# User-Oriented QoS in Packet Video Delivery

Olivier Verscheure and Xavier Garcia, Swiss Federal Institute of Technology

Gunnar Karlsson, Royal Institute of Technology (KTH)

Jean-Pierre Hubaux, Swiss Federal Institute of Technology

**Abstract**

We focus on packet video delivery, with an emphasis on the quality of service perceived by the end-user. A video signal passes through several subsystems, such as the source coder, the network and the decoder. Each of these can impair the information, either by data loss or by introducing delay. We describe how each of the subsystems can be tuned to optimize the quality of the delivered signal, for a given available bit rate in the network. The assessment of end-user quality is not trivial. We present recent research results, which rely on a model of the human visual system.

## 1 Introduction

The field of telecommunications is a driving force in today's society. Market deregulation facilitates the development of communication services and applications. Among these, interactive multimedia applications are now gaining interest owing to the proliferation of web technology. The development of corporate intranets opens the possibility for video conferencing and distributed collaborative work. Residential access, via technologies such as digital subscriber loop techniques (xDSL) and cable modem, is being encouraged by the entertainment and personal computer industries. Broadband access to the Internet and a large choice of on-demand services could enable new mass markets.

Multimedia can be placed in the intersection of traditionally separated industries as depicted in Fig. 1. This intersection reflects the integration of multiple media in a single application. The transmission of such applications requires a network capable of handling different types of data.

For several years, the solution has been considered to be the *asynchronous transfer mode* (ATM). ATM is the network technology for the broadband integrated services digital network (B-ISDN). Now the role of ATM is being challenged by the success of the Internet and other IP-based networks due to the new developments of integrated and differentiated services.

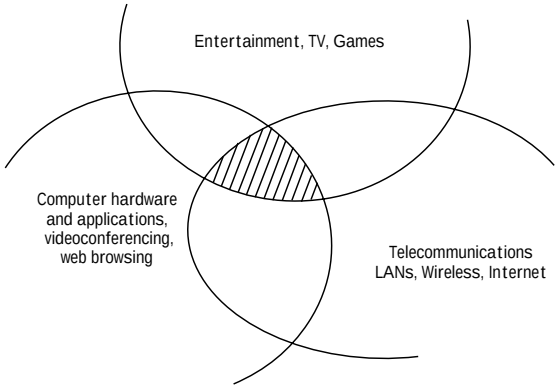


Figure 1: *Convergence of multimedia enabling sectors.*

A truly integrated network will have to cope with different traffic characteristics and quality requirements in terms of delay, delay jitter and data loss. Providing integration of heterogeneous traffic and adequate QoS to users has been proven difficult to achieve.

Work remains to be done to optimize multimedia applications so they can be offered at attractive prices. In other words, the user expects an adequate audio-visual quality at the lowest possible cost. From the user's viewpoint, in the case of video transmission over packet networks, both the encoding and the transmission processes affect the quality of service. The most economic offering can thus only be found by considering the entire system and not by optimization of individual system components in isolation [1].

The rest of the paper is organized in the following way. In Sec. 2, we summarize video compression, with a focus on MPEG-2. In Sec. 3, we explain how compressed video is conveyed over packet networks (e.g. ATM and IP). In Sec. 4, we define the concept of user-perceived QoS and recommend the use of models that take the human vision into account. Finally, Sec. 5 details the impact of MPEG-2 encoding rate and data loss on the user perceived quality.

# 2 Major Video Compression Standards

## 2.1 Overview

The purpose of source coding (or compression) is data rate reduction. For example, the data rate of an uncompressed NTSC [1] TV-resolution video stream is close to 170 MBits/s, which corresponds to less than 30 seconds of recording time on a regular compact disk (CD).

The choice of a compression standard mostly depends on the available transmission or storage capacity as well as the features required by the application. The most cited video standards are H.263, H.261, MPEG-1 and MPEG-2 [2]. They are based on the techniques of discrete cosine transform (DCT) and motion prediction (see Fig. 2), even though they target different applications (i.e. encoding rates and qualities). These applications range from desktop video-conferencing to TV channels broadcast over satellite, cable, and other broadcast channels [2]. The former typically uses H.261 or H.263 while MPEG-2 is the most appropriate compression standard for the video broadcast applications.

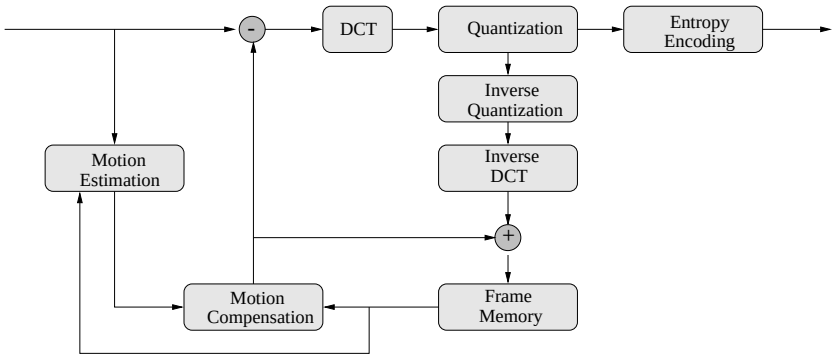


Figure 2: *Block diagram of an encoder.*

The MPEG-2 standard is an audio-visual standard developed by the International Organization for Standards (ISO) together with the International Electrotechnical Commission (IEC) [3]. The video part of MPEG-2 permits data rates up to 100 Mbps and also supports interlaced video formats and a number of advanced features, including those supporting HDTV[3]. MPEG-2 is capable of compressing NTSC or PAL TV-resolution video into an average bit rate of 3 to 7 Mbps with a quality comparable to analog broadcast TV [4].

[1] NTSC stands for National Television Systems Committee

[2] MPEG stands for Moving Picture Experts Group

[3] HDTV stands for High-Definition TeleVision

## 2.2 MPEG-2 Background

An MPEG-2 video stream is hierarchically structured as illustrated in Fig. 3. The smallest entity defined by the standard is the *block*, which is an area of $8 \times 8$ pixels of luminance or chominance. A *macroblock* ($16 \times 16$ pixels) contains four blocks of luminance samples and two, four or eight blocks of chrominance samples, depending on the chrominance format. A variable number of macroblocks is encapsulated in an entity called a *slice*. A new slice always starts on each new line of macroblocks. Slices occur in the bitstream in the order in which they are encountered. Thus, each *picture* is composed of a variable number of slices.

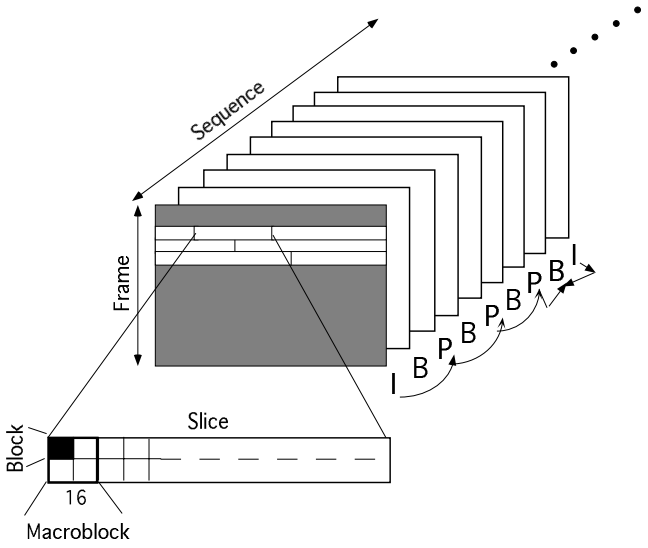


Figure 3: *MPEG-2 video structure.*

The MPEG-2 video syntax defines three different types of pictures :

- Intra-coded (or I-) pictures are coded without reference to preceding or upcoming pictures in the sequence. A picture is divided into $8 \times 8$ blocks of pixels and a two-dimensional discrete cosine transform (DCT) is applied to each block. The resulting DCT coefficients are quantized and variable-length coded. The quantization is performed by dividing each DCT coefficient by a quantizer and by rounding the result to an integer. The quantizer applied to a DCT coefficient comes from the multiplication of a quantizer scale, the so-called MQUANT, and the corresponding quantizer matrix entry. A different MQUANT value may be used for each macroblock. As the MQUANT increases, the quality decreases in favor of the increased compression factor (see Sec. 5.2). Intra-coding provides a moderate compression rate while

allowing random access into the compressed video data.

- Predicted (or P-) pictures are coded with respect to the nearest previous I- or P-picture. The predictions of the best-matching macroblocks are indicated by motion vectors that describe the displacement between them and the target macroblocks. The difference between the best-matching and the target macroblocks, called the *prediction error*, is encoded using the DCT-based intraframe technique summarized above. The motion vectors, as well as the quantized DCT coefficients, are variable-length coded. It is noteworthy that individual macroblocks may still be intra-coded in P-pictures (i.e. do not use motion estimation).

- Bidirectional (or B-) pictures use both past and future I or P pictures as reference. Motion compensation is also applied here. However, both forward and backward motion vectors may be used for each macroblock since B-pictures are coded in relation to two reference pictures. Like in P-pictures, macroblocks in B-pictures may be intra-coded, and furthermore, some of the motion-estimated macroblocks might use only one motion vector (i.e. forward or backward motion vector). Bidirectional encoding provides the highest compression rate while introducing some delay.

The use of these three picture types allows MPEG-2 to be robust to packet loss (I-pictures provide stop points for the error propagation) and efficient (B- and P-pictures allow good compression). Furthermore, the MPEG-2 standard does not specify how I-, P- and B-pictures are mixed together. As mentioned, all coding modes can even be chosen per macroblock, which allows fine-tuned tradeoffs of robustness and efficiency.

The MPEG-2 system document specifies two systems. The first multiplexes video, audio and data of a single program together for relatively error-free environments such as storage systems. The resultant aggregate is called a *program stream*. The other system creates a *transport stream* (TS) which can be used for broadcast, video-on-demand and cable TV. The transport stream defines a packet-based protocol for transmission on digital networks. It allows multiplexing of multiple MPEG-2 compressed channels with a fixed-length (188 bytes) format (so called TS packets). It also includes a program clock reference in the TS header, as well as presentation and decoding time stamps.

It is worth noting that a header with syntactic information is inserted before each of the following

information elements: sequence, GOP [4], picture, slice and TS packets.

# 3 Major Networking Technologies

## 3.1 Overview

The most interesting network protocol suites to consider today for broadband communication are the asynchronous transfer mode (ATM) and the internet protocol (IP). The asynchronous transfer mode combines the circuit-switched routing of telephony networks with the asynchronous time-division multiplexing of traditional packet switching. This is accomplished by establishing a virtual channel through the network before accepting any traffic. Data are sent in 53-octet long cells. The network guarantees that all the cells of a call follow the same route and are delivered in the same order as sent. The internet protocol differs in two major respects from ATM. First, IP does not synchronize between the establishment of a route and the start of a session. Also, IP packets are of variable length (up to 65,535 octets). These packets may consequently arrive out of order if the routing decision has changed during the session.

It is inevitable to have delays and losses during transfers across both ATM and IP networks. The delay is chiefly caused by propagation and queuing. The queuing delay depends on the load of the links. Loss of information is mainly caused by a multiplexing overload of such magnitude and duration that the buffers in the nodes overflow. Loss may also be caused by misrouting due to bit-errors in the addresses, but this is less probable. Quality of service usually means that the probability of packet loss and the maximum delay are bounded to specified values. There could also be bounds on the delay variations as part of a service contract.

Service quality is ensured by regulation of the network load. Connection requests state the needed sending rate and its variations. The network only admits as much traffic as it can sustain at the desired quality for a given path.

## 3.2 ATM Networks

There are two types of ATM service classes suitable for transmission of audio/video streams: constant bit rate (CBR) service and variable bit rate (VBR) service (called deterministic bit rate and

[4] The Group of Pictures (GOP) concept as defined by MPEG-1 is not required for MPEG-2.

statistical bit rate transfer capabilities in the ITU Recommendations). CBR service means that virtual channels are allocated portions of capacity that are at least equal to their declared peak rates. Note that CBR service only bounds the bit rate from above and that it may vary within that bound. The ITU Recommendation does not state the associated quality of service but loss-free service with low maximum delay is possible. The VBR service means that a rate below the declared peak rate is allocated for the connection [5]. The specific value depends on the peak and sustainable rates, the maximum burst size that the connection initiator declares, and the quality level that the operator maintains for the service class. It is not clear what quality levels will be offered by network providers. The ATM Forum has separated the variable bit rate service into real-time VBR (rt-VBR) and non-realtime VBR (nrt-VBR). Delay limits are only guaranteed for rt-VBR. The ITU does not make this distinction.

The transport of compressed video over ATM can be done over two ATM adaptation layers (AAL) depending on the service required. If constant-rate video is to be transmitted, AAL1 may be used [6]. It provides a CBR service and features specific to real-time applications such as clock recovery and forward-error correction. Since an equivalent adaptation layer able to transport VBR data does not exist, AAL5 is considered as the second alternative for video transmission [7]. Even if AAL5 does not provide such real-time specific functions, it is generally accepted because it is simple and is able to handle both CBR and VBR traffic.

For MPEG-2, these real-time specific functions are provided by a network adaptation layer, as described in the following.

### Adaptation of MPEG-2 streams: Network Adaptation Layer

Several functions required by multimedia applications are not provided by the ATM adaptation layers. Multimedia applications involve the transmission of synchronized audio, video and data flows. Recommendation H.222.1 [8] specifies a network adaptation layer (NAL) that provides multiplexing and synchronization functions. In particular, the NAL provides multiplexing, timebase recovery, error reporting and priorities on a packet basis. It also describes the mapping of access data units from the applications to the AALs. The NAL is not totally generic. It provides functions specific to MPEG-2 and embeds its systems layer as shown in Fig. 4.

The encapsulation of MPEG-2 TS packets into AAL5-SDUs is defined in the ATM Forum's

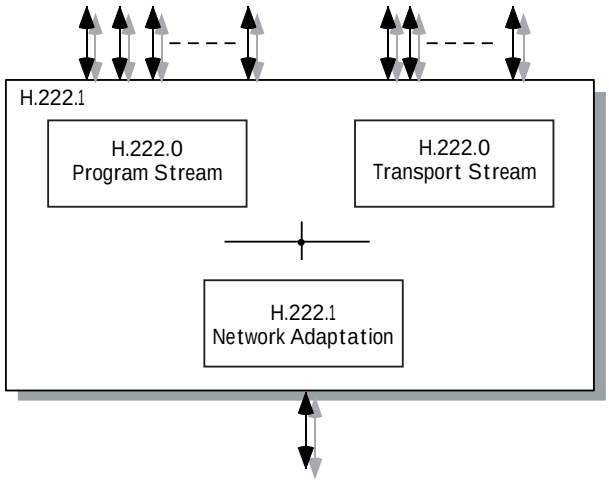


Figure 4: *H.222.1 overview.*

Video on Demand Specification 1.1. The proposed scheme encapsulates two single program transport stream (SPTS) packets, regardless of their information content (i.e. audio, video or timing data) into a single AAL5-SDU. The addition of the AAL5 8 byte trailer results in a AAL5-PDU that is segmented into exactly 8 ATM cells.

For further readings on MPEG-2 over ATM, please refer to [9].

### 3.3 IP-based Networks

The integrated-services architecture for IP has two classes, which can be loosely compared to the CBR and VBR service classes of ATM: the guaranteed service and the controlled-load service [10].

- Guaranteed service (GS) gives a lossless transfer with tight delay bounds.
- Controlled load service (CLS) is supposed to yield a quality corresponding to a lightly loaded IP network at best effort; it is not expressed quantitatively. The admission control is based on the peak rate declared by a session initiator and on measurements of the load in the network. This could lead to higher network efficiency when compared to admission control based only on declared source descriptors.

Both GS and CLS connections can be established by RSVP signaling.

### Adaptation of MPEG-2 streams: RTP encapsulation

The IETF has developed the real-time transport protocol (RTP) [11] suitable for transmitting real-time data such as audio and video, over multicast or unicast networks. RTP does not address resource reservation for providing quality of service for real-time services. It provides payload-type identification, sequence numbers, time stamps and delivery reports (see Fig. 5). RTP relies on lower layers for multicast, timely data delivery and quality of service in general.

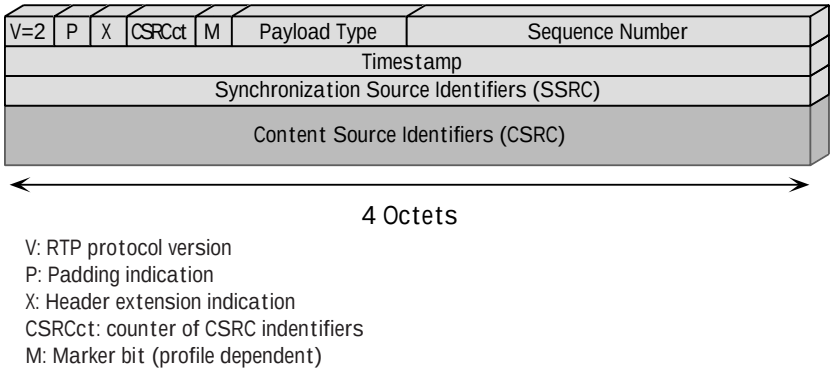


Figure 5: *RTP packet format.*

The real-time control protocol (RTCP), part of the RTP specification, monitors the QoS and conveys information about the participants in a multicast session. RTCP is not intended to support all of an application's control signaling requirements. A higher level session control protocol, may therefore be needed.

RTCP is based on a periodic transmission of control packets to all participants in a session, using the same distribution mechanism as the data packets. The underlying protocol *must provide* multiplexing of the data and control packets.

RTP is intended to be tailored to any particular application. To achieve the needed flexibility, the specification is deliberately incomplete. It defines a set of core functions but allows for extensions and customization depending on the target application. Companion documents have been written which specify the extensions needed for specific applications (called profiles). RFC 2250 defines the packet format for MPEG-1 and MPEG-2 audio and video [12]. It specifies payload identifier and encapsulation schemes for the different packet formats (i.e. TS or program stream).

# 4 User-Oriented QoS in MPEG-2 Video Communications

In MPEG-2 delivery, the video information flows through several subsystems (e.g. coder, network, decoder). Each of these subsystems may degrade the video quality, either by data loss or by introducing some delay. We call *user-oriented QoS* the video quality as perceived by the end-user.

In this section we analyze (i) what may affect the user-oriented QoS, (ii) how to improve it and finally, (iii) how to measure it. We further refine the analysis by individually considering each of the subsystems previously mentioned.

## 4.1 What may affect the QoS

In general, the quality of service a customer perceives results from both the encoding artifacts and delays, as well as from the packet losses, delays and delay jitters caused by transmission.

### 4.1.1 Encoding: artifacts and delays

All lossy compression schemes both distort and delay the signal.

Degradations come from the quantization which is the only irreversible process in a coding scheme. In general, the higher the quantization step, the higher the degradation (see Sec. 5.2 for details). The most usual coding artifacts are ringing around contours, small stains around edges (so called *mosquito noise*), blurring of textured areas and visible block boundaries in almost uniform areas.

The amount of delay introduced is related to the size of the encoding buffer. The bigger the buffer, the smoother the bit rate may become, but it is at the expense of higher delay. For example, completely constant bit rate encoding [5] introduces a maximum delay of around 500 ms while variable bit rate encoding might have smoothing delays as low as a frame time (33 ms for NTSC). In VBR encoding, a trade-off exists between the smoothing delay and bit rate variations in the output stream. Moreover, the regulation of the encoder to avoid overflow of the smoothing buffer causes quality variations in the decoded video stream.

[5] Assuming the output of the encoder is sent over the network at a constant rate

### 4.1.2 Transmission: loss and delay

In an MPEG-2 video stream, data loss that reduces the quality is dependent on the importance of the lost information type. For example, losses in headers affect the quality more than losses of DCT coefficients and motion vectors. The quality degradation depends also on the picture type of the lost video data because of the predictions used for MPEG-2.

Figure 6 shows how network losses map into visual information losses in different types of pictures. Data loss spreads within a single picture up to the next resynchronization point (e.g. picture or slice headers) due to the variable-length coding. This is referred to as spatial propagation. When loss occurs in a reference picture (I- or P- picture), the lost macroblocks will affect the predicted macroblocks in subsequent frame(s). This is known as temporal propagation.

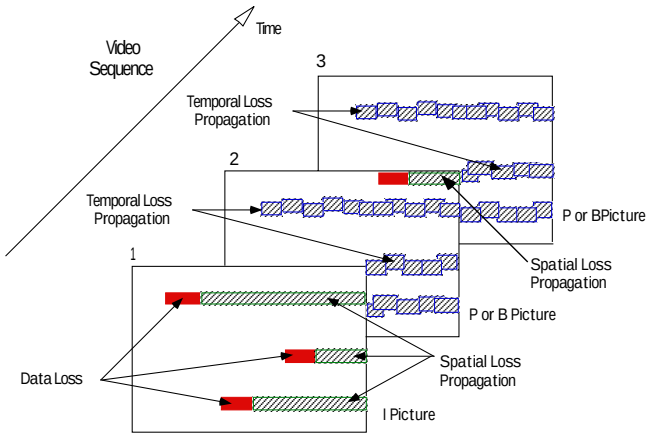


Figure 6: *Data loss propagation in MPEG-2 video streams.*

The impact of the loss of syntactic data is, in general, more important and more difficult to recover than the loss of semantic information. For instance when a frame header is lost, the entire frame is skipped since the decoder is not able to detect its beginning. If the skipped frame is a reference picture, the temporal error propagation may greatly reduce the perceptual quality. So, when a header is lost, in general, the whole information it precedes is skipped. Some headers are thus more important than others.

## 4.2 How to Improve the QoS

Algorithms aiming at improving the quality of service may be implemented in both the application layer (encoder and decoder) and the network protocol stack.

For the remainder of this paper, we focus the analysis on data loss only. Indeed, delay jitter can be solved via buffering techniques. However some delayed packets may fail to meet their respective decoding schedule and are therefore lost.

### 4.2.1 Encoder

We present three major algorithms. Two of these, syntactic protection and layered coding, reduce loss sensitivity. Another algorithm, adaptive quantization, increases video quality without modifying the average bit rate.

**Adaptive quantization:** Adaptive quantization aims at spatially and temporally uniformizing the coding noise by adjusting the $MQUANT$ value on a macroblock basis [13]. Therefore, the same video quality may be reached at a lower average bit rate [14].

**Syntactic protection:** Loss sensitivity may be dramatically reduced by properly structuring video data and headers. For instance, slice headers and intra-coded macroblocks act as resynchronization points for spatial and temporal propagation of errors. Algorithms have been proposed in the literature to optimize the tradeoff between sensitivity to loss and amount of headers [15, 16].

**Layered coding:** Layered (or hierarchical) video coding means that the signal is separated into components of differing visual importance [17, 18, 19, 20, 21]. The idea is that error protection and quality provisioning could be selected for the properties of each individual layer rather than for the entire bitstream. For instance, error-control codes of varying strengths and rates could be used independently for the layers to reach a suitable level of error recovery at a lower overhead compared to a protection of all data by a single code.

For a network that supports separate service classes, it is possible to transfer each layer with a service level commensurate with its importance. Vital layers may thus be transferred in a class with guaranteed quality, while a signal layer that enhances the quality could be sent "best effort". The hope is that the overall transfer is more economical than if the transfer was done over one channel with a service quality determined by the most sensitive part of the information. Layering therefore assumes that a set of connections with different capacities and qualities of service is cheaper than one connection for the aggregate stream.

Layered coding is also useful when a specific target bit rate or quality level cannot be stated *a priori* for the transfer. By layering, the sender can provide a range of bit rates and qualities in one and the same encoding of the information, and the particular point in that range can be chosen dynamically. It can, for instance, be beneficial for stored programs and for multicast [22, 23, 24].

The MPEG-2 video coding provides several layering options. The SNR scalability, and the related non-standardized data partitioning, are suitable for error-control purposes. SNR scalability is a multiple description technique that uses both a coarse and a fine quantizer for the DCT coefficients [25]. The basic layer contains the coarsely quantized coefficients. The upper layer contains the needed refinements to yield the coefficients according to the fine quantizer. It is to be noted that MPEG-2 further provides spatial and temporal scalabilities.

### 4.2.2 Network Adaptation

**Forward error correction:** From the transmission standpoint, video delivery can be improved mainly by providing error correction mechanisms. Two major techniques exist: retransmission and forward error correction (FEC). Retransmission requires the receiver to inform the sender about the data that must be repeated. It has the advantage of error-free delivery, but at a large cost in unpredictable delay. FEC means that redundancy is added to the data so that the receiver can recover from losses or errors without any further intervention from the sender.

Considering the delay requirements for interactive video and real-time applications in general, FEC is more appropriate than retransmission because it meets the timing constraints.

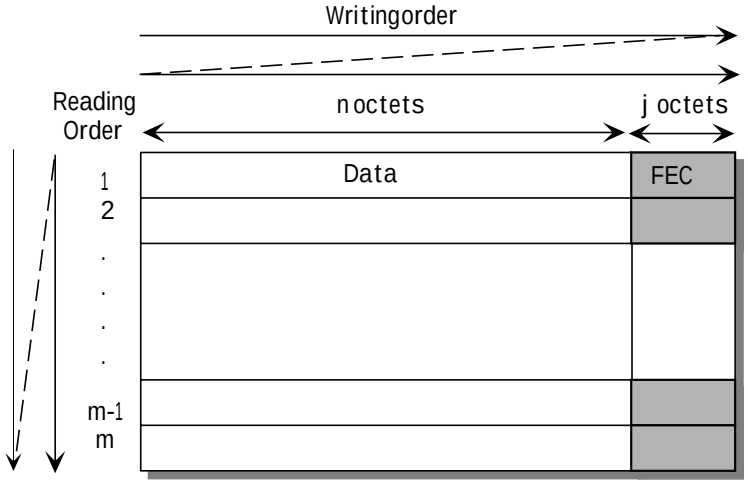


Figure 7: *Octet-based interleaver.*

Coding theory defines an error as a corrupted symbol with an unknown value in an unknown location. Similarly, an erasure is defined as a symbol with an unknown value, but in a known location. When the error location is known, the error correcting power of a given code is doubled. Popular codes for FEC are the *Reed-Solomon codes* (RSC) which can correct bit errors as well as erasures (data loss). A new set of codes called *burst erasure codes* have been derived from the RSC, which are simple to implement [26] and are able to correct only erasures. Burst erasure codes rely on the fact that the basic transmission unit is known. For ATM, erasures are of known size and are located at cell boundaries. The position is known if a sequence number is inserted in all cells. However, only AAL1 implements this feature (see Fig. 7). If AAL5 is used, the coding must be applied on packets in the upper layers.

The Internet community is also considering the use of FEC for video transport over RTP.

To reduce the coding overhead, selective protection may be used. As described in Sec. 4.1.2, the perceptual impact of data loss depends on the type of information lost. It is therefore possible to reduce the overhead by selectively protecting the most important data [27, 28].

### 4.2.3 Decoder

**Error concealment techniques:** Error concealment is used to reduce the impact of data loss on the visual information. These algorithms include, for example, spatial interpolation, temporal interpolation and early resynchronization (see Fig. 8). The MPEG-2 standard proposes an elementary error concealment algorithm based on motion compensation. It estimates the vectors for the lost macroblock by using the motion vectors of neighbouring macroblocks in the affected picture (provided these have not also been lost). This improves the concealment of moving picture areas. There is however an obvious problem with lost macroblocks whose neighbours are intra-coded, because there are ordinarily no motion vectors associated with them. To get around this problem, the encoding can include motion vectors also for intra-coded macroblocks [6].

Error concealment may, in general, efficiently decrease the visibility of data loss. However, severe data loss may still lead to annoying degradations in the decoded video quality.

[6] Some MPEG-2 encoder chips automatically produce concealment motion vectors for all macroblocks.

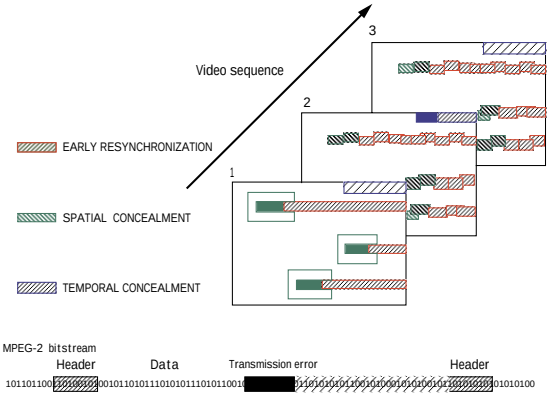


Figure 8: *Error concealment techniques.*

## 4.3 How to Measure the QoS

A quality metric often used for audio-visual signals is the peak signal-to-noise ratio (PSNR). Many studies have shown that this metric is poorly correlated with human perception as it does not take visual masking into consideration. In other words, every errored pixel contributes to a decrease in PSNR even if the error cannot be perceived. Recent research has therefore addressed the issue of video quality assessment by means of metrics based on the properties of the human visual system. All these metrics fall into one of the following categories: (i) metrics based on a mathematical fit of a subjective rating function obtained by intensive psychovisual experiments and (ii) metrics relying on a model of the human visual system. An example of the former category is $\widehat{S}$ from ITS [29]. However, metrics belonging to the latter category usually perform better [30]. These include Sarnoff JND Vision Model [31], MPQM [32] and PDM [33].

In [31, 32, 33], spatio-temporal models of human vision were developed for the assessment of video coding quality. These three models are 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.

In the remainder of this paper, we present results by using the model developed in [32], and a computational quality metric built upon that model called the moving pictures quality metric (MPQM) [30]. This metric was proven to behave consistently with human judgments. First, it decomposes the original sequence and a distorted version of it into perceptual channels. A channel-based distortion measure is then computed while accounting for contrast sensitivity and masking. Finally, the data is pooled over all the channels to compute the quality rating which is then scaled from 1 to 5 [34] (see Fig. 9). This quality scale is used for subjective testing in the engineering community (see Tbl. 1).

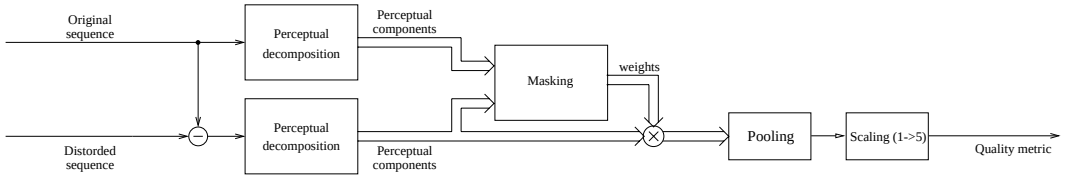


Figure 9: *Moving pictures quality metric (MPQM) block diagram*

| 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

# 5 Perceptual Impact of MPEG-2 Rate and Data Loss

The combined effect of the coding bit rate and the network impairments on the user-perceived quality is still not well understood. However these results are needed for the design and deployment of packet video services. One of the common misconceptions is that increasing the coder bit rate always enhances the perceived image quality.

In this section, we study how the video quality is affected by the MQUANT value (MPEG-2 encoding parameter) and the packet loss ratio measured while transmitting MPEG-2 streams over ATM- or IP-based networks [35] (see Sec. 3). We first analyze how the user-perceived quality is related to the average encoding bit rate for VBR MPEG-2 video. We then show why simple distortion metrics may lead to inconsistent interpretations. Next, we analyze, for a given coder setup, the effect of packet loss on the user-level quality. Finally, when studying the joint impact of coding bit rate and packet loss, the quality exhibits one optimal coding rate for a given packet loss ratio.

## 5.1 Experimental setup

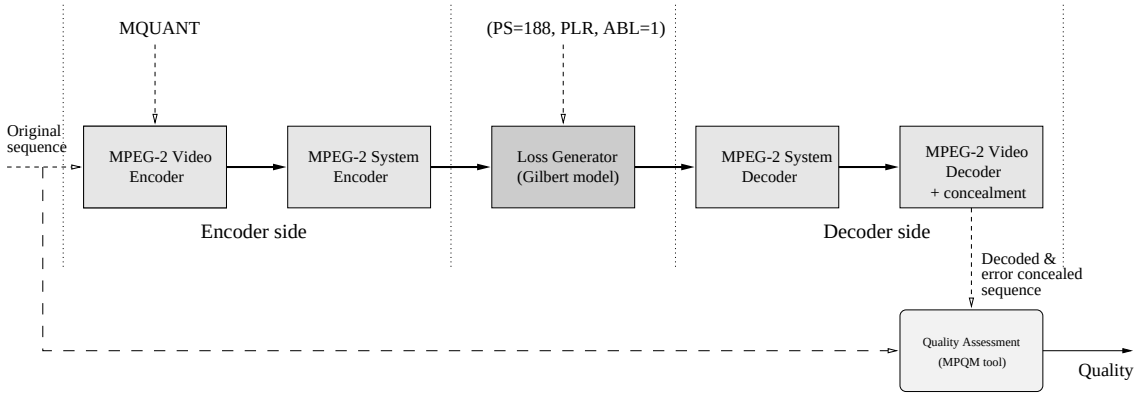


Figure 10: *Experimental testbed.*

The experimental testbed is composed of four parts (see Fig. 10):

- Our MPEG-2 software encoder consisted of an open-loop VBR (OL-VBR) TM5 video encoder [36] and a transport stream encoder. Three 100 frame-long (Football, News, and Barcelona) and one 1000 frame-long (ski) sequences conforming to the ITU-R 601 format were used. All these sequences differ in terms of spatial and temporal complexities. They

were encoded in an OL-VBR mode, as interlaced video, with a structure of 11 images between each pair of I-pictures and 2 B-pictures between every reference picture. The following MQUANT values were used: 6, 10, 16, 20, 28, 32, 36, 40 and 48. Motion vectors were generated for all intra-coded macroblocks. The introduction of these extra motion vectors do not affect the OL-VBR encoding quality. Before being transmitted, each MPEG-2 video bitstream was 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.

- A model-based data loss generator was used to simulate packet network losses. For this purpose, we used a two-state Markovian model (Gilbert model [37], see Fig. 11).

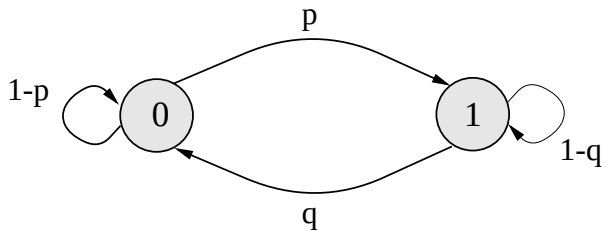


Figure 11: *Two-state Markov chain: Gilbert model.*

States 0 and 1 respectively correspond to the correct reception and loss of a packet. The transition rates between the states control the lengths of the bursts of errors. Hence, there are three parameters to be controlled: the packet size (PS), the packet loss ratio ($PLR = \frac{p}{p+q}$) and the average number of packets lost in a burst of errors ($ABL = \frac{1}{q}$). In our simulations, we imposed a non-bursty TS packets loss process ($ABL = 1$, $PS = 188$ bytes) and varied the packet loss ratio between $10^{-2}$ and $10^{-7}$.

- Video quality was evaluated by means of the MPQM tool presented in Sec. 4.3. The per-frame quality values given by the MPQM tool were gathered together by means of a Minkowski summation [32] (i.e. weighted average). This summation, along with the correct exponent, gives a result that is more accurate than the simple average quality, which is too optimistic [33] (i.e. the subjective quality evaluated over a set of frames is lower than the average of the per-frame quality values).

- The last part is an MPEG-2 software decoder, which constitutes both a TS decoder and a video decoder. The video decoder provides the motion compensated concealment technique

briefly explained in Sec 4.2.3. This technique was chosen for different reasons. The first is to be consistent with real implementations. The second is to be able to perform the perceptual measurements. Indeed, the vision model currently developed and the derived metrics have been tested for errors below what is called the *suprathreshold* [7]. Therefore, a problem occurs when the degradation due to data loss generates highly visible artifacts (i.e. holes) in the sequence since these errors may be above this suprathreshold. By using error concealment techniques, most of the artifacts may be considered as being below the suprathreshold of vision, making the perceptual measure accurate.

## 5.2 MPEG-2 VBR Encoding Impact on Video Quality

We first study how the encoding process influences video quality. Figures 12 and 13 show how the quality is affected by the MQUANT parameter when measured by the PSNR metric and the MPQM. While the PSNR versus MQUANT curve may be represented by a decreasing exponential [38], it is to be noted that the MPQM metric exhibits a linear relationship with MQUANT. Such an important behavior has been verified for all of the four sequences. The same characteristic has recently been observed through users' subjective evaluations [39]. Moreover, the ITS metric [29] also shows the same characteristic but on a smaller range of MQUANT values.

The slopes of the lines are directly related to the complexity of the sequence: the higher the encoding complexity, the higher the slope. For instance, the video sequence "News" is a *Head and Shoulder* type of sequence and does not contain any high spatio-temporal complexities. The absolute value of the slope is therefore smaller.

We now have an idea of how the encoding quality depends on the value of MQUANT. Then, we need to study how the average output bit rate is affected by the MQUANT. In [38], it has been demonstrated that $\overline{R} = c \times MQUANT^{-d}$ is a good approximation of the relation between the quantizer scale factor and the average bit rate. $\overline{R}$ represents the average output bit rate and the parameters $c$ and $d$ are related to the encoding complexity of the scene. This behavior is illustrated in Fig. 14. The parameters $c$ and $d$ have been obtained by minimizing the mean square error. Finally, by combining the equation above with the linear relationship of MPQM and MQUANT, a model describing how the video quality behaves according to the average encoding bit rate may be

[7] Two to three times above the threshold of vision which corresponds to the threshold of visibility of the noise

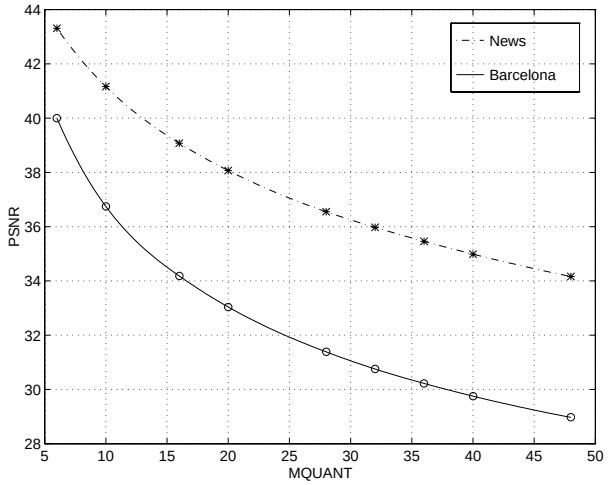


Figure 12: *PSNR versus quantizer scale factor for 2 different scenes.*

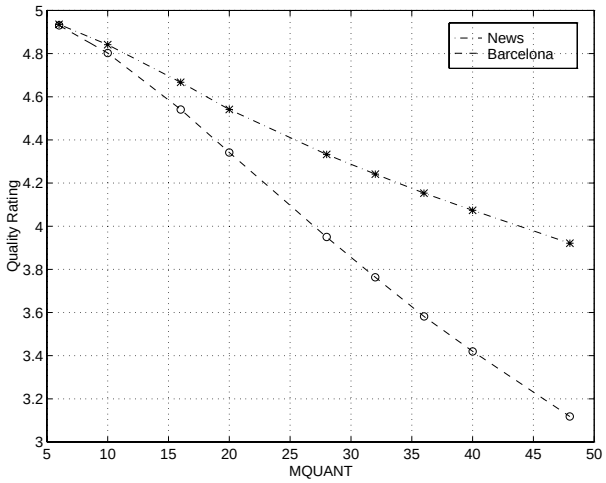


Figure 13: *MPQM versus quantizer scale factor for 2 different scenes.*

derived:

$$Q = a.\left(\frac{\overline{R}}{c}\right)^{-\frac{1}{d}} + b \tag{1}$$

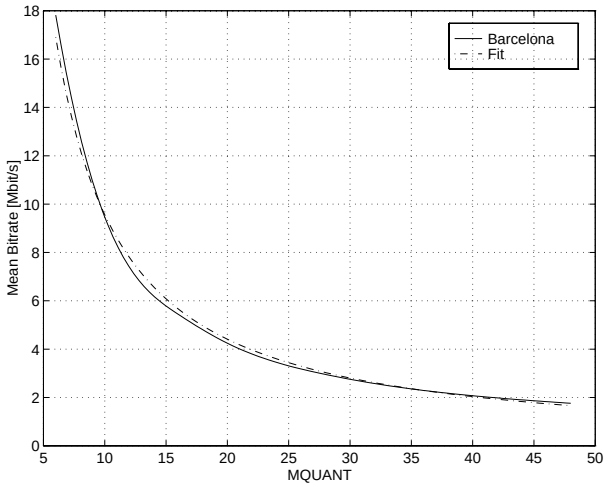


Figure 14: *Average output encoding bit rate versus quantizer scale factor (MQUANT) for Barcelona. Fitting parameters: (c=124.7615, d=1.1156)*

The three main parameters $a$, $c$ and $d$ are somehow related to the spatio-temporal complexity of the sequence ($b$ is always close to 5.0 which is the maximal quality). Results from computer simulations and the corresponding curve given by the equation above are plotted in Fig. 15.

The graph illustrates that the perceptual quality saturates at high bit rates. Increasing the bit

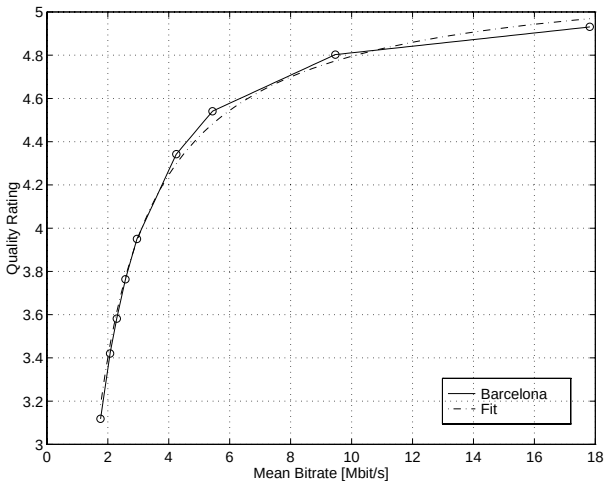


Figure 15: *MPQM video quality versus average output encoding bit rate for the "Barcelona" sequence. Fitting parameters for Eq. 1: (a=-0.0448, b=5.2254, c=124.7615, d=1.1156)*

rate may at some point result in a waste of bandwidth since the end user does not perceive an improvement in quality. Such a quality saturation is however not captured well by the PSNR.

The average bit rate after which the quality does not increase significantly may be reduced by means of an adaptive quantization scheme (see Sec. 4.2.1).

## 5.3 Impact of Data Loss on Video Quality

Up to this point, we did not consider the degradation of video streams by network losses. Figure 16 illustrates how the video quality is affected by uniformly distributed TS packet losses over a 1000-frame long MPEG-2 transport stream [8]. It is shown that, on a semi-logarithmic scale and for a given MQUANT (average bit rate), the video quality first remains constant with the PLR. This constant value corresponds to the encoding quality. Then, beyond a certain PLR, the perceptual quality drops fast.

The relation between video quality and PLR may therefore be represented as $Q = e + f \times PLR$ where $e$ corresponds to the encoding quality and $f$ depends on both the complexity of the sequence and the average bit rate [35]. In other words, for a given sequence and a fixed MQUANT, the video quality, averaged over the whole sequence decreases linearly with the PLR. This behavior is also captured by the $\widehat{S}$ metric from ITS.

It is to be noted that the PLR value after which the quality quickly drops may be increased by

[8] Every simulation has been run five times with different packet loss patterns

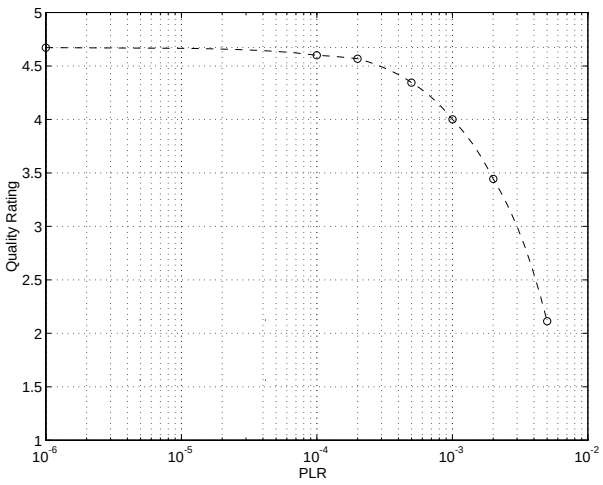


Figure 16: *MPQM versus PLR (ABL=1, PS=188) for MQUANT=28 using the "Ski" sequence.*

means of some techniques presented in Sec. 4.2 (i.e. syntactic protection, layered coding, FEC and error concealment techniques).

### 5.4 Joint Impact Analysis

In this section, we analyze how the PLR and the average encoding bit rate (i.e. packet rate) are intimately related to each other when considering their impact on video quality in MPEG-2 video communications. The results presented below still apply to MPEG-2 CBR transmission [40].

We have already demonstrated how the perceived video quality saturates as the encoding bit rate increases in an error-free environment. Moreover, for a given encoding bit rate, we have shown that the video quality dropped dramatically after a certain PLR value. We now show that the higher the bit rate, the lower the PLR after which the video quality drops, and inversely. The PLR is indeed defined as the number of lost packets per time unit divided by the number of packets transmitted during that time unit. In MPEG-2 video delivery, the packet size does not depend on the encoding bit rate [8, 12]. Therefore, the higher the encoding bit rate, the higher the number of packets transmitted per time unit. Thus, for a given PLR, the higher the encoding bit rate, the higher the number of packets lost per time unit [9].

Therefore, the relation between quality and the encoding bit rate for a given non-zero PLR should somehow exhibit an optimal value. Such a behavior is illustrated in Fig. 17 using the "Ski" sequence. It is indeed shown that the video quality first increases with the average bit rate and

[9]The number of video frames transmitted per time unit is independent of the encoding bit rate

then decreases after around 6 Mbps. This optimal average bit rate directly depends on the content type of the sequence.

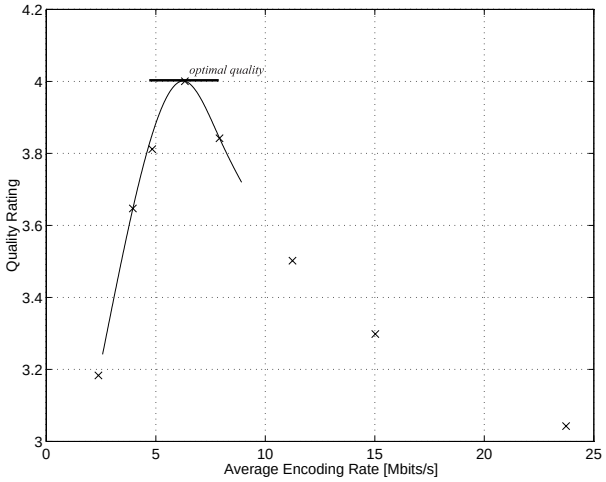


Figure 17: *MPQM versus average encoding bit rate for* $PLR = 10^{-3}$ *for the "Ski" sequence.*

Moreover, in most packet networks where losses occur due to congestion (i.e. buffer overflows), the PLR may be somehow related to the rate sent throughout the network. Therefore, when increasing the encoding bit rate, the PLR may increase and cause the number of lost packets per time unit to increase even more, resulting in very annoying video degradations.

Hence, image quality cannot be improved by only acting on the MPEG-2 coding bit rate: increasing the bit rate above a certain threshold results in quality degradations. For a given packet loss ratio, there is a quality-optimal coding rate that has to be found. Although the relationship between coding bit rate, packet loss ratio and user-level quality is intrinsically complex, it can be characterized by a simple expression and a set of parameters [35].

# 6 Conclusion

Because of the increasing availability of Internet and ATM networks, packet video is expected to become common in the coming years. It is therefore important to fully understand the parameters that may affect the quality of the image delivered to the end-user, and how to cope with these impairments.

In this tutorial, we have shown how the quality of service can be assessed from the perspective of the end-user. We have also shown how this assessment technique can be used to analyze the

impact of the encoding bit rate and the data loss. Based on these results, we have explained how the optimal bit rate can be found over a lossy packet network.

There are, however, several issues that were not possible to include in this tutorial. They encompass synchronization of audio and video, multicast, as well as video over wireless networks. All these topics are currently under intense investigation by the research community.

## 7 Acknowledgments

The authors are grateful to Erin Farr from IBM Corporation, Ismail Dalgic from 3Com Corporation and Pascal Frossard from the Signal Processing Laboratory at EPFL for their relevant comments and remarks on the paper.

## References

[1] G. Karlsson, "Asynchronous Transfer of Video", *IEEE Communications Magazine*, vol. 34, pp. 118–126, August 1996.

[2] U. Reimers, "Guideline for the Use of DVB Specifications and Standards", Technical Report TR 101 200, The DVB Project, June 1998.

[3] ISO/IEC JTC 1, *Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - Part 1, 2 and 3*, ISO/IEC JTC 1, 1996.

[4] B. G. Haskell, A. Puri and A. N. Netravali, *Digital Video: an Introduction to MPEG-2*, Digital Multimedia Standards Series. Chapman and Hall, 1997.

[5] G. Karlsson and G. Djuknic, "On the Efficiency of Statistical Bit Rate Service for Video", *in IFIP International Conference on Performance of Information and Communications Systems*, Lund, Sweden, May 25-28 1998.

[6] ITU-T Study Group 13, editor, *Recommendation I.363.1 B-ISDN ATM Adaptation Layer Specification, Type 1*, ITU-T, Geneva, Aug. 1996.

[7] ITU-T Study Group 13, editor, *Recommendation I.363.5 B-ISDN ATM Adaptation Layer Specification, Type 5*, ITU-T, Geneva, Aug. 1996.

[8] ITU-T, editor, *Recommendation H.222.1 Multimedia Multiplex and Synchronization for Audiovisual Communications in ATM Environments*, ITU-T, Geneva, Mar. 1998.

[9] S. Gringeri, B. Khasnabish, A. Lewis, K. Shuaib, R. Egorov and B. Bash, "Transmission of MPEG-2 Video Streams over ATM", *IEEE Multimedia*, vol. 5, pp. 58–71, January-March 1998.

[10] P. Almquist, "Type of service in the internet protocol", Technical Report RFC 1349, IETF, July 1992.

[11] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", Technical Report RFC 1889, IETF, 1996.

[12] V. Goyal D. Hoffman, G. Fernando and R. Civanlar, "RTP Payload Format for MPEG1/MPEG2 Video", Technical Report RFC 2250, IETF, january 1998.

[13] J.L. Mitchell, W.B. Pennebaker, C.E. Fogg and D. LeGall, *MPEG Video Compression Standard*, Chapman and Hall, 1997.

[14] O. Verscheure and C. van den Branden Lambrecht, "Adaptive Quantization Using a Perceptual Visibility Predictor", *in IEEE ICIP*, 1997.

[15] P. Frossard and O. Verscheure, "AMIS: Adaptive MPEG-2 Information Structuring", *in SPIE International Symposium on Voice, Video, and Data Communications*, Boston, USA, November 1998.

[16] O.A. Aho and J. Juhola, "Error resilience techniques for mpeg-2 compressed video signal", *in IEE International Broadcasting Convention (IBC)*, pp. 327–332, January 1994.

[17] G. Karlsson and M. Vetterli, "Sub-band Coding of Video for Packet Networks", *Optical Engineering*, vol. 27, pp. 574–586, July 1988.

[18] M. Ghanbari, "Two-layer Coding of Video Signals for VBR Networks", *IEEE Journal on Selected Areas in Communications*, vol. 7, pp. 771–781, June 1989.

[19] T. Chiang and D. Anastassiou, "Hierarchical Coding of Digital Television", *IEEE Communications Magazine*, vol. 32, pp. 38–45, May 1994.

[20] E. Chang W. Tan and A. Zakhor, "Real Time Software Implementation of Scalable Video Codec", *in Proceedings of the International Conference on Image Processing*, Lausanne, Switzerland, September 1996.

[21] D. Taubman and A. Zakhor, "Multirate 3D Subband Coding of Video", *IEEE Transactions on Image Processing*, vol. 3, September 1994.

[22] S. McCanne, V. Jacobson, M. Vetterli, "Receiver-driven Layered Multicast", *ACM Computer Communication Review*, vol. 26, pp. 117–130, October 1996.

[23] N. Shacham, "Multipoint Communication by Hierarchically Encoded Data", *in INFOCOM*, pp. 2107–2114, 1992.

[24] M. Sudan and N. Shacham, "Gateway based Approach for Managing Multimedia Sessions over Heterogeneous Signaling Domains", *in INFOCOM*, 1997.

[25] A. A. El Gamal and T. M. Cover, "Achievable Rates for Multiple Descriptions", *IEEE Transactions on Information Theory*, vol. 28, pp. 851–857, November 1982.

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

[27] X. Adanez and O. Verscheure, "New Network Adaptation and ATM Adaptation Layers for Interactive MPEG-2 Video Communications: A Performance Study Based on Psychophysics", *Interoperable Communication Networks (ICON)*, vol. 1, pp. 145–178, January 1998.

[28] B. Lamparter, O. Böhrer, W. Effelsberg, and V. Turau, "Adaptable Forward Error Correction for Multimedia Data Streams", Technical report, University of Mannheim, September 1993.

[29] A. Webster, C. Jones, M. Pinson, S. Voran and S. Wolf, "An Objective Video Quality Assessment System Based on Human Perception", *in SPIE - Human Vision, Visual Processing and Digital Display*, vol. 1913, pp. 15–26, 1993.

[30] C. J. van den Branden Lambrecht and O. 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.

[31] J. Lubin and D. Fibush, *Sarnoff JND Vision Model*, T1A1.5/97-612, Contribution to T1 Standards project, August 1997.

[32] C. 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 1996.

[33] S. Winkler, "A Perceptual Distortion Metric for Digital Color Images", *in Proceedings of the International Conference on Image Processing*, October 1998.

[34] 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.

[35] O. Verscheure, P. Frossard and M. Hamdi, "MPEG-2 Video Services over Packet Networks: Joint Effect of Encoding Rate and Data Loss on User-Oriented QoS", *in 8th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, Cambridge, U.K., July 8-10 1998.

[36] C. Fogg, "mpeg2encode/mpeg2decode", *in MPEG Software Simulation Group*, 1996.

[37] J. W. Roberts, J. Guibert and A. Simonian, "Network Performance Considerations in the Design of a VBR Codec", *in Queuing Performance and Control in ATM*, pp. 77–82. J. W. Cohen and C. D. Pack, June 1991.

[38] S. Sakazawa, Y. Takishima, M. Wada and Y. Hatori, "Coding Control Scheme for a Multi-Encoder System", *in 7th International Workshop on Packet Video*, pp. 83–88, March 1996.

[39] K. Fukuda, N. Wakamiya, M. Murata and H. Miyahara, "On Flow Aggregation for Multicast Video Transport", *in 6th IFIP International Workshop on Quality of Service (IWQoS)*, May 1998.

[40] O. Verscheure, P. Frossard and M. Hamdi, "Joint Impact of MPEG-2 Encoding Rate and ATM Cell Losses on Video Quality", *in IEEE GLOBECOM*, November 1998.

# 8 Biographies

OLIVIER VERSCHEURE (Olivier.Verscheure@epfl.ch) is a Ph.D. candidate at the Institute for computer Communications and Applications (ICA) at the Swiss Federal Institute of Technology, Lausanne. He received his B.S. from the *Faculté Polytechnique de Mons*, Belgium and his M.S. from the Swiss Federal Institute of Technology, both in Electrical Engineering. He was a visiting researcher at the Hewlett-Packard Laboratories (Palo Alto, CA) during the summer of 1997. His research lies within the areas of video communications and vision science.

XAVIER GARCIA ADANEZ (Xavier.Garcia@adventis.ch) obtained his Ph.D. in communication systems at the Institute for computer Communications and Applications (ICA) in March 1998. He received his B.S. and M.S. in Electrical Engineering from the Swiss Federal Institute of Technology in Lausanne. His main topics of interest are related to video communications and Quality of Service; high speed reliable protocols for audiovisual communications and perceptual quality metrics as network performance tools.

Since August 1998 he is working as a consultant in Adventis Communications Engineering in Switzerland.

He is the Assistant Vice Chair for the European Activities of the ATM Forum User Group. He is also a member of the IEEE Communications Society.

GUNNAR KARLSSON (Gunnar.Karlsson@sics.se) is professor at the Department of Teleinformatics at the Royal Institute of Technology (KTH) in Stockholm, Sweden. Before joining KTH in 1998 he worked six years at the Swedish Institute of Computer Science and three years at the IBM Zurich Research Laboratory. He holds a Ph.D from Columbia University and a M.Sc. from Chalmers University of Technology. Gunnar's research interests are mainly in the areas of packet video and switch/router architecture.

JEAN-PIERRE HUBAUX (Jean-Pierre.Hubaux@epfl.ch) has been an associate professor at the Swiss Federal Institute of Technology - Lausanne since 1990; he is co-founder and co-director of the Institute for computer Communications and Applications (ICA, icawww.epfl.ch). His research activity is focused on service engineering, with a special emphasis on multimedia and security services. He has authored and co-authored more than 30 publications in this area and holds several patents.

He is currently on a sabbatical in the US. The first part of it was spent at the IBM T.J. Watson Research Center in Hawthorne (NY), during the Spring of 1998; it was focused on a convergence solution for telecom and datacom services. The second part is spent at the EECS Department of the University of California in Berkeley (CA), until January 1999.

Prior to this position, he spent 10 years in France with Alcatel, where he was involved in R&D activities, primarily in the area of switching systems architecture and software. He is a senior

member of IEEE.