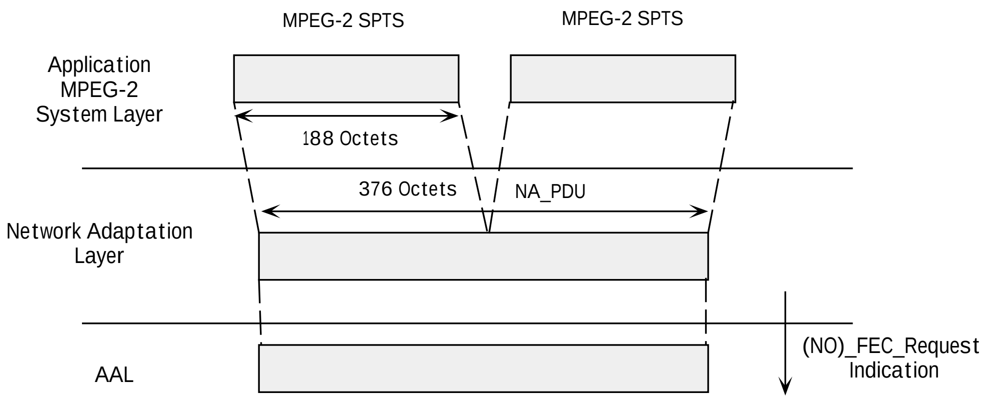

Figure 7: *Proposed Network Adaptation Layer for MPEG-2 Based Applications.*

5 Simulation Experiments

This section first presents the simulation setup including system and video coding, network, system and video decoding, and quality assessment setups. We present next some of the main quality assessments obtained on MPEG-2 video streams transmitted over ATM network using both AAL-5 and our proposed AAL.

5.1 Simulation Setup

5.1.1 Video Coding Setup

The traffic under test (TUT) consists of a ski sequence of 1000 frames (720×576) encoded at 4 Mbits/s, with a software simulator of the Test Model v5 (TM-5) of MPEG-2 [7], as interlaced video, with a structure of 12 images per GOP and 2 B pictures between every reference picture, and a single slice per line (i.e. 45 macroblocs per slice).

Before being transmitted, the MPEG-2 bitstream is encapsulated into 18800-bytes length Packetized Elementary Stream (PES) packets and divided into fixed length Transport Stream (TS) packets by the MPEG-2 system encoder.

5.1.2 Network Setup

To test the efficiency of the perceptual metric as a performance tool, we have carried out a set of simulation experiments. The simulation framework shown in Fig. 8 is composed of four multimedia workstations and two ATM switches. Both switching modules are implemented as multiplexers with limited buffer size. The multimedia workstations are connected as two point-to-point communications. One of the connections uses AAL5 to transmit the MPEG-2 bit stream while the other uses the proposed AAL. To generate cell losses in the multimedia streams, we load the multiplexing stages with background traffic provided by several On-Off sources. This type of source model is widely used to simulate a multiplex of traffic such as the one that could be found at the entrance of an ATM switch. Moreover, two state Markov source models encompass the peak cell rate parameter which is currently the most important traffic contract parameter [16]. To guarantee the same CLRs to both cell streams, the background traffic is replicated and sent simultaneously to both multiplexing stages.

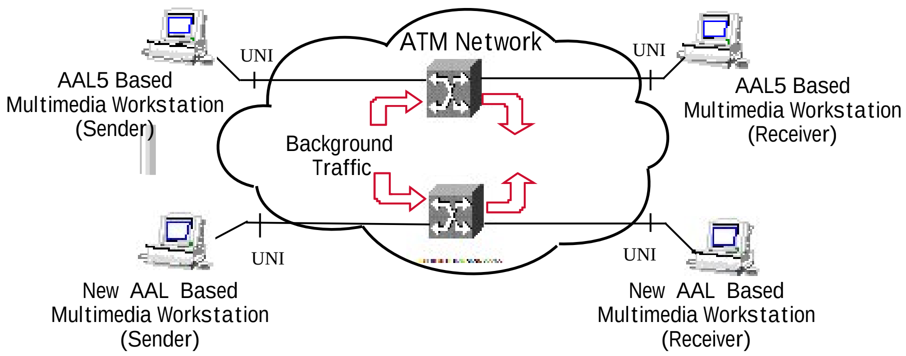

Figure 8: *Simulation Setup.*

The TS packets that form the Traffic Under Test (TUT) are delivered at a constant packet rate to the AAL. The resulting traffic is sent to the switching stage at a constant cell rate where it is multiplexed with the background traffic. Since the switch buffer is limited in size, some of the TUT or background cells may be lost. The TUT is then routed to the receiver end system where the data is reassembled prior to decoding. The background traffic is assumed not to interfere further with the TUT and thus is directly routed to a traffic sink after leaving the switching stage.

The transmission of video over AAL5 is based on the approved ATM Forum Video on Demand specification [17]. This scheme packetizes two TS packets into a single AAL5-SDU. The AAL5 adds its 8-byte trailer and the resulting AAL5-PDU is segmented into 8 ATM cells without any padding (see Fig. 4).

The transmission of video over the proposed AAL is based on the segmentation described in Sec. 4. Two TS packets are passed from the Network Adaptation layer to the AAL. The AAL-PDU is then segmented into 47-byte payloads giving exactly 8 ATM cells. When FEC is applied, we add a single redundancy cell obtained by XORing the 8 data cells. A single cell loss per PDU can be recovered.

5.1.3 Video Decoding and Quality Assessment Setup

When the MPEG-2 decoder detects data loss in semantic information, it tries to resynchronize as quick as possible by applying early resynchronization algorithms presented in [9]. For not resynchronized areas, it uses classical spatial and temporal concealment techniques [11].

When a frame has been lost because of the lack of a frame header, the previous decoded frame is then considered instead. In general, when a system or video header is missing, the decoder tries to recover as much video data as it can.

As stated before in Sec. 3.2, error concealment techniques reduce the impact of cell loss on the perception of decoded video sequences. This implies that, for a given confidence interval, the mean value of the perceptual quality can be estimated on a lower number of frames. The quality estimation using the vision model has been performed on 100 frames (720×576) out of the 1000 ski pictures. The mean quality value is then calculated.

5.2 Simulation Results

To test the efficiency of the perceptual metric as a performance tool, we have carried out a set of simulation experiments for different background loads varying between 79% and 86%. These background loads generate cell losses measured at the receiver which, for AAL5, includes all the data lost due to the packet discard mechanism (see Fig. 9).

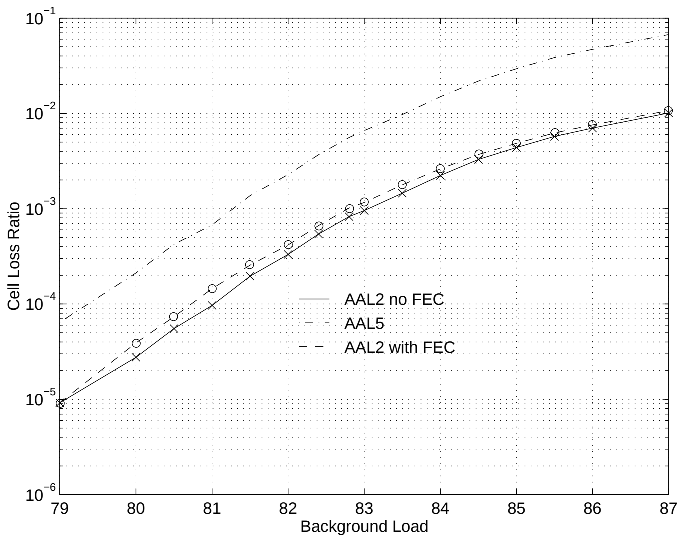

Figure 9: *Cell Loss Ratios seen by receivers.*

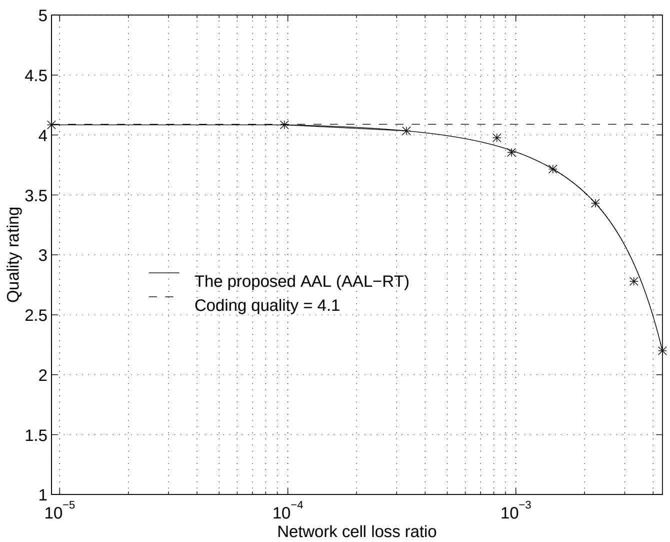

Figure 10: *Quality rating versus network cell loss ratio for the proposed AAL.*

Figure 10 shows the impact of cell losses on the user perception of the decoded sequence using the proposed AAL. While increasing the CLR, the perceptual quality remains first constant at a maximum value corresponding to the quality of the MPEG-2 coded sequence compared to the original not coded one. This is mainly due to the use of error concealment techniques which are efficient for low cell losses. Beyond a certain CLR, the perceptual quality drops smoothly as only a few number of frames are lost, keeping efficient the error masking mechanisms.

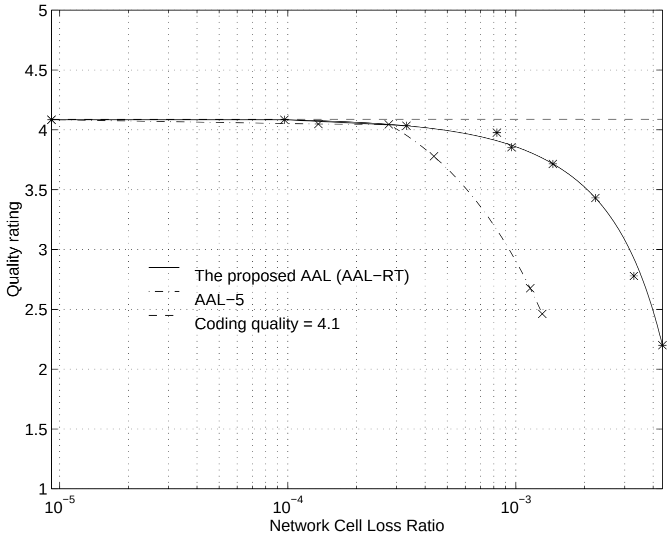

Figure 11: *Quality rating versus network cell loss ratio for the proposed AAL and AAL-5.*

On Fig. 11, one can notice that AAL5 leads to a slightly different behaviour. Indeed, due to the packet discard mechanism of the AAL5 (see Sec. 4), the quality drops faster for lower CLRs since an increasing number of pictures are lost. Moreover, as the efficiency of error concealment algorithms decreases with the increasing amount of lost video data in a row, AAL5 appears then to be less interesting than our proposed AAL.

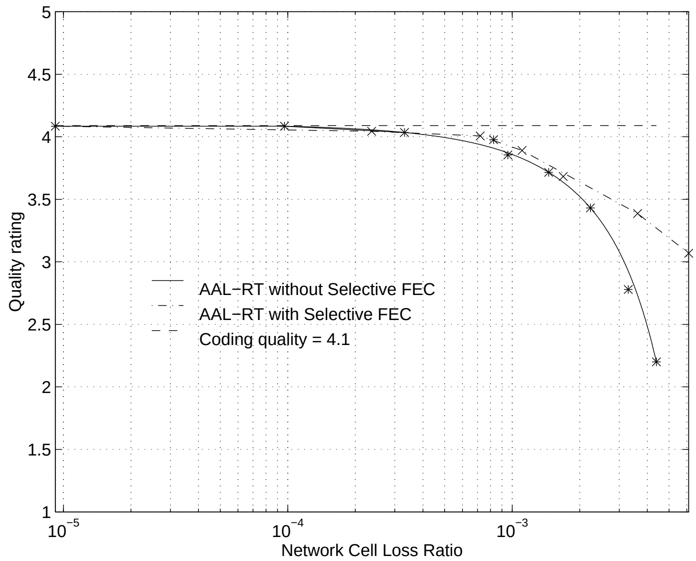

Figure 12: *Quality rating versus network cell loss ratio for the proposed AAL with and without using selective FEC mechanisms.*

Figure 12 presents the gain in perceptual quality we obtained by protecting with FEC cells crucial syntactic data, namely sequence, picture and PES headers. The overhead due to this mechanism is very small ($4.75e^{-3}$). It is obvious that, beyond the critical CLR, the quality drops slower than the one obtained without any protection. Indeed, thanks to the selective FEC mechanism, a lower amount of video data and frames are lost.

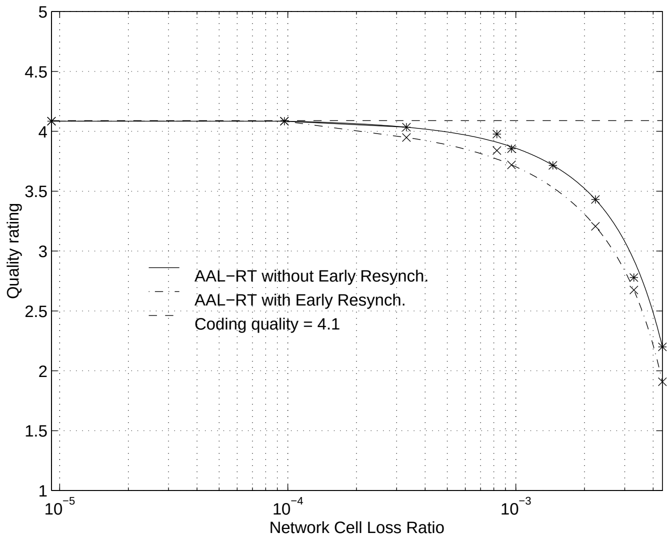

Figure 13: *Quality rating versus network cell loss ratio for the proposed AAL with and without using early resynchronization algorithms.*

As stated before, early resynchronization techniques work only with intra-coded frames. In order to show the importance of such coded frames, we show in Fig. 13 the difference in perceptual quality when early resynchronization algorithms are used or not with our proposed AAL. It is clear that, even if there are only 84 I-frames out of 1000, the difference in quality is significant.

6 Conclusion

This paper studied the perceptual impact of data loss on MPEG-2 video coded streams transmitted over an ATM network on top of AAL5 and new versions of both network and ATM adaptation layers. We presented results showing that the proposed AAL behaved better in terms of both network performance and perceived quality of the MPEG-2 decoded sequence.

Future studies will be done with improved data protection schemes to increase the robustness of audiovisual applications to data loss [18].

References

[1] Serge Comes, *Les traitements perceptifs d'images numerisées*, PhD thesis, Université Catholique de Louvain, 1995.

[2] Christian J. van den Branden Lambrecht, "A Working Spatio-Temporal Model of the Human Visual System for Image Restoration and Quality Assessment Applications", *in*

Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, pp. 2293–2296, Atlanta, GA, May 7-10 1996, available on `http://ltswww.epfl.ch/~vdb`.

[3] Christian J. van den Branden Lambrecht and Olivier Verscheure, "Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System", *in Proceedings of the SPIE*, vol. 2668, pp. 450–461, San Jose, CA, January 28 - February 2 1996, available on `http://tcomwww.epfl.ch/~verscheu`.

[4] M. Ardito and M. Barbero and M. Stroppiana and M. Visca, "Compression and Quality", *in Proceedings of the International Workshop on HDTV 94*, Brisbane, Australia, October 1994.

[5] Pär Lindh and Christian J. van den Branden Lambrecht, "Efficient Spatio-Temporal Decomposition for Perceptual Processing of Video Sequences", *in Proceedings of the International Conference on Image Processing*, Lausanne, Switzerland, September 16-19 1996, accepted for publication, available on `http://ltswww.epfl.ch/~vdb`.

[6] David J. Heeger and Patrick C. Teo, "A Model of Perceptual Image Fidelity", *in Proceedings of the International Conference on Image Processing*, pp. 343–345, Washington, DC, October 23-26 1995.

[7] Chadd Fogg, "mpeg2encode/mpeg2decode version 1.1. available via anonymous ftp at ftp.netcom.com", *in MPEG Software Simulation Group*, June 1994.

[8] H. Leopold, G. Blair, A. Campbell, G. Coulson, P. Dark, F. Garcia, D. Hutchinson, N. Singer, and N. Williams, "Distributed Multimedia Communication System Requirements", OSI 95 Consortium, May 1992.

[9] Carlos Lopez Fernandez, Andrea Basso and Jean-Pierre Hubaux, "Error Concealment and Early Resynchronization Techniques for MPEG-2 Video Streams Damaged by Transmission over ATM Networks", *in SPIE*, San Jose, USA, February 1996.

[10] S. Lee, J. Youn and S. Jang, "Transmission error detection, resynchronization, and error concealment for mpeg-2 video decoder", *in SPIE*, 1993.

[11] S. Aign and K. Fazel, "Temporal and spatial error concealment techniques for hierarchical mpeg-2 video coded", *in ICC'95*, 1995.

[12] X. Garcia Adanez, "Study of AAL5 and a New AAL Segmentation Mechanism for MPEG-2 Video Over ATM", *in 7th Workshop on Packet Video*, Brisbane, Australia, March 1996.

[13] ISO/IEC JTC 1, *Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - Part 1: Systems Specification*, ISO/IEC JTC 1, 1994.

[14] Xavier Garcia Adanez, Olivier Verscheure and Jean-Pierre Hubaux, "New Network and ATM Adaptation Layers for Real-Time Multimedia Applications: A Performance Study Based on Psychophysics", *in COST-237 workshop*, Barcelona, Spain, November 1996, submitted paper.

[15] A.J. McAuley, "Reliable broadband communications using a burst erasure correcting code", *in ACM SIGCOMM'90*, Philadelphia, USA, 1990.

[16] Rastislav Slosiar, *Performance Analysis Methods of ATM-Based Broadband Access Networks Using Stochastic Traffic Models*, PhD thesis, TCOM Laboratory, Swiss Federal Institute of Technology, Lausanne, 1995.

[17] The ATM Forum, "Video on Demand Specification 1.0", Jan. 1996.

[18] Olivier Verscheure, "Local Video Activity Metric and its Impact on Video Services Optimization", *in COST-237 workshop*, Barcelona, Spain, November 1996, submitted paper.