



专题：视频技术的理论与实践

视频用户体验理论与实践

贺甜甜^{1,2}, 刘彦凯^{1,2}, 宋利^{1,2}

(1. 上海交通大学图像通信与网络工程研究所, 上海 200240;
2. 未来媒体网络协同创新中心, 上海 200240)

摘要: 随着视频服务逐渐成为人们获取信息的主要途径之一, 消费者对观看体验的要求不断提高, 视频用户体验质量已经成为视频服务的主要竞争因素。首先对用户体验质量理论进行了系统的阐述, 指出了用户体验质量与服务质量之间的差别和联系, 同时给出了用户体验评价方法的主要步骤和视频质量评价的具体方法。进一步地, 对我国现阶段视频用户体验评价研究及标准化的进展进行详细介绍, 概括了目前我国视频服务的用户体验现状并给出了相关的改进方向意见。

关键词: 用户体验质量; 视频服务; 客观质量评价

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017247

Theory and practice of quality of experience in video services

HE Tiantian^{1,2}, LIU Yankai^{1,2}, SONG Li^{1,2}

1. Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
2. Future Medianet Innovation Center, Shanghai 200240, China

Abstract: As videos becoming the major sources of acquiring information in current era, consumers have raised their expectations for video quality experience, which has become the most critical competitiveness for video service providers. A thorough review of quality of experience (QoE) was presented, and then detailed analysis of video QoE and relevant no reference assessment methods were provided. Further, the standardization process of video QoE in China and the evaluation results of Chinese video services quality were introduced. Last but not the least, the characteristics of Chinese video services was summarized and advices of improvement were presented.

Key words: quality of experience, video service, no reference video quality assessment

1 QoE 理论体系

ITU-T P.10/G.100 (Amd5) 标准^[1]将用户体验质量 (quality of experience, QoE) 定义为“用户对某项应用或者服务的满意程度, 该满意程度来

自于用户对于此项应用或服务功能或质量的期望的实现程度, 与该用户的个人喜好和感受相关”。简单来说, 就是终端用户对应用或者服务整体的主观可接受程度, 其影响因素主要可以分为 3 个层面: 服务层面、环境层面和用户层面, 如图 1

收稿日期: 2017-07-10; 修回日期: 2017-08-07

基金项目: 国家自然科学基金资助项目 (No.61671296, No.61527804)

Foundation Items: The National Natural Science Foundation of China (No.61671296, No.61527804)



所示。服务层面的影响因素又包括应用层、传输层和服务层的参数；环境层面的影响因素包括自然环境（如光照条件、噪声条件、环境的固定和移动）、社会与人文环境（如社会观念、文化规范）以及服务运行环境（包括软硬件环境等）；用户层面的因素包括用户的期望、体验经历、身心状态和自身背景（如年龄、性别和受教育程度等）。

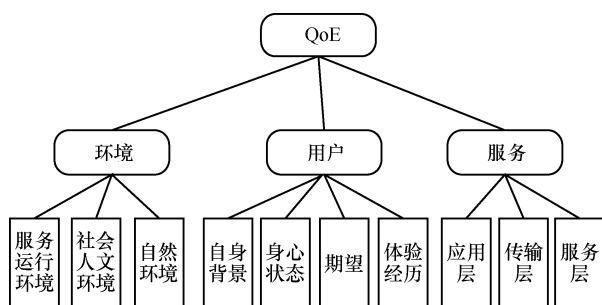


图1 QoE的影响因素

QoE 概念是在用户服务质量（quality of service, QoS）的基础上发展起来的。ITU-T Rec E.800 标准^[2]将 QoS 定义为“某项服务满足使用该服务的用户的明示或暗示需求的综合能力”，指一个网络能够利用各种基础技术，为指定的网络通信提供更好的服务能力。从定义上来看，QoE 和 QoS 的概念较为抽象，具体到网络服务质量层面来说，QoS 可以狭义地理解为基于底层分组数据传输的关键性能指标（key performance indicator, KPI），而 QoE 是建立在由 KPI 生成的关键质量指标（key quality indicator, KQI）之上的，并且引入了与用户行为相关的数据。具体的 QoE 与 QoS 体系关系如图 2 所示。

视频服务 QoE 评价体系如图 3 所示，视频服务 KPI 一般指基于网元设备的关键指标，主要包括终端性能、网络传输质量、视频传输质量、音频体验质量、媒体的封装质量和视频的编码参数。KQI 从业务和应用层面出发反映业务使用体验，例如视频源质量，包括视频源本身的编码格式、视频清晰度和媒体封装等；观看体验质量，即视

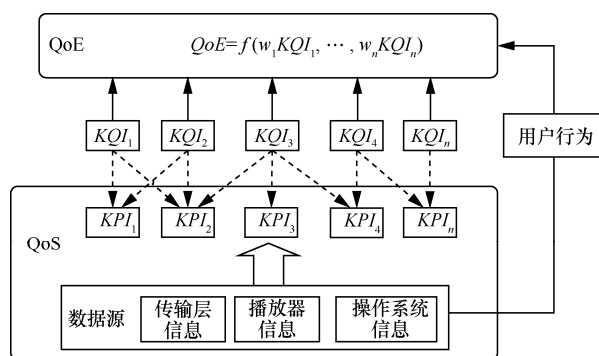


图2 QoE与QoS体系关系

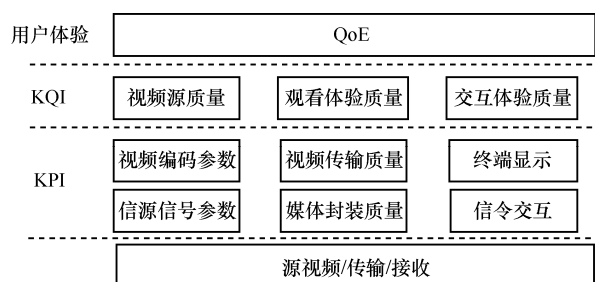


图3 视频服务 QoE 评价体系

频在播放过程中是否有出现缓冲卡顿、花屏和马赛克的情况；交互体验质量，是指视频在观看过程中业务交互的情况，包括直播业务的交互成功率以及交互时延等。KQI 一般通过 KPI 处理得到。QoE 指标从用户层面出发，可以定义为终端用户对视频业务的总体主观感知。客观评估 QoE 通常通过一组 KQI/KPI 进行映射^[3]，通过可靠性和舒适性两个主要的方面对视频业务的用户体验进行评价，并得到用户观看视频过程中体验的综合评分。

值得注意的是，QoE 引入的用户主观感受本质上需要通过主观评价方法加以测定，然而因为主观评价方法对于时间和参与测试人数有着一定的要求，较为耗时耗力，目前主观测试结果主要用于对客观模型输出进行校验和训练，QoE 问题的研究方向在于如何设计一套有效的客观评价办法，使得评价系统输出尽量接近人的主观感受，这样的客观评价算法在实际应用中才具有广泛的使用价值，而视频服务即目前对于客观评价算法有迫切需求的领域之一。

2 视频服务与用户体验评价

随着视频编码、显示技术的发展,国内外视频服务都进入了高速发展期,对宽带建设和网络提速起到显著的拉动作用,带动了视频通信、视频消费和视频监控等业务的发展,共同构建了大视频业务体系并成为支撑智慧城市、先进制造等工作的重要基础。随着技术的进步和带宽的提升,市场提供的视频服务的主流分辨率从过去以标清为主演变到目前以高清为主,并将进一步过渡到以超高清 4K 为主,未来还会向 8K、AR/VR 等更高水平发展。视频服务供应商纷纷采取各类措施来提升视频服务的用户体验。因此用户体验质量与视频服务各环节之间关系和实际检测方法具有十分重要的研究意义。本文将从用户体验质量理论出发,介绍视频质量评价的具体方法,并结合我国目前视频用户体验标准化工作对视频用户体验理论与实践做出详细介绍。

2.1 视频服务体验主观评价方法

如上文所述,QoE 模型的最终输出是对用户的主观感受进行评价,因此主观评价方法是 QoE 研究不可缺少的一环。ITU 对于不同的应用场景设计了不同的主观测试标准,对于视频服务应用最广泛的有 ITU-T P.910^[4]以及 ITU-R BT.500^[5]标准,其中前者侧重视频会议、视频通话场景,后者侧重家庭电视环境,在基本方法上二者有交叠,主要差异在于应用场景不同。两个标准被包含的主要测试方法有:单刺激绝对分级(absolute category rating, ACR)法、成对比较(pair comparison, PC)法、双刺激损伤分级(double stimulus impairment scale, DSIS)法、双刺激损伤连续评价(double stimulus continuous quality scale, DSCQS)法。每类测试方法在开始前均需要对被测试者讲解流程,BT.500 标准建每次测试至少需要 15 名被测人员以达到统计有效。在获得每个被测者的打分后,需要对其有效性进行筛查,其根本原理

是检查一个测试者的数据能否落在由其余所有测试者的数据计算出的置信区间内,详细筛选算法请参见 BT.500 标准^[4]。通过数据筛查后的打分平均处理即可得到每个测试序列的主观得分。下面对不同的主观打分方法进行简要介绍。

ACR 方法中被测者每次只观看一段视频并对观看的每段视频分别打分,每一位测试者采用不同的随机播放顺序观看不同的视频。通常打分采用五级离散质量量表,该量表中分值为 MOS (mean opinion score) 分值,表示是对“视频质量”进行打分,认为该视频质量越高,则分数越高。即优秀=5,良好=4,一般=3,较差=2,极差=1。有时该方法也会采用九级、十一级离散量表(最高分值为 9 分或 11 分)或 100 级连续量表(为后续数据处理方便)等。

PC 方法中一对视频同时呈现给被测者且组合排序随机,被测者须使用七级离散量表判别第二段视频与第一段相比的质量优劣,即极好=3,好=2,较好=1,无差别=0,较差=1,差=-2,极差=-2,也可以采用两级量表,判定第二段视频好或差于第一段视频。

DSIS 方法类似 PC 方法,但强调测试素材既包含测试序列,也包含相应的原始序列。测试时,测试者先观看源参考视频,再观看测试视频。测试者评分的对象,是后一个视频相较于前一个视频的差别,即测试视频相对源参考视频的损伤度,即损伤不可察觉=5,轻微可察觉但可接受=4,轻微不可接受=3,不可接受=2,极端不可接受=1。

DSCQS 方法用于比较一个视频源与它的一个损伤版本之间的质量关系,二者随机排列并且呈现给被测者两次,观看完两轮后被试者使用 ACR 方法的评价量表分别对两个视频独立打分。DSCQS 方法的优点是可以比较出测试视频与源参考视频的细微差别,它的缺点在于测试时间几乎是单刺激法的两倍。



2.2 KPI 与视频服务体验

飞速发展的网络技术与多媒体技术相结合,出现了多种多样的网络视频服务形式。网络视频已经成为主流的视频服务方式之一,其形成过程如图 4 所示。原始的视频数据通过编码器进行编码压缩并封装,然后通过网络进行传输,接收端的解码器收到网络视频流后对数据进行解分组并解码,最后将解码后的视频序列在终端设备上播放显示。



图 4 网络视频形成过程

在上述过程中,网络传输性能和视频压缩性能成为了影响视频质量和用户体验的关键性能指标。保证网络视频质量成为网络视频服务的关键,视频质量评价方法的研究也受到了广泛的关注。

2.2.1 基于网络传输性能的视频质量评价

网络传输质量的评价是利用可量化的评价指标对视频质量受网络损伤的影响程度进行合理有效的评定。由于网络的传输环境错综复杂且具有时变性,因此存在多种多样的因素能影响到传输视频的质量,其中影响较大的因素包括分组丢失、端到端时延、时延抖动和带宽等方面^[6]。同时由于网络性能中的抖动、时延所导致的主要后果就是产生数据分组丢失,即视频流信息丢失,数据分组丢失对视频质量有直接的影响,包含语义信息(例如运动向量、量化参数等)的数据丢失会对与之相关的视频内容产生影响,而与系统信息和分组头信息有关的语法信息的丢失可能导致整块数据无法解码的严重后果,甚至会导致一个或更多片的丢失。此外,由于编码/解码的特性,数据分组丢失引起的误差可能会在时间上或/和空间上扩散。采用有差别、可变长的编码的方式会导致数据分组丢失在单个帧(任何类型的帧)

中延续,直到遇到下一个重新同步点,引起分组丢失误差在空间上的扩散。时间上的误差扩散是由于帧间预测产生的,主要指在参考帧(I 帧或 P 帧)中发生分组丢失时,以该帧为参考的帧会受到影响。无线传输的引入使得移动视频服务质量受到更多的传输差错影响^[7]。与网络传输性能有关的参数通常包含在分组头中的信息中。

网络传输性能对于视频服务质量的影响已经在国际上引起广泛关注,一系列与之相关的标准和模型应运而生。ITU-T G.1000 系列标准主要关注通用的和与用户相关的多媒体服务质量和性能。ITU-T G.1050^[8]提供了 IP 网络上多媒体传输性能的网络评估模型,该模型着重对分组时延变化和分组丢失特性进行分析,可以用于评估来自任何类型网络设备的 IP 媒体数据流。ITU-T G.1080^[9]对 IPTV 的用户体验质量和影响网络传输和应用层行为的信息进行了定义。ITU-T G.1081^[10]定义了 IPTV 服务中的性能监测点。ITU-T G.1082^[11]提供了基于实时测量结果,用于提高 IPTV 性能顽健性。此外,ITU-T P.1200 系列标准提供了流媒体质量评估模型,其中部分模型对于基于网络传输的视频质量进行了评估,例如 ITU-T P.1201.2^[12]。

2.2.2 基于视频压缩性能的视频质量评价

压缩编码对于视频的传输来说是必不可少的,视频的压缩编码是由视频编码器来实现的。视频编码器的作用是将输入的原始像素数据编码成码流数据并输出,其最终的目的在于实现视频的数据量的压缩。视频编码过程是一个有损压缩的过程,对视频的质量产生决定性的影响。视频编码器输出的视频质量与编码标准、编码器种类、视频内容、视频码率(比特率)和其他编码参数(如帧率、分辨率等)有密切的关系:通常情况下,按照不同编码标准(如 H.264、HEVC 等)对视频进行压缩获得的视频质量不同。不同编码器的性能存在差异——即使采用相同的编码标准,不同

编码器之间性能也有区别。同时,不同的视频内容在相同的编码情况下会有不同的质量。对同一编码器压缩的视频,视频质量和码率大体上成正比的关系。此外,视频编码的过程并不是一个标准化定义的过程,用户可以对编码参数进行自主设置,只要编码产生的码流符合解码规范即可。因此,采用不同的编码参数设置输出的视频的质量往往可能相差很多。

基于视频压缩性能的视频质量评价就是根据视频压缩过程中的关键影响因素对视频质量的压缩损伤进行评价。现有的一些主流的视频质量评价模型,包括国际电信联盟(ITU)提出的标准化视频质量评价模型和其他相关学者提出的模型,针对不同的应用场景对视频压缩质量进行了评估。这些模型引用了不同的参数,见表1,使用相同的视频序列集对这些模型的性能进行验证,根据均方根误差(RMSE)、样本离群率(OR)和皮尔森相关系数(PCC)3个统计指标可以看出,各个模型在实际的应用中对视频质量评价的准确性、一致性和线性相关性等方面存在差异^[21]。

在现实的应用,如IP电视(IPTV)或OTT等视频服务中,视频编码过程使用的参数通常是

不规范的,一个视频序列的GOP长度和结构通常是变化的。因此,基于GOP的模型,如上述P.1201.2的性能可能受影响,而不是基于GOP的模型具有更好的性能。其中,新提出的模型uVESModel和P1203.1 Mode 3在视频质量评价过程中考虑了终端播放参数的影响,在实际应用中表现出了更优的性能,这为未来提出性能更优、泛化能力更强的视频质量评价模型指明了方向。面临更加复杂的应用环境和更多样的视频服务类型,视频质量评价模型需要做出适当的调整(例如引入更多的有效参数)以跟上实际应用变化和发展的步伐。

2.3 KQI与视频服务体验

2.3.1 超高清视频服务

2012年国际电信联盟无线电通信部门(ITU-R)颁布了面向新一代超高清视频(ultra-high definition, UHD)制作与显示系统的BT.2020标准^[22],重新定义了电视广播与消费电子领域关于超高清视频显示的各项参数指标,促进4K超高清家用显示设备进一步走向规范化。其中最为关键的是,BT.2020标准指出超高清视频显示系统包括4K与8K两个阶段,其中4K的像素分辨率为

表1 模型比较

模型	参数				性能		
	码率	帧率	视频内容	终端显示参数	PCC	RMSE	OR
ITU-T G1070 ^[13]	是	是	否	否	0.842 2	0.269 4	45.22%
ITU-T P.1201.1 ^[14]	是	是	是—内容复杂度	否	0.741 1	0.350 7	65.22%
ITU-T P.1201.2 ^[12]	是	是	是—内容复杂度(场景检测)	否	0.765 1	0.906 7	73.91%
ITU-T P.1203.1 Mode 3 ^[15]	是	是	是—QP, 宏块跳过程	是—设备类型, 显示像素	0.890 1	0.251 3	43.48%
Yamagishi ^[16]	是	是	否	否	0.835 1	0.248 8	50.43%
Rises ^[17]	是	是	是—内容类型	否	0.824 8	0.258 3	53.91%
Joskowicz ^[18]	是	是	是—SAD	否	0.808 1	0.272 5	57.39%
Takagi ^[19]	是	是	是—内容复杂度	否	0.881 3	0.273 2	36.52%
uVES Mode 1 ^[20]	是	是	是—内容复杂度(运动矢量、QP、宏块跳过程、帧闪烁)	是—屏幕大小	0.887 5	0.283 5	39.13%

注:“是”表示模型引用了该参数,“否”表示没有引用,后3列加粗显示各项指标中最优的两项。



3 840 dpi×2 160 dpi, 而 8K 则为 7 680 dpi×4 320 dpi。之所以超高清视频显示系统会有两个阶段, 实际上是因为全球各个地区超高清视频显示系统发展差异性所造成的, 例如在电视广播领域技术领先的日本就直接发展 8K 电视广播技术, 避免由 4K 过渡到 8K 可能出现的技术性障碍。而在世界的其他地区, 多数还是以 4K 技术作为下一代的电视广播发展标准。超高清视频服务代表着新一轮视频技术革新已经历将分辨率提升到 4K 的第一阶段发展, 到第二阶段包含分辨 (3 840 dpi×2 160 dpi)、帧率 (50/60/120 帧/s)、对比度 (high dynamic range, HDR)、广色域 (wide color gamut, WCG)、位深 (10 bit、12 bit) 5 方面全方位的提升。

此外, 下一代视频编码标准, 5G 通信技术等保证了超高清视频服务的落地, 由此为用户带来了更为沉浸式的体验, 同时也对 QoE 模型的发展提出了更高的要求。最为显著的是, 目前视频服务的质量评价应由传统的误差检测过渡到性能检测, 即从区别“好”与“坏”, 过渡到区别“好”与“更好”。从用户角度衡量视频服务质量需要进一步考虑人眼视觉系统对画面的感知, 设计基于图像层信息的关键质量指标, 以此提升 QoE 体系。

2.3.2 3C 概念与视频质量评价

作为工业界高端视频生产和研究行业组织, Ultra HD Forum 在 2016 年提出了基于关键质量指

标的超高清图像质量“3C”KQI 概念, 即颜色度、对比度、清晰度 (colorfulness, contrast, clarity), 颜色度侧重图像还原真实世界色彩的能力, 对比度代表画面呈现亮区和暗区细节的能力, 清晰度强调图像没有细节质量的损伤, 具体可细化为模糊度、块效应程度、噪点度, 即“3C”概念对应着 5 类人眼视觉相关评价指标: 模糊度、块效应程度、噪点度、颜色丰富度与噪点度。

视频质量评价是图像识别领域的一个研究方向, 包含主观质量评价与客观质量评价两部分, 而如第 2.1 节所述, 主观质量评价虽然可以真实反映用户对于服务的感受, 但其需要大量人员参加测试并且进行多轮以实现统计显著性, 主观评价需要极高的人力物力资本, 故多数用于模型验证等, 不适用于实际应用环境。客观评价的目的即通过提取视频信息, 设计相关评价算法, 通过模型拟合对视频进行客观评价以预测出主观感受分数, 是实际应用中更具价值的评价方案。根据评价算法是否需要无质量损失的参考源, 客观质量评价可以分为无参考、半参考、全参考 3 类, 其中以无参考客观质量评价算法的应用场景更为广泛。无参考客观评价根据模型输入的不同可分为传输层模型、码流模型及混合模型, 而“3C”概念即对应需要输入视频解码后图像层信息的混合客观评价模型, 如图 5 所示。

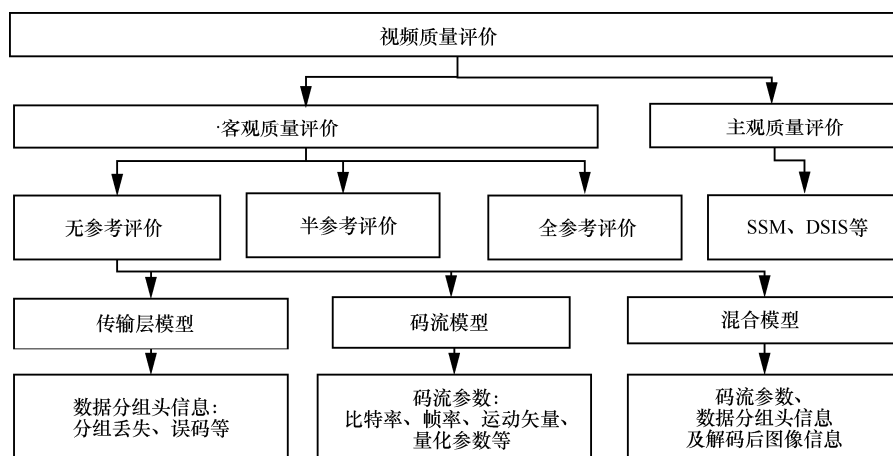


图 5 视频质量评价框架

3 我国视频用户体验标准化进程

由中国信息通信研究院、国家新闻出版广电总局广播电视规划院牵头，中国电信、华为公司、上海交通大学等深度参与的中国视频服务用户体验标准工作组，与2016年9月发布了国内首个《视频服务用户体验评估标准（1.0）》（以下简称《评估标准》）（如图6所示），此外工作组还组织相关单位在国内部分地区开展了针对IPTV视频业务和互联网视频业务的视频服务用户体验评测工作，以验证标准的实用性和有效性。

3.1 视频服务用户体验评估标准

依据图1中的QoE影响因素，由于用户层面和环境层面变化多种多样，对某个视频业务整体的用户体验评估主要从服务层面进行。服务层面影响因素可分为应用层、传输层和服务层。视频体验的应用层面主要为音视频质量，包括音视频的信源质量和编码质量等，传输层反映视频的传输状况，包括分组丢失、抖动和时延等，表现为用户的观看体验。服务层主要为用户的交互体验，如初始加载、快进快退等。《评估标准》将视频终端用户对视频业务的总体主观体验分数 U_{ves} 定义为 $U_{ves}=f(Q_s, Q_a, Q_i, Q_v)$ ，即用户体验综合评分为视频体验质量(Q_s)、音频体验质量(Q_a)、交互体验质量(Q_i)、观看体验质量(Q_v)的函数关系。标准区分了不同的应用场景，包括点播、直播以及视频监控等，其余场景正在逐步扩充中。

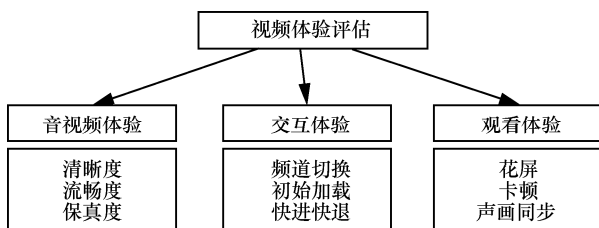


图6 《视频服务用户体验评估标准（1.0）》评价结构

3.1.1 视频体验质量

视频体验质量按照其处理信号层级主要包含

Model0、Model1、Model2 三层模型。其中 Model0 显式包含比特率 ($BitRate$)、视频显示屏幕尺寸 ($ScreenSize$)、电视机分辨率 (PPI) 和编码器类型 ($CodecType$) 变量。Model1 需要从编码数据分组及比特流中采集视频帧关键编码信息，包含视频分辨率 ($Resolution$) 及帧率 ($FrameRate$)、每一帧的编码类型 ($FrameType$) 及帧大小 ($BytesPerFrame$)、编码量化参数 (QP)、运动矢量信息 (MV) 以及每一帧内编码单元跳过比例 ($SkipRatio$)。Model2 需要从播放器连续采集视频帧的图像层关键质量信息，包含模糊度 ($Blurriness$)、块效应 ($Blockiness$)、对比度 ($Contrast$)、噪点度 ($Noise$) 和颜色度 ($Color$)。三层模型所需的输入信息按获取难度递增，且评价精细化程度递增，采用分层形式有利于根据实际应用条件中的灵活调整（如不具备完全解码能力时可仅采用基于码流的前两层模型）：

$$Q_s = f(Q_{sModel0}, Q_{sModel1}, Q_{sModel2})$$

$$Q_{sModel0} = f(PPI, ScreenSize, CodecType, BitRate)$$

$$Q_{sModel1} = f(Resolution, FrameType, FrameRate, BytesPerFrame, QP, MV, SkipRatio)$$

$$Q_{sModel2} = f(Blurriness, Blockiness, Contrast, Noise, Color) \quad (1)$$

3.1.2 音频体验质量

音频体验质量同样依据其处理信号层级分为三级。其中 Model0 包含码率和编码类型变量，Model1 包含码率、编码类型 (E-AC3、AAC、MPEG1 Layer2、Layer3)、音频数据每秒采样点数 ($SampleRate$)、声道数 ($NumberOfChannels$) 和 AAC 编码中的尺度因子 ($ScaleFactor$)，Model2 需要从原始音频或播放器连续采集音频，获得响度 ($Loudness$)、动态范围 ($Dynamic Range$) 和左右声道相位差 ($Skewing$) 等关键质量信息：

$$Q_a = f(Q_{aModel0}, Q_{aModel1}, Q_{aModel2})$$

$$Q_{aModel0} = f(BitRate, CodecType)$$

$$Q_{aModel1} = f(BitRate, SampleRate, CodecType,$$



NumberOfChannels, ScaleFactor)

$$Q_{\text{Model2}} = f(\text{Loudness}, \text{DynamicRange}, \text{Skewing}, \text{Sonicboom}, \text{MuteLength}) \quad (2)$$

3.1.3 交互体验质量

交互体验是指用户在视频业务使用过程中业务操作的便捷性和效率,包括直播、点播等业务操作的成功率以及交互时延等指标。用户的交互体验主要受视频系统的响应速度影响;对于不同视频业务,用户关注的交互体验具体指标有所不同。

直播交互体验受传统电视使用习惯的影响,用户对换台时延的期望值较高;因此对直播业务操作时,频道切换的时延是影响交互体验最重要的因素;频道切换时延 t_{loading} 越长,用户体验越差。由于点播业务在传统操作上的反应就比较慢,用户对其交互体验的初始缓冲时延 t_{zapping} 容忍度比较高;业界普遍采用“2 s 定律”来衡量;即大部分用户对在点播操作后的 2 s 内能完成视频初始加载的服务是可以接受的;超过 2 s 之后,点播操作后的视频初始加载时间每增加 1 s 约有 10% 的用户会选择放弃;当交互时延达到 10 s 时大多数用户会选择放弃。此外对于点播和直播应用,均需要考虑当前已播放时长 t 和最大遗忘时 T (默认为系统参数,用以构建遗忘曲线刻画用户对时延的忍耐性):

$$\begin{aligned} Q_{\text{IVOD}} &= f(t_{\text{zapping}}, t, T) \\ Q_{\text{IVOD}} &= f(t_{\text{zapping}}, t, T) \end{aligned} \quad (3)$$

3.1.4 观看体验质量

观看体验即视频在播放过程中是否有出现视频图像不连续,图像出现异常等质量劣化的情况,包括花屏、马赛克、卡顿、声画不同步等。造成这些情况的原因主要是网络和业务平台的能力是否满足要求,是否能协调一致。比如视频直播时会因网络分组丢失发生花屏现象,而点播时会因为网络或平台缓存等原因造成数据分组到达延迟而引起视频卡顿。观看体

验质量首先考虑视频卡顿现象,一般代表着视频重新进行缓冲,因此模型的输入变量为缓冲平均时长 (*Duration*, 所有的缓冲时长之和除以缓冲次数 *Frequency*) 与多次缓冲情况下缓冲间隔的平均值 (*Interval*)。其次考虑视频流中花屏时间 (*BlockingTimeRatio*) 占比和花屏面积占比 (*BlockingAreaRatio*):

$$Q_v = f(\text{Duration}, \text{Interval}, \text{Frequency}, \text{BlockingTimeRatio}, \text{BlockingAreaRatio})$$

$$\text{Duration} = \frac{\sum \text{BufferLength}}{\text{Frequency}}$$

$$\text{Interval} =$$

$$\frac{\sum \text{BufferStartTime}_i - (\text{BufferStartTime}_{i-1} + \text{BufferLength}_{i-1})}{\text{Frequency} - 1} \quad (4)$$

3.2 部分地区视频服务体验现状

为了验证《评估标准》的实用性,工作组在标准 (1.0) 版本发布后首先以视频播放服务为对象进行了标准验证,国内主要电信运营商和互联网视频服务商联合在国内开展了首次视频播放的大范围用户体验评测活动。国内目前的视频播放服务主要包括 IPTV 业务、互联网视频业务和有线电视业务,由于时间原因,本次用户体验评测先采集了部分地区的 IPTV 业务 (中国电信和中国联通等部分省电信运营商的 IPTV 业务) 和互联网视频业务 (包括优酷土豆、爱奇艺、腾讯视频、乐视、芒果 TV 和风行等主流服务商) 的样本数据,而有线电视业务的样本数据采集的工作正在进行中。本次测试的报告于 2016 年 12 月发布在《中国视频服务体验白皮书》中。根据报告中数据分析, IPTV 业务的用户体验基本得到保障,互联网视频体验受网络因素影响较大;图 7 为我国 IPTV 及互联网视频服务 uVES 体验得分, IPTV 的样本用户的 uVES 得分平均为 3.06 (良),而互联网视频服务在 PC 端的 uVES 得分平均为 1.79 (差),互联网视频手机端的 uVES 得分平均为

2.53 (中); 应该说相比 IPTV 业务, 互联网视频服务仍然有很大的提升空间, 其主要原因是 IPTV 全程通过专网传送, 网络传输质量能得到保障, 而互联网视频服务通过公网传送, 需要和其他互联网业务共享网络资源, 在出现流量高峰时用户体验难以保障。

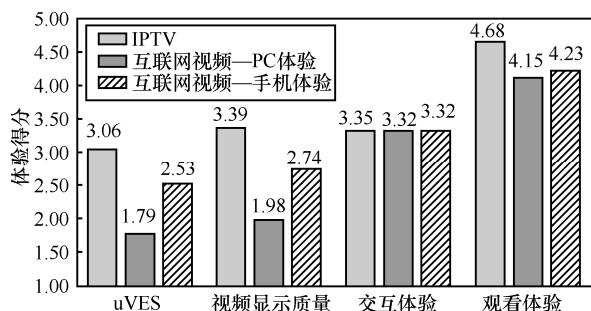


图7 我国IPTV及互联网视频服务uVES体验得分
(2016年12月)

此外, 华为 mLAB 基于 SpeedVideo App 2016 年全球 vMOS 测试数据, 对视频体验及网络能力现状进行了深入挖掘, 并且从全球区域、国内省、国内一线城市典型场景等维度展开分析, 与 2017 年 2 月发布了《全球 vMOS 洞察报告 2017Q1》。报告指出, 全球区域维度视频体验: 非洲东南部和中国区靠前, 其平均 vMOS 约为 3.8; 拉丁美洲北部、南太平洋地区和亚洲中部靠后, 其平均 vMOS 不到 3.6, 且南太平洋地区 vMOS 分布呈现多样化。中国区 vMOS Top30 城市中东部城市占比 66.7%, 超过了东部城市个数的比例 56%; 西部城市占比 3.3%, 少于西部城市个数的比例 12%; 中部城市占比 30.0%, 基本持平中部城市个数的比例 32%。在中国区大部分一线城市中按照场景 vMOS 得分对比看, 高速和景区场景 vMOS 总体排名靠前, 可能与该场景下 4G 覆盖普遍较好且负荷较轻有一定关系; 酒店、校园场景排名总体上处于中等水平; 而高铁和地铁场景排名则整体比较靠后, 可能与该场景下 4G 信号穿透损耗大、用户较多且并发率高从而导致 4G 负荷较重有关。

4 视频用户体验提升措施

对于国内专网传输的 IPTV 业务, 由于网络环境能够得到有效保障, 用户体验总体表现良好, 但 IPTV 点播业务的 4K 内容的交互体验却较低, 这是由于 4K 点播内容相对高清 (1080P) 内容在码率上要求更高, 在今后一段时期高清和 4K 内容占比不断增加的情况下, 需要开展网络端到端的优化, 以提升交互体验和观看体验。

对于互联网服务业务用户体验目前发展水平不均衡现象, 不能通过简单的提升视频码率的方式来取得较好的视频服务用户体验, 因为互联网视频服务的高码率得不到高带宽的保障, 码率提升后播放出现卡顿的概率反而较大, 导致用户观看体验评分会下降。若单纯增加初始视频加载时间, 缓冲更多的播放数据来保障用户在观看过程中的流畅性, 也会造成点播的初始视频加载时间的延长, 影响用户的交互体验。因此互联网厂商应同时面向视频显示质量、交互体验和观看体验进行均衡配置, 在初始视频加载时间不大的情况下, 尽力保证用户观看的流畅度; 综合采用多种技术手段, 如 CDN 下沉尽量靠近用户、新的编码技术、基于主观感受的码率优化及终端播放性能优化等。

5 结束语

视频服务流量目前已经占到全球互联网流量的 75% 以上, 伴随着超高清视频服务的全面开展, 用户在选择视频服务时已经从网络质量转向内容体验质量, 准确的用户体验质量评价技术具有迫切的产业发展需求。本文对用户体验质量理论与视频质量评价技术做出了详尽的介绍, 总结了目前我国视频体验评价标准的进展, 随着学术界与工业界的共同参与, 视频用户体验评价技术将为今后大视频业务体系的高质量发展提供坚实的保证。



参考文献:

- [1] ITU. Amendment 5: New definitions for inclusion in recommendation: ITU-T P.10/G.100[S]. 2006.
- [2] ITU. Definitions of terms related to quality of service: ITU-T Rec.E.800[S]. 2009.
- [3] 姚良, 奚溪. 三网融合下视频业务质量评估体系的研究[J]. 电信科学, 2011, 27(3):27-31.
YAO L, XI X. Research of video service quality evaluation system in triple-play Filed[J]. Telecommunications Science, 2011, 27(3): 27-31.
- [4] ITU. Subjective audiovisual quality assessment methods for multimedia applications: ITU-T Rec.P.910[S]. 2008.
- [5] ITU. Methodology for the subjective assessment of the quality of television pictures: ITU-R Rec. BT.500-13[S]. 2012.
- [6] 韩建亭, 张凤. 基于智能终端的视频通信业务服务质量评测模型研究[J]. 电信科学, 2013, 29(4): 27-32.
HAN J T, ZHANG S. Study on the assessment model of video communication quality based on smart terminal[J]. Telecommunications Science, 2013, 29(4): 27-32.
- [7] 魏耀都, 谢湘, 匡镜明, 等. 移动视频质量评价方法及发展趋势[J]. 电信科学, 2010, 26(3): 57-63.
WEI Y D, XIE X, KUANG J M, et al. Testing method for routers of network-on-chip[J]. Telecommunications Science, 2010, 26(3): 57-63.
- [8] ITU. Network model for evaluating multimedia transmission performance over internet protocol: ITU-T Rec.G.1050[S]. 2016.
- [9] ITU. Quality of experience requirements for IPTV services: ITU-T Rec. G.1080[S]. 2008.
- [10] ITU. Performance monitoring points for IPTV: ITU-T Rec. G.1081[S]. 2008.
- [11] ITU. Measurement-based methods for improving the robustness of IPTV performance: ITU-T Rec. G.1082[S]. 2009.
- [12] ITU. Parametric non-intrusive assessment of audiovisual media streaming quality-higher resolution application area: ITU-T Rec. P.1201.2[S]. 2012.
- [13] ITU. Opinion model for video-telephony applications: ITU-T Rec. G.1070[S]. 2012.
- [14] ITU. Parametric non-intrusive assessment of audiovisual media streaming quality-lower resolution application area: ITU-T Rec. P.1201.1[S]. 2012.
- [15] ITU. Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport-Video quality estimation module: ITU-T R.P.1203.1[S]. 2016.
- [16] YAMAGISHI K, KAWANO T, HAYASHI T. Hybrid video-quality-estimation model for IPTV services[C]// Global Telecommunications Conference (GLOBECOM 2009), Nov 30-Dec 4, 2009, Honolulu, USA. New Jersey: IEEE Press, 2009: 1-5.
- [17] RIES M, CRESPI C, NEMETHOVA O, et al. Content based video quality estimation for H.264/AVC video streaming[C]// 2007 Wireless Communications and Networking Conference(WCNC 2007), March 11-15, 2007, Kowloon, China. New Jersey: IEEE Press, 2007: 2668-2673.
- [18] JOSKOWICZ J, L' OPEZARDAO J C. Combining the effects of frame rate, bit rate, display size and video content in a parametric video quality model[C]//The 6th Latin America Networking Conference(LANC 2011), Oct 12-13, 2011, Quito, Ecuador. New York: ACM Press, 2011: 4-11.
- [19] TAKAGI M, FUJII H, SHIMIZU A, et al. Subjective video quality estimation to determine optimal spatio-temporal resolution[C]//2013 Picture Coding Symposium, Dec 8-11, 2013, San Jose, USA. New Jersey: IEEE Press, 2013: 422-425.
- [20] 中国视频服务用户体验标准工作组. 中国视频服务体验[M]. 北京: 人民邮电出版社, 2017.
China Video Service User Experience Standards Working Group. China video service experience[M]. Beijing: Posts & Telecom Press, 2017.
- [21] HE T T, LIU Y K, XIE R, et al. Evaluation of no reference bitstream-based video quality assessment methods[J]. arXiv: 1706.10143.
- [22] ITU. Parameter values for ultra-high definition television systems for production and international programme exchange: ITU-T Rec. BT.2020-2[S]. 2015.

[作者简介]



贺甜甜 (1993-), 女, 上海交通大学图像通信与网络工程研究所、未来媒体网络协同创新中心硕士生, 主要研究方向为终端视频质量评价。



刘彦凯 (1992-), 男, 上海交通大学图像通信与网络工程研究所、未来媒体网络协同创新中心硕士生, 主要研究方向为超高清视频质量评价。



宋利 (1975-), 男, 上海交通大学图像通信与网络工程研究所、未来媒体网络协同创新中心研究员、博士生导师, IEEE 高级会员, 主要研究方向为视频编码、图像处理与计算机视觉。