International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.1203.3

(12/2016)

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

Recommendation ITU-T P.1203.3

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
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.1203.3

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

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 bitrate 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 bitrate video streaming.

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

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T P.1203.3	2016-12-22	12	11.1002/1000/13161

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

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

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 bitrate 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.

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

Video sequence duration	60 seconds - 5 minutes
Initial-buffering	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.

[1] This Recommendation includes an electronic attachment containing the full 20 trees described in clause 8.4.

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), *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), *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), *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), *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]

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

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

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

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

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

3.2.5 initial buffering: Refers to the time in seconds between the initiation of video playback by the user and the actual start of the playback.

3.2.6 media adaptation: Events where the player switches video playback between a known set of media quality levels while adapting to network conditions.

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) scores and 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) and 4) perceptual buffering indication (O.23).

The model details are described in clauses 7 and 8.

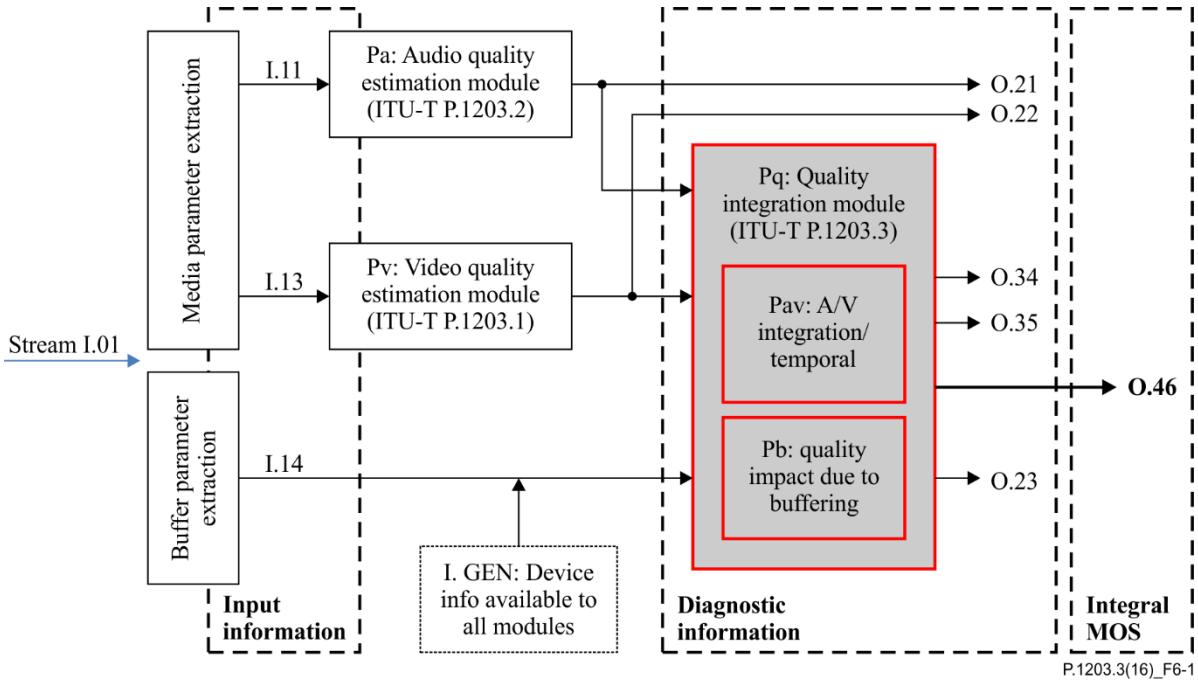

Figure 6-1 – *Pq* module in context of building blocks of the ITU-T P.1203 model

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 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 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].

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 buffering (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 buffering 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 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.

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 module outputs the following parameters:

- numStalls: number of stalling events
- totalBuffLen: total length of stalling events
- avgBuffLen: average interval between stalling events

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

Each individual stalling event has a weight assigned to it depending on the time between evaluation time and when the stalling event occurs. This length of the stalling event is multiplied with the weight when calculating the totalBuffLen sum. The weight is calculated as:

$$w\_buff_i = c_{ref7} + \left(1 - c_{ref7}\right) * e^{-\left(x * \left(\frac{\log(0.5)}{-c_{ref8}}\right)\right)}$$

where "*i*" is the *i*:th stalling event and "*x*" is the difference between the length of the clip and the stalling event time-stamp.

Table 8-1 shows coefficient sets for w_buff_i.

Table 8-1 – Coefficient sets for w_buff_i

C_{ref7}	0.48412879
C_{ref8}	10

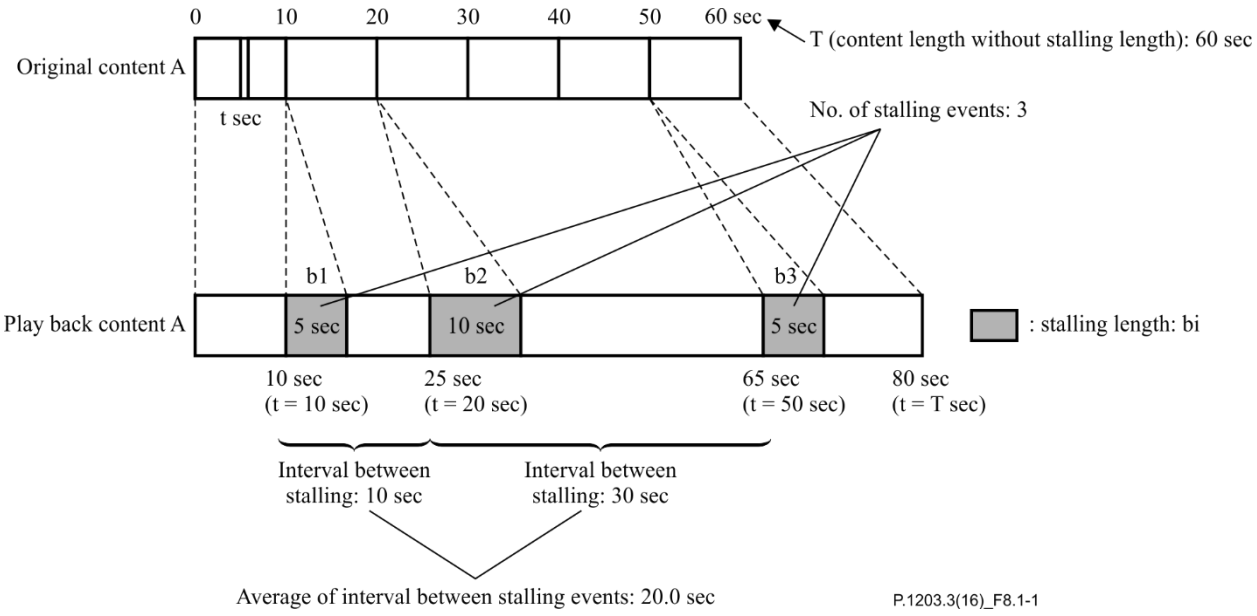

Figure 8-1 – A part of buffer parameter extraction

8.1.2 Parameters related to audiovisual quality

The module outputs the following parameters:

- negativeBias
- vidQualSpread
- vidQualChangeRate
- qDirChangesTot
- qDirChangesLongest

8.1.2.1 negativeBias

The *negBias* factor is calculated in two steps. First calculate the difference between the O.35 score and individual O.34 scores, weighted with a logarithm based on the time that has passed since the specific O.34 location. If the 10%-percentile border (*negPerc*) of that list is negative the *currnegBias* is *–negPerg* * *c*23, otherwise it is 0.

$$O.34\_diff[t] = O.34[t] * w_{diff}[t]$$

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

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

$$negPerc = 10th\ percentile\ of\ "O.34\_diff"$$

$$negBias = \max(0, -negPerc) * c23$$

Table 8-2 shows coefficient sets for *negBias*.

Table 8-2 – Coefficient sets for *negBias*

$c1$	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, O.22.

8.1.2.3 vidQualChangeRate

The vidQualChangeRate is calculated by considering each of the video quality scores, O.22, in time order starting at the second value and counting the number of times that the difference between the current and previous MOS score is greater than 0.2. The total number of differences is then divided by the overall clip duration to give the video quality change rate.

8.1.2.4 qDirChangesTot

This indicates the total number of quality direction changes. To calculate this firstly the video scores, O.22, are padded 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 the $p = 0$ 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 length of the session 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])
quality_direction_changes_longest = max(distances) * s
else:
quality_direction_changes_longest = T
```

8.1.3 Parameters related to machine learning module

The machine learning *Pq* model is based on 14 features. These 14 features are listed below:

NOTE – In order to calculate features 7, 8 and 9 in the following list, the input vector of one second *Pv MOS* scores is divided into three parts. Similarly, features 13 and 14 are calculated by dividing a vector of *Pa* scores into two parts.

1) Media length (*mediaLength*): This feature is simply the minimum of the length of *Pa* and *Pv MOS* score vectors provided as input to the machine learning *Pq* module.

$$mediaLength = \min(PaLength, PvLength)$$

2) Number of rebufferings (*reBuffCount*): This feature is the total number of stalling events occurring in the media streaming, excluding the initial stalling event.

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

$$stallDur = \frac{1}{3} \cdot initBuffDur + \sum_{i=1}^{n} reBuffDur_i$$

4) Frequency of rebuffering events (*reBuffFreq*): This feature is calculated by taking the number of stalling events (excluding the initial buffering) and dividing by the length of media.

$$reBuffFreq = \frac{reBuffCount}{mediaLength}$$

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

$$stallRatio = \frac{stallDur}{mediaLength}$$

6) Time between position of last rebuffering event to the end of media streaming (*timeLastRebuffToEnd*): 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 media length if there is no rebbuffering in the session.

$$timeLastRebuffToEnd = mediaLength - timeLastRebuff$$

7) Average *Pv MOS* score of the first part of media streaming (*averagePvScoreOne*): This feature is calculated by taking the average of all the *Pv* scores that correspond to the first part of a *Pv* scores vector.

$$averagePvScoreOne = \frac{\sum_{i=1}^{n/3} Pv_i}{n/3}$$

8) Average *Pv MOS* score of the second part of media streaming (*averagePvScoreTwo*): This feature is calculated by taking the average of all the *Pv MOS* scores of the second half of a *Pv* scores vector.

$$averagePvScoreTwo = \frac{\sum_{i=\frac{n}{3}+1}^{2\cdot n/3} Pv_i}{n/3}$$

9) Average *Pv MOS* score of the third part of media streaming (*averagePvScoreThree*): This feature is calculated by taking the average of all the *Pv MOS* scores of the second half of a *Pv* scores vector.

$$averagePvScoreThree = \frac{\sum_{i=2\cdot\frac{n}{3}+1}^{n} Pv_i}{n/3}$$

10) 1st percentile of *Pv MOS* scores (*1PercentilePvScore*): This feature is calculated by taking 1st percentile *MOS* score of *Pv MOS* scores vector.

11) 5th percentile of *Pv MOS* scores (*5PercentilePvScore*): This feature is calculated by taking 5th percentile *MOS* score of *Pv MOS* scores vector.

12) 10th percentile of *Pv MOS* scores (*10PercentilePvScore*): This feature is calculated by taking 10th percentile MOS score of *Pv MOS* scores vector.

13) Average *Pa MOS* score of the first part of media streaming (*averagePaScoreOne*): To calculate this feature, all the *Pa MOS* scores corresponding to the first half of media streaming are averaged.

$$averagePaScoreOne = \frac{\sum_{i=1}^{n/2} Pa_i}{n/2}$$

14) Average *Pa MOS* score of the second part of media streaming (*averagePaScoreTwo*): To calculate this feature, all the *Pa MOS* scores corresponding to the second half of media streaming are averaged.

$$averagePaScoreTwo = \frac{\sum_{i=\frac{n}{2}+1}^{n} Pa_i}{n/2}$$

Each of the above features is given a unique ID as listed in Table 8-3.

Table 8-3 – Mapping of feature IDs and feature names

Feature ID	Feature Name
0	*reBuffCount*
1	*stallDur*
2	*reBuffFreq*
3	*stallRatio*
4	*timeLastRebuffToEnd*
5	*averagePvScoreOne*
6	*averagePvScoreTwo*
7	*averagePvScoreThree*
8	*1PercentilePvScore*
9	*5PercentilePvScore*
10	*10PercentilePvScore*
11	*averagePaScoreOne*
12	*averagePaScoreTwo*
13	*mediaLength*

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:

$$\boldsymbol{O.34(t) = max(min(av_1 + av_2 \times O.21(t) + av_3 \times O.22(t) + av_4 \times O.21(t) \times O.22(t), 5), 1),} \quad \text{(Eq. 8-1)}$$

Note that in the above equation, t = [1, 2, ..., T] and T is the media length, i.e., the minimum of the length of O.21 and O.22.

The coefficients for the calculation of O.34 (i.e., *av*1, ..., *av*4) 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

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:

$$O.35 = O.35_{baseline} - negBias - oscComp - adaptComp, \quad \text{(Eq. 8-2)}$$

$$O.35_{baseline} = \frac{\sum_t w_1(t) \cdot w_2(t) \cdot O.34[t]}{\sum_t w_1(t) \cdot w_2(t)}, \quad \text{(Eq. 8-3)}$$

$$w_1(t) = t_1 + t_2 \cdot e^{t/(T/t_3)}, \quad \text{(Eq. 8-4)}$$

$$w_2(t) = t_4 - t_5 \cdot O.34[t], \quad \text{(Eq. 8-5)}$$

$$negBias = \text{see clause 8.1.2.1} \quad \text{(Eq. 8-6)}$$

$$oscComp = \begin{cases} 0, & oscTest\ is\ false \\ min\left(qDiff \cdot e^{min(c1 \cdot qDirChangesTot + c2, 1.5)}\right), & oscTest\ is\ true \end{cases} \quad \text{(Eq. 8-7)}$$

$$oscTest = \left(\frac{qDirChangesTot}{dur} < 0.25\right)\ AND\ (qDirChangesLongest < 30) \quad \text{(Eq. 8-8)}$$

$$adaptComp = \begin{cases} 0, & adaptTest\ is\ false \\ min(c3 \cdot vidQualSpread \cdot vidQualChangeRate + c4), & adaptTest\ is\ true \end{cases}. \quad \text{(Eq. 8-9)}$$

$$adaptTest = \left(\frac{qDirChangesTot}{dur} < 0.25\right) \quad \text{(Eq. 8-10)}$$

$$qDiff = max(0.1 + \log_{10}(vidQualSpread + 0.01)) \quad \text{(Eq. 8-11)}$$

The coefficients of O.35 (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

*t*1	0.00666620027943848
*t*2	0.0000040018840273729
*t*3	0.156497800436237
*t*4	0.143179744942738
*t*5	0.0238641564518876
*c*1	0.67756080
*c*2	–8.05533303
*c*3	0.17332553
*c*4	–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 (totalBuffLen), the average of the interval between stalling events (avgBuffInterval), the final audiovisual compression quality (O.35) and media length as input.

O.46 is calculated as follows:

$$O.46 = 0.75 * (1 + (O.35 - 1) * SI) + 0.25 * RF_{prediction} \quad \text{(Eq. 8-12)}$$

$$SI = \exp(\text{- numStalls / s1}) \cdot \exp(\text{- (}\frac{\text{totalBuffLen}}{\text{T}}\text{) / s2}) \cdot \exp(\text{- (}\frac{\text{avgBuffInterval}}{\text{T}}\text{) / s3}) \quad \text{(Eq. 8-13)}$$

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

Table 8-6 – Coefficient sets for O.46

*s*1	9.35158684
*s*2	0.91890815
*s*3	11.0567558

A final adjustment to O.46 is added to compensate for differences in subjective ratings due to the heterogeneity of tests across different laboratories. 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 8-14.

$$O.46 = 0.02833052 + 0.98117059 \cdot O.46 \quad \text{(Eq. 8-14)}$$

Calculation of RF_prediction.

The machine learning module for audiovisual quality estimation takes vectors of *Pa* and *Pv* scores, video stall locations and video stall durations information as inputs. Based on this information, it calculates values of several constructed features, which are then passed on to the ensemble of 20 decision trees, i.e., random forest. Each decision tree has a maximum depth of 6. It passes through each decision tree with calculated feature vector and gets an estimate of MOS score per decision tree. The module then takes arithmetic mean of all the MOS scores from 20 decision trees and provides one MOS score as *Pq* output.

Each feature is described in clause 8.1.3.

A feature vector with feature IDs and their corresponding values for a media streaming session are provided as input to all the 20 decision trees. The module then runs through all the decision trees one by one in a sequence.

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.

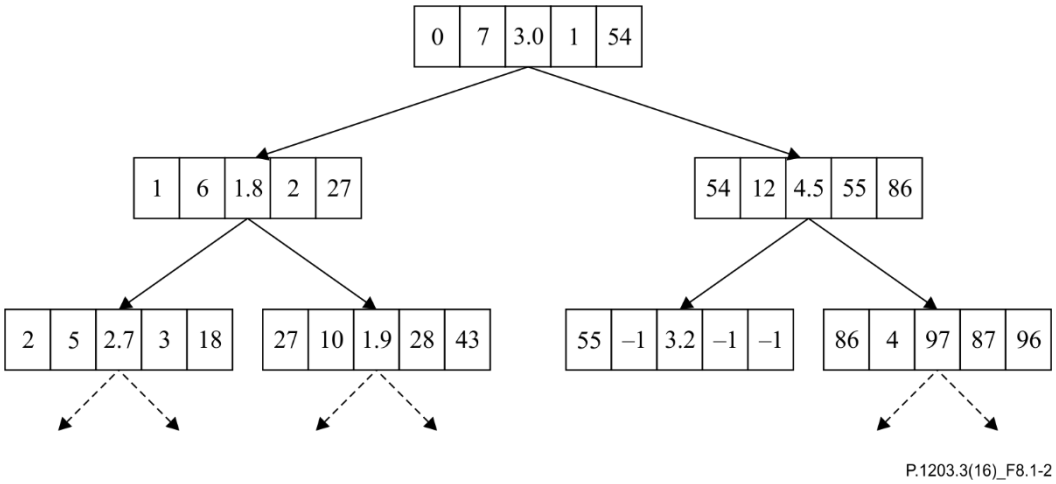

Figure 8-2 – Decision tree

This tree generates an output that is RF_prediction which is used to calculate the final O.46 as described above. The trees used in the RF implementation are stored as 20 separate CSV-files, an example of some 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 buffering indication O.23 is calculated based on the number of stalling events (numStalls), the total length of stalling events (totalBuffLen) and the media length

O.23 is calculated as follows:

$$O23 = 1 + 4 \cdot SI \, . \qquad \text{(Eq. 8-15)}$$

Where *SI* is given in Equation 8-7.

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 telecommunication network management 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, 2017