International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.1201

Amendment 2
(12/2013)

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

Amendment 2: New Appendix III – Use of ITU-T P.1201 for non-adaptive, progressive download type media streaming

Recommendation ITU-T P.1201 (2012) – Amendment 2

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

Amendment 2

New Appendix III – Use of ITU-T P.1201 for non-adaptive, progressive download type media streaming

Summary

Amendment 2 to Recommendation ITU-T P.1201 (2012) introduces Appendix III.

Appendix III to ITU-T P.1201 describes a method allowing ITU-T P.1201 to be used for quality predictions of TCP-based, non-adaptive streaming typically referred to as progressive download (PD). The appendix can be used standalone, since it reproduces the required portions of model algorithms from ITU-T P.1201.1 and ITU-T P.1201.2.

Here, the following changes or additions have been applied to the ITU-T P.1201 scope:

– Longer duration source sequences of between 30 s and 60 s compared to 10 s to 16 s long source sequences as in the case of the ITU-T P.1201.1, ITU-T P.1201.2, ITU-T P.1202.1 and ITU-T P.1202.2 models;

– TCP-based transport instead of UDP-based transport;

– Inclusion of longer stalling events as they may occur in case of playout-buffer underrun;

– Inclusion of initial buffering, before the actual playout starts.

It is noted that the model has been developed based on a relatively small subjective test dataset, as compared to the large database used for the ITU-T P.1201.x series model training and validation procedure.

What is not addressed by this model:

– Adaptive streaming;

– Sequences longer than 60 s;

– User-interaction with the player (stop, play, rewind, fast forward, jump to specific part of sequence, etc.).

The goal of this appendix is to provide the market with an initial model capable of estimating quality related with PD-type streaming applications but with a limited scope.

The appendix is targeted to be usable standalone, and therefore the required parts from ITU-T P.1201 and the model descriptions in ITU-T P.1201.1 and ITU-T P.1201.2 are provided here for the implementers' convenience.

History

Edition	Recommendation	Approval	Study Group	Unique ID[*]
1.0	ITU-T P.1201	2012-10-14	12	11.1002/1000/11727
1.1	ITU-T P.1201 (2012) Amd. 1	2013-03-28	12	11.1002/1000/11940
1.2	ITU-T P.1201 (2012) Amd. 2	2013-12-12	12	11.1002/1000/12109

Keywords

Audio, audiovisual, mean opinion score (MOS), monitoring, multimedia, progressive download, QoE, stalling, video.

[*] 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/.

Recommendation ITU-T P.1201

Parametric non-intrusive assessment of audiovisual media streaming quality

Amendment 2

New Appendix III – Use of ITU-T P.1201 for non-adaptive, progressive download type media streaming

1) Appendix III

Introduce Appendix III after Appendix II of this Recommendation, as shown below.

Appendix III

Use of ITU-T P.1201 for non-adaptive, progressive download type media streaming

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

III.1 Scope

This appendix describes an objective parametric quality assessment model that predicts the impact of observed IP network impairments on quality experienced by the end user in multimedia mobile streaming and fixed network applications using progressive download.

The model described is restricted to information provided to it by an appropriate packet- or bitstream-analysis module. The model is applicable for the effects due to audio- and video-coding as well as initial buffering and re-buffering (which are both perceivable as stalling of the media) as the typical degradations associated with progressive download.

The model predicts mean opinion scores (MOS) on a 5-point ACR scale (see [ITU-T P.910]) as a global multimedia MOS score (as defined in [ITU-T P.911], for instance). In addition, audio-, video-only and audio-visual MOS according to ITU-T P.1201 as well as a perceptual stalling quality indicator (MOS) are provided for diagnostic purposes.

The primary applications for this model are monitoring of transmission quality for operations and maintenance purposes. The ITU-T P.1201 model for non-adaptive, progressive download type media streaming may be deployed both in end-point locations and at mid-network monitoring points. The location of the model, together with the location of the measurement probe determines the mode of operation. Note, however, that the present document only describes the model that maps input parameters obtained from a probe located at a specific point in the network or in the client to the global multimedia MOS-scores and related quality diagnostic indicators, as described above.

This model cannot provide a comprehensive evaluation of transmission quality as perceived by a particular end user because its scores reflect the impairments on the IP network being measured, which may only be part of the end-to-end connection. Furthermore, the scores predicted by a parametric model necessarily reflect an average perceptual impairment. Note also that the model is developed with a specific encoder and decoder pair. If a different encoder and decoder pair is used in a monitoring situation the scores will not reflect that.

The effects of audio level, noise, delay (and corresponding similar video factors) and other impairments related to the payload are not reflected in the scores computed by this model. Therefore, it is possible to have high scores with this model, yet have a poor quality stream overall. 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 model still enables estimation of payload-related information, and thus allows for the provision of valid and in most cases accurate predictions, presuming that it is 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. The implementations that can be assessed with ITU-T P.1201 for non-adaptive, progressive download type media streaming include the audio and video encoding bitrates, the employed video GOP-structure, frame-rate, resolution and the audio codec type.

The application areas of the ITU-T P.1201 model for non-adaptive, progressive download type media streaming are summarized in Tables III.1, III.2 and III.3:

Table III.1 – Application areas, test factors and coding technologies where ITU-T P.1201 for non-adaptive, progressive download type media streaming has been verified and are known to produce reliable results[a]

Applications for which the model is intended
In-service monitoring of audiovisual, TCP-based video and audio. Both so called over-the-top (OTT) services (for example YouTube), and operator managed video services (over TCP), using the protocols HTTP/TCP/IP and RTMP/TCP/IP. This model is intended to be used for video services typically using container formats such as Flash (FLV), MP4, WebM and 3GP. Note that this model is agnostic to the type of container format
Performance and quality assessment of live networks (including codecs) including the effect due to encoding bitrate, and transmission problems causing long initial buffering and stalling events
Test factors for which the model has been validated
Encoding (compression) degradation of audio and video with a variety of bitrates: Video: 200-6000 kbit/s (HVGA), 2-16 Mbit/s (HD) Audio: 24-128 kbit/s
Stalling (re-buffering) degradation (audio-only and video-only re-buffering not validated) and initial buffering
Video contents of different spatio-temporal complexity
Different video keyframe and frame-rates and GOP lengths for HVGA resolution • Frame rate: 12-30 Hz • GOP lengths: 2 s For HD resolution: • Frame rate: 24 and 30 Hz • GOP lengths: M3N24 (about 1 s)
Different video resolutions: HVGA, HD (1080i50, 1080p24, 1080i60, 1080p30)
Interlaced and progressive scan for HD resolution

Table III.1 – Application areas, test factors and coding technologies where ITU-T P.1201 for non-adaptive, progressive download type media streaming has been verified and are known to produce reliable results[a]

Coding technologies on which the model has been trained
ITU-T H.264 (MPEG4 Part 10)
Audio for HD resolution sequences: HE-AACv1/v2 Audio for HVGA resolution sequences: AMR-WB, AAC-LC, HE-AACv1
[a] For details about the settings, cf. clause III.8.

Table III.2 – Application areas, test factors and coding technologies for which further investigation of ITU-T P.1201 for non-adaptive, progressive download type media streaming is needed

Applications where the model can be used, but the results may not be reliable
–
Test factors where the model can be used but the results may not be reliable (conditions not included in subjective tests underlying the model development)
Video resolutions: QCIF, CIF, SD (PAL/NTSC), HD (720p50, 720p60)
Encoding (compression) degradation of audio and video with bitrates slightly wider range than what has been verified: Video: 200 kbit/s-30 Mbit/s Audio: 4.75-576 kbit/s
Additional keyframe, frame-rates and GOP structures for lower resolutions: • Frame rate: 5-30 Hz • GOP lengths: 2-10 s For HD resolution: • Different video group-of-pictures (GOP) structures and video frame-rates
Coding technologies where the models can be used but the results may not be reliable
MPEG4 Part 2
For lower resolution videos: AMR-NB, HE-AACv2 For higher resolution videos: AAC-LC, AC3, MPEG1-LII
Specific implementations of video en- and decoders other than the codecs used in the development, cf. clause III.8

Table III.3 – Application areas, test factors, and coding technologies for which ITU-T P.1201 for non-adaptive, progressive download type media streaming is not intended to be used

Applications for which the model is not intended
Direct comparison/benchmarking of encoder and decoder implementations, and thus of services that employ different en- or decoder implementations
Evaluation of audio-visual quality including display/device properties
Test factors for which the model is not intended
Audio-visual streaming with significant rate adaptation (such as used in DASH/HTTP streaming)
Transcoding situations
The effects of audio level, noise, delay (and corresponding similar video factors)
Coding technologies for which the model is not intended
ITU-T H.261, MPEG-2, ITU-T H.263, ITU-T H.265, etc. (Note)
NOTE – For the exact set of codecs the models have been validated for, see clause III.8.

III.2 References

This appendix uses the following references:

[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.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*.

III.3 Definitions

This appendix uses the following terms:

III.3.1 Terms defined elsewhere

This appendix uses the following term defined elsewhere:

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

III.3.2 Terms defined in this Recommendation

This appendix uses the following terms defined in this Recommendation:

III.3.2.1 bitstream: The part of an IP-based transmission where the actual audiovisual, video or audio content is available in encoded and packetized form.

III.3.2.2 compression artefacts: Artefacts introduced due to lossy compression of the encoding process.

III.3.2.3 model, model algorithm: An algorithm with the purpose of estimating the subjective (perceived) quality of a media sequence.

III.3.2.4 rebuffering or stalling artefacts: Artefacts coming from rebuffering or stalling events at the client side, which could be a result of video data arriving late. Usually, rebuffering/stalling events are indicated to the viewer, e.g., in the form of a spinning wheel. This is also referred to as freezing without skipping.

III.3.3 Terms defined in this appendix

This appendix defines the following terms:

III.3.3.1 initial buffering: The state of the video client starting when the user selects to play a video and ending when the playout actually starts. During this time the client's buffer is filled with video payload data (packets) to avoid having stalling (re-buffering) events during the playout.

III.3.3.2 sequence: A short decoded audio, video or audiovisual portion of a stream, of 30 s to 60 s duration.

III.4 Abbreviations and acronyms

This appendix 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
DASH	Dynamic Adaptive Streaming over HTTP
GOP	Group Of Pictures
HD	High Definition (television)
HE-AAC	High-Efficiency Advanced Audio Coding
HTTP	Hypertext Transfer Protocol
HVGA	Half Video Graphics Array
I-	Inline-(frame)

MOS	Mean Opinion Score
MPEG	Moving Pictures Expert Group
NB	Narrowband
NTSC	National Television Systems Committee
PAL	Phase Alternating Line
PCC	Pearson Correlation Coefficient
QCIF	Quarter Common Intermediate Format
QoE	Quality of Experience
QVGA	Quarter Video Graphics Array
RMSE	Root Mean Square Error
SD	Standard Definition
TS	Transport Stream
UDP	User Datagram Protocol
VSP	Visual Simple Profile
WB	Wideband

III.5 Conventions

None.

III.6 Areas of application

The application area for using ITU-T P.1201 for non-adaptive, progressive download type media streaming is:

- Progressive download streaming (TCP-based, non-adaptive), which includes:
  - Both so called over-the-top (OTT) services (for example YouTube), and operator managed video services (over TCP);
  - Video over both mobile and fixed connections;
  - The protocols HTTP/TCP/IP and RTMP/TCP/IP;
  - The Appendix III model is intended to be used for video services typically using container formats such as Flash (FLV), MP4, WebM and 3GP. Note that the model is agnostic to the type of container format.

III.7 Building blocks

The model layout is depicted in Figure III.1. The red boxes represent the new building blocks and the blue boxes the building blocks based on the ITU-T P.1201 Recommendations.

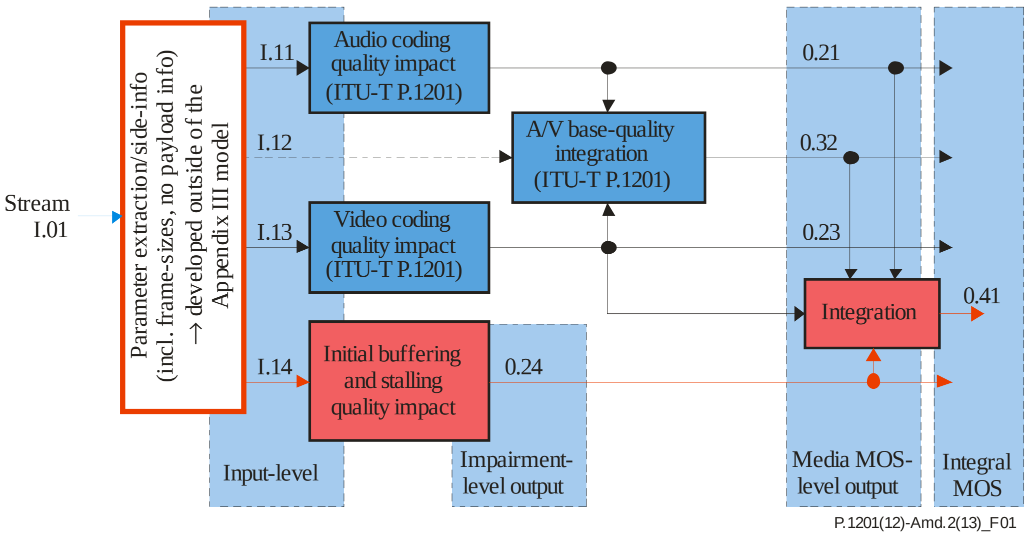
Stream
I.01
Parameter extraction/side-info (incl. frame-sizes, no payload info) → developed outside of the Appendix III model
I.11
I.12
I.13
I.14
Audio coding quality impact (ITU-T P.1201)
A/V base-quality integration (ITU-T P.1201)
Video coding quality impact (ITU-T P.1201)
Initial buffering and stalling quality impact
0.21
0.32
0.23
0.24
0.41
Integration
Input-level
Impairment-level output
Media MOS-level output
Integral MOS
P.1201(12)-Amd.2(13)_F01

Figure III.1 – Building blocks of the ITU-T P.1201 for non-adaptive, progressive download type media streaming. Red blocks are new, and blue are based on the ITU-T P.1201 or ITU-T P.1202 Recommendation building blocks

Note that the parameter extraction is not explicitly included.

III.8 Model input

The model receives media information and prior knowledge about the media stream. In various modes of operation the following inputs may be extracted or estimated in different ways, which is outside the scope of this Recommendation but may be added in future annexes. The core model receives the following input signals regardless of the mode of operation:

I.11: audio coding information, as specified in clause III.8.1

I.12: video resolution, as specified in clause III.8.1

I.13: video coding information, as specified in clause III.8.1

I.14: stalling events as described in clause III.8.2

Note that fault correction techniques, such as ARQ and FEC used for UDP based streaming are not applicable for this case, where the streaming is TCP based. In TCP-based transport all retransmissions and packet loss information is typically handled transparently by the transport layer and while it can be available to the models described in this Recommendation it is not needed. Any packet loss or packet retransmissions are conveyed to the models described in this Recommendation as latency.

III.8.1 Specification of inputs I.11, I.12 and I.13

Table III.4 presents a description of I.11, I.12 and I.13 inputs.

Table III.4 – Description of I.11, I.12 and I.13 inputs

ID	Description	Values	Exemplary name
I.11			
1	Average audio bitrate in kbit/s		AvBitrateAKbit/s (float)
2	Audio codec	– For ITU-T P.1201.1 it is one of: (AAC-LC, AAC-HEv1, AAC-HEv2, AMR-NB, AMR-WB+) – For ITU-T P.1201.2 it is one of: (MPEG1-L2, AC3 AAC-LC, AAC-HEv2)	AudioCodec (string)
I.12 and I.13			
3	Video resolution	– For ITU-T P.1201.1 it is one of: (QCIF, QVGA, HVGA) – Not needed for use with ITU-T P.1201.2	VideoResolution (string)
I.13			
4	Video frame rate in fps		FrameRateFps (float)
5	Video codec and profile	– For ITU-T P.1201.1 it is one of: ITU-T H.264, baseline; MPEG4, vsp – For ITU-T P.1201.2 it is one of: ITU-T H.264, high; ITU-T H.264, main	VideoCodec (string) VideoProfile (string)
6	The type of scanning used for the video	Possible values are: progressive, interlaced	VideoScanType (string)
I.13 per-frame inputs			
7	Video frame number in the encoding order		VideoFrameNum (int)
8	Size of each video picture in the payload in bytes		PicSizeByt (int)
9	Type of each picture	Possible values are "I", "P", "B", "b". "B" stands for reference B frames and "b" for non-reference B frames	PicType (char)

As an example, the inputs I.11, I.12 and I.13 that are not per-frame are provided in text format below:

```
videoCodec          H264
videoCodecProfile   HIGH
videoResolution     HD1080
scanningType        PROGRESSIVE
videoFrameRate      24
audioCodec          AAC-HE v2
audioBitRate        128
```

And an example of the per-frame inputs for I.13 (inputs 7, 8, 9 in Table III.4) is provided below:

```
I,322294
B,119172
B,144588
P,179837
B,121676
B,95552
P,95122
B,64548
B,57143
P,75649
…
```

III.8.2 I.14 input specification

I.14 consists of a list of buffering or stalling events. Each event contains a start time and a duration, both in seconds. The start time is expressed in terms of a "normalized time" which is calculated from the start of the original video sequence (i.e., the sequence without any buffering or stalling). Initial buffering has a start time of 0 and stalling events have a start time greater than zero.

I.14 input information can be provided in a file containing one line per buffering event with each line containing the buffering start time in seconds followed by the buffering duration in seconds. Such a file may contain the following (tab-delimited in this example):

```
0	3.0
2.5	9.8
63.2	2.0
```

In this example there are 3 seconds of initial buffering, and 9.8 seconds of stalling after 2.5 seconds of the video playing and then a further 2 seconds of stalling after 63.2 seconds of the original video has played.

Note that user interactions (such as pausing, seeking, user initiated quality change, user initiated play or user initiated end) are NOT considered at all.

III.9 Model output

The ITU-T P.1201 model for non-adaptive, progressive download type media streaming outputs O.21, O.32, O.23, O.24 and O.41 (see Figure III.1) are the estimated multimedia (audiovisual) MOS (O.32), a separate video score (O.23), a separate audio score (O.21), a stalling and initial buffering degradation score (O.24) and the integral media session score (O.41) as the main model output. All further outputs can serve for diagnostic information on the contributions of individual processing parts on the integral MOS.

Note that outputs O.21, O.23 and O.32 are updated once at the end of the sequence. For the time being the final value of these outputs is the average of all results produced by the underlying ITU-T P.1201 model (if there is more than one result). The averaging method may be modified at a later stage if other methods seem more appropriate.

There is no definition of the update interval for O.24.

Only the final value of O.41 will be evaluated, but it may be updated several times during the measurement. Its final value should reflect the estimated quality over the entire duration of the sequence.

Further optional QoE diagnostics parameters can be defined.

III.9.1 Model output O.21

In the case of ITU-T P.1201.1 (lower resolution application area)

$$O.21 = A\_MOSC \qquad \text{(Eq. 6-35 in [ITU-T P.1201.1])}$$

with:

$$A\_MOSC = 1 + \left(a1 - \frac{a1}{1 + \left(\frac{A\_BR}{a2} \right)^{a3}} \right) \qquad \text{(Eq. 6-36 in [ITU-T P.1201.1])}$$

Where A_BR is the audio bitrate in kbit/s.

Coefficients a1, a2 and a3 depend on the audio codec and are provided in Table III.5.

Table III.5 – Coefficient sets for audio (coding degradations only), from Table 6-21 of ITU-T P.1201.1

	AAC-LC	AAC-HEv1	AAC-HEv2	AMR-NB	AMR-WB+
$a1$	3.36209	3.19135	3.13637	1.33483	3.19158
$a2$	16.46062	4.17393	7.45884	6.42499	5.7193
$a3$	2.08184	1.28241	2.15819	3.49066	1.63208

In the case of ITU-T P.1201.2 (higher resolution application area)

$$O.21 = MOSfromR(QA) \qquad \text{(Eq. 13d in [ITU-T P.1201.2])}$$

with:

$$QA = 100 - QcodA \qquad \text{(Eq. 13c in [ITU-T P.1201.2],}$$

with coding degradations only, i.e., with QtraA = 0)

where:

$$QcodA = a1A \times \exp(a2A \times Bitrate) + a3A \qquad \text{(Eq. 13a in [ITU-T P.1201.2])}$$

Bitrate is the audio bitrate in kbit/s.

The function MOSfromR is given in clause 6.4 of [ITU-T P.1201.2] and is provided below:

```
function MOS = MOSfromR(Q)

set MOS_MAX = 4.9;
set MOS_MIN = 1.05;


if (Q > 0 & Q < 100),
       MOS = (MOS_MIN+(MOS_MAX-MOS_MIN)/100×Q+Q×(Q-60)×(100-Q)×7.0E-6);
elseif (Q >= 100),
      MOS = MOS_MAX;
else
      MOS = MOS_MIN;
end
```

Coefficients a1A, a2A and a3A depend on the audio codec. They are provided in Table III.6:

Table III.6 – Audio model coefficients for different audio codecs (coding degradations only), from Table 1 of ITU-T P.1201.2

Audio codec	a1A	a2A	a3A
MPEG1 L2	100.0	–0.02	15.48
AC3	100.0	–0.03	15.70
AAC-LC	100.0	–0.05	14.60
HE-AAC v2	100.0	–0.11	20.06

III.9.2 Model output O.23

In the case of use of ITU-T P.1201 for non-adaptive, progressive download type media streaming (lower resolution application area)

$$O.23 = V\_MOSC \qquad \text{(clause 6.4.2 in ITU-T P.1201.1)}$$

The computation of V_MOSC depends on `videoFrameRate`, the video frame rate

In pseudo-code:

```
IF videoFrameRate >= 24 THEN
    $V\_MOSC = MOS\_MAX - V\_DC$        (Eq. 6-38 in ITU-T P.1201.1)
ELSE
    $V\_MOSC = (MOS\_MAX - V\_DC) \times \left(1 + v1 \times V\_CCF - v2 \times V\_CCF \times \log\left(\frac{1000}{videoFrameRate}\right)\right)$
        (Eq. 6-39 in [ITU-T P.1201.1])
ENDIF
```

where:

MOS_MAX = 5.0, and *MOS_MIN* = 1.0,

`V_DC` is the video distortion due to compression and is calculated as follows (it is initialized to 0.0):

$$V\_DC = \frac{MOS\_MAX - MOS\_MIN}{1 + \left(\frac{V\_NBR}{v3 \times V\_CCF + v4}\right)^{(v5 \times V\_CCF + v6)}}$$

(Eq. 6-40 in [ITU-T P.1201.1])

`V_NBR` is the normalized video bitrate and is calculated as follows:

$$V\_NBR = \frac{V\_BR \times 8 \times 30}{1000 \times Min(30, videoFrameRate)}$$

(Eq. 6-31 in [ITU-T P.1201.1])

Where `V_BR` is the video bitrate in kbit/s and `videoFrameRate` is the video frame rate.

`V_CCF` is the video content complexity factor. It describes the content's spatio-temporal complexity. The maximum value is 1.0, the initial value is 0.5, and is calculated as shown in the pseudocode below:

```
IF (V_ABIF > 0.0) THEN
    $V\_CCF = Min\left(\sqrt{\frac{V\_BR}{V\_ABIF \times 15.0}}, 1.10\right)$   (Eq. 6-32 in ITU-T P.1201.1)
ENDIF
```

Where V_BR is the video bitrate in kbit/s and V_ABIF is the average number of bytes per I-frame.

*v*1, *v*2, *v*3, *v*4, *v*5, and *v*6 are coefficients as shown in Table III.7.

Table III.7 – Coefficient sets for *V_MOSC* and *V_DC* video quality estimation – Table 6-24 in ITU-T P.1201.1

	ITU-T H.264			MPEG4		
	QCIF	**QVGA**	**HVGA**	**QCIF**	**QVGA**	**HVGA***
v1	3.4	2.49	2.505	2.43	1.6184	1.6184
v2	0.969	0.7094	0.7144	0.692	0.4611	0.4611
v3	104.0	324.0	170.0	0.01	280.0	280.0
v4	1.0	3.3	130.0	134.0	11.0	11.0
v5	0.01	0.5	0.05	0.01	1.69	1.69
v6	1.1	1.2	1.1	1.7	0.02	0.02

* Provisional values, since this condition was not included in the test plan.

In the case of use of ITU-T P.1201 for non-adaptive, progressive download type media streaming (higher resolution application area)

$$O.23 = MOSfromR(QV) \qquad \text{(Eq. 14f in [ITU-T P.1201.2])}$$

with:

$$QV = 100 - QcodV \qquad \text{(Eq. 14e in [ITU-T P.1201.2],}$$

with coding degradations only, i.e., with QtraV = 0),

where:

$$QcodV = a1V \times \exp(a2V \times BitPerPixel) + a3V \times ContentComplexity + a4V$$

(Eq. 14a in [ITU-T P.1201.2])

BitPerPixel is the number of bits per pixel and is computed as follows:

$$BitPerPixel = \frac{Bitrate \times 10^6}{NumPixelPerFrame \times FrameRate}$$

(Eq. 1 in [ITU-T P.1201.2])

where Bitrate is the overall video bitrate (in Mbit/s), NumPixelPerFrame is the number of pixels per frame and FrameRate is the video frame rate.

ContentComplexity captures the impact of the content spatio-temporal complexity in case of no loss. It is given by:

$$ContentComplexity = \frac{\sum_{sc} Nw}{\sum_{sc} S^{I}_{sc} \times Nw} \times \frac{NumPixelPerFrame \times FrameRate}{1000}$$

(Eq. 2 in [ITU-T P.1201.2])

where `NumPixelPerFrame` is the number of pixels per frame and `FrameRate` the video frame.

$\boldsymbol{s}^{I}_{sc}$ is a vector containing the I frame sizes averaged per scene, i.e., $\boldsymbol{s}^{I}_{sc} = (s^{I}_{sc1}, s^{I}_{sc2}, s^{I}_{sc3}, \ldots) \in IR^{S}$, where S is the number of scenes in the measurement window and s^{I}_{sci} the I frame sizes averaged over scene *sci*. The first I frame of the first scene in the measurement window is ignored. The vector length corresponds to the number of scenes in the measurement window. Scene cuts are detected using the pseudo-code provided in clause 6.3.1.7 of [ITU-T P.1201.2]. This pseudo-code is reproduced at the end of this clause.

The function MOSfromR is provided in clause III.9.1.

Moreover, *Nw* is computed as follows: if **N** is a S-dimensional vector containing the number of GOPs per scene and with S being the number of scenes in the measurement window, i.e., $\mathbf{N} = (n_{sc1}, n_{sc2}, n_{sc3}, \ldots) \in IR^{S}$, and if *m* is the index of the scene having the lowest s^{I}_{sci} value, and *s* is the index of the scene then:

$$Nw(s) = \begin{cases} N(s) \times 16, & s = m \\ N(s), & s \neq m \end{cases}$$

(Eq. 2a in [ITU-T P.1201.2])

Note that the number of GOPs for the first scene includes the first GOP of the first scene.

The coefficients a1V, a2V, a3V and a4V are provided for the ITU-T H.264 codec in Table III.8:

Table III.8 – Video model coefficients for the different video resolutions, from Table 2 of ITU-T P.1201.2

Video resolution	a1V	a2V	a3V	a4V
SD (PAL, NTSC)	61.28	–11.00	6.00	6.21
HD (HD1080, HD720)	51.28	–22.00	6.00	6.21

Pseudocode of scene cut detection (clause 6.3.1.7 in [ITU-T P.1201.2])

```
// define thresholds
set I_1 = 1.50;
set I_2 = 0.80;
set I_3 = 1.21;
set I_4 = 0.85;

set P_1 = 0.70;
set P_2 = 1.35;
set P_3 = 0.65;
set P_4 = 1.55;

set b_1 = 0.75;
set b_2 = 1.30;
set b_3 = 0.67;
set b_4 = 1.42;


set SceneCutData.SceneCutNum = 1; //initialize the scene-cut data with first frame
set SceneCutData.SceneCutPos = 1;

for (each I-frame i>2)            // loop over I-frames (starting from the third one)
{
   set Ipos[i] to the frame index of the current I-frame;
```

```
  set P_prev to the number of P-frames in the previous GOP;
  set b_prev to the number of b-frames in the previous GOP;

  set P_curr to the number of P-frames in the current GOP;
  set b_curr to the number of b-frames in the current GOP;

  if (!P_curr)
       continue;

  set Num_P_frames = min(min(P_prev,P_curr), 6);
  set Num_b_frames = min(min(b_prev,b_curr), 6);

          // compute scaling factor of previous I-frame
     set PScaleNum to min(P_prev, 4);
     set Pmedian to the median of the previous PScaleNum P-frames of the previous GOP ;
     set Pmean to the mean of the previous PScaleNum P-frames of the previous GOP;
     set Iscale = Pmedian/Pmean;

  set Ir = VideoFrames[Ipos[i]].FrameSize / (VideoFrames[Ipos[i-1]].FrameSize × Iscale);
  set I_p = I_b = 1.0;

  if (Ir > I_1 || Ir <I_2)
  {
   if (P_prev && Num_P_frames>1)
    {
     set Pmean_prev to the mean of the P_prev P-frames in the previous GOP ;
     set Pmean_curr to the mean of the P_curr P-frames in the current GOP;
     set I_P = Pmean_prev/Pmean_curr;
    }
   if (b_prev && Num_b_frames>1)
    {
     set bmean_prev to the mean of the b_prev b-frames in the previous GOP;
     set bmean_curr to the mean of the previous b_curr b-frames in the current GOP;
     set I_b = bmean_prev/bmean_curr;
    }
     if (  (I_P > P_1) && (I_P < P_2) && (I_b > b_1) && (I_b < b_2) )
       continue;
     else
     {
       SceneCutData.SceneCutPos = Ipos[i];
       SceneCutData.SceneCutNum ++;

     }
  }
  else if (Ir > I_3 || Ir <I_4)
  {
   if (P_prev && Num_P_frames>1)
    {
     set Pmean_prev to the mean of the P_prev P-frames in the previous GOP ;
     set Pmean_curr to the mean of the P_curr P-frames in the curr GOP;
     set I_P = Pmean_prev/Pmean_curr;
    }
   if (b_prev && Num_b_frames>1)
    {

     set bmean_prev to the mean of the b_prev b-frames in the previous GOP;
     set bmean_curr to the mean of the b_curr b-frames in the current GOP;
     set I_b = bmean_prev/bmean_curr;
    }
     if (  (I_P > P_3) && (I_P < P_4) && (I_b > b_3) && (I_b <b_4) )
       continue;
     else
     {
       SceneCutData.SceneCutPos = Ipos[i];
       SceneCutData.SceneCutNum ++;
     }
  }
}
```

III.9.3 Model output O.32

In the case of use of ITU-T P.1201 for non-adaptive, progressive download type media streaming, the output O.32 or the "A/V base-quality integration (P.1201)" block is computed as follows:

a) From ITU-T P.1201.1 (lower resolution application area):

$$O.32 = AV\_MOSC \qquad \text{(Eq. 6-55 of [ITU-T P.1201.1])}$$

with:

$$AV\_MOSC = av1 \times V\_MOSC + av2 \times A\_MOSC + av3 \times V\_MOSC \times A\_MOSC + av4$$

(Eq. 6-46 of [ITU-T P.1201.1))

and where *av*1, *av*2, *av*3 and *av*4 are coefficients as shown in Table III.9:

Table III.9 – Coefficients sets for AV_MOSC audiovisual quality estimation – Table 6-30 of ITU-T P.1201.1

	QCIF	QVGA	HVGA
*av*1	0.7977	0.7495	0.6419
*av*2	0.03732	0.09736	0.1362
*av*3	0.02472	0.006725	0.016
*av*4	0.1657	0.3186	0.5694

b) From [ITU-T P.1201.2] (higher resolution application area):

$$O.32 = MOSfromR(QAV) \qquad \text{(Eq. 15d of [ITU-T P.1201.2])}$$

with:

$$QAV = r + s \times QcodA + t \times QcodV + u \times QcodA \times QcodV \qquad \text{(Eq. 15c1 of [ITU-T P.1201.2))}$$

where Equation 15c1 is computed from Equations 15a, 15b and 15c of [ITU-T P.1201.2], with QtraA = QtraV = 0, since there is no packet-loss in the case where ITU-T P.1201 is used for progressive download type media streaming.

The function MOSfromR is provided in clause III.9.1.

The coefficients r, s, t and u are computed from the ITU-T P.1201.2 audiovisual model coefficients (Table III.11) using eEquations 15a, 15b and 15c. Coefficients are provided in Table III.10.

Table III.10 – Audiovisual model coefficients

r	s	t	U
100.8670	–0.3590	–0.9210	0.00135

The same set of four coefficients is used for all higher video resolutions (SD, HD720, HD1080).

III.9.4 Model output O.24

This part of the Appendix III model is based on the stalling and initial buffering quality model proposed in Appendix I of [ITU-T P.1201].

Here, *DegStall* is the quality impact due to stalling events occurring during media playout. Initial buffering and respective delay before the playout starts in the beginning of a media sequence are handled below by the quality impact due to initial loading quality impact *DegT0*.

The quality impact due to stalling *DegStall* can be calculated as:

$$DegStall = \max\left(\min\left(s_4 + s_1 \times \exp\left(\left(s_2 \times L + s_3\right) \times N\right), 4\right), 0\right) \qquad \text{(III-1)}$$

L is the averaged stalling duration in sec (Note that this does not include initial buffering).

N is the number of stalling events (Note that this does not include initial buffering).

DegT0 is the quality impact due to initial loading. It is expressed as:

if *T0* > 1 – d_2,

$$DegT0 = \max\left(\min\left(d_1 \times \lg\left(T_0 + d_2\right), 4\right), 0\right) \qquad \text{(III-2)}$$

else

$$DegT0 = 0 \qquad \text{(III-3)}$$

end

T0 is the initial loading time in sec.

The following coefficients are recommended for Equations III.1 to III.3:

Table III.11 – Coefficient values for Equations III.1 to III.3

s1	s2	s3	s4	d1	d2
–1.72	–0.04	–0.36	1.66	0.29	–3.29

The output O.24, the buffer-related perceptual indicator *PBufInd*, is presented on a 1 to 5 point scale, as are all other outputs O.21, O.32, O.23 and O.41.

PBufInd is calculated as:

$$O.24 = PBufInd = 5 - \max\left(\min\left(\left(DegStall + DegT0\right), 4\right), 0\right) \qquad \text{(III-4)}$$

III.9.5 Model output O.41

The final output of the appendix III model is the overall MOS-score for the respective 30 s to 60 s long media session, Q_{ms}. Q_{ms} is computed as follows from the intermediate outputs O.32 and O.24 using:

$$O.24 = Q_{ms} = \max\left(\min\left(\left(O.32 - 5 + O.24\right), 5\right), 1\right) \qquad \text{(III-5)}$$

III.10 Diagnostic information

Complementary to the overall session quality score Q_{ms} = O.41, the Appendix III model can provide additional diagnostic information on the cause of possible quality problems.

At the highest levels, initial identification of problems can be related with the output indicators O.21 (audio quality indicator), O.23 (video quality indicator), O.32 (audiovisual quality indicator) and O.24 (perceptual buffering-related indicator). Here, a lower value of any of these scores provides an indication of where the cause of low overall quality may lie.

At lower levels, the input information to the respective modules depicted in Figure III.1 can be used for further identifying the cause of why the outputs O.21, O.23, O.32 and O.24 are low.

Table III.12 – Overview of diagnostic information that can be used for diagnosing quality problems in relation with the appendix scores

Output	Diagnostic parameter	Level	Information
O.21: Audio quality indicator	Audio quality indicator	Indicator level	Information on contribution of audio quality to overall media session quality score
	See ITU-T P.1201.1 App. I, and ITU-T P.1201.2 App. I	Input parameter level	Diagnostic information regarding possible audio quality problems
O.23: Video quality indicator	Video quality indicator	Indicator level	Information on contribution of video quality to overall media session quality score
	See ITU-T P.1201.2 App. I and ITU-T P.1201.1 App. I	Input parameter level	Diagnostic information regarding possible video quality problems
O.32: Audiovisual quality indicator	Audiovisual quality indicator	Indicator level	Information on contribution of audiovisual quality to overall media session quality score
O.24: Buffer-related perceptual indicator	Buffer-related perceptual indicator	Indicator level	Information on contribution of buffer-related perceptual indicator to overall media session quality score
	DegStall	Intermediate level	Stalling degradation
	DegT0	Intermediate level	Initial buffering degradation
	L	Input parameter level	Average duration of stalling events (in seconds)
	N	Input parameter level	Number of stalling events
	T0	Input parameter level	Initial buffering duration (in seconds)

III.11 Performance analysis

The performance of the appendix III model has been evaluated during a cross-validation and training procedure on a database of five quality tests conducted during the development of this model.

The following databases have been created:

Table III.13 – Test databases used for training and cross-validation of this appendix[a]

	Resolution	Sequence	#PAVS
PD01	HD1080	30 s-45 s	68
PD02	HVGA	30 s-45 s	64
PD03	HVGA	30 s-45 s	64
PD04	HD1080	45 s-1 min	56
PD05	HVGA	45 s-1 min	48

[a] PAVS refers to the processed audiovisual sequences used in the respective test.

As for the performance evaluation of the ITU-T P.1201.1, ITU-T P.1201.2, ITU-T P.1202.1 and ITU-T P.1202.2 models, the RMSE defined in [ITU-T P.1401] is computed as a model performance indicator. The Pearson correlation coefficient ('r' in Table III.14 and Figure III.2) is also provided as additional information. These performance indicators have been calculated applying the model to the entire database-set and applying a 1st order mapping per database on the objective scores using the subjective test results as target value according to [ITU-T P.1401] and to compensate for potential biases between subjective test databases.

Table III.14 – Appendix III model performance results

r	RMSE
0.89	0.44

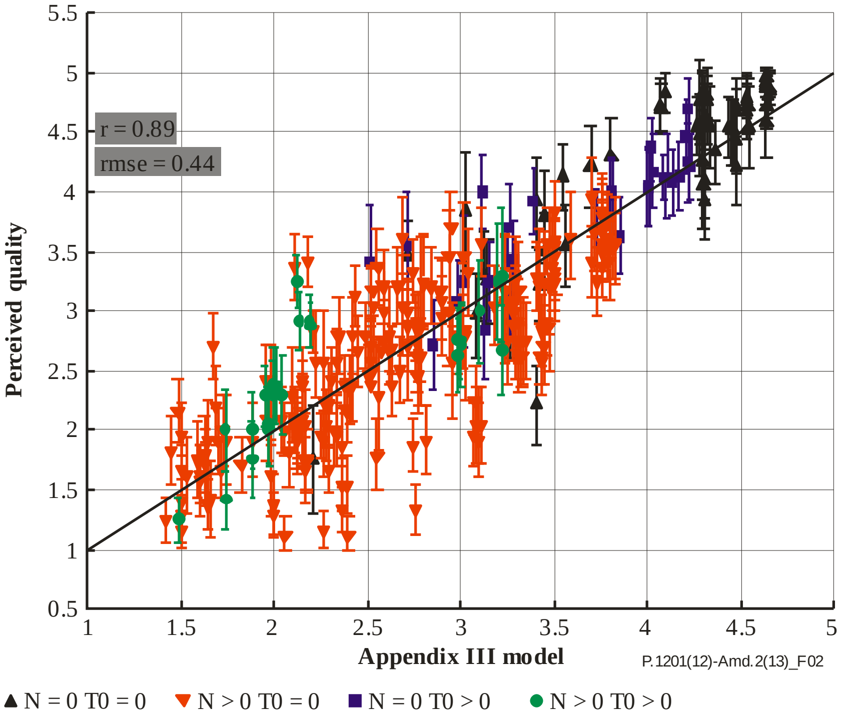

Figure III.2 – Model performance (*N*: number of stalling events; *T0*: initial loading duration in sec)

Bibliography

[b-Hossfeld et al.] Hossfeld, T., Schatz, R., Biersack, E., and Plissonneau, L. (2013), Internet Video Delivery in YouTube: From Traffic Measurements to Quality of Experience. In: Biersack, E. Callegari, C. & Matijasevic, M., eds. *Data Traffic Monitoring and Analysis, Lecture Notes in Computer Science*. Springer, Berlin/Heidelberg, pp. 264-301.

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, 2014