International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.1204.5

(01/2020)

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 transport and received pixel information

Recommendation ITU-T P.1204.5

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 | 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 |
| | | P.700 |
| Methods for objective and subjective assessment of speech quality | Series | P.80 |
| Methods for objective and subjective assessment of speech and video quality | Series | 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 |
| Methods for objective and subjective assessment of quality of services other than speech and video | Series | P.1500 |

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

# Recommendation ITU-T P.1204.5

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


### Summary

Recommendation ITU-T P.1204.5 describes the hybrid no-reference video quality estimation model 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), quick user datagram protocol internet connections (QUIC)). The estimate is validated for videos encoded with H.264, H.265 or video payload type 9 (VP9) codecs at any resolution up to 4K/ultra-high definition-1 (UHD-1) resolution for personal computer (PC) monitors and television (TV) and up to 2 560 × 1 440 for mobile (MO) and tablet (TA) displays.

The ITU-T P.1204 series of Recommendations provide 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, for example, ITU-T P.1203, ITU-T P.1203.2, ITU-T P.1203.3).

Recommendation ITU-T P.1204.5 was developed in collaboration with the Video Quality Experts Group (VQEG).

The 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.1204.5 | 2020-01-13 | 12 | 11.1002/1000/14158 |

### Keywords

Adaptive streaming, IPTV, mean opinion score, mobile video, mobile TV, monitoring, multimedia, MOS, 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, 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.1204.5

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

### 1 Scope

This Recommendation describes the hybrid no-reference video quality assessment model 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 modules can also be used stand-alone as a video quality prediction model.

The model defined covers the use-case of monitoring video quality where the video payload is fully encrypted and the pixel information is available e.g., from the client side.

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), 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 certain performance and do not include specific terminal device 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.

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 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 (2019), *Advanced video coding for generic audiovisual services*.

[ITU-T H.265] Recommendation ITU-T H.265 (2019), *High efficiency video coding*.

[ITU-T P.910] Recommendation ITU-T P.910 (2008), *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.*

[ITU-T P.1204] Recommendation ITU-T P.1204 (2020), *Video quality assessment of streaming services over reliable transport for resolutions up to 4K.*

## 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)** [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:

ACR Absolute Category Rating

AV1 AOMedia Video 1

AVC Advanced Video Coding

AVI Audio Video Interleave

CC Content Complexity

CRF Constant Rate Factor

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

MOS Mean Opinion Score

MP4 MPEG-4 Part 14

MPEG Moving Pictures Expert Group

MPEG2-TS MPEG-2 Transport Stream

OTT Over The Top

PC Personal Computer

QUIC Quick User datagram protocol Internet Connections

RMSE Root Mean Square Error

RTP Real-time Transport Protocol

RTMP Real-Time Messaging Protocol

TCP Transmission Control Protocol

UDP User Datagram Protocol

UHD Ultra High Definition

VP9 Video Payload type 9

VVC Versatile Video Coding

## 5 Conventions

This Recommendation uses the following conventions:

– 4K: Video resolution of 4 096 × 2 160 or 3 840 × 2 160;

– Pv designates the video quality estimation module (as specified in this Recommendation for the case of hybrid prediction, see [ITU-T P.1204] for alternative implementations such as bitstream based and pixel based);

– Reliable transport: Reliable delivery with protocols guaranteeing no loss of information.

## 6 Areas of application

### 6.1 Application range for the model

Table 1 shows the application range of the model in this Recommendation based on what the model has actually been developed for and Table 2 lists areas where it is not applicable. Table 3 lists test factors and coding technologies for which this Recommendation has been validated.

**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. |

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

| **Ares 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 (AOMedia Video 1 (AV1), 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**

<table>
<tr><th colspan="5">Video test factors for which the model has been validated</th></tr>
<tr><td>Video content</td><td colspan="4">Movies and movie trailers, sports videos, documentaries, computer-generated graphics/games, etc.</td></tr>
<tr><td>Input video length</td><td colspan="4">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.</td></tr>
<tr><td>Bitstream Container</td><td colspan="4">AVI, MP4, MKV, WebM</td></tr>
<tr><td>Encoder types (and implementation, see Note 1)</td><td colspan="4">H.264/advanced video coding (AVC) (libx264), H.265/HEVC (libx265), VP9 (libvpx-vp9)</td></tr>
<tr><td>Encoder profiles</td><td colspan="4">H.264 (MPEG-4 Part 10): Constrained baseline, Main, Hi, Hi10, Hi422.<br>H.265: Main, Main10, range extension (Rext).<br>VP9: 0, 1, 2, 3.</td></tr>
<tr><td rowspan="5">Video resolution and bitrate</td><td>Resolution definition</td><td>Video height range</td><td>Personal computer/ television (PC/TV)</td><td>Mobile/tablet (MO/TA)</td></tr>
<tr><td>Below SD</td><td>180-270</td><td>–</td><td>90 Kbps-1 Mbps</td></tr>
<tr><td>SD</td><td>360-540</td><td>150 Kbps-4 Mbps</td><td>150 Kbps-4 Mbps</td></tr>
<tr><td>HD</td><td>720-1 080</td><td>500 Kbps-15 Mbps</td><td>500 Kbps-15 Mbps</td></tr>
<tr><td>Above HD</td><td>1 440-2 160</td><td>1.5 Mbps-45 Mbps</td><td>1.5 Mbps-20 Mbps</td></tr>
<tr><td>Video aspect ratio</td><td colspan="4">16:9, see Note 2</td></tr>
<tr><td>Group of pictures (GoP)</td><td colspan="4">Variable. Average GoP length can be between 0.5 s and chunk duration.</td></tr>
<tr><td>Bit-depth</td><td colspan="4">8 bits or 10 bits</td></tr>
<tr><td>Chroma subsampling</td><td colspan="4">YUV 4:2:0 and YUV 4:2:2</td></tr>
<tr><td>OTTs</td><td colspan="4">Online providers that offer Video on Demand and video encoding as a service. It should be noted that the models are applicable for similar OTTs.</td></tr>
<tr><td>Display resolution and frame rate</td><td colspan="4">PC/TV; 2160p, up to 60 frames/s.<br>MO/TA: 1440p, up to 60 frames/s.</td></tr>
<tr><td>Viewing distances</td><td colspan="4">PC/TV: 1.5H to 3H (H: Screen height), see Note 3<br>MO/TA: 4H to 6H</td></tr>
</table>

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

| Video test factors for which the model has been validated |
|---|
| NOTE 1 – During training and validation, FFmpeg 3.2.2 was used with x264 snapshot 20170202-2245, x265 v2.2, libvpx 1.6.1. |
| NOTE 2 – For original content with a larger aspect ratio, letterboxing of up to 30% was allowed, that is 1512 pixels height for video coded at 2160 pixels height. Video content with 1.89:1 aspect ratio (e.g., cinema 4K) may also be used. |
| NOTE 3 – It is noted that for PC/MO, the model output is conservative and should be interpreted to correspond to a viewing distance of 1.5*H* to 1.6*H*. |

## 7 Building blocks

The module layout of the ITU-T P.1204 model is depicted in Figure 1.

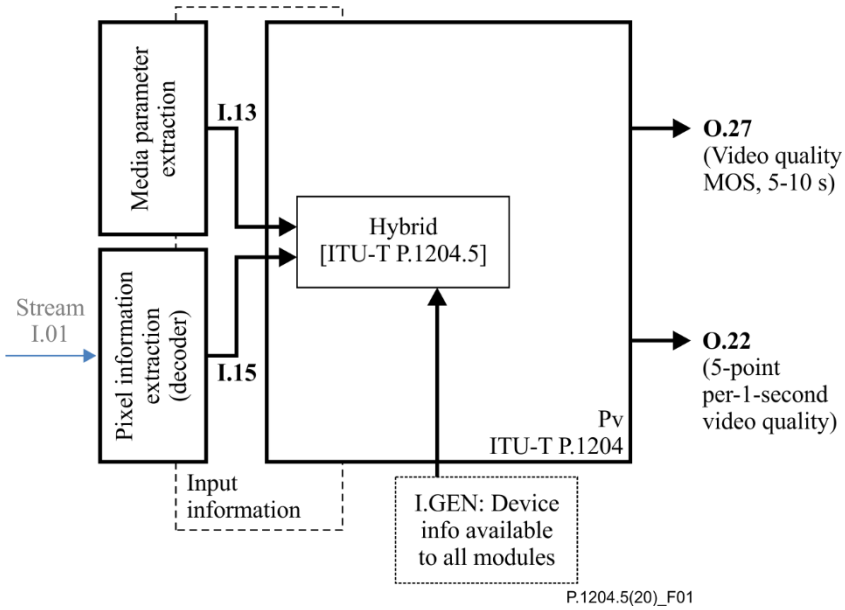


**Figure 1 – Building blocks of the ITU-T P.1204 video quality model used stand-alone for the case of this Recommendation**

### 7.1 Model input interfaces

The ITU-T P.1204 model will receive media information and prior knowledge about the media stream or streams. For the different types of model, 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 model receives the following input information, depending on its type (for details see Table 4):

**I.GEN** Display resolution and device type. The device type is defined as follows:

- PC/TV: screen size 24 inch or larger and less than or equal to 100 inch;
- MO/TA: screen size 13 inch or smaller.

**I.13** Video coding information.

**I.15** Degraded video pixel information.

## 7.2 Specification of inputs I.GEN, I.13 and I.15

**Table 4 – I.GEN, I.13 and I.15 inputs description**

| ID | Description | Values | Frequency | Notation |
|---|---|---|---|---|
| ***I.GEN*** | | | | |
| 0 | The resolution of the image displayed to the user | Number of pixels ($W \times H$) in displayed video | Per media chunk | *disRes* |
| 1 | The device type on which the media is played | "PC", "TV", "MO", "TA" | Per media chunk | *device* |
| ***I.13*** | | | | |
| 4 | Video bitrate | Bitrate in kilobits per second | Per media chunk | *bitrate* |
| 5 | Video frame rate | Frame rate in frames per second | Per media chunk | *framerate* |
| 6 | Segment duration | Duration in seconds | Per media chunk | *duration* |
| 7 | Video encoding resolution | Number of pixels ($W \times H$) in transmitted video | Per media chunk | *codRes* |
| 8 | Video codec and profile | H.264 (MPEG-4 Part 10): Constrained baseline, Main, Hi, Hi10, Hi422<br>H.265: Main, Main10, Rext<br>VP9: 0, 1, 2, 3 | Per media chunk | *codec*, *codecProfile* |
| ***I.15*** | | | | |
| 26 | Degraded video | The raw pixels (YUV file including metadata required for parsing; width, height, frame rate and pixel format) of the processed video, i.e., the video decoded and upscaled (using the bicubic upscaling method) to display resolution without buffering/stalling. | Per media chunk | *degVid* |

Some examples of the inputs:

*disRes*: The video display resolution in number of pixels.

For instance, PC monitor and TV UHD screens have a video display resolution of:

$$disRes = 3840 \times 2160 = 8294400\ pixels$$

For instance, MO/TA screens have a video display resolution of:

$$disRes = 2560 \times 1440 = 3686400\ pixels$$

*codRes*: The video encoding resolution in pixels. The video resolution used to encode the video, for instance, $codRes = 854 \times 480 = 409920\ pixels$.

## 7.3 Model output information

The model provides one score per second and one overall video quality score for the chunk under consideration.

There should not be any output score for frames at the end of a sequence, when those frames do not add up to a complete second. The quality score is calculated at the closest frame boundary at or after each integer second from the start of the stream.

For all outputs, a quality scale is used, ranging from 1, meaning "bad", to 5, meaning "excellent", as specified in [ITU-T P.910].

The ITU-T P.1204 Pv-model outputs are as follows:

- O.22: Video coding quality per second
  - Per-second scores provided per chunk and on a quality scale of 1 to 5.
- O.27: Final video session quality score
  - Single score for the chunk, on a quality scale of 1 to 5.

## 8 Model algorithm

The video model defined in this Recommendation has one output, O.27. It provides a single output value on the five-point ACR scale (MOS) for a short video. The core model algorithm is described in clause 8.1.

### 8.1 Core model

#### 8.1.1 Relative raw bitrate factor (*relRawBitrateRatio*)

For a parametric model, *bitrate* carries the most important information about the quality of the video. However, *bitrate* only makes sense together with information about the encoded chroma subsampling format (content complexity (CC) estimate). This is because the same video in YUV420 or YUV422 would need a slightly different bitrate in order to be encoded to the same quality. The same is also true for 8 bit or 10 bit reference videos.

*relRawBitrateRatio* is defined as:

$$relRawBitrateRatio = \frac{RawBitrateActual}{RawBitrate8bitYUV420} \tag{1}$$

$$RawBitrate8bitYUV420 = codRes * framerate * 1.5 * 8$$

$RawBitrateActual$ for different CC values:

- (yuv420p) 8 bit YUV420 = $codRes * framerate * 1.5 * 8.0$
- (yuv422p) 8 bit YUV422 = $codRes * framerate * 2.0 * 8.0$
- (yuv420p10le) 10 bit YUV420 = $codRes * framerate * 1.5 * 10.0$
- (yuv422p10le) 10 bit YUV422 = $codRes * framerate * 2.0 * 10.0$

Thus:

$$relRawBitrateRatio = f(CC) = \begin{cases} yuv420p & 1.0 \\ yuv422p & 2.0/1.5 \\ yuv420p10le & 10.0/8.0 \\ yuv422p10le & (10.0 * 2.0)/(8.0 * 1.5) \end{cases} \tag{2}$$

#### 8.1.2 *codecProfile* to chroma subsampling type mapping:

Separate *codecProfile* to chroma subsampling mappings are used for each codec. Note that this information may not be known to the model in all cases, in which case it can use a pre-assumed default chroma subsampling type:

H.264:

$$CC = g(codecProfile) = \begin{cases} ConstrainedBaseline & yuv420p \\ Main & yuv420p \\ Hi & yuv420p \\ Hi10 & yuv420p10le \\ Hi422 & yuv422p \\ others/unknown & yuv422p \end{cases}$$

H.265:

$$CC = g(codecProfile) = \begin{cases} Main & yuv420p \\ Main10 & yuv422p10le \\ Rext & yuv422p \\ others/unknown & yuv422p \end{cases}$$

VP9:

$$CC = g(codecProfile) = \begin{cases} 0 & yuv420p \\ 1 & yuv422p \\ 2 & yuv420p10le \\ 3 & yuv422p10le \\ others/unknown & yuv422p \end{cases}$$

### 8.1.3 Adjusted log bitrate (*logBitrate*)

$$bitrateAdj = bitrate * \exp\left(-h_0 * (relRawBitrateRatio - 1)\right) \quad (3)$$

or

$$bitrateAdj = bitrate * \exp\left(-h_0 * \left(f\left(g(codecProfile)\right) - 1\right)\right) \quad (4)$$

and

$$logBitrate = \log10(bitrateAdj) \quad (5)$$

where $h_0$ is specified in Table 5.

**Table 5 – *logBitrate* coefficients**

| | PC/TV | MO/TA |
|---|---|---|
| | $h_0$ | $h_0$ |
| H.264 | 1.1776641027814067e-09 | 0.5923649958216682 |
| H.265 | 0.1648644781080738 | 0.6286917954823384 |
| VP9 | 1.4370415811329779e-15 | 0.3595185885781488 |

### 8.1.4 Upscaling feature (*scaleFactor*)

*scaleFactor* defines the degree of upscaling after the decoding for display purposes.

$$scaleFactor = \max\left(\frac{disRes}{codRes}, 1\right) \quad (6)$$

The larger the *scaleFactor* value, the greater the upscaling performed at the display device.

### 8.1.5 Temporal feature (*framerateFactor*)

The maximum frame rate the model was tested for is 60 frames/s. The coded frame rate can be less than 60 frames/s. *framerateFactor* is defined as

$$framerateFactor = \max\left(\frac{60.0}{framerate}, 1\right) \quad (7)$$

The larger the *framerateFactor* value, the greater the frame rate up-conversion during playback.

### 8.1.6 Source complexity feature (*contentFactor*)

The model uses a codec-based source complexity estimation method. Using the constant rate factor (CRF) coding recipe of the libvpx-vp9 encoder, *degVid* is encoded at a certain quality *Q* to an encoded file *degVidEncoded*, where *Q* is an unknown quality value resulting from the CRF encoding of the *degVid* at a CRF value of 32.

To determine the *contentFactor*, the degraded video sequence is encoded using the following command line:

```
ffmpeg –i degVid.avi -pix_fmt yuv420p -an -c:v libvpx-vp9 -crf 32 -b:v 0
degVidEncoded.mp4
```

For the static build of the ffmpeg implementation used, see [b-ffmpeg_3.4], which uses the source code [b-libvpx-vp9 v1.6.1].

The *contentFactor* is determined as follows:

$$norm\_crf\_bitrate = \frac{sizebytes(degVidEncoded)*1000}{framerate*T*disRes} \quad (8)$$

where *sizebytes* is the size in bytes including the MP4 overhead.

The source complexity is computed as a function of $norm_{crf_{bitrate}}$

$$srcComplexity = 7.273 * log10(norm\_crf\_bitrate) \quad (9)$$

and

$$contentFactor = c_1 * srcComplexity + c_2 \quad (10)$$

where the constants $c_1$ and $c_2$ are specified in Tables 6 and 7.

**Table 6 – *contentFactor* coefficients for PC /TV**

| | **PC /TV** | |
|---|---|---|
| | $c_1$ | $c_2$ |
| H.264 | 0.026020856130385718 | 0.18771981049276384 |
| H.265 | 0.321901099557003 | –0.9339240842451443 |
| VP9 | 0.027131654431210638 | –0.07758026781152491 |

**Table 7 – *contentFactor* coefficients for MO/TA**

| | **MOS/TA** | |
|---|---|---|
| | $c_1$ | $c_2$ |
| H.264 | 0.03304059217693778 | 0.5191195117506 |
| H.265 | 0.054392293564817444 | –0.4752924970529189 |
| VP9 | 0.01703446988358945 | –0.09703179546863315 |

### 8.1.7 Feature integration

The constants *a*, *b* and *c* are computed as:

$$a = a_0 - a_s * \log10(u_a * (scaleFactor - 1) + 1) - a_f * framerateFactor - a_c * contentFactor \quad (11)$$

$$b = b_0 - b_s * \log10(u_b * (scaleFactor - 1) + 1) + b_f * framerateFactor + b_c * contentFactor \quad (12)$$

$$c = c_0 - c_s * \log10(u_c * (scaleFactor - 1) + 1) - c_f * framerateFactor + c_c * contentFactor \quad (13)$$

with the constraint on a minimum value of $b$ as in Equation 14:

$$b = \max(0, b) \quad (14)$$

The three computed constants along with another constant $k_0$ can be combined to a score value $S$ using Equation 15.

$$S = a * \left(\frac{1-\exp(-k_0*(logBitrate-c))}{1+\exp(-b*(logBitrate-c))}\right) \quad (15)$$

All the model constants used in Equation 11 and constants used in the derivation of $a$, $b$ and $c$ are specified in Tables 8 and 9.

**Table 8 – Feature Integration constants, PC /TV**

| Constant | PC /TV | | |
|---|---|---|---|
| | **H.264** | **H.265** | **VP9** |
| $a_0$ | 5.677728847992967 | 5.03853891104581 | 4.859699233665362 |
| $b_0$ | 3.4712005807048745 | 2.0993542290664227 | 2.6541304260526557 |
| $c_0$ | 2.326478357956036 | 2.8334365643929855 | 2.9399953618001136 |
| $a_s$ | 1.8350235211981674 | 2.558825165003877 | 2.3476224402785877 |
| $b_s$ | 1.4141232302855393 | 0.5098792603744106 | 7.255415776808229e-11 |
| $c_s$ | 0.23475280755478767 | 0.22681818096833914 | 0.2873320369663877 |
| $u_a$ | 0.1778191362520981 | 0.08444039691348859 | 0.12643591444328875 |
| $u_b$ | 0.156900730863524 | 1.5410279574057658e-36 | 0.004818194829532265 |
| $u_c$ | 42.406080941967936 | 2.0059093997172757 | 2.0509739990614357 |
| $a_f$ | 0.39159165912177857 | 0.2525211972777661 | 0.15581905716465846 |
| $b_f$ | 2.6729710558144443e-28 | 2.6688343545615205e-21 | 6.690412679884795e-15 |
| $c_f$ | 0.29490002469830306 | 0.21402618037698756 | 0.20483793964560515 |
| $a_c$ | 1.6943267545826664e-13 | 0.0431077938951142 | 1.668359219633742e-14 |
| $b_c$ | 7.0362956885089e-14 | 0.43792733573736864 | 4.093588017285955 |
| $c_c$ | 3.678498383915767 | 0.358852205906036 | 4.3023537324911105 |
| $k_0$ | 1.4419774585129321 | 2.9400708635994275 | 2.9195734718894553 |

**Table 9 – Feature integration constants, MO/TA**

| Constant | MO/TA | | |
|---|---|---|---|
| | **H.264** | **H.265** | **VP9** |
| $a_0$ | 5.268960765324393 | 5.0474497689434275 | 4.984684538764142 |
| $b_0$ | 3.970252547227931 | 1.26707140012788e-21 | 5.2136891589367425 |
| $c_0$ | 0.955861731604233 | 2.884571319491612 | 2.7840703793378223 |
| $a_s$ | 4.36888019813821 | 3.0455666232932663 | 5.803265994082781 |
| $b_s$ | 2.1125548778844156 | 0.00017290708274250087 | 1.4701594292800126 |
| $c_s$ | 0.40383887688983744 | 0.10996363240734348 | 0.21040175571457492 |

**Table 9 – Feature integration constants, MO/TA**

| Constant | MO/TA | | |
|---|---|---|---|
| | **H.264** | **H.265** | **VP9** |
| $u_a$ | 0.024553971967259326 | 0.04988189636286348 | 0.01833878302910475 |
| $u_b$ | 0.5557309759968077 | 5.020735385579775 | 25.189492746842372 |
| $u_c$ | 1.4393665855340954 | 3.351799514986455 | 4.425914043223159 |
| $a_f$ | 0.23654971807507216 | 0.2118845114345596 | 0.20658178681704242 |
| $b_f$ | 8.69531265907939e-37 | 3.1098630749524796 | 0.9720701616151223 |
| $c_f$ | 0.19146906019485413 | 0.1515064042031239 | 0.14910953368910074 |
| $a_c$ | 0.26458342387745737 | 7.844661892720165e-36 | 1.9881820627248652e-24 |
| $b_c$ | 1.4427813426296531e-33 | 1.5165682395521835e-10 | 0.0017425312678303107 |
| $c_c$ | 2.953357298372877 | 2.0316300541234864 | 6.80531487679437 |
| $k_0$ | 2.7475799851849545 | 2.20751587008015 | 2.5709237715026094 |

### 8.1.8 Device-based linear mapping

A final linear mapping is used to map the *S* score to the predicted MOS:

$$\widehat{MOS} = \min(\max(m_1 * S + m_2, 1), 5) \quad (16)$$

where $m_1$ and $m_2$ are specified in Table 10.

**Table 10 – Device-based linear mapping coefficients**

| | $m_1$ | $m_2$ |
|---|---|---|
| PC monitor | 0.967 | 0.153 |
| TV | 1.051 | –0.187 |
| MO | 0.942 | 0.146 |
| TA | 1.080 | –0.330 |

# Appendix I

## Performance figures

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

In this section, the root mean square errors (RMSEs) of Pv models are reported. Note that the numbers are reported after a final per-database mapping between the model output and the subjective scores of a database. This linear mapping is used to account for scale and bias variations between different databases.

**Table I.1 – Validation performance of Pv model: The *submitted Model* is the model trained on the exchanged training databases and frozen before creation of validation data. Models were retrained using a five-fold cross-validation approach, with their validation performance listed to show stability of the performance indicating no over-fitting**

| | | | | | | |
|---|---|---|---|---|---|---|
| **Hybrid No reference mode 0 model** | *Submitted model* | 0.452 | | | | |
| | *Five-fold cross-validation* | 0.451 | 0.440 | 0.441 | 0.443 | 0.441 |

- All training and validation databases were merged to obtain in total 26 different short databases (18 PC/TV and 8 MO/TA).
- A level of difficulty of prediction for each database was determined based on average prediction error over all models.
- A 50:50 training-validation split was determined randomly, but respecting the level of difficulty. In total, five different splits were defined. Each split had a balanced distribution of databases based on difficulty in both the training and validation.
- The 50:50 split was separately performed for PC/TV and MO/TA cases.
- The final model coefficients correspond to the best performing split.

The final selected model is the model from cross-validation set 2 (RMSE 0.440). This is because the model resulting from this split has the best overall performance.

# Bibliography


[b-ITU-T G.1022] Recommendation ITU-T G.1022 (2016), *Buffer models for media streams on TCP transport.*

[b-ITU-T P.800.1] Recommendation ITU-T P.800.1 (2016), *Mean opinion score (MOS) terminology*.

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

[b-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*.

[b-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*.

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

[b-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*.

[b-ITU-T P.1203] Recommendation ITU-T P.1203 (2017), *Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport*.

[b-ITU-T P.1203.2] Recommendation ITU-T P.1203.2 (2017), *Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport – Audio quality estimation module.*

[b-ITU-T P.1204.3] Recommendation ITU-T P.1204.3 (2020), *Video quality assessment of streaming services over reliable transport for resolutions up to 4K with access to full bitstream information.*

[b-ITU-T P.1204.4] Recommendation ITU-T P.1204.4 (2020), *Video quality assessment of streaming services over reliable transport for resolutions up to 4K with access to full and reduced reference pixel information.*

[b-ITU-T P.1401] Recommendation ITU-T P.1401 (2020), *Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models*.

[b-ffmpeg_3.4] FFmpeg developers (2000-2017). ffmpeg version 3.4, built with gcc 7.2.0 (GCC). Available [viewed 2020-02-28] at: https://ffmpeg.zeranoe.com/builds/win64/static/ffmpeg-3.4-win64-static.zip

[b-libvpx-vp9 v1.6.1] Github (2020). libvpx-vp9 v1.6.1 source code. Available [viewed 2020-02-28] at: https://github.com/webmproject/libvpx/releases/tag/v1.6.1.

## SERIES OF ITU-T RECOMMENDATIONS

| | |
|---|---|
| Series A | Organization of the work of ITU-T |
| Series D | Tariff and accounting principles and international telecommunication/ICT economic and policy issues |
| 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 | Environment and ICTs, climate change, e-waste, energy efficiency; 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** | **Telephone transmission quality, telephone installations, local line networks** |
| Series Q | Switching and signalling, and associated measurements and tests |
| 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, next-generation networks, Internet of Things and smart cities |
| Series Z | Languages and general software aspects for telecommunication systems |

Printed in Switzerland
Geneva, 2020