International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.1204.5

Amendment 1

(01/2021)

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

Amendment 1: New Appendix II: Long term integration module (Pq) for ITU-T P.1204.5

Recommendation ITU-T P.1204.5 (2020) – Amendment 1

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.1204.5

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

Amendment 1

New Appendix II: Long term integration module (Pq) for ITU-T P.1204.5

Summary

Amendment 1 to Recommendation ITU-T P.1204.5 introduces Appendix II, which describes a long term integration module for ITU-T P.1204.5.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T P.1204.5	2020-01-13	12	11.1002/1000/14158
1.1	ITU-T P.1204.5 (2020) Amd. 1	2021-01-07	12	11.1002/1000/14593

* 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/.

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

Amendment 1

New Appendix II: Long term integration module (Pq) for ITU-T P.1204.5

1) *Add the following appendix after Appendix I.*

Appendix II

Long term integration module (Pq) for ITU-T P.1204.5

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

The current approach for long-term integration described in this appendix is to be considered as an intermediate solution. It has been developed based on a total of 6 databases (2 for training and 4 for validation), while the short-term video quality model (Pv) described in the normative part of this Recommendation was developed using 26 databases (13 for training and 13 for validation).

As a consequence of the relatively low number of long-sequence databases used, the integration module presented in this appendix does not form an integral part of the Recommendation, and is considered for information. It is planned to be superseded in the future by a more comprehensive integration module trained and validated on a higher number of databases, which is currently under development.

The streaming parameters ranges used in the training and validation tests of the long-term integration model presented in this appendix are summarized in Table II.1.

Table II.1 – Parameters range used in the tests for the development ITU-T P.1204.5 long-term integration model

Video sequence duration	60 seconds – 5 minutes
Initial loading delay	0-30 seconds
Total stalling duration	0-26 seconds
Number of stalling events	0-5
Total number of quality level switches	0-39

The ranges of parameter settings of other test factors related to each of the segments in a video session are summarized in Table 3 of ITU-T P.1204.5. It should be noted that audio quality was not varied in any of the 6 tests used to train and validate the long-term integration model presented in this appendix. For the PC/TV case, the audio from the source video was encoded using 16-bit PCM with 2 channels and 48 kHz sampling frequency. For the case of mobile/tablet, AAC codec was used for audio with a bit rate of 512 kbits/s.

II.1 Model input

The model must receive the following input signals regardless of the mode of operation:

– O.21: audio quality per output sampling interval, as specified in clause 7 of [ITU-T P.1204].

– O.22: video quality per output sampling interval, as specified in ITU-T P.1204.5 – see clause 7.3.

– I.14: stalling events, as described in [ITU-T P.1204] clause 7.

– I.GEN: device type (either of "PC", "TV", "Mobile" or "Tablet"), as specified in [ITU-T P.1204].

II.2 Model output

The Pq model outputs the following information:

– O.23: perceptual buffering indication

– O.34: audiovisual segment coding quality per output sampling interval

– O.35: final audiovisual coding quality score

– O.46: final media session quality score

The values O.23, O.35 and O.46 will be output once per session.

The value O.34 will be output once per output sampling interval.

II.3 Model description

II.3.1 Parameters

1 Initial-loading time in seconds – $initialLoadingLen$. This can be extracted from I.14, i.e., the stalling duration value at playout time 0. If not specified in I.14 the default value is 0.0.

2 Total stalling length in seconds – $totalBuffLen$. This can be computed from I.14, i.e., sum of all stalling values not including $initialLoadingLen$. If not specified in I.14 the default value is 0.

3 Number of stalling – $numStalls$. This can be computed from I.14, i.e., number of all stalling events (not including initial-loading event). If not specified in I.14 the default value is 0.

4 Total buffering-free video duration in seconds – T. This is the length of the O.22 list.

5 Time between end of last buffering to the end of video in seconds – $timeSinceLastBuff$. For a stalling-free video this is equal to T.

II.3.2 Computation of O.34

The audiovisual quality score O_i for the *i*th output sampling interval is derived from the linear combination of *i*th O.21 (i.e., a_i) and the *i*th O.22 (i.e., v_i) score as follows

$$O_i = 0.05 * a_i + 0.95 * v_i$$

Note that, it is assumed that audio signal is always present, i.e., O21 and O22 arrays have equal length, and audio is coded with high quality, i.e., audio MOS of 4.5 or above. Let $O.34 = [O_1, O_2, O_3, \dots, O_T]$ denote an array of O.34 scores.

II.3.3 Computation of O.35

1 First a sliding window of size 30 is used with a step size of 1 over the $O.34$ array to yield overlapping pieces of O.34 scores. Each piece contains 30 scores. The window sliding is stopped if the scores contained in the window are less than 30. For each piece a normalized histogram of the O.34 quality scores is computed using the bin boundaries given below:

$$\text{histBins1} = (1.0, 1.5), (1.5, 2.5), (2.5, 3.5), (3.5, 4.5), (4.5, 5.0)$$

Let $qualHistogramList = [h_0, h_1, h_2, \ldots, h_N]$ denotes the array of histograms, where h_i denotes the histogram resulting from the ith shift of the sliding window, $0 \le i \le N$. The pseudo code for computing the histogram can be at the end of this section.

2 A complementary array of quality changes $\Delta O.34 = [\Delta O_1, \Delta O_2, \Delta O_3, \ldots, \Delta O_{T-1}]$ is prepared using the difference between successive O.34 scores, i.e., $\Delta O_i = O_{i+1} - O_i$ for $1 \le i \le T - 1$. Like (1), a sliding window of size 30 is used to create normalized histograms based on $\Delta O.34$ array, using the bin boundaries

$$\text{histBins2} = (-4.5, -3.5), (-3.5, -2.5), (-2.5, -1.5), (-1.5, -0.5), (-0.5, 0.5), (0.5, 4.0)$$

Similar to (1), let $qualChangeHistogramList = [g_0, g_1, g_2, \ldots, g_{N-1}]$ denote the array of quality change histograms. The pseudo code for computing the histogram can be at the end of this section.

3 Let h_{ij} and g_{ij} denote the jth element of histograms of h_i and g_i, respectively. Create an aggregate array $F = [f_0, f_1, f_2, \ldots, f_{N-1}]$ by combining $qualHistogramList$ and $qualChangeHistogramList$ as follows:

$$f_i = \sum_{j=1}^{5} a_j * h_{ij} + \sum_{j=1}^{6} b_j * g_{ij}$$

where $0 \le i \le N - 1$.

Table II.2 – Values of histogram related constants

a_1	1.7036144962372886
a_2	1.6281208003842298
a_3	2.14625868168416
a_4	3.154522195465948
a_5	3.1811440812907144
b_1	-12.892854165904497
b_2	-6.205923716980252
b_3	-2.477111070479436
b_4	-0.9875867258584734
b_5	0.778247340510056
b_6	0.4101562929016858

4 Compute a feature list and linearly weight the features to compute the O.35 score

$$L = [\min(F), \max(F), \text{median}(F), \text{mean}(F), f_{N-1}]$$

$$O.35 = \sum_{j=1}^{5} w_j * l_j$$

where l_j denote the jth feature of the feature list L.

Table II.3 – Values of O.35 related constants

w_1	0.29508584543387967
w_2	0.00146837942360000
w_3	0.00118943982340000
w_4	0.35482926488923905
w_5	0.34742707042988136

Pseudocode for histogram computation

```
function hist = histogram (x, bins):
   h = An array of zeros with size one less than bins
   for every value x_i of x:
      for every bin b of bins:
         h[b] += max(0, 1 - abs((bins[b + 1] + bins[b]) / 2 – x_i))
   hist = normalize each element of h by sum(h)
```

II.3.4 Computation of O.46

First by using a four dimensional exponential function we create an intermediate variable $InitLoadAndStallImpact$ which combines the effect of initial-loading, stalling and recency due to the last occurred stall.

$$InitLoadAndStallImpact = \exp(-s_1 * numStalls) * exp\left(-s_2 * \frac{initialLoadingLen}{T}\right) * \exp\left(-s_3 * \frac{totalBuffLen}{T}\right) * \exp\left(-s_4 * \frac{T - timeSinceLastBuff}{T}\right)$$

Table II.4 – Values of initial-loading and stalling related constants

s_1	0.08768743173928367
s_2	0.7167602031580045
s_3	0.06981494241303295
s_4	0.30959519998764706

$InitLoadAndStallImpact$ is the factor by which the buffering-free quality O.35 is reduced, i.e.,

$$Q = 1 + (O.35 - 1) * InitLoadAndStallImpact$$

The intermediate aggregated quality value Q is mapped to the final O.46 using a linear mapping and a limiting function which constraints the output value between 1 and 5.

$$\text{O}.46 = \min(5.0, \max(1.0, m * \text{Q} + c))$$

where constants m and c depend on the display device type – see the Table below.

Table II.5 – Values of display devices related constants

Display Device	*M*	*C*
TV/ PC-Monitor	1.11	−0.232
Mobile/Tablet	1.0	−0.25

II.3.5 Computation of O.23

The perceptual buffering indication O.23 is calculated based only on the impacts of initial-loading, buffering and the media length as follows:

$$O.23 = 1 + 4 * InitLoadAndStallImpact$$

II.4 Performance figures

In this clause, the aggregated RMSE of the model. Aggregated RMSE is defined as:

$$RMSE = \frac{1}{W} \sum_{k=1}^{M} w_k \cdot RMSE_k$$

Where M represents the total number of (training and validation) databases, w_k and $RMSE_k$ the weight and root mean square error for database k respectively, and $W = \sum_{k=1}^{M} w_k$. Training and validation databases have different weights:

$$w_{training} = 0.1$$

$$w_{validation} = 0.9$$

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 II.6 – Performance of the Pq described above, based on the submitted version of ITU-T P.1204.5

RMSE	**Training**		**Validation**	
	Nr. of databases	***Nr. of Samples***	***Nr. of databases***	***Nr. of Samples***
0.553	2	82	4	134

Table II.7 – Performance of the Pq described above, based on the standardized ITU-T P.1204.5

RMSE	**Training**		**Validation**	
	Nr. of databases	***Nr. of Samples***	***Nr. of databases***	***Nr. of Samples***
0.529	2	82	4	134

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, 2021