ITU-T

International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.1201
(10/2012)

SERIES P: TERMINALS AND SUBJECTIVE AND OBJECTIVE ASSESSMENT METHODS

Models and tools for quality assessment of streamed media

Parametric non-intrusive assessment of audiovisual media streaming quality

Recommendation ITU-T P.1201

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.1201

Parametric non-intrusive assessment of audiovisual media streaming quality

Summary

Recommendation ITU-T P.1201 provides an overview of algorithmic models for non-intrusive monitoring of the audio, video and audiovisual quality of IP-based video services based on packet-header information. The ITU-T P.1201-series of Recommendations addresses two application areas:

– ITU-T P.1201.1 specifies the model algorithm for the lower resolution (LR) application area, including services such as mobile TV.

– ITU-T P.1201.2 specifies the model algorithm for the higher resolution (HR) application area, including services such as IPTV.

The two ITU-T P.1201 model algorithms are no-reference (i.e., non-intrusive) models which operate by analysing packet header 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, such as the video resolution, which may not be available from packet header information, is provided to the model algorithm out-of band, for example in the form of stream-specific side information.

As output, the model algorithms provide individual estimates of audio, video and audiovisual quality in terms of the five-point absolute category rating (ACR) mean opinion score (MOS) scale. Further, diagnostic information on causes of quality degradations can also be made available.

Complementary to the ITU-T P.1201 models, there are further models described in the ITU-T P.1202-series of Recommendations. These so-called ITU-T P.1202 models are bitstream-based video quality models. The main differences with ITU-T P.1201 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 the 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.1201	2012-10-14	12
1.1	ITU-T P.1201 (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.1201

Parametric non-intrusive assessment of audiovisual media streaming quality

1 Scope

This Recommendation provides an overview of two recommended objective parametric quality assessment models that predict the impact of observed IP network impairments on quality experienced by the end user in multimedia mobile streaming and IPTV applications over transport formats such as: RTP (over UDP), MPEG2-TS (over UDP or RTP/UDP), 3GPP-PSS (over RTP).

As an umbrella Recommendation, ITU-T P.1201 provides the framework for the actual two algorithmic models described in [ITU-T P.1201.1] and [ITU-T P.1201.2]. These algorithms are aimed at monitoring the audio, video and audiovisual quality of IP-based video services based on packet-header information, with two application areas:

– [ITU-T P.1201.1] specifies the model algorithm for the lower resolution application area, including services such as mobile TV.
– [ITU-T P.1201.2] specifies the model algorithm for the higher resolution application area, including services such as IPTV.

These models are restricted to input information contained in packet headers, prior and static knowledge about the media stream and dynamic buffering information from the client. Information from decoding the bit-stream or parsing the packet payload is not used.

These models predict mean opinion scores (MOS) on a five-point ACR scale (see [ITU-T P.910]) for audio and video parts of the stream, as well as a global audiovisual MOS score (as defined in [ITU-T P.911], for instance).

The primary application for these models is the monitoring of transmission quality during in-service operation or for maintenance purposes. The ITU-T P.1201 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 not based on the information in the audio and video coding layer (payload) of the stream being analysed, but on information on the payload as it is available from packet headers and additional side-information.

This Recommendation cannot provide a comprehensive evaluation of audiovisual 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, for example, satellite networks, display properties, etc., are not considered. Hence, also the effects of audio level, noise, delay (and corresponding similar video factors) and other impairments related to the media payload are not reflected in the scores computed by such a model. Further, the quality impact due to a specific audio or 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, but yet to have a poor quality of the stream as it is perceived by actual users, or vice versa. Moreover, the scores predicted by a parametric model (i.e., without access to payload information) necessarily reflect a somewhat simplified representation of the perceptual impairment of the considered stream.

However, with only using packet-header information, the models still enable estimation of payload-related information, and thus allow to provide valid and, in most cases, accurate predictions, presuming that they are applied in an appropriate manner, following this Recommendation.

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 below;

– benchmarking of different service implementations. However, it cannot be used for direct benchmarking of different encoder implementations, but only the effect of different encoding bitrates and transmission errors for a given decoder-based packet loss concealment. The implementations that can be assessed with ITU-T P.1201 include the audio and video encoding bitrates, the employed video GOP-structure, frame rate, resolution, potential packet loss and the audio codec type.

The audio model algorithm is considered as a part of ITU-T P.1201 that may be employed in combination with other than the ITU-T P.1201 video model algorithm, for example, with the bitstream-based ITU-T P.1202 model algorithms (ITU-T P.1202-series of Recommendations), or be used as a stand-alone audio model. Consequently, care is taken that the application range for the audio model algorithm is designed to be slightly larger than necessary for a component of an audiovisual model applied to services as considered for the ITU-T P.1201 activity.

The application areas of the ITU-T P.1201 model algorithms are summarized in Tables 1, 2, and 3 below:

Table 1 – Application areas, test factors and coding technologies where [ITU-T P.1201.1] and [ITU-T P.1201.2] have been verified and are known to produce reliable results. For details about the settings, see clause 6

[ITU-T P.1201.1] – Lower resolution (LR)	[ITU-T P.1201.2] – Higher resolution (HR)
Applications the models are intended for	
In-service monitoring of audiovisual, video, and audio UDP-based streaming	(same for both models, see entry on the left)
Performance and quality assessment of live networks (including codecs) including the effect due to encoding bitrate, and transmission errors	(same for both models, see entry on the left)
Test factors the models have been validated for	
Encoding (compression) degradation of audio and video with a variety of bitrates Video: 40-6000 kbit/s Audio: 4.75-576 kbit/s	Encoding (compression) degradation of audio and video with a variety of bitrates Video: 0.5-30 Mbit/s Audio: 16-384 kbit/s
Packet loss degradation of audio and video (both random and bursty packet loss patterns)	Packet loss degradation of audio and video (both random and bursty packet loss patterns)
Re-buffering degradation (audio-only re-buffering not validated)	–
Video contents of different spatio-temporal complexity	(same for both models, see entry on the left)

Table 1 – Application areas, test factors and coding technologies where [ITU-T P.1201.1] and [ITU-T P.1201.2] have been verified and are known to produce reliable results. For details about the settings, see clause 6

[ITU-T P.1201.1] – Lower resolution (LR)	[ITU-T P.1201.2] – Higher resolution (HR)
Different video keyframe and frame rates Frame rates: 5-30 Hz GOP lengths (1/keyframe rate): 2-10 s	Different video group of picture (GOP) structures and video frame rates Supporting variable GOP structures with all frame types
Different video resolutions: HVGA, QVGA, QCIF	Different video resolutions: SD (PAL/NTSC), HD (720p50, 720p60, 1080i50, 1080p25, 1080i60, 1080p30)
Different decoder-side packet loss concealment strategies (freezing with skipping, one slice per RTP packet/frame)	Different decoder-side packet loss concealment strategies (freezing with skipping, one/multiple slices per frame – based on the slicing settings chosen in the encoder)
–	Interlaced and progressive scan
Coding technologies the models have been trained on	
Video: MPEG4 Part 2, ITU-T H.264 (MPEG4 Part 10)	Video: ITU-T H.264 (MPEG4 Part 10)
Audio: AMR-NB/WB+, AAC-LC, HE-AACv1/v2	Audio: AAC-LC, HE-AACv2, AC3, MPEG-LII

Table 2 – Application areas, test factors and coding technologies for which further investigation of ITU-T P.1201.1 and ITU-T P.1201.2 models is needed

[ITU-T P.1201.1] – Lower resolution (LR)	[ITU-T P.1201.2] – Higher resolution (HR)
Applications where the models can be used, but the results may not be reliable	
In-service monitoring of live network audiovisual, video, and audio TCP-based streaming (assuming that parameter extraction from TCP-based streaming is implemented)	In-service monitoring of live network audiovisual, video, and audio TCP-based streaming without re-buffering (assuming that parameter extraction from TCP-based streaming is implemented)
Test factors where the models can be used but the results may not be reliable	
–	–
Coding technologies where the models can be used but the results may not be reliable	
Specific implementations of video encoders and decoders other than the codecs used in the development, see clause 6	(same for both models, see entry on the left)

Table 3 – Application areas, test factors, and coding technologies for which [ITU-T P.1201.1] and [ITU-T P.1201.2] are not intended to be used

[ITU-T P.1201.1] – Lower resolution (LR)	[ITU-T P.1201.2] – Higher resolution (HR)
Applications the models are not intended for	
Direct comparison/benchmarking of encoder and decoder implementations, and thus of services that employ different encoder or decoder implementations	(same for both models, see entry on the left)
Evaluation of audiovisual quality including display/device properties	(same for both models, see entry on the left)
Test factors the models are not intended for	
Audiovisual streaming with significant rate adaptation (such as used in dynamic adaptive streaming over HTTP (DASH) streaming)	(same for both models, see entry on the left)
Transcoding situations	(same for both models, see entry on the left)
The effects of audio level, noise, delay (and corresponding similar video factors)	(same for both models, see entry on the left)
–	Re-buffering degradation of audio, video and audiovisual
Coding technologies the models are not intended for	
ITU-T H.261, MPEG-2, ITU-T H.263, ITU-T H.265, etc. (Note)	
NOTE – For the exact set of codecs for which the models have been validated, see clause 6.	

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 (2012), *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.911] Recommendation ITU-T P.911 (1998), *Subjective audiovisual quality assessment methods for multimedia applications*.

[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] Recommendation ITU-T P.1202 (2012), *Parametric non-intrusive bitstream assessment of video media streaming quality*.

[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.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 terms 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 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:

AAC Advanced Audio Coding

AAC-LC Advanced Audio Coding – Low Complexity

AC3 Audio Coding 3

ACR Absolute Category Rating

AMR-NB	Adaptive Multi-Rate – Narrowband
AMR-WB	Adaptive Multi-Rate – Wideband
API	Application Programming Interface
ARQ	Automatic Repeat Request
DASH	Dynamic Adaptive Streaming over HTTP
FB	Fullband
FEC	Forward Error Correction
GOP	Group Of Pictures
HD	High Definition (television)
HE-AAC	High-Efficiency Advanced Audio Coding
HR	Higher Resolution
HRC	Hypothetical Reference Circuit
HTTP	Hypertext Transfer Protocol
HVGA	Half Video Graphics Array
IDR-	Instantaneous Decoder Refresh (frame)
I-	Inline-(frame)
LR	Lower Resolution
MBMS	Multimedia Broadcast/Multicast Service
MOS	Mean Opinion Score
MPEG	Motion Pictures Expert Group
NB	Narrowband
NTSC	National Television Standard Committee
PAL	Phase Alternating Line
PCAP	Packet Capture Format
PCC	Pearson Correlation Coefficient
PES	Packetized Elementary Stream
PLC	Packet Loss Concealment
PSS	Packet-switched Streaming Service
PVS	Processed Video Sequence
QCIF	Quarter Common Intermediate Format
QoE	Quality of Experience
QVGA	Quarter Video Graphics Array
RMSE	Root Mean Square Error
RTP	Real-time Transport Protocol
RTSP	Real Time Streaming Protocol
SD	Standard Definition
SRC	Source Reference Channel or Circuit

SWB	Superwideband
TS	Transport Stream
UDP	User Datagram Protocol
VSP	Visual Simple Profile
WB	Wideband

5 Conventions

None.

6 Areas of application

The two application areas for ITU-T P.1201 are:

– ITU-T P.1201.1 (lower resolution (LR) mode):

  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).

– ITU-T P.1201.2 (higher resolution (HR) mode): SD and HD television, mostly for IPTV with the sub-application areas:

  • Linear broadcast TV (includes transmission over MPEG2-TS/RTP/UDP/IP, MPEG2-TS/UDP/IP (see [ITU-T P.1201.2]), and is assumed to be applicable to RTP/UDP/IP transport with similar, but so far unverified, accuracy as compared to MPEG2-TS/RTP/UDP/IP and MPEG2-TS/UDP/IP)

  • Video on-demand (includes transmission over MPEG2-TS/RTP/UDP/IP, MPEG2-TS/UDP/IP (see [ITU-T P.1201.2]), and is assumed to be applicable to RTP/UDP/IP transport with similar, but so far unverified, accuracy as compared to MPEG2-TS/RTP/UDP/IP and MPEG2-TS/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 developed 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.1201.1 and ITU-T P.1201.2 model algorithms

	ITU-T P.1201.1 – Lower resolution (LR)	ITU-T P.1201.2 – 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	The model has been validated on source sequence lengths of: 10 s: no rebuffering – PVS length = SRC length It is expected that the model will give reliable prediction results for sequence durations of approximately 8-16 seconds
Packetization	3GPP MBMS, PSS or using RTSP directly (all three over RTP/UDP/IP)	MPEG2-TS/RTP/UDP/IP RTP/UDP/IP (Note 3) MPEG2-TS/UDP/IP (Note 3)
Video codec	MPEG4 visual simple profile (VSP) ITU-T H.264 baseline profile	ITU-T H.264 main profile, ITU-T H.264 high profile
Video size	QCIF, QVGA, HVGA	SD: PAL, NTSC HD: 720p, 1080p, 1080i (High profile: 1080; main profile: 720, SD)
Audio codec	AMR-NB, AMR-WB+, AAC-LC, HE-AAC (v1, v2)	MPEG-4 AAC-LC MPEG-4 HE-AAC (V1 and V2 = 3GPP enhanced AAC+) MPEG-1 Layer 2 AC3 MPEG-2 audio
Coded video bitrate	[ITU-T H.264] QCIF: 32-1000 kbit/s QVGA: 80-3000 kbit/s HVGA: 192-6000 kbit/s MPEG4 QCIF: 40-1500 kbit/s QVGA: 90-3500 kbit/s HVGA: 192-6000 kbit/s	HD [ITU-T H.264]: 0.5 up to 30 Mbit/s SD [ITU-T H.264]: 0.5 up to 9 Mbit/s

Table 4 – Factors and application ranges of the ITU-T P.1201.1 and ITU-T P.1201.2 model algorithms

	ITU-T P.1201.1 – Lower resolution (LR)	**ITU-T P.1201.2 – Higher resolution (HR)**
Application information	**Value range, unit**	
Coded audio bitrate	AMR-NB: 4.75-12.2 kbit/s AMR-WB+: 10.4-48 kbit/s AAC-LC: 16-128+ kbit/s HE-AAC (v1, v2): 32-128 kbit/s	AAC-LC: 32-576 kbit/s HE AACv2: 16-96 kbit/s MPEG-1 Layer 2: 64-384 kbit/s AC3: 64-384 kbit/s
Video decoder packet loss concealment	Two types of assumed decoder behaviour are covered: 1) freezing with skipping; 2) slicing with: MPEG4: 1 slices/frame [ITU-T H.264]: 1 slice/packet Both MPEG4 and [ITU-T H.264]: Fixed PLC (using fixed decoder, details and settings)	Types of decoder behaviour: two dimensions: slicing, PLC 1) freezing with skipping (duration(source)=duration(processed sequence)); 2) slicing with 1 slice per frame; 3) slicing with 1 slice per macroblock row, PLC with zero-motion copy (temporal from same region of previous good frame).
Audio decoder packet loss concealment	Decoder default modes: Codec-implementation specific loss concealment	(same for both models, see entry on the left)
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 for further study.	(No rebuffering) 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 for further study.
Encoder implementation	The model has been developed using the following encoders (Note 1): • MPEG4 Part 2: ffmpeg • ITU-T H.264 (MPEG4 Part 10): x264 • AMR-NB/WB+: According to standard • AAC: Nero	The model has been developed using the following encoders (Note 1): Video: • ITU-T H.264 (MPEG-4 Part 10): x264 Audio: • AAC-LC, HE-AAC v2: Nero • MPEG1-LII and AC3: ffmpeg

Table 4 – Factors and application ranges of the ITU-T P.1201.1 and ITU-T P.1201.2 model algorithms

	ITU-T P.1201.1 – Lower resolution (LR)	**ITU-T P.1201.2 – Higher resolution (HR)**
Application information	**Value range, unit**	
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).	(same for both models, see entry on the left)
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)	Supporting default modes for typical GOP structures E.g., M = 3, N = 15 Length: fixed, variable, adaptive Structure (e.g., IBBPBB...PBBI) NOTE – GOP structure is explicitly estimated from stream.
Frame rate	5, 8.33, 12.5, 15, 20, 25, 30 fps	SD: 50i (PAL), 59.94i (NTSC) HD: 50p, 59.94p, 60p, 50i, 59.94i, 60i, 25p, 29.97p, 30p
Marker bit in RTP header	Model can handle that marker bit means end of frame	For TS, see TS-related information below
Audio channel number	1 (diotic mono), 2 (stereo)	2 (stereo)
# of Audio frames per RTP packet	1 to 5 audio frames	Bitrate-specific (depending on both the audio and video bitrate)
Audio-video multiplexed?	Default: No, at RTP-level; no audio-video asynchrony	In MPEG-2 TS/RTP/UDP and MPEG-2 TS/UDP: Supported
Encrypted payload	Only payload level	Transport streams: Both PES- and TS-level encryption, i.e., including and excluding PES-headers, respectively
Usage of: payload-unit-start indicator in TS headers	Not applicable	Information on whether payload-unit-start indicator in TS header means the start of frame is provided to the model as side information (Boolean)
Usage of: marker bit in RTP header	"End of frame" (True/false)	Not applicable
Usage of: random-access indicator in TS header	Not applicable	Information on whether random-access indicator in TS header means the start of I-frame or of IDR-frame is provided to the model as side information (Boolean)

Table 4 – Factors and application ranges of the ITU-T P.1201.1 and ITU-T P.1201.2 model algorithms

	ITU-T P.1201.1 – Lower resolution (LR)	ITU-T P.1201.2 – Higher resolution (HR)
Application information	**Value range, unit**	
Usage of: elementary-stream-propriety indicator in TS header	Not applicable	Information on whether elementary-stream-propriety indicator in TS header means the start of I-frame or of IDR-frame is provided to the model as side information (Boolean)
Packet loss degradation, video	Uniform loss: 0-10% Burst loss: 0-10% (4-state Markov model)	Uniform loss: 0-2% Burst loss: 0-2% (4-state Markov model)
Packet loss degradation, audio	Uniform loss: 0-10% Burst loss: 0-10% (4-state Markov model)	Uniform loss: 0-6% Burst loss: 0-6% (4-state Markov model)
Symmetrical versus asymmetrical handling of audio and video in audiovisual case	Model application: Symmetrical, but can handle asymmetric cases due to specific model development process	(same for both models, see entry on the left)
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, including set-top boxes, is for further study. NOTE 3 – The ITU-T P.1201.2 model has been trained on MPEG-2 TS/RTP/UDP. However, due to the design of the ITU-T P.1201.2 algorithm, it is also applicable to MPEG2-TS/UDP/IP. Further, to the model's design, it is assumed to also work for RTP/UDP/IP transport with similar, but so far unverified, accuracy as compared to MPEG2-TS/RTP/UDP/IP.		

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 models as described in [ITU-T P.1201.1] and [ITU-T P.1201.2] only support one of the four possible modes (the so-called CC mode). Additional adaptation is required to use the ITU-T P.1201 for the other modes.

Table 5 – Modes of operations of ITU-T P.1201

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:

Media stream

Signalling protocol

Static information

Buffering information

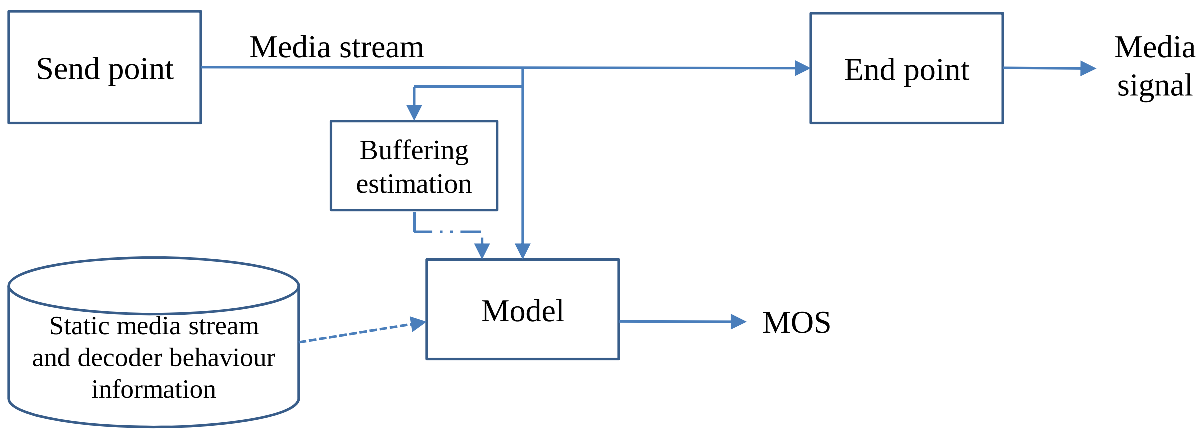

Figure 1-a – Static operation mode (NN) inside the network

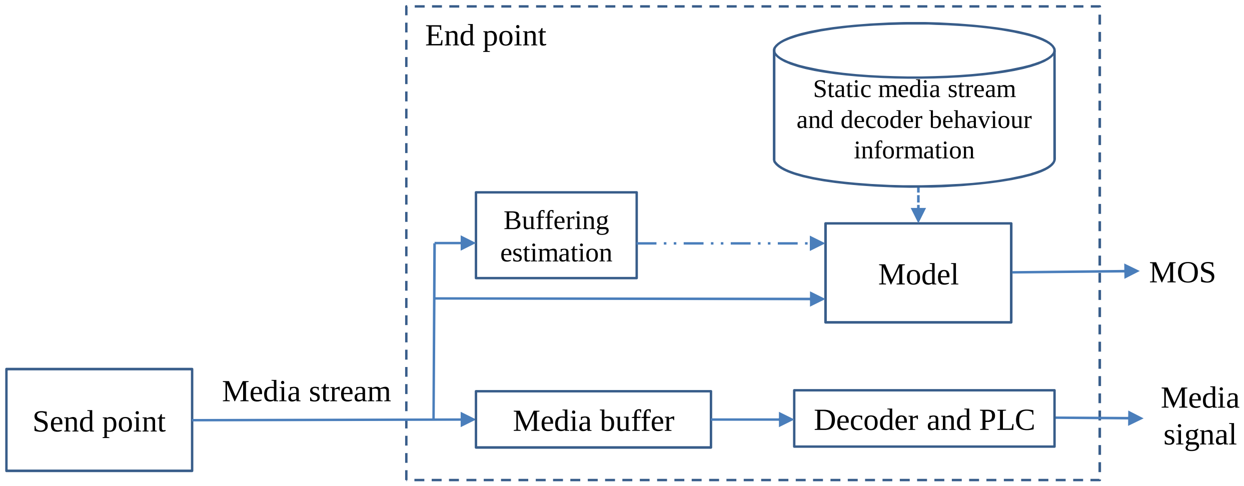

Figure 1-b – Static operation mode (NN) inside a terminal

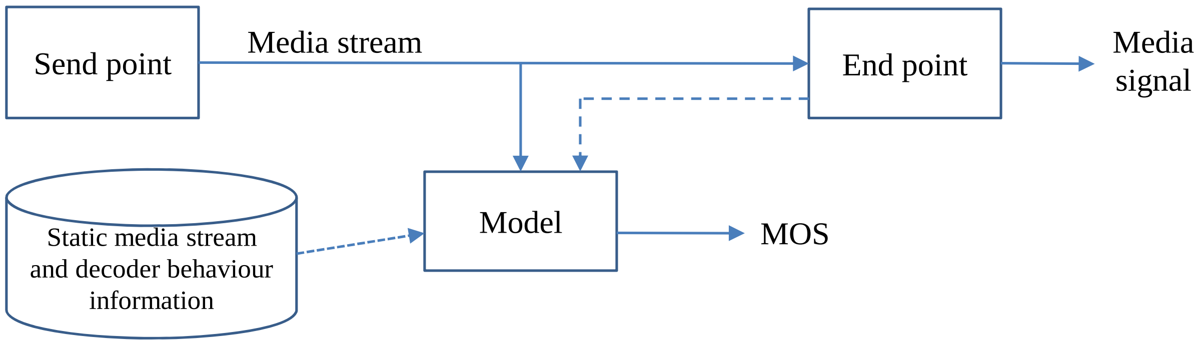

Figure 1-c – Non-embedded dynamic operation mode (BN)

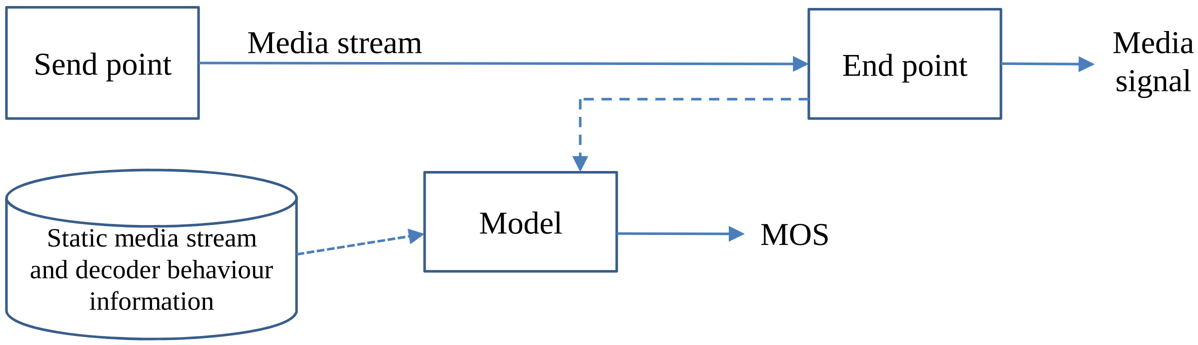

Figure 1-d – Non-embedded distributed operation mode (CN)

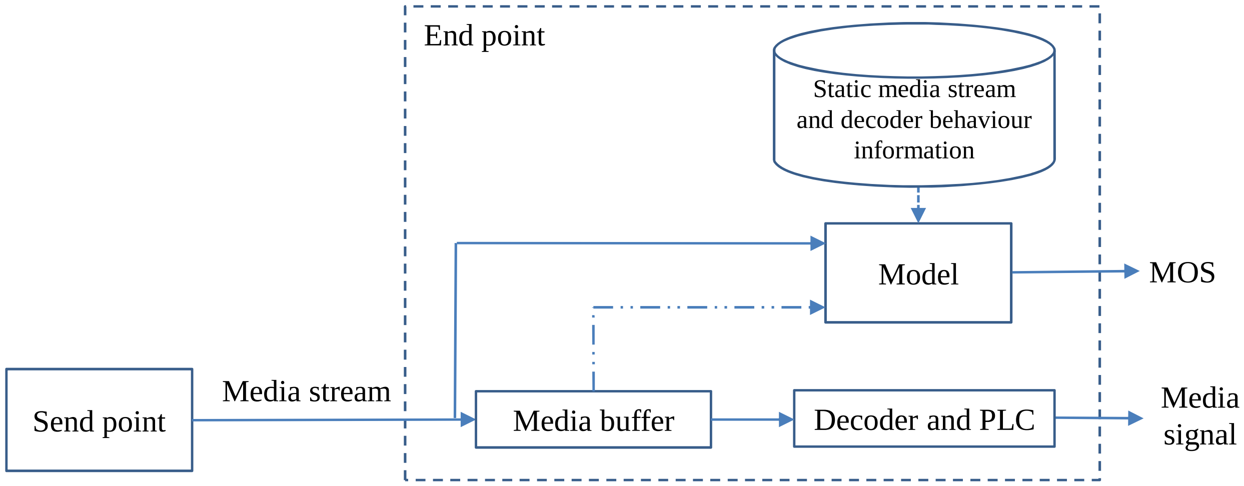

Figure 1-e – Embedded operation mode (CC)

7 Model input interfaces

The ITU-T P.1201 models will receive dynamic transport layer information and static side information. For the models as they are described in [ITU-T P.1201.1] and [ITU-T P.1201.2], the dynamic transport layer input information is expected to be provided in PCAP file format with transport header information. However, in practical implementations, realizations of the transport layer information extraction other than PCAP-based realizations can be envisaged. For the 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 below:

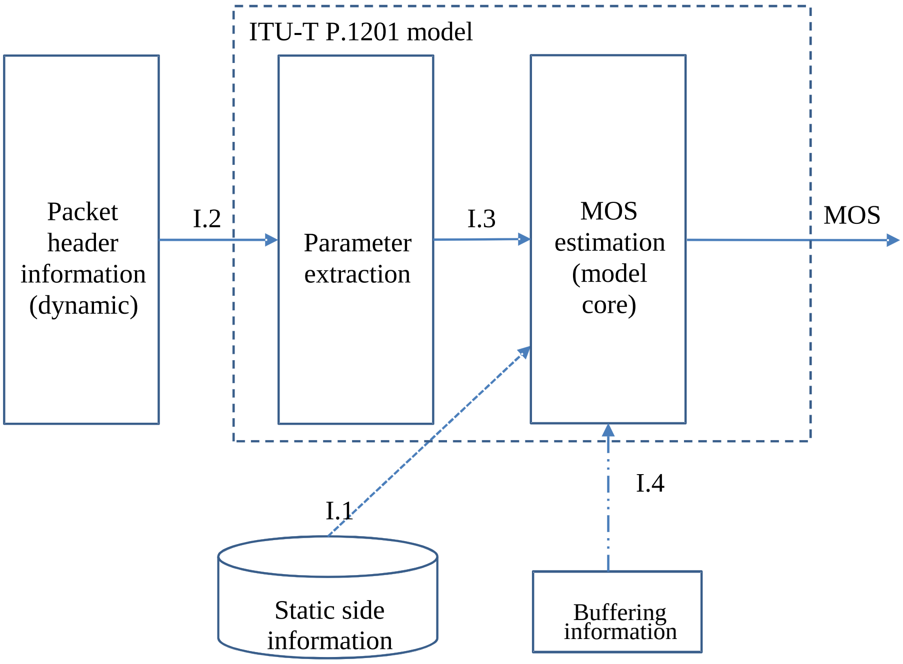

Figure 2 – Overview of ITU-T P.1201 model interfaces

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.1201 model has three main inputs:

– **Packet-header information**: This input can be taken from a PCAP file, or streamed from a network interface or from a parameter extraction module located in the end point in a CN mode implementation.
– **Buffering information**: This input can be 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.1201 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 decoder. Detailed description in Table 7.	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 interface I.3 (see Figure 2)	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 interface I.3 (see Figure 2)	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 not required)	PCAP file (payload not required)	Not available in this mode	PCAP file (payload not required)
Interface 3 (I.3)	Parameters extracted from the PCAP file	Output from a parameter extraction module for transport header information located in the end point and transferred to the model with a protocol.	Output from a parameter extraction module for transport header information 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.1201 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

The following Tables 7 and 8 give more detailed examples of information that is provided at the input interfaces I.1 (see Figure 2 above).

Table 7 – Input to the ITU-T P.1201.1 model algorithm (LR)

Input	Typical values
Dynamic input	
Media stream	PCAP file or other capture format
Rebuffering information	Text file containing rebuffering information
Static information	
Audio destination port	1234
Video destination port	5678
Video codec	ITU-T H.264, MPEG4
Video codec profile	Simple, baseline
Video resolution	QCIF, QVGA, HVGA
Video scanning type	Progressive
Video frame rate	5, 8.33, 10, 12.5, 15, 25, 30
Video packet loss concealment	Slicing, freezing
Audio codec	AMRNB; AMRWB+; MPEG1L2; AAC-LC; HE-AAC v1,v2; AC3
Audio channel number	1, 2
Audio bandwidth	NB, WB, SWB, FB
Audio frame length	20 to 42.67
RTP marker bit usage	"Last packet of frame" (TRUE/FALSE)

Table 8 – Input to the ITU-T P.1201.2 model algorithm (HR)

Input	Typical values
Dynamic input	
Media stream	PCAP file or other capture format
Static information	
Audio destination port	1234
Video destination port	5678
Video codec	ITU-T H.264
Video codec profile	Main, High
Video resolution	PAL, NTSC, HD720, HD1080
Video scanning type	Progressive, Interlaced
Video frame rate	24, 25, 29.97, 30, 50, 60
Video packet loss concealment	Slicing, freezing
Number of slices per frame	1, 30, 34, 68
Audio codec	MPEG1L2, AAC-LC, HE-AAC v2, AC3

Table 8 – Input to the ITU-T P.1201.2 model algorithm (HR)

Input	Typical values
Audio channel number	2
Audio bandwidth	FB
Audio frame length	21.33 to 42.67 ms
Transport stream random access indicator	"First TS-packet of I-frame", "Not applicable"
Elementary stream priority	"First TS-packet of I-frame", "Not applicable "
Payload unit start indicator	"First TS-packet of a frame", "Not applicable"

The models described in [ITU-T P.1201.1] and [ITU-T P.1201.2] 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. Two principal approaches are conceivable to capture this case, which are both based on prior knowledge:

a) The packet stream S is 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 (PCAP). This step results in a converted stream S', see Figure 3-c.

b) The input parameters to the model are converted into values that reflect the application of FEC, ARQ and/or dejitter buffering (module "loss-related parameter mapping" in 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 extracted internal model parameters are mapped to reflect the properties of a corrected stream. Note that the "Loss-related parameter mapping" module is considered transparent for all non loss-related parameters. In Figure 3-c 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.

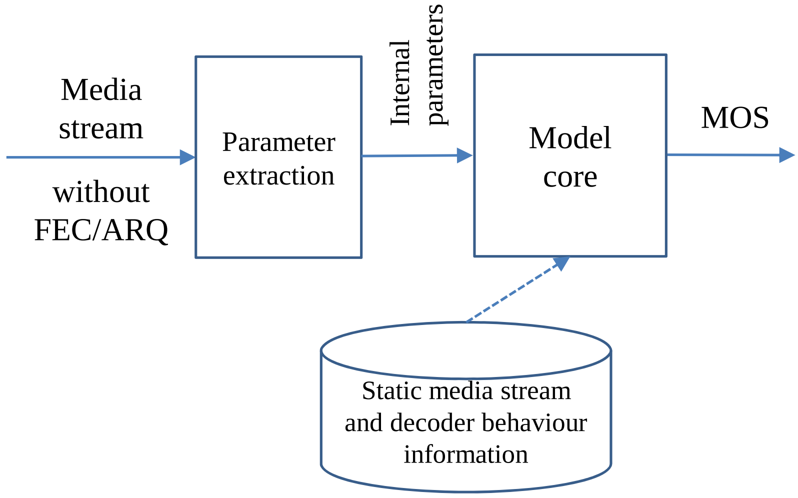

Figure 3-a – Model structure when error resilience methods (FEC/ARQ) are not used

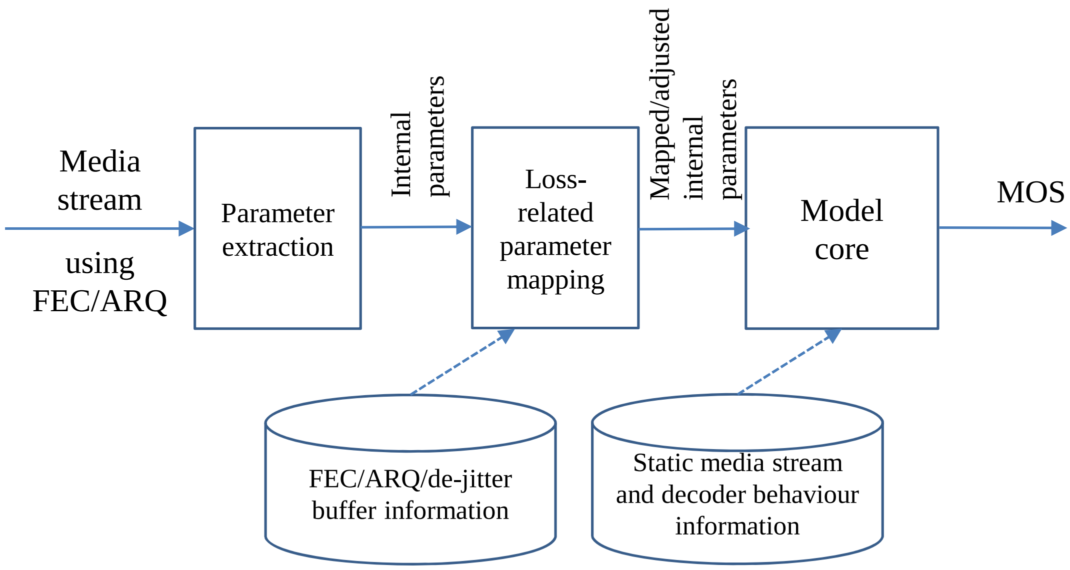

Figure 3-b – Model structure when FEC/ARQ are in use and extracted parameters are mapped

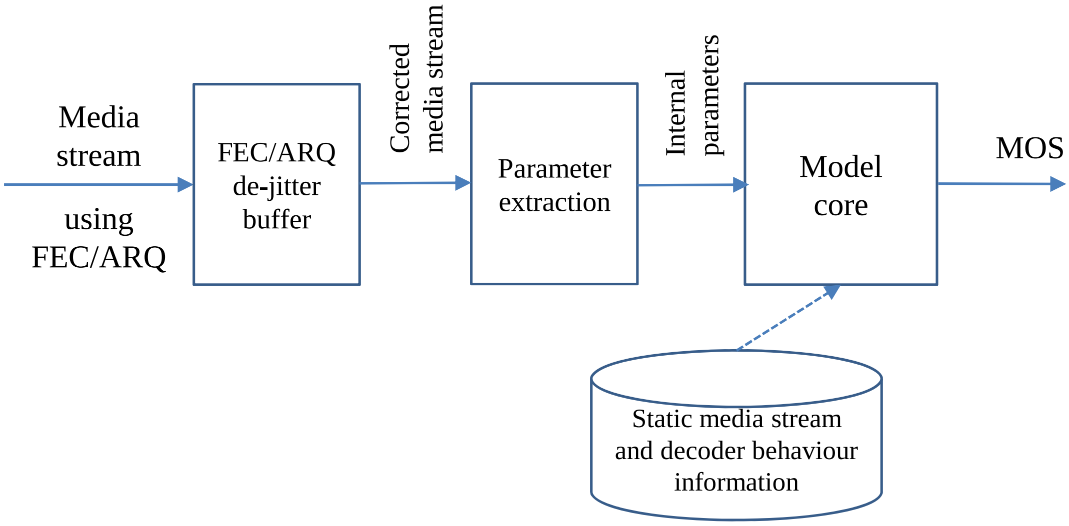

Figure 3-c – Model structure when FEC/ARQ are used and the stream is corrected before parameter extraction

8 Model output information and performance details

The ITU-T P.1201.1 and ITU-T P.1201.2 models have three output parameters:

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

2) Estimated video MOS on the 1 to 5 scale, which is an estimation of the perceived video quality (without audio present). Note that the model is able to give a video score for a stream without audio, and a stream including audio.

3) Estimated audio MOS on the 1 to 5 scale, which is an estimation of the perceived audio quality (without video present). Note that the model is able to give an audio score for a stream without video, and a stream including video.

The performance information for the ITU-T P.1201 models can be found in Table 9 and in Appendix I. The statistical metrics root mean square error (RMSE) and Pearson correlation are used to describe the performance, see [ITU-T P.1401]. Note that for these 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 9 – Performance information for ITU-T P.1201.1 (LR)

	RMSE	**Pearson correlation**
Audiovisual	0.470 (based on 1166 samples)	0.852 (based on 1166 samples)
Video	0.535 (based on 1430 samples)	0.830 (based on 1430 samples)
Audio	0.351 (based on 690 samples)	0.941 (based on 690 samples)

Table 10 – Performance information for ITU-T P.1201.2 (HR)

	RMSE	Pearson correlation
Audiovisual	0.435 (based on 3190 samples, PES and TS encrypted)	0.911 (based on 3190 samples, PES and TS encrypted)
Video	0.461 (based on 6138 samples, PES and TS encrypted)	0.902 (based on 6138 samples, PES and TS encrypted)
Audio	0.336 (based on 1360 samples, PES and TS encrypted)	0.949 (based on 1360 samples, PES and TS encrypted)

9 Description of the ITU-T P.1201 model algorithms

The ITU-T P.1201 lower resolution model is described in [ITU-T P.1201.1], and the description of the ITU-T P.1201 higher resolution model algorithm can be found in [ITU-T P.1201.2].

Appendix I

Detailed performance figures for the ITU-T P.1201.1 and ITU-T P.1201.2 algorithms

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

Table I.1 – Detailed performance figures for the ITU-T P.1201.1 algorithm

Media	Video codec	Audio codec	RMSE	Pearson correlation	# files
All	Overall		0.478	0.867	3286
Audio	Overall		0.351	0.941	690
	–	AAC-LC	0.300	0.955	207
	–	AAC-HEv1	0.354	0.960	24
	–	AAC-HEv2	0.429	0.808	16
	–	AMRNB	0.263	0.755	161
	–	AMRWBP	0.419	0.853	282
Video	Overall		0.535	0.830	1430
	H264 (QCIF)	–	0.531	0.848	207
	MPEG4 (QCIF)	–	0.461	0.829	184
	H264 (QVGA)	–	0.529	0.851	375
	MPEG4 (QVGA)	–	0.523	0.708	264
	H264 (HVGA)	–	0.584	0.832	400
Audiovisual	Overall		0.470	0.852	1166
	H264 (QCIF)	AAC-LC	0.360	0.951	87
	H264 (QCIF)	AMRNB	0.480	0.701	64
	H264 (QCIF)	AMRWBP	0.499	0.882	80
	MPEG4 (QCIF)	AAC-LC	0.420	0.754	48
	MPEG4 (QCIF)	AMRNB	0.379	0.917	48
	MPEG4 (QCIF)	AMRWBP	0.474	0.823	56
	H264 (QVGA)	AAC-LC	0.463	0.888	153
	H264 (QVGA)	AMRNB	0.443	0.811	71
	H264 (QVGA)	AMRWBP	0.639	0.729	100
	MPEG4 (QVGA)	AAC-LC	0.645	0.724	42
	MPEG4 (QVGA)	AMRNB	0.495	0.706	73
	MPEG4 (QVGA)	AMRWBP	0.395	0.749	104
	H264 (HVGA)	AAC-LC	0.422	0.927	80

Table I.1 – Detailed performance figures for the ITU-T P.1201.1 algorithm

Media	Video codec	Audio codec	RMSE	Pearson correlation	# files
	H264 (HVGA)	AMRNB	0.465	0.784	72
	H264 (HVGA)	AMRWBP	0.455	0.868	88

Table I.2a – Detailed performance figures for the ITU-T P.1201.2 algorithm

Media	Video codec	Audio codec	RMSE	Pearson correlation	# files
All	Overall		0.439	0.911	10688
Audio	Overall		0.336	0.949	1360
	–	AAC-HEv2	0.463	0.861	188
	–	AAC-LC	0.328	0.950	514
	–	MPEG1L2	0.317	0.941	450
	–	AC3	0.258	0.975	208
Video	Overall		0.461	0.902	6138
	H264 (SD)	–	0.465	0.898	1396
	H264 (HD720)	–	0.465	0.897	1438
	H264 (HD1080)	–	0.458	0.905	3304
Audiovisual	Overall		0.435	0.911	3190
	H264 (SD)	AAC-LC	0.408	0.921	502
	H264 (SD)	AC3	0.519	0.857	316
	H264 (HD720)	AAC-HEv2	0.318	0.943	48
	H264 (HD720)	AAC-LC	0.427	0.932	256
	H264 (HD720)	MPEG1L2	0.454	0.881	482
	H264 (HD720)	AC3	0.411	0.884	160
	H264 (HD1080)	AAC-LC	0.444	0.920	744
	H264 (HD1080)	MPEG1L2	0.406	0.920	682

Table I.2b – Detailed performance figures for the ITU-T P.1201.2 algorithm

Media	Encryption level	RMSE	Pearson correlation	# files
Audio	PES	0.337	0.949	680
	M2TS	0.336	0.949	680
Video	PES	0.459	0.903	3069
	M2TS	0.463	0.901	3069
Audiovisual	PES	0.433	0.913	1595
	M2TS	0.438	0.91	1595

Appendix II

Performance figures for the ITU-T P.1201.2 algorithm in the case of MPEG2-TS/UDP

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

In order to validate the ITU-T P.1201.2 algorithm in the case of MPEG2-TS/UDP, two sets of PCAP files were used:

A) MPEG2-TS/RTP/UDP case: The PES-encrypted ITU-T P.1201.1 and ITU-T P.1201.2 training databases for higher resolutions (HRs) with MPEG2-TS/RTP/UDP as packetization scheme.

B) MPEG2-TS/UDP case: The same PCAP files as A). However, this time, a MPEG2-TS/UDP packetization scheme was simulated by:

- not reading the RTP-header (to ensure a true simulation none of the parameters of the RTP-header were actually read at all);
- setting a CTS-flag, which implies a replacement of the RTP-timestamp by the video-DTS, as described in clause 4.2.3 of [ITU-T P.1201.2];
- not using the RTP-sequence number for the calculation of the number of lost packets, as described in clause 4.2.2 of [ITU-T P.1201.2]. Instead, a different (new) function was used that utilized the CC of the TS-headers in order to estimate the number of lost packets.

Both sets A) and B), that were used for the validation, contain the following number of databases and PCAP files:

– One audio database: 240 files

– Five video-only databases: 1200 files

– Three audiovisual databases: 714 files

As performance indicators, the Pearson correlation coefficient (PCC) and the root-mean-square-error (RMSE) were computed between the MOS score predictions on sets A) and B). Performance results are reported in Table II.1. The percentages of PCAP files with a MOS estimation difference smaller than 0.05 between the MPEG2-TS/RTP/UDP and MPEG2-TS/UDP PCAP files are also provided.

Table II.1 – Performance results for the ITU-T P.1201.2 algorithm in the case of MPEG2-TS/UDP

Media	PCC	RMSE	% files with MOS diff < 0.05
Audio for audio database	0.99	0.01	98.75%
Audio for audiovisual databases	0.99	0.01	99.86%
Video for video databases	0.99	0.06	90.42%
Video for audiovisual databases	0.99	0.05	93.56%
Audiovideo	0.99	0.03	95.10

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