ITUPublications
Recommendations

International Telecommunication Union
Standardization Sector

Recommendation

ITU-T P.1211 (10/2023)

SERIES P: Telephone transmission quality, telephone installations, local line networks

Models and tools for quality assessment of streamed media

Derivation procedure for contribution values to quality degradation of adaptive audiovisual streaming services

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.1211

Derivation procedure for contribution values to quality degradation of adaptive audiovisual streaming services

Summary

Recommendation ITU-T P.1211 describes a derivation procedure for contribution values that provide information about a relationship between quality-influencing factors (e.g., media quality levels and stalling) and the final media session quality score of adaptive audiovisual streaming services. More precisely, contribution values show how much each quality-influencing factor affects (i.e., decreases) the final media session quality score provided by a quality estimation model (e.g., that of ITU-T P.1203 or ITU-T P.1204) using Shapley theory, which is a key concept in cooperative game theory for distributing profits in a game among participants. This derivation procedure can be used to monitor the streaming service since operators can ascertain which quality-influencing factors decrease the final media session quality score.

History *

Edition	Recommendation	Approval	Study Group	Unique ID
1.0	ITU-T P.1211	2023-10-29	12	11.1002/1000/15695

Keywords

Adaptive streaming, audiovisual, audiovisual quality estimation, contribution value, mean opinion score (MOS), monitoring, multimedia, QoE, Shapley theory.

---

* To access the Recommendation, type the URL https://handle.itu.int/ in the address field of your web browser, followed by the Recommendation's unique ID.

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/.

Table of Contents

Recommendation ITU-T P.1211

Derivation procedure for contribution values to quality degradation of adaptive audiovisual streaming services

1 Scope

This Recommendation describes a derivation procedure for contribution values that provides information about a relationship between a quality-influencing factor (i.e., media quality levels and stalling) and media session quality of adaptive audiovisual streaming services.

In streaming services, the quality experienced by the end user can be assessed by using quality estimation models such as:

– A metadata-based model, which uses audio and video bitrates, framerate, resolution ITU-T P.1203.1 mode 0, [ITU-T P.1203.2], [ITU-T P.1203.3];

– A bitstream-based model that uses meta-data and any information from the video bitstream ITU-T P.1203.1 mode 3, [ITU-T P.1204.3], [ITU-T P.1203.2], [ITU-T P.1203.3];

– A pixel-based model that uses information derived from reference video [ITU-T P.1204.4], [ITU-T P.1203.2], [ITU-T P.1203.3];

– A hybrid metadata-based and pixel-based model [ITU-T P.1204.5], [ITU-T P.1203.2], [ITU-T P.1203.3].

These models estimate a mean opinion score (MOS) on a five-point absolute category rating scale [ITU-T P.910] as a global multi-media MOS score [ITU-T P.911] by taking into account quality-influencing factors such as media quality level (i.e., set of audio-coding or video degradations due to coding, spatial re-scaling or variations in video frame rates) and stalling (i.e., delivery degradations due to initial loading delay and rebuffering at the client) [ITU-T P.1203]. In this Recommendation, the output of these quality estimation models (i.e., global multi-media MOS score) is referred to as the final media session quality score as specified in [ITU-T P.1203].

The derivation procedure described provides contribution values that show how much each media quality level and stalling affects (i.e., decreases) the final media session quality score provided by quality estimation models such as [ITU-T P.1203] as previously described for each media session. This contribution value is calculated as the degree of quality degradation due to each media quality level and stalling by comparing with the maximum final media session quality score (i.e., the highest audiovisual quality and no stalling during the media session) based on Shapley theory, which is a key solution concept in cooperative game theory for distributing profits in a game among participants [b-Shapley]. Here, for example, different media quality levels selected or stalling occurrences during the media session result in different contribution values of each media quality level and stalling. Also, if the final media session quality score of a certain media session is different because of the use of a different quality estimation model, the contribution values of each media quality level and stalling are also different.

The derivation procedure described is agnostic to the specified quality estimation models such as the ITU-T P.1203 type of model. This means that, although a metadata-based model (i.e., ITU-T P.1203 mode 0), bitstream-based model (i.e., ITU-T P.1203 mode 3), pixel-based model (i.e., [ITU-T P.1204.4]) and hybrid model (i.e., [ITU-T P.1204.5]), and their respective appendices for long term integration, generally exist to estimate the final media session quality score, this Recommendation can be used with all types of quality estimation model. As examples, the procedures when ITU-T P.1203 mode 0 (metadata-based model), ITU-T P.1203 mode 3 (bitstream-based model) and [ITU-T P.1204.4] (pixel-based model) are used as quality estimation model are described in Appendices I, II and III, respectively.

This Recommendation primarily applies to the analysis of the quality aspect of progressive download or adaptive audiovisual streaming so that operators can ascertain degradation factors by deploying the derivation procedure described in end-point locations, mid-network monitoring points or a streaming server.

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.910] Recommendation ITU-T P.910 (2023), *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.1203] Recommendation ITU-T P.1203 (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 (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 (2017), *Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport – Audio 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.*

[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*.

[ITU-T P.1204.4] Recommendation ITU-T P.1204.4 (2022), *Video quality assessment of streaming services over reliable transport for resolutions up to 4K with access to full and reduced reference pixel information*.

[ITU-T P.1204.5] Recommendation ITU-T P.1204.5 (2023), *Video quality assessment of streaming services over reliable transport for resolutions up to 4K with access to transport and received pixel information*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 adaptation set [ITU-T P.1203]: A set of distinct media quality levels to be used for HTTP adaptive streaming, between which the player can perform media adaptation.

3.1.2 mean opinion score (MOS) [b-ITU-T P.10]: The mean of opinion scores.

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

3.1.4 segment [ITU-T P.1203]: An audiovisual file belonging to one particular media quality level.

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

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 maximum media session quality score: The final media session quality score for a certain sequence in which only the highest media quality level is selected, and no stalling occurs.

3.2.2 modified sequence: A sequence in which specified selected media quality levels during a media session are replaced by the highest media quality level in an adaptation set or the number of stalling occurrences is replaced by zero.

3.2.3 stalling: Stalling is caused by initial loading delay and 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, and result in stalling of the media playout.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

AAC Advanced Audio Coding

AC3 Audio Compression 3

FQ Final Quality

HE High Efficiency

HRC Hypothetical Reference Circuit

HTTP Hyper Text Transfer Protocol

LC Low Complexity

MO Mobile

MOS Mean Opinion Score

PC Personal Computer

PVS Processed Video Sequence

TA Tablet

TV Television

VP9 Video Payload type 9

5 Conventions

None.

6 Field of application

This Recommendation analyses the quality aspect of progressive download streaming or adaptive bitrate streaming. Limitations regarding the transport protocol, video length or encoding protocol and other attributes related to streaming services are the same as those of the quality estimation model used since the procedure described requires the final media session quality score from it.

The procedure described is used on a server providing streaming service, as shown in Figure 1. After finishing a video playback on a client device, stalling information is sent to the streaming server. The contribution values are then calculated by a derivation procedure described in this Recommendation using a quality estimation model and selected quality level during media sessions that are on the streaming server.

The contribution values are generally calculated on the streaming server so that operators can ascertain degradation factors. This is because the contribution calculation needs bitstream or pixel information of the highest media quality level at each segment, which is stored on a streaming server and not on a client device, if the bitstream-based (i.e., ITU-T P.1203 mode 3) or pixel-based (i.e., [ITU-T P.1204.4]) quality estimation model is used. However, if the metadata-based quality estimation model (i.e., ITU-T P.1203 mode 0) is used, the contribution values can be calculated on the client device since the metadata-based quality estimation model can be executed there and the metadata of the highest media quality level is also found on that site.

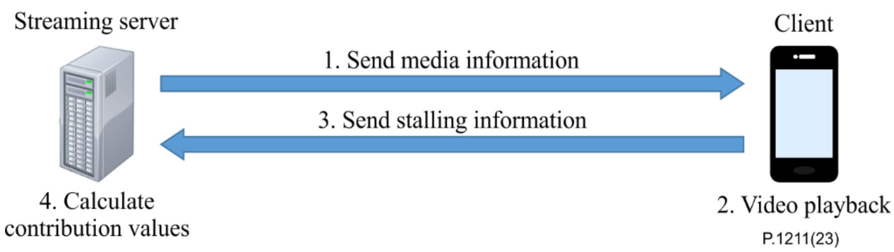

Figure 1 – Sequence diagram

7 Building blocks

The building blocks are depicted in Figure 2. This Recommendation uses the selected quality level sequence and stalling information extracted from the sequence and final media session quality score output from a quality estimation module. In addition to that, if the bitstream-based or pixel-based quality estimation model is used, bitstream information or pixel information is needed as shown in Figure 2. The details are described in clause 7.1.

During the calculation of contribution values, since the derivation procedure requires its final media session quality score, the modified sequence is sent to the quality estimation module. Here, in a modified sequence, specified selected media quality levels during media session are replaced by the highest media quality level in an adaptation set or the number of stalling occurrences is replaced by zero. The details of the modified sequence are explained in clause 8. The derivation procedure described also needs adaptation set information, which is prepared at the streaming server.

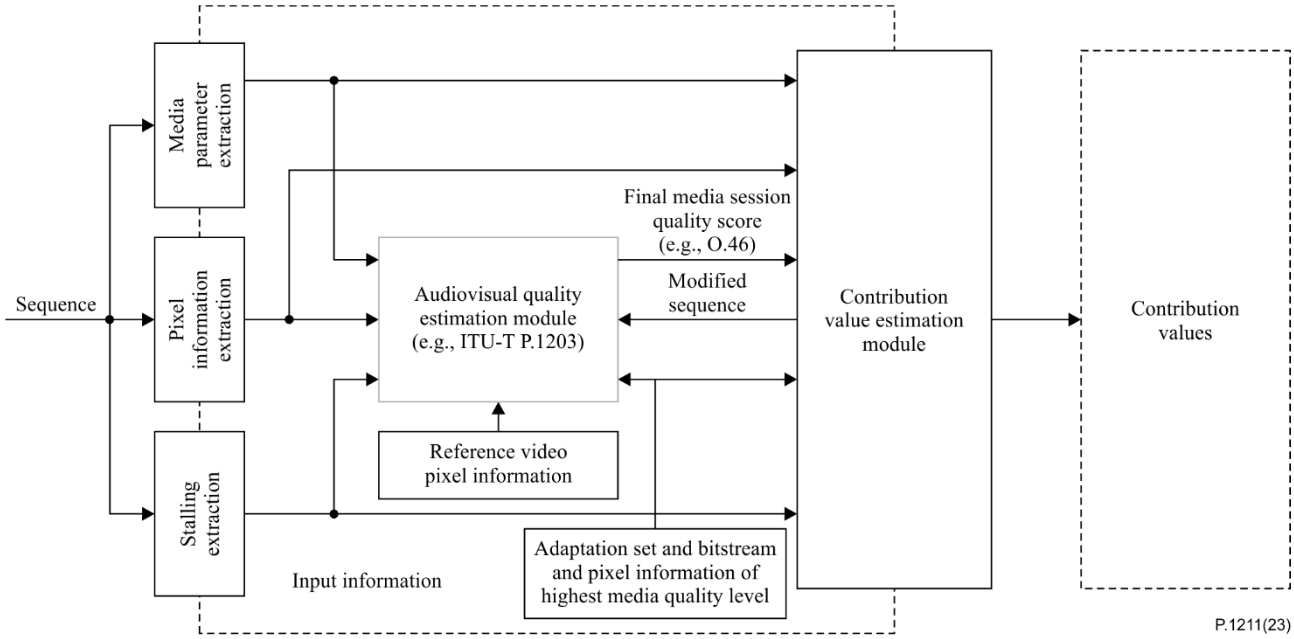

Figure 2 – Building blocks

7.1 Input data

This Recommendation uses the adaptation set information at a streaming server [ITU-T P.1203], which consists of multiple media quality levels (i.e., multiple levels of tuple of audio bitrate, video bitrate, video resolution, and video frame rate), a sequence that is a history of selected media quality levels during a media session, stalling information tuple (i.e., start time and duration of stalling [ITU-T P.1203.3]) during a media session, and the final media session quality score. In addition, if the bitstream-based quality estimation model is used, bitstream information of the selected quality levels during the media session is extracted from the sequence and the bitstream information of the highest media quality level at each segment, is needed as shown in Figure 2. When a pixel-based quality estimation model is used, video pixel information of the selected quality levels is extracted from the sequence and pixel information of the highest media quality level, and video reference pixel information obtained as input, are needed. Appendices II and III describe how to use the information.

Let $N = \{r_1, \ldots, r_R, s\}$ be a set of an adaptation set (i.e., prepared media quality levels on a streaming server) and stalling information, where r_1 is the lowest media quality level, r_R is the highest media quality level, R is the number of media quality levels, and s represents all stalling events. Here, media quality level is the tuple of audio bitrate, video bitrate, video resolution, and video frame rate. Let $x_j \in N$ be the jth element in $N, j = 1, \ldots R + 1$. H denotes the sequence, which consists of selected media quality levels r_i and all stalling events s during video playback. The notation used in this Recommendation is listed in Table 1.

Note that, since the quality estimation model is used during the contribution value calculation, the input data in this Recommendation is specified by the input data of the quality estimation model. As examples, the specific input data is described if the ITU-T P.1203 mode 0, ITU-T P.1203 mode 3 or ITU-T P.1204 is used in Appendices I, II and III.

Table 1 – Notation used in this Recommendation

Symbol	Description
$N = \{r_1, \dots, r_R, s\}$	A set of adaptation set and stalling information. r_1 is the lowest media quality level, r_R is the highest media quality level, R is the number of media quality levels, and s represents all stalling events. Here, media quality level is the tuple of audio bitrate, video bitrate, video resolution, and video frame rate
x_j	jth element in $N, j = 1, \dots R + 1$.
Z	All of subsets of N without x_j.
z_i	One of the elements of Z. $z_i \in Z, i = 1, \dots, \lvert Z \rvert$
H	A sequence, that consists of selected media quality level r_i and all stalling events s during video playback.
H' (or H'')	A modified sequence of an input sequence H and z_i, in which a part of the selected media quality level during a media session is replaced with the highest media quality level in an adaptation set.
f	A function that takes a given H and subset z_i (or $z_i \cup x_j$) as input and outputs a modified sequence H' (or H'').
ν	A quality estimation model such as that of [ITU-T P.1203] or [ITU-T P.1204], which estimates the quality of an input sequence, H.
c_j	Contribution value of x_j.

8 Derivation procedure

The derivation procedure for contribution values consists of three steps, which are summarized in Table 2.

Table 2 – Calculation steps of contribution values

Step	*Description*
1	Pick up one element x_j in N and iterate steps 2 and 3 for over all elements x_j in $N, j = 1, \dots R + 1$.
2	Determine the set Z consisting of all subsets of N without x_j.
3	Calculate the contribution value, c_j, of x_j by adapting the Shapley value calculation given in Equation 1.

The first step is to pick up one element x_j and iterate steps 2 and 3 for over all elements x_j in N. Here, the order of picking up element in N does not affect the calculation results. For simplicity, the case x_j is explained in the following. However, the procedure described is executed over all elements in N to calculate all contribution values, $c_j, j = 1, \dots, R, R + 1$. The second step is to determine the composition of all subsets Z of N without x_j. Here, Z is all possible patterns of subsets induced from N without x_j, and the cardinality of Z is $2^{|N|-1}$. The third step is to calculate the contribution value, c_j, of x_j by adapting the Shapley value calculation.

The details of the third step, in which the contribution value, c_j, of x_j is calculated, are as follows. Let f be a function that takes a given sequence H and subset $z_i \in Z, i = 1, \dots, |Z|$ (or $z_i \cup x_j$) as input and outputs a modified sequence, H' (or H'').Let ν be a quality estimation model such as that of [ITU-T P.1203] or [ITU-T P.1204], which estimates the quality of an input sequence, H. From here, although the contribution value calculation algorithm is explained only for a given subset, z_i, it is also applicable to other subsets, $(z_i \cup x_j)$. The modification procedure is described in the following paragraphs. The contribution value, c_j, of x_j is calculated by adopting the Shapley value calculation algorithm as follows.

$$c_j(v, f, H) = \sum_{i=1}^{|Z|} \frac{|z_i|!(|N|-|z_i|-1)!}{|N|!} \Big(v(f(z_i, H)) - v\big(f\big(z_i \cup x_j, H\big)\big) \Big) \quad (1)$$

where $|Z|$ and $|N|$ denote the cardinality of Z and N, respectively, and the sum extends over all subsets, $z_i \in Z$, which are the subsets of N that do not include x_j.

The function f maps a subset z_i and a sequence of quality levels $H = (h_t)$ to a new sequence of quality levels $H' = (h'_t)$ by replacing all occurrences of quality levels in z_i by the highest quality level r_R,

$$h_t^{'} = \begin{cases} r_R & \text{if } h_t \in z_i \\ h_t & \text{otherwise} \end{cases} \quad (2)$$

For the stalling, if s is in z_i, f generates H' in which the stalling events are removed, while if s is not in z_i, f generates H' in which stalling information is the same as the original H.

For the case, $z_i \cup x_j$, f outputs H'' similarly to modifying H to H'. By generating sequences H' and H'' on the basis of a given subset z_i and calculating the quality decrease over all possible subsets using Equation (1), the contribution value, c_j, of x_j is calculated.

Appendix I

Example of the deviation procedure with ITU-T P.1203 mode 0 (metadata-based model) as a quality estimation model

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

This appendix describes the example of the deviation procedure when ITU-T P.1203 mode 0 is used as a quality estimation model. Especially, prepared processed video sequences (PVSs), input data specification, and contribution values are described.

Prepared PVSs

In this example, four PVSs listed under this heading are used to show the derivation procedure and the contribution values for various kinds of hypothetical reference circuit (HRC). These PVSs are generated on the basis of the HRC of an ITU-T P.1203 open dataset [b-Robitza] by using a prepared source video, which is 1 min long. The characteristics of these PVSs follow and the HRCs used are shown in Figures I.1 and I.2. Table I.1 shows the media quality level of the HRC.

PVS01: HRC80 in TR04 [b-Robitza] in which only one media quality level is selected during a media session with no stalling.

PVS02: HRC03 TR04 [b-Robitza] in which multiple media quality levels are selected during a media session with no stalling.

PVS03: HRC85 in TR04 [b-Robitza] in which multiple media quality levels are selected and a stalling occurs during a media session.

PVS04: HRC02 in TR04 [b-Robitza] in which multiple media quality levels are selected and stalling occurs multiple times during a media session.

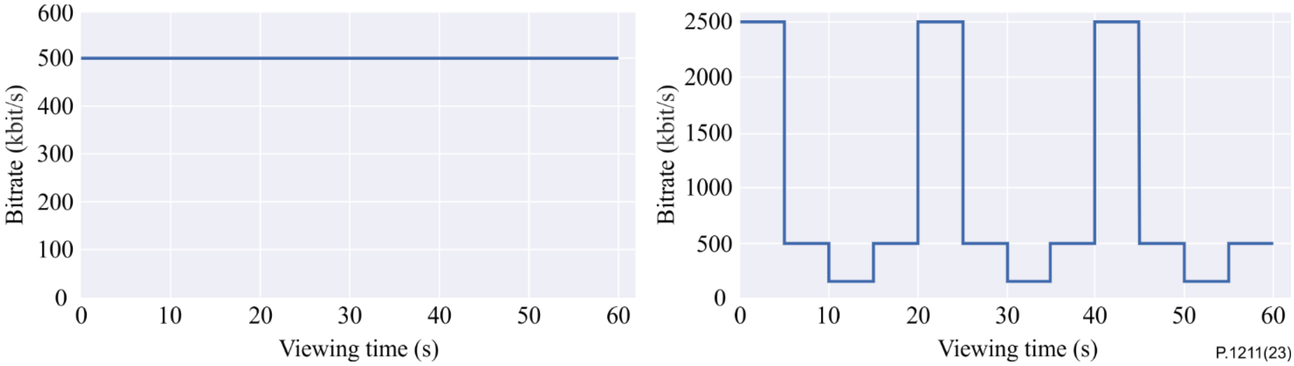

Figure I.1 – HRC80 and HRC03 in TR04 in ITU-T P.1203 open dataset [b-Robitza]

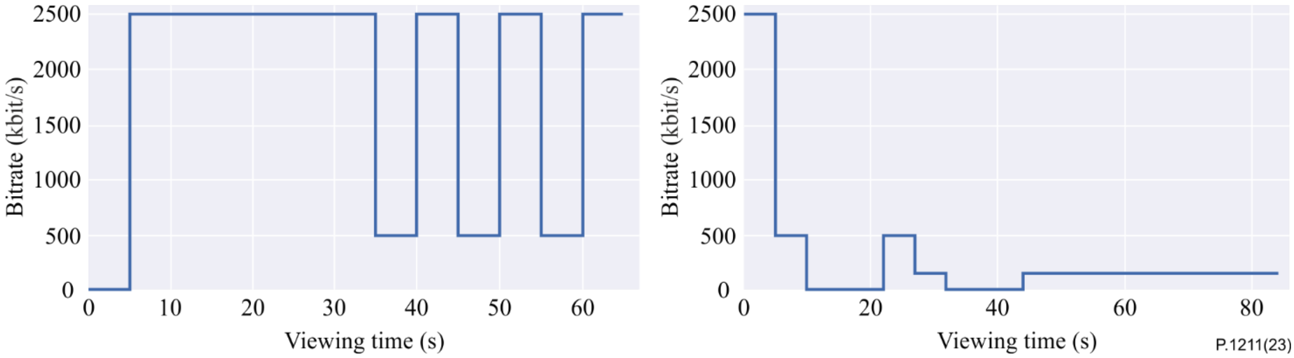

Figure I.2 – HRC85 and HRC02 in TR04 in ITU-T P.1203 open dataset [b-Robitza]

Table I.1 – Media quality level information of PVSs

NOTE – Quality levels QL1, QL3, and QL5 are not defined in this dataset [b-Robitza].

Quality level	Video bitrate (kbit/s)	Resolution (pixel)
QL7	10 000	1 080
QL6	2 500	1 080
QL4	500	480
QL2	150	240

Input data specifications

The input data of this Recommendation is the adaptation set information, a sequence that is a history of selected media quality levels during a media session, stalling information during a media session and the final media session quality score. The input data are specified in Table I.2, where the listed input data is based on ITU-T P.1203.1 mode 0, [ITU-T P.1203.2] and ITU-T P.1203.3].

Table I.2 – Input specifications

Description	Values	Frequency
The resolution of the image displayed to the user	Number of pixels ($W \times H$) in displayed video	Per media chunk
The device type on which the media is played	"PC", "TV", "MO", "TA"	Per media chunk
Device display size	Display size (diagonal in inches)	Per media chunk
Relative viewing distance in multiple of display height	Relative viewing distance	Per media chunk
Target Audio bitrate	Bitrate in kilobits per second	Per media segment
Segment duration	Duration in seconds	Per media segment
Audio codec	One of: AAC-LC, AAC-HEv1, AAC-HEv2, AC3	Per media segment
Audio sampling frequency	In hertz	Per media segment
Number of audio channels	2	Per media segment

Table I.2 – Input specifications

Description	Values	Frequency
Video target bitrate	Bitrate in kilobits per second	Per media chunk
Video frame rate	Frame rate in frames per second.	Per media chunk
Segment duration	Duration in seconds	Per media chunk
Video encoding resolution	Number of pixels ($W \times H$) in transmitted video	Per media chunk
Video codec and profile	ITU-T H.264-high	Per media chunk
Stalling or initial loading event start	The start time of the stalling event in seconds relative to the start of the original video clip, expressed in media time (not wall clock time) NOTE – This is 0 for initial loading delay.	Per stalling event
Event duration	The duration of the stalling event in seconds	Per stalling event
Final media session quality score	Scores provided per session and on a quality scale of 1 to 5	Per session
Legend – AAC: advanced audio coding; AC3: audio compression 3; HE: high efficiency; LC: low complexity; MO: mobile; MPEG: Moving Picture Experts Group; PC: personal computer; TA: tablet; TV: television: VP9: video payload type 9.		

Example of calculation steps

An example calculation step is described. Assume a sequence and adaptation set as follows.

$$H = [QL4, QL6, QL2, QL2, QL7], N = \{QL7, QL6, QL4, QL2, s\},$$

where, H has no stalling. The first step of the calculation is to select x_j from N and determine all of subset Z. Assume that the selected x_j in the iteration of step 1 is $QL2$,

$$x_j = QL2,$$

and Z will be determined, according to Table 2, as all the patterns of subsets of N without x_j, which is

$$Z = \{z_1 = \{QL4\}, z_2 = \{QL6\}, z_3 = \{QL7\}, z_4 = \{s\}, z_5 = \{QL4, QL6\}, z_6 = \{QL4, QL7\}, z_7 = \{QL4, s\}, z_8 = \{QL6, QL7\}, z_9 = \{QL6, s\}, z_{10} = \{QL7, s\}, z_{11} = \{QL4, QL6, QL7\}, z_{12} = \{QL4, QL6, s\}, z_{13} = \{QL4, QL7, s\}, z_{14} = \{QL6, QL7, s\}, z_{15} = \{QL4, QL6, QL7, s\}, z_{16} = \{\}\}.$$

Then, z_i is selected from Z to modify the sequence H. For instance, if z_1 is selected,

$$z_1, \{z_1 \cup x_j\} = \{QL2, QL4\}.$$

Then, by using Equation (2), the modified sequence is generated. In this case, since QL2 and QL4 are in subsets z_1 and $\{z_1 \cup x_j\}$, H' and H'' become as follows.

$$H' \quad = [\boldsymbol{QL7}, QL6, QL2, QL2, QL7], v(H') = 2.880$$

$$H'' = [\boldsymbol{QL7}, QL6, \boldsymbol{QL7}, \boldsymbol{QL7}, QL7], v(H'') = 4.885$$

The bold segments are replaced with the highest media quality level in N and the final media session quality score of both sequence H' and H'' is calculated using ITU-T P.1203 mode 0. Then, the part of the contribution values is calculated as follows.

$$\frac{|z_1|!\,(|N| - |z_1| - 1)!}{|N|!}\Big(v(f(z_1, H)) - v(f(z_1 \cup x_j, H)\Big) = \frac{1!\,3!}{5!}(2.880 - 4.885) = -0.100.$$

Note that, the outputs of v is rounded to the fourth decimal place.

This process is executed over all elements in Z as follows. From here onwards, the term

$$\frac{|z_i|!\,(|N| - |z_i| - 1)!}{|N|!}\Big(v(f(z_i, H)) - v(f(z_i \cup x_j, H)\Big)$$

is denoted E.

– For z_2
  - $z_2, \{z_2 \cup x_j\} = \{QL2, QL6\}$.
  - $H' = [QL4, \boldsymbol{QL7}, QL2, QL2, QL7], v(H') = 2.822,$
  - $H'' = [QL4, \boldsymbol{QL7}, \boldsymbol{QL7}, \boldsymbol{QL7}, QL7], v(H'') = 4.425$
  - $E = \frac{1!3!}{5!}(2.822 - 4.425) = -0.080$

– For z_3
  - $z_3, \{z_3 \cup x_j\} = \{QL2, QL7\}$.
  - $H' = [QL4, QL6, QL2, QL2, QL7], v(H') = 2.822,$
  - $H'' = [QL4, QL6, \boldsymbol{QL7}, \boldsymbol{QL7}, QL7], v(H'') = 4.423$
  - $E = \frac{1!3!}{5!}(2.822 - 4.423) = -0.080$

– For z_4
  - $z_4, \{z_4 \cup x_j\} = \{QL2, s\}$.
  - $H' = [QL4, QL6, QL2, QL2, QL7], v(H') = 2.822,$
  - $H'' = [QL4, QL6, \boldsymbol{QL7}, \boldsymbol{QL7}, QL7], v(H'') = 4.423$
  - $E = \frac{1!3!}{5!}(2.822 - 4.423) = -0.080$

– For z_5
  - $z_5, \{z_5 \cup x_j\} = \{QL2, QL4, QL6\}$.
  - $H' = [\boldsymbol{QL7}, \boldsymbol{QL7}, QL2, QL2, QL7], v(H') = 2.880,$
  - $H'' = [\boldsymbol{QL7}, \boldsymbol{QL7}, \boldsymbol{QL7}, \boldsymbol{QL7}, QL7], v(H'') = 4.896$
  - $E = \frac{2!2!}{5!}(2.880 - 4.896) = -0.067$

– For z_6
  - $z_6, \{z_6 \cup x_j\} = \{QL2, QL4, QL7\}$.
  - $H' = [\boldsymbol{QL7}, QL6, QL2, QL2, QL7], v(H') = 2.880,$
  - $H'' = [\boldsymbol{QL7}, QL6, \boldsymbol{QL7}, \boldsymbol{QL7}, QL7], v(H'') = 4.885$
  - $E = \frac{2!2!}{5!}(2.880 - 4.885) = -0.067$

– For z_7
  - $z_7, \{z_7 \cup x_j\} = \{QL2, QL4, s\}$.

- $H' = [\boldsymbol{QL7}, QL6, QL2, QL2, QL7], v(H') = 2.880,$
- $H'' = [\boldsymbol{QL7}, QL6, \boldsymbol{QL7}, \boldsymbol{QL7}, QL7], v(H'') = 4.885$
- $E = \frac{2!2!}{5!}(2.880 - 4.885) = -0.067$

– For z_8

- $z_8, \{z_8 \cup x_j\} = \{QL2, QL6, QL7\}.$
- $H' = [QL4, \boldsymbol{QL7}, QL2, QL2, QL7], v(H') = 2.822,$
- $H'' = [QL4, \boldsymbol{QL7}, \boldsymbol{QL7}, \boldsymbol{QL7}, QL7], v(H'') = 4.425$
- $E = \frac{2!2!}{5!}(2.822 - 4.425) = -0.053$

– For z_9

- $z_9, \{z_9 \cup x_j\} = \{QL2, QL6, s\}.$
- $H' = [QL4, \boldsymbol{QL7}, QL2, QL2, QL7], v(H') = 2.822,$
- $H'' = [QL4, \boldsymbol{QL7}, \boldsymbol{QL7}, \boldsymbol{QL7}, QL7], v(H'') = 4.425$
- $E = \frac{2!2!}{5!}(2.822 - 4.425) = -0.053$

– For z_{10}

- $z_{10}, \{z_{10} \cup x_j\} = \{QL2, QL7, s\}.$
- $H' = [QL4, QL6, QL2, QL2, QL7], v(H') = 2.822,$
- $H'' = [QL4, QL6, \boldsymbol{QL7}, \boldsymbol{QL7}, QL7], v(H'') = 4.423$
- $E = \frac{2!2!}{5!}(2.822 - 4.423) = -0.053$

– For z_{11}

- $z_{11}, \{z_{11} \cup x_j\} = \{QL2, QL4, QL6, QL7\}.$
- $H' = [\boldsymbol{QL7}, \boldsymbol{QL7}, QL2, QL2, QL7], v(H') = 2.880,$
- $H'' = [\boldsymbol{QL7}, \boldsymbol{QL7}, \boldsymbol{QL7}, \boldsymbol{QL7}, QL7], v(H'') = 4.896$
- $E = \frac{3!1!}{5!}(2.880 - 4.896) = -0.100$

– For z_{12}

- $z_{12}, \{z_{12} \cup x_j\} = \{QL2, QL4, QL6, s\}.$
- $H' = [\boldsymbol{QL7}, \boldsymbol{QL7}, QL2, QL2, QL7], v(H') = 2.880,$
- $H'' = [\boldsymbol{QL7}, \boldsymbol{QL7}, \boldsymbol{QL7}, \boldsymbol{QL7}, QL7], v(H'') = 4.896$
- $E = \frac{3!1!}{5!}(2.880 - 4.896) = -0.100$

– For z_{13}

- $z_{13}, \{z_{13} \cup x_j\} = \{QL2, QL4, QL7, s\}.$
- $H' = [\boldsymbol{QL7}, QL6, QL2, QL2, QL7], v(H') = 2.880,$
- $H'' = [\boldsymbol{QL7}, QL6, \boldsymbol{QL7}, \boldsymbol{QL7}, QL7], v(H'') = 4.885$
- $E = \frac{3!1!}{5!}(2.880 - 4.885) = -0.100$

– For z_{14}

- $z_{14}, \{z_{14} \cup x_j\} = \{QL2, QL6, QL7, s\}.$
- $H' = [QL4, \boldsymbol{QL7}, QL2, QL2, QL7], v(H') = 2.822$
- $H'' = [QL4, \boldsymbol{QL7}, \boldsymbol{QL7}, \boldsymbol{QL7}, QL7], v(H'') = 4.425$
- $E = \frac{3!1!}{5!}(2.822 - 4.425) = -0.080$

– For z_{15}

- $z_{15}, \{z_{15} \cup x_j\} = \{QL2, QL4, QL6, QL7, s\}$.
- $H' = [\boldsymbol{QL7}, \boldsymbol{QL7}, QL2, QL2, QL7], v(H') = 2.880,$
- $H'' = [\boldsymbol{QL7}, \boldsymbol{QL7}, \boldsymbol{QL7}, \boldsymbol{QL7}, QL7], v(H'') = 4.896$
- $E = \frac{4!0!}{5!}(2.880 - 4.896) = -0.403$

– For z_{16}

- $z_{16}, \{z_{16} \cup x_j\} = \{QL2\}$.
- $H' = [QL4, QL6, QL2, QL2, QL7], v(H') = 2.822,$
- $H'' = [QL4, QL6, \boldsymbol{QL7}, \boldsymbol{QL7}, QL7], v(H'') = 4.423$
- $E = \frac{0!4!}{5!}(2.822 - 4.423) = -0.320$

Finally, the contribution value c_j of x_j (i.e., QL2) can be obtained by summing E from $z_1$to z_{16}. In this example case, c_j is −1.807. To calculate contribution values for QL4, QL6, QL7, and stalling, the preceding process is repeated.

Note that as described in clause 6, since all information to estimate the quality of modified sequence H' and H'' is from the metadata-based model, the contribution calculation can be executed on the client device.

Calculation results

Table I.3 shows the calculated contribution values. In Table I.3, final quality (FQ) means the final media session quality score of corresponding PVS calculated by ITU-T P.1203 mode 0. QL7, QL6, QL4, QL2, and stalling columns mean the contribution value of each media quality level. The column total means the total value of contribution value.

In PVS01, since QL7, QL6 and QL2 are not selected and do not induce any quality degradation, the contribution value equals 0. Also, since stalling does not occur or induce any quality degradation, the contribution value equals 0. On the other hand, since QL4 was selected during a media session, it has −1.375 points. This means that QL4 reduces the final media session quality score by approximately 1.375 compared with the maximum media session quality score. In PVS02, QL6, QL4 and QL2 were selected. Thus, each media quality level reduced the FQ, and QL2 reduced the quality the most. In PVS03, since stalling occurred, the contribution value of the stalling was −0.375. In PVS04, since several quality levels were selected, each quality level reduces the quality. Specifically, since stalling occurred twice and media quality level QL2 was selected many times in this PVS, the stalling and QL2 reduced the final media session quality score. These contribution values of each PVS also indicate that if the PVS differs, the contribution values of each PVS also differ since the selected media quality level or stalling occurrence differs between PVSs.

Table I.3 – Calculated results of contribution values for PVS01, PVS02, PVS03, and PVS04

	FQ	**QL7**	**QL6**	**QL4**	**QL2**	**Stalling**	**Total**
PVS01	3.531	0	0	−1.357	0	0	−1.357
PVS02	2.749	0	−0.008	−0.974	−1.157	0	−2.139
PVS03	3.881	0	−0.036	−0.595	0	−0.375	−1.007
PVS04	1.599	0	−0.003	−0.203	−2.320	−0.763	−3.288

Appendix II

Example of the deviation procedure with ITU-T P.1203 mode 3 (bitstream-based model) as a quality estimation model

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

This appendix describes an example of the deviation procedure when ITU-T P.1203 mode 3 is used as a quality estimation model. Especially, PVSs used, input data specification and contribution values are described.

Prepared PVSs

In this example, the same PVSs as those of Appendix I are used. Thus, detailed information about the PVSs is omitted.

Input data specifications

The input data of this Recommendation is the adaptation set information, a sequence that is a history of selected media quality level during a media session, stalling information during a media session, and the final media session quality score. In addition to that, since ITU-T P.1203 mode 3 is a bitstream-based model, the bitstream information at each segment is used. The specifications of input data are in Table II.1, where the listed input data is based on ITU-T P.1203 mode 3, [ITU-T P.1203.2], [ITU-T P.1203.3].

Table II.1 – Input specifications

Description	Values	Frequency
The resolution of the image displayed to the user	Number of pixels ($W \times H$) in displayed video	Per media chunk
The device type on which the media is played	"PC", "TV", "MO", "TA"	Per media chunk
Device display size	Display size (diagonal in inches)	Per media chunk
Relative viewing distance in multiple of display height	Relative viewing distance	Per media chunk
Target audio bitrate	Bitrate in kilobits per second	Per media segment
Segment duration	Duration in seconds	Per media segment
Audio frame number	Integer, starting with 1	Per media segment
Audio frame size	Size of the frame in bytes	Per audio frame
Audio frame duration	Duration in seconds	Per audio frame
Audio codec	One of: AAC-LC, AAC-HEv1, AAC-HEv2, AC3	Per media segment
Audio sampling frequency	In hertz	Per media segment
Number of audio channels	2	Per media segment
Audio bitstream	Encoded audio bytes for the frame	Per audio frame
Video bitrate	Bitrate in kilobits per second	Per media chunk

Table II.1 – Input specifications

Description	Values	Frequency
Video frame rate	Frame rate in frames per second	Per media chunk
Segment duration	Duration in seconds	Per media chunk
Video encoding resolution	Number of pixels ($W \times H$) in transmitted video	Per media chunk
Video codec and profile	ITU-T H.264-high	Per media chunk
Video frame number	Integer, starting at 1, denoting the frame sequence number in encoding order	Per video frame
Video frame duration	Duration of the frame in seconds	Per video frame
Video frame size	The size of the encoded video frame in bytes	Per video frame
Type of each picture	"I"/"P"/"B"	Per video frame
Video bitstream	Encoded video bytes for the frame	Per video frame
Stalling or initial loading event start	The start time of the stalling event in seconds relative to the start of the original video clip, expressed in media time (not wall clock time) NOTE – This is 0 for initial loading delay.	Per stalling event
Event duration	The duration of the stalling event in seconds	Per stalling event
Final media session quality score	Scores provided per session and on a quality scale of 1 to 5	Per session

Example of calculation steps

Most parts of the calculation step are the same as in Appendix I. Since the bitstream information is needed in ITU-T P.1203 mode 3, generation of the modified sequence is slightly different. Assume that the same N, x_j, S as in Appendix I and a sequence H as follows.

$$H = \left[(QL4, b_{1,ql4}), (QL6, b_{2,ql6}), (QL2, b_{3,ql2}), \left(QL2, b_{4,ql2}\right), (QL7, b_{5,ql7})\right],$$

where $b_{j,qlk}$ represents a tuple of video frame size, type of each picture and quantization parameter values of media quality level k at segment j that are parsed from video bitstream. Then, the modified sequences become as follows.

$$H' = \left[(\boldsymbol{QL7}, \boldsymbol{b_{1,ql7}}), (QL6, b_{2,ql6}), (QL2, b_{3,ql2}), \left(QL2, b_{4,ql2}\right), (QL7, b_{5,ql7})\right],$$

$$H'' = \left[(\boldsymbol{QL7}, \boldsymbol{b_{1,ql7}}), (QL6, b_{2,ql6}), (\boldsymbol{QL7}, \boldsymbol{b_{3,ql7}}), \left(\boldsymbol{QL7}, \boldsymbol{b_{4,ql7}}\right), (QL7, b_{5,ql7})\right].$$

Here, $b_{j,ql7}$is obtained as an input. Different from the metadata-based model, ITU-T P.1203 mode 0 in Appendix I, the media quality level and bitstream information at bold segments are replaced with the highest media quality level and bitstream information of the highest media quality level at those segments. Here, since the bitstream information of the highest quality level at each segment is placed at the streaming server, the contribution calculation is executed at the streaming server. The other steps are the same as those in Appendix I.

Calculation results

The calculated contribution values are listed in Table II.2, in which FQ means the final media session quality score of the corresponding PVS calculated by ITU-T P.1203 mode 3. QL7, QL6, QL4, QL2, and stalling columns mean the contribution value of each media quality level. The column total means the total value of the contribution value.

Most calculated contribution values tend to be the same as those of ITU-T P.1203 mode 0 in Appendix I. However, since the quality estimation model and final quality media session quality score differ, the contribution values also differ from those of ITU-T P.1203 mode 0.

In PVS01, since QL7, QL6, and QL2 are not selected and do not induce any quality degradation, the contribution value equals 0. Also, since stalling does not occur or induce any quality degradation, the contribution value equals 0. On the other hand, since QL4 was selected during the media session, it was −2.197 points. In PVS02, QL6, QL4, and QL2 were selected. Thus, each media quality level reduced the FQ. In PVS03, since a stalling occurred, the contribution value of the stalling was −0.337. In PVS04, since several quality levels were selected, each quality level reduces the quality. Specifically, since stalling occurred twice and media quality level QL2 was selected many times in this PVS, the stalling and QL2 reduced the final media session quality score.

Table II.2 – Calculated results of contribution values for PVS01, PVS02, PVS03 and PVS04

	FQ	**QL7**	**QL6**	**QL4**	**QL2**	**Stalling**	**Total**
PVS01	2.631	0	0	−2.197	0	0	−2.197
PVS02	2.310	0	−0.146	−1.383	−0.990	0	−2.518
PVS03	3.192	0	−0.200	−1.099	0	−0.337	−1.636
PVS04	1.569	0	−0.008	−0.275	−2.256	−0.721	−3.259

Appendix III

Example of the deviation procedure with [ITU-T P.1204.4] (pixel-based model) as a quality estimation model

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

This appendix describes an example of the deviation procedure when [ITU-T P.1204.4] is used as a quality estimation model. Especially, the PVSs used, input data specification and contribution values are described.

Prepared PVSs

In this example, the same PVSs as those in Appendix I are used. Thus, detailed information about PVSs is omitted.

Input data specifications

The input data of this Recommendation is the adaptation set information, a sequence that is a history of selected media quality level during a media session, stalling information during a media session, and the final media session quality score. In addition to that, since [ITU-T P.1204.4] is a pixel-based model, the video pixel information at each segment is used. The input data are specified in Table III.1, where the listed input data is based on [ITU-T P.1204.4], [ITU-T P.1203.2], [ITU-T P.1203.3].

Table III.1 – Input specifications

Description	Values	Frequency
The resolution of the image displayed to the user	Number of pixels ($W \times H$) in displayed video	Per media chunk
The device type on which the media is played	"PC", "TV", "MO", "TA"	Per media chunk
Device display size	Display size (diagonal in inches)	Per media chunk
Relative viewing distance in multiples of display height	Relative viewing distance	Per media chunk
Target audio bit-rate	Bitrate in kilobits per second	Per media segment
Segment duration	Duration in seconds	Per media segment
Audio codec	One of: AAC-LC, AAC-HEv1, AAC-HEv2, AC3	Per media segment
Audio sampling frequency	In hertz	Per media segment
Number of audio channels	2	Per media segment
Stalling or initial loading event start	The start time of the stalling event in seconds relative to the start of the original video clip, expressed in media time (not wall clock time) NOTE – This is 0 for initial loading delay.	Per stalling event

Table III.1 – Input specifications

Description	Values	Frequency
Event duration	The duration of the stalling event in seconds	Per stalling event
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 to display resolution without buffering or stalling. The frame information in degraded video and reference video is synchronized, i.e., no frame misalignments are present	Per media chunk
Reference video information	The reference-side information extraction module takes as input the reference video and outputs the side information file. The reference model side channel bandwidth limit is 256 kbit/s. Thus, the side information of the reference model for a video sequence v is stored in a file with size at most 256/8 * t_v kB, where t_v is the duration of video v in seconds.	Per media chunk
Final media session quality score	Scores provided per session and on a quality scale of 1 to 5	Per session

Example of calculation steps

Most parts of the calculation step are the same as those in Appendix I. Since the pixel information is needed in [ITU-T P.1204.4], a sequence and modified sequences are slightly different. Assume the same N, x_j, S, H as in Appendix I. In addition to that in the pixel-based model, the pixel information of that sequence H,

$$H = \left[(QL4, p_{1,ql4,}), (QL6, p_{2,ql6}), (QL2, p_{3,ql2}), \left(QL2, p_{4,ql2}\right), (QL7, p_{5,ql7})\right],$$

where $p_{j,qlk}$ represents the degraded video pixel information of media quality level k at segment j . Then, the modified sequences become as follows.

$$H' = \left[(\boldsymbol{QL7}, \boldsymbol{p_{1,ql7}}), (QL6, p_2), (QL2, p_3), (QL2, p_4), (QL7, p_5)\right],$$

$$H'' = \left[(\boldsymbol{QL7}, \boldsymbol{p_{1,ql7}}), (QL6, p_2), (\boldsymbol{QL7}, \boldsymbol{p_{3,ql7}}), \left(\boldsymbol{QL7}, \boldsymbol{p_{4,ql7}}\right), (QL7, p_5)\right].$$

Also, when the final media session quality score of H, H' and H'' is calculated using [ITU-T P.1204.4], reference video pixel information is used.

Here, since the pixel information of the highest media quality level at each segment is placed at the streaming server, the contribution calculation is executed at the streaming server. The other steps are the same as those in Appendix I.

Calculation results

Table III.2 shows the calculated contribution values. In Table III.2, FQ means the final media session quality score of the corresponding PVS calculated by [ITU-T P.1204.4]. QL7, QL6, QL4, QL2, and stalling columns mean the contribution value of each media quality level. The column total means the total value of the contribution value.

Most calculated contribution values tend to be the same as those of ITU-T P.1203 mode 0 in Appendix I. However, since the quality estimation model and final quality media session quality score differ, the contribution values also differ from those of ITU-T P.1203 mode 0.

In PVS01, since QL7, QL6, and QL2 are not selected and do not induce any quality degradation, the contribution value equals 0. Also, since stalling does not occur or induce any quality degradation, the contribution value equals 0. On the other hand, since QL4 was selected during a media session, it was the −0.965 points. In PVS02, QL6, QL4 and QL2 were selected. Thus, each media quality level reduced the FQ. In PVS03, since stalling occurred, its contribution value was −0.051. In PVS04, since several quality levels were selected, each quality level reduced the quality. Specifically, since stalling occurred twice and media quality level QL2 was selected many times in this PVS, the stalling and QL2 reduced the final media session quality score.

Table III.2 – Calculated results of contribution values for PVS01, PVS02, PVS03, and PVS04

	FQ	**QL7**	**QL6**	**QL4**	**QL2**	**Stalling**	**Total**
PVS01	3.442	0	0	−0.965	0	0	−0.965
PVS02	3.243	0	−0.064	−0.511	−0.590	0	−1.164
PVS03	3.775	0	−0.225	−0.375	0	−0.051	−0.632
PVS04	1.871	0	−0.002	−0.040	−1.675	−0.818	−2.536

Bibliography

[b-ITU-T H.264] Recommendation ITU-T H.264 (2021), *Advanced video coding for generic audiovisual services*.

[b-ITU-T H.265] Recommendation ITU-T H.265 (2023), *High efficiency video coding*.

[b-ITU-T P.10] Recommendation ITU-T P.10/G.100 (2017), *Vocabulary for performance, quality of service and quality of experience*.

[b-Robitza] Robitza, W., Göring, S., Raake, A., Lindegren, D., Heikkilä, G., Gustafsson,J., List, P., Feiten, B., Wüstenhagen, U., Garcia, M.-N., Yamagishi, K., Broom, S. (2018). HTTP adaptive streaming QoE estimation with ITU-T Rec. P.1203 – Open databases and software. In: *9th ACM Multimedia Systems Conference*, Amsterdam, 2018-06-12/15, pp. 466-471. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3204949.3208124

[b-Shapley] Shapley, L.S. (1953). A value for *n*-person games. In: Kuhn, H., Tucker, A., editors. *Contributions to the theory of games*, vol. 2, pp. 307–317. Princeton, NJ: Princeton University Press. https://doi.org/10.1515/9781400881970-018

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

Published in Switzerland
Geneva, 2023