

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.1204.4

(07/2022)

SERIES P: TELEPHONE TRANSMISSION QUALITY,
TELEPHONE INSTALLATIONS, LOCAL LINE
NETWORKS

Models and tools for quality assessment of streamed
media

**Video quality assessment of streaming services
over reliable transport for resolutions up to 4K
with access to full and reduced reference pixel
information**

Recommendation ITU-T P.1204.4

ITU-T P-SERIES RECOMMENDATIONS

TELEPHONE TRANSMISSION QUALITY, TELEPHONE INSTALLATIONS, LOCAL LINE NETWORKS

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	P.10–P.19
Voice terminal characteristics	P.30–P.39
Reference systems	P.40–P.49
Objective measuring apparatus	P.50–P.59
Objective electro-acoustical measurements	P.60–P.69
Measurements related to speech loudness	P.70–P.79
Methods for objective and subjective assessment of speech quality	P.80–P.89
Voice terminal characteristics	P.300–P.399
Objective measuring apparatus	P.500–P.599
Measurements related to speech loudness	P.700–P.709
Methods for objective and subjective assessment of speech and video quality	P.800–P.899
Audiovisual quality in multimedia services	P.900–P.999
Transmission performance and QoS aspects of IP end-points	P.1000–P.1099
Communications involving vehicles	P.1100–P.1199
Models and tools for quality assessment of streamed media	P.1200–P.1299
Telemeeting assessment	P.1300–P.1399
Statistical analysis, evaluation and reporting guidelines of quality measurements	P.1400–P.1499
Methods for objective and subjective assessment of quality of services other than speech and video	P.1500–P.1599

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

Recommendation ITU-T P.1204.4

Video quality assessment of streaming services over reliable transport for resolutions up to 4K with access to full and reduced reference pixel information

Summary

Recommendation ITU-T P.1204.4 describes the reduced-reference and full-reference video quality estimation model for Recommendation ITU-T P.1204 for monitoring the video quality for streaming using reliable transport (e.g., hypertext transfer protocol- (HTTP-) based adaptive streaming (HAS) over the transmission control protocol (TCP), and quick user datagram protocol Internet connections (QUIC)). The estimate is validated for videos encoded with H.264, H.265, video payload type 9 (VP9), or AOMedia Video 1 (AV1) codecs at any resolution up to 4K/ultra-high definition (UHD) ($3\,840 \times 2\,160$) resolution for personal computer (PC) monitors and television (TV) and up to $2\,560 \times 1\,440$ for smartphone and tablet displays.

The ITU-T P.1204 series of Recommendations provides sequence-related (between 5 s and 10 s) and per-1-second video-quality estimation. In principle, the per-one-second outputs of these video-quality models can be used together with an audio model for integration into audiovisual quality and, together with information about initial loading delay and media playout stalling events, further into a final per-session model output, an estimate of integral per-session quality (see e.g., Recommendations ITU-T P.1203, ITU-T P.1203.2, ITU-T P.1203.3).

Recommendation ITU-T P.1204.4 was developed in collaboration with the Video Quality Experts Group (VQEG).

This ITU-T P.1204-series of Recommendations addresses three application areas:

- large-screen presentation as with fixed-network video streaming;
- mobile streaming on handheld devices such as smartphones;
- presentation on tablet-type devices.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T P.1204.4	2020-01-13	12	11.1002/1000/14157
1.1	ITU-T P.1204.4 (2020) Amd. 1	2021-01-07	12	11.1002/1000/14592
2.0	ITU-T P.1204.4	2022-07-29	12	11.1002/1000/15007

Keywords

Adaptive streaming, IPTV, mean opinion score (MOS), mobile video, mobile TV, monitoring, multimedia, OTT, progressive download, QoE, TV, 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/software copyrights, 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 appropriate ITU-T databases available via the ITU-T website at <http://www.itu.int/ITU-T/ipr/>.

© ITU 2022

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

Table of Contents

	Page
1 Scope	1
2 References.....	2
3 Definitions	3
3.1 Terms defined elsewhere	3
3.2 Terms defined in this Recommendation.....	3
4 Abbreviations and acronyms	3
5 Conventions	4
6 Areas of application.....	5
6.1 Application range for the models	5
7 Video module in the ITU-T P.1204 context	7
8 Model input.....	7
9 Model algorithm and output	8
10 Model description	8
10.1 Notation	8
10.2 Multi-resolution frame pyramid	9
10.3 Frame edge representation.....	9
10.4 Local edge statistic	11
10.5 Sharpness.....	12
10.6 Features of reference (side information)	13
10.7 Features of test video.....	13
10.8 Additional features	14
10.9 Step functions	16
10.10 Transformations.....	17
10.11 Core model	17
10.12 Viewing distance	19
10.13 Parameters	19
10.14 Device-dependent model parameters	20
10.15 S-transformation parameters.....	20
Appendix I – Performance figures	22
Appendix II – Long term integration module (Pq) for ITU-T P.1204.4.....	23
II.1 Model input	23
II.2 Model output	23
II.3 Model description.....	24
II.4 Performance figures.....	26
Bibliography.....	28

Recommendation ITU-T P.1204.4

Video quality assessment of streaming services over reliable transport for resolutions up to 4K with access to full and reduced reference pixel information

1 Scope

This Recommendation describes a full-reference/reduced-reference video quality assessment module that together with audio and integration modules can be used to form a complete model to predict the impact of audio and video media encodings and observed Internet protocol (IP) network impairments on quality experienced by the end-user in multimedia streaming applications. The streaming techniques addressed comprise progressive download and adaptive streaming, for both mobile and fixed network streaming applications. The video quality module can also be used stand-alone as a video quality prediction model.

The model described here is a pixel-based reduced-reference model, which can be seen as a special form of a full-reference model. It can be used from monitoring to accurate measurement of client-side quality of experience (QoE) for network or service optimization or benchmarking purposes. The model developed may be deployed both in end-point locations and at mid-network monitoring points, as long as the reference information can be made available at that location. The model in this Recommendation can also be used for laboratory testing of video systems.

The model described here is applicable to progressive download and adaptive streaming or other streaming applications with reliable transport, where the quality experienced by the end user is affected by video degradations due to coding, spatial re-scaling or variations in video frame rates. Quality assessment of adaptive streaming includes aspects of media adaptation that may be handled in integration modules such as [ITU-T P.1203.3], and not in the video modules in this Recommendation. This Recommendation is able to handle various video codecs (i.e., H.264, H.265/high-efficiency video coding (HEVC), video payload type 9 (VP9), and AV1, resolutions up to 4K/ ultra-high definition-1 (UHD-1) and frame rates up to 60 frames/s. The video-quality module Pv of [b-ITU-T P.1203], i.e., [ITU-T P.1203.1], only addresses H.264 and full high definition (HD) with up to 30 frames/s.

The model predicts a mean opinion score (MOS) on a five-point absolute category rating (ACR) scale (see [ITU-T P.910]) as an overall video quality MOS (5 s to 10 s). In addition to the overall quality score, this video quality model produces a per-one-second quality score, suitable for diagnostics or integration into an integral quality score for longer sessions (see, for example, [ITU-T P.1203.3] for 1 min to 5 min duration sessions).

The model associated with this Recommendation cannot provide a comprehensive evaluation of the video quality as perceived by an *individual end-user* because the scores reflect the perceived impairments due to coded video media data being transmitted over an IP connection with a certain performance and do not include specific terminal devices or user-specific information. The scores predicted by such a general quality model necessarily reflect *average perceptual quality*.

Effects due to source generations, such as signal noise, video shake, certain colour properties (and other similar video factors) and other impairments related to the payload, are not reflected in the scores computed by this model.

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

- in-service quality monitoring for specific IP-based audiovisual services, as specified in more detail in clause 6.1;
- performance and quality assessment of live networks (including video encoding) considering the effect due to encoding bitrate, encoding resolution and encoding frame rate;



- laboratory testing of video systems;
- benchmarking of different service implementations;
- benchmarking of different encoder implementations;
- evaluation of transcoding solutions.

In particular, targeted applications are progressive download streaming and adaptive streaming (using reliable transport), which includes the following.

- Over-the-top (OTT) services, as well as operator-managed video services (over the transmission control protocol (TCP)).
- Video over both mobile and fixed connections.
- The streaming protocols HTTP live streaming (HLS) or dynamic adaptive streaming over HTTP (DASH) used with the hypertext transfer protocol (HTTP) or HTTP2 over TCP/IP or quick user datagram protocol Internet connections (QUIC), or real-time messaging protocol (RTMP) over TCP/IP. Note that the model is agnostic to the specific application or transport layer protocol, with the exception that it assumes reliable delivery of video packets.
- Video services typically using container formats based on the ISO/IEC base media file format such as Moving Picture Experts Group-4 (MPEG-4) Part 14 (MP4), or other container formats such as audio video interleave (AVI), Matroska video (MKV), WebM, Third Generation Partnership (3GP), and MPEG-2 transport stream (MPEG2-TS). Note that the model is agnostic to the type of container format.

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 (2021), *Advanced video coding for generic audiovisual services*.
- [ITU-T H.265] Recommendation ITU-T H.265 (2021), *High efficiency video coding*.
- [ITU-T P.910] Recommendation ITU-T P.910 (2022), *Subjective video quality assessment methods for multimedia applications*.
- [ITU-T P.1203.1] Recommendation ITU-T P.1203.1 (2019), *Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport – Video quality estimation module*.
- [ITU-T P.1203.3] Recommendation ITU-T P.1203.3 (2019), *Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport – Quality integration module*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following term defined elsewhere:

3.1.1 bitstream [ITU-T H.264]: A sequence of bits that forms the representation of coded pictures and associated data forming one or more coded video sequences. Bitstream is a collective term used to refer either to a NAL unit stream or a byte stream.

3.1.2 integral quality [b-ITU-T P.1203]: The quality as perceived by a subject in a subjective test, which corresponds to the scope of this Recommendation. Artefacts presented in the subjective tests typically include a combination of audio compression, video compression, and stalling effects.

3.1.3 mean opinion score (MOS) [b-ITU-T P.1204]: The mean of opinion scores, which are values on a predefined scale that subjects assign to their opinion of the performance of the telephone transmission system used either for conversation or for listening to spoken material.

NOTE – Paraphrased from clause 7 of [b-ITU-T P.800.1].

3.1.4 media adaptation [b-ITU-T P.1203]: Events where the player switches video playback between a known set of media quality levels while adapting to network conditions, by downloading and decoding individual segments in sequence.

3.1.5 media quality level [b-ITU-T P.1203]: A particular encoding setting applied to a video or audio stream.

3.1.6 model, model algorithm [b-ITU-T P.1203]: An algorithm with the purpose of estimating the subjective (perceived) quality of a media sequence.

3.1.7 sequence [b-ITU-T P.1203]: An audiovisual stream composed of multiple non-overlapping segments.

3.1.8 video chunk [b-ITU-T G.1022]: A contiguous set of samples for one track of a video.

3.2 Terms defined in this Recommendation

None.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

3GP	Third Generation Partnership
AAC	Advanced Audio Coding
ACR	Absolute Category Rating
AV1	AOMedia Video 1
AVC	Advanced Video Coding
AVI	Audio Video Interleave
DASH	Dynamic Adaptive Streaming over HTTP
GoP	Group of Pictures
HAS	HTTP-based adaptive streaming
HD	High Definition
HEVC	High-Efficiency Video Coding
HLS	HTTP live streaming

HTTP	Hypertext Transfer Protocol
IP	Internet Protocol
MKV	Matroska Video
MO	Mobile
MOS	Mean Opinion Score
MP4	MPEG-4 Part 14
MPEG	Moving Pictures Expert Group
MPEG2-TS	MPEG-2 Transport Stream
NAL	Network Abstraction Layer
PC	Personal Computer
QoE	Quality of Experience
QUIC	Quick User datagram protocol Internet Connections
RMSE	Root Mean Square Error
RTMP	Real-Time Messaging Protocol
RTP	Real-time Transport Protocol
TA	Tablet
TCP	Transmission Control Protocol
TS	Transport Stream
UDP	User Datagram Protocol
UHD	Ultra-High Definition
VVC	Versatile Video Coding

5 Conventions

This Recommendation uses the following conventions:

- 4K: Video resolution of $4\,096 \times 2\,160$ or $3\,840 \times 2\,160$;
- P_v designates the video quality estimation module (as specified in this Recommendation for the case of full and reduced reference pixel-based prediction, see [b-ITU-T P.1204] for alternative implementations such as bitstream based and hybrid);
- Reliable transport: Reliable delivery with protocols guaranteeing no loss of information.

6 Areas of application

6.1 Application range for the models

Table 1 shows the application range of the model in this Recommendation based on what the model has actually been developed for. Table 2 lists areas where it is not applicable. Table 3 lists test factors and coding technologies for which this Recommendation has been validated for H.264, H.265, VP9, and correspondingly Table 4 lists test factors for the case of AV1 encoding.

Table 1 – Application areas for which this Recommendation is applicable

Areas for which the model is applicable
In-service monitoring of video sent over reliable transport. Both OTT services and operator-managed video services, using reliable delivery with protocols such as HTTP or HTTP2 over TCP/IP or QUIC, or RTMP over TCP/IP. Note that this model is agnostic to the type of container format.
Performance and quality assessment of live networks (including video encoding) considering impairments due to encoding bitrate, encoding resolution, and encoding frame rate.
Laboratory testing of video systems.
Benchmarking of different service implementations.
Benchmarking of different encoder implementations. Note that only the full and reduced reference pixel-based model type can be used for direct benchmarking of this type.
Evaluation of transcoding solutions.

Table 2 – Application areas for which this Recommendation is not applicable

Areas for which the model is not applicable
In-service monitoring of video streaming using unreliable transport (e.g., real-time transport protocol/user datagram protocol (RTP/UDP), where packet loss introduces visible quality degradations.
Evaluation of visual quality of display/device properties.
Evaluation of audio/video sync distortions.
Evaluation of video codecs for which the model is not validated (MPEG-I Part 3 [versatile video coding (VVC)], etc.).
Evaluation of the effects of noise, delay, colour correctness or other content-production-related aspects.

Table 3 – Test factors, and coding technologies for which this Recommendation has been validated

Video test factors for which the model has been validated	
Video content	Movies and movie trailers, sports videos, documentaries, computer-generated graphics/games, etc.
Input video length	The video modules were trained and validated to produce one overall video-quality score for a chunk of ~7-9 s and also provide the per-second scores. Optimal performance for ~8 s. Models are assumed to provide valid overall video-quality estimations for 5-10 s long sequences.
Bitstream container	AVI, MP4, MKV, WebM
Encoder types (and implementation), see Note 1	H.264/advanced video coding (AVC) (libx264), H.265/HEVC (libx265), VP9 (libvpx-vp9), AV1 (libaom-av1)