International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.1203.3

Amendment 1

(05/2020)

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

Models and tools for quality assessment of streamed media

Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport – Quality integration module

Amendment 1

Recommendation ITU-T P.1203.3 (2019) – Amendment 1

ITU-T

International Telecommunication Union

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.1203.3

Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport – Quality integration module

Amendment 1

Summary

Recommendation ITU-T P.1203.3 specifies the quality integration module for Recommendation ITU-T P.1203. The ITU-T P.1203 series of ITU-T Recommendations specify modules for monitoring the audio, video and audiovisual quality of video services such as adaptive bit-rate video streaming. The respective ITU-T work item has formerly been referred to as P.NATS (parametric non-intrusive assessment of TCP-based multimedia streaming quality). The ITU-T P.1203.3 part of Recommendation ITU-T P.1203 can be applied to the monitoring of performance and quality of experience (QoE) of video services such as adaptive bit-rate video streaming. Besides stream-based input information, the ITU-T P.1203.3 quality integration module takes the per-one-second video- and audio-quality scores calculated according to ITU-T P.1203.1 and ITU-T P.1203.2, respectively, as input.

Only one quality integration module is recommended for all four modes 0 to 3 of the Recommendation ITU-T P.1203 model series, unique across all modes.

This Recommendation includes an electronic attachment containing the 20 trees described in clause 8.4.

This revision to Recommendation ITU-T P.1203.3 (2007) replaces equation (20) in clause 8.3.

Amendment 1 introduces an adjustment of the audiovisual quality of ITU-T P.1203.3 for the case of very low audio quality and long stalling events.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T P.1203.3	2016-12-22	12	11.1002/1000/13161
2.0	ITU-T P.1203.3	2017-10-29	12	11.1002/1000/13402
3.0	ITU-T P.1203.3	2019-01-13	12	11.1002/1000/13846
3.1	ITU-T P.1203.3 (2019) Amd. 1	2020-05-29	12	11.1002/1000/14284

Keywords

Adaptive streaming, audio, audiovisual, IPTV, mean opinion score (MOS), mobile video, mobile TV, monitoring, multimedia, 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

Electronic attachment: Decision trees to calculate the model output described in clause 8.4.

Editorial note for Amd.1 – The electronic attachment has can be downloaded bundled with the main version of ITU-T P.1203.3 (2019).

Recommendation ITU-T P.1203.3

Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport – Quality integration module

Amendment 1

Editorial note: This is a complete-text publication. Modifications introduced by this Amendment are shown in revision marks relative to Recommendation ITU-T P.1203.3 (2019).

1 Scope

This Recommendation[1] describes an objective parametric quality assessment model as part of the ITU-T P.1203 series of ITU-T Recommendations. The model specified herein predicts the impact of audiovisual quality variations and stalling events on quality experienced by the end user in multimedia mobile streaming and fixed network applications using adaptive bit-rate streaming, based on a previous estimation of audio and video quality and information on stalling events during the media session.

The model predicts mean opinion scores (MOS) on a 5-point ACR scale (see [ITU-T P.910]) as a final audiovisual quality MOS score (as defined in [ITU-T P.911], for instance). The model also outputs a perceptual stalling quality indicator, a final audiovisual compression quality score and a list of integrated audiovisual quality scores for diagnostic purposes.

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. Note that user interactions (such as pausing, seeking, user initiated quality change, user initiated play or user initiated end) are NOT considered in this model either.

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.

The application ranges of the ITU-T P.1203.3 model are summarized in Table 1.

[1] This Recommendation includes an electronic attachment containing the full 20 decision trees to calculate the model output described in clause 8.4.

Table 1 − Factors and application ranges of the ITU-T P.1203.3 model algorithms

Video sequence duration	60 seconds – 5 minutes
Initial loading delay and stalling	0-10 seconds
Maximum number of stalling events	5
Maximum length of a single stalling event	15 seconds
Total stalling duration	30 seconds
Other details	No stalling within 5 seconds of the start of the video playing.

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 P.800.1] Recommendation ITU-T P.800.1 (2016), *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] Recommendation ITU-T P.1201 (2012), *Parametric non-intrusive assessment of audiovisual media streaming quality*.

[ITU-T P.1203] Recommendation ITU-T P.1203 (~~2016~~2017), *Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport*.

[ITU-T P.1203.1] Recommendation ITU-T P.1203.1 (~~2016~~2019), *Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport – Video quality estimation module*.

[ITU-T P.1203.2] Recommendation ITU-T P.1203.2 (~~2016~~2017), *Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport – Audio quality estimation module*.

[ITU-T P.1401] Recommendation ITU-T P.1401 (~~2012~~2020), *Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following term defined elsewhere:

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

Further terms are defined in Recommendation [ITU-T P.1203].

3.2 Terms defined in this Recommendation

This Recommendation defines the following term:

3.2.1 T (media length): The length (in seconds) of the input media signal under consideration (O.21 or O.22). If both audio and video are considered, and they are not of equal length, the shorter duration of both inputs shall be used. The longer input shall be truncated at the end to match the duration of the shorter one:

$$\boldsymbol{T} = \boldsymbol{min}(length(O.21), length(O.22))$$

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

MOS Mean Opinion Score

5 Conventions

None.

6 Building blocks

The overall ITU-T P.1203 architecture is shown in Figure 6-1. The model described in this Recommendation is highlighted as the "*Pq*: Quality integration module".

The model takes as input:

1) audio coding quality per output sampling interval (O.21),
2) video coding quality per output sampling interval (O.22),
3) stalling positions and their length (I.14).

An internal media parameter extraction module extracts audio- and video-related parameters. The model outputs:

1) final media session quality score (O.46),
2) audiovisual segment coding quality per output sampling interval (O.34),
3) final audiovisual coding quality score (O.35),
4) perceptual stalling indication (O.23).

The model details are described in clauses 7 and 8.

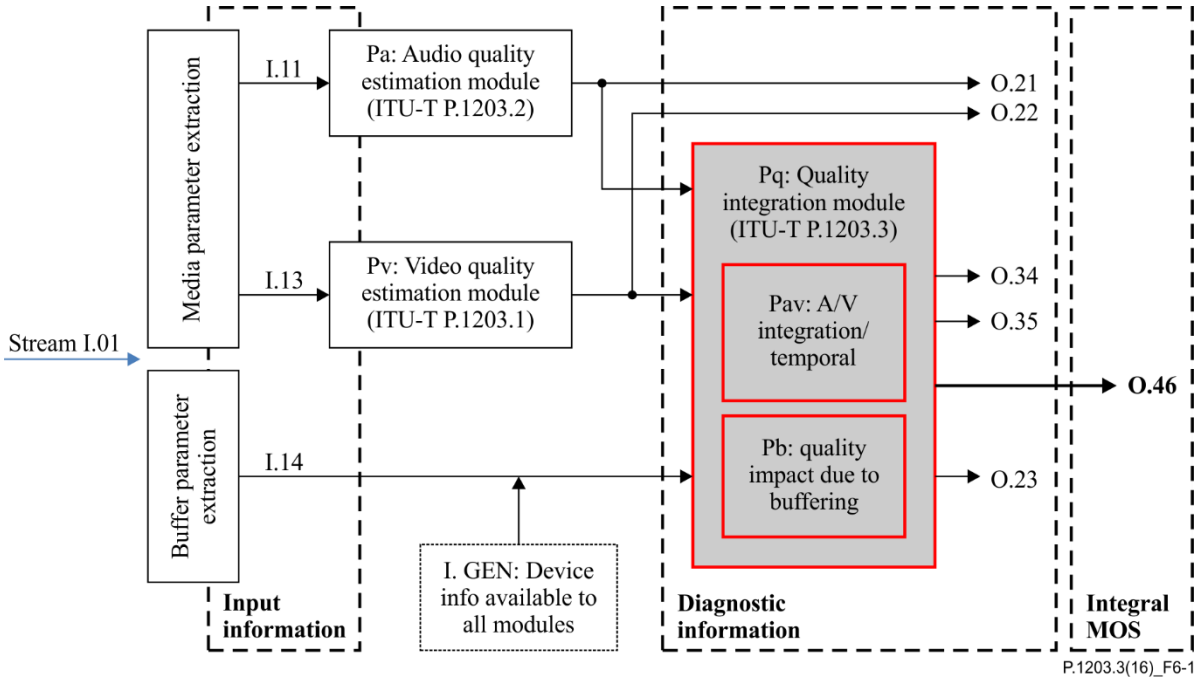

Figure 6-1 – *Pq* module in context of building blocks of the ITU-T P.1203 model

7 Model input

The model must receive information about the estimated quality of the audio and video streams and the occurrence of stalling events during playback. In various modes of operation, the following inputs may be extracted or estimated from network transmissions in different ways, which are outside the scope of this Recommendation:

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 [ITU-T P.1203.2].
- O.22: video quality per output sampling interval, as specified in [ITU-T P.1203.1].
- I.14: stalling events, as described in clause 7.1.
- I.GEN: device type (either "PC" or "Mobile"), as specified in [ITU-T P.1203].

The audio and video coding quality estimations (O.21 and O.22) may be generated using [ITU-T P.1203.2] and [ITU-T P.1203.1] modules, respectively, but may in principle also be generated using other audiovisual quality prediction models, provided that they 1) operate on the same output score range and 2) output sampling interval as defined in [ITU-T P.1203]. It is noted that in case that short-term video and/or audio quality modules are used, other than the ones specified in [ITU-T P.1203.1] and [ITU-T P.1203.2], respectively, the performance figures will be different from the ones described in [ITU-T P.1203]. Also, to be fully compliant with [ITU-T P.1203], the implementation of a complete set of modules Pv, Pa and Pq according to [ITU-T P.1203.1], [ITU-T P.1203.2] and this Recommendation is required.

7.1 I.14 input specification

I.14 consists of a vector of stalling events observed in the media session. Each event consists of a tuple of *start time* and *duration*, both measured in seconds.

The start time is expressed in terms of a "normalized time"/"media time" which is calculated from the start of the original sequence (i.e., the sequence without any stalling). Initial loading delay (i.e., stalling occurring at the beginning of the sequence before playback) has a start time of 0. All other stalling events have a start time greater than zero.

In practice, I.14 input information may be represented as a file containing one line per stalling event, with each line containing the stalling start time in seconds, followed by a whitespace and the stalling duration in seconds. Such a file may contain the following information (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 loading delay and 9.8 seconds of stalling after 2.5 seconds of the video playing. Then, there are 2 seconds of stalling at the mark of 63.2 seconds of the original video.

8 Model output

The ITU-T P.1203.3 model outputs the following information:

- O.23: perceptual stalling 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 (see [ITU-T P.1203], clause 3.2.6).

8.1 Media parameter extraction

Before calculating the outputs O.23, and O.46, the model calls a media parameter extraction module, which calculates parameters derived from the input information O.21, O.22 and I.14, as well as O.34 and O.35.

8.1.1 Parameters related to stalling

The media parameter extraction module outputs the following parameters:

- *numStalls*: number of stalling events
- *totalStallLen*: weighted sum of stalling events
- *avgStallInterval*: average interval between stalling events

8.1.1.1 Calculation of *totalStallLen*

For calculating *totalStallLen*, each individual stalling event *i* must first be assigned a weight w_stall_i, depending on its location within the media session. Note that if the number of stalling events is 0, the *totalStallLen* is 0.

The weight is calculated as:

$$w_{stall_i} = c_{ref7} + \left(1 - c_{ref7}\right) * e^{-\left(stallPositionFromEnd_i * \left(\frac{\log(0.5)}{-c_{ref8}}\right)\right)}, \qquad i \in [1, \ldots, numStalls] \quad (1)$$

where *stallPositionFromEnd* is the difference between the media length of the clip *T* and the stalling event start time:

$$stallPositionFromEnd_i = T - startTime_i, \qquad i \in [1, \ldots, numStalls] \quad (2)$$

Then, *totalStallLen* is:

$$totalStallLen = \sum_{i=1}^{numStalls} w_{stall_i} * stallLen_i \quad (3)$$

Table 8-1 shows coefficient sets for w_stall_i.

Table 8-1 – Coefficient sets for w_stall_i

C_{ref7}	0.48412879
C_{ref8}	10

8.1.1.2 Calculation of *avgStallInterval*

The calculation procedure for *avgStallInterval* can be found in Figure 8-1. Note that if the number of stalling events is 0 or 1, the average interval between stalling events is 0.

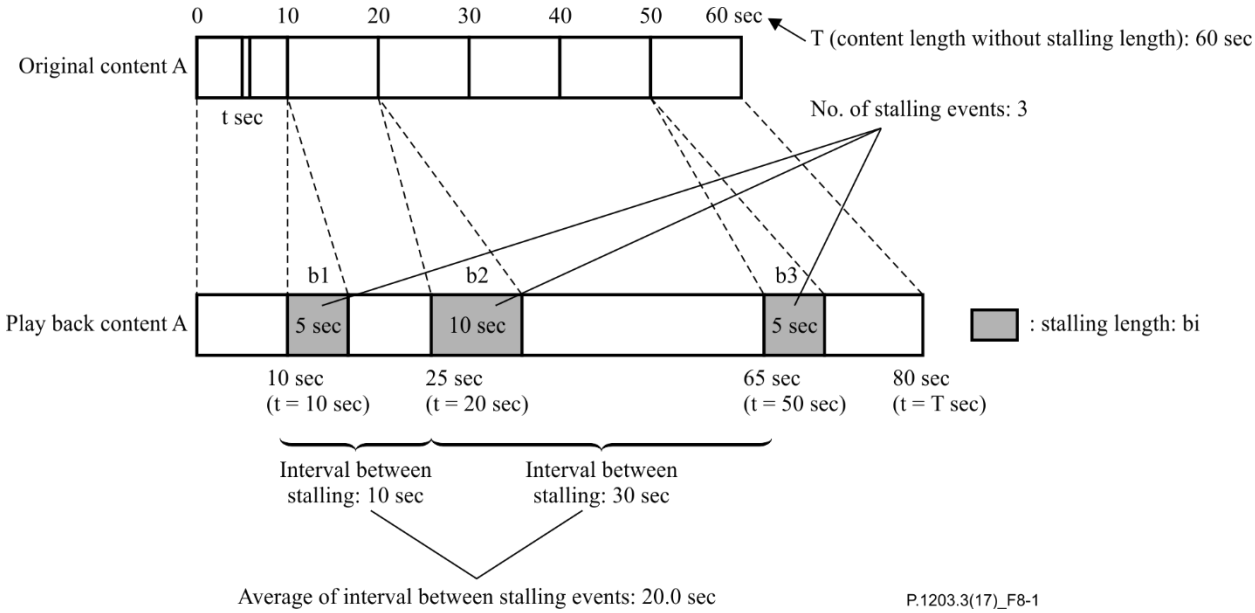

Figure 8-1 – Calculation of avgStallInterval

8.1.2 Parameters related to audiovisual quality

The module outputs the following parameters:

- *negativeBias*
- *vidQualSpread*
- *vidQualChangeRate*
- *qDirChangesTot*
- *qDirChangesLongest*

8.1.2.1 negativeBias

First, the difference between the O.35 score and individual O.34 scores must be calculated on a per-second basis weighted with a logarithm based on the time that has passed since the specific O.34 location.

$$O.34_{diff[t]} = (O.34[t] - O.35_{baseline}) * w_{diff}[t] \quad (4)$$

Where $O.35_{baseline}$ is calculated according to Eq. 19-21 (see clause 8.3), and:

$$w_{diff}[t] = c1 + (1 - c1) * e^{-\left((T-t)*\frac{\log(0.5)}{-c2}\right)} \quad (5)$$

where *t* is the temporal position in the O.34 list and *T* is the media length in seconds.

If the 10%-percentile border (*negPerc*) of that list is negative, the *negativeBias* is $-negPerc * c23$, otherwise it is 0:

$$negPerc = 10th\ percentile\ of\ 0.34\_diff \quad (6)$$

$$negativeBias = \max(0, -negPerc) * c23 \quad (7)$$

Table 8-2 shows the coefficient sets for *negativeBias*.

Table 8-2 – Coefficient sets for *negativeBias*

$c1$	1.87403625
$c2$	7.85416481
$c23$	0.01853820

8.1.2.2 vidQualSpread

The *vidQualSpread* parameter is calculated as the difference between the maximum and minimum video quality scores within O.22.

8.1.2.3 vidQualChangeRate

The *vidQualChangeRate* is calculated by iterating through the scores in O.22, starting from the second entry and counting the number of times that the difference between the current and previous O.22 value is greater than 0.2.

The total number of these occurrences must then be divided by the overall media length to give the *vidQualChangeRate*.

8.1.2.4 qDirChangesTot

This parameter indicates the total number of quality direction changes. To calculate this, firstly the video scores, O.22, are padded, and then transformed with a moving average filter of kernel size $k = 5$ into *O.22MA*.

Note that padding is needed in order to obtain meaningful moving averages for the beginning and the end of the session; the first and last audiovisual quality score in O.22 will be padded at the beginning and the end of O.22 such that the first (k-1) scores in padded O.22 are equal to the first score in O.22 and the last (k-1) scores in padded O.22 are equal to the last score in O.22.

To estimate the direction changes of the quality scores (positive = 1, negative = –1, equal = 0), the following algorithm is applied on O.22MA to result in a list *QC*.

1) Start with $p = 0$ and $q = 3$.
2) Calculate the difference $O.22MA(p) - O.22MA(q)$.
3) If the difference is above a threshold $t = 0.2$, append "1" to *QC*.
4) If the difference is between −0.2 and 0.2, append "0" to *QC*.
5) If the difference is below −0.2, append "−1" to *QC*.
6) Increase the values of p and q by the step size $s = 3$.
7) Continue with step 2.

The list *QC* now contains an estimation of the quality changes.

To compute *qDirChangesTot*, first, all zero values are removed from *QC*. Consecutive entries of identical values (1 or −1) then correspond to one region. *qDirChangesTot* is the total number of regions.

For example:

if *QC*=[0,0,1,1,1,0,0, −1, −1,0,1] then $qDirChangesTot = 3$.

if *QC*=[0,0,1,0,0, −1,0,0,0] then $qDirChangesTot = 2$.

8.1.2.5 qDirChangesLongest

This parameter is the duration in seconds of the longest period without positive or negative quality change. To calculate this value, the following algorithm is applied:

1) For every nonzero value in *QC*, the index and the value are stored in a tuple in a list *qc_len*, but only if it is the first nonzero value in a consecutive part of the same values. For example, for *QC*=[0,0,1,1,1,0,0,−1,−1,0,1] and the index starting at 1, the list *qc_len* contains [[3, 1], [8, −1], [11, 1]].

2) Then, after padding the list *qc_len* with zero value entries at the beginning and the end, the distances between consecutive indices is stored as distances. In the above example, *qc_len* becomes [[1, 0], [3, 1], [8, −1], [11, 1], [12, 0]] and the list of distances is [2, 5, 3, 1].

3) The value of *qDirChangesLongest* is equal to the maximum in the list of distances *qc_len* multiplied by the step size $s = 3$, or in the case of no quality changes the media length T.

The following pseudocode explains the calculation of the parameter *qDirChangesLongest*:

```
qc_len    = []
distances = []
T         = Length of O.22

for index in [0, ..., T],value in qc:
    if value != 0:
       if qc_len and qc_len[len(qc_len)][2] != value:
         qc_len.append([index, value])
       if not qc_len:
          qc_len.append([index, value])

if qc_len not empty:
  qc_len.prepend([1, 0])
  qc_len.append([length(qc) + 1, 0])
  for i in [2, ..., length(qc)]:
    distances.append(qc[i] – qc[i – 1])
  q_dir_changes_longest = max(distances) * s
else:
  q_dir_changes_longest = T
```

8.1.3 Parameters related to machine learning module

For the machine learning part of the *Pq* module, the following 14 features must be calculated:

0) *stallCountWithoutInitial*: total number of stalling events occurring in the media session, excluding the initial stalling event.

1) *stallDur*: This is the sum of duration of all the stalling events. The duration of stalling event corresponding to the initial stalling is given a weight of 1/3 as compared to the other stalling events:

$$stallDur = \frac{1}{3} \cdot initStallDur + \sum_{i=0}^{stallCountWithoutInitial} stallDur_i \qquad (8)$$

Where *initStallDur* is the duration of the initial stalling and $stallDur_i$ is the duration of the i-th stalling event within the media session (excluding initial stalling).

2) *stallFreq* (Frequency of stalling events): the number of stalling events (excluding the initial stalling) and dividing by the length of media.

$$stallFreq = \frac{stallCountWithoutInitial}{T} \quad (9)$$

3) *stallRatio* (Ratio of stalling duration): The ratio of stalling events is ratio of *stallDur* to the total media length (*T*).

$$stallRatio = \frac{stallDur}{T} \quad (10)$$

4) *timeLastStallToEnd*: This feature corresponds to the time elapsed since the start of occurrence of last stalling event to the end of video. The initial stalling event is excluded from the calculation of this feature. The value of this feature is set to *T* if there is no stalling in the session.

$$timeLastStallToEnd = T - stallPos_{stallCountWithoutInitial} \quad (11)$$

Where $stallPos_i$ is the position of the i-th stalling event within the media session (excluding initial stalling).

5) *averagePvScoreOne*: average of all the *O.22* scores that correspond to the first third of the *O.22* scores vector.

$$averagePvScoreOne = \frac{\sum_{i=1}^{\frac{n}{3}} O.22_i}{\frac{n}{3}} \quad (12)$$

6) *averagePvScoreTwo*: average of all the *O.22* scores of the second third of the *O.22* scores vector.

$$averagePvScoreTwo = \frac{\sum_{i=\frac{n}{3}+1}^{2\cdot\frac{n}{3}} O.22_i}{\frac{n}{3}} \quad (13)$$

7) *averagePvScoreThree*: average of all the *O.22* scores of the third third of the *O.22* scores vector.

$$averagePvScoreThree = \frac{\sum_{i=2\cdot\frac{n}{3}+1}^{n} O.22_i}{\frac{n}{3}} \quad (14)$$

8) *1PercentilePvScore*: 1st percentile of *O.22*.

9) *5PercentilePvScore*: 5th percentile of *O.22*.

10) *10PercentilePvScore*: 10th percentile of *O.22.*

11) *averagePaScoreOne*: all the *O.21* scores corresponding to the first half of the session are averaged:

$$averagePaScoreOne = \frac{\sum_{i=1}^{\frac{n}{2}} O.21_i}{\frac{n}{2}} \quad (15)$$

12) *averagePaScoreTwo*: all the *O.21* scores corresponding to the second half of the session are averaged:

$$averagePaScoreTwo = \frac{\sum_{i=\frac{n}{2}+1}^{n} O.21_i}{\frac{n}{2}} \quad (16)$$

13) T: the length of the media, see clause 3.2.2.

The parameters can also be referred to by a unique "feature ID" that is the numbered position in this list.

8.2 Model output O.34

The audiovisual quality per output sampling interval (O.34) is derived from O.21 and O.22 as follows, where $t = [1,2,\ldots,T]$:

$$O34'_t \text{~~}O34_t\text{~~} = max(min(av_1 + av_2 * O21_t + av_3 * O22_t + av4 * O21_t * O22_t, 5), 1) \quad (17)$$

In addition, the audiovisual quality O.34 is adjusted by a factor handling audio quality below *a_threshold*

$$O34_t = (1 - \max(0, a\_threshold - O21_t)) * (O34'_t - 1) + 1 \quad (17a)$$

The coefficients for the calculation of $O34_t$ ~~O.34~~ (i.e., *av*1, ..., *av*4, *a_threshold*) are listed in Table 8-4.

Table 8-4 – Coefficient sets for O.34

*av*1	–0.00069084
*av*2	0.15374283
*av*3	0.97153861
*av*4	0.02461776
a_threshold	2.0

8.3 Model output O.35

The final audiovisual coding quality (O.35) takes into account the audiovisual quality per output sampling interval (O.34) as well as any temporal effects and media length.

O.35 is calculated as follows:

$$O.35 = O.35_{baseline} - negBias - oscComp - adaptComp \quad (18)$$

$$O.35_{baseline} = \frac{\sum_t w_1(t) \cdot w_2(t) \cdot O.34[t]}{\sum_t w_1(t) \cdot w_2(t)} \quad (19)$$

$$w_1(t) = t_1 + t_2 \cdot e^{\left(\frac{t-1}{T}\right)/t_3} \quad (20)$$

$$w_2(t) = t_4 - t_5 \cdot O.34[t] \quad (21)$$

$$negativeBias = \text{see clause 8.1.2.1} \quad (22)$$

$$oscComp = \begin{cases} 0, & oscTest\ is\ false \\ \max\left(0, \min\left(qDiff \cdot e^{c1 \cdot qDirChangesTot + c2}, 1.5\right)\right), & oscTest\ is\ true \end{cases} \quad (23)$$

$$oscTest = \left(\frac{qDirChangesTot}{dur} < 0.25\right)\ AND\ (qDirChangesLongest < 30) \quad (24)$$

$$adaptComp = \begin{cases} 0, & adaptTest\ is\ false \\ \max(0, \min(c3 \cdot vidQualSpread \cdot vidQualChangeRate + c4, 0.5)), & adaptTest\ is\ true \end{cases}. \quad (25)$$

$$adaptTest = \left(\frac{qDirChangesTot}{dur} < 0.25\right) \quad (26)$$

$$qDiff = max(0, 1 + \log_{10}(vidQualSpread + 0.001)) \quad (27)$$

The coefficients for the above calculations (i.e., *t*1, ..., *t*5 and *c*1, ..., *c*4) are listed in Table 8-5.

Table 8-5 – Coefficient sets for O.35

$t1$	0.00666620027943848
$t2$	0.0000404018840273729
$t3$	0.156497800436237
$t4$	0.143179744942738
$t5$	0.0238641564518876
$c1$	0.67756080
$c2$	−8.05533303
$c3$	0.17332553
$c4$	−0.01035647

8.4 Model output O.46

The final media session quality score is calculated based on the number of stalling events (*numStalls*), the total length of stalling events (*totalStallLen*), the average of the interval between stalling events (*avgStallInterval*), the final audiovisual compression quality (O.35), the Random Forest model prediction ~~*(RFPrediction*, see clause 8.4.1)~~, and media length.

~~First, a temporary O.46 score is calculated as follows:~~

~~$$O.46\_\text{temp} = 0.75 * (1 + (O.35 - 1) * SI) + 0.25 * RFPrediction \quad (28)$$~~

$$O.46'_{\text{temp}} = 0.75 * (1 + (O.35 - 1) * SI) + 0.25 * RFPrediction \quad (28)$$

$$O.46_{\text{temp}} = 1 + \left(O.46'_{\text{temp}} - 1 \right) * q_{fac} \quad (28\text{-a})$$

$$q'_{fac} = a1 * totalStallLen + a2 \quad (28\text{-b})$$

$$q_{fac} = minimum(maximum(q'_{fac}, 0), 1) \quad (28\text{-c})$$

$$SI = \exp\left(-\frac{numStalls}{s1}\right) * exp\left(-\frac{\frac{totalStallLen}{T}}{s2}\right) * \exp\left(-\frac{\frac{avgStallInterval}{T}}{s3}\right) \quad (29)$$

The coefficients for *O.46_temp* (i.e., *s*1, *s*2, *s*3) are listed in Table 8-6.

Table 8-6 – Coefficient sets for O.46

$s1$	9.35158684
$s2$	0.91890815
$s3$	11.0567558
a1	–0.066667
a2	2.0

A final adjustment to O.46 is added to compensate for differences in subjective ratings due to the heterogeneity of tests across different laboratories.

$$O.46 = 0.02833052 + 0.98117059 \cdot O.46\_temp \quad (30)$$

The coefficients for this adjustment were calculated by performing a first order linear regression between subjective MOS and O.46 for each test and validation database separately. The 30 resulting coefficient pairs (slope and intercept) were then averaged in order to get the resulting coefficients used in Equation 30.

8.4.1 Calculation of RF prediction

The machine learning module for audiovisual quality estimation passes the features (with their IDs and values) described in clause 8.1.3 on to an ensemble of 20 decision trees, i.e., a Random Forest. Each decision tree has a maximum depth of 6.

It passes through each decision tree with the calculated feature vector and gets an estimate of MOS score per decision tree. The module then takes the arithmetic mean of all the MOS scores from 20 decision trees and provides one MOS score as output *RFPrediction*.

Every node in the decision tree is programmed with 5 pieces of information, i.e., node ID, feature ID, feature threshold (or MOS value, see below), left child node ID and right child node ID. The parsing of each decision tree starts at node 0 where the value of listed feature ID is checked. If the value of a given feature is below the feature threshold on the corresponding node, it moves on to the node with ID matching the left child node ID. Conversely, if the value of a same feature is above the feature threshold, it moves on to the node with ID same as the right child node ID. This process continues recursively until the node with feature ID of −1 is reached. The feature ID of –1 indicates that the leaf node is reached. The value in the feature threshold column of leaf node is the decision of corresponding tree, i.e., the MOS value associated to that particular decision tree.

An example of a part of a decision tree is shown in Figure 8-2 with each node consisting of 5 columns as mentioned above.

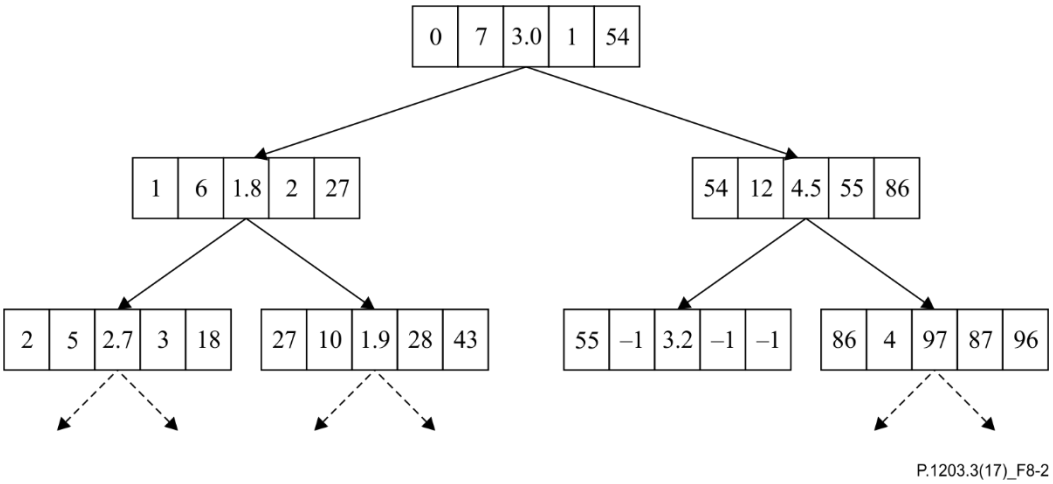

Figure 8-2 – Decision tree

This tree generates an output *RFprediction*, which is used to calculate the final O.46 as described above. The trees used in the RF implementation are stored as 20 individual comma-separated files. An example of initial rows in one file is shown below, where the columns correspond to the information as seen in each node above:

```
0, 10, 2.060699941, 1, 46
1, 1, 15.333333970, 2, 31
2, 12, 4.543439866, 3, 18
3, 5, 2.681250096, 4, 11
4, 9, 1.051000000, 5, 8
5, 5, 1.562726737, 6, 7
6,-1, 1.571538391, -1, -1
7,-1, 2.154570338, -1, -1
8, 6, 1.753901125, 9, 10
```

The full 20 trees can be found in the electronic attachment to this Recommendation.

8.5 Model output O.23

The perceptual stalling indication O.23 is calculated based on the number of stalling events (*numStalls*), the weighted sum of stalling events (*totalStallLen*) and the media length.

O.23 is calculated as follows:

$$O.23 = 1 + 4 * SI \tag{31}$$

Where *SI* is given in Equation 29.

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