International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.1202

(10/2012)

SERIES P: TERMINALS AND SUBJECTIVE AND OBJECTIVE ASSESSMENT METHODS

Models and tools for quality assessment of streamed media

Parametric non-intrusive bitstream assessment of video media streaming quality

Recommendation ITU-T P.1202

ITU-T

International Telecommunication Union

ITU-T P-SERIES RECOMMENDATIONS

TERMINALS AND SUBJECTIVE AND OBJECTIVE ASSESSMENT METHODS

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	Series	P.10
Voice terminal characteristics	Series	P.30
		P.300
Reference systems	Series	P.40
Objective measuring apparatus	Series	P.50
		P.500
Objective electro-acoustical measurements	Series	P.60
Measurements related to speech loudness	Series	P.70
Methods for objective and subjective assessment of speech quality	Series	P.80
		P.800
Audiovisual quality in multimedia services	Series	P.900
Transmission performance and QoS aspects of IP end-points	Series	P.1000
Communications involving vehicles	Series	P.1100
Models and tools for quality assessment of streamed media	**Series**	**P.1200**
Telemeeting assessment	Series	P.1300
Statistical analysis, evaluation and reporting guidelines of quality measurements	Series	P.1400

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T P.1202

Parametric non-intrusive bitstream assessment of video media streaming quality

Summary

Recommendation ITU-T P.1202 provides an overview of algorithmic models for non-intrusive monitoring of the video quality of IP-based video services based on packet-header and bitstream information. The ITU-T P.1202-series of Recommendations addresses two application areas:

– ITU-T P.1202.1 specifies the model algorithm for the lower resolution (LR) application area, including services such as mobile TV.

– ITU-T P.1202.2 specifies the model algorithm for the higher resolution (HR) application area, which includes services such as IPTV.

The ITU-T P.1202 model algorithms are no-reference (i.e., non-intrusive) models which operate by analysing packet header and bitstream information as available from respective packet trace data provided to the model algorithms in the packet capture format (PCAP). Further input information on more general aspects of the stream, which may not be available from packet header and bitstream information, is provided to the model algorithm out-of-band, for example in the form of stream-specific side information.

ITU-T P.1202.1 describes one model, the model for the LR application area. ITU-T P.1202.2 describes two models for the HR application area corresponding to two modes: mode 1 and mode 2, which are both no-reference (i.e., non-intrusive) models. Mode 1 refers to a parsing mode; the model operates by analysing information in the video bitstream without fully decoding the bitstream (i.e., no pixel information is used) for MOS estimation. Mode 2 refers to a full decoding mode, in addition to the bitstream information which mode 1 uses, the model can also decode parts or all of the video bitstream (i.e., pixel information is used) for MOS estimation. Further client specific information, such as concealment type, is provided to the algorithm out-of-band, for example in the form of stream-specific side information. As output, the model algorithms provide individual estimates of video quality in terms of the five-point absolute category rating (ACR) mean opinion score (MOS). Further, diagnostic information on causes of quality degradations can be made available, too, since different types of performance parameters are derived during model calculations.

Complementary to the ITU-T P.1202 models, there are two further models described in Recommendations ITU-T P.1201.1 and ITU-T P.1201.2. The respective entry-Recommendation for these models is ITU-T P.1201. It describes packet-header-only-based video, and audio and audiovisual quality models. The main differences with ITU-T P.1202 can be summarized as follows:

– The ITU-T P.1201 models provide audio, video and audiovisual quality estimates, while the ITU-T P.1202-only models provide video quality estimates.

The ITU-T P.1201 models use packet header information, while the ITU-T P.1202 models exploit further bitstream information, such as coding-related information. As a consequence, the ITU-T P.1202 models can be more accurate in their quality predictions. In turn, they require non-encrypted streams to enable access to payload information. Since the ITU-T P.1202 models are more complex, they also require more computational power to estimate the video quality.

History

Edition	Recommendation	Approval	Study Group
1.0	ITU-T P.1202	2012-10-14	12
1.1	ITU-T P.1202 (2012) Amd. 1	2013-03-28	12

Keywords

Audio, audiovisual, IPTV, mean opinion score (MOS), mobile TV, monitoring, multimedia, QoE, video.

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/.

Table of Contents

Recommendation ITU-T P.1202

Parametric non-intrusive bitstream assessment of video media streaming quality

1 Scope

This Recommendation describes recommended objective models for non-intrusive monitoring of the video quality of IP-based video services based on packet-header and bitstream information. This Recommendation addresses two application areas:

– [ITU-T P.1202.1] specifies the model algorithm for the lower resolution (LR) application area, including services such as mobile TV.

– [ITU-T P.1202.2] specifies the model algorithm for the higher resolution (HR) application area, which includes services such as IPTV. This application area is currently under study but not yet completed.

These models are restricted to information contained in packet headers, the packet bitstream (payload information), prior and static knowledge about the media stream and dynamic buffering information from the client.

[ITU-T P.1202.1] consists of one model. [ITU-T P.1202.2] consists of two models corresponding to two modes: mode 1 and mode 2, which both are no-reference (i.e., non-intrusive) models. Mode 1 refers to a parsing mode; the model operates by analysing information in the video bitstream without fully decoding the bitstream (i.e., no pixel information is used) for MOS estimation. Mode 2 refers to a full decoding mode, in addition to the bitstream information which mode 1 uses, the model can also decode parts or all of the video bitstream (i.e., pixel information is used) for mean opinion score (MOS) estimation. Further client specific information, such as concealment type, is provided to the algorithm out-of-band, for example in the form of stream specific side information.

These models predict video quality in terms of MOSs on a five-point ACR scale (see [ITU-T P.910]).

The primary application for these models is the monitoring of transmission quality during service operation or for maintenance purposes. The ITU-T P.1202 model may be deployed both in end-point locations and at mid-network monitoring points. The location of the model and the location of the measurement probe together determine the *mode of operation*, as described in more detail in clause 6.1.

The primary quality prediction made by such models is based on the payload of the stream being analysed. Therefore, this Recommendation can provide a comprehensive evaluation of quality as perceived by a particular end-user because its scores can reflect the impairments on the coding and the Internet Protocol (IP) network being measured, which differ from user to user. This Recommendation cannot provide a comprehensive evaluation of video quality as perceived by a particular end-user, because its scores reflect the impairments due to encoding and the subsequent IP network being assessed, which may only be one part of the end-to-end connection. An explicit inclusion of processing steps such as content contribution from e.g., satellite networks, display properties etc. are not considered. See Table 3 for more information. Further, the quality-impact due to a specific video encoder implementation or a specific decoder-side packet loss concealment implementation is not explicitly addressed. Instead, the models have been developed for a set of dedicated service implementations, which are assumed to be meaningful representations of today's IP-based streaming video services. As a consequence, however, in case of significant deviations of a given service being assessed from the service configurations used for developing this standard, it is possible to obtain high quality scores with this Recommendation and yet to have a poor quality of the stream as it is perceived by actual users.

As a consequence, this Recommendation can be used for applications such as:

– in-service quality monitoring for specific IP-based audiovisual services, as specified in more detail in Tables 4 to 7;
– benchmarking of different service implementations. However, it cannot be used for direct benchmarking of different decoder implementations. The implementations that can be assessed with ITU-T P.1202 include the encoding aspects and potential packet loss.

The application areas of the ITU-T P.1202 model algorithms are summarized in Tables 1, 2, and 3 below:

Table 1 – Application areas, test factors, and coding technologies for which [ITU-T P.1202.1] has been verified and is known to produce reliable results. For details about the settings, see clause 6

ITU-T P.1202.1 lower resolution (LR)	Higher resolution (HR)
Applications for which the models are intended	
In-service monitoring of video UDP-based streaming	Under study
Performance and quality assessment of live networks including the effect due to encoding and transmission errors	Under study
Test factors for which the models have been validated	
Encoding (compression) degradation of video with a variety of bitrates Video: 50 – 6000 kbit/s	Under study
Packet loss degradation of video (both random and bursty packet loss patterns)	Under study
Re-buffering degradation	–
Video contents of different spatio-temporal complexity	Under study
Different video keyframe and frame-rates Frame rates: 12.5-30 Hz GOP lengths (1/keyframe rate): 2-10 s	Under study
Different video resolutions: HVGA, QVGA, QCIF	Under study
Different decoder-side packet loss concealment strategies (freezing with skipping, one/multiple slices per frame)	Under study
Coding technologies on which the models have been trained	
Video: ITU-T H.264/AVC baseline profile	Under study

Table 2 – Application areas, test factors, and coding technologies for which further investigation of ITU-T P.1202 models is needed

ITU-T P.1202.1 lower resolution (LR)	Higher resolution (HR)
Applications for which the models can be used, but the results may not be reliable	
In-service monitoring of live network video TCP based streaming (assuming that parameter extraction from TCP based streaming is implemented)	Under study
Direct comparison/benchmarking of encoder implementations, and thus of services that employ different encoder implementations	Under study
Test factors for which the models can be used but the results may not be reliable	
–	–
Coding technologies for which the models can be used but the results may not be reliable	
–	–

Table 3 – Application areas, test factors, and coding technologies for which ITU-T P.1202 models are not intended to be used

ITU-T P.1202.1 lower resolution (LR)	Higher resolution (HR)
Applications for which the models are not intended	
Direct comparison/benchmarking of decoder implementations, and thus of services that employ different decoder implementations	Under study
Evaluation of visual quality including display/device properties	Under study
Test factors for which the models are not intended	
Video streaming with significant rate adaptation (such as used in dynamic adaptive streaming over HTTP (DASH)	Under study
Transcoding situations	Under study
The effects of noise, delay, colour correctness	Under study
Audio visual streaming	Under study
Coding technologies for which the models are not intended	
[ITU-T H.261], MPEG-2, MPEG-4, ITU-T H.263, ITU-T H.265, etc.	

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 (2011), *Advanced video coding for generic audiovisual services*.

[ITU-T P.800.1] Recommendation ITU-T P.800.1 (2006), *Mean Opinion Score (MOS) terminology*.

[ITU-T P.910] Recommendation ITU-T P.910 (2008), *Subjective video quality assessment methods for multimedia applications*.

[ITU-T P.1201] Recommendation ITU-T P.1201 (2012), *Parametric non-intrusive assessment of audiovisual media streaming quality*.

[ITU-T P.1201.1] Recommendation ITU-T P.1201.1 (2012), *Parametric non-intrusive assessment of audiovisual media streaming quality – lower resolution application area*.

[ITU-T P.1201.2] Recommendation ITU-T P.1201.2 (2012), *Parametric non-intrusive assessment of audiovisual media streaming quality – higher resolution application area*.

[ITU-T P.1202.1] Recommendation ITU-T P.1202.1 (2012), *Parametric non-intrusive bitstream assessment of video media streaming quality – lower resolution application area*.

[ITU-T P.1202.2] Recommendation ITU-T P.1202.2 (2013), *Parametric non-intrusive bitstream assessment of video media streaming quality – Higher resolution application area*.

[ITU-T P.1401] Recommendation ITU-T P.1401 (2012), *Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following term defined elsewhere:

3.1.1 mean opinion score (MOS): [ITU-T P.800.1].

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 model, model algorithm: An algorithm with the purpose of estimating the subjective (perceived) quality of a media sequence.

3.2.2 sequence: A short decoded audio, video or audiovisual portion of a stream, typically shorter than 30 seconds.

3.2.3 bitstream: The part of an IP-based transmission where the actual audiovisual, video or audio content is available in encoded and packetized form.

3.2.4 compression artefacts: Artefacts introduced due to lossy compression of the encoding process.

3.2.5 slicing artefacts: Artefacts introduced when packet losses are concealed using a packet-loss concealment (PLC) scheme trying to repair erroneous frames.

3.2.6 freezing artefacts: Artefacts introduced when the packet-loss concealment (PLC) scheme of the receiver replaces the erroneous frames (either due to packet loss or error propagation) with the previous error free frame until a decoded picture without errors has been received. Since the erroneous frames are not displayed, this type of artefact is also referred to as freezing with skipping.

3.2.7 rebuffering artefacts: Artefacts coming from rebuffering events at the client side, which could be a result of video data arriving late. Usually, rebuffering events are indicated to the viewer, e.g., in the form of a spinning wheel. This is also referred to as freezing without skipping.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

DASH Dynamic Adaptive Streaming over HTTP

GOP Group of Pictures

HD High Definition (television)

HRC Hypothetical Reference Circuit

HVGA Half Video Graphics Array

IP Internet Protocol

MBAFF Macroblock-Adaptive Frame-Field

MBMS Multimedia Broadcast Multicast Service

MOS Mean Opinion Score

MPEG Motion Pictures Expert Group

NTSC National Television Standard Committee

PAL Phase Alternating Line

PCAP Packet Capture format

PSS Packet Switched Streaming

PVS Processed Video Sequence

QCIF Quarter Common Intermediate Format

QoE Quality of Experience

QVGA Quarter Video

RTP Real-time Transport Protocol

SD Standard Definition

SRC Source Reference Channel or Circuit

UDP User Datagram Protocol

5 Conventions

None.

6 Areas of application

The two application areas for ITU-T P.1202 are:

– [ITU-T P.1202.1] (lower resolution mode (LR)):

  QCIF-QVGA-HVGA, mostly for mobile TV and streaming with the sub-application areas:

  • Linear mobile TV over RTP (includes mobile TV over a 3G mobile network with MBMS and with unicast, transport over RTP/UDP/IP).

  • Multimedia streaming (includes 3GPP PSS with transport over RTP/UDP/IP).

– Higher resolution mode, (HR): SD and HD television, mostly for IPTV with the sub-application areas (this mode is currently under study):

  • Linear broadcast TV (includes transmission over MPEG2-TS/RTP/UDP/IP, and is assumed to be applicable to MPEG2-TS/UDP/IP and RTP/UDP/IP transport with similar, but so far unverified accuracy as compared to MPEG2-TS/RTP/UDP/IP).

  • Video on-demand (includes transmission over MPEG2-TS/RTP/UDP/IP, and is assumed to be applicable to MPEG2-TS/UDP/IP and RTP/UDP/IP transport with similar, but so far unverified accuracy as compared to MPEG2-TS/RTP/UDP/IP).

6.1 Application range for the models

Table 4 below shows the application range of the models based on what the models have actually been trained for. Note that all cases represent the CC mode of operation, see clause 6.2 for more details about the modes.

Table 4 – Factors and application ranges of the ITU-T P.1202 model algorithms

	ITU-T P.1202.1 lower resolution (LR)	**Higher resolution (HR)**
Application information	**Value range, unit**	
Sequence duration (Ts)	The model has been validated on source sequence lengths of: 10 s: no rebuffering 16 s: rebuffering No rebuffering: PVS length = SRC length Rebuffering: PVS length = SRC length + rebuffering length (no rebuffering at end and start) It is expected that the model will give reliable prediction results for sequence durations within the range 8-24 seconds	Under study
Packetization	3GPP MBMS, PSS or using RTSP directly (all three over RTP/UDP/IP)	Under study
Video codec	ITU-T H.264/AVC baseline profile	Under study
Video size	QCIF, QVGA, HVGA	Under study
Coded video bitrate	QCIF: 50-1000 kbit/s QVGA: 80-3000 kbit/s HVGA: 192-6000 kbit/s	Under study

Table 4 – Factors and application ranges of the ITU-T P.1202 model algorithms

	ITU-T P.1202.1 lower resolution (LR)	Higher resolution (HR)
Application information	**Value range, unit**	
Video decoder packet loss concealment	Two types of assumed decoder behaviour are covered: 1) freezing with skipping, 2) slicing with: ITU-T H.264: 1 slice/packet. Fixed PLC (using fixed decoder, details and settings)	Under study
Retransmission mechanisms (ARQ); Forward Error Correction (FEC) Client jitter buffer behaviour	Rebuffering handling, particular to LR-case: without skipping of length 0 to 8 seconds Developed models represent CC-mode (see clause 6.1), hence applied as if dejitter buffer, ARQ and FEC mechanisms have already corrected the stream For other modes of operation, see clause 6.1, appropriate methods to correct the streams in ways reflecting the expected FEC, ARQ and dejitter buffer behaviour are under study.	Under study
Encoder implementation	The model has been trained using the following video encoders (Note 1): – ITU-T H.264/AVC: x264	Under study
Decoder implementation	Reference decoder was a proprietary decoder provided by one proponent, which also performed de-packetization and audio/video-demultiplexing. The ITU-T H.264-decoding is standard-conformant, with the PLC as described above (Note 2).	Under study
Group of pictures (GOP)	GOP-structure is estimated from the stream. Typical GOP structure for which the model has been trained: M = 1, N = 40 (typically no B frames for mobile case) Length: fixed, variable, adaptive Structure (e.g., IPPP...PPPI)	Under study
Frame rate	12.5, 15, 20, 25, 30 fps	Under study
Usage of: Marker bit in RTP header	"End of frame" (True/false)	Under study
Encrypted payload	Not applicable	Not applicable
Packet loss degradation, video	Uniform loss: 0-6% Burst loss: 0-6% (4-state Markov model)	Under study

Table 4 – Factors and application ranges of the ITU-T P.1202 model algorithms

	ITU-T P.1202.1 lower resolution (LR)	Higher resolution (HR)
Application information	**Value range, unit**	
NOTE 1 – It is assumed that the model can be used for estimating quality when other encoder implementations for the given codec have been used. However, if the encoder performance is significantly worse or better than for the encoder used, the model prediction accuracy will be reduced. NOTE 2 – One aspect not covered by decoder packet loss concealment is postfiltering. Guidance on how to adjust internal model parameters for specific other decoders incl. set-top boxes is for further study.		

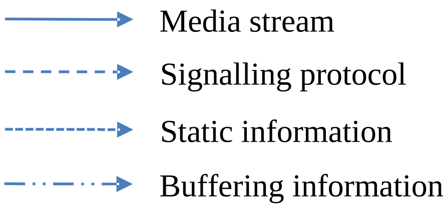
6.2 Modes of operation

The four modes of operation are described in Table 5 and Figures 1-a to 1-e below. Note that the model as described in [ITU-T P.1202.1] support one of the four possible modes (the so-called CC mode). Additional adaptation is required to use the ITU-T P.1202 for the other modes.

Table 5 – Modes of operations of ITU-T P.1202

Class	Name	Mode abbreviation*	Description
Mid-point or End-point	Static operation	NN	The model uses information from the local transport layer, prior knowledge about coding and prior knowledge about the end-point
Mid-point	Non-embedded dynamic operation	BN	The model uses information from the local transport layer, prior knowledge about coding and information about the end-point collected through measurement reporting protocols
Mid-point	Non-embedded distributed operation	CN	The model, located inside the network, uses information from the transport layer measured at an end-point and collected through signalling protocols, prior knowledge about coding and information about the end-point collected through signalling protocols
End-point	Embedded operation	CC	The model uses information from the local transport layer, information from the end-point, and prior knowledge about coding
*Mode abbreviation naming scheme: XY, where X corresponds to place of measurement (N: Network, C: Client, B: Both network and client) Y corresponds to place of model (N: Network, C: Client)			

In Figures 1-a to 1-e below the following arrow style is used:



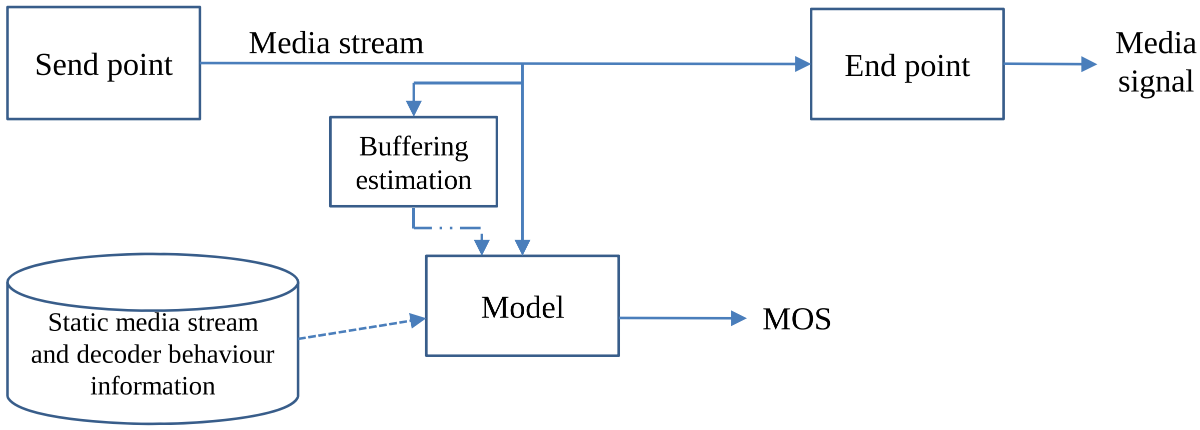


Figure 1-a – Static operation mode (NN) inside the network

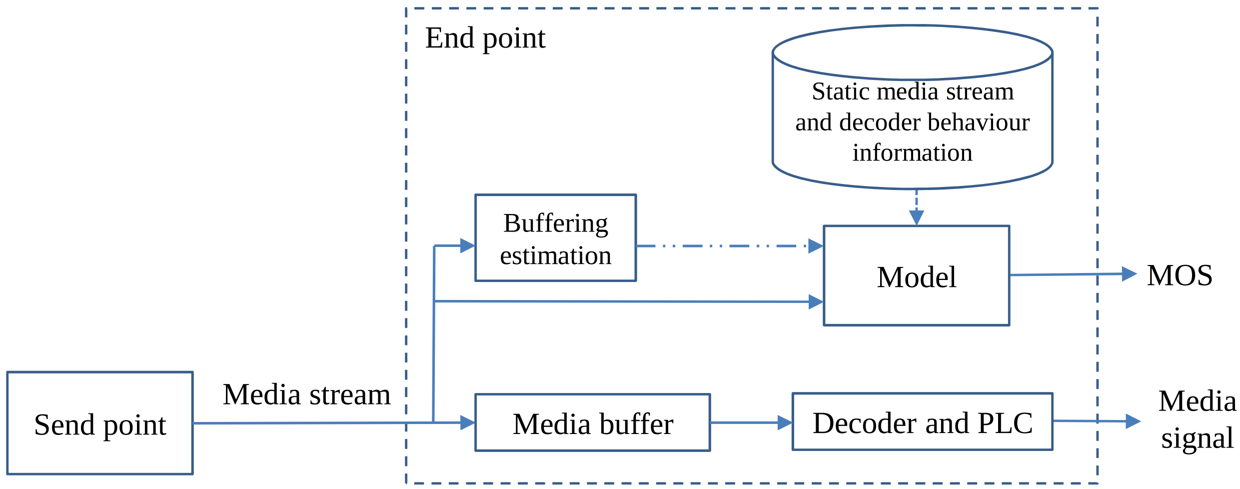


Figure 1-b – Static operation mode (NN) inside a terminal

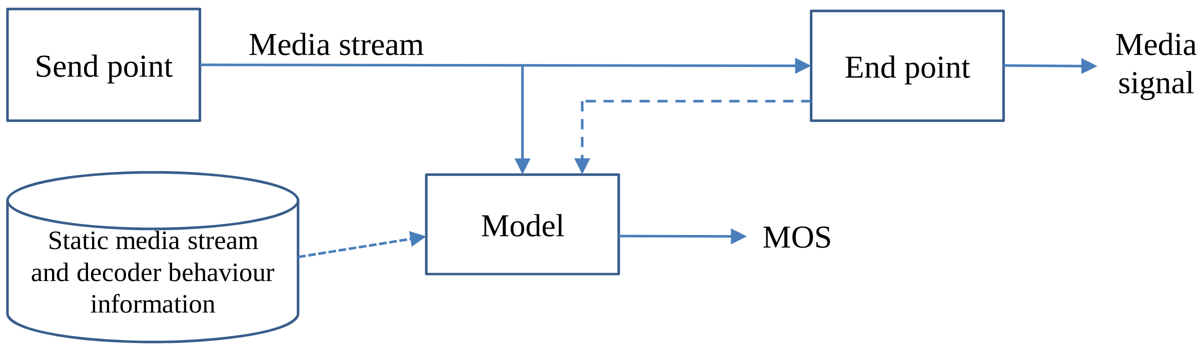


Figure 1-c – Non-embedded dynamic operation mode (BN)

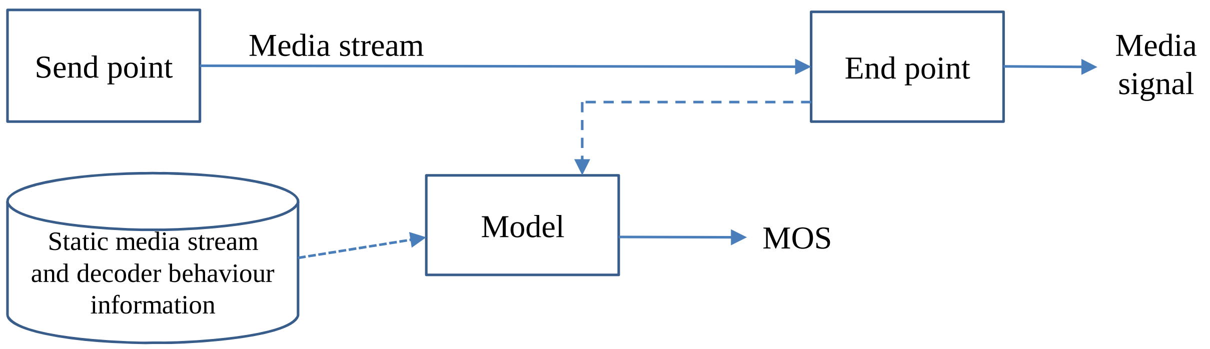


Figure 1-d – Non-embedded distributed operation mode (CN)

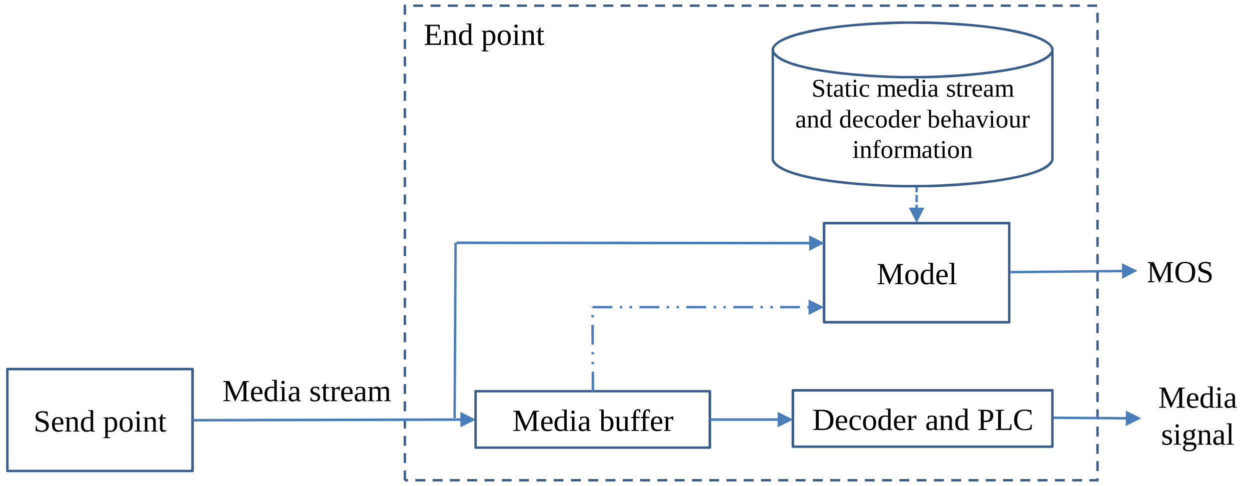


Figure 1-e – Embedded operation mode (CC)

7 Model input interfaces

The ITU-T P.1202 model will receive the encoded bitstream and static side information. For the model as it is described in [ITU-T P.1202.1], the encoded bitstream is expected to be provided in a PCAP file format with transport header information. However, in practical implementations, other than PCAP-based realizations of the transport layer packet extraction can be envisaged. For models as they are described here, the PCAP file could be created based on packets being captured at a network interface. The static side information is information about the media stream and the decoder behaviour. The overview information per application area and mode is described in Table 6 and Figure 2.

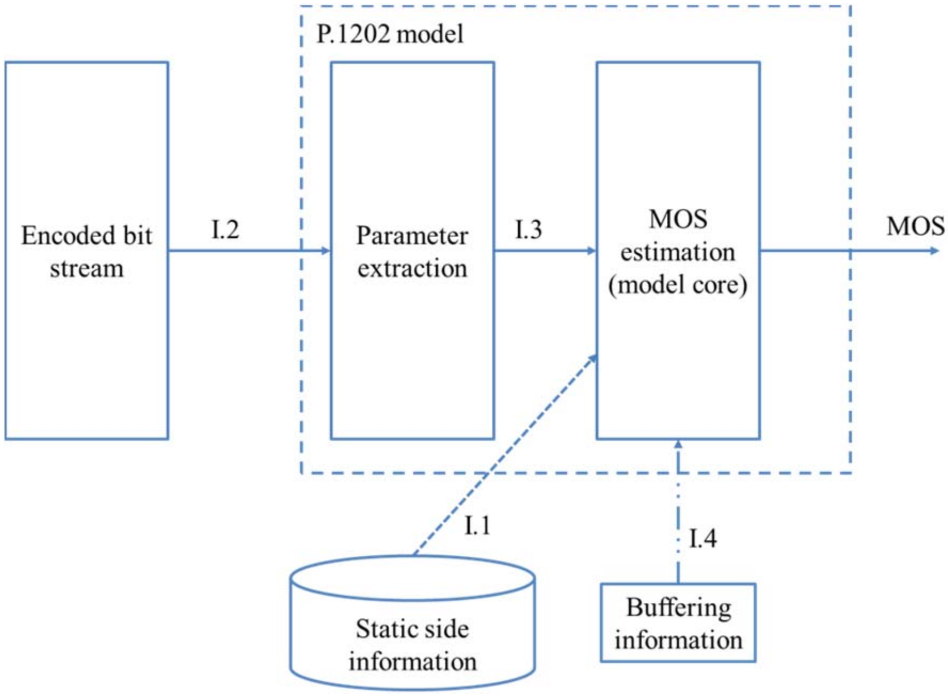


Figure 2 – Overview of ITU-T P.1202 model interfaces

The I.3 interface in the CN mode conveys the signalling information from the parameter extraction module located in the end-point.

The ITU-T P.1202 model has three main inputs:

– Encoded video bitstream: This input can be extracted from a PCAP file, or streamed from a network interface. Parameters from the bitstream can be extracted from the bitstream in the end-point of a CN mode implementation and sent back to the MOS estimation (model core) located in the "model" node.

– Buffering information: This input taken from the media buffer in the client, or estimated by a buffering estimation module using packet information.

– Static media- and decoder information: This input is obtained from packet information or from a player API.

Table 6 – Overview of input to the ITU-T P.1202 model for the modes of operation

	Static operation (NN)	Non-embedded dynamic operation (BN)	Non-embedded distributed operation (CN)	Embedded operation (CC)
Interface 1 (I.1)	Static information about the media stream (codec, usage of flags etc.), and static information about the decoder. Detailed description in Table 7	Static information about the media stream (codec, usage of flags etc.), and static information about the decoder. Detailed description in Table 7 NOTE – Static decoder behaviour information might be provided via signalling and included in I.3	Static information about the media stream (codec, usage of flags, etc.), and static information about decoder. Detailed description in Table 7 NOTE – Static decoder behaviour information might be provided via signalling and included in I.3	Static information about the media stream (codec, usage of flags, etc.), and static information about decoder. Detailed description in Table 7
Interface 2 (I.2)	PCAP file (payload required)	PCAP file (payload required)	PCAP file (payload required)	PCAP file (payload required)
Interface 3 (I.3)	Parameters extracted from the PCAP file	Output from a parameter extraction module located in the end-point and transferred to the model with a protocol	Output from a parameter extraction module located in the end-point and transferred to the model with a protocol	Parameters extracted from the PCAP file
Interface 4 (I.4) **Only available for the ITU-T P.1202 LR model**	Rebuffering information (estimated)	Rebuffering parameters measured/extracted in the end-point and transferred to the model with a protocol	Rebuffering parameters measured/extracted in the end-point and transferred to the model with a protocol	Rebuffering information from the media buffer

Table 7 gives more detailed examples of information that is provided at the input interfaces I.1 (see Figure 2).

Table 7 – Input and typical values to ITU-T P.1201 models

Input	Typical value
Dynamic input	
Media stream	PCAP file or other capture format
Rebuffering information	Text file containing rebuffering information
Static information	
Video destination port	1234
Video codec	H264
Video codec profile	Baseline
Video resolution	QCIF, QVGA, HVGA

Table 7 – Input and typical values to ITU-T P.1201 models

Input	Typical value
Video scanning type	Progressive
Video frame rate	12.5, 15, 25, 30
Video packet loss concealment	Slicing, freezing

The models described in ITU-T P.1202 have been validated assuming that the available information already reflects the impact of any error resilience methods, such as forward error correction (FEC), or packet re-transmission mechanisms such as automatic repeat request (ARQ), and of the impact due to the dejitter buffer. This is equivalent to the parameter extraction module being located behind these processing steps, that is, implementing the CC mode of operation. In case of the NN mode, the measurement point is located prior to the actual FEC, ARQ and dejitter buffer mechanisms. Since these mechanisms may have a very strong impact on factors such as the packet loss seen by the decoder, the NN-mode requires an explicit handling of these mechanisms, to reflect a CN or BN type of behaviour. To capture this case, the packet stream may be converted into a stream that reflects an assumed behaviour of FEC, ARQ, and/or jitter de-buffering, reflecting the input format to be provided to the parameter extraction module. This step results in a converted stream, see Figure 3-b.

Figure 3 shows how the case of error resilience methods such as forward error correction (FEC) and automatic repeat request (ARQ) could be handled in case of the NN/CN/BN modes. Figure 3-a shows the model architecture when FEC/ARQ is not used. In Figure 3-b the media stream is first corrected using a FEC/ARQ dejitter buffer and then the parameters are extracted in exactly the same way as in the case without FEC/ARQ.

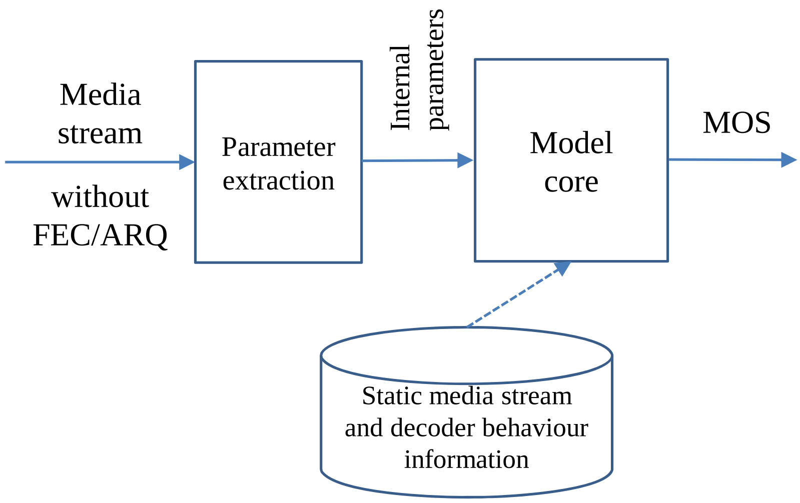


Figure 3-a – Model structure when error resilience methods (FEC/ARQ) are not used

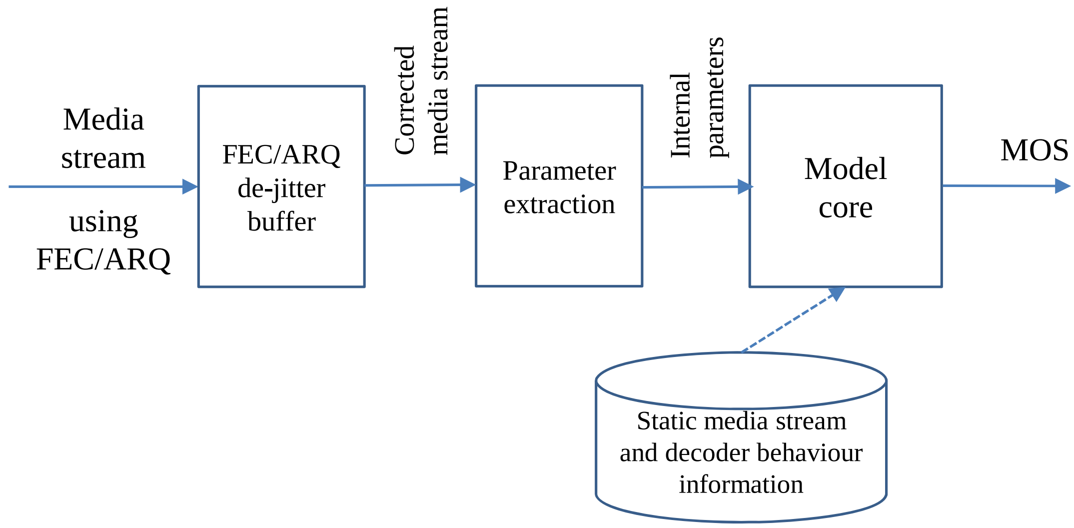


Figure 3-b – Model structure when FEC/ARQ is used and the stream is corrected before parameter extraction

8 Model output information and performance details

The ITU-T P.1202 models have one output parameter:

- Estimated video MOS on the 1 to 5 scale, which is an estimation of the perceived video quality.

The performance information for the ITU-T P.1202.1 model can be found in Table 8 and in Appendix I. The performance information for the ITU-T P.1202.2 models can be found in Table 9, Table 10, Appendix II and Appendix III. The statistical metrics RMSE (root mean square error) and Pearson correlation are used to describe the performance, see [ITU-T P.1401]. Note that for those performance figures, the subjective ratings have been mapped to the model scores using a linear, i.e., 1st-order mapping function, at a per-database level. This has been done in order to avoid misalignment due to bias in the different subjective tests, e.g., as a result of different test settings.

Table 8 – Performance information for ITU-T P.1202.1

	RMSE	**Pearson correlation**
Overall performance	0.397 (based on 982 sequences)	0.918 (based on 982 sequences)

Table 9 – Performance information for ITU-T P.1202.2 mode 1

	RMSE	**Pearson correlation**
Overall performance	0.357 (based on 3069 sequences)	0.938 (based on 3069 sequences)

Table 10 – Performance information for ITU-T P.1202.2 mode 2

	RMSE	**Pearson correlation**
Overall performance	0.353 (based on 3069 sequences)	0.940 (based on 3069 sequences)

9 Description of the ITU-T P.1202 model algorithm

The ITU-T P.1202 lower resolution model is described in [ITU-T P.1202.1]. The ITU-T P.1202 higher resolution model is described in [ITU-T P.1202.2].

Appendix I

Detailed performance figures for the ITU-T P.1202.1 algorithm

(This appendix does not form an integral part of this Recommendation.)

	ITU-T P.1202.1 (lower resolution)
Overall video RMSE	0.397 (based on 982 sequences)
RMSE for video packet loss conditions causing freezing	0.400 (based on 104 sequences)
RMSE for video packet loss conditions causing slicing	0.508 (based on 422 sequences)
RMSE for video rebuffering	0.299 (based on 168 sequences)
RMSE for pure compression conditions	0.284 (based on 288 sequences)
Pearson correlation	0.918 (based on 982 sequences)

Media	**Codec**	**Degradation type**	**RMSE**	**PC**	**# files**
Video	Overall		0.397	0.918	982
	H264 (QCIF)	Compression, Slicing, Freezing, Rebuffering	0.442	0.895	207
	H264 (QVGA)	Compression, Slicing, Freezing, Rebuffering	0.365	0.931	375
	H264 (HVGA)	Compression, Slicing, Freezing, Rebuffering	0.402	0.921	400

Appendix II

Detailed performance figures for the ITU-T P.1202.2 mode 1 algorithm

(This appendix does not form an integral part of this Recommendation.)

	ITU-T P.1202.2 (higher resolution) mode 1
Overall video RMSE	0.357 (based on 3069 sequences)
RMSE for video packet loss conditions causing freezing	0.313 (based on 687 sequences)
RMSE for video packet loss conditions causing slicing	0.396 (based on 1374 sequences)
RMSE for pure compression conditions	0.325 (based on 1008 sequences)
Pearson correlation	0.938 (based on 3069 sequences)

Media	**Codec**	**Degradation type**	**RMSE**	**PC**	**# files**
Video	Overall		0.357	0.938	3069
	H264 (SD)	Compression, Slicing, Freezing	0.337	0.940	698
	H264 (720P)	Compression, Slicing, Freezing	0.354	0.938	719
	H264 (HD)	Compression, Slicing, Freezing	0.368	0.937	1652

Appendix III

Detailed performance figures for the ITU-T P.1202.2 mode 2 algorithm

(This appendix does not form an integral part of this Recommendation.)

	ITU-T P.1202.2 (higher resolution) mode 2
Overall video RMSE	0.353 (based on 3069 sequences)
RMSE for video packet loss conditions causing freezing	0.287 (based on 687 sequences)
RMSE for video packet loss conditions causing slicing	0.392 (based on 1374 sequences)
RMSE for pure compression conditions	0.337 (based on 1008 sequences)
Pearson correlation	0.940 (based on 3069 sequences)

Media	**Codec**	**Degradation type**	**RMSE**	**PC**	**# files**
Video	Overall		0.353	0.940	3069
	H264 (SD)	Compression, Slicing, Freezing	0.335	0.943	698
	H264 (720P)	Compression, Slicing, Freezing	0.346	0.942	719
	H264 (HD)	Compression, Slicing, Freezing	0.364	0.937	1652

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	**Terminals and subjective and objective assessment methods**
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects and next-generation networks
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2013