International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.1203.2

(10/2017)

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 – Audio quality estimation module

Recommendation ITU-T P.1203.2

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	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.1203.2

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

Summary

Recommendation ITU-T P.1203.2 specifies the short-term audio quality estimation module for Recommendation ITU-T P.1203. The ITU-T P.1203 series of ITU-T Recommendations specifies modules for a set of model algorithms for monitoring the integral media session quality for transport control protocol (TCP) type video streaming. The models comprise modules for short-term video-quality and audio-quality estimation (the latter specified in this Recommendation). The per-one-second outputs of these short-term modules are integrated into estimates of audio-visual quality and together with information about initial loading delay and media playout stalling events, they are further integrated into the final model output, the estimate of integral quality. The respective ITU-T work item has formerly been referred to as "Parametric non-intrusive assessment of TCP-based multimedia streaming quality" or "P.NATS". The Recommendation ITU-T P.1203.2 part of Recommendation ITU-T P.1203 provides details for the module for bitstream-based, short-term audio quality estimation.

Only one audio module is recommended for all four modes 0 to 3 of the Recommendation ITU-T P.1203 model series, corresponding to mode 0. The model is identical to the audio coding quality estimation component of the user datagram protocol (UDP) streaming related prediction model described in Recommendation ITU-T P.1201.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T P.1203.2	2016-11-29	12	11.1002/1000/13160
2.0	ITU-T P.1203.2	2017-10-29	12	11.1002/1000/13401

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.1203.2

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

1 Scope

This RecommendatioSn describes the short term audio quality estimation module which is an integral part of the ITU-T P.1203 series. [ITU-T P.1203] describes a set of objective parametric quality assessment modules. Combined, these modules can be used to predict the impact of audio and video media encodings as well as Internet protocol (IP) network impairments on the quality experienced by an end-user of multi-media streaming applications.

The addressed streaming techniques comprise progressive download as well as adaptive streaming, for both mobile and fixed network streaming applications over transport control protocol (TCP) or other TCP like protocols which are not affected by transmission errors.

The model described is restricted to information provided to it by an appropriate packet- or bitstream-analysis module. The overall ITU-T P.1203 model is applicable for the effects due to audio- and video-coding as well as initial loading delay and stalling (which are both caused by rebuffering at the client) as the typical degradations associated with progressive download. As final output, the ITU-T P.1203 series models target integral audio-visual media quality scores.

This Recommendation describes only one audio quality module. With regard to the required input data, this audio module corresponds to mode 0 of [ITU-T P.1203].

The same, purely header-based/bitrate-based audio quality module is also specified in [ITU-T P.1201.2]. Using a large number of subjective experiments, it was validated that this model also leads to accurate predictions within the scope of [ITU-T P.1203].

The audio module predicts mean opinion scores (MOS) on a 5-point absolute category rating (ACR) scale (see [ITU-T P.910]) as a per-one-second MOS score.

During the development of [ITU-T P.1201], explicit short-term audio quality tests were carried out in order to validate the stand-alone use of the audio module for the estimation of audio-only quality. It could be shown within the scope of [ITU-T P.1201] that this is possible.

It must be noted however, that since the subjective tests conducted for [ITU-T P.1201] included packet loss degradations, range-equalization and other biases may need to be considered (see for example [b-Zielinski_2008]) if the module is to be used stand-alone within the scope of [ITU-T P.1203].

This model cannot provide a comprehensive evaluation of audio transmission quality as perceived by an *individual* end user because its scores reflect the impairments due to audio coding only. Furthermore, the scores predicted by a parametric model necessarily reflect an average perceptual impairment. Note also that the model was developed and validated for one specific encoder and decoder implementation. If a different encoder and decoder pair is used in a monitoring situation the scores may not reflect that.

Effects such as audio level or noise (and corresponding similar audio factors) or other impairments related to the audio signals are not reflected in the scores computed by this model. Moreover, the scores predicted by a parametric model (i.e., without access to payload information, such as the audio signals) necessarily reflect a somewhat simplified representation of the perceptual impairment of the considered stream.

However, presuming that it is applied in an appropriate manner, according to this Recommendation, the model still enables estimation of some coding quality related information and thus valid and in most cases accurate predictions.

Tables 1.1 and 1.2 indicate the areas and parameter ranges for which the Pa module specified in this Recommendation has been validated and for which applications it can be used, with some caution.

Table 1.1 – Application areas, test factors and coding technologies where ITU-T P.1203.2 for adaptive streaming and progressive download has been verified and is known to produce reliable results

Applications for which the model is intended	
In-service monitoring of TCP-based audio. Both so called over the top (OTT) services (for example YouTube) and operator managed video services (over TCP), using the protocols HTTP/TCP/IP and RTMP/TCP/IP. Note that this model is agnostic to the type of container format (e.g. Flash (FLV), MP4, WebM or 3GP.	
Performance and quality assessment of live networks (including codecs) considering the effect due to encoding bit rate.	
Audio test factors for which the model has been validated	
Input audio length	Maximum 20 seconds. The video model produces a per-second score considering input data from a measurement window of max. 20 s length.
Bitstream container	Coded audio bitstream contained in MPEG-2 transport stream (TS) segments
Encoder/Decoder implementation	The model has been trained using the following audio encoder: – AAC-LC: libfdk_aac, low complexity (LC) mode (ffmpeg). – A common framework was developed based on the above codec, all the test data was generated using the common framework.
Audio sample rate	48 000 samples/s
Audio bit rate	16, 32, 64 and 98 kBit/s/channel Audio bit rate was always varied in a correlated fashion with the video bit rate, i.e., high video bit rate corresponds to high audio bit rate and vice versa. Bearing to this condition it has been observed that audio quality has very little effect on the overall audio-visual quality.
Segment length	1-9 seconds NOTE – The segment length determines how often the audio quality can be adapted.
Audio channels	2 (stereo)

Table 1.2 – Application areas, test factors and coding technologies for which ITU-T P.1203.2 is assumed to give valid results

Test factors where the model can be used but the results may not be reliable (conditions not included in subjective tests underlying the model development)
All factors as indicated in Table 1.1, with additions as described below: Codecs: HE-AACv2, AC3, MPEG-LII Bit rates: 4.75-576 kbit/s
NOTE – ITU-T P.1203 was tested on AAC-LC only. The audio module alone has been tested with the codecs mentioned above with dedicated audio-quality tests during [ITU-T P.1201] development.

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.1201.1] Recommendation ITU-T P.1201.1 (2012), *Parametric non-intrusive assessment of audiovisual media streaming quality – Lower resolution application area*.

[ITU-T P.1201.2] Recommendation ITU-T P.1201.2 (2012), *Parametric non-intrusive assessment of audiovisual media streaming quality – Higher resolution application area*.

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

[ITU-T P.1202.1] Recommendation ITU-T P.1202.1 (2012), *Parametric non-intrusive bitstream assessment of video media streaming quality – Lower resolution application area*.

[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.1203.3] Recommendation ITU-T P.1203.3 (2016), *Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport –Quality integration 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].

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

3.2 Terms defined in this Recommendation

None.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

AAC Advanced Audio Coding
AAC-LC Advanced Audio Coding – Low Complexity
AC3 Audio Coding 3
ACR Absolute Category Rating
ARQ Automatic Repeat Request
FEC Forward Error Correction
HE-AAC High-Efficiency Advanced Audio Coding
HTTP Hypertext Transfer Protocol
IP Internet Protocol
MOS Mean Opinion Score
MPEG Moving Pictures Expert Group
OTT Over The Top
TCP Transport Control Protocol
TS Transport Stream
UDP User Datagram Protocol

5 Conventions

None.

6 Pa module in ITU-T P.1203 context

The overall model structure is shown in Figure 6-1, highlighting the position of the *Pa* module. More details on the general structure can be found in the introductory [ITU-T P.1203].

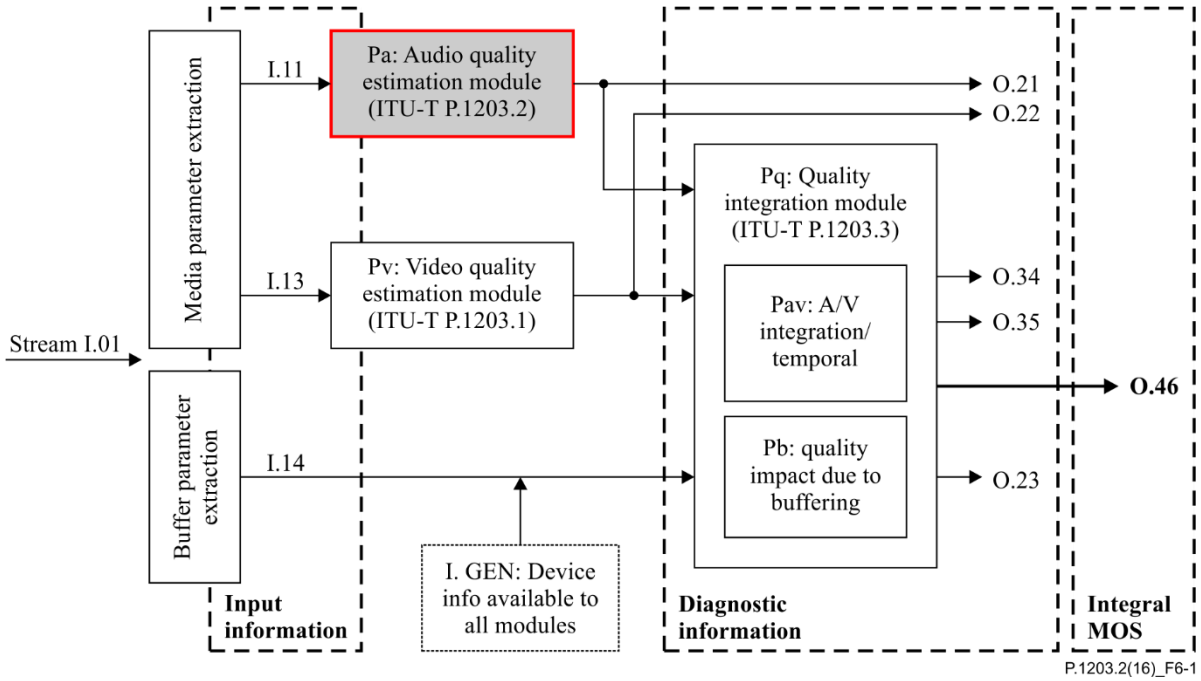

Figure 6-1 – *Pa* module in context of building blocks of the ITU-T P.1203 model

6.1 *Pa* module modes

The modes of operation for ITU-T P.1203.2 are defined in the Table 6-1. Detailed information on exactly which inputs are available for each mode is provided in Table 7-1. A single model is specified for all modes and is described in clause 8.

Table 6-1 –ITU-T P.1203.2 modes

Mode	Encryption	Input	Complexity	Comments
0	Encrypted media payload and media frame headers	Meta-data	Low	Module defined in this Recommendation
1	Encrypted media payload	Meta-data and frame header information	Low (see comments)	Same as mode 0
2	No encryption	Meta-data and up-to 2% of the media stream	Medium (see comments)	Same as mode 0
3	No encryption	Meta-data and any information from the video stream	Unlimited (see comments)	Same as mode 0

7 Model input

The model receives media information and prior knowledge about the media stream. The audio quality module receives the following input signals, regardless of the mode of operation, following the measurement window-based procedure as specified in [ITU-T P.1203], clause 7.4:

I.11: Audio coding information, as specified in [ITU-T P.1203], clause 7.1.

Note that fault correction techniques, such as automatic repeat request (ARQ) and forward error correction (FEC) used for user datagram protocol (UDP) based streaming are not applicable for this case, where the streaming is TCP based. In TCP-based transport all retransmissions and packet loss information is typically handled transparently by the transport layer and while it can be available to

the models described in this Recommendation it is not needed. The only information provided to the model that may implicitly include effects such as packet loss and respective retransmission is the initial loading delay and stalling information provided to the quality integration module (see [ITU-T P.1203.3]).

7.1 I.11 input specification

Since the audio quality module for mode 0 will be used as a component for the ITU-T P.1203 series models, I.11 consists of two parameters:

– Audio codec

– Bit rate in kbit/s

Details can be found in Table 7-1.

Note that the actual information available to the module at a specific output sample timestamp is restricted by the measurement window as defined in [ITU-T P.1203], clause 7.4.

Table 7-1 – Description of I.11

ID	Description	Values	Frequency	Modes available
I.GEN				
0	The resolution of the image displayed to the user	Number of pixels (WxH) in displayed video	Per media session	All
1	The device type on which the media is played	PC or mobile	Per media session	All
I.11				
2	Target audio bit rate	Bit rate in kbit/s.	Per media segment	All
3	Segment duration	Duration in seconds	Per media segment	All
4	Audio frame number	Integer, starting with 1	Per media segment	1,2,3
5	Audio frame size	Size of the frame in bytes	Per audio frame	1,2,3
6	Audio frame duration	Duration in seconds	Per audio frame	1,2,3
7	Audio codec	One of: AAC-LC, AAC-HEv1, AAC-HEv2, AC3	Per media segment	All
8	Audio sampling frequency	Hz	Per media segment	All
9	Number of audio channels	2	Per media segment	All
10	Audio bit-stream	Encoded audio bytes for the frame	Per audio frame	2,3

8 Model algorithm and output

The [ITU-T P.1203.2] model for audio has one output, O.21. It provides output values on the 5-point ACR scale ("MOS") per output sampling interval.

One single audio quality module is recommended to be used in the ITU-T P.1203 series models. This audio quality module algorithm is the same as the one specified in [ITU-T P.1201.2]. It is summarized here for completeness:

$$O.21 = MOSfromR(QA) \qquad \text{(Eq. 13d in [ITU-T P.1201.2])}$$

with:

$$QA = 100 - QcodA \qquad \text{(Eq. 13c in [ITU-T P.1201.2])}$$

with coding degradations only, i.e., with QtraA = 0)

where:

$$QcodA = a1A \times \exp(a2A \times Bitrate) + a3A \qquad \text{(Eq. 13a in [ITU-T P.1201.2])}$$

Bit rate is the audio bit rate in kBit/s.

The function $MOSfromR$ is given in Annex E of [ITU-T P.1203.1] and is provided below:

$$MOSfromR: \mathbb{R} \longmapsto \mathbb{R}$$

$$Q \longmapsto MOS := MOSfromR(Q)$$

$$MOS = MOS_{MIN} + (MOS_{MAX} - MOS_{MIN}) * \frac{Q}{100} + Q * (Q - 60) * (100 - Q) * 0.000007 \quad (E.1)$$

$$MOS = \min(MOS_{MAX}, \max(MOS, MOS_{MIN})) \quad (E.2)$$

where $MOS_{MAX} = 4.9$ and $MOS_{MIN} = 1.05$.

Coefficients a1A, a2A and a3A depend on the audio codec. These audio model coefficients are provided in Table 8-1:

Table 8-1 – Audio model coefficients for different audio codecs (coding degradations only), adapted from Table 1 of [ITU-T P.1201.2]

Audio codec	a1A	a2A	a3A
MPEG1 L2	100.0	–0.02	15.48
AC3	100.0	–0.03	15.70
AAC-LC	100.0	–0.05	14.60
HE-AAC v2	100.0	–0.11	20.06

Bibliography

[b-Zielinski_2008] Slawomir Zielinski, Soren Bech and Francis Rumsey (2008), *On some biases encountered in modern audio quality listening tests – A review, Journal Audio Engineering Society (JAES)*, 56(6), 427-451.
<http://www.aes.org/e-lib/browse.cfm?elib=14393>

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, 2017