Multimedia Quality of Service and performance – Generic and user-related aspects

Opinion model for network planning of video and audio streaming applications

Recommendation ITU-T G.1071
<table>
<thead>
<tr>
<th>Classification</th>
<th>Recommendation Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>International Telephone Connections and Circuits</td>
<td>G.100–G.199</td>
</tr>
<tr>
<td>General Characteristics Common to All Analogue Carrier-Transmission Systems</td>
<td>G.200–G.299</td>
</tr>
<tr>
<td>Individual Characteristics of International Carrier Telephone Systems on Metallic Lines</td>
<td>G.300–G.399</td>
</tr>
<tr>
<td>General Characteristics of International Carrier Telephone Systems on Radio-Relay or Satellite Links and Interconnection with Metallic Lines</td>
<td>G.400–G.449</td>
</tr>
<tr>
<td>Coordination of Radiotelephony and Line Telephony</td>
<td>G.450–G.499</td>
</tr>
<tr>
<td>Transmission Media and Optical Systems Characteristics</td>
<td>G.600–G.699</td>
</tr>
<tr>
<td>Digital Terminal Equipments</td>
<td>G.700–G.799</td>
</tr>
<tr>
<td>Digital Networks</td>
<td>G.800–G.899</td>
</tr>
<tr>
<td>Digital Sections and Digital Line System</td>
<td>G.900–G.999</td>
</tr>
<tr>
<td>Multimedia Quality of Service and Performance – Generic and User-Related Aspects</td>
<td>G.1000–G.1999</td>
</tr>
<tr>
<td>Transmission Media Characteristics</td>
<td>G.6000–G.6999</td>
</tr>
<tr>
<td>Data over Transport – Generic Aspects</td>
<td>G.7000–G.7999</td>
</tr>
<tr>
<td>Packet over Transport Aspects</td>
<td>G.8000–G.8999</td>
</tr>
<tr>
<td>Access Networks</td>
<td>G.9000–G.9999</td>
</tr>
</tbody>
</table>

For further details, please refer to the list of ITU-T Recommendations.
Recommendation ITU-T G.1071

Opinion model for network planning of video and audio streaming applications

Summary
Recommendation ITU-T G.1071 provides algorithmic models for network planning of IP-based video services. This Recommendation addresses two application areas:
– the higher resolution (HR) application area, including services such as IPTV;
– the lower resolution (LR) application area, including services such as mobile TV.
The algorithmic model addressing the HR application area is described in Annex A.
The algorithmic model addressing the LR application area is described in Annex B.
The application of the models is limited to QoE/QoS planning. Other applications such as quality benchmarking and monitoring are outside the scope of this Recommendation.
As input, the models take network planning assumptions, for example, the video resolution, the audio and video codec types and profiles, the audio and video bitrates, the packet-loss-rate and the packet-loss distribution.
As output, the model algorithms provide individual estimates of audio, video and audiovisual quality in terms of the five-point absolute category rating (ACR) mean opinion score (MOS) scale. Further diagnostic information on causes of quality degradations can also be made available.

History

<table>
<thead>
<tr>
<th>Edition</th>
<th>Recommendation</th>
<th>Approval</th>
<th>Study Group</th>
<th>Unique ID*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>ITU-T G.1071</td>
<td>2015-06-29</td>
<td>12</td>
<td>11.1002/1000/12512</td>
</tr>
</tbody>
</table>

Keywords
Audio, audiovisual, IPTV, mean opinion score (MOS), mobile TV, network planning, quality, QoE, 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/.

© ITU 2015

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Scope</td>
<td>1</td>
</tr>
<tr>
<td>2  References</td>
<td>1</td>
</tr>
<tr>
<td>3  Definitions</td>
<td>2</td>
</tr>
<tr>
<td>3.1 Terms defined elsewhere</td>
<td>2</td>
</tr>
<tr>
<td>3.2 Terms defined in this Recommendation</td>
<td>2</td>
</tr>
<tr>
<td>4  Abbreviations and acronyms</td>
<td>2</td>
</tr>
<tr>
<td>5  Conventions</td>
<td>3</td>
</tr>
<tr>
<td>6  Areas of application</td>
<td>4</td>
</tr>
<tr>
<td>7  Model framework</td>
<td>7</td>
</tr>
<tr>
<td>8  Model output information and performance details</td>
<td>7</td>
</tr>
<tr>
<td>Annex A – Description of the ITU-T G.1071 model algorithm for HR application area</td>
<td>9</td>
</tr>
<tr>
<td>A.1 Audio module</td>
<td>9</td>
</tr>
<tr>
<td>A.2 Video module</td>
<td>11</td>
</tr>
<tr>
<td>A.3 Audiovisual module</td>
<td>16</td>
</tr>
<tr>
<td>Annex B – Description of the ITU-T G.1071 model algorithm for LR application area</td>
<td>17</td>
</tr>
<tr>
<td>B.1 Audio module</td>
<td>17</td>
</tr>
<tr>
<td>B.2 Video module</td>
<td>19</td>
</tr>
<tr>
<td>B.3 Audiovisual module</td>
<td>24</td>
</tr>
</tbody>
</table>
Recommendation ITU-T G.1071

Opinion model for network planning of video and audio streaming applications

1 Scope
This Recommendation provides models which deliver estimates of the impact of typical IP network impairments on the quality experienced by the end user in multimedia mobile streaming and IPTV applications over transport formats such as: RTP (over UDP), MPEG2-TS (over UDP or RTP/UDP), 3GPP-PSS (over RTP).

The models are network planning tools. They are of help in selecting IP-network transmission settings such as the audio and video format, the audio and video codecs and the audio and video bitrates, under the assumption that the network is prone to packet loss.

This Recommendation targets the same services as [ITU-T P.1201] and [ITU-T P.1202]. In particular, this Recommendation covers the same coding technologies and unreliable network mechanisms such as user datagram protocol (UDP). However, this Recommendation is limited to QoE/QoS planning, while [ITU-T P.1201] and [ITU-T P.1202] are dedicated to service monitoring and benchmarking.

The following Recommendations have also been developed for QoE/QoS planning, but for different applications:
- [ITU-T G.107]: speech (telephone band)
- [ITU-T G.107.1]: speech (wideband)
- [ITU-T G.1070]: videophone
- [ITU-T G.1030]: Appendix II: web browsing

For a summary of the services, encoding and network characteristics covered by the ITU-T G.1071 model algorithms, refer to Table 1 of clause 6.

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.


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]: The mean of opinion scores, i.e., of the values on a predefined scale that subjects assign to their opinion of the performance of the telephone transmission system used either for conversation or for listening to spoken material.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 model, model algorithm: An algorithm used for estimating the subjective (perceived) quality of a media sequence.

3.2.2 sequence: A short decoded audio, video or audiovisual portion of a stream, typically shorter than 30 seconds.

3.2.3 compression artefacts: Artefacts that are introduced due to lossy compression of the encoding process.

3.2.4 slicing artefacts: Artefacts that are introduced when packet losses are concealed through use of a packet loss concealment (PLC) scheme to repair erroneous frames.

3.2.5 freezing artefacts: Artefacts that are introduced when the packet loss concealment (PLC) scheme of the receiver replaces the erroneous frames (either due to packet loss or error propagation) with the previous error-free frame until a decoded picture without errors has been received. Since the erroneous frames are not displayed, this type of artefact is also referred to as freezing with skipping.

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
AMR-NB Adaptive Multi-Rate – Narrowband
AMR-WB Adaptive Multi-Rate – Wideband
ARQ Automatic Repeat Request
DASH Dynamic Adaptive Streaming over HTTP
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>FB</td>
<td>Fullband</td>
</tr>
<tr>
<td>FEC</td>
<td>Forward Error Correction</td>
</tr>
<tr>
<td>GOP</td>
<td>Group Of Pictures</td>
</tr>
<tr>
<td>HD</td>
<td>High Definition (television)</td>
</tr>
<tr>
<td>HE-AAC</td>
<td>High-Efficiency Advanced Audio Coding</td>
</tr>
<tr>
<td>HR</td>
<td>Higher Resolution</td>
</tr>
<tr>
<td>HRC</td>
<td>Hypothetical Reference Circuit</td>
</tr>
<tr>
<td>HTTP</td>
<td>Hypertext Transfer Protocol</td>
</tr>
<tr>
<td>HVGA</td>
<td>Half Video Graphics Array</td>
</tr>
<tr>
<td>LR</td>
<td>Lower Resolution</td>
</tr>
<tr>
<td>MBMS</td>
<td>Multimedia Broadcast/Multicast Service</td>
</tr>
<tr>
<td>MOS</td>
<td>Mean Opinion Score</td>
</tr>
<tr>
<td>MPEG</td>
<td>Motion Pictures Expert Group</td>
</tr>
<tr>
<td>NB</td>
<td>Narrowband</td>
</tr>
<tr>
<td>NTSC</td>
<td>National Television Standard Committee</td>
</tr>
<tr>
<td>PAL</td>
<td>Phase Alternating Line</td>
</tr>
<tr>
<td>PCC</td>
<td>Pearson Correlation Coefficient</td>
</tr>
<tr>
<td>PES</td>
<td>Packetized Elementary Stream</td>
</tr>
<tr>
<td>PLC</td>
<td>Packet Loss Concealment</td>
</tr>
<tr>
<td>PVS</td>
<td>Processed Video Sequence</td>
</tr>
<tr>
<td>QCIF</td>
<td>Quarter Common Intermediate Format</td>
</tr>
<tr>
<td>QoE</td>
<td>Quality of Experience</td>
</tr>
<tr>
<td>QVGA</td>
<td>Quarter Video Graphics Array</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root Mean Square Error</td>
</tr>
<tr>
<td>RTP</td>
<td>Real-time Transport Protocol</td>
</tr>
<tr>
<td>SD</td>
<td>Standard Definition</td>
</tr>
<tr>
<td>SRC</td>
<td>Source Reference Channel or Circuit</td>
</tr>
<tr>
<td>SWB</td>
<td>Superwideband</td>
</tr>
<tr>
<td>TS</td>
<td>Transport Stream</td>
</tr>
<tr>
<td>UDP</td>
<td>User Datagram Protocol</td>
</tr>
<tr>
<td>VSP</td>
<td>Visual Simple Profile</td>
</tr>
<tr>
<td>WB</td>
<td>Wideband</td>
</tr>
</tbody>
</table>

### 5 Conventions

None.
## Areas of application

Table 1 below shows the application range of the models based on what the models have actually been developed for.

### Table 1 – Factors and application ranges of the ITU-T G.1071 model algorithm

<table>
<thead>
<tr>
<th>Application information</th>
<th>Value range, unit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sequence duration (Ts)</strong></td>
<td><strong>ITU-T G.1071 – Lower resolution (LR)</strong></td>
</tr>
<tr>
<td>It is expected that the model will give reliable prediction results for sequence durations within the range 8-24 seconds</td>
<td>It is expected that the model will give reliable prediction results for sequence durations of approximately 8-16 seconds</td>
</tr>
<tr>
<td><strong>Packetization</strong></td>
<td><strong>ITU-T G.1071 – Lower resolution (LR)</strong></td>
</tr>
<tr>
<td>MPEG2-TS/RTP/UDP/IP</td>
<td>RTP/UDP/IP (Note 3)</td>
</tr>
<tr>
<td>MPEG2-TS/UDP/IP (Note 3)</td>
<td></td>
</tr>
<tr>
<td><strong>Video codec</strong></td>
<td><strong>ITU-T G.1071 – Lower resolution (LR)</strong></td>
</tr>
<tr>
<td>MPEG4 visual simple profile (VSP)</td>
<td>ITU-T H.264 main profile, ITU-T H.264 high profile</td>
</tr>
<tr>
<td>ITU-T H.264 baseline profile</td>
<td></td>
</tr>
<tr>
<td><strong>Video size</strong></td>
<td><strong>ITU-T G.1071 – Lower resolution (LR)</strong></td>
</tr>
<tr>
<td>SD: PAL, NTSC</td>
<td>HD: 720p, 1080p, 1080i</td>
</tr>
<tr>
<td>(High profile: 1080; main profile: 720, SD)</td>
<td></td>
</tr>
<tr>
<td><strong>Audio codec</strong></td>
<td><strong>ITU-T G.1071 – Lower resolution (LR)</strong></td>
</tr>
<tr>
<td>AMR-NB, AMR-WB+, AAC-LC, HE-AAC (v1, v2)</td>
<td>MPEG-4 AAC-LC</td>
</tr>
<tr>
<td>MPEG-4 HE-AAC (V1 and V2 = 3GPP enhanced AAC+)</td>
<td></td>
</tr>
<tr>
<td>MPEG-1 Layer 2</td>
<td></td>
</tr>
<tr>
<td>AC3</td>
<td></td>
</tr>
<tr>
<td><strong>Coded video bitrate</strong></td>
<td><strong>ITU-T G.1071 – Lower resolution (LR)</strong></td>
</tr>
<tr>
<td>QCIF: 32-1000 kbit/s</td>
<td>HD</td>
</tr>
<tr>
<td>QVGA: 80-3000 kbit/s</td>
<td></td>
</tr>
<tr>
<td>HDVGA: 192-6000 kbit/s</td>
<td></td>
</tr>
<tr>
<td>MPEG4</td>
<td></td>
</tr>
<tr>
<td>QCIF: 40-1500 kbit/s</td>
<td></td>
</tr>
<tr>
<td>QVGA: 90-3500 kbit/s</td>
<td></td>
</tr>
<tr>
<td>HDVGA: 192-6000 kbit/s</td>
<td></td>
</tr>
<tr>
<td><strong>Coded audio bitrate</strong></td>
<td><strong>ITU-T G.1071 – Lower resolution (LR)</strong></td>
</tr>
<tr>
<td>AMR-NB: 4.75-12.2 kbit/s</td>
<td>AAC-LC: 32-576 kbit/s</td>
</tr>
<tr>
<td>AMR-WB+: 10.4-48 kbit/s</td>
<td></td>
</tr>
<tr>
<td>AAC-LC: 16-128+ kbit/s</td>
<td></td>
</tr>
<tr>
<td>HE-AAC (v1, v2): 32-128 kbit/s</td>
<td></td>
</tr>
<tr>
<td>HE AACv2: 16-96 kbit/s</td>
<td></td>
</tr>
<tr>
<td>MPEG-1 Layer 2: 64-384 kbit/s</td>
<td></td>
</tr>
<tr>
<td>AC3:64-384 kbit/s</td>
<td></td>
</tr>
</tbody>
</table>
Table 1 – Factors and application ranges of the ITU-T G.1071 model algorithm

<table>
<thead>
<tr>
<th>Application information</th>
<th>ITU-T G.1071 – Lower resolution (LR)</th>
<th>ITU-T G.1071 – Higher resolution (HR)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Value range, unit</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Video decoder packet loss concealment</td>
<td>Two types of assumed decoder behaviour are covered: 1) freezing with skipping; 2) slicing with: MPEG4: 1 slices/frame [ITU-T H.264]: 1 slice/packet Both MPEG4 and [ITU-T H.264]: Fixed PLC (using fixed decoder, details and settings)</td>
<td>Types of decoder behaviour: two dimensions: slicing, PLC 1) freezing with skipping (duration(source)=duration(processed sequence)); 2) slicing with 1 slice per frame; 3) slicing with 1 slice per macroblock row, PLC with zero-motion copy (temporal from same region of previous good frame).</td>
</tr>
<tr>
<td>Audio decoder packet loss concealment</td>
<td>Decoder default modes: Codec-implementation specific loss concealment</td>
<td>Decoder default modes: Codec-implementation specific loss concealment</td>
</tr>
<tr>
<td>Retransmission mechanisms (ARQ); forward error correction (FEC); client jitter buffer behaviour</td>
<td>Developed models assume that the dejitter buffer, ARQ and FEC mechanisms have already corrected the stream.</td>
<td>(No rebuffering) Developed models assume that the dejitter buffer, ARQ and FEC mechanisms have already corrected the stream.</td>
</tr>
<tr>
<td>Encoder implementation</td>
<td>The model has been developed using the following encoders (Note 1): Video: • MPEG4 Part 2: ffmpeg • ITU-T H.264 (MPEG4 Part 10): x264 Audio: • AMR-NB/WB+: According to standard • AAC-LC, HE-AAC(v1, v2): Nero</td>
<td>The model has been developed using the following encoders (Note 1): Video: • ITU-T H.264 (MPEG4 Part 10): x264 Audio: • AAC-LC, HE-AAC v2: Nero • MPEG1-LII and AC3: ffmpeg</td>
</tr>
<tr>
<td>Decoder implementation</td>
<td>Reference decoder was a proprietary decoder provided by one proponent, which also performed de-packetization and audio-video-demultiplexing. The ITU-T H.264-decoding is standard-conformant, with the PLC as described above (Note 2).</td>
<td>Reference decoder was a proprietary decoder provided by one proponent, which also performed de-packetization and audio-video-demultiplexing. The ITU-T H.264-decoding is standard-conformant, with the PLC as described above (Note 2).</td>
</tr>
<tr>
<td>Group of pictures (GOP)</td>
<td>Typical GOP structure for which the model has been trained: M = 1, N = 40 (typically no B frames for mobile case) Length: fixed, variable, adaptive Structure (e.g., IPPP...PPPI)</td>
<td>Supporting default modes for typical GOP structures E.g., M = 3, N = 15 Length: fixed, variable, adaptive Structure (e.g., IBBPBB...PBBI) NOTE – GOP structure is explicitly estimated from stream.</td>
</tr>
</tbody>
</table>
Table 1 – Factors and application ranges of the ITU-T G.1071 model algorithm

<table>
<thead>
<tr>
<th>Application information</th>
<th>ITU-T G.1071 – Lower resolution (LR)</th>
<th>ITU-T G.1071 – Higher resolution (HR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value range, unit</td>
<td></td>
</tr>
<tr>
<td>Frame rate</td>
<td>5, 8.33, 12.5, 15, 20, 25, 30 fps</td>
<td>SD: 50i (PAL), 59.94i (NTSC)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HD: 50p, 59.94p, 60p, 50i, 59.94i, 60i, 25p, 29.97p, 30p</td>
</tr>
<tr>
<td>Audio channel number</td>
<td>1 (diotic mono), 2 (stereo)</td>
<td>2 (stereo)</td>
</tr>
<tr>
<td># of Audio frames per RTP packet</td>
<td>1 to 5 audio frames</td>
<td>Bitrate-specific (depending on both the audio and video bitrate)</td>
</tr>
<tr>
<td>Audio-video multiplexed?</td>
<td>Default: No, at RTP-level; no audio-video asynchrony</td>
<td>In MPEG-2 TS/RTP/UDP and MPEG-2 TS/UDP: Supported</td>
</tr>
<tr>
<td>Packet loss degradation, video</td>
<td>Uniform loss:0-10% Burst loss:0-10% (4-state Markov model)</td>
<td>Uniform loss:0-2% Burst loss:0-2% (4-state Markov model)</td>
</tr>
<tr>
<td>Packet loss degradation, audio</td>
<td>Uniform loss:0-10% Burst loss:0-10% (4-state Markov model)</td>
<td>Uniform loss:0-6% Burst loss:0-6% (4-state Markov model)</td>
</tr>
<tr>
<td>Symmetrical versus asymmetrical handling of audio and video in audiovisual case</td>
<td>Model application: Symmetrical, but can handle asymmetric cases due to specific model development process</td>
<td>Model application: Symmetrical, but can handle asymmetric cases due to specific model development process</td>
</tr>
</tbody>
</table>

NOTE 1 – It is assumed that the model can be used for estimating quality when other encoder implementations for the given codec have been used. However, if the encoder performance is significantly worse or better than for the encoder used, the model prediction accuracy will be reduced.

NOTE 2 – One aspect not covered by decoder packet loss concealment is post-filtering. Guidance on how to adjust internal model parameters for specific other decoders, including set-top boxes, is for further study.

NOTE 3 – The ITU-T P.1201.2 model has been trained on MPEG-2 TS/RTP/UDP. However, due to the design of the ITU-T P.1201.2 algorithm, it is also applicable to MPEG2-TS/UDP/IP. Further, to the model's design, it is assumed to also work for RTP/UDP/IP transport with similar, but so far unverified, accuracy as compared to MPEG2-TS/RTP/UDP/IP.

Factors and applications not covered by the model are:
- Evaluation of audiovisual quality including display/device properties
- Audiovisual streaming with significant rate adaptation (such as that used in dynamic adaptive streaming over HTTP (DASH) streaming)
- Transcoding situations
- The effects of audio level, noise and delay (and corresponding similar video factors)
- Re-buffering degradation of audio, video and audiovisual
- Coding technologies the models are not intended for: [ITU-T H.261], MPEG-2, [ITU-T H.263], [ITU-T H.265], etc.
7 Model framework

As shown in Figure 1, the ITU-T G.1071 models follow the same structure as the ITU-T P.1201 models. Indeed, they are composed of three modules: the audio module, the video module and the audiovisual module. However, in contrast to the ITU-T P.1201 models, the ITU-T G.1071 models take as inputs network planning assumptions instead of the encrypted bitstream. Some of the ITU-T P.1201 input parameters are available as planning assumptions as well. For the ITU-T P.1201 input parameters which are not available as planning assumptions, the ITU-T G.1071 models provide a set of rules to convert planning assumptions into these ITU-T P.1201 input parameters (see block "conversion rules").

![Figure 1 – Model framework](image)

8 Model output information and performance details

ITU-T G.1071 has three output parameters:

1) Estimated audiovisual MOS on the 1 to 5 scale, which is an estimation of the perceived audiovisual quality.

2) Estimated video MOS on the 1 to 5 scale, which is an estimation of the perceived video quality (without audio present). Note that the model is able to give both a video score for a stream without audio and a stream including audio.

3) Estimated audio MOS on the 1 to 5 scale, which is an estimation of the perceived audio quality (without video present). Note that the model is able to give both an audio score for a stream without video and a stream including video.

The performance information for the ITU-T G.1071 models can be found in Table 2 and Table 3 for the HR and LR application areas respectively. The statistical metrics root mean square error (RMSE) and the Pearson correlation are used to describe the performance, see [ITU-T P.1401]. (Note that for these performance figures, the subjective ratings have been mapped to the model scores using a linear, i.e., 1st-order, mapping function at a per-database-level. This has been done in order to avoid misalignment due to bias in the different subjective tests, e.g., as a result of different test settings).

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
<th>Pearson correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audiovisual</td>
<td>0.51 (based on 1595 samples)</td>
<td>0.87 (based on 1595 samples)</td>
</tr>
<tr>
<td>Video</td>
<td>0.53 (based on 3069 samples)</td>
<td>0.86 (based on 3069 samples)</td>
</tr>
<tr>
<td>Audio</td>
<td>0.37 (based on 680 samples)</td>
<td>0.93 (based on 680 samples)</td>
</tr>
</tbody>
</table>

Table 2 – Performance information for ITU-T G.1071 (HR)

Samples were taken from the ITU-T P.1201.2 training and validation databases
Table 3 – Performance information for ITU-T G.1071 (LR)
Samples were taken from the ITU-T P.1201.1 training and validation databases

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
<th>Pearson correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audiovisual</td>
<td>0.50 (based on 1166 samples)</td>
<td>0.83 (based on 1166 samples)</td>
</tr>
<tr>
<td>Video</td>
<td>0.60 (based on 1430 samples)</td>
<td>0.78 (based on 1430 samples)</td>
</tr>
<tr>
<td>Audio</td>
<td>0.38 (based on 690 samples)</td>
<td>0.93 (based on 690 samples)</td>
</tr>
</tbody>
</table>
Annex A

Description of the ITU-T G.1071 model algorithm for HR application area

(This annex forms an integral part of this Recommendation.)

A.1 Audio module

The HR application area audio module is depicted in Figure A.1.

![Audio module diagram]

Figure A.1 – Audio module

\[ QA = 100 - QcodA - QtraA \]  
(1.1)

\[ MOSA = MOSfromR(QA) \]  
(1.2)

Where:

\[ QcodA = a1A \cdot e^{a2A \cdot \text{Bitrate}^A} + a3A \]  
(1.3)

And

\[ QtraA = (b1A - QcodA) \cdot \frac{\text{FrameLoss}^A}{\text{FrameLoss}^A + b2A \cdot \text{Burstiness}^A + b3A} \]  
(1.4)

from [ITU-T P.1201.2].

The following conversion rules are applied:

\[ \text{FrameLoss}^A = c1A \cdot \text{Bitrate}^A \cdot T\text{PacketLoss}^A + c2A \cdot T\text{PacketLoss}^A \]  
(1.4a)

\[ \text{Burstiness}^A = d1A \cdot T\text{Burstiness}^A + d2A \cdot \text{Bitrate}^A \cdot T\text{Burstiness}^A + d3A \]  
(1.4b)

where:

- \( QA \) is the overall estimated audio quality, expressed on a 100-point scale [0,100], where 0 is the worst quality and 100 the best quality
- \( MOSA \) is the overall estimated audio quality, expressed on the MOS scale [1,5], where 1 is the worst quality and 5 the best quality
- \( QcodA \) is the estimated audio quality for audio compression artefacts
- \( QtraA \) is the estimated audio quality for audio transmission errors
BitrateA is the audio bitrate (in kbit/s)
FramelossA is the audio frame loss percentage
TSp PacketLossA is the percentage of audio-TS-packet loss
BurstinessA is the audio frame loss "burstiness" (average number of consecutively lost audio frames),
T SburstinessA is the audio-TS-packet "loss burstiness" (average number of consecutively lost audio-TS packets)

The coefficient values $a_{1A}, a_{2A}, a_{3A}, b_{1A}, b_{2A}, b_{3A}, c_{1A}, c_{2A}, d_{1A}, d_{2A}$ and $d_{3A}$ depend on the audio codec. Coefficient values are provided in Tables A.1 and A.2.

MOSfromR transform quality scores expressed on the 100-point scale to the MOS scale; MOSfromR is expressed as follows (from [ITU-T P.1201.2]):

\[
\text{function } \text{MOS} = \text{MOSfromR}(Q) \\
\text{set } \text{MOS\_MAX} = 4.9; \\
\text{set } \text{MOS\_MIN} = 1.05; \\
\text{if } (Q > 0 \& Q < 100), \\
\text{MOS} = (\text{MOS\_MIN} + (\text{MOS\_MAX} - \text{MOS\_MIN})/100 \times Q + Q \times (Q-60) \times (100-Q) \times 7.0E-6); \\
\text{elseif } (Q >= 100), \\
\text{MOS} = \text{MOS\_MAX}; \\
\text{else} \\
\text{MOS} = \text{MOS\_MIN}; \\
\text{end}
\]

TSp PacketLossA and T SburstinessA are computed from RTP-layer parameters as follows:

If

\[
\frac{\text{BitrateA}}{\text{BitrateV} \times 10^3} = \frac{(7-\text{NTSV})}{\text{NTSV}}
\]

with:

\[
\text{NTSV} = [1.6] \\
\text{NTSV integer}
\]

then

\[
\text{TSp PacketLossA} = \text{RTP\_packetLoss} \tag{1.4c}
\]

\[
\text{T SburstinessA} = 7 \cdot \frac{\text{BitrateA}}{\text{BitrateA} + \text{BitrateV} \times 10^3} \cdot \text{RTP\_burstiness} \tag{1.4d}
\]

If

\[
\frac{\text{BitrateA}}{\text{BitrateV} \times 10^3} = \frac{1}{6+7 \cdot D}
\]

then

\[
\text{TSp PacketLossA} = \text{RTP\_packetLoss} \tag{1.4e}
\]

\[
\text{T SburstinessA} = \frac{7 \cdot \text{BitrateA}}{\text{BitrateV} \times 10^3 + \text{BitrateA}} \cdot \text{burstLengthA} \cdot \text{RTP\_burstiness} \tag{1.4f}
\]

If each RTP packet contain audio TS packets only, then

\[
\text{TSp PacketLossA} = \text{RTP\_packetLoss} \tag{1.4g}
\]

\[
\text{T SburstinessA} = 7 \cdot \text{RTP\_burstiness} \tag{1.4h}
\]
where:

- $Bitrate_A$ is the audio bitrate (in kbit/s)
- $Bitrate_V$ is the video bitrate (in Mbit/s)
- $TS\text{packetLoss}_A$ is the percentage of audio-TS-packet loss
- $RTP\text{packetLoss}$ is the percentage of RTP packet loss
- $TS\text{burstiness}_A$ is the audio-TS-packet "loss burstiness" (average number of consecutively lost audio-TS packets),
- $RTP\text{burstiness}$ is the RTP packet loss "burstiness" (average number of consecutively lost RTP packets)
- $NTSV$ is the average number of video TS packets into one RTP packet
- $D$ is the number of RTP packets which contain video TS packets only between two RTP packets containing a single audio TS packet
- $burstLength_A$ is the average number of audio TS packets in one RTP packet for RTP packets containing audio TS packets.

### A.1.1 Coefficient values and parameter ranges

**Table A.1 – Audio model coefficients for the different audio codecs**
(from [ITU-T P.1201.2])

<table>
<thead>
<tr>
<th>Audio codec</th>
<th>$a_1A$</th>
<th>$a_2A$</th>
<th>$a_3A$</th>
<th>$b_1A$</th>
<th>$b_2A$</th>
<th>$b_3A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mp2</td>
<td>100.0</td>
<td>−0.02</td>
<td>15.48</td>
<td>100.0</td>
<td>1.51</td>
<td>1.64</td>
</tr>
<tr>
<td>AC3</td>
<td>100.0</td>
<td>−0.03</td>
<td>15.70</td>
<td>100.0</td>
<td>0.2</td>
<td>2.40</td>
</tr>
<tr>
<td>AacLC</td>
<td>100.0</td>
<td>−0.05</td>
<td>14.60</td>
<td>101.32</td>
<td>0.1</td>
<td>4.09</td>
</tr>
<tr>
<td>HeAac</td>
<td>100.0</td>
<td>−0.11</td>
<td>20.06</td>
<td>105.68</td>
<td>0.1</td>
<td>5.92</td>
</tr>
</tbody>
</table>

**Table A.2 – Audio model coefficients to convert network planning parameters into ITU-T P.1201.2 parameters**

<table>
<thead>
<tr>
<th>Audio codec</th>
<th>$c_1A$</th>
<th>$c_2A$</th>
<th>$d_1A$</th>
<th>$d_2A$</th>
<th>$d_3A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mp2</td>
<td>0.006</td>
<td>1.124</td>
<td>0.682</td>
<td>−0.001</td>
<td>0.908</td>
</tr>
<tr>
<td>AC3</td>
<td>0.016</td>
<td>0.973</td>
<td>0.277</td>
<td>−0.003</td>
<td>0.974</td>
</tr>
<tr>
<td>AacLC</td>
<td>0.005</td>
<td>0.976</td>
<td>0.486</td>
<td>−0.001</td>
<td>0.923</td>
</tr>
<tr>
<td>HeAac</td>
<td>0.026</td>
<td>0.482</td>
<td>−0.627</td>
<td>0.012</td>
<td>0.984</td>
</tr>
</tbody>
</table>

### A.2 Video module

The HR application area video module is depicted in Figure A.2.
From [ITU-T P.1201.2]

\[ QV = 100 - Q_{\text{codV}} - Q_{\text{traV}} \]  \hspace{1cm} (2.1)

\[ MOSV = MOS_{\text{fromR}}(QV) \]  \hspace{1cm} (2.2)

From [ITU-T P.1201.2]

\[ Q_{\text{codV}} = a1V \cdot e^{a2V \cdot \text{BitPerPixel}} + a3V \cdot \text{ContentComplexity} + a4V \]  \hspace{1cm} (2.3)

From [ITU-T P.1201.2]

\[ \text{BitPerPixel} = \frac{\text{BitrateV} \cdot 10^6}{\text{NumPixelPerFrame} \cdot \text{FrameRate}} \]  \hspace{1cm} (2.3a)

In the case of freezing:

\[ Q_{\text{traV}} = b1V \cdot \log(b2V \cdot \text{FreezingRatio}_{E} + 1) \]  \hspace{1cm} (2.4)

From [ITU-T P.1201.2]

\[ (\text{where } \text{FreezingRatio}_{E} = \text{FreezingRatio} \cdot \text{BitPerPixel} \text{ in [ITU-T P.1201.2]} \) \]

In the case of slicing:

\[ Q_{\text{traV}} = c1V \cdot \log(c2V \cdot \text{LossMagnitude}_{E} + 1) \]  \hspace{1cm} (2.5)

From [ITU-T P.1201.2]

\[ (\text{where } \text{LossMagnitude}_{E} = \frac{\text{LossMagnitude}}{Q_{\text{codVn}}}, \text{ with} \) \]

\[ Q_{\text{codVn}} = \begin{cases} 1 \quad , & Q_{\text{codV}} < 20 \\ 0.1125 \cdot Q_{\text{codV}} - 1.25, & Q_{\text{codV}} \geq 20 \end{cases} \text{ in [ITU-T P.1201.2]} \]

\[ QV \] is the overall estimated video quality, expressed on a 100-point scale [0,100], where 0 is the worst quality and 100 the best quality

\[ MOSV \] is the overall estimated video quality, expressed on the MOS scale [1,5], where 1 is the worst quality and 5 the best quality

\[ Q_{\text{codV}} \] is the estimated video quality for video compression artefacts

\[ Q_{\text{traV}} \] is the estimated video quality for video transmission errors

\[ \text{BitPerPixel} \] is the average number of bits per pixel (see clause A.2.1)
Bitrate\textsubscript{V} is the video bitrate in Mbps

NumPixelPerFrame is the number of pixels per video frame

FrameRate is the video frame rate

ContentComplexity captures the spatio-temporal complexity of the video sequence (see clause A.2.1)

FreezingRatio\textsubscript{E} captures the degradation when freezing is applied as packet loss concealment (see clause A.2.2)

LossMagnitude\textsubscript{E} captures the degradation when slicing is applied as packet loss concealment (see clause A.2.3)

The coefficient values \( a_{1V}, a_{2V}, a_{3V}, a_{4V}, b_{1V}, b_{2V}, c_{1V}, \) and \( c_{2V} \) depend on the video resolution. Coefficient values are provided in Table A.3. MOS\textsubscript{fromR} has been defined in clause A.1.

A.2.1 Video compressions

The ContentComplexity parameter of equation 2.3 is estimated from network planning parameters as follows:

\[
\text{ContentComplexity} = a_{31} \cdot \exp(a_{32} \cdot \text{BitPerPixel}) + a_{33}
\]  
(2.3b)

where BitPerPixel is the already defined average number of bits per pixel.

If BitPerPixel > 0.1 OR medium content complexity is assumed for all contents:

\[
\text{ContentComplexity} = a_{31} \cdot \exp(a_{32} \cdot \text{BitPerPixel}) + a_{33}
\]

Else if BitPerPixel \( \leq \) 0.1

If the network planner assumes that contents have high spatio-temporal (ST) complexity:

\[
\text{ContentComplexity} > a_{31} \cdot \exp(a_{32} \cdot \text{BitPerPixel}) + a_{33}
\]  
(2.3c)

If the network planner assumes that contents have low to medium ST complexity:

\[
\text{ContentComplexity} \leq a_{31} \cdot \exp(a_{32} \cdot \text{BitPerPixel}) + a_{33}
\]  
(2.3d)

The coefficient values \( a_{31}, a_{32}, \) and \( a_{33} \) depend on the video resolution. They are provided in Table A.4.

A.2.2 Video freezing

The FreezingRatio\textsubscript{E} parameter of equation 2.4 is estimated from network planning parameters as follows:

\[
\text{FreezingRatio\textsubscript{E}} = p_1 \cdot \exp(p_2 \cdot \text{FreezinRarioNP})
\]  
(2.4a)

where

\[
\text{FreezingRatioNP} = (b_{21} - I\text{codn}) \cdot \frac{T\text{PacketLossV}}{I\text{codn}(b_{22}T\text{burstinessV} + b_{23}) + T\text{PacketLossV}}
\]  
(2.4b)

and

\[
I\text{codn} = \begin{cases} 
Q\text{codV}, & Q\text{codV} \leq 65 \\
65, & Q\text{codV} > 65 
\end{cases}
\]  
(2.4c)

\( T\text{PacketLossV} \) is the percentage of lost TS video packets in the measurement window, and \( T\text{burstinessV} \) is the average number of consecutively lost video TS packets in the measurement window.

\( Q\text{codV} \) is expressed in equation 2.3

The coefficient values \( p_1, p_2, b_{21}, b_{22}, \) and \( b_{23} \) are the same for both SD and HD. Detailed values are provided in Table A.5.
\( T\text{PacketLoss}_V \) and \( T\text{Sburstiness}_V \) are computed from RTP-layer parameters as follows:

If

\[
\frac{\text{Bitrate}_A}{\text{Bitrate}_V \times 10^3} = \frac{(7-\text{NTSV})}{\text{NTSV}}
\]

with \( \text{NTSV} = [1,6] \), \( \text{NTSV} \) integer

then

\[
T\text{PacketLoss}_V = \text{RTPpacketLoss}
\]

\[
T\text{Sburstiness}_V = \frac{7 \times \text{Bitrate}_V \times 10^3}{\text{Bitrate}_A + \text{Bitrate}_V \times 10^3} \cdot \text{RTPburstiness}
\]

If

\[
\frac{\text{Bitrate}_A}{\text{Bitrate}_V \times 10^3} = \frac{1}{6+7 \times D}
\]

then

\[
T\text{PacketLoss}_V = \text{RTPpacketLoss}
\]

\[
T\text{Sburstiness}_V = \text{RTPburstiness} \cdot \left(7 - \frac{7 \times \text{burstLength}_A \times \text{Bitrate}_A}{\text{Bitrate}_V \times 10^3 + \text{Bitrate}_A} \right)
\]

If each RTP packet contains video TS packets only, then

\[
T\text{PacketLoss}_V = \text{RTPpacketLoss}
\]

\[
T\text{Sburstiness}_V = 7 \cdot \text{RTPburstiness}
\]

where

\( \text{Bitrate}_A \) is the audio bitrate (in kbit/s)

\( \text{Bitrate}_V \) is the video bitrate (in Mbit/s)

\( T\text{PacketLoss}_V \) is the percentage of lost TS video packets in the measurement window, and

\( T\text{Sburstiness}_V \) is the average number of consecutively lost video TS packets in the measurement window

\( \text{RTPpacketLoss} \) is the percentage of RTP packet loss

\( \text{RTPburstiness} \) is the RTP packet loss "burstiness" (average number of consecutively lost RTP packets)

\( \text{NTSV} \) is the average number of video TS packets into one RTP packet

\( D \) is the number of RTP packets which contain video TS packets only between two RTP packets containing a single audio TS packet

\( \text{burstLength}_A \) is the average number of audio TS packets in one RTP packet for RTP packets containing audio TS packets.

A.2.3 Video slicing

The \( \text{LossMagnitude}_E \) parameter of equation 2.5 is estimated from network planning parameters as follows:

\[
\text{LossMagnitude}_E = q_1 \cdot \exp(q_2 \cdot \text{LossMagnitude}_{NP}) - q_1
\]

where

\[
\text{LossMagnitude}_{NP} = (c_{21} - I\text{codn}) \cdot \frac{T\text{PacketLoss}_V}{I\text{codn} \cdot (c_{22} \times T\text{Sburstiness}_V + c_{23}) + T\text{PacketLoss}_V}
\]
\[ I_{\text{cdn}} = \begin{cases} Q_{\text{cdn}}, & Q_{\text{cdn}} \leq 65 \\ 65, & Q_{\text{cdn}} > 65 \end{cases} \]  
(2.5c)

\( T_{\text{PacketLoss}} \) is the percentage of lost TS video packets in the measurement window, and 
\( T_{\text{Burstiness}} \) is the average number of consecutively lost TS video packets in the measurement window, 
\( Q_{\text{cdn}} \) is expressed in equation 2.3

The coefficient values \( q_1, q_2, c_{21}, c_{22}, \) and \( c_{23} \) are the same for SD and HD, but they depend on number of slices per frame. Detailed values are provided in Table A.6.

To compute \( T_{\text{PacketLoss}} \) and \( T_{\text{Burstiness}} \) from RTP-layer parameters, the same conversion rules as for Video Freezing (equations 2.4d to 2.4i) are used.

### A.2.4 Coefficient values and parameter ranges

#### Table A.3 – Video model coefficients for the different video resolutions  
(from ITU-T P.1201.2)

<table>
<thead>
<tr>
<th>Video resolution</th>
<th>( a_1 )</th>
<th>( a_2 )</th>
<th>( a_3 )</th>
<th>( a_4 )</th>
<th>( b_{1V} )</th>
<th>( b_{2V} )</th>
<th>( c_{1V} )</th>
<th>( c_{2V} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD</td>
<td>61.28</td>
<td>-11.00</td>
<td>6.00</td>
<td>6.21</td>
<td>12.70</td>
<td>907.36</td>
<td>17.73</td>
<td>123.08</td>
</tr>
<tr>
<td>HD</td>
<td>51.28</td>
<td>-22.00</td>
<td>6.00</td>
<td>6.21</td>
<td>12.70</td>
<td>907.36</td>
<td>17.73</td>
<td>123.08</td>
</tr>
</tbody>
</table>

#### Table A.4 – Coefficient values for the ContentComplexity parameter

<table>
<thead>
<tr>
<th></th>
<th>SD (PAL, NTSC)</th>
<th>HD (HD720, HD1080)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_{31} )</td>
<td>0.91</td>
<td>3.92</td>
</tr>
<tr>
<td>( a_{32} )</td>
<td>-9.39</td>
<td>-27.54</td>
</tr>
<tr>
<td>( a_{33} )</td>
<td>0.10</td>
<td>0.26</td>
</tr>
</tbody>
</table>

#### Table A.5 – Coefficient values for the FreezingRatioE parameter

<table>
<thead>
<tr>
<th></th>
<th>SD (PAL, NTSC), HD (HD720, HD1080)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_1 )</td>
<td>0.0001661</td>
</tr>
<tr>
<td>( p_2 )</td>
<td>0.1166</td>
</tr>
<tr>
<td>( b_{21} )</td>
<td>69.39</td>
</tr>
<tr>
<td>( b_{22} )</td>
<td>0.00019</td>
</tr>
<tr>
<td>( b_{23} )</td>
<td>0.00082</td>
</tr>
</tbody>
</table>

#### Table A.6 – Coefficient values for the LossMagnitudeE parameter

<table>
<thead>
<tr>
<th></th>
<th>1 slice / frame</th>
<th>&gt; 1 slice/frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>( q_1 )</td>
<td>0.018</td>
<td>0.018</td>
</tr>
<tr>
<td>( q_2 )</td>
<td>0.040</td>
<td>0.040</td>
</tr>
<tr>
<td>( c_{21} )</td>
<td>80.61</td>
<td>67.15</td>
</tr>
<tr>
<td>( c_{22} )</td>
<td>0.00046</td>
<td>0.00144</td>
</tr>
<tr>
<td>( c_{23} )</td>
<td>0.00147</td>
<td>0.00147</td>
</tr>
</tbody>
</table>
A.3 Audiovisual module

As shown in Figure A.3, the audiovisual module of ITU-T G.1071 is identical to the audiovisual module in [ITU-T P.1201.2].

\[
QAV = 0.7 \times QQAV + 0.3 \times QQFAV
\]

(3.1)

from [ITU-T P.1201.2]

\[
MOSAV = MOSfromR(QAV)
\]

(3.2)

from [ITU-T P.1201.2]

with

\[
QQAV = \alpha + \beta \cdot QV + \gamma \cdot QA \cdot QV
\]

(3.3)

from [ITU-T P.1201.2]

\[
QQFAV = a - b \cdot QcodA - c \cdot QcodV - d \cdot QtraA - e \cdot QtraV - f \cdot QtraA \cdot QtraV - g \cdot QcodV \cdot QtraA - h \cdot QcodA \cdot QtraV
\]

(3.4)

from [ITU-T P.1201.2]

where

\(QAV, QA, \) and \(QV\) are the overall estimated audiovisual, audio and video quality, expressed on an 100-point scale \([0,100]\), where 0 is the worst quality and 100 the best quality

\(MOSAV, MOSA, \) and \(MOSV\) are the overall estimated audiovisual, audio and video quality, expressed on the MOS scale \([1,5]\), where 1 is the worst quality and 5 the best quality

\(QcodA\) is the estimated audio quality for audio compression artefacts (see clause I.1)

\(QtraA\) is the estimated audio quality for audio transmission errors (see clause I.1)

\(QcodV\) is the estimated video quality for audio compression artefacts (see clause I.2)

\(QtraV\) is the estimated video quality for audio transmission errors (see clause I.2)

\(MOSformR\) has been defined in clause I.1. Coefficient values \(\alpha, \beta, \gamma, a, b, c, d, e, f, g, h\) are provided in Table A.7.

Table A.7 – Audiovisual model coefficients (from [ITU-T P.1201.2])

<table>
<thead>
<tr>
<th>(\alpha)</th>
<th>(\beta)</th>
<th>(\gamma)</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
<th>(e)</th>
<th>(f)</th>
<th>(g)</th>
<th>(h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.89</td>
<td>0.52</td>
<td>0.0045</td>
<td>100.0</td>
<td>0.32</td>
<td>0.9</td>
<td>0.705</td>
<td>1.02</td>
<td>−0.007</td>
<td>−0.010</td>
<td>−0.008</td>
</tr>
</tbody>
</table>
Annex B

Description of the ITU-T G.1071 model algorithm for LR application area
(This annex forms an integral part of this Recommendation.)

B.1 Audio module

The LR application area audio module is depicted in Figure B.1.

\[
A_{MOS} = 1 + (A_{MOSC} - 1) \times MA
\]  

(4.1)

from [ITU-T P.1201.1]

where

\[
A_{MOSC} = 1 + \left( a_1 - \frac{a_1}{1 + \left( \frac{a_3}{a_2} \right)^{a_3}} \right)
\]  

(4.2)

from [ITU-T P.1201.1]

and

\[
MA = (1 - a_4) \exp\left(-\frac{10 \times A_{LFL}}{a_5 \times A_{MT}}\right) + a_4 \exp\left(-\frac{10 \times A_{LFL}}{a_6 \times A_{MT}}\right)
\]  

(4.3)

from [ITU-T P.1201.1]

\[
A_{LFL} = A_{PLEF} \times \text{MAX} \left( \text{AudioFrameLength}, A_{LFLpP} \times \frac{A_{ABPLL} + A_{NPpTS} - 1}{A_{NPpTS}} \right)
\]  

(4.4)

from [ITU-T P.1201.1]

The following conversion rules are applied:
\[ A_{\text{ABPLL}} = \text{IPPacketAverageBurstLength} \]  
\[ A_{\text{LFLpP}} = \text{AudioFrameLength} \times \text{NumberofAudioFramesperPacket} \]  
\[ A_{\text{NPpTS}} = \left[ \frac{A_{\text{BR}} \times \text{AudioFrameLength}}{8 \times \text{MaximumofDataSizeperPacket}} \right] \]  
\[ A_{\text{PLEF}} = \frac{1000 \times A_{\text{NPpTS}} \times \text{IPPacketLossRate}}{A_{\text{LFLpP}} \times A_{\text{ABPLL}}} \]  

The value of the following parameter is fixed
\[ A_{\text{MT}} = 1 \]

where
- \( A_{\text{MOS}} \) is the overall estimated audio quality, expressed on an MOS scale [1,5], where 1 is the worst quality and 5 is the best quality
- \( A_{\text{MOSC}} \) is the estimated audio quality for audio compression artefacts
- \( \text{AudioBitRate}(A_{\text{BR}}) \) is the audio bitrate (in kbit/s)
- \( \text{AudioCodec} \) is the audio codec
- \( M_A \) is the degree of degradation for audio transmission errors on a scale [0,1], where 0 is no degradation for audio transmission errors and 1 is the maximum degradation for audio transmission errors
- \( A_{\text{MT}} \) is the measurement time for audio (in seconds); this value is fixed to 1
- \( A_{\text{LFL}} \) is the lost audio frame length per \( A_{\text{MT}} \)
- \( A_{\text{LFLpP}} \) is the audio frame length per 1 audio RTP packet
- \( A_{\text{ABPLL}} \) is the average burst IP packet-loss length for an audio stream
- \( A_{\text{NPpTS}} \) is the number of audio RTP packets per 1 audio frame
- \( A_{\text{PLEF}} \) is the number of audio packet-loss-events per \( A_{\text{MT}} \)
- \( \text{AudioFrameLength} \) is the audio frame length (in milliseconds)
- \( \text{NumberofAudioFramesperPacket} \) is the number of audio frames in 1 RTP packet (to be provided by a network/service planner)
- \( \text{IPPacketLossRate} \) is the loss rate of IP packets
- \( \text{IPPacketAverageBurstLength} \) is the average burst IP packet-loss length; this value is fixed to 1 if \( \text{IPPacketLossRate} \) is 0
- \( \text{MaximumofDataSizeperPacket} \) is the maximum size for audio stream per 1 RTP packet (Bytes)

The coefficient values \( a_1, a_2, a_3, a_4, a_5, \) and \( a_6 \) depend on the audio codec. Coefficient values are provided in Table B.1.

### B.1.1 Coefficient values and parameter ranges

<table>
<thead>
<tr>
<th>Audio codec</th>
<th>( a_1 )</th>
<th>( a_2 )</th>
<th>( a_3 )</th>
<th>( a_4 )</th>
<th>( a_5 )</th>
<th>( a_6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAC-LC</td>
<td>3.36209</td>
<td>16.46062</td>
<td>2.08184</td>
<td>0.352</td>
<td>508.83419</td>
<td>37.78354</td>
</tr>
<tr>
<td>AAC-HEv1</td>
<td>3.19135</td>
<td>4.17393</td>
<td>1.28241</td>
<td>0.68955</td>
<td>6795.99773</td>
<td>186.76692</td>
</tr>
<tr>
<td>AAC-HEv2</td>
<td>3.13637</td>
<td>7.45884</td>
<td>2.15819</td>
<td>0.61993</td>
<td>3918.639</td>
<td>153.3399</td>
</tr>
<tr>
<td>AMR-NB</td>
<td>1.33483</td>
<td>6.42499</td>
<td>3.49066</td>
<td>0</td>
<td>723.3661</td>
<td>1</td>
</tr>
<tr>
<td>AMR-WB+</td>
<td>3.19158</td>
<td>5.7193</td>
<td>1.63208</td>
<td>0</td>
<td>826.7936</td>
<td>1</td>
</tr>
</tbody>
</table>

Table B.1 – Audio model coefficients for the different audio codecs (from [ITU-T P.1201.1])
B.2 Video module

The LR application area video module is depicted in Figure B.2.

![Figure B.2 – Video module]

The final video quality is estimated as follows:

IF No Packet-loss AND No rebuffering THEN

\[ V_{MOS} = V_{MOSC} \] (5.1)

from [ITU-T P.1201.1]

ELSEIF Packet-loss AND No rebuffering

\[ V_{MOS} = V_{MOSP} \] (5.2)

from [ITU-T P.1201.1]

ELSE

\[ V_{MOS} = V_{MOSR} \] (5.3)

from [ITU-T P.1201.1]

ENDIF

\( V_{MOSC} \) is calculated as follows:

IF videoFrameRate >= 24 THEN

\[ V_{MOSC} = MOS_{MAX} - V_{DC} \] (5.4)

from [ITU-T P.1201.1]

ELSE
\[
V_{MOSC} = (MOS_{MAX} - V_{DC}) \cdot \left(1 + v1 \cdot V_{CCF} - v2 \cdot V_{CCF} \cdot \log \left( \frac{1000}{videoFrameRate} \right) \right)
\]
(5.5)

from [ITU-T P.1201.1]

ENDIF

\[V_{DC} \text{ is calculated as follows:} \]
\[
V_{DC} = \frac{MOS_{MAX} - MOS_{MIN}}{1 + \left( \frac{V_{NBR}}{v3 \cdot V_{CCF} + v4} \right)^{(v5 \cdot V_{CCF} + v6)}}
\]
(5.6)

from [ITU-T P.1201.1]

\[V_{MOSP} \text{ is calculated as follows:} \]
\[
V_{MOSP} = V_{MOSC} - V_{DP}
\]
(5.7)

from [ITU-T P.1201.1]

\[IF \ videoPLC==SLICING\]

\[
V_{DP} = (V_{MOSC} - MOS_{MIN}) \cdot \frac{\left( \frac{V_{AIRF} \cdot V_{IR}}{v7 \cdot V_{CCF} + v8} \right)^{v9} \cdot \left( \frac{V_{PLEF}}{v10 \cdot V_{CCF} + v11} \right)^{v12}}{1 + \left( \frac{V_{AIRF} \cdot V_{IR}}{v7 \cdot V_{CCF} + v8} \right)^{v9} \cdot \left( \frac{V_{PLEF}}{v10 \cdot V_{CCF} + v11} \right)^{v12}}
\]
(5.8)

from [ITU-T P.1201.1]

\[IF \ videoPLC==FREEZING\]

\[
V_{DP} = (V_{MOSC} - MOS_{MIN}) \cdot \frac{\left( \frac{V_{IR}}{v7 \cdot V_{CCF} + v8} \right)^{v9} \cdot \left( \frac{V_{PLEF}}{v10 \cdot V_{CCF} + v11} \right)^{v12}}{1 + \left( \frac{V_{IR}}{v7 \cdot V_{CCF} + v8} \right)^{v9} \cdot \left( \frac{V_{PLEF}}{v10 \cdot V_{CCF} + v11} \right)^{v12}}
\]
(5.9)

from [ITU-T P.1201.1]

\[V_{MOSR} \text{ is calculated as follows:} \]
\[
V_{MOSR} = Video_{Quality} - V_{DR}
\]
(5.10)

from [ITU-T P.1201.1]

\[
V_{DR} = (Video_{Quality} - MOS_{MIN}) \cdot \frac{\left( \frac{NRE}{v13} \right)^{v14} \cdot \left( \frac{ARL}{v15} \right)^{v16} \cdot \left( \frac{MREEF}{v17} \right)^{v18}}{1 + \left( \frac{NRE}{v13} \right)^{v14} \cdot \left( \frac{ARL}{v15} \right)^{v16} \cdot \left( \frac{MREEF}{v17} \right)^{v18}}
\]
(5.11)

from [ITU-T P.1201.1]

\[IF \ rebuffering \ AND \ packet-loss \ THEN \]

\[Video_{Quality} = V_{MOSP}\]
(5.12)

from [ITU-T P.1201.1]

ELSE
Video Quality = V_MOSC \tag{5.13}

from [ITU-T P.1201.1]

ENDIF

The conversion rule of $V_{CCF}$

$$V_{CCF} = \begin{cases} a \ln V_{BR} + b & \text{if } V_{BR} < \text{Threshold} \\ c & \text{if } V_{BR} \geq \text{Threshold} \end{cases} \tag{5.5a}$$

 Threshold = d \cdot \text{Width} \cdot \text{Height}

Table B.2 – Coefficient values for the $V_{CCF}$ parameter

<table>
<thead>
<tr>
<th></th>
<th>ITU-T H.264</th>
<th>MPEG4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QCIF</td>
<td>QVGA</td>
</tr>
<tr>
<td>a</td>
<td>0.1077</td>
<td>0.0975</td>
</tr>
<tr>
<td>b</td>
<td>0.0207</td>
<td>0.0001</td>
</tr>
<tr>
<td>c</td>
<td>0.91</td>
<td>0.85</td>
</tr>
<tr>
<td>d</td>
<td></td>
<td>0.02</td>
</tr>
</tbody>
</table>

The conversion rule of $V_{PLEF}$

\text{IF } V_{ratio} < 1 \text{ THEN}

$$V_{PLEF} = \frac{\text{TotalPktNum} \cdot V_{LossRate}}{V_{Burst}}$$

\text{ELSE}

$$V_{PLEF} = \frac{\text{TotalPktNum} \cdot V_{LossRate}}{V_{Burst}} \cdot \frac{V_{Burst}}{V_{PktpF}} = \frac{\text{TotalPktNum} \cdot V_{LossRate}}{V_{PktpF}} \tag{5.5b}$$

Where TotalPktNum is to be provided by the network/service planner otherwise it can be derived from the video bitrate and measurement time as follows:

$$\text{TotalPktNum}_{tmp} = V_{BR} \cdot 1000 \cdot \text{MeasureTime} \div 1000 \cdot 8$$

\text{IF}

$$\text{TotalPktNum}_{tmp} < \text{TotalFrameNum} \text{ THEN}$$

$$\text{TotalPktNum} = \text{TotalFrameNum} + \frac{\text{TotalPktNum}_{tmp}}{10}$$

\text{ELSE}

$$\text{TotalPktNum} = \text{TotalPktNum}_{tmp}$$

Where $V_{Ratio}$ can be derived from the $V_{Burst}$ and $V_{PktpF}$ as follows:

$$V_{ratio} = \frac{V_{Burst}}{V_{PktpF}}$$

Where $V_{PktpF}$ can be derived from TotalPktNum and TotalFrameNum as follows:

$$V_{PktpF} = \frac{\text{TotalPktNum}}{\text{TotalFrameNum}} \cdot \frac{\text{TotalPktNum}}{\text{FrameRate} \cdot \text{MeasureTime}}$$
Where $V_{AIRF}$ can be derived from $V_{LossRate}$ and $V_{PktpF}$ as follows:

$$V_{AIRF} = \frac{1}{1-(1-V_{LossRate})^{V_{PktpF}}} - \frac{1-V_{LossRate}}{V_{LossRate} \cdot V_{PktpF}} \tag{5.5c}$$

Where $V_{IR}$ can be derived from $V_{LossRateFrame}$ and $GopLength$ as follows:

$$V_{IR} = 1 - \left(\frac{1-V_{LossRateFrame}}{V_{LossRateFrame} \cdot GopLength}\right)^{GopLength} \tag{5.5d}$$

Where

$$V_{LossRateFrame} = 1 - (1-V_{LossRate})^{V_{PktpF}}$$

Where

$V_{MOS}$ is the final video quality, expressed on an MOS scale [1,5], where 1 is the worst quality and 5 is the best quality

$V_{MOSC}$ is the video quality due to compression

$V_{MOSP}$ is the video due to packet loss

$V_{MOSR}$ is the video quality due to rebuffering

$V_{DC}$ is the video distortion quality due to compression

$V_{DP}$ is the video distortion quality due to packet-loss

$V_{DR}$ is the video distortion quality due to rebuffering

$V_{CCF}$ is the video content complexity

$V_{PLEF}$ is the video packet-loss event frequency

$V_{AIRF}$ is the average impairment rate of video frame

$V_{IR}$ is the impairment rate of video stream

$VideoCodec$ is the video codec

$MeasureTime$ is the measurement time in the pre-determined interval, in seconds

$V_{BR}$ is the bit rate of the video stream, in kbps

$GopLength$ is the length of GOP

$V_{Burst}$ is the average burst IP packet-loss length per video stream.

$V_{PktpF}$ is the average number of video RTP packets per video frame

$TotalPktNum$ is the total number of video RTP packets

$Framerate$ is the video frame rate, in frames per second

$V_{LossRate}$ is the loss rate of IP packets

$V_{LossRateFrame}$ is the loss rate of video frame

$NRE$ is the number of rebuffering events

$ARL$ is the average rebuffering length

$MREFF$ is the multiple rebuffering events effect factor

$MOS_{MAX}$ is the maximum value of video quality, i.e., 5.0

$MOS_{MIN}$ is the minimum value of video quality, i.e., 1.0

$Width$ is the video width

$Height$ is the video height
The coefficient values $v_1$-$v_{18}$ depend on the video codec and resolution, and are provided in Tables B.3 to B.6. They are identical to those in [ITU-T P.1201.1].

### Table B.3 – Coefficient sets for $V_{MOSC}$ and $V_{DC}$ video quality estimation

<table>
<thead>
<tr>
<th></th>
<th>ITU-T H.264</th>
<th></th>
<th></th>
<th>MPEG4</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QCIF</td>
<td>QVGA</td>
<td>HVGA</td>
<td>QCIF</td>
<td>QVGA</td>
<td>HVGA*</td>
<td></td>
</tr>
<tr>
<td>$v_1$</td>
<td>3.4</td>
<td>2.49</td>
<td>2.505</td>
<td>2.43</td>
<td>1.6184</td>
<td>1.6184</td>
<td></td>
</tr>
<tr>
<td>$v_2$</td>
<td>0.969</td>
<td>0.7094</td>
<td>0.7144</td>
<td>0.692</td>
<td>0.4611</td>
<td>0.4611</td>
<td></td>
</tr>
<tr>
<td>$v_3$</td>
<td>104.0</td>
<td>324.0</td>
<td>170.0</td>
<td>0.01</td>
<td>280.0</td>
<td>280.0</td>
<td></td>
</tr>
<tr>
<td>$v_4$</td>
<td>1.0</td>
<td>3.3</td>
<td>130.0</td>
<td>134.0</td>
<td>11.0</td>
<td>11.0</td>
<td></td>
</tr>
<tr>
<td>$v_5$</td>
<td>0.01</td>
<td>0.5</td>
<td>0.05</td>
<td>0.01</td>
<td>1.69</td>
<td>1.69</td>
<td></td>
</tr>
<tr>
<td>$v_6$</td>
<td>1.1</td>
<td>1.2</td>
<td>1.1</td>
<td>1.7</td>
<td>0.02</td>
<td>0.02</td>
<td></td>
</tr>
</tbody>
</table>

### Table B.4 – Coefficient sets for $V_{DP}$ for SLICING video quality estimation

<table>
<thead>
<tr>
<th></th>
<th>videoPLC = SLICING</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ITU-T H.264</td>
<td></td>
<td></td>
<td>MPEG4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>QCIF</td>
<td>QVGA</td>
<td>HVGA</td>
<td>QCIF</td>
<td>QVGA</td>
<td>HVGA*</td>
<td></td>
</tr>
<tr>
<td>$v_7$</td>
<td>$-0.63$</td>
<td>$-0.64$</td>
<td>$-0.05$</td>
<td>$-0.01$</td>
<td>$-0.01$</td>
<td>$-0.01$</td>
<td></td>
</tr>
<tr>
<td>$v_8$</td>
<td>1.4</td>
<td>0.81</td>
<td>0.42</td>
<td>0.99</td>
<td>0.76</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>$v_9$</td>
<td>0.01</td>
<td>0.4</td>
<td>0.72</td>
<td>0.34</td>
<td>0.39</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td>$v_{10}$</td>
<td>$-14.4$</td>
<td>$-9.0$</td>
<td>$-3.3$</td>
<td>$-0.1$</td>
<td>$-0.01$</td>
<td>$-0.01$</td>
<td></td>
</tr>
<tr>
<td>$v_{11}$</td>
<td>19.0</td>
<td>11.5</td>
<td>7.0</td>
<td>15.5</td>
<td>10.0</td>
<td>10.0</td>
<td></td>
</tr>
<tr>
<td>$v_{12}$</td>
<td>1.04</td>
<td>0.4</td>
<td>0.49</td>
<td>0.66</td>
<td>0.86</td>
<td>0.86</td>
<td></td>
</tr>
</tbody>
</table>

* Provisional values, since this condition was not included in the test plan.

### Table B.5 – Coefficient sets for $V_{DP}$ for FREEZING video quality estimation

<table>
<thead>
<tr>
<th></th>
<th>videoPLC = FREEZING</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ITU-T H.264</td>
<td></td>
<td></td>
<td>MPEG4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>QCIF</td>
<td>QVGA</td>
<td>HVGA</td>
<td>QCIF*</td>
<td>QVGA*</td>
<td>HVGA*</td>
<td></td>
</tr>
<tr>
<td>$v_7$</td>
<td>$-0.115$</td>
<td>$-0.05$</td>
<td>$-0.05$</td>
<td>$-0.115$</td>
<td>$-0.05$</td>
<td>$-0.05$</td>
<td></td>
</tr>
<tr>
<td>$v_8$</td>
<td>0.25</td>
<td>0.53</td>
<td>0.32</td>
<td>0.25</td>
<td>0.53</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>$v_9$</td>
<td>2.05</td>
<td>0.6</td>
<td>0.24</td>
<td>2.05</td>
<td>0.6</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>$v_{10}$</td>
<td>$-0.7$</td>
<td>$-0.1$</td>
<td>$-0.1$</td>
<td>$-0.7$</td>
<td>$-0.1$</td>
<td>$-0.1$</td>
<td></td>
</tr>
<tr>
<td>$v_{11}$</td>
<td>1.5</td>
<td>11.5</td>
<td>1.0</td>
<td>1.5</td>
<td>11.5</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>$v_{12}$</td>
<td>0.45</td>
<td>0.01</td>
<td>1.16</td>
<td>0.45</td>
<td>0.01</td>
<td>1.16</td>
<td></td>
</tr>
</tbody>
</table>

* Provisional values, since these conditions were not included in the test plan.
Table B.6 – Coefficient sets for video rebuffering quality estimation

<table>
<thead>
<tr>
<th></th>
<th>Single rebuffering event</th>
<th>Multiple rebuffering event</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QCIF</td>
<td>QVGA</td>
</tr>
<tr>
<td>v13</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>v14</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>v15</td>
<td>9.8</td>
<td>20.6</td>
</tr>
<tr>
<td>v16</td>
<td>0.85</td>
<td>0.37</td>
</tr>
<tr>
<td>v17</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>v18</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

B.3 Audiovisual module

As shown in Figure B.3, the LR application area audiovisual module of ITU-T G.1071 is identical to the audiovisual module for [ITU-T P.1201.1].

![Audiovisual module diagram]

Figure B.3 – Audiovisual module

AV_MOS is calculated as shown in the pseudocode below:

IF No Packet-loss AND No rebuffering
THEN

AV_MOS = AV_MOSC

(6.1)

from [ITU-T P.1201.1]

ELSEIF Packet-loss AND No rebuffering

AV_MOS = AV_MOSP

(6.2)

from [ITU-T P.1201.1]

ELSE

AV_MOS = AV_MOSR

(6.3)

from [ITU-T P.1201.1]

ENDIF

AV_MOSC is calculated as follows:

\[
AV_MOSC = av1 \cdot V_{MOSC} + av2 \cdot A_{MOSC} + av3 \cdot V_{MOSC} \cdot A_{MOSC} + av4
\]

(6.4)

from [ITU-T P.1201.1]

AV_MOSP is calculated as follows:

\[
AV_MOSP = AV_MOSC - AV_{DP}
\]

(6.5)
Where \( AV_{DP} \) is calculated as follows:

\[
AV_{DP} = (AV_{MOSC} - MOS_{MIN}) \cdot AV_{DF}
\]  
(6.6)

from [ITU-T P.1201.1]

where

\[
AV_{DF} = \frac{av5 \cdot AV_{DFV} + av6 \cdot AV_{DFA}}{1 + av5 \cdot AV_{DFV} + av6 \cdot AV_{DFA}}
\]  
(6.7)

from [ITU-T P.1201.1]

\[
AV_{DFV} = \frac{V_{DP}}{V_{MOSC}} \quad \text{or} \quad AV_{DFV} = \frac{V_{MOSC} - Video_{MOS}}{V_{MOSC}}
\]  
(6.8)

from [ITU-T P.1201.1]

\[
AV_{DFA} = \frac{A_{DP}}{A_{MOSC}} \quad \text{or} \quad AV_{DFA} = \frac{A_{MOSC} - Audio_{MOS}}{A_{MOSC}}
\]  
(6.9)

from [ITU-T P.1201.1]

\[
IF \text{ video packet-loss of audiovisual stream has occurred THEN} \quad Video_{MOS} = V_{MOSP}
\]  
(6.10)

from [ITU-T P.1201.1]

ELSE

\[
Video_{MOS} = V_{MOSC}
\]  
(6.11)

from [ITU-T P.1201.1]

ENDIF

\[
IF \text{ audio packet-loss of audiovisual stream has occurred THEN} \quad Audio_{MOS} = A_{MOS}
\]  
(6.12)

from [ITU-T P.1201.1]

ELSE

\[
Audio_{MOS} = A_{MOSC}
\]  
(6.13)

from [ITU-T P.1201.1]

ENDIF

\( AV_{MOSR} \) is calculated as follows:

\[
AV_{MOSR} = Audiovisual\ _Quality - AV\ _DR
\]  
(6.14)

from [ITU-T P.1201.1]

\[
AV_{DR} = (Audiovisual\ _Quality - MOS\ _MIN) \cdot \left( \frac{NRE}{av7} \right)^{m8} \cdot \left( \frac{ARL}{av9} \right)^{m10} \cdot \left( \frac{MREEF}{av11} \right)^{av12}
\]  
\[
1 + \left( \frac{NRE}{av7} \right)^{av8} \cdot \left( \frac{ARL}{av9} \right)^{av10} \cdot \left( \frac{MREEF}{av11} \right)^{av12}
\]  
(6.15)

from [ITU-T P.1201.1]
where

IF rebuffering AND packet-loss THEN

\[ Audiovisual\_Quality = AV\_MOSP \] (6.16)

from [ITU-T P.1201.1]

ELSE

\[ Audiovisual\_Quality = AV\_MOSC \] (6.17)

from [ITU-T P.1201.1]

ENDIF

where

\( AV\_MOS \) is final audiovisual MOS, expressed on the MOS scale [1,5], where 1 is the worst quality and 5 the best quality

\( AV\_MOSC \) is the audiovisual quality due to compression

• \( AV\_MOSP \) is the audiovisual quality due to packet-loss

• \( AV\_MOSR \) is the audiovisual quality due to rebuffering

• \( V\_MOSC \) is the estimated video quality due to compression

• \( V\_MOSP \) is the estimated video quality due to packet loss

• \( V\_DP \) is the video distortion quality due to packet-loss

• \( A\_DP \) is the audio distortion quality due to packet-loss

• \( A\_MOS \) is the estimated audio quality

• \( A\_MOSC \) is the estimated audio quality for audio compression artefacts

• \( NRE \) is the number of rebuffering events

• \( ARL \) is the average rebuffering length

• \( MREFF \) is the multiple rebuffering events effect factor

• \( AV\_DP \) is the audiovisual distortion quality due to packet-loss

• \( AV\_DFV \) represents the packet-loss distortion factor for video

• \( AV\_DFA \) represents the packet-loss distortion factor for audio

• \( AV\_DF \) represents the packet-loss distortion factor for audiovisual stream

• \( AV\_DR \) is the audiovisual distortion quality due to rebuffering

Coefficient values \( av1-av12 \) are provided in Tables B.7 to B.9 and are identical to those in [ITU-T P.1201].
Table B.7 – Coefficients sets for *AV_MOSC* audiovisual quality estimation

<table>
<thead>
<tr>
<th></th>
<th>QCIF</th>
<th>QVGA</th>
<th>HVGA</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>av1</em></td>
<td>0.7977</td>
<td>0.7495</td>
<td>0.6419</td>
</tr>
<tr>
<td><em>av2</em></td>
<td>0.03732</td>
<td>0.09736</td>
<td>0.1362</td>
</tr>
<tr>
<td><em>av3</em></td>
<td>0.02472</td>
<td>0.006725</td>
<td>0.016</td>
</tr>
<tr>
<td><em>av4</em></td>
<td>0.1657</td>
<td>0.3186</td>
<td>0.5694</td>
</tr>
</tbody>
</table>

Table B.8 – Coefficient sets for *AV_DP* audiovisual quality estimation

<table>
<thead>
<tr>
<th></th>
<th>QCIF</th>
<th>QVGA</th>
<th>HVGA</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>av5</em></td>
<td>2.908</td>
<td>1.541</td>
<td>1.94</td>
</tr>
<tr>
<td><em>av6</em></td>
<td>0.4755</td>
<td>0.96</td>
<td>2.178</td>
</tr>
</tbody>
</table>

Table B.9 – Coefficient sets for audiovisual rebuffering estimation

<table>
<thead>
<tr>
<th></th>
<th>Single rebuffering event</th>
<th>Multiple rebuffering event</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QCIF</td>
<td>QVGA</td>
</tr>
<tr>
<td><em>av7</em></td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><em>av8</em></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>av9</em></td>
<td>79.6</td>
<td>12.6</td>
</tr>
<tr>
<td><em>av10</em></td>
<td>0.3</td>
<td>0.26</td>
</tr>
<tr>
<td><em>av11</em></td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><em>av12</em></td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

* Provisional values, since these conditions were not included in the test plan.
# SERIES OF ITU-T RECOMMENDATIONS

<table>
<thead>
<tr>
<th>Series</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Organization of the work of ITU-T</td>
</tr>
<tr>
<td>D</td>
<td>General tariff principles</td>
</tr>
<tr>
<td>E</td>
<td>Overall network operation, telephone service, service operation and human factors</td>
</tr>
<tr>
<td>F</td>
<td>Non-telephone telecommunication services</td>
</tr>
<tr>
<td>G</td>
<td><strong>Transmission systems and media, digital systems and networks</strong></td>
</tr>
<tr>
<td>H</td>
<td>Audiovisual and multimedia systems</td>
</tr>
<tr>
<td>I</td>
<td>Integrated services digital network</td>
</tr>
<tr>
<td>J</td>
<td>Cable networks and transmission of television, sound programme and other multimedia signals</td>
</tr>
<tr>
<td>K</td>
<td>Protection against interference</td>
</tr>
<tr>
<td>L</td>
<td>Environment and ICTs, climate change, e-waste, energy efficiency; construction, installation and protection of cables and other elements of outside plant</td>
</tr>
<tr>
<td>M</td>
<td>Telecommunication management, including TMN and network maintenance</td>
</tr>
<tr>
<td>N</td>
<td>Maintenance: international sound programme and television transmission circuits</td>
</tr>
<tr>
<td>O</td>
<td>Specifications of measuring equipment</td>
</tr>
<tr>
<td>P</td>
<td>Terminals and subjective and objective assessment methods</td>
</tr>
<tr>
<td>Q</td>
<td>Switching and signalling</td>
</tr>
<tr>
<td>R</td>
<td>Telegraph transmission</td>
</tr>
<tr>
<td>S</td>
<td>Telegraph services terminal equipment</td>
</tr>
<tr>
<td>T</td>
<td>Terminals for telematic services</td>
</tr>
<tr>
<td>U</td>
<td>Telegraph switching</td>
</tr>
<tr>
<td>V</td>
<td>Data communication over the telephone network</td>
</tr>
<tr>
<td>X</td>
<td>Data networks, open system communications and security</td>
</tr>
<tr>
<td>Y</td>
<td>Global information infrastructure, Internet protocol aspects and next-generation networks</td>
</tr>
<tr>
<td>Z</td>
<td>Languages and general software aspects for telecommunication systems</td>
</tr>
</tbody>
</table>