An Experimental Study on the Quality of Service of Video Encoded Sequences over an Emulated Rain-Faded Satellite Channel

Nedo Celandroni^, Franco Davoli*, Erina Ferro^,
Stefano Vignola*, Sandro Zappatore*, Andrea Zinicola*

^ISTI-CNR Institute of National Research Council (C.N.R.),
C.N.R. Pisa Research Area, Via Moruzzi 1, 56124 Pisa, Italy
Nedo.Celandroni, Erina.Ferro@cnuce.cnr.it

*CNIT – Consorzio Nazionale Interuniversitario per le Telecomunicazioni
National Multimedia Communications Laboratory, Via Diocleziano 328, Napoli, Italy
Franco.Davoli, Stefano.Vignola, Sandro.Zappatore, Andrea.Zinicola@cnit.it

Abstract. Video coding techniques, widely used in videoconferencing and streaming applications over the Internet, may face degradation in performance when traversing satellite channels. The aim of the present paper is to provide objective measurements of the quality of video encoded sequences for two of the most popular video coders in this setting (namely, H.261 and MPEG-2), after transmission over faded satellite channels. In particular, the emphasis is on the Ka band, which is becoming an attractive alternative for commercial applications, but is more prone to quality degradation than the currently widely used Ku band. In order to test the behaviour of the coders in repeatable experimental conditions and over different data link platforms, two transmission chains have been implemented in the laboratory, by careful emulation of the required environment. The data link layer has been based on HDLC-like and DVB protocols, respectively. The experiments have been performed in the presence of additive white Gaussian noise (AWGN), and by using fading patterns derived by real world data. The results obtained highlight some often neglected aspects in the behaviour of the coders under examination, both in relation to their comparative performance and to their adaptability to different underlying data link protocols.

Preferred address for correspondence:

Prof. Franco Davoli
DIST-University of Genoa
Via Opera Pia 13
I-16145 Genova, Italy

Phone: +39 010 3532732
Fax: +39 010 3532154
E-mail: franco@dist.unige.it

Work partially supported by the Italian Space Agency (ASI) in the framework of the National Project “Study, Design, and Implementation of a Reconfigurable Satellite Network with Guaranteed QoS for Multimedia Applications” and by the Italian National Research Council (C.N.R.) under the “5%” Multimedia Project.

1. Introduction

The recent years have seen an increase in the use of satellite networks, especially those based on geostationary (GEO) spacecrafts, in packet service delivery. At the same time, there has been a growing popularity of video transmission over the Internet, and of Direct Video Broadcasting (DVB) for the commercial diffusion of digital TV. Both aspects have greatly contributed to the diffusion of video coding applications that have to face potentially harsher transmission environments than the ones they were originally designed to work with, especially as regards traversing packet networks with the presence of satellite links. In view of the forthcoming exploitation of the Ka satellite band (20 – 30 GHz), where the effects of fading caused by adverse atmospheric conditions are more pronounced than in the currently widely used Ku band, the need arises to test the performance of commonly used video coders in such operating environments.

The two most widely used standard video coding techniques are ITU-T H.261 [1] and ISO MPEG-2 [2]. In particular, the former is utilized in a number of videoconferencing packages in use over the Internet Protocol Suite (e.g., VIC [3]). The latter is the standard in use for digital TV broadcasting over DVB-S (Digital Video Broadcasting – Satellite) [4]. Numerous applications over packet networks rely on the use of one or the other coder, especially as regards distance learning, telemedicine, cooperative working, and video streaming for entertainment. Due to the increasing use of satellite channels, several such applications are transported over them, either for the entire network path or for a part of it. A recent experience in this sense, among others, is the one of the Italian National Consortium for Telecommunications (CNIT), which has undertaken a number of projects, based on the TCP/IP protocol suite over a heterogeneous network platform, which also includes Ka band satellite links [5-8]. The multimedia applications employed in these projects make extensive use of video encoding techniques, like H.261 and MPEG2-MPEG4, which are transported over a RTP/UDP/IP platform.

In general, when packet data transmission is performed over a noisy satellite link, we may have different video encoders whose frames can be segmented in various ways, according to the network and data link layer characteristics. The latter may range from IP over HDLC-like data link protocols, to IP over DVB, or to direct transmission of MPEG or H.261 over DVB.

As a matter of fact, in an environment characterized by a non-negligible bit error probability (BER), these different scenarios may have a strong impact on the quality of the decoded video sequence. The literature is poor in the analysis of the effects produced by corrupted bits in compressed video bit streams. The results of a transmission experiment of MPEG-2 coded video data over a satellite link affected by noise can be found in [9], where it was investigated under which conditions this type of transmission is economically feasible. Within a similar setting, we have undertaken an evaluation of the impact of the satellite channel characteristics on the performance of both H.261 and MPEG-2 video coded streams, derived from TV news, in different satellite channel quality conditions. In the case of H.261, both IP and direct DVB have been used in the packetization of the encoded sequence. As the experiments with H.261 have been carried on as first ones, the results obtained addressed our decision toward the choice of testing MPEG-2 with DVB encapsulation only. The goal is to quantify the quality of the received video stream, both in an objective and in a subjective mode. In order to obtain repeatable experiments to compare the various solutions in the same setting, we have performed the tests in a laboratory environment, where an accurate emulation of the Italsat Ka band satellite channel has been carried on(1). The characteristics of H.261 and MPEG-2 are sketched in Section 2, in particular as regards their use on error-prone channels. Section 3

(1) At the time of writing, the Italsat satellite was no longer operative.

presents the laboratory testbed, while the results of the experiments are described and commented in Section 4. Section 5 contains our conclusions.

2. Packetized H.261 and MPEG-2 video sequences in error-prone environments

The video coding techniques (such as MPEG-x and H.26x families), described in the main international standards, are highly sensitive to errors during transmission. The process of removing spatial and temporal redundancy [10] means that the remaining data are very important to the correct reconstruction of the video sequence in the decoder. It means that if any part of the coded data is lost or corrupted, the decoded sequence may be significantly distorted. An error during the transmission of a coded video sequence may have a very significant effect, because decoding and decompressing the coded data may lead to a magnification of the error. The resulting distorted area may propagate spatially and temporally through the decoded sequence, leading to a large area of distortion in a series of frames, thus significantly reducing the visual quality and, in some cases, ending up in a decoded sequence that disturbs the observer's sight.

In order to evaluate the effect of different levels and patterns of errors on the quality of the decoded video, it is necessary to use subjective and/or analytical testing methods to determine the visual quality of the sequence. The commonly used subjective assessment is the perceptual quality, called the *mean opinion score* (MOS)[(2)], which requires several observers and many tests in order to provide a reasonable statistical spread of results.

The reference measures used for an objective video quality assessment are the *mean squared error* (MSE), calculated on the difference signal between the original error-free sequence and the received decoded sequence, and the *peak signal to noise ratio* (PSNR), measured in dB. More specifically, given an M x N matrix of pixels, representing a single picture, we have, for the picture MSE,

$$MSE_P = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left[s(i,j) - \hat{s}(i,j) \right]^2 \qquad (1)$$

where $s(i,j)$ and $\hat{s}(i,j)$ are the ij-th pixel of the original and of the decoded image, respectively, and the MSE of the whole sequence is obtained by averaging over time.

Then, the PSNR is defined as:

$$PSNR = 10 \ \log_{10} \frac{(2^n - 1)^2}{MSE} \qquad [dB] \qquad (2)$$

where n is the number of bits required to represent each pixel.

To evaluate MSE and PSNR it would be necessary that the transmission medium delivered the whole content of the video stream, including the corrupted parts, without discarding data blocks affected by errors that cannot be recovered by the FEC code. This is possible in the DVB environment, where the receiver can also accept corrupted packets without discarding them. At this point, the problem of handling the corrupted sequence is delegated to the entropic decompression module of the video decoder. The latter stops working at the first error occurrence, and resumes only when receiving the appropriate synchronization block (which depends on the specific video decoder in use). Thus, both the H.261 and MPEG-2 video decoders are not able, in general, to handle errored sequences. For this reason, even though the data link protocol would deliver the entire received bit stream, synchronization losses might occur at the video decoder, which would require a synchronization recovery

(2) Recommendation 500-5 of the ITU-R.

mechanism. The temporary loss of data in this situation also requires a choice on the strategy to apply to replace a missing data block. The solution depends on the structure of the video coding algorithm and on the type of packetization adopted.
H.261 video is organized into a hierarchy of layers, namely, from top level: Picture, Group Of Blocks (GOB), Macro Block (MB), and Block. Each layer is built from the lower one, and it contains its data payload and a header composed by the parameters used for the bit stream generation. As a result of this hierarchical structure, in order to decode a MB, the knowledge of the picture and GOB headers, the MB belongs to, is sufficient. A Picture or even a GOB can sometimes be too large to fit into a single packet. Therefore it is reasonable to choose the MB as unit of fragmentation: packets have to start and to end at a MB boundary. To allow independent processing of packets, some state information derived from the frame and the GOB headers, which is necessary to decode the MBs in the packet, is carried within each packet [11]. In our experiment, as far as H.261 is concerned, we have chosen the MB as fragmentation unit, and to freeze the previous MB [1] correctly received and to repeat it for substituting the missing one, in case of loss.

An MPEG encoder uses two basic techniques: block-based motion compensation for the reduction of temporal redundancy (*inter-frame* coding), and transform domain-based compression for the reduction of residual spatial redundancy (*intra-frame* coding). Three main picture types are defined: intra-pictures (I), predicted pictures (P), and bidirectionally-predicted pictures (B). The first ones are self-contained, since they use only transform coding, and provide access points to the coded sequence where decoding can begin. They are coded without reference to other pictures and with only moderate compression, and are used for predicting P and B pictures in inter-frame coding. They give the lowest compression ratios within MPEG. P pictures are generally used for further prediction and are coded more efficiently, using forward predictive coding, where the actual frame is coded with reference to the previous frame (I or P). The compression ratio of P frames is significantly higher than that of the I frames. Also, P pictures are used in the inter-frame prediction of other P and B pictures. B pictures provide the highest degree of compression. They are coded using two reference frames, a past and a future frame (I or P frames) for motion compensation. Furthermore they are never used as a temporal reference for other frames. The effect of errors in the bit stream on the decoding process strongly depends on the bit stream structure itself and on the choice of coding parameters. MPEG-2 does not impose particular requirements on the choice of coding parameters. It does not specify either the decoder architecture or the error handling and correction strategies. Thus, different decoders can exhibit quite different performance in case of errors. The robustness to the errors is obtained by organizing the bit stream in a hierarchy of levels, six for the video bit stream. The upper layers start with *start codes*, which cannot be found in any other position inside the bit stream. The decoder, when an error is found, re-synchronizes itself on the next start code. The information between the errored element in the bit stream and the following start code cannot be recovered. The effect on an error on the decoded sequence depends on the layer where the error occurred: the upper the layer is, the worse the distorsion introduced will be (in the worst case, many frames could be lost).

3. The Laboratory Testbed

In order to have repeatable experiments, we have emulated the satellite link by using the telecommunication systems testbed of the CNIT National Laboratory for Multimedia Communications in Naples. The experimental layout is shown in Fig. 1. We have set various signal to noise ratio (S/N) values suitable for evaluating MSE and PSNR at constant bit rates. Furthermore, taking the Italsat satellite characteristics [12] and the CNIT earth station parameters (1.8 m antenna and a 1 W SSPA), we derived the link budget, which then resulted of up-link noise predominant type. We also assumed that the up-link fade were perfectly compensated by the up-link power control, and we applied real attenuation data on the down-link by means of the fading generator. The modems used were of Fairchild SM290 type, operating in QPSK, at an information rate of 2048 kbps, with a sequential FEC (forward error correction) of 3/4.

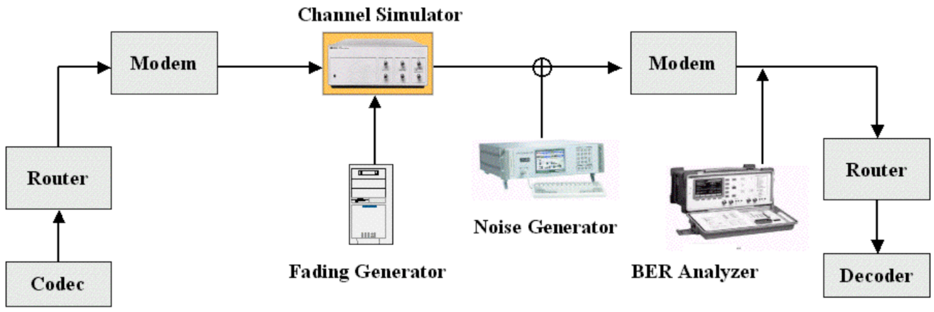


Fig. 1. The laboratory experimental setup.

The logical data flow diagram of the H.261 experiment is shown in Fig. 2, which highlights both situations of transmission chains that have been implemented: i) relatively long IP packets (1500 byte Maximum Transfer Unit - MTU, with average length of 347 bytes), carried within a PPP (point-to-point protocol) data link, which discards corrupted PDUs; ii) DVB frames, carrying a payload of 184 bytes per packet. In this second case, packets in error are not discarded but, for the previously said synchronization loss reason, the H.261 decoder is generally unable to decode them.

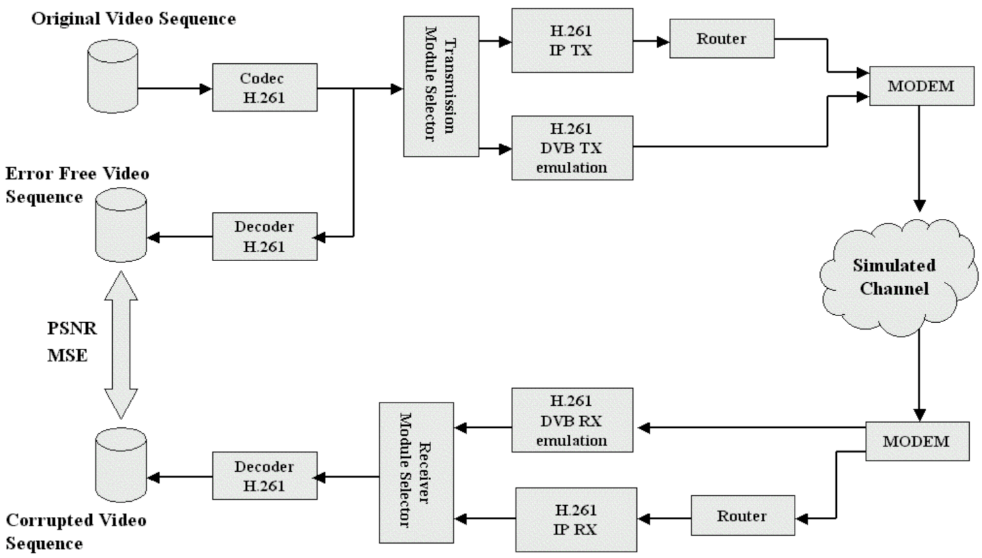


Fig. 2. The logical data flow diagram of the experiment for the H.261 coder.

When MPEG-2 is used, the logical data flow diagram is substantially the same as in Fig. 2. The only difference is that in this case there is no division between IP and DVB packets, because only DVB encapsulation is present. In the experiments, we have used a Main Profile Main Level MPEG-2 video sequence, coded at constant bit rate (900 kbps), with a GOP length of 13, with B frames, and a distance between two consecutive P frames of 2 (IBBPBBPBBPBBP). Whenever the loss of an entire frame occurs, we have chosen to keep the previous frame.

4. The Measurement Results

Several tests have been performed over the emulated environment. The video sequence utilized is taken from a 5 minutes recording of TV news. The chosen sequence is particularly critical, as it contains moving and still images, text (headline news), and it is characterized by a number of zooming movements. The original and the decoded noisy sequences can be seen at http://www.labnet.cnit.it/Ka-test. The results are organized as follows: we show the H.261 case first, in both IP-over-PPP and DVB packetizations, and then the MPEG-2 over DVB case, also in comparison with the previous one.

4. 1. H.261 case

Figures from 3 to 6 refer to measurements done under a constant BER level during the entire duration of the transmission.

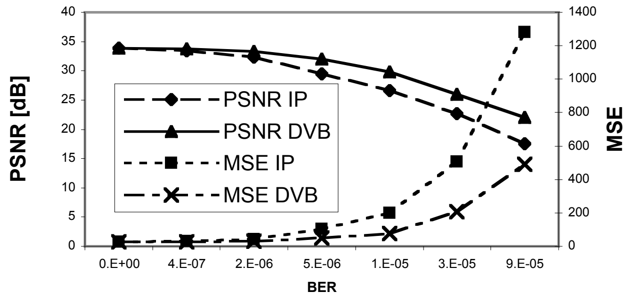


Fig. 3. PSNR (left axis) and MSE (right axis) vs. BER for transmissions of a H.261 coded sequence. Two cases are reported: the first one refers to an IP stream over PPP, while in the second one the stream is directly packetized according to a DVB format.

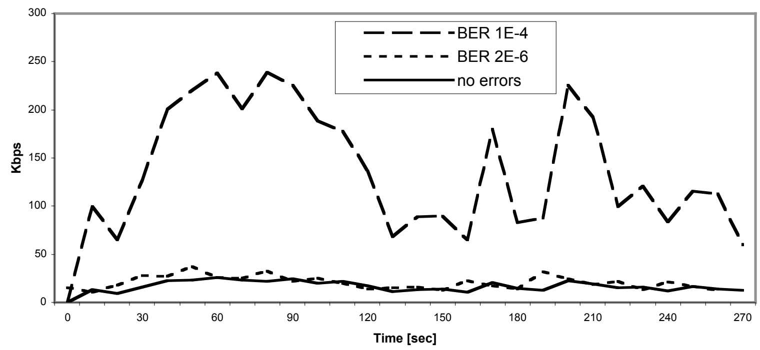
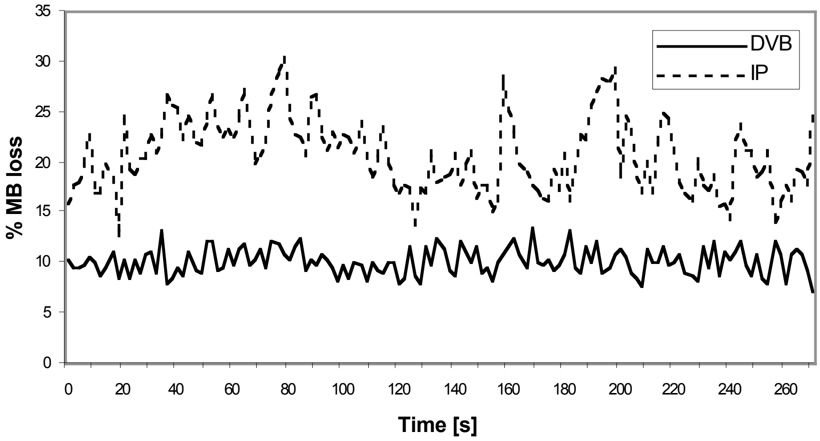
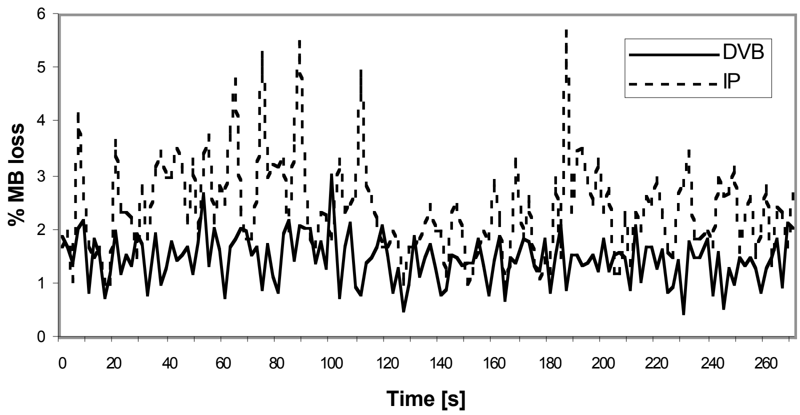


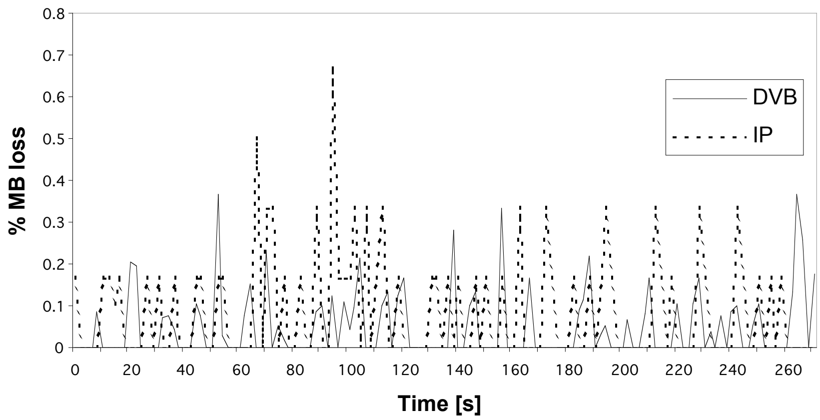
Fig. 4. Goodput difference between DVB and IP transmissions for two BER levels. The "no errors" line represents the overhead introduced in the DVB case.



a)



b)



c)

Fig. 5. Percentage of lost Macro Blocks vs. time during a transmission of IP over PPP and direct DVB on a satellite link with BER 10^{-4} (a), 10^{-5} (b) and 10^{-7} (c).

Figure 3 reports the results of MSE and PSNR, calculated at different BER values, by averaging the same quantities over all video frames in the sequence. The 33.9 dB PSNR at BER=0 is caused by the coding/decoding process, which is lossy. For all other BER values, the DVB case exhibits a lower degradation.
In order to explain this result, we have to make the following considerations. In the IP case, packets affected by at least one errored bit are discarded, and the decoder looses all the MBs they contain. In the DVB case, packets in error are not discarded; however, the decoder looses the corresponding MBs, owing to the possible synchronization loss mentioned above. As DVB packets have shorter length, the number of unused MBs will be anyway less than in the previous IP case. The gain resulting from this effect actually justifies the additional overhead that is necessary in order to make the MBs contained in a DVB packet identifiable (i.e., attributable to the corresponding GOB and Picture). As a matter of fact, the corresponding goodput (in terms of MBs actually available to the decoder) is always in favour of the DVB case, as evidenced in Fig. 4 for two BER values.
Figure 5 shows the behaviour of the percentage of lost MBs, averaged over 2-second intervals, at different BER values. The comparison is practically always in favour of DVB.

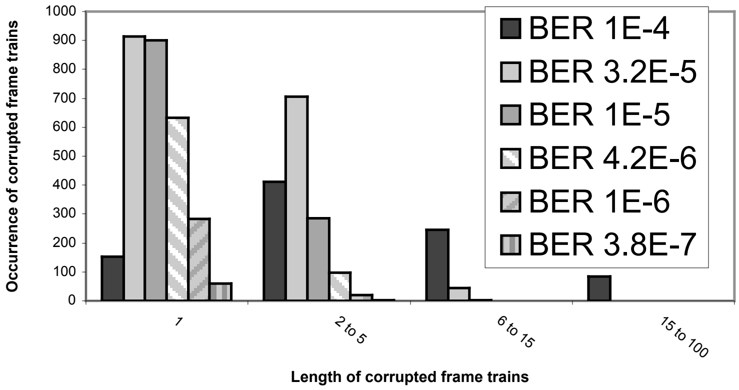


Fig. 6. Number of occurrences of consecutive corrupted frames for different levels of BER.

In Fig. 6 the corrupted frames are grouped in four classes according to the length of corrupted sequences. The first class takes into account isolated corrupted frames, the second class contains sequences formed by consecutive corrupted frames from 2 to 5, and so on.

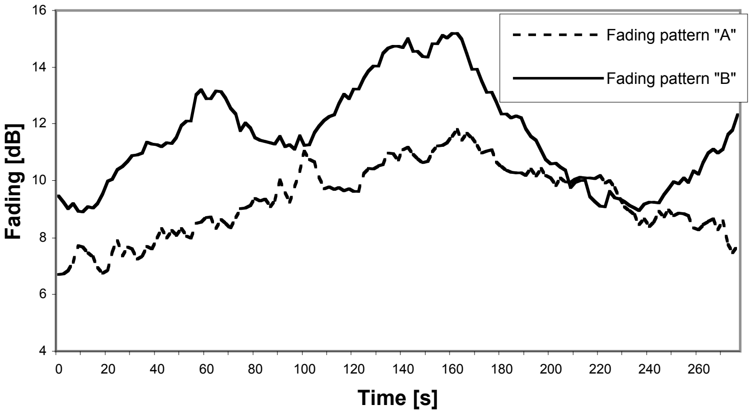


Fig. 7. Plot of the fading vs. time in the two cases studied: the pattern indicated as “B”, which corresponds to a quite severe fading, causes a temporary sync loss at the demodulator.

The set of figures from 7 to 13 refer to the situation where a specific fading pattern, taken from real Ka band data, has been applied.
Figure 7 shows the two rain attenuation patterns used, which are taken from the results of the propagation experiment carried out in Ka band on the Olympus satellite by the CSTS (Centro Studi sulle Telecomunicazioni Spaziali) Institute, on behalf of the Italian Space Agency (ASI). The attenuation samples considered were 1-second averages, expressed in dB, of the signal power attenuation with respect to clear sky conditions. Patterns A and B are taken from data recorded at the Spino d'Adda (Northern Italy) station on September 21st and 23th, 1992, respectively. These patterns have been used in the experiments whose results are reported in Figs. 8-13 below. In Figs. 8 and 9 the MB loss is reported in the IP and DVB cases for both fade patterns, while Figs. 10-13 show the calculated MSE and PSNR. When fade pattern B is present, the modem experiences periods of synchronization losses lasting 5 seconds on average. In general, the results confirm the better performance of the DVB case.

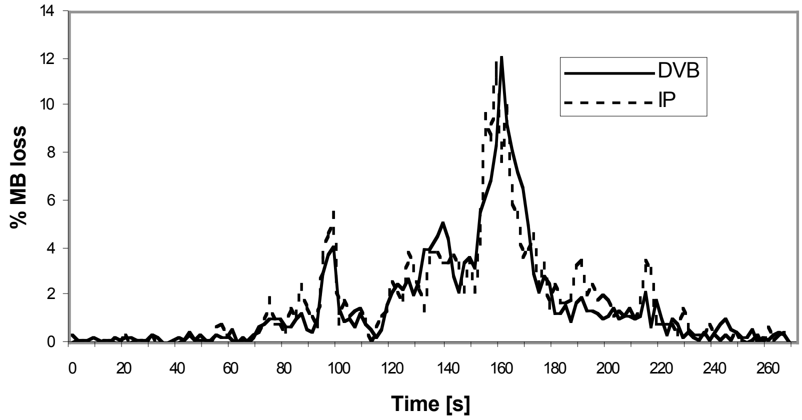


Fig. 8. Percentage of lost Macro Blocks vs. time for a transmission of IP over PPP and direct DVB, when the satellite link is affected by the fading pattern "A" reported in Fig. 7.

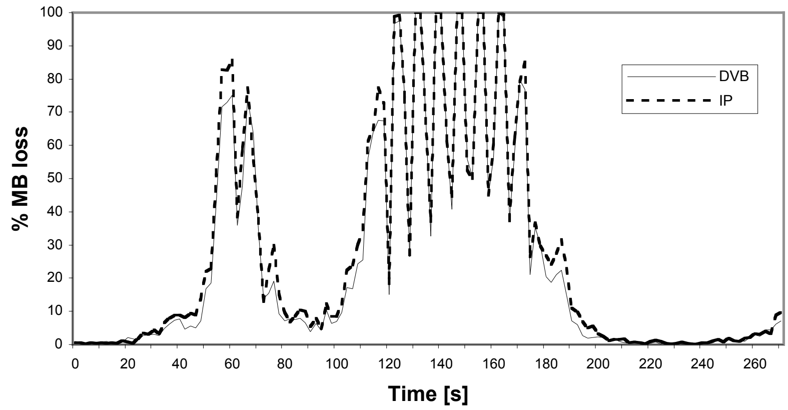


Fig. 9. Percentage of lost Macro Blocks vs. time for a transmission of IP over PPP and direct DVB when the satellite link is affected by the fading pattern "B" reported in Fig. 7.

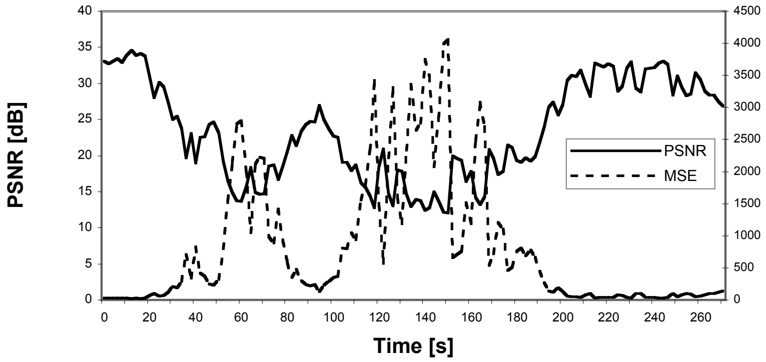


Fig. 10. PSNR and MSE vs. time for a DVB transmission when the satellite link is affected by the fading pattern "B" reported in Fig. 7.

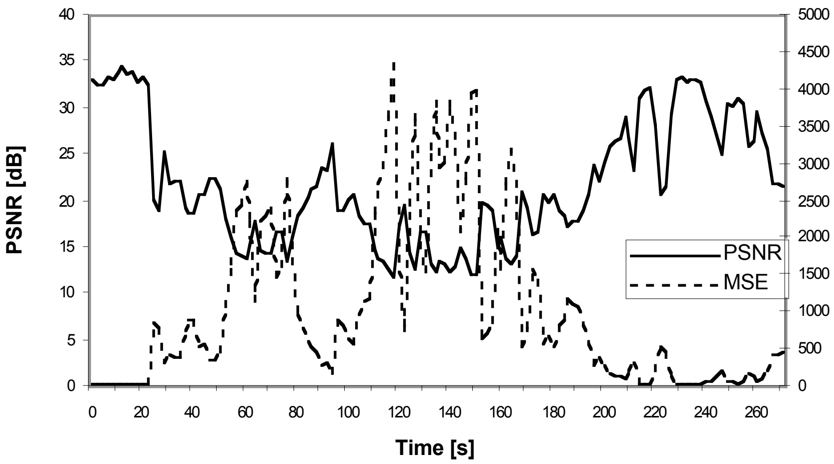


Fig. 11. PSNR and MSE vs. time for an IP over PPP transmission, when the satellite link is affected by the fading pattern "B" reported in Fig. 7.

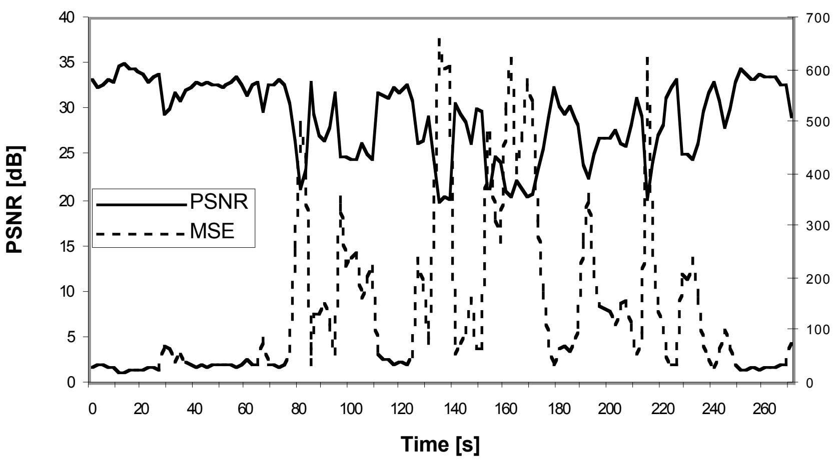


Fig. 12. PSNR and MSE vs. time for an IP over PPP transmission, when the satellite link is affected by the fading pattern "A" reported in Fig. 7.

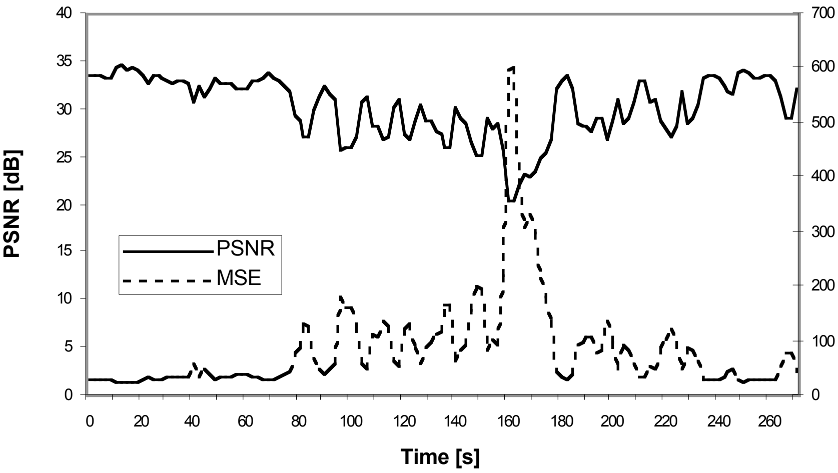


Fig. 13. PSNR and MSE vs. time for a DVB transmission, when the satellite link is affected by the fading pattern "A" reported in Fig. 7.

4.2. MPEG-2 case

The indication derived from the H.261 experiments, i.e., that DVB encapsulation performs better than IP over PPP, suggested to directly test the performance of MPEG-2 over DVB, and thus to compare the two video coding techniques in the DVB case only. Figure 14 reports the results of MSE and PSNR, for DVB transmissions, calculated at different BER values, by averaging the same quantities over all video frames in the sequence. For comparison, besides the MPEG-2 results, in the same figure we show again the H.261 case, as well. However, since we want to compare the *relative* degradation between the two different coding methods, in this case we plot the PSNR values after subtraction of the ones at BER=0 (no channel degradation). These values correspond to 33.9 dB and 26.8 dB for H.261 and MPEG-2, respectively. As we have already noted, they are caused by the lossy coding/decoding processes. On the other hand, in spite of the 7 dB difference in favour of H.261, our MOS computation at BER=0 resulted more than 1/2 point higher in favour of MPEG-2. This confirms (see, e.g., [13]) that there is no objective measurement system that is currently able to replace subjective testing, and that PSNR is adequate to quantify the effects of transport solutions within the same coding algorithm, but it is not suitable to compare different coding algorithms. What is apparent from the differential PSNR values in Fig. 14 is that the rate of variation of the relative degradation of MPEG-2 with respect to increasing BER values is higher, thus confirming a lower resilience of this coding method when used on error-prone channels. This effect is also evidenced by the comparison in terms of lost MBs, reported in Fig. 15 for a fixed BER value of 10^{-6}. In fact, even in this case of rather low BER, the behaviour of MB losses for MPEG-2 has a much more bursty shape. The reason for this is that the loss of a MB causes the loss of all successive MBs in the same slice of the frame.
As regards the evaluation in dynamic fading conditions, we show the PSNR of the MPEG-2 case, again compared with H.261 over DVB on the same graph, under the two fading patterns A and B, in Figs. 16 and 17, respectively. In general, it can be observed that, for most of the time, the variation of the MPEG-2 curve with respect to the BER=0 value is higher than that of H.261. The region in Fig. 17, where the two curves partially overlap, is characterized by such a severe fading to cause sync losses at the demodulator. The PSNR in this area is practically meaningless.

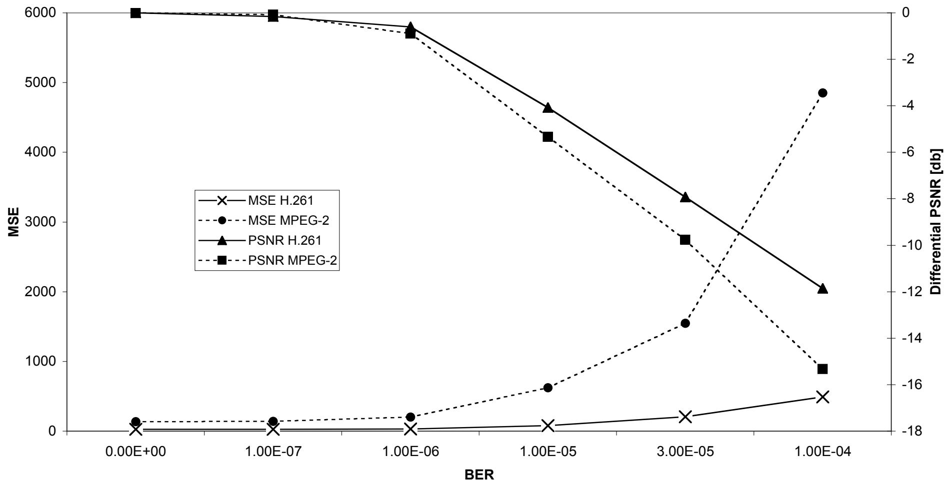


Fig. 14. MSE (left axis) and Differential PSNR (right axis) vs. BER for DVB transmission of the H.261 and MPEG-2 coded sequence.

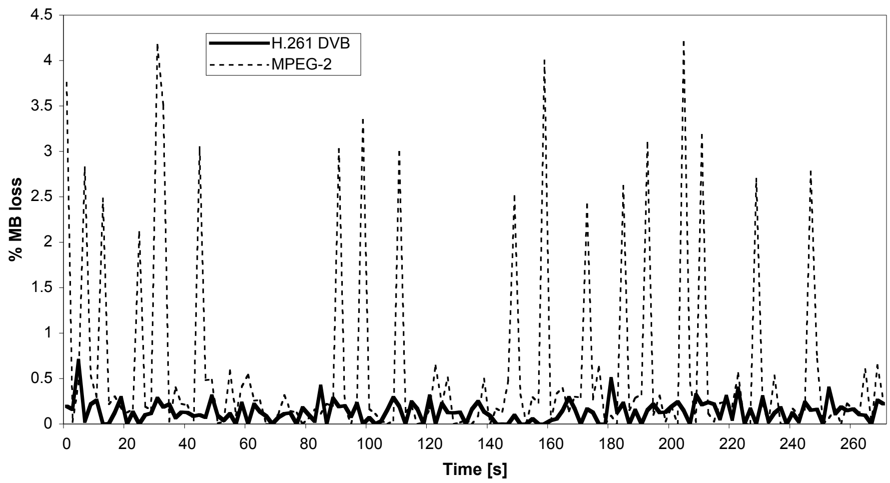


Fig. 15. Percentage of lost MBs vs. time for for DVB transmission of H.261 and MPEG-2 at constant BER=10^{-6}.

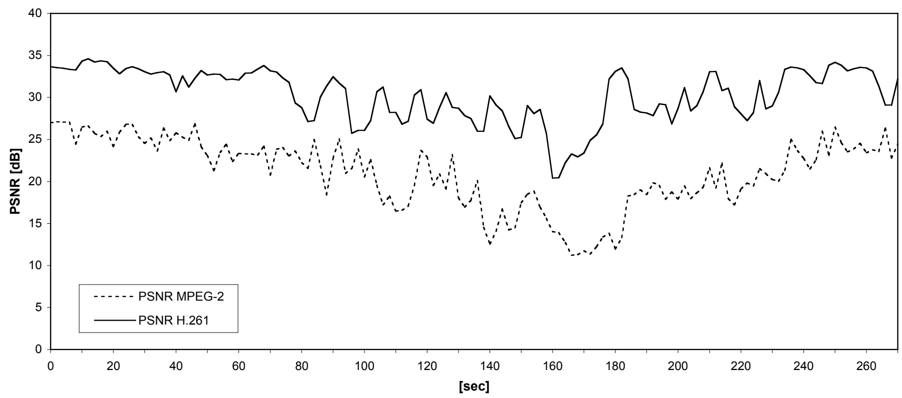


Fig. 16. PSNR vs. time: comparison between DVB transmissions of the H.261 and MPEG-2 coded sequences, when the satellite link is affected by the fading pattern “A” reported in Fig. 7.

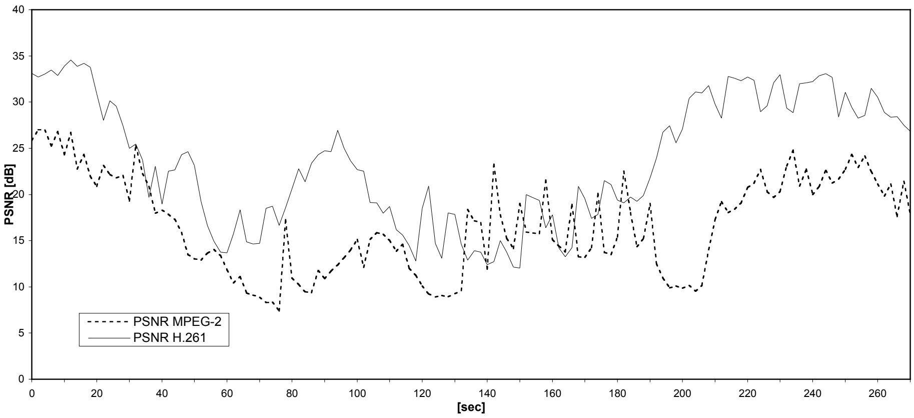


Fig. 17. PSNR vs. time: comparison between DVB transmissions of the H.261 and MPEG-2 coded sequences, when the satellite link is affected by the fading pattern "B" reported in Fig. 7.

The results of the MOS evaluation over the BER=0 case and the two corrupted sequences, obtained from the average scores of 20 independent observers, are reported in Table I. They are basically in agreement with our previous discussion. As a matter of fact, the better subjective quality of MPEG-2 over H.261 for the uncorrupted decoded sequences does not hold under rather severe fading conditions.

	MOS
H.261 – BER=0	3.2
H.261- Fading A	2.3
H.261- Fading B	1.1
MPEG-2 – BER=0	3.8
MPEG-2 - Fading A	1.1
MPEG-2 - Fading B	0.2

Table I. MOS evaluation.

5. Conclusions

In this experiment, both the H.261 and the MPEG-2 video stream coders have been taken into account, as standard coding algorithms that are of widespread use in several applications. The CNIT National Multimedia Communications Laboratory in Naples has offered a perfect environment to carry on the experimental evaluation. The tests have required some modifications in the decoders' software of both video coding algorithms, in order to make the comparison feasible, to choose the correct packet sizes and to implement a recovery strategy for lost frames. As far as H.261 tests are concerned, the direct DVB encapsulation has always exhibited better performance, with respect to IP over PPP. When H.261 is compared with MPEG-2 (both with DVB encapsulation), the results confirm that there is no objective measurement system that is currently able to replace subjective testing, and that PSNR is adequate to measure transport solutions within the same coding algorithm, but it is not suitable to compare different coding algorithms. Thus, since PSNR is often not completely adequate to express the actual quality of a video stream perceived by the user, the MOS

subjective metric has also been employed in order to evaluate the effects of fading on multimedia transmissions. However, a comparison in terms of differential PSNR has shown a higher robustness of H.261 when operating over degraded satellite channels.
The results obtained also suggest the possible (partial) re-implementation of some classic video coders, with the aim of improving the robustness of the streams transmitted over the satellite channel.

References

[1] ITU-T Recommendation H.261, Video codec for audiovisual services at p x 64 kbit/s, 1993.

[2] ISO/IEC 13818-2, ITU-T Rec. H.262, "Generic coding of moving pictures and associated audio information: video", 1995.

[3] http://www-nrg.ee.lbl.gov/vic/

[4] ETSI EN 300 421 V1.1.2, Digital Video Broadcasting, frame structure, channel coding and modulation for 11-12 GHz satellite services, 1997.

[5] D. Adami, M. Marchese, L.S. Ronga, "TCP/IP-based multimedia applications and services over satellite links: experience from an ASI/CNIT project", IEEE Personal Communications Mag., vol. 8, no. 3 , pp. 20-27, June 2001.

[6] G. Prati, E. Del Re, S. Pupolin, L.S. Ronga, M. Marchese, D. Adami, P. Castoldi, S. Vignola, "The Teledoctorate project: a CNIT experience", Proc. 2nd COST 276 Workshop on Information and Knowledge Management for Integrated Media Communication, Florence, Italy, March 2002.

[7] F. Davoli, S. Zappatore, "The LABNET project: infrastructure and software architecture", Proc. 2nd COST 276 Workshop on Information and Knowledge Management for Integrated Media Communication, Florence, Italy, March 2002.

[8] E. Ferro, G. Maral, L. Franck, "COST272: packet-oriented service delivery via satellite: Part II", Global Commun. Newsletter, pp. 2-4, in IEEE Commun. Mag., vol. 40, no. 10, Oct. 2002.

[9] N. Celandroni, E. Ferro, F. Potortì, A. Chimienti, M. Lucenteforte, R. Picco, "MPEG-2 coded video traces transmitted over a satellite link: scalable and non-scalable solutions in rain fading condition", Multimedia Tools and Applications, vol. 10, no. 1, pp. 73-97, 2000.

[10] A. Webster, C. Jones, M. Pinson, S. Voran, S. Wolf, "An objective video quality assessment system based on human perception", Proc. SPIE Human Vision, Visual Processing and Digital Display IV, Feb. 1993, San Jose, CA, vol. 1913.

[11] RFC 2032 " RTP Payload format for H.261 Video", October 1996.

[12] F. Carducci, M. Francesi, "The Italsat Satellite System", International Journal of Satellite Communications, vol. 13, pp. 49-81, 1995.

[13] A. M. Rohaly, *et al.*, "Video Quality Experts Group: current results and future directions ", Proc. SPIE The International Society for Optical Engineering, 4067, pt. 1-3, pp. 742-753, 2000.