中 关 村 现 代 信 息 消 费 应 用 产 业 技 术 联 盟 团 体 标 准 T/INFOCA 004-2022

移动短视频体验质量指标与评测方法

Mobile short video experience quality indicators and evaluation methods

2022年10 月12日发布 2022年10 月 12 日实施

中关村现代信息消费应用产业技术联盟 发布

目 次

目 次 2

[前 言 0](#_Toc111555957)

[移动短视频体验质量指标与评测方法 1](#_Toc111555958)

[1 范围 1](#_Toc111555959)

[2 规范性引用文件 1](#_Toc111555960)

[3 术语、定义和缩略语 2](#_Toc111555961)

[3.1 缩略语 2](#_Toc111555962)

[3.2 术语和定义 3](#_Toc111555963)

[4 短视频QoE评估架构 3](#_Toc111555964)

[4.1 关键指标 3](#_Toc111555965)

[4.2 短视频QoE评估架构 5](#_Toc111555966)

[5 短视频QoE指标构成 6](#_Toc111555967)

[5.1 视频质量输入参数 6](#_Toc111555968)

[5.2 音频质量输入参数 7](#_Toc111555969)

[5.3 呈现体验质量输入参数 8](#_Toc111555970)

[5.4 交互体验质量输入参数 8](#_Toc111555971)

[5.5 模型输出参数 9](#_Toc111555972)

[6 短视频功能用户体验评估算法 10](#_Toc111555973)

[6.1 总体模型综述 10](#_Toc111555974)

[6.2 模型三大模块综述 10](#_Toc111555975)

[6.2.1. 视听体验 10](#_Toc111555984)

[6.2.2. 呈现体验 13](#_Toc111555985)

[6.2.3. 交互体验 0](#_Toc111555986)

[附录A](#_Toc111555987)[ITU-T P.1203工作模式 1](#_Toc111555987)

[附录B 关于PPD计算方法 1](#_Toc111555988)

前 言

本标准是针对移动终端（智能手机，平板电脑）场景下短视频QoE评估算法和参数。

本标准参考国内和国际相关标准，并结合国内网络的实际情况制定。

本标准按照GB/T 1.1-2009给出的规则起草。

注意本文件的某些内容可能涉及专利，本文件的发布机构不承担识别这些专利的责任。 本标准由中关村现代信息消费应用产业技术联盟提出并归口。

本标准起草单位：中关村现代信息消费应用产业技术联盟、华为技术有限公司、上海兆言网络科技有限公司、TCL华星光电技术有限公司、北京邮电大学、联通研究院、中国信息通信研究院、西安电子科技大学、北京市博汇科技股份有限公司、深圳市中兴微电子技术有限公司、北京快手科技有限公司、中恒达（北京）软件测评科技有限公司、中国电信集团有限公司、中国移动通信有限公司研究院、优酷网络技术（北京）有限公司、德科仕通信（上海）有限公司。

本标准主要起草人：陈红、宋祖平、武亮平、段涛、黄卫东、闫石、贾武、王亚军、杨崑、杨付正、张家斌、孔德辉、马英武、陈劼联、陈戈、张世俊、李静、吴雪波、张春蕾、吴敬芳、郑川川、余大力、刘璇。

移动短视频体验质量指标与评测方法

1 范围

本标准规定了在固定和移动宽带网络中基于移动终端（智能手机，平板电脑）观看短视频（ShortVideo）时的QoE评估模型，分析了影响短视频QoE的主要因素和参数，定义了短视频QoE的评估计算方法。

本标准适用于公用电信网、公用互联网环境下短视频的QoE进行综合评估，以及对影响短视频QoE的因素进行定位和分析。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅所注日期的版本适用于本文 件。凡是不注日期的引用文件，其最新版本（包括所有的修改单） 适用于本文件。

ITU-R BT.500-11，Methodology for the subjective assessment of the quality of television pictures

ITU-T P.913 (2021) Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment.

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 (2017) 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 (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 (2017) Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport – Quality integration module.

3 术语、定义和缩略语

3.1 缩略语

以下缩略语适用于本文件。

缩略语 中文全称 英文全称

QoE 用户体验质量 Quality of User Experience

HTTP 超文本传输协议 Hypertext Transfer Protocol

HLS Apple的动态码率自适应技术 HTTP Live Streaming

DASH 动态自适应流媒体国际标准 Dynamic Adaptive Streaming over HTTP

QUIC 基于UDP的低时延的互联网传输层协议 Quick UDP Internet Connection

RTMP 实时消息传输协议 Real Time Messaging Protocol

TCP 传输控制协议 Transmission Control Protocol

UDP 用户数据报协议 User Datagram Protocol

HDR 高动态范围 High-Dynamic Range

WCG 宽色域 Wide Color Gamut

MOS 主观体验评分 Mean Opinion Score

PPI 每英寸所拥有的像素个数 Pixels Per Inch

PPD 角分辨率 Pixels Per Degree

BPP 每像素编码比特数 (coded) Bits Per Pixel

ITU 国际电信联盟 International Telecommunication Union

3.2 术语和定义

下列术语和定义适用于本文件。

3.2.1

用户体验质量 Quality of User Experience

用户感受到的服务过程中硬件操作、服务操作和内容展示的质量水平。

3.2.2

输出采样间隔 Output Sampling Interval

指解析后的视频或音频每隔1秒钟（不考虑卡顿）输出一次，且音频质量（QA）和视频质量（QV）的输出采样间隔必须与质量整合模块（QIgt）期望的输入相匹配。

3.2.3

卡顿 Stalling Event

在观看视频过程中，出现画面滞帧现象（通常在手机App中以转圈，出现实时下载速率等形式存在）。

3.2.4

4K

4K分辨率是指水平方向每行像素值达到或者接近4096个，多数情况下特指4096\*2160分辨率。而根据使用范围的不同，4K分辨率也有各种各样的衍生分辨率，例如Full Aperture 4K的4096\*3112、Academy 4K的3656\*2664以及UHDTV标准的3840\*2160等，都属于4K分辨率的范畴。

3.2.5

短视频

在各种[新媒体](https://baike.baidu.com/item/%E6%96%B0%E5%AA%92%E4%BD%93/6206%22%20%5Ct%20%22_blank)平台（如抖音、快手等）上播放的、适合在移动状态和休闲状态下观看的、高频推送的视频内容（以竖屏方式观看内容为主），几秒到几分钟不等。

4 短视频QoE评估架构

4.1 关键指标

当用户发出观看请求时，流媒体服务系统就会根据用户喜好，将存放在片源库中的节目信息主动推送，以视频和音频流文件的方式，通过高速传输网络传送到用户终端。短视频一般采用单播网络来实现，所谓单播就是利用一种协议将IP数据包从一个信息源传送到一个目的地，此时信息的接收和传递只在两个节点之间进行。短视频通过互联网服务商的移动OTT视频业务系统实现，本标准涉及的短视频涵盖了如下主流传输协议： HTTP/TCP/IP， DASH/QUIC/UDP/IP（或其他可靠UDP传输协议）；采用以下视频编码格式： MP4等；显示设备：智能手机、平板电脑等屏幕尺寸在10英寸以下的移动终端。

用户使用短视频的过程包括启动终端/播放器，通过滑屏选择观看内容，观看等操作。影响短视频用户体验的因素来自终端的软硬件、传输协议、编解码、网络、平台、内容源质量等多个方面； 其中需要考虑的关键因素可以被归纳到视听体验质量、用户的交互体验质量及观看体验质量等三个方面。

* 视频体验质量，取决于视频源的清晰度、流畅度、保真度（色调、对比度） 等因素，涵盖了视频源的分辨率、帧率、码率、内容、编码和终端六个维度的指标，可以编码参数、分辨率、帧率、 码率、信号层质量等客观指标度量, 相对于PPI而言，PPD可以更好地适应曲面屏、非平面的折叠屏等终端形态的演进需要，因此通过PPD来综合衡量视频分辨率、屏幕分辨率、屏幕尺寸等参数对观看清晰度的影响。
* 音频体验质量，取决于音频源的清晰度、流畅度、保真度等因素，可以音频采样率、声道数、码率、编码方式、编码参数等客观指标度量。
* 呈现体验质量，即用户的收视/收听体验，取决于观看过程中出现的节目信号质量，影响因素包括卡顿、音画不同步等，可以通过视音频信息的传输性能和质量损伤等客观指标进行衡量。
* 用户的交互体验质量，即用户的互动过程体验，取决于系统对用户交互操作的响应速度，涵盖了平台、网络、终端性能指标，可以用户一次观看行为中对视频首帧响应时延、拖拽响应时延、X倍速响应时延、短视频切换响应时延等客观指标度量。其它人机界面的因素也非常重要。



图 1 影响短视频用户体验的关键因素

此外，使用环境、人感受的惯性、终端的大小、设备的安装维护、服务的丰富性和连贯性等其它部

分的指标也对用户体验有一定的影响。

* 开展短视频用户体验评估时的环境会影响终端用户的感受，在移动环境下用户手持移动终端 观看视频时和用户在静止环境下在电视或者电脑前观看视频时的感受有着明显差异，如对卡顿的忍耐程度会不同。
* 不同大小的终端，用户对视频清晰度的要求也不一样。
* 近因/遗忘效应：呈现体验和交互体验客观指标对用户视频体验质量的影响权重随着观看行为的持续而弱化，当卡顿事件出现时，视频体验质量立即下降，当卡顿结束，恢复正常播放时，用户体验就会逐渐缓慢恢复。假如后续一直能持续正常播放，实时的质量体验会逐渐恢复到正常值。

为了科学的评估和量化短视频的用户体验质量，把评估场景分为会话场景（表征用户一次完整的 观看行为的体验质量） 和实时场景（表征用户实时观看过程中的体验质量） 。对用户在一定周期内（周、 月、年） 会话场景下的体验质量进行统计分析可以反映用户使用短视频的真实感知体验。

开展短视频用户体验评价需要综合考虑音视频源质量、操作体验质量、播放体验质量，使用环境、 时间持续性和频度等各方面因素。用户的视频体验质量得分，综合了视听体验质量、呈现体验质量、交互体验质量，是真实反映用户体验感知的平均主观评分（MOS） 。

4.2 短视频QoE评估架构

短视频QoE评估模型架构如图2所示：

1. 视听体验质量用于表征不受卡顿、等体验损伤影响情况下，收看短视频时所见所听的主观感受。视听体验质量主要由视频质量和音频质量构成。视频质量主要与内容清晰度（以PPD表征，受内容分辨率、屏幕尺寸、屏幕分辨率、观看距离等因素影响）、内容流畅度（以帧率表征）和编码参数（如编码算法、码率、BPP）有关。音频质量主要与音频采样率、声道数（如单声道、立体声、杜比音效）、编码参数（如编码算法、码率）有关。
2. 呈现体验质量用于表征受网络传输质量影响的体验损伤，主要与收看短视频过程中因传输带宽、时延抖动和丢包导致的卡顿等问题有关。此外，音画不同步，以及X倍速播放对用户理解造成的影响也归入到呈现体验质量进行表征。
3. 交互体验质量用于表征短视频交互行为（如视频初始加载、拖拽、X倍速播放）的影响，主要与响应时延有关。值得一提的是，因拖拽、X倍速播放导致的卡顿问题仍然归入到呈现体验质量进行表征。



图 2 短视频QoE评估模型架构

根据图2，除QIgt（Quality integration module，质量整合模块）之外的其他模块，模型输入和输出均比较清晰，此处不做展开。这里重点介绍一下比较复杂的QIgt模块。QIgt模块由QAVE、QPE和QInE三个子模块构成。

* QAVE（Audio-Visual Experience Quality，视听体验质量）子模块，其输入是：1）每输出采样间隔的视觉体验质量（O.21）；2）每输出采样间隔的音频质量（O.22）。其输出是：1）基于每输出采样间隔的视听体验质量（O.31）；2）基于会话的视听体验质量（O.32）。
* QPE（Presenting Experience Quality，呈现体验质量）子模块，其输入是：1）会话级的卡顿感知质量（QStall）； 2）每输出采样间隔的音视频同步（*TA2V*）参数；3）每输出采样间隔的倍率（*XSp*，当X倍速播放时）参数。其输出是：基于会话的呈现体验质量（O.33）。关于QPE子模块的更多细节请参考本标准6.2.2小节的相关描述。
* QInE（Interactive Experience Quality，交互体验质量）子模块，其输入是：1）每输入采样间隔的交互响应参数（I.15）；其输出是：1）基于每输出采样间隔的交互体验质量（O.34）；2）基于会话的交互体验质量（O.35）。关于QInE子模块的更多细节请参考本标准6.2.3小节的相关描述。

短视频QoE指标从用户层面出发，可以定义为终端用户对短视频的综合主观感知，反映用户观看短视频的体验综合评分，用短视频\_MOS表示。

* 定义短视频\_MOS = f (QAVE, QPE, QInE)，即用户体验综合评分为（视听体验质量QAVE，呈现体验质量QPE，交互体验质量 QInE）的函数关系

5 短视频QoE指标构成

5.1 视频质量输入参数

视频质量提取的输入参数包括I.11和I.GEN，其中I.11表征内容源中视频的编码信息，I.GEN则主要表征终端的显示能力（如屏幕分辨率，刷新率等）。

表1 视频质量输入参数

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 参数 | 数据源 | 定义/取值 | 备注 |
| I.11 |
| 1 | 视频码率 | 终端播放器 | 视频编码时单位时间取样的视频数据量（Kbps） | VideoBitrate，缩写为*BrV* |
| 2 | 视频帧率 | 终端播放器 | 视频编码时每秒的帧数（fps） | VideoFrameRate，缩写为*FRV* |
| 3 | 视频水平分辨率 | 终端播放器 | 视频编码时水平方向的像素点个数 | VideoHorizontalResolution，缩写*Rh* |
| 4 | 视频垂直分辨率 | 终端播放器 | 视频编码时垂直方向的像素点个数 | VideoVerticalResolution，缩写为*Rv* |
| 5 | 编码算法 | 终端播放器 | 视频的编码算法： H.265、H.264、VP9; | VideoCodecAlgorithm，缩写为*CodecV* |
| 6 | metadata参数（Mode 0） | 终端播放器 | 色深（Bit depth） | HDR+WCG典型参数色深：10颜色初选：BT.2020（WCG）传输特性：HLG（HDR）SDR典型参数色深：8颜色初选：BT.709 |
| 颜色初选（Color primaries） |
| 矩阵系数（Matrix coefficients） |
| 传输特性（Transfer characteristics） |
| 信号层参数（Mode 2） | 源端检测点终端播放器 | 块度（Blockiness） | Mode 0：只能粗略地定性评估HDR+WCG的影响；Mode 2：在Mode 0基础上结合信号层指标（如对比度，色彩度等）可以完整地定量评估HDR+WCG对MOS的影响 |
| 模糊度（Blurriness） |
| 色彩度（Chroma） |
| 对比度（Contrast） |
| 噪点度（Noise） |
| I.GEN |
| 7 | 屏幕水平分辨率 | 终端OS | 终端屏幕水平方向的像素点个数 | ScreenHorizontalResolution，缩写为*RSh* |
| 8 | 屏幕垂直分辨率 | 终端OS | 终端屏幕垂直方向的像素点个数 | ScreenVerticalResolution，缩写为*RSv* |
| 9 | 屏幕尺寸 | 终端OS | 终端屏幕对角线的长度（inch） | ScreenSize，缩写为*ScrSize* |
| 10 | 屏幕刷新率 | 终端OS | 终端屏幕每秒刷新的次数（Hz） | ScreenRefreshRate，缩写为*ScrRR* |
| 11 | 观看距离 | 经验数据 | 用户观看视频时距离终端屏幕的最佳观看距离（cm） | DistanceToScreen，缩写为*D2Scr* |

5.2 音频质量输入参数

音频质量提取的输入参数是I.12，表征内容源中音频的编码信息。

表2 音频质量输入参数

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 参数 | 数据源 | 定义/取值 | 备注 |
| I.12 |
| 12 | 音频码率 | 终端播放器 | 单位时间传输的音频数据量（Kbps） | AudioBitrate，缩写为*BrA* |
| 13 | 音频声道 数 | 终端播放器 | 声道数是声音录制时的音源数量或 回放时相应的扬声器数量。 | NumberOfChannels，缩写为*NoCA* |
| 14 | 编码方式 | 终端播放器 | 音频的编码算法：涵盖业界主流如AAC-LC v4、AC3、HE-AAC v4、E-AC-3等 | AudioCodecAlgorithm，缩写为*CodecA* |

5.3 呈现体验质量输入参数

呈现体验质量提取的输入参数包括I.13和I.14，其中I.13表征卡顿事件信息， I.14表征音画同步、倍率（X倍速播放时可能影响用户对内容的理解）等信息。

表3 呈现体验质量输入参数

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 参数 | 数据源 | 定义/取值 | 备注 |
| I.13 |
| 15 | 第i次卡顿事件开始时间 | 终 端 播 放 器 （解码器） | 在观看短视频过程中第i次发生缓冲停顿的开始时间（相对于视频剪辑而言，s） | StartTimeInStallEventi，缩写为*StartTimei*(i = 1,2,…,5) |
| 16 | 第i次卡顿事件持续时长 | 终端播放器 | 在观看短视频过程中第i次发生缓冲停顿的持续时长（s） | StallLengthInStallEventi，缩写为*StallLeni* |
| 17 | 卡顿次数 | 终端播放器 | 在观看短视频过程中发生缓冲停顿的次数 | StallNumber，缩写为*StallNum* |
| I.14 |
| 18 | 音画同步时延 | 终端播放器 | 短视频的声音和画面的同步时延（ms） | AudioDelayRelativeToVideo，缩写为*TA2V* |
| 19 | 倍率（倍速播放时） | 终端播放器 | 倍速播放的倍率，如0.75X、1X、1.25X、1.5X、2X、3X等 | X-Speed，缩写为*XSp* |

5.4 交互体验质量输入参数

交互体验质量提取的输入参数是I.16，表征短视频交互行为（初始加载，拖拽，X倍速）的响应时延。

表4 交互体验质量输入参数

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 参数 | 数据源 | 定义/取值 | 备注 |
| I.15 |
| 20 | 视频初始加载时长 | 终端播放器 | 从点击开始播放按钮到短视频首帧画面出现的时长（s） | InitialLoadingDelay，缩写为*TIniLoad* |
| 21 | 交互行为标识 | 终端播放器 | 根据短视频播放过程中的交互行为，用从低到高4个比特位分别表示初始加载、拖拽、X倍速播放和短视频切换时延。如：1-仅初始加载, 2-初始加载和拖拽, 3-X倍速响应时延，4-短视频切换时延 | InteractionFlag，缩写为*ItractFlag* |
| 22 | 拖拽响应时延 | 终端播放器 | 从进度条拖拽后的位置开始，到拖拽后首帧画面出现结束，两者之间持续的时长（s） | DragResponseDelay，缩写为*TDragRsp* |
| 23 | X倍速响应时延 | 终端播放器 | 从切换到X倍速选项开始，到完成X倍速音画效果结束，两者之间持续的时长（s） | XSpeedResponseDelay，缩写为*TXSp* |
| 24 | 短视频切换时延 | 终端播放器 | 从手指滑屏开始计时，到屏幕上开始显示另一个视频画面结束，二者之间的时延。 | ShortVideoSwitchingDelay，缩写为*TShortVideoSw* |

5.5 模型输出参数

短视频QoE评估模型的输出信息，如表5所示。

表5 短视频QoE评估模型的输出参数

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 编号 | 参数 | 定义/取值 | 频次 | 备注 |
| 1 | O.21 | 对短视频的画质评分，包括受终端显示影响的视频质量。取值范围：1-5 | 每一个短视频会话每一个输出采样间隔（如：每1秒） | Visual Quality，缩写为*QV* |
| 2 | O.22 | 对短视频的音质评分。取值范围：1-5 | 每一个短视频会话每一个输出采样间隔（如：每1秒） | Audio Quality，缩写为*QA* |
| 3 | O.31 | 对短视频的视听质量的评分。取值范围：1-5 | 每一个短视频会话每一个输出采样间隔（如：每1秒） | 　 |
| 4 | O.32 | 对短视频的视听质量的评分。取值范围：1-5 | 每一个短视频会话 | Visual Audio Experience Quality，缩写为*QAVE* |
| 5 | O.33 | 网络传输质量（如带宽、丢包、时延）对短视频体验质量影响的评分。取值范围：1-5 | 每一个短视频会话 | Presenting Experience Quality，缩写为*QPE* |
| 6 | O.34 | 交互响应时延对短视频体验质量影响的评分。取值范围：1-5 | 每一个短视频会话每一个输出采样间隔（如：每1秒） | 　 |
| 7 | O.35 | 交互响应时延对短视频体验质量影响的评分。取值范围：1-5 | 每一个短视频会话 | Interaction Experience Quality，缩写为*QInE* |
| 8 | *QStall* | 对短视频的卡顿感知质量。取值范围：1-5 | 每一个短视频会话 | Stall Quality |
| 9 | O.41 | 对短视频体验质量的综合评分。取值范围：1-5 | 每一个短视频会话 | 短视频\_MOS |

6 短视频功能用户体验评估算法

6.1 总体模型综述

即总体模型为三个模块（视听体验质量*QAVE*，呈现体验质量*QPE*，交互体验质量*QInE*）的函数关系，用于评估会话场景（用户一次短视频观看行为时长建议不低于30s，至少包含一次滑屏行为）的用户体验 。

（1）

公式（1）中*v*1~*v*2 分别是呈现体验质量和交互体验质量的动态加权系数。初始权重系数值由大数据调研的结果得到，同时叠加惩罚因子，当分项的得分变化时，权值系数也会相应调整。

6.2 模型三大模块综述

1.
2.
3.
4.
5.
6. 1.
	2.

### 视听体验

根据ITU-T P.1203视频体验建模规范，本标准主要基于附录A所述的Mode 0方式，也即基于元数据信息的轻量级建模方式，构建视频质量O.21和音频质量O.22的评估模型。O.21和O.22的模型参数T为媒体长度，以秒为单位。如果同时考虑了音频和视频，并且它们的长度不相等，则应使用两个输入的较短持续时间。 较长的输入应在末尾被截断以匹配较短输入的持续时间。输入/输出采样的时间t的取值范围：0 – T，单位是秒。

6.2.1.1 视频质量评估

*QV*= *f* (*ScreenSize*, *ScreenRefreshRate*, *ScreenResolutionHorizontal*, *ScreenResolutionVertical*, *DistanceFromEyeToScreen*, *VideoResolutionHorizontal*, *VideoResolutionVertical*, *VideoBitrate*, *VideoFrameRate*, *VideoCodec*, *Bitdepth*,*TransferCharacteristics*,*ColorPrimaries*)

当基于Mode 0（从元数据中提取分辨率、码率、帧率、编码算法等信息）时，视频质量*QV*（O.21）主要考虑了短视频编码、时域失真、因视频与屏幕分辨率不一致导致的空域失真等因素的影响。

（2）

（3）

（4）

（5）

（6）

（7）

（8）

（9）

公式（3）~（7）中根据不同的视频编码算法（H.265、H.264、VP9）对应不同的模型系数v3~v15取值。

注：

1. 由于本标准主要基于Mode 0方式构建，从metadata提取的HDR+WCG参数（如Bit depth，Transfer characteristics，Color Primaries等）只能定性描述对视频质量的影响；或者通过增加一个补偿项来近似表征HDR+WCG特性对视频质量的影响。如要完整地定量描述HDR+WCG特性对视频质量的影响，需要基于Mode 2（因为重建像素值就必须解析码流）。
2. 关于PPD计算方法的推导过程，可参考附录B。
3. 当视频和屏幕水平分辨率（即像素个数）不同时，公式（8）中应取二者的较小值，以便反映用户实际看到的水平像素个数。
4. 竖屏观看时（如抖音、快手之类的短视频），上述公式中，公式如下所示：

6.2.1.2 音频质量评估

*QA*  *f (* *AudioBitrate*, *NumOfAudioChannels*, *AudioSamplingRate*, *AudioCodec* )

当基于Mode 0（从元数据中提取采样率、码率、声道数、编码算法等信息）时，音频质量*QA*（O.22）评估模型为：

（10）

（11）

公式（11）中根据不同的音频编码算法（AAC-LC、AC3、HE-AAC、E-AC3）和声道数对应不同的模型系数v16~v20取值。

6.2.1.3 视听体验质量评估

基于输出采样间隔的视听体验质量（O.31）兼顾了下述两个方面：视频质量*QV*（O.21），音频质量*QA*（O.22），其评估模型为：

（12）

公式（12）中*v*21~*v*24是模型系数。

最后，基于会话的视听体验质量（O.32，即*QAVE*）评估是在当前实时评分和上一时刻会话评分的基础上进行计算得到的。

（13）

公式（13）中*v*31是模型系数。

\

### 呈现体验

6.2.2.1 卡顿感知质量评估

在呈现体验中与卡顿相关的过程参数有3个，由原始入参通过一定运算得到：

* *StallNum*：卡顿次数，系原始入参
* *TotalStallLen*：卡顿时长的加权求和
* *AvgStallInterval*：卡顿事件的平均时间间隔

1、TotalStallLen的计算

 为了计算*TotalStallLen*，必须首先为每个单独的卡顿事件i分配一个权重*w\_stalli*，具体取决于其在媒体会话中的位置，根据遗忘曲线进行衰减。

（14）

（15）

（16）

2、AvgStallInterval的计算

（17）

注：

1. 如果卡顿事件数（*StallNum*）为0或1，则卡顿事件的平均时间间隔为0。
2. 参数*StallNum*仅统计卡顿事件（不包括初始加载事件）。如果卡顿事件数为0，则*TotalStallLen*为0。

最后，卡顿感知质量（*QStall*）评估模型为：

（18）

（19）

公式（14）~（19）中*v*32~*v*39是模型系数。

6.2.2.2 呈现体验质量评估

当终端对丢包的处理基于无丢包掩盖机制时，呈现体验损伤卡顿；当终端对丢包的处理基于有丢包掩盖机制时，呈现体验损伤仅包括卡顿（即*PsIpmFlag*=0）。

此外，考虑到音画同步/X倍速播放的影响，故呈现体验质量（O.33）评估模型为：

O.33

（24）

（25）

（26）

公式（24）~（26）中*v*25~*v*30 及 *v*45是模型系数。

### 交互体验

6.2.3.1 交互体验质量评估

对于短视频，基于输出采样间隔的交互体验质量（O.34）的公式为：

（27）

公式（27）中，*DMOSom*(*ItractFlag*)为交互响应时延导致的交互一致性下降量。

其中，短视频*DMOSom*(*ItractFlag*)与交互响应时延的关系可以表示为：

（28）

公式（28）中*v*46~*v*57是模型系数。

最后，基于会话的交互体验质量（O.35）评估模型为：

（29）

（30）

（31）

公式（31）中*v*58~*v*61是模型系数。

附录A ITU-T P.1203工作模式

根据ITU-T P.1203及P.1203.1/2规范，本建议标准的QV/QA评估模型基于Mode 0方式构建：

* Mode 0：从元数据中获得的信息，例如从DASH中使用的清单文件中获取有关编解码器、编码和显示分辨率、码率、帧率，以及卡顿的信息。
* Mode 1：来自模式0的所有信息，以及基于数据包报头解析的其他视频和音频帧信息（例如视频帧的大小和持续时间，以及视频帧的类型（区分I帧和非I帧））。
* Mode 2：来自模式1的所有信息，以及高达2%（以字节为单位）基于深度数据包解析的整体媒体流信息和部分码流解析。
* Mode 3：来自模式1的所有信息，以及基于码流解析的完整媒体流信息。

不同工作模式之间的关系图3所示。



图 3 ITU-T P.1203.1工作模式

附录B 关于PPD计算方法

如图4所示，以手机屏幕为例，横向PPD计算公式如下：

（1）

（2）

（3）

一般地，手机公开的基本参数是对角线尺寸*S*（单位：英寸），水平和垂直分辨率（像素个数），故根据公式（2）和（3）可知，

（4）

将公式（4）代入公式（1）可得：

（5）

同理可知，纵向PPD公式可得：

（6）

至此，公式（5）即为当横屏观看时手机屏幕PPD的最终计算公式, 公式（6）即为当竖屏观看时手机屏幕PPD的最终计算公式。

此外，常见移动终端（智能手机、平板电脑）的舒适观看距离D典型值分别为30cm和41cm。



图 4 PPD计算示意图

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |
|  |
|  |
|  |
|  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |
|  |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |