

ICS 33.020  
CCS M01

T/INFOCA 8-2022

# 团 体 标 准

T/INFOCA8-2022

---

## 移动直播视频用户体验质量（QoE）观众端评测方法

Quality of Experience (QoE) evaluation method for mobile live broadcast audience

2022年10月31日发布

2022年10月31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.....	1
前 言.....	2
移动直播视频用户体验质量 (QoE) 观众端评测方法.....	1
1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义、缩略语.....	1
3.1 术语和定义.....	1
3.1 缩略语.....	2
4 移动直播QoE评估模型架构.....	3
4.1 移动直播评估概述.....	3
4.2 观众端关键体验指标.....	3
4.3 移动直播QoE评估模型架构.....	4
5 移动直播质量指标构成.....	6
5.1 观众端视频质量输入参数.....	6
5.2 观众端音频质量输入参数.....	7
5.3 观众端呈现体验质量输入参数.....	7
5.4 观众端交互体验质量输入参数.....	8
5.5 观众端QoE模型输出参数.....	9
6 移动直播观众端用户体验评估算法.....	10
6.1 总体算法综述.....	10
6.2 四大模块算法综述.....	10
6.2.1. 视听体验.....	11
6.2.2. 呈现体验.....	14
6.2.3. 交互体验.....	16
附录A (资料性) ITU-T P. 1203工作模式.....	17
附录B (资料性) 关于PPD计算方法.....	18
参考文献.....	20
图 1 影响直播观众端体验的关键因素.....	4
图 2 移动直播观众端QoE评估模型架构.....	5
表 1 观众端视频质量输入参数.....	6
表 2 观众端音频质量输入参数I.12.....	7
表 3 观众端呈现体验质量输入参数.....	7
表 4 交互体验质量输入参数I.16.....	8
表 5 直播观众端QoE评估模型的输出参数.....	9

## 前 言

本文件按照GB/T 1.1-2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请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专利的责任。

本文件由中关村现代信息消费应用产业技术联盟提出并归口。

本文件起草单位：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、北京快手科技有限公司、北京密境和风科技有限公司、优酷网络技术（北京）有限公司、北京爱奇艺科技有限公司、聚好看科技股份有限公司、中移（杭州）信息技术有限公司、中国电信集团有限公司、中国移动通信有限公司研究院、中国联通研究院、中国信息通信研究院、北京值得买科技股份有限公司、智者四海（北京）技术有限公司、北京邮电大学、上海交通大学、北京理工大学、北京市博汇科技股份有限公司、TCL华星光电技术有限公司、上海艾策通讯科技有限公司、西安电子科技大学、深圳市中兴微电子技术有限公司、中恒达（北京）软件测评科技有限公司、德科仕通信（上海）有限公司、北京未来媒体科技股份有限公司。

本文件主要起草人：陈红、宋祖平、武亮平、段涛、马英武、高玮泽、唐赓、李静、崔艳春、史东平、刘帅、任子健、罗传飞、陈戈、尹之帆、张世俊、贾武、杨琛、黄挺、王亚军、杨崑、刘峰、郭锦程、闫石、宋利、翁冬冬、杨付正、张家斌、黄卫东、刘文泉、孔德辉、陈劼联、吴雪波、王倩、缪旭阳、张春蕾、吴敬芳、郑川川、余大利、姜卫平、郭忠武、钟琳、刘璇。

# 移动直播视频用户体验质量（QoE）观众端评测方法

## 1 范围

本文件提供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）进行移动直播时的用户体验质量（QoE）评估模型，给出了影响移动直播用户体验质量（QoE）的指标构成，描述了移动直播用户体验质量（QoE）的评估计算方法。

本文件适用于公用电信网、公用互联网等环境下，通过互联网服务商的固定或移动OTT视频业务系统实现的移动直播场景下用户体验质量（QoE）综合评估，以及对影响移动直播QoE的因素进行定位和分析。

## 2 规范性引用文件

本文件没有规范性引用文件。

## 3 术语和定义、缩略语

### 3.1 术语和定义

下列术语和定义适用于本文件。

#### 3.1.1

**用户体验质量** quality of user experience

用户体验移动直播的服务过程中相应的视听体验质量、交互体验时延、呈现体验的质量水平。

#### 3.1.2

**移动直播** live streaming video services in mobile network

通过互联网服务商的音视频业务系统，使用移动终端设备（如智能手机、平板电脑等），主播将产生的音视频流实时地推流到直播云平台，观众实时地接收主播的音视频流进行观看的过程。

#### 3.1.3

**主播** host

直播频道内进行实时音视频流推流到直播云平台的用户。

#### 3.1.4

**观众** audience

直播频道内从直播云平台通过拉流拉取直播内容进行观看的用户。

#### 3.1.5

**推流** push stream

把采集阶段封包好的内容传输到直播平台服务器的过程。

### 3.1.6

拉流 pull stream

播放端用户通过拉流地址从服务器拉取直播内容的过程。

### 3.1.7

卡顿 freeze

在直播过程中，出现音频或视频的播放不连续、不流畅、画面滞帧等现象。

### 3.1.8

花屏 blurred screen

在直播过程中，画面出现图像紊乱，异常颜色的方块图等部分画面无法有效显示的问题。

### 3.1.9

输出采样间隔 output sampling interval

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（ $Q_A$ ）和视频质量（ $Q_V$ ）的输出采样间隔与质量整合模块（ $Q_{igt}$ ）期望的输入相匹配。

### 3.1.10

主播连麦PK anchors are connected to the microphone pk

主播与其他主播间进行连麦互动PK。

### 3.1.11

音画同步时延 audio-to-video synchronization delay

直播中观众端播放器接收的音频流声音和观看到的画面的之间时间戳的差值。

## 3.1 缩略语

下列缩略语适用于本文件。

AAC: 高级音频编码 (Advanced Audio Coding)

BPP: 每像素编码比特数 ((coded) Bits Per Pixel)

IPTV: IP电视 (Internet Protocol Television)

ITU: 国际电信联盟 (International Telecommunication Union)

ITU-T: 国际电信联盟标准化部门 (International Telecommunication Union-Telecommunication Standardization Sector)

MOS: 主观体验评分 (Mean Opinion Score)

OS: 操作系统 (Operating System)

OTT: 互联网向用户提供各种应用服务 (Over The Top)

PK: 比赛、挑战、淘汰 (PlayerKilling)

PPD: 每一度视场角的像素数 (Pixels Per Degree)

PPI: 每英寸的像素数 (Pixels Per Inch)

QoE: 用户体验质量 (Quality of User Experience)

## 4 移动直播QoE评估模型架构

### 4.1 移动直播评估概述

移动直播的过程为主播端将产生的音视频流实时地发送到音视频云平台, 观众实时接收主播的音视频流进行观看。移动直播过程中主播和观众间会频繁地进行实时音视频互动, 或者主播之间进行连麦PK, 互动性比较强, 对端到端的时延要求比较高。直播中的用户角色分为主播和观众。

开展移动直播观众端用户体验评价综合了观众端视听体验质量、交互体验质量、呈现体验质量, 是真实反映移动直播观众端用户体验感知的平均主观评分 (MOS) 。

### 4.2 观众端关键体验指标

直播中观众端实时地接收主播端产生的音视频流进行观看。直播中衡量观众端体验的主要指标有视听体验质量、交互体验质量、呈现体验质量, 见图1。视听体验质量包含音频体验质量和视频体验质量。

音频体验质量, 取决于观众端音频的清晰度、流畅度等因素, 可以声道数、码率、编码方式、编码参数、信号层质量等客观指标度量。

视频体验质量, 取决于观众端视频的清晰度、流畅度等因素, 涵盖了客户端视频的分辨率、帧率、码率、编码和终端多个维度的指标, 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 (每英寸的像素数) 而言, PPD (每一度视场角的像素数) 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, 并且PPD可以考虑到观看距离的影响因素。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。

交互体验质量, 取决于观众端端到端的时延 (观众端端到端时延是指数据从主播端采集到观众端播放需要的时间)。观众端交互体验质量也包含观众端在直播间的跳转时延、观众进入直播间的首屏时延等。观众端在直播间的跳转时延是指在主播连麦PK场景下 (如主播A与主播B连麦PK) 主播A直播间的观众从主播A的直播间跳转到主播B的直播间所需的时间。观众进入直播间的首屏时延是指从观众端点击进入主播直播间到直播间首帧画面成功播放显示的时间。

观众端呈现体验质量, 可以音频卡顿率、视频卡顿率、花屏、音画同步时延等指标进行衡量。

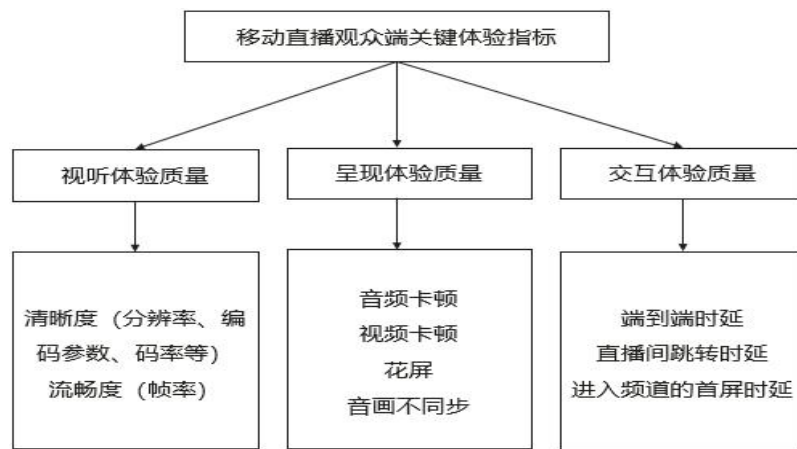


图 1 影响直播观众端体验的关键因素

#### 4.3 移动直播QoE评估模型架构

评估场景分为会话场景和实时场景，会话场景表征观众端用户一次完整的观看行为的体验质量，实时场景表征观众端用户实时观看过程中的体验质量。

移动直播观众端QoE评估模型架构的输入参数、输出质量模块及综合质量如图2所示：

- a) 视听体验质量用于表征直播中观众端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：
  - 1) 视频质量主要与内容清晰度（受分辨率等因素影响）、内容流畅度（以帧率表征）和编码参数（如编码算法、码率）有关。
  - 2) 音频质量主要与声道数（如单声道、双声道（立体声））、编码参数（如编码算法、码率）有关。
- b) 呈现体验质量用于表征受网络传输质量影响的体验损伤，主要与直播过程中因传输带宽、时延抖动和丢包导致的卡顿、花屏和音画不同步等问题有关；
- c) 交互体验质量用于表征直播中观众端交互行为的影响，主要与响应时延有关，如观众端端到端时延、观众在直播间跳转时延、观众进入直播间的首屏时延等。

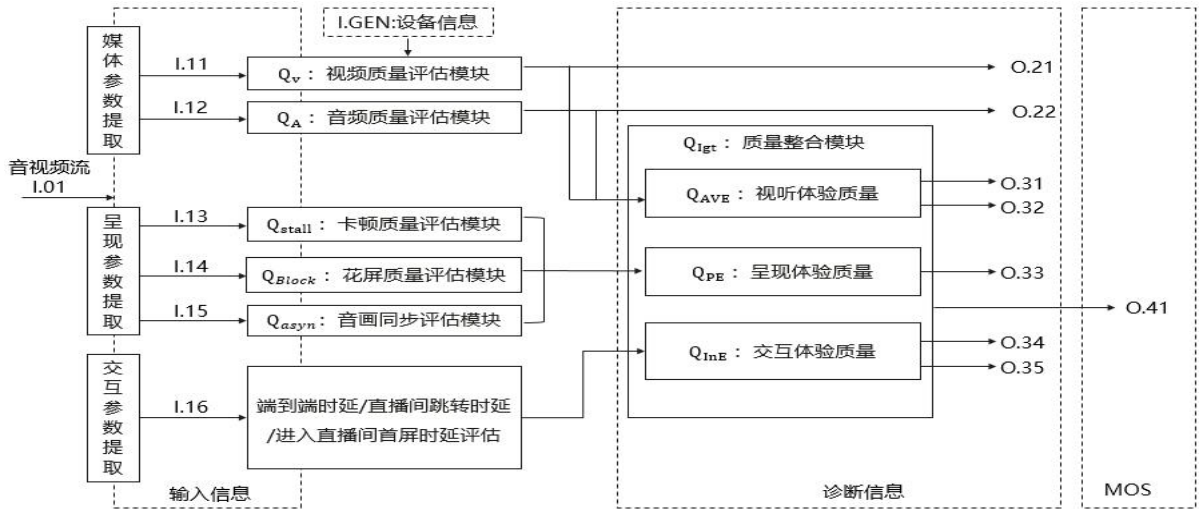


图 2 移动直播观众端 QoE 评估模型架构

根据图2，除Q<sub>igt</sub>（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，这里重点介绍一下比较复杂的Q<sub>igt</sub>模块。Q<sub>igt</sub>模块由如下Q<sub>AVE</sub>、Q<sub>PE</sub>和Q<sub>InE</sub>三个子模块构成。

● Q<sub>AVE</sub>（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）。

Q<sub>AVE</sub>（Audio-Visual Experience Quality, 视听体验质量）子模块其输出是：

- 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）。

关于Q<sub>AVE</sub>子模块请参考本文件6.2.1的相关描述。

● Q<sub>PE</sub>（Presenting Experience Quality, 呈现体验质量）子模块，其输入是：

- 1) 卡顿感知质量（Q<sub>Stall</sub>）；
- 2) 花屏感知质量（Q<sub>Block</sub>）；
- 3) 每输出采样间隔的音视频同步（Q<sub>asyn</sub>）参数。

Q<sub>PE</sub>（Presenting Experience Quality, 呈现体验质量）子模块其输出是：

- 1) 呈现体验质量（O.33）。

关于Q<sub>PE</sub>子模块请参考本文件6.2.2的相关描述。

● Q<sub>InE</sub>（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.16）；

Q<sub>InE</sub>（Interactive Experience Quality, 交互体验质量）子模块其输出是：

- 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Q<sub>InE</sub>子模块请参考本文件6.2.3的相关描述。

移动直播观众端的QoE指标从用户层面出发，可以定义为终端用户观众端对直播业务的综合主观感知，反映移动直播的体验综合评分，用MOS表示，即用户体验综合评分为：视听体验质量（Q<sub>AVE</sub>），呈现体验质量（Q<sub>PE</sub>），交互体验质量（Q<sub>InE</sub>）的函数关系，见公式（1）。

$$MOS = f(Q_{AVE}, Q_{InE}, Q_{PE}) \dots\dots\dots (1)$$



## 5 移动直播质量指标构成

### 5.1 观众端视频质量输入参数

观众端视频质量提取的输入参数I.11（表1中参数1~6）表征直播中观众端视频的质量，I.GEN（表1中参数7~11）则主要表征观众端终端的显示能力（如屏幕分辨率，刷新率等）。观众端视频质量输入参数详见表1。

表 1 观众端视频质量输入参数

序号	参数	英文	数据源	定义/取值
1	视频码率	VideoBitrate, 缩写为 $Br_v$	终端播放器	视频单位时间取样的视频数据量(Kbps), 直播业务观众端视频的接收码率
2	视频帧率	VideoFrameRate, 缩写为 $FR_v$	终端播放器	视频每秒的帧数。直播业务观众端观看直播视频的帧率
3	视频水平分辨率	VideoHorizontalResolution, 缩写为 $R_h$	终端播放器	视频水平方向的像素点个数。直播业务观众端观看直播视频水平分辨率
4	视频垂直分辨率	VideoVerticalResolution, 缩写为 $R_v$	终端播放器	视频垂直方向的像素点个数。直播业务观众端观看直播视频垂直分辨率
5	编解码算法	VideoCodecAlgorithm, 缩写为 $Codec_v$	终端播放器	直播业务, 观众端视频的编解码算法。如H.264、H.265、VP9等
6	信号层参数		终端播放器	块状度 模糊度 对比度 噪点度 色彩度
7	屏幕水平分辨率	ScreenHorizontalResolution, 缩写为 $RS_h$	终端OS	终端屏幕水平方向的像素点个数 直播业务观众端移动设备的屏幕水平分辨率
8	屏幕垂直分辨率	ScreenVerticalResolution, 缩写为 $RS_v$	终端OS	终端屏幕垂直方向的像素点个数 直播业务观众端移动设备的屏幕垂直分辨率
9	屏幕尺寸	ScreenSize, 缩写为 $ScrSize$	终端OS	终端屏幕对角线的长度(inch) 直播业务观众端移动设备的屏幕尺寸

表 1 观众端视频质量输入参数（续）

序号	参数	英文	数据源	定义/取值
10	屏幕刷新率	ScreenRefreshRate, 缩写为ScrRR	终端OS	终端屏幕每秒刷新的次数（Hz） 直播业务观众端移动设备的屏幕刷新率
11	观看距离	DistanceToScreen, 缩写为D2Scr	经验数据	观众端用户观看直播时距离终端屏幕的最佳观看距离（cm）

### 5.2 观众端音频质量输入参数

音频质量提取的输入参数I.12，表征直播中观众端音频的质量，详见表2。

表 2 观众端音频质量输入参数I.12

序号	参数	英文	数据源	定义/取值
1	音频码率	AudioBitrate, 缩写为Br <sub>A</sub>	终端播放器	单位时间传输的音频数据量（Kbps） 直播过程中观众端的音频码率
2	音频声道数	NumberOfChannels, 缩写为NoC <sub>A</sub>	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。 直播过程中观众端音频声道数
3	编解码方式	AudioCodecAlgorithm, 缩写为Codec <sub>A</sub>	终端播放器	音频的编解码算法。涵盖主流的音频编解码算法，如AAC、Opus等。
4	信号层参数		终端播放器	响度 动态范围 左右声道相位差 爆音 静音长度

### 5.3 观众端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、I.14和I.15，其中I.13表征卡顿事件信息，I.14表征花屏事件信息（基于丢包近似评估花屏感知），I.15表征音画同步信息，详见表3。

表 3 观众端呈现体验质量输入参数

序号	参数	英文	数据源	定义/取值
卡顿事件I.13				

1	音频卡顿		终端播放器	在直播过程中，观众端出现音频播放不连续、不流畅等现象。指从主播端到观众端端到端的音频卡顿。可以取音频卡顿的时长。
2	视频卡顿		终端播放器	在直播过程中，观众端出现视频播放不连续、不流畅、画面滞帧等现象。指从主播端到观众端端到端的视频卡顿。可以取视频卡顿的时长。
花屏事件I.14				
3	数据包丢失率	PacketLossRatioInSamplingInterval, 缩写为 $pLR_t$	终端播放器	在每一个输入采样间隔内，视频包丢失百分比 (%)
4	呈现损伤标识	PresentImpairmentFlag, 缩写为 $PslpmFlag$	终端播放器	根据终端对丢包处理是否有丢包掩盖机制，可能存在3种取值：0-仅卡顿（缺省），1-仅花屏，2-同时存在卡顿和花屏
音画同步I.15				
5	音画同步时延	AudioDelayRelativeToVideo, 缩写为 $T_{asyn}$	终端播放器	直播中观众端接收的音频流声音和观看看到的画面的同步时延 (ms)

#### 5.4 观众端交互体验质量输入参数

观众端交互体验质量提取的输入参数I.16，详见表4，表征直播过程中观众端的交互行为的响应时延，涵盖观众端端到端的时延（端到端时延是指数据从主播端采集到观众端播放需要的时间），观众端交互体验质量也包含主播连麦PK场景下观众端在直播间的跳转时延、观众进入直播间的首屏时延。

表 4 交互体验质量输入参数 I.16

序号	参数	英文	数据源	定义/取值
1	交互行为标识	InteractionFlag, 缩写为 $ItractFlag$	终端播放器	根据直播过程中的观众的交互行为，端到端时延、直播间跳转时延、进入直播间的首屏时延等。

表 4 交互体验质量输入参数I.16（续）

序号	参数	英文	数据源	定义/取值
----	----	----	-----	-------

2	观众端端到端时延	$T_{delay}$	终端播放器	从主播端采集到观众端播放需要的时间
3	观众在直播间跳转时延	$T_{jump}$	终端播放器	主播连麦PK场景下，观众从一个主播直播间切换到另外一个主播直播间需要的时间
4	观众进入直播间的首页时延	$T_{firp}$	终端播放器	观众初始进入直播间到首帧出图的时间

### 5.5 观众端QoE模型输出参数

直播观众端QoE评估模型的输出信息，如表5所示。

表 5 直播观众端 QoE 评估模型的输出参数

序号	参数	英文	定义/取值	频次
1	O.21	Visual Quality, 缩写为 $Q_V$	对直播视频的画质评分。取值范围：1-5	直播每一个输出采样间隔（如：每1秒）
2	O.22	Audio Quality, 缩写为 $Q_A$	对直播视频的音质评分。取值范围：1-5	直播每一个输出采样间隔（如：每1秒）
3	O.31	Visual Audio Experience Quality	对直播视频的视听质量的评分。取值范围：1-5	直播每一个输出采样间隔（如：每1秒）
4	O.32	Visual Audio Experience Quality, 缩写为 $Q_{AVE}$	对直播视频的视听质量的评分。取值范围：1-5	每一个直播会话
5	O.33	Presenting Experience Quality, 缩写为 $Q_{PE}$	网络传输质量（如带宽、丢包、时延）对直播视频体验质量影响的评分。取值范围：1-5	每一个直播会话
6	O.34	Interaction Experience Quality	交互响应时延对直播体验质量影响的评分。取值范围：1-5	直播每一个输出采样间隔（如：每1秒）
序号	参数	英文	定义/取值	频次
7	O.35	Interaction Experience Quality, 缩写为 $Q_{InE}$	交互响应时延对直播视频体验质量影响的评分。取值范围：1-5	每一个直播会话

8	$Q_{Stall}$	Stall Quality, 缩写为 $Q_{Stall}$	直播视频的卡顿感知质量。取值范围: 1-5	每一个直播会话
9	$Q_{Block}$	Block Quality, 缩写为 $Q_{Block}$	直播视频的花屏感知质量。取值范围: 1-5	每一个直播会话
10	$Q_{asyn}$	Audio-to-video synchronization Quality, 缩写为 $Q_{asyn}$	直播视频的音画同步质量。取值范围: 1-5	每一个直播会话
11	O.41	live streaming video MOS, 缩写为MOS	对直播视频体验质量的综合评分。取值范围: 1-5	每一个直播会话

表 5 直播观众端QoE评估模型的输出参数 (续)

## 6 移动直播观众端用户体验评估算法

### 6.1 总体算法综述

$MOS = f(Q_{AVE}, Q_{PE}, Q_{InE})$ , 即总体模型为三个模块质量 (视听体验质量 $Q_{AVE}$ , 呈现体验质量 $Q_{PE}$ , 交互体验质量 $Q_{InE}$ ) 的函数关系, 用于评估移动直播场景中观众端的用户体验。

$$MOS = (Q_{AVE} - 1) \times (1 - v_1 \times (5 - Q_{PE}) - v_2 \times (5 - Q_{InE})) + 1$$

..... (2)

式中:

$v_1$  —— 体验质量的动态加权系数;

$v_2$  —— 交互体验质量的动态加权系数;

$Q_{AVE}$  —— 视听体验质量;

$Q_{PE}$  —— 呈现体验质量;

$Q_{InE}$  —— 交互体验质量;

模型通过主观标注的打分数据库进行数据拟合或模型训练获得, 数据库足够覆盖典型使用场景。

### 6.2 四大模块算法综述

### 6.2.1. 视听体验

本文件主要基于附录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—媒体长度（T），单位是秒。

#### 6.2.1.1 视频质量

$Q_V = f(\text{ScreenSize}, \text{ScreenRefreshRate}, \text{ScreenResolutionHorizontal}, \text{ScreenResolutionVertical}, \text{DistanceFromEyeToScreen}, \text{VideoResolutionHorizontal}, \text{VideoResolutionVertical}, \text{VideoBitrate}, \text{VideoFrameRate}, \text{VideoCodec}, \text{Bitdepth}, \text{TransferCharacteristics}, \text{ColorPrimaries})$

..... (3)

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 $Q_V$ （O.21）主要考虑了移动直播观众端视频编解码、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。

$$Q_V(\text{O.21}) = \min(\max(f_1(\text{Quant}) \cdot f_2(\text{FR}) \cdot f_3(\text{PPD}), 1), 5)$$

..... (4)

$$f_1(\text{Quant}) = v_3 + v_4 \times \exp(v_5 \times \text{Quant})$$

..... (5)

$$f_2(\text{FR}) = v_6 + v_7 \times \exp(v_8 \times \text{FR})$$

..... (6)

$$f_3(\text{PPD}) = 1 + v_9 - \frac{v_9}{1 + \left(\frac{\text{PPD}}{v_{10}}\right)^{v_{11}}}$$

..... (7)

$$\text{BPP} = \frac{Br_v}{(R_h \times R_v) \times FR_V}$$

..... (8)

$$Quant = \max(v_{12} + v_{13} \times \ln(v_{14} + \ln(Br_V) + \ln(Br_V \times BPP + v_{15})), 0) \quad \dots\dots\dots (9)$$

$$PPD = RoundUp \left( \frac{\min(R_h, RS_h)}{\left( \frac{\left( \frac{ScrSize}{\sqrt{1 + (RS_v/RS_h)^2}} \right) / 2}{(D2Scr/2.54)} \right) \times (180/\pi) \times 2} \right), 0 \quad \dots\dots\dots (10)$$

$$FR = \begin{cases} FR_V & FR_V \leq RR \\ RR & FR_V > RR \end{cases} \quad \dots\dots\dots (11)$$

公式（5）~（9）中根据不同的视频编码算法对应不同的模型系数 $v_3 \sim v_{15}$ 取值。

PPD计算方法的推导过程，再参考附录B。

当视频和屏幕水平分辨率（即像素个数）不同时，公式（10）中取二者的较小值，以便反映用户实际看到的水平像素个数。

上述公式（10）PPD计算是以手机横屏观看直播为例，对于竖屏观看，PPD计算方法需要相应进行变更，上述公式中 $RS_v$ 为 $RS_h$ ， $RS_h$ 为 $RS_v$ ， $\min(R_h, RS_h)$ 为 $\min(R_v, RS_v)$ ，公式如下所示：

$$PPD = RoundUp \left( \frac{\min(R_v, RS_v)}{\left( \frac{ScrSize}{\sqrt{1 + (RS_k/RS_v)^2}} \right)^2 \times (180/\pi) \times 2} \right), 0$$

..... (12)

6.2.1.2 音频质量

$$Q_A = f(Br_A, NoC_A, Codec_A)$$

..... (13)

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  $Q_A$  (O.22) 评估模型为：

$$Q_A(O.22) = f_4(Br_A)$$

..... (14)

$$f_4(Br_A) = v_{19} \times \left( 1 + v_{16} - \frac{v_{16}}{1 + \left( \frac{Br_A}{v_{17}} \right)^{v_{18}}} \right) + v_{20}$$

..... (15)

公式 (15) 中根据不同的音频编码算法和声道数对应不同的模型系数  $v_{16} \sim v_{20}$  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  $Q_V$  (O.21)，音频质量  $Q_A$  (O.22)，其评估模型为：



$$0.31 = v_{21} \times Q_V + v_{22} \times Q_A + v_{23} \times Q_V \times Q_A + v_{24} \dots \dots \dots (16)$$

式中： $v_{21} \sim v_{24}$ 是算法系数。

最后，基于会话的视听体验质量（O.32，即 $Q_{AVE}$ ）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。

$$0.32_t = v_{25} \times 0.32_{t-1} + (1 - v_{25}) \times 0.31_t \dots \dots \dots (17)$$

式中： $v_{25}$ 是算法系数。

## 6.2.2. 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2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。

#### a) TotalStallLen的计算

为了计算**TotalStallLen**，应首先为每个单独的卡顿事件*i*分配一个权重 $w\_stall_i$ ，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。

$$w\_stall_i = v_{26} + v_{27} \times e^{(-v_{28} \times StallPosFromEnd_i)}, i \in [1, \dots, StallNum] \dots \dots \dots (18)$$

$$StallPosFromEnd_i = T - StartTime_i, i \in [1, \dots, StallNum] \dots \dots \dots (19)$$

$$TotalStallLen = \sum_{i=1}^{StallNum} w\_stall_i \times StallLen_i \dots \dots \dots (20)$$

注：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。

卡顿感知质量（ $Q_{Stall}$ ）评估模型为：

$$Q_{Stall} = f_5(StallNum, TotalStallLen, T) \dots \dots \dots (21)$$

$$f_5(StallNum, TotalStallLen, T) = \min(v_{31} + v_{32} \times \left( e^{-\frac{StallNum}{v_{29}}} \cdot e^{-\frac{TotalStallLen/T}{v_{30}}} \right), 5)$$

..... (22)

6.2.2.2 花屏感知质量评估

在呈现参数提取模块中输出（花屏相关的）参数是 $PLR$ ，系通过对 $pPLR_t$ 进行时变加权求和得到。其中，权重依赖于它在媒体会话中的位置，根据遗忘曲线进行衰减。

$$PLR = \frac{\sum_t (w\_block(T-t) \times pLR_t)}{\sum_t w\_block(T-t)}$$

..... (23)

$$w\_block(T-t) = \begin{cases} -v_{33} \times \ln(T-t) + v_{34} & T > t \\ 1.4512 & T = t \end{cases}$$

..... (24)

花屏感知质量（ $Q_{Block}$ ）评估模型为：

$$Q_{Block} = f_6(PLR)$$

..... (25)

$$f_6(PLR) = \min(\max(v_{35} \times e^{(v_{36} \times PLR/100)} + v_{37}, 1), 5)$$

..... (26)

式中： $v_{33} \sim v_{37}$ 是算法系数。

6.2.2.3 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，呈现体验损伤可能同时包括卡顿、花屏（此时同时存在 $Q_{Stall}$ 和 $Q_{Block}$ 子体验项），也可能仅包括花屏；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。

此外，考虑到音画同步 $Q_{asyn}$ 的影响，故呈现体验质量（O.33）评估模型为：

$$O.33 = \begin{cases} Q_{Stall} \times Q_{asyn}, \text{基于有丢包掩盖机制（卡顿）} \\ Q_{Block} \times Q_{asyn}, \text{基于无丢包掩盖机制（花屏）} \\ ((Q_{Stall} - 1) \times (1 - v_{38} \times (5 - Q_{Block})) + 1) \times Q_{asyn}, \text{基于无丢包掩盖机制（卡顿 + 花屏）} \end{cases}$$

..... (27)

$$Q_{asyn} = \min(v_{39} \times \exp(v_{40} \times T_{asyn}) + v_{41} \times \exp(v_{42} \times T_{asyn}), 5)$$

式中： $v_{38} \sim v_{42}$ 是算法系数。 ..... (28)

6.2.3. 交互体验

对于移动直播，基于输出采样间隔的交互体验质量 (O.34) 的公式为：

$$O.34 = 5 - \text{各种交互响应时延损伤量} = 5 - \sum_{i=0}^2 DMOS_{om}(ItractFlag) \dots\dots\dots (29)$$

式中： $DMOS_{om}(ItractFlag)$ ——交互响应时延导致的交互一致性下降量。

其中，移动直播 $DMOS_{om}(ItractFlag)$ 与交互响应时延的关系可以表示为：

$$DMOS_{om}(ItractFlag) = \begin{cases} \max(\min(v_{43} \times \ln(T_{delay} + v_{44}) + v_{45}, 4), 0), & \text{包含端到端时延} \\ \max(\min(v_{46} \times \ln(T_{jump} + v_{47}) + v_{48}, 4), 0), & \text{包含直播间跳转时延} \\ \max(\min(v_{49} \times \ln(T_{firp} + v_{50}) + v_{51}, 4), 0), & \text{包含首屏时延} \\ 0, & \text{else} \end{cases} \dots\dots\dots (30)$$

式中： $v_{43} \sim v_{51}$ 是算法系数。

最后，基于会话的交互体验质量 (O.35) 评估模型为：

$$O.35 = \frac{\sum_t (w_{InE}(T-t) \times Q_{MA\_InE}(t))}{\sum_t w_{InE}(T-t)} \dots\dots\dots (31)$$

$$Q_{MA\_InE}(t) = \frac{\sum_{j=0}^{N-1} O.34(t+j)}{N} \dots\dots\dots (32)$$

$$w_{InE}(T-t) = 1 / \left( \frac{v_{52} - v_{58}}{1 + \left(\frac{T-t}{v_{54}}\right)^{v_{53}}} + v_{55} \right) \dots\dots\dots (33)$$

式中： $v_{52} \sim v_{55}$ 是算法系数。

## 附录 A

(资料性)

## ITU-T P. 1203 工作模式

根据ITU-T P. 1203规范，本建议标准的QV/QA评估算法基于Mode 0方式构建：

- a) 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。
- b) 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））；
- c) 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。
- d) 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。

不同操作模式之间的关系图A. 1所示。



图 A. 1 ITU-T P. 1203.1 操作模式

附录 B  
 (资料性)  
 关于 PPD 计算方法

如图 B.1 所示, 以手机屏幕横屏为例, PPD 计算公式如下:

$$PPD = \frac{R_h}{2 \times \tan^{-1} \left( \frac{W/2}{D/2.54} \right) \times 180/\pi} \dots\dots\dots (B.1)$$

$$\frac{R_h}{R_v} = \frac{W}{H} \dots\dots\dots (B.2)$$

$$S = \sqrt{W^2 + H^2} \dots\dots\dots (B.3)$$

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 (B.2) 和 (B.3) 可知,

$$w = \frac{S}{\sqrt{1 + \left( \frac{R_v}{R_h} \right)^2}} \dots\dots\dots (B.4)$$

将公式 (B.4) 代入公式 (B.1) 可得:

$$PPD = \frac{R_h}{2 \times \tan^{-1} \left( \frac{\frac{S}{\sqrt{1 + \left( \frac{R_v}{R_h} \right)^2}}}{D/2.54} \right) \times 180/\pi} \dots\dots\dots (B.5)$$

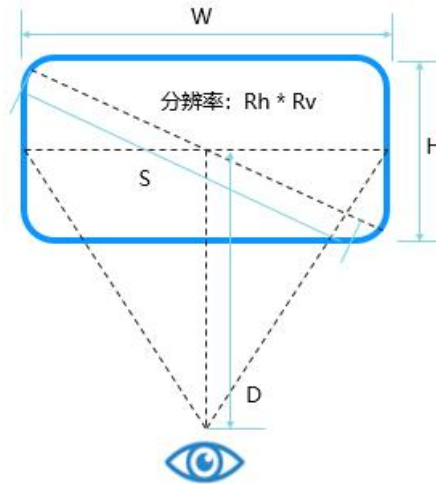
同理可知, 纵向 PPD 公式可得:

$$PPD = \frac{R_v}{2 \times \tan^{-1} \left( \frac{\frac{S}{\sqrt{1 + \left(\frac{R_h}{R_v}\right)^2}}}{D/2.54} \right)} \times 180/\pi$$

..... (B.6)

至此，公式 (B.5) 即为当横屏观看时手机屏幕角分辨率PPD的最终计算公式，公式 (B.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B. 1 PPD计算示意图

参考文献

- [1] ITU-T P. 1203 (2017), Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport. 瑞士: 国际电信联盟(ITU), 2017
-