

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

J.343.6

(11/2014)

SERIES J: CABLE NETWORKS AND TRANSMISSION
OF TELEVISION, SOUND PROGRAMME AND OTHER
MULTIMEDIA SIGNALS

Measurement of the quality of service - Part 3

**Hybrid-FR objective perceptual video quality
measurement for HDTV and multimedia
IP-based video services in the presence of a full
reference signal and non-encrypted bitstream
data**

Recommendation ITU-T J.343.6

ITU-T

Recommendation ITU-T J.343.6

Hybrid-FR objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a full reference signal and non-encrypted bitstream data

Summary

Recommendation ITU-T J.343.6 provides hybrid full reference (Hybrid-FR) objective perceptual video quality measurement methods for HDTV and multimedia when a full reference signal and non-encrypted bitstream data are available. The following list shows example applications that can use this Recommendation:

- potentially real-time, in-service quality monitoring at the headend;
- video television streams over cable/IPTV networks including those transmitted over the Internet using Internet protocol;
- video quality monitoring at the receiver when a full reference signal and non-encrypted bitstream data are available;
- video quality monitoring at measurement nodes located between point of transmission and point of reception when a full reference signal and non-encrypted bitstream data are available;
- quality measurement for monitoring of a transmission system that utilizes video compression and decompression techniques, either a single pass or a concatenation of such techniques;
- lab testing of video transmission systems.

This Recommendation includes test vectors, including video sequences, bitstream files and predicted objective model scores.

This Recommendation includes an electronic attachment containing test vectors, including video sequences, bitstream files and predicted objective model scores.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T J.343.6	2014-11-29	9	11.1002/1000/12321

* To access the Recommendation, type the URL <http://handle.itu.int/> in the address field of your web browser, followed by the Recommendation's unique ID. For example, <http://handle.itu.int/11.1002/1000/11830-en>.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at <http://www.itu.int/ITU-T/ipr/>.

© ITU 2015

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

Table of Contents

		Page
1	Scope.....	1
	1.1 Applications.....	2
	1.2 Limitations.....	2
2	References.....	3
3	Definitions	3
	3.1 Terms defined elsewhere	3
	3.2 Terms defined in this Recommendation.....	3
4	Abbreviations and acronyms	3
5	Conventions	4
6	Performance metrics	4
7	Description of the hybrid reduced-reference methodology	4
8	Models	5
	Annex A – Description of video quality measure PEVQ-S.....	6
	A.1 Introduction	6
	A.2 Description of the PEVQ-S algorithm.....	6
	Annex B – YHyFR.....	38
	B.1 Introduction	38
	B.2 Hybrid-FR VQM computation	38
	Bibliography.....	43

Electronic attachment: Test vectors, including video sequences, bitstream files and predicted objective model scores.

Recommendation ITU-T J.343.6

Hybrid-FR objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a full reference signal and non-encrypted bitstream data

1 Scope

This Recommendation¹ describes algorithmic models for measuring the visual quality of IP-based video services.

The models are hybrid full reference (Hybrid-FR) models, which use the unimpaired reference video signal, non-encrypted bitstream data and video image data captured at the video player.

As output the models provide an estimate of visual quality on the [1,5] difference mean opinion score (DMOS) scale, derived from five-point absolute category rating (ACR) as in [ITU-T P.910]. The models address low-resolution (VGA/WVGA) application areas, including services such as mobile TV, as well as high-resolution (HD) application areas, including services such as IPTV.

This Recommendation is to be used with videos encoded using [ITU-T H.264] and media payload encapsulated in RTP/UDP/IP packets for the low resolution and encapsulated in MPEG-TS/RTP/UDP/IP for the high resolution.

The models in this Recommendation measure the visual effect of spatial and temporal degradations as a result of video coding, erroneous transmission or video rescaling. The models may be used for applications such as to monitor the quality of deployed networks to ensure their operational readiness or to benchmark service quality. The models in this Recommendation can also be used for lab testing of video transmission systems.

The models identified in this Recommendation have limited precision. Therefore, directly comparing model results can be misleading. The accuracy of models has to be understood and taken into account (e.g., using [ITU-T J.149]).

The validation test material consisted of video encoded using different implementations of [ITU-T H.264]. It included media transmitted over wired and wireless networks, such as WIFI and 3G mobile networks. The transmission impairments included error conditions such as dropped packets, packet delay, both from simulations and from transmission over commercially operated networks.

The following source reference channel (SRC) conditions were included in the validation test:

- 1080i 60 Hz (29.97 fps);
- 1080p (25 fps);
- 1080i 50 Hz (25 fps);
- 1080p (29.97 fps);
- SRC duration: HD: 10s, VGA/WVGA: 10 s or 15 s (rebuffering);
- VGA at 25 and 30 fps;
- WVGA at 25 and 30 fps.

¹ This Recommendation includes an electronic attachment containing test vectors, including video sequences, bitstream files and predicted objective model scores.

The following HRC conditions were included in the validation test for each resolution:

Test factors
Video resolution: 1920 × 1080 interlaced and progressive
Video frame rates 29.97 and 25 frames per second
Video bitrates: 1 to 30 Mbit/s (HD), 100 kbit/s to 3 Mbit/s (VGA/WVGA)
Temporal frame freezing (pausing with skipping) of up to 50% of video duration
Transmission errors with packet loss
Rebuffering (VGQ/WVGA only): up to 50% of SRC
Coding technologies
[ITU-T H.264]/AVC (MPEG-4 Part 10)
Tandem coding

1.1 Applications

The applications for the estimation model described in this Recommendation include, but are not limited to:

- potentially real-time, in-service quality monitoring at the headend;
- video television streams over cable/IPTV networks including those transmitted over the Internet using Internet protocol;
- video quality monitoring at the receiver when a copy of the unimpaired reference video and non-encrypted bitstream data are available;
- video quality monitoring at measurement nodes located between point of transmission and point of reception when a copy of the unimpaired reference video and non-encrypted bitstream data are available;
- quality measurement for monitoring of transmission systems that utilize video compression and decompression techniques, including concatenations of such techniques;
- lab testing of video transmission systems.

1.2 Limitations

The video quality estimation models described in this Recommendation cannot be used to fully replace subjective testing.

When frame freezing was present, the test conditions had frame freezing durations up to 50% of SRC duration. The models in this Recommendation were validated for measuring video quality in a re-buffering condition (i.e., video that has a steadily increasing delay or freezing without skipping) only for VGA/WVGA. The models were not tested on other frame rates than those used in TV systems (i.e., 29.97 fps and 25 fps, in interlaced or progressive mode).

If forward error correction techniques are employed, the models in this Recommendation may not be used.

It is important that no additional transmission errors occur between the collection point of the bitstream data and the capture point of the processed video sequence (PVS).

It should be noted that in case of new coding and transmission technologies producing artifacts, which were not included in this evaluation, the objective model may produce erroneous results. Here, a subjective evaluation is required.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

- [ITU-T H.264] Recommendation ITU-T H.264 (2014), *Advanced video coding for generic audiovisual services*.
- [ITU-T J.149] Recommendation ITU-T J.149 (2004), *Method for specifying accuracy and cross-calibration of Video Quality Metrics (VQM)*.
- [ITU-T J.342] Recommendation ITU-T J.342 (2011), *Objective multimedia video quality measurement of HDTV for digital cable television in the presence of a reduced reference signal*.
- [ITU-T J.343] Recommendation ITU-T J.343 (2014), *Hybrid perceptual bitstream models for objective video quality measurements*.
- [ITU-T P.910] Recommendation ITU-T P.910 (2008), *Subjective video quality assessment methods for multimedia applications*.
- [ITU-R BT.601] Recommendation ITU-R BT.601 (2011), *Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following term defined elsewhere:

3.1.1 hybrid full reference model [ITU-T J.343]: An objective video quality model that predicts subjective quality using the reference video, the decoded video frames, packet headers, and the video payload. Such models cannot analyse encrypted video.

3.2 Terms defined in this Recommendation

None.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

CODEC	Coder-Decoder
DMOS	Difference Mean Opinion Score
DTS	Decoding Time Stamp
FR	Full Reference
HRC	Hypothetical Reference Circuit
HVS	Human Visual System
Hybrid-FR	Hybrid Full Reference
LUT	Look-Up Table

MOS	Mean Opinion Score
MPEG	Moving Picture Experts Group
MSE	Mean Square Error
PES	Packetized Elementary bitStream
PEVQ-S	Perceptual Evaluation of Video Quality for Streaming
PID	Packet Identifier Field
PTS	Presentation Time Stamp
PVS	Processed Video Sequence
ROI	Region Of Interest
SEQ	Sequence index
SRC	Source Reference Channel (or Circuit)
TS	Transport Stream
VQEG	Video Quality Experts Group
VQM	Video Quality Metrics

5 Conventions

None.

6 Performance metrics

A summary of this and other hybrid models may be found in [ITU-T J.343]. See [b-VQEG Hybrid] for the complete analysis of the models included in this Recommendation.

7 Description of the hybrid reduced-reference methodology

This Recommendation specifies objective video quality measurement methods that use both processed video sequences and bitstream data. The bitstream data may be provided in the forms of elementary bitstream (ES), packetized elementary bitstream (PES) or packet video (Figure 1).

In addition, the Hybrid-FR models also use reference video sequences. Figure 2 shows a Hybrid-FR model. The Hybrid-FR model need the source video sequence (SRC).

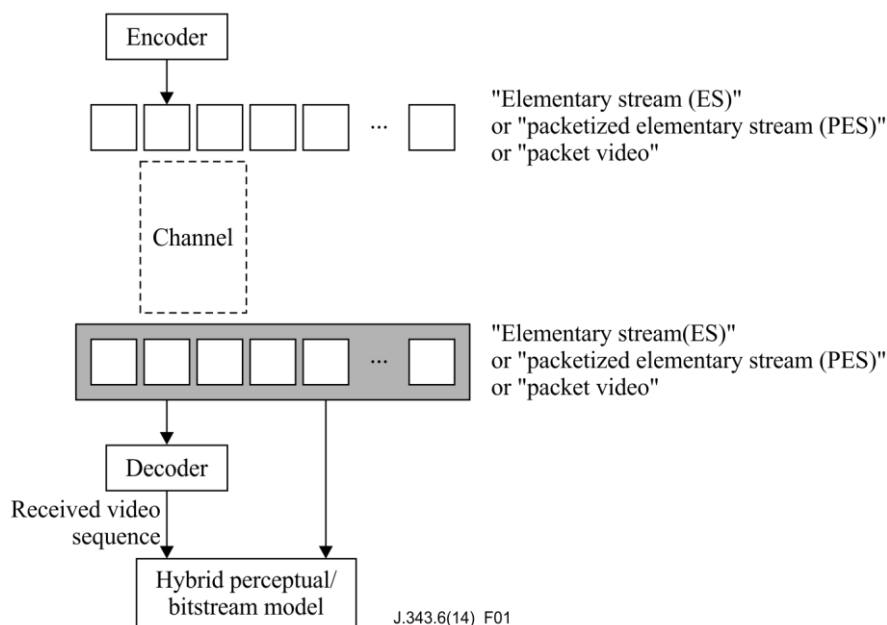
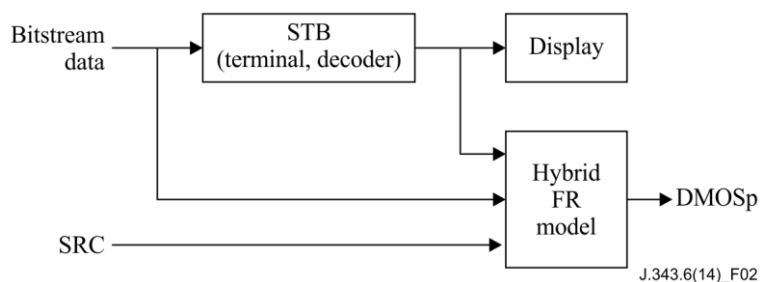


Figure 1 – Block-diagram depicts the core concept of hybrid perceptual bitstream models



DMOSp: predicted DMOS by the model

Figure 2 – Block-diagram depicts the Hybrid-FR model

8 Models

Annexes A and B contain full disclosures of all models included in this Recommendation. Those models are perceptual evaluation of video quality for streaming (PEVQ-S) and YHyFR.

Annex A

Description of video quality measure PEVQ-S

(This annex forms an integral part of this Recommendation.)

A.1 Introduction

PEVQ-S is a very robust model that is designed to predict the effects of transmission impairments on the video quality as perceived by a human subject. Its main targets are mobile applications and multimedia applications. The key features of PEVQ-S are:

- temporal alignment of the input sequences based on multi-dimensional feature correlation analysis with limits that reach far beyond those tested by VQEG, especially with regard to the amount of time clipping, frame freezing and frame skipping which can be handled;
- full-frame spatial alignment;
- colour alignment algorithm based on cumulative histograms;
- detection and perceptually correct weighting of frame freezes and frame skips;
- perceptual estimation of degradations;
- degradation caused by packet loss and bit errors in the transmission stream;
- handling of encrypted and non-encrypted bitstreams.

Only five indicators are used to detect the video quality. Those indicators are motivated by the human visual system (HVS). Perceptual masking properties of the HVS are modelled at several stages of the algorithm. These indicators are integrated using a sophisticated spatial and temporal integration algorithm.

In its first stage the algorithm collects information required for the alignment, concerning monochrome and frozen or skipped frames. In the second step the frames are synchronized in the temporal, spatial and colour domain, followed by comparison for visual differences in the luminance as well as in the chrominance domain and taking masking and motion effects into account. In addition to the image comparison, the transmitted bitstream is analysed for transmission errors. The above steps result in a set of indicators where each indicator describes a certain quality aspect.

The last step is finally the integration of the indicators by non-linear functions in order to derive the final mean opinion score (MOS).

Due to the low number of indicators and the resulting low degree of freedom the model can hardly be over trained and is very robust. PEVQ-S can be efficiently implemented without sacrificing the accuracy.

A.2 Description of the PEVQ-S algorithm

A.2.1 Overview

The basic idea of the PEVQ-S model is given in Figure A.1. Since PEVQ-S is a full reference hybrid algorithm, it requires three input signals. The first must be the undistorted source reference channel (SRC), S, the second must be the PVS, P, and third must be the collected bitstream file. It is assumed that the PVS is the result of the processed SRC through an HRC, and the bitstream file is collected at the player side.

Both sequences must have the same resolution (e.g., VGA, HD 1080p, HD 1080i). The signals should follow the YCbCr colour representation, as defined in [ITU-R BT.601].

The data storage for the images is planar oriented, no down sampling is used (sampling 4:4:4). Plane 1 is the Y component, Plane 2 the Cb component and Plane 3 the Cr component. The pixel values are stored as floating-point data.

Within the analysis two main modules can be found. A bitstream analysis module and a pixel based analysis module. Within the bitstream analysis module the bitstream is analysed for packet loss resulting in one indicator representative for additional temporal and spatial distortions.

Within the pixel-based analysis module uncompressed reference and degraded video sequences are used for detailed distortion analysis. The outcome of this analysis is indicators describing the severity of distortions in the temporal luminance and chrominance domain.

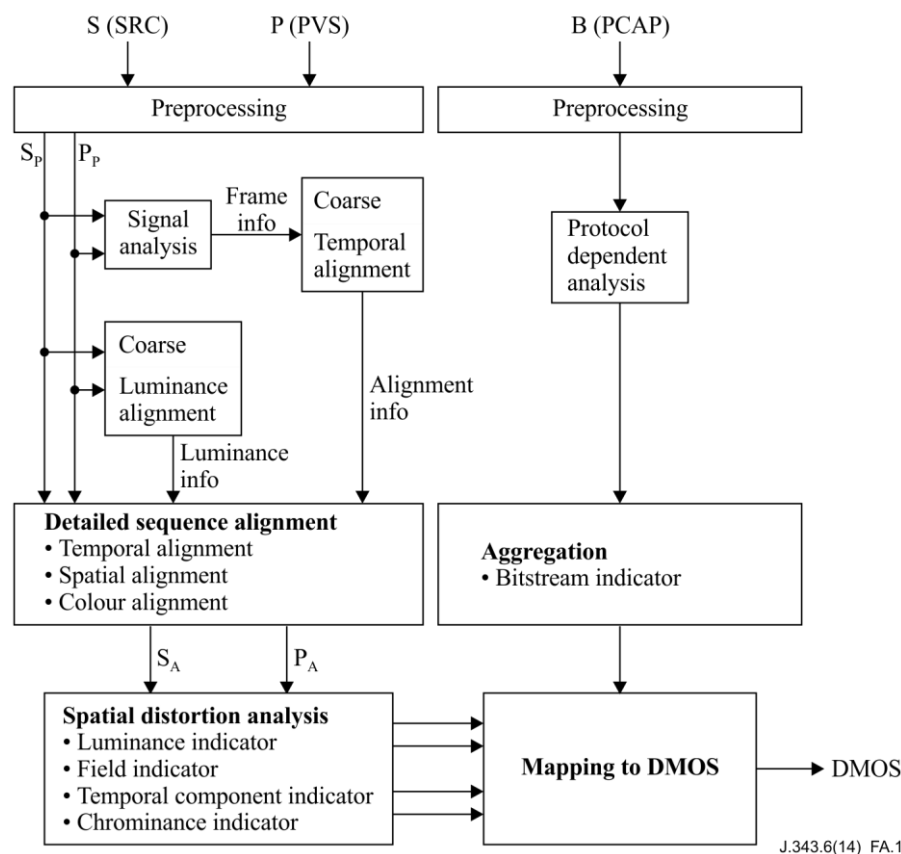


Figure A.1 – Overview of PEVQ-S

In the last step 'Mapping to DMOS' in Figure A.1 of PEVQ-S the five indicators that were derived above are weighted by logistic functions and combined to form the final PEVQ-S score, which correlates highly with a MOS obtained from the subjective tests.

A.2.2 Bitstream analysis

The task of the bitstream analysis is to produce an indicator that reflects the distortion in the video signal caused by packet loss (PL) during the transmission. This indicator is used within the OPTICOM model, together with the indicators from the pixel analysis, to predict the overall video quality of a given video stream.

The input into the bitstream analysis module is the bitstream B, which first is segmented in a list of packets $P(i)$, $i \in [0..I - 1]$, where I denotes the total number of packets arrived at the measurement point. This means the arrival times of the packets increase with increasing i .

The bitstream analysis module supports [ITU-T H.264] as video codec and two different protocol stacks:

- protocol stack S_1 : Ethernet (ETH) / IPv4 / UDP / RTP
- protocol stack S_2 : Ethernet (ETH) / IPv4 / UDP / RTP / MPEG2-TS

This means the packets $P(i)$ containing the video data have the structure shown in Figure A.2 (a) and (b) for stack S_1 and S_2 respectively.

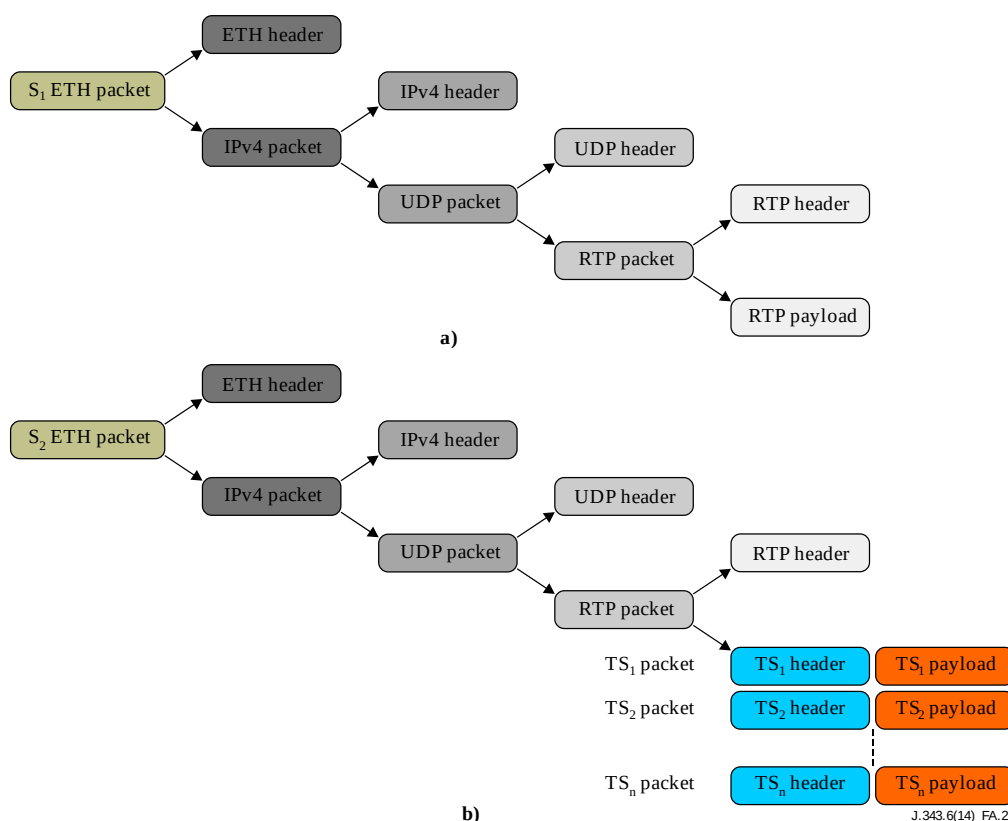


Figure A.2 – Payload structure with the Ethernet packets
(a) Protocol Stack S_1 (b) Protocol Stack S_2

The bitstream analysis can handle specific bitstream encryption cases. For stack S_1 the RTP payload may be encrypted, which means that the RTP header is still accessible to the model.

For stack S_2 two encryption modes are supported: transport stream (TS) payload encryption and PES payload encryption. For PES payload encryption, the PES headers are accessible to the model; for TS payload encryption, only RTP and TS headers are available to the model.

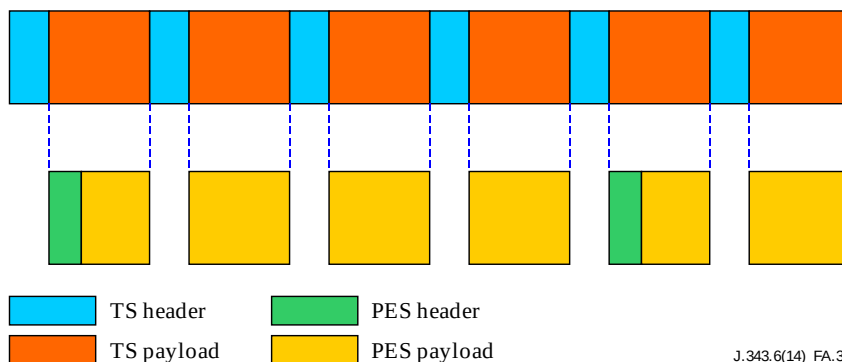


Figure A.3 – PES payload arrangement within one or more TS packets in protocol stack S_2

An audio stream may be present in the bitstream but is not taken into account for the prediction of the perceived video quality.

A.2.2.1 Bitstream analysis module

The bitstream analysis module runs through the following three main functions sequentially:

- pre-processing (common for both supported protocol stacks);
- protocol dependent bitstream analysis;
- aggregation of results from previous steps to a single indicator.

A.2.2.1.1 Pre-processing

The first step of the bitstream analysis is the pre-processing of the received packets. As depicted by the flow diagram in Figure A.4, the pre-processing of the received bitstream is done in two phases, which are explained below.

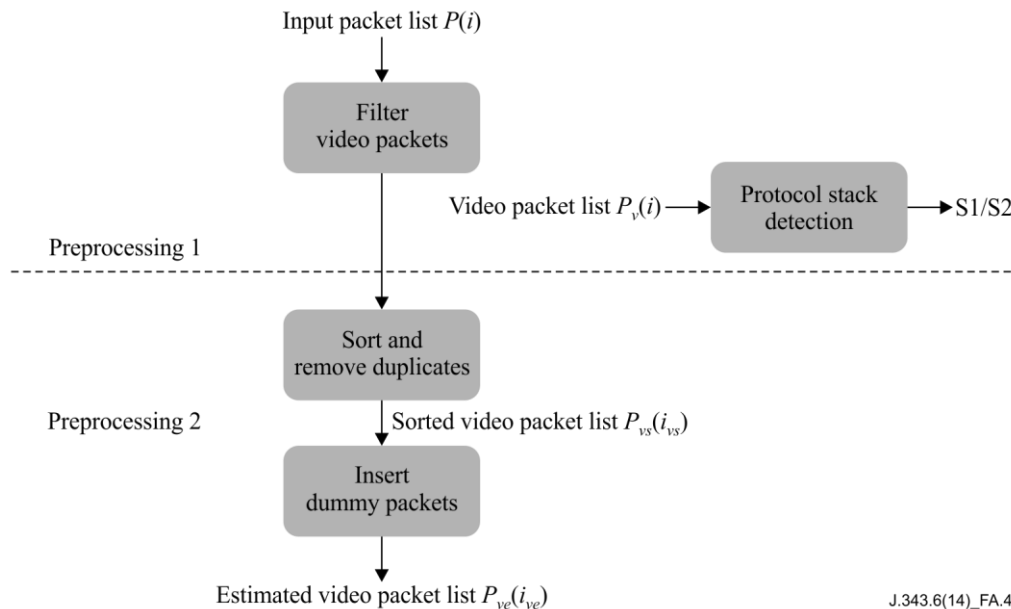


Figure A.4 – Flow diagram of the bitstream pre-processing step

A.2.2.1.2 Filtering and protocol stack detection

The input stream may contain packets that are irrelevant for the analysis and therefore need to be removed. Packets are deemed relevant for the analysis if they belong to the RTP stream containing the video packets.

In general, video packets represent the majority of the input stream. Therefore, the UDP port of the RTP video stream at the client side *clientUdpVideoPort* is equal to the destination UDP port that is present in the highest number of packets in the input stream.

Consequently, the input stream is filtered for UDP packets with destination port *clientUdpVideoPort*. The list of packets that fulfills these requirements is called $P_v(i_v)$ with $i_v \in [0..I_v - 1]$; all other packets are disregarded for further analysis.

Based on the video stream packets P_v the complete protocol stack is determined by probing the RTP packets. If the RTP payload consists of MPEG2-TS packets the protocol stack of the video stream is considered to be S_2 otherwise it is S_1 .

A RTP packet is classified as carrying MPEG2-TS packets if size of the payload of the RTP packet is a multiple of 188 bytes and if every potential TS packet in the RTP payload starts with a byte value of 0x47.

A.2.2.1.3 RTP SEQ unwrapping and packet reordering

As the packets may have arrived out of order due to travelling over different network paths, the next step restores the sending order of the packets.

For S_1 the RTP header carries a sequence index (SEQ) field for this purpose. The SEQ field is a counter that is increased for every sent RTP packet. Therefore, the sending order can be restored by sorting the packets by ascending SEQ number.

The word length of the SEQ field is 16 bits; this means the highest RTP SEQ number that can be represented is 65'535. As the counter overflows if the stream contains more packets the SEQ number must be unwrapped prior to sorting.

After unwrapping the SEQ number and sorting the packets, the sending order is restored. It is possible that packets got duplicated while travelling over the network. These duplicates can be found by searching for packets with the same SEQ number. Removing the duplicates yields the list of ordered, unique packets $P_{vs}(i_{vs})$.

If no packet loss has occurred while transmitting the video stream the SEQ number of each packet $P_{vs}(x)$ fulfills:

$$P_{vs}(x).RTP_Header.SEQ == P_{vs}(x-1).RTP_Header.SEQ + 1 \quad (A.2-1)$$

Otherwise one or more packets have been lost during transmission. The number of lost packets $PL(x)$ between two packets $P_{vs}(x)$ and $P_{vs}(x+1)$ can be determined through comparison of their SEQ numbers.

$$PL(x) = P_{vs}(x+1).RTP_Header.SEQ - P_{vs}(x).RTP_Header.SEQ - 1 \quad (A.2-2)$$

For every lost packet a dummy packet is inserted into $P_{vs}(i_{vs})$ which yields the estimated list of video packets $P_{ve}(i_{ve})$ with $i_{ve} \in [0..I_{ve}-1]$ that have been sent by the streaming server.

The video packet lists P_{vs} and P_{ve} will be used in the following steps to extract features from which the MOV of the bitstream module can be created.

A.2.2.2 Protocol dependent analysis

The input to the protocol dependent analysis is the packet lists P_{vs} and P_{ve} . The result is the video frame rate and the number of video frames in the original streamed video (i.e., $nrOrigFrames$).

A.2.2.2.1 Specific analysis for stack S_1

It is assumed that in case of protocol stack S_1 the time stamp in the RTP header can be set according to two schemes:

- decoding time stamp (DTS);
- presentation time stamp (PTS).

In both cases the RTP time stamp is the same for all packets carrying bytes of the same coded picture. For the DTS scheme the timestamp represents a clock that runs at 90 kHz and reflects the relative time when a coded picture needs to be decoded. For the PTS scheme the timestamp represents a clock that runs at 90 kHz and reflects the relative time when a coded picture needs to be displayed on the screen.

As the coded pictures are stored in decoding order in the bitstream, the RTP timestamp is monotonically increasing with the coded picture number for DTS scheme.

The table below gives an example for DTS and PTS scheme for a video with a frame rate of 30 fps.