Kamil Koniuch[a]*, Sabina Baraković[b] , Jasmina Baraković Husić[c]
KatrienDe Moor[d] , Lucjan Janowski[a] , Michał Wierzchoń[e,f,g]

[a] 通信技术研究所，*AGH*科技大学，波兰克拉科夫；[b] 萨拉热窝大学交通与通信学院，波斯尼亚和 *vo,* 那
萨拉热窝 *Zn* 挪威科技大学信息安全与通信技术系，挪威特隆赫姆；[d] 雅盖隆大学心理学研究所意识实验室，波兰克拉科夫；[f] 雅盖隆大学脑研究中心，波兰克拉科夫；[g] 雅盖隆大学以人为中心的人工智能实验室，波兰克拉科夫

*通讯作者 *koniuch@agh.edu.pl,*

自2020年以来，Kamil Koniuch一直是AGH科技大学通信技术研究所的博士生。他是AGH视频体验质量（QoE）团队的一员。他的研究重点是体验质量测量的生态效度。2019年，他获得心理学硕士学位。他毕业于雅盖隆大学心理学研究所。

ORCID：000-0002-8243-7155

Sabina Baraković是萨拉热窝大学副教授，也是波斯尼亚和黑塞哥维那安全部的专家顾问。她是研究实验室Little Mama Labs的联合创始人。她曾/现任多个国家和国际项目（地平线2020、北约、国际电信联盟、Erasmus+、欧洲委员会、欧洲安全与合作组织等）的首席研究员，并与世界各地的合作者开展合作（如欧盟成员国、东南欧地区国家、卡塔尔等）。她的研究兴趣包括多媒体多维体验质量建模、移动网络应用和统一通信、智能环境与物联网/万物网中的网络安全，以及服务质量/体验质量/生活质量管理等。

ORCID: *0000-0001-8121-9765*

Jasmina Baraković Husić 是萨拉热窝大学的正教授。她受雇于萨拉热窝的 BH Telecom 股份有限公司，担任规划部及

技术与服务开发部的专家。她是研究实验室 Little Mama Labs 的联合创始人。她曾担任多个国家和国际项目（北约、Erasmus+、教育部、RAI 等）的首席研究员，并与世界各地的合作者开展合作（例如欧盟成员国、东南欧地区国家等）。她的研究兴趣涵盖网络多媒体服务、服务质量/体验质量/生活质量管理以及 5G物联网通信等多个领域。

ORCID： *0000-0001-6119-6447*

Katrien De Moor 自2019年起担任挪威科技大学信息安全与通信技术系副教授。她于2012年获得根特大学社会科学博士学位，随后在挪威科技大学获得玛丽·居里博士后奖学金。自2018年起，她担任多学科期刊《质量与用户体验》（Springer）的联合主编，也是挪威青年科学院20位创始成员之一。她的主要研究兴趣和工作涉及人技体验（例如基于视频的体验和沉浸式体验……）以及人技行为、相关的方法学挑战（包容性和用户多样性、生态效度等），以及伦理影响（例如能动性、设计过程中的权力动态、信息与通信技术的生态足迹）。

ORCID：0000-0002-8752-3351

Lucjan Janowski 于2006年获得波兰克拉科夫AGH科技大学电信博士学位。2007年，[illegible] 2010至2011年，他在瑞士日内瓦的日内瓦大学担任博士后研究员。2014至2015年，他在奥地利维也纳的维也纳电信研究中心担任博士后研究员。他目前是AGH科技大学通信技术研究所的助理教授。他的研究兴趣包括受试者统计学与概率建模，以及用于体验质量评估的主观评分。

ORCID：0000-0002-3151-2944

Michał Wierzchoń 是一位社会科学教授。目前，他担任脑研究中心主任及哲学院科研事务副院长。他还是意识实验室（C-Lab）的创始人兼负责人。近期，他获提名为欧洲认知心理学会候任主席（2023—2024年）。他的研究聚焦于意识的神经元和认知机制，包括有意识知觉、记忆和自我意识。他采用多种研究方法，

例如行为研究、神经影像学、定性研究和计算建模。在应用研究方面，他的工作聚焦于为盲人开发感觉替代系统，以及评估媒体质量。

ORCID：0000-0002-7347-2696

视频体验质量研究中的自上而下与自下而上方法：综述及新模型提案

现代视频流媒体服务需要对所呈现的视听材料进行质量保障。质量保障机制使流媒体平台能够以尽可能少的数据传输量提供被认为足以提升用户满意度的质量水平。人们开发了多种措施和方法来控制视频质量，例如根据网络状况进行调整。其中包括质量的客观指标，以及通过主观感知判断确定的阈值。近期，（多）媒体研究人员开始关注前一类指标。他们将这一研究领域称为“体验质量”（QoE）。本文回顾了体验质量的理论模型，并讨论了这些模型的属性及其对该领域的意义。我们认为，其中大多数模型都采用自下而上的建模方法。这类模型着重描述尽可能多的变量，但研究变量之间因果关系的能力有限，因此研究结果在实践中的适用性也受到限制。为推动该领域的发展，我们提出一种视频体验质量的结构化自上而下模型，用以描述变量之间的因果关系。我们希望该框架有助于设计该领域中可比较的实验。

关键词：词语；另一个词语；除名称外均用小写

学科分类代码：如果期刊要求，请在此处填写

引言

多媒体质量是众多领域从业者和研究人员关注的课题，涉及电信工程、语音、音频与视频处理、心理物理学、人机交互、心理学、人体工学、人因研究，以及创新与经济学（Raake & Egger, 2014年）。体验质量（QoE）作为更广泛的多媒体研究领域的一部分，关注各种多媒体服务的主观质量感知。更具体地说，体验质量研究旨在识别对实现或阻碍

良好用户体验的关键因子和特征，并对其进行优化，使所提供的质量水平被感知为较高，从而满足用户的期望，并促使用户未来愿意再次使用该服务（Möller & Raake, 2014 年）。

多年来，关于体验质量（QoE）的概念解释和定义，相关文献一直较为零散（例如，可参见 De Moor（2012年）的综述）。然而，作为 Qualinet 项目的一部分，近期一次社区层面的努力促成了一个新的整体性定义（Brunnström, 等, 2013）。根据该定义，体验质量（QoE）是“应用程序或服务用户的愉悦或烦恼程度。它源于用户的个性和当前状态，以及应用程序或服务的效用和/或享受是否满足了其期望”。这一定义非常宽泛，涵盖了一般意义上的多媒体。实际上，不同类型多媒体领域的专家（例如，网页浏览（Baraković & Skorin-Kapov, 2015年；Baraković & Skorin-Kapov, 2017年；Baraković & Skorin-Kapov, 2017年）、统一通信（Baraković Husić, 等, 2020；Barakovic Husic, Barakovic, Krejcar, & Maresova, 2021）、声音（Ragano, Benetos, & Hines, 2019）和游戏（Laghari, 等, 2019））会分别研究特定的多媒体服务。这是因为，不同的使用场景和服务类型需要理解其中发挥作用的关键因子，并将这些因子纳入考虑，以改善体验质量（QoE）。这些关键因子往往取决于具体的应用程序和使用场景，因此会导致一组多样化的变量，需要对其进行操作化、捕获和理解。

因此，针对某一种多媒体类型描述专门的模型可能会有所裨益，因为这类模型能够为特定类型的服务提供可行的见解。本文着眼于视频流生态系统中各方对相关见解的需求，以及视频服务所带来的互联网流量占比不断攀升这一趋势，因此将只关注视频流媒体情境下的体验质量。我们的

目标则是进一步完善并调整本文提出的模型，使其未来也适用于其他类型的多媒体服务。

在视频质量的情境下，现代服务利用研究成果来改善和优化用户体验。例如，用户评估促进了视频多方法评估融合（VMAF）的开发，这是一种能够自动预测主观评分的客观全参考指标（Li、Aaron、Katsavounidis、Moorthy 和 Manohara，2016）。这类指标被用于开发新的视频压缩算法，旨在降低能源消耗。因此，质量感知优化可能有助于提高视频点播服务的可持续性（Lee、Lee 和 Song，2019）。由此，QoE研究可以成为日益增长的研究的一部分，用于探讨视频服务对自然环境的影响（Ejembi 和 Bhatti，2015年；Batmunkh，2022；Chandaria、Hunter 和 Williams，2011年）。

在此情境下，评估与服务可持续性相关的方面（例如能源效率）是可行的，因为QoE概念不仅描述用户与服务之间的交互，还提供了一个用于优化的框架。具体而言，标准化测量与测试协议使这一点成为可能，因为这些协议允许研究人员测量质量主观判断。在传统的视频体验质量标准化实验中，使用无声的短视频片段作为刺激（国际电信联盟无线电通信部门，2020）。每呈现一个刺激后，便收集一次主观质量判断。这一流程会针对数十个视频重复进行，且相同内容会多次呈现。对此方法的一项主要批评是，这种研究范式过于简单，限制了所研究变量的范围。实际上，它虽能提供可靠的结果，却难以深入揭示人们在自然环境中如何感知视频质量；在这种环境中，影响用户的不仅仅是视频压缩质量，还有许多其他因素。

因此，基于这一范式收集的数据集得出的指标，可能会误判日常用户体验，包括相关的认知加工及其对用户行为的影响。

为应对这一挑战，研究人员对体验质量（QoE）的影响因素（IFs）进行了研究。影响因素（IFs）的定义是“用户、系统、服务、应用程序或情境的任何特征，其实际状态或设置可能会影响用户的体验质量”（Brunnström, 等, 2013）。在视频服务的情境中，存在大量可能影响用户的变量。此外，研究人员还研究了体验质量（QoE）的感知维度（PDs）或“感知特征”（Baraković Husić & Baraković, 2022）。这些变量的定义是“个人对某项服务的经验中可感知、可识别且可命名的特征，这些特征有助于形成该服务的质量”（Jekosch, 2005）。感知维度（PDs）的概念凸显了刺激或情境的客观变化与人们对这些变化的主观感知之间的差异。

因此，体验质量（QoE）可视为一个多维概念，在设计实验时需要考虑许多变量。由于人类感知是体验质量的关键组成部分，其中许多变量都是潜在的。因此，需要坚实的理论基础，以便提供可比较的操作化方式和实验程序。为此，人们提出了一系列详尽的理论模型来描述体验质量的复杂特性（Brunnström, 等, 2013; Raake & Egger, 2014; Möller, Wältermann, & Garcia, 2014; Reichl, 等, 2015; Robitza, Schönfellner, & Raake, 2016; Schmitt, Bulterman, & Cesar, 2018; Reiter, 等, 2014）；（Baraković Husić & Baraković, 2022; Geerts, 等, 2010; Egger, Reichl, & Schoenenberg, 2014）。这些模型涵盖了许多变量，并对信息因子（IFs）和感知维度（PDs）进行了分类。

为了研究哪些信息因子（IFs）对体验质量（QoE）最为重要，研究人员采用探索性研究设计，这类设计通常基于自我报告方法，例如问卷（Baraković Husić & Baraković, 2022; Zhu, Heynderickx, & Redi, 2015）。通过这种方法，研究人员能够在自然情境中测量大量变量（Staelens, 等, 2010）。这种研究信息因子（IFs）和感知维度（PDs）的方法可称为自下而上（见图1）。该方法的目标是尽可能研究更多变量。

然而，采用自下而上方法设计的研究往往无法探究模型中变量之间的因果关系。缺少这方面的信息时，模型可能会因误解混杂变量而错误估计某些影响因素或感知维度的重要性（McElreath，2020）。自上而下方法更适合探究有关变量间因果关系的信息如何改变对其影响的估计（图1）。这种方法不仅能够探究预测变量的影响，还能以图的形式探究它们之间的相互关系。此外，用于描述变量间关系的图示有助于厘清实验设计过程。通过直观呈现假设和变量，研究人员更容易找到合适的统计方法，并与其他研究人员交流结论。因此，设计出可与以往研究相比较的实验也更加容易。尽管自上而下方法显然具有这些优势，但据我们所知，采用这类方法的体验质量研究数量有限；这类研究所用的模型从预先设定的变量列表出发，并检验这些变量的影响。

鉴于上述情况，本文旨在提出一种可通过实验数据验证的视频体验质量模型。此外，我们的模型为研究影响因素与感知维度之间的因果关系提供了框架。最后，我们希望我们的模型

能够通过在 QoE 研究中提供清晰的分类体系，促进相关讨论。本文将聚焦于视频体验质量，但我们的模型也可调整以适用于其他使用场景。本文其余部分安排如下：下一节将详细比较并讨论这两种建模方法及其优缺点。此外，我们将介绍模型的组成部分及其操作化。最后，我们将提出一个广义视频体验质量模型，描述变量之间的因果关系，并讨论其对视频体验质量研究的启示。

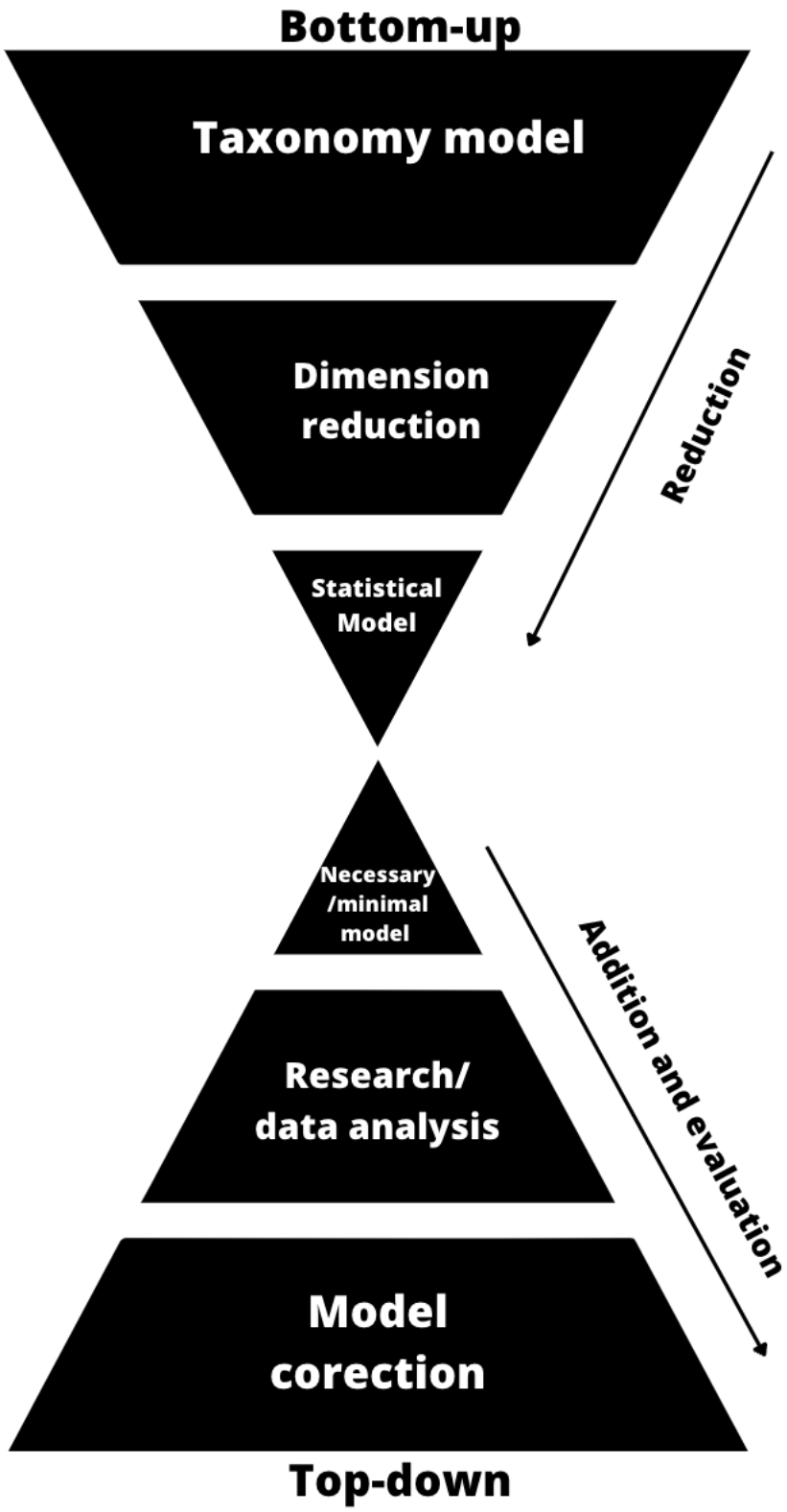


图*1.* 建模方法。图形大小表示模型中的变量数量，箭头

自上而下方法与自下而上方法

为了直观呈现先前发表的体验质量模型与我们的模型之间的差异，

我们用自下而上和自上而下 方法来描述它们（图1）。这些术语用于描述不同领域中的模型，从认知过程（例如注意；参见（Connor、Egeth和Yantis，2004））到主观幸福感的因果方向（Headey、Veenhoven和Weari，2005）。本文中，我们使用这些术语来强调不同方法在纳入变量数量及其分析类型上的差异。模型复杂性与其所提供信息量之间的权衡，是研究人员必须解决的关键问题之一。良好的模型应尽可能少地依赖假设和变量，同时尽可能准确地进行预测（Simon，2001；Gabaix和Laibson，2008）。模型的这一特性称为简约性。要实现简约性，必须确定最重要的变量。为了识别模型的关键组成部分，可以先尽可能多地纳入变量，再找出其中最重要的变量。我们称之为自下而上方法（见图1）。另一方面，也可以先确定描述该现象所必需且足够的最少变量，再据此构建模型。我们称之为自上而下方法。下文将进一步详细讨论这些方法。

自下而上方法

自下而上方法通常用于探索 数据集。其主要目标是找出最重要的变量或模式，并在可能的情况下识别该现象的潜在结构。这种方法的成果是由最重要变量构成的模型。这类研究常被用作后续实验设计的假设来源。下面，我们简要概述体验质量研究中使用的自下而上方法。

多年来，人们开发了大量统计方法，以深入了解数据集。在体验质量研究的情境中，统计分析

方法小组致力于在该领域实施新方法（VQEG，2022）。主成分分析（主成分分析）（Ketykó、De Moor、Joseph、Martens 和 De Marez，2010；Msakni 和 Youssef，2016；Baraković Husić 和 Baraković，2022）以及因子分析（FA）（例如，Özer、Argan 和 Argan，2013；Ende，2015）等方法通常用于探究数据中的潜在结构。方差、协方差和相关性分析也可用于研究数据模式。近年来，机器学习和网络分析方法已成为分析复杂非线性关系的热门解决方案（Youssef、Mellouk、Afif 和 Tabbane，2016）。上述所有方法均可用于包含许多变量和因子的数据集。不过，不可能在同一实验设计中纳入所有潜在的信息因子和感知维度。因此，为了涵盖这些差异，研究中常使用问卷（Baraković Husić 和 Baraković，2022；Ketykó、De Moor、Joseph、Martens 和 De Marez，2010；Msakni 和 Youssef，2016；De Pessemier、Martens 和 Joseph，2013）。

研究	分析方法	数据类型	样本
（Ketykó, De Moor, Joseph, Martens, & De Marez，2010)	主成分分析和相关分析	问卷，经验抽样方法	19名参与者
（Msakni 和 Youssef, 2016）	主成分分析和相关性分析	情绪量表，ACR量表	72名参与者
（Youssef、Mellouk, Afif和Tabbane，2016）	PCA 机器学习	ACR量表	45 位参与者
(Baraković Husić & Baraković, 2022)	PCA 多元线性	问卷，ACR量表	233 位参与者

	回归 （多元线性回归）		
（Özer、Argan 和 Argan, 2013）	PCA 验证性因子分析 （验证性因子分析）	问卷	1000 参与者
（Ende，2015年）	探索性因子分析 CFA t检验 相关分析	问卷, ACR量表	探索: 107名参与者 验证: 100名参与者
(De Pessemier, Martens 和 Joseph, 2013)	MLR	经验抽样 日记 问卷	29名参与者

除了探索性研究中常见的统计方法外，这种方法也可用于定性研究。对问卷中的访谈和开放式问题进行分析，能够提供丰富的信息来源，而研究人员必须对这些信息进行精简和整理（Ickin 等，2012年；Staelens 等，2010）。同样，文献计量研究也能帮助我们深入了解大量既有成果和结论（Zhao、Liu 和 Chen，2016；Vega、Perra、De Turck 和 Liotta，2018；Maia、Yehia 和 de Errico，2015年；Skorin-Kapov、Varela、Hoßfeld 和 Chen，2018；Alreshoodi 和 Woods，2013；Alreshoodi 和 Woods，2013；Juluri、Tamarapalli 和 Medhi，2016）。此外，混合

方法（如开放式质量剖析（OPQ）（Strohmeier、Jumisko-Pyykkö、Kunze 和 Bici，2011年））可以基于定性数据得出定量结果。在其他情况下，研究人员采用混合方法，通过后续的定性研究加深对定量数据的理解（例如，参见 De Moor、Borge 和 Heegaard，2019；Øie、Koniuch、Cieplińska 和 De Moor，2021）。上述所有方法都有相似的目标：更好地理解收集到的数据。这类洞见对于研究体验质量等复杂现象尤其有用。

然而，这种方法存在一些局限性。自下而上方法很少提供有关变量之间因果关系的信息，例如体验质量的影响因素或感知维度。实际上，人们很容易误判某些变量的重要性。此外，可能存在同时影响预测变量和预测值的混杂变量。分析数据的特性也可能导致解释误差。所有测量都受到某种未知程度测量误差的影响。对于自我描述数据，科学家需要格外谨慎。许多因素都可能影响参与者的回答。另一种减少模型中变量的方法是采用简化的实验设计。受控实验室实验旨在实现这类简化（国际电信联盟无线电通信部门, 2020）。在这种方法中，研究人员主要关注对技术因素进行操控，以便更好地预测这些因素对感知视频质量的影响。事实证明，这种方法有助于构建 VMAF 等质量指标（Li, Aaron, Katsavounidis, Moorthy, & Manohara, 2016），也适用于评估新的压缩算法。不过，VMAF 预测的生态效度仍未确定。上述实验与用户的日常经验相去甚远，采用这种方法构建的指标可能会错误估计用户满意度。

总而言之，自下而上建模在体验质量研究中占据主导地位。许多理论模型试图汇集所有潜在的影响因素和感知维度（Brunnström, 等, 2013; Raake & Egger, 2014; Möller, Wältermann, & Garcia, 2014; Reichl, 等, 2015; Robitza, Schönfellner, & Raake, 2016; Schmitt, Bulterman, & Cesar, 2018; Reiter, 等, 2014）；（Baraković Husić & Baraković, 2022; Geerts, 等, 2010; Egger, Reichl, & Schoenenberg, 2014）。这些模型突显了人类感知的复杂性，并为影响因素和感知维度提供了分类体系。这些模型启发了许多研究，用以探究体验质量最重要的预测变量；其中，系统层面和技术影响因素受到了广泛关注。尽管文献中已有呼吁要求探讨这些关系（例如，见（Reiter, 等, 2014）），但影响因素、感知维度与体验质量之间的因果关系仍未得到充分研究。为此，自上而下方法可能更有用。

自上而下方法

自上而下方法是一种用于模型比较的理论驱动方法。借助这种方法，可以在不同的理论模型和统计模型之间作出选择。通过这种思路，可以利用观察数据或实验数据验证理论等抽象概念。进而，可以通过融入实验带来的新见解来构建模型。最重要的是，这种方法可以验证数据驱动模型得出的结论。此外，借助这种方法，还可以研究变量之间的因果关系，并确定混杂变量。

目前，用于验证理论模型的统计方法正在迅速发展。各种方法在技术细节上有所不同，但都基于相似的假设。所有这些方法都以图为基础，用来表示所研究变量之间的

关系，这种方法称为路径分析。因此，路径分析不仅可以添加新的预测变量，还可以确定它们之间的因果关系。通过测量模型与实验中收集的数据的拟合程度，可以验证模型。最常用的路径分析模型是验证性因子分析（CFA），它属于结构方程建模和因果结构模型。

这种方法的主要缺点是可供研究的变量数量有限。图中包含的单元越多，就必须表示并分析它们之间越多潜在的联系。此外，根据模型比较提供的统计学结果得出结论也更加困难。例如，信息量更大的模型可能具有较差的统计拟合度。最后，能够提供充分数据的实验十分复杂，而且必须以大样本开展。

然而，要解决现代体验质量问题，自上而下建模是必要的。标准化实验缺乏生态效度，因此需要测量大量因子并进行更复杂的操控。要理解的不仅是测试者的评分，还有消费者的整体行为，这就需要更深入地了解数据集中变量之间的关系。此外，如果不研究影响消费者行为的变量之间的因果关系，研究人员就容易受到混杂因素的影响。

然而，要在数据建模中采用理论驱动方法，就需要简化当前的体验质量模型。因此，我们提出一种体验质量建模的加性方法。在下一节中，我们将介绍一个视频体验质量图模型，其中仅包含必要且最重要的少量变量。通过进一步添加影响因素和感知维度，可以扩展该模型，同时保持其核心不变。换言之，我们提出了一个通用模型，研究人员可以对其进行调整，以研究

特定情境。虽然我们提出该模型是用于视频体验质量，但也可以对其进行调整，以适用于其他使用场景。

基于路径图的视频体验质量模型

我们的模型基于对体验质量一般定义第一部分的解读：“体验质量是应用程序或服务用户的愉悦或烦恼程度。”因此，我们没有在图2中将体验质量表示为一个独立单元。相反，我们将“愉悦或烦恼”作为视频服务体验中典型变量交互的结果。在现实生活中，很难想象用户的愉悦或烦恼并非由视频内容引起的情境。根据互联网效能的不同，这种依赖于内容的体验可能会受到视频质量下降的调节。基于这一思路，我们构建了一个体验质量路径模型。

沿用以往的体验质量模型，我们将影响因素和感知维度区分为一般用户满意度的预测变量。影响因素定义为“用户、系统、服务、应用程序或情境的任何特征，其实际状态或设置可能会影响用户的体验质量”（Reiter 等，2014年）。感知维度指“个人服务体验中可感知、可识别且可命名，并有助于提升服务质量的特征。”此外，我们将这些变量描述为潜变量，并提供了测量示例。下面，我们将简要介绍模型中各变量的操作化方式。

组件操作化

服务质量服务质量（QoS）是衡量网络整体性能的指标。它通常用于描述网络为其用户提供稳定服务水平的能力

。国际电信联盟电信标准化部门将服务质量定义为“电信服务特性的总和，这些特性关系到该服务满足用户明示和隐含需求的能力”（建议书，2008年）。服务质量可以通过吞吐量、时延、抖动和丢包等性能指标来表示。用于服务质量评估的专门指标正在不断开发。也有一些模型描述服务质量对体验质量的影响（Fiedler、Hossfeld 和 Tran-Gia，2010；Chen、Wu 和 Zhang，2014；Kim 和 Choi，2014）。服务质量是体验质量的一个知名信息因子。

内容

内容特征具有多维性。根据研究问题的不同，可以考虑不同的变量来分析其影响。由于内容具有复杂性，因此可以将其视为系统或情境信息因子。例如，运动、细节数量、亮度和计算复杂度等特征已在体验质量研究中得到广泛认可（Song & Tjondronegoro, 2014；Duanmu, Ma, & Wang, 2018；Khan, Sun, & Ifeachor, 2011）。这些特征被归类为影响体验质量的系统因素。另一方面，用户选择的内容类型可归类为情境信息因子，并可用于预测用户满意度（Agboma & Liotta, 2012）。此外，YouTube 等具有影响力的服务提供商也有自己的内容分类和量化指标。用于区分该服务中视频的一项关键统计指标是“参与度”，其操作化定义为平均观看时长占比。该指标可用于操作化内容的感知维度，例如感知参与度或对视频的兴趣。此外，还可使用自我描述方法评估参与者的兴趣程度（Song, Yang, Zhou, Wan, & Wu, 2016）或动机（Kobayashi, Masuda, & Hayashi, 2012）。此外，还有标准化的视觉刺激数据集，可用于评估内容唤起的情绪的影响（Marchewka, Żurawski, Jednoróg, &

Grabowska, 2014；Di Crosta, et al., 2020；Baveye, Dellandrea, Chamaret, & Chen, 2015；Baveye, Dellandrea, Chamaret, & Chen, 2015）。

多媒体信号质量

在我们的模型中，多媒体信号质量代表视觉刺激的客观属性。在视频流媒体情境下，它指显示在用户设备上的视频。评估视频质量的方法有很多，其中一些是客观的，另一些则是主观的。因此，我们建议将其划分为媒体信号客观质量（QoMS）及其感知维度（PD）——下文所述的感知媒体信号质量。在这种方法中，QoMS 是信息因子（IF），代表源信号（内容）的再现质量。它具有源内容的属性，这些属性会受到网络和用户硬件效能的调节。可使用信噪比或 VMAF 等客观指标对其进行评估（García, Gortázar, Gallego, & Hines, 2020; García, López-Fernández, Gortázar, & Gallego, 2019; Li, Aaron, Katsavounidis, Moorthy, & Manohara, 2016）。在现有模型中（Brunnström, 等, 2013; Raake & Egger, 2014; Robitza, Schönfellner, & Raake, 2016），QoMS 被描述为信号的物理表征。

多媒体信号的感知质量

我们使用“感知媒体信号质量”（PQoMS）一词，旨在强调在视频体验质量研究中感知所发挥的作用，因为这类研究聚焦于用户对质量所做的主观评估。通常，研究人员使用 5 分制绝对类别评分（ACR）量表来估计服务质量（ITUT，2016）。此外，目前还有一项新工作，旨在构建兼顾服务质量与感知媒体信号质量的矩阵，以预测用户满意度（Li、Aaron、Katsavounidis、Moorthy 和 Manohara, 2016）。在提出的模型中（Raake & Egger，2014；Brunnström, 等, 2013），这些描述是质量形成过程的结果。

愉悦或烦恼程度

我们假设，用户的愉悦或烦恼状态（DoA）是质量和内容属性共同作用的结果。根据一般定义，这实际上就是体验质量的衡量指标。例如，可以使用经过调整的差异情绪量表进行评估（De Moor, Quintero, Strohmeier, & Raake, 2013）。在自然情境中，技术因素（Nam, Kim, & Schulzrinne, 2016）和内容相关因素（Laiche, Ben Letaifa, Elloumi i Aguili, 2021）都可能导致体验质量发生变化，进而引起用户行为的转变（Robitza, Schönfellner, & Raake, 2016）。

行为

根据研究范围的不同，行为可能是对质量相关事件的短期反应（Robitza & Raake, 2016; Fogelberg, 2020; Laiche, Ben Letaifa i Aguili, 2021）、习惯评估（Robitza, et al., 2020），甚至是消费者态度（Reichl, et al., 2015; Perkis, Reichl, & Beker, 2014）的预测变量。通常，在零售情境中，愉悦和唤醒是趋近—回避行为的良好预测变量（Kenhove & Desrumaux, 1997）。鉴于网络服务提供商相关的长期行为可能受到一系列其他重要变量（例如定价）的影响，我们将重点放在短期行为上。在我们的模型中，行为是 DoA 的结果，可通过与服务的交互变化（例如，切换视频）来观察。

变量与模型假设之间的关系

我们以图示的形式描述了广义体验质量模型（见图2），其中包含变量之间的因果路径。按照因果路径分析方法，箭头表示这样一种假设：变量A可能是B的直接原因。例如，在图2中，DoA 可能由感知媒体信号质量和内容引起。除此之外，箭头的表示并不意味着该过程很简单。A与B之间的关系

可能是复杂的、多阶段的和非线性的；我们假设的是这种关系的方向。此外，通过绘制从A指向B的箭头，我们并不假设A是导致B的唯一因素。事实上，这种方法假设每个变量都可能受到图中未表示的未指定因子的影响。这使得我们能够在模型中纳入测量误差和未知因子的影响，并在未来更新模型时添加新的变量。然而，如果我们知道模型中的各个单位具有共同原因，就必须在图中予以表示。

另一方面，没有箭头则代表一个更强的假设，即节点所代表的变量之间不存在关系（Paulewicz, Siedlecka, & Koculak, 2020）。例如，在图2中，愉悦或烦躁的程度不会改变服务质量；因此，影响方向似乎无从谈起。

在我们的模型中，我们将视频信号的物理表征（服务质量）描述为网络效能（服务质量）与内容交互的结果。我们假设，从用户的角度来看，除了服务质量之外，服务质量与内容之间不存在其他关系。换言之，如果网络效能对视频内容产生了某种影响，这种影响只能通过服务质量观察到。因此，服务质量不能直接影响 DoA 或行为。用户必须看到服务质量发生变化，才能得出网络效率存在问题的结论。此外，从服务质量指向感知媒体信号质量的箭头代表人类感知。事实上，这一过程可以用质量形成模型来描述，例如（Raake & Egger, 2014; Brunnström, 等, 2013）。此外，视觉敏感度等认知因素也可能对其产生影响（Banitalebi-Dehkordi, Ebrahimi-Moghadam, Khademi, & Hadizadeh, 2019）。另外，我们不能假设质量感知不会受到 DoA 的调节。这就是服务质量与 DoA 之间的箭头是双向的原因。

最重要的是，我们假设在现实场景中，DoA 是内容和感知媒体信号质量共同作用的函数。对于相同的服务质量，用户可能会因内容类型不同而做出不同反应。

最后，用户的行为是这种 DoA 的结果（Seufert、Wassermann 和 Casas，2019；Hu 等，2020）。人们可能会因内容和 PQoMS 而感到不满。这可能导致人们采取行动来提升网络效能、更换内容、进行补偿，或放弃活动。

本文提出的模型具有通用性。这意味着，我们将其视为一个框架，可针对特定使用场景和实验设置进行具体化。如前所述，在这种情况下，可以为模型添加其他变量。这一过程需要对图中的新单元进行操作化并将其纳入图中。例如，如果我们想加入参与者兴趣，可以将其置于从内容指向 DoA 的箭头上。在这种情况下，我们假设兴趣由内容引起，而非由其他变量引起。此外，内容只能通过兴趣水平影响总体满意度。如果我们的假设是兴趣并非总会影响内容与 DoA 之间的因果路径，就应通过增加一条路径将兴趣纳入模型。另一种假设可能是，感知媒体信号质量受视觉敏感度影响。在这种情况下，我们必须将视觉敏感度作为图外的一个新单元加入，并画一条线指向感知质量。模型中的其他影响因子（Reiter 等，2014年）也可以用类似方式添加。

基于这一因果结构，我们可以提出一个仅适用于视频服务的体验质量操作性定义。体验质量是指由内容引发、并受视频感知质量调节的、与行为相关的用户 DoA 的程度。

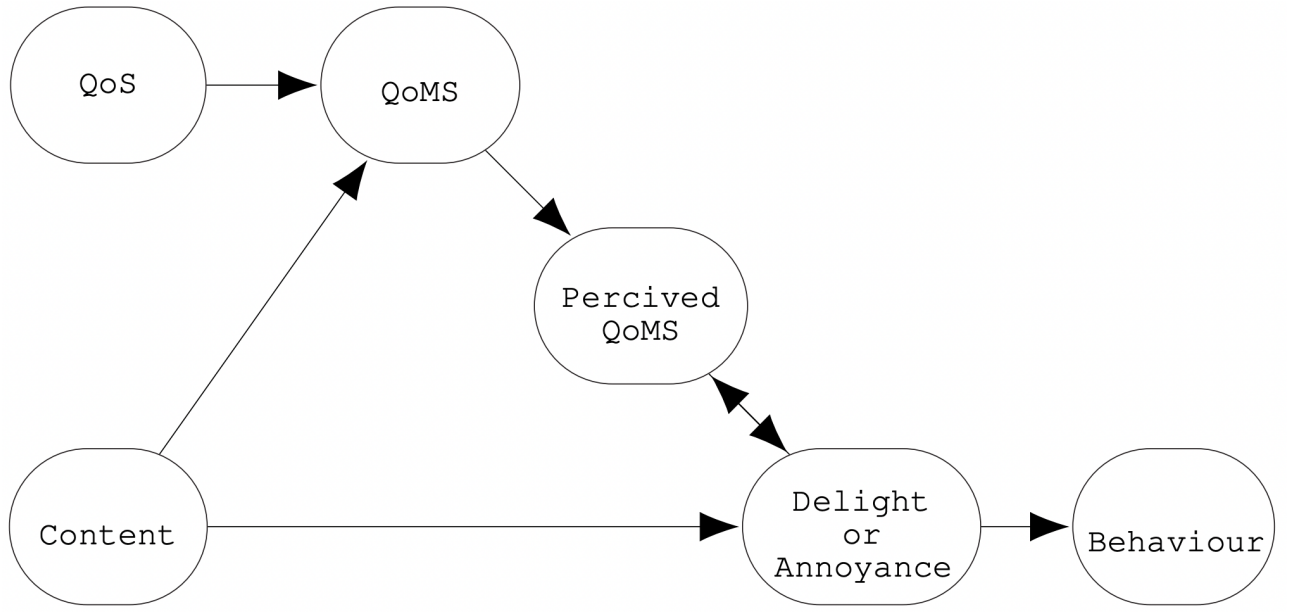


图2 视频体验质量路径模型

启示与未来方向

如上一节所述，我们的模型可用于对新变量进行操作化、构建假设，并检验实验设置中的假定。由于模型具有因果结构，且变量的操作化定义清晰，因此可以利用实验数据对其进行验证。此外，我们用箭头表示的假设也可以得到验证。可采用路径分析，例如结构方程建模或因果结构模型，来验证这些假设。不过，模型所包含的单元数量有限。未来对模型的完善应包括验证模型，并纳入自下而上的体验质量研究中已知的其他因子。这样便能以累加的方式构建更复杂的模型。比较这些模型需要采用新的实验设计，并纳入范围更广的变量。我们希望，这一模型不仅有助于区分哪种压缩方法更好，还能为更复杂的问题提供答案，例如：“质量有多重要？”以及“人们为什么会停止观看视频？”例如，这是因为.....吗？

是因为内容吸引力不足，还是视频质量不够？我们相信，回答这些问题将有助于为未来的视频流媒体提供更可持续的解决方案。尽管我们的模型描述的是视频体验质量，但未来也可以调整应用于不同的多媒体内容。最后，我们希望我们的模型能为体验质量领域的新假设和实验研究提供灵感。应用我们的模型或可为体验质量讨论提供框架，并促成可比较、可复现范式的形成。这一新框架或许能促进不同领域专家之间围绕体验质量展开对话。

致谢，避免在同行评审前透露任何作者身份

利益披露声明：作者未报告任何潜在利益冲突。资金

本研究成果的相关研究获得了挪威金融机制 2014年-2021年的资助，项目编号为 2019/34/H/ST6/00599。

参考文献

国际电信联盟电信标准化部门。（2016）。P.800.1（07/16）：平均意见分（MOS）术语。国际电信联盟：日内瓦，瑞士。

Laiche, F., Ben Letaifa, A., & Aguili, T.（2021）。基于用户行为的视频流媒体服务感知体验质量的流量监测。《并发与计算：实践与经验》，e6678。

Möller, S., Wältermann, M., & Garcia, M.-N.（2014）。体验质量的特征。载于 S. Möller, & A. Raake（编），体验质量：高级概念、应用与方法（第 73–84 页）。Cham：施普林格国际出版公司。doi:10.1007/978-3-319-02681-7_5

Laiche, F., Ben Letaifa, A., Elloumi, I., & Aguili, T. (2021). 当机器学习算法遇上用户参与度参数以预测视频体验质量。《无线个人通信》, *116*, 2723–2741。

Raake, A., & Egger, S. (2014年). 质量与体验质量. 载于 S. Möller, & A. Raake (编), 体验质量：高级概念、应用与方法 (pp. 11–33). Cham：施普林格国际出版公司. doi:10.1007/978-3-319-02681-7 2_

Möller, S., & Raake, A. (2014年). 体验质量的特征. 载于 A. Raake, & S. Egger, 体验质量：高级概念、应用与方法(pp. 73-84). Springer.

De Moor, K. (2012年)。工程师来自火星，用户来自金星吗？弥合体验质量研究中的鸿沟：关于跨学科历程的思考与经验。博士论文，根特大学。

Brunnström, K., Beker, S. A., De Moor, K., Dooms, A., Egger, S., Garcia, M.-N., . . . 等人（2013）。Qualinet 关于体验质量定义的白皮书。

Baraković, S., & Skorin-Kapov, L.（2017）。建模移动网页浏览中设计/性能因素与影响体验质量的感知特征之间的关系。《人类行为中的计算机》，*74C*，311–329。doi:10.1016/j.chb.2017.04.046

Baraković, S., & Skorin-Kapov, L.（2015）。移动网页浏览体验质量的多维建模。《人类行为中的计算机》，*50*，314–332。doi:10.1016/j.chb.2015.03.071

Baraković Husić, J., Baraković, S., Cero, E., Slamnik, N., Oćuz, M., Dedović, A., & Zupčić, O. (2020). 统一通信的体验质量：综述。《国际网络管理期刊》，*30*。doi:10.1002/nem.2083

Barakovic Husic, J., Barakovic, S., Krejcar, O., & Maresova, P. (2021). 基于网页的统一通信感知维度体验质量建模。《信号、图像与视频处理》，1–9。doi:10.1007/s11760-020- 01822-0

Ragano, A.、Benetos, E. 和 Hines, A.（2019 年）。为音频档案评估调整体验质量框架。*2019* 年第十一届国际

多媒体体验质量会议（*QoMEX*），（第 1-3 页）。
doi:10.1109/QoMEX.2019.8743302

Laghari, A. A., He, H., Memon, K. A., Laghari, R. A., Halepoto, I. A., & Khan, A.（2019）。云游戏模型中的体验质量（QoE）：综述。多智能体与网格系统，*15*，289–304。

Li, Z., Aaron, A., Katsavounidis, I., Moorthy, A., & Manohara, M.（2016）。迈向实用的感知视频质量指标。*Netflix* 技术博客，*6*。

Lee, D., Lee, J., & Song, M.（2019）。用于降低视频服务器转码能耗的视频质量自适应。*IEEE Access*，*7*，126253–126264。

Batmunkh, A.（2022）。最受欢迎社交媒体平台的碳足迹。*Sustainability*，*14*，2195。

Chandaria, J., Hunter, J., & Williams, A.（2011年）。观看电视的碳足迹：数字地面电视与视频点播的比较。*2011*年 *IEEE* 可持续系统与技术国际研讨会论文集，（第1–6页）。

Baraković Husić, J., & Baraković, S. (2022). 视频流媒体体验质量的多维建模。《人类行为中的计算机》, *129*。doi:10.1016/j.chb.2021.107155

Jekosch, U. (2005). 为声音赋予意义——产品声音设计情境中的符号学。载于通信声学（第193–221页）。Springer。

Reichl, P., Egger, S., Möller, S., Kilkki, K., Fiedler, M., Hoßfeld, T., . . . Asrese, A. (2015年). 面向体验质量与用户行为建模的综合框架。*2015*年第七届多媒体体验质量国际研讨会 *(QoMEX)*，（第 1–6 页）。

Robitza, W., Schönfellner, S., & Raake, A. (2016). 多媒体服务中体验质量的形成与用户行为的理论研究。第*5*届 *ISCA/DEGA* 系统感知质量研讨会，（第 39–43 页）。

Schmitt, M., Bulterman, D. C., & Cesar, P. S. (2018). 对比效应：同时混合不同视频质量的体验质量。《质量与用户体验》, *3*, 1–17.

Reiter, U., Brunnström, K., Moor, K. D., Larabi, M.-C., Pereira, M., Pinheiro, A., . . . Zgank, A. (2014年). 影响体验质量的因素。载于《体验质量》（第 55–72 页）。Springer。

Egger, S., Reichl, P., & Schoenenberg, K. (2014年). 体验质量与交互性。载于《体验质量》（第 149–161 页）。Springer。

Geerts, D., De Moor, K., Ketykó, I., Jacobs, A., Van den Bergh, J., Joseph, W., . . . De Marez, L. (2010). 将综合框架与适当的方法相结合以测量体验质量。*2010* 第二届多媒体体验质量国际研讨会（*QoMEX*），（第 158-163 页）。doi:10.1109/QOMEX.2010.5516292

Zhu, Y., Heynderickx, I., & Redi, J. A. (2015年). 了解社会情境和用户因素在视频体验质量中的作用。《人类行为中的计算机》，*49*, 412–426.

Staelens, N., Moens, S., Van den Broeck, W., Marien, I., Vermeulen, B., Lambert, P., . . . Demeester, P. (2010). 在真实环境中评估 IPTV 和视频点播服务的体验质量。《*IEEE*广播汇刊》，*56*, 458–466.

McElreath, R.（2020）。众多变量与虚假的华夫饼。载于《统计学再思考》（第 123–160 页）。Chapman and Hall/CRC。

Connor, C. E., Egeth, H. E., & Yantis, S.（2004）。视觉注意：自下而上与自上而下。《当代生物学，14》，R850–R852。

Headey, B., Veenhoven, R., & Weari, A.（2005年）。主观幸福感的自上而下理论与自下而上理论。载于《社会指标研究引文经典》（第401–420页）。Springer。

Simon, H. A.（2001）。科学追求简约性，而非简单性：探索现象中的模式。简单性、推断与建模：保持精妙的简洁，32–72。

Gabaix, X., & Laibson, D.（2008）。优秀模型的七项性质。实证与规范经济学基础：手册，292–319。

VQEG。（2022年11月30日）。视频质量专家组（*VQEG*）*SAM*小组。取自VQEG：https://www.vqeg.org/projects/statistical-analysis-methods-sam

Ketykó, I., De Moor, K., Joseph, W., Martens, L., & De Marez, L.（2010）。在实际的3G网络中进行体验质量测量。*2010 IEEE*宽带多媒体系统与广播国际研讨会（*BMSB*），（第1–6页）。

Msakni, H. G., & Youssef, H. (2016). 用户情绪和视频内容对视频体验质量的影响。第*5*届 *ISCA/DEGA* 系统感知质量研讨会（*PQS 2016*）论文集，（第 97–101 页）。

Özer, A.、Argan, M. T. 和 Argan, M.（2013）。移动服务质量维度对客户满意度的影响。《*Procedia*：社会与行为科学》*, 99*，428–438。

Ende, N. V.（2015年）。体验音视频质量：对感知视频质量与参与度之间关系的研究。 Uitgeverij BOXPress。

Youssef, Y. B., Mellouk, A., Afif, M., & Tabbane, S. (2016). 基于统计选择方法的影响体验质量因素依赖关系的视频质量评估。*2016 IEEE*全球通信大会*(GLOBECOM)*,（第 1–6 页）。

De Pessemier, T., Martens, L., & Joseph, W. (2013). 生活实验室情境下移动视频观看主观质量评价的建模。*2013 IEEE*宽带多媒体系统与广播国际研讨会*(BMSB)*,（第 1–5 页）。

Ickin, S., Wac, K., Fiedler, M., Janowski, L., Hong, J.-H., & Dey, A. K. (2012). 影响常用移动应用体验质量的因素。*IEEE*通信杂志, *50*, 48–56.

Zhao, T., Liu, Q., & Chen, C. W. (2016). 视频传输中的体验质量：一种以用户体验为驱动的策略。*IEEE Communications Surveys & Tutorials, 19*, 285–302.

Vega, M. T., Perra, C., De Turck, F., & Liotta, A. (2018). 视频流媒体服务中预测性体验质量管理综述。《*IEEE*广播汇刊》，*64*，432–445。

Maia, O. B., Yehia, H. C., & de Errico, L. (2015年). 视频流媒体体验质量评估简明综述。计算机通信，*57*，1– 12。

Skorin-Kapov, L., Varela, M., Hoßfeld, T., & Chen, K.-T. (2018). 多媒体服务体验质量管理的新兴概念与挑战综述。*ACM*多媒体计算、通信与应用学报*(TOMM), 14*, 1–29.

Alreshoodi, M., & Woods, J. (2013). 多媒体服务体验质量 \backslash{}{Q}{o}S相关模型综述。 *arXiv*预印本 *arXiv:1306.0221*.

Strohmeier, D., Jumisko-Pyykkö, S., Kunze, K., & Bici, M. O.（2011）。用于以用户为中心的体验质量评估的扩展 OPQ方法：一项关于通过 DVB-H 进行移动3D视频广播的研究。*EURASIP*《图像与视频处理杂志》，*2011*年，1–24。

De Moor, K., Borge, H. K., & Heegaard, P.（2019）。幼儿与视频聊天的使用：对体验质量研究的启示。*2019*年第十一届多媒体体验质量国际会议（*QoMEX*），（第 1–3 页）。

Øie, E. B., Koniuch, K., Cieplińska, N., & De Moor, K.（2021）。影响视频会诊体验质量的因子。*2021*年第十三届多媒体体验质量国际会议（*QoMEX*），（第 137–140 页）。

国际电信联盟无线电通信部门. (2020). 500-14 (10/2019): 电视图像质量主观评价方法. 国际电信联盟：日内瓦，瑞士.

建议书, I. T. (2008). E. 800, 与服务质量相关的术语定义。国际电信联盟电信标准化部门 *(*国际电信联盟电信标准化部门*)* 标准.

Fiedler, M., Hossfeld, T., & Tran-Gia, P. (2010). 体验质量与服务质量之间的一般定量关系。 *IEEE Network, 24*, 36–41.

Chen, Y., Wu, K., & Zhang, Q. (2014). 从服务质量到体验质量：视频质量评估教程。 *IEEE Communications Surveys & Tutorials, 17*, 1126–1165.

Kim, H. J., & Choi, S. G. (2014年). 用于多媒体流媒体服务的体验质量评估模型：基于服务质量参数。《多媒体工具与应用》, *72*, 2163– 2175。

Song, W., & Tjondronegoro, D. W. (2014年). 面向移动视频的基于可接受性的体验质量模型。*IEEE* 多媒体汇刊, *16*, 738–750。

Duanmu, Z., Ma, K., & Wang, Z. (2018年). 自适应流媒体视频的体验质量：一种基于期望确认理论的方法。*IEEE* 图像处理汇刊, *27*, 6135–6146。

Khan, A., Sun, L., & Ifeachor, E. (2011年). 体验质量预测模型及其在 U MTS 网络视频质量自适应中的应用。*IEEE* 多媒体汇刊, *14*, 431–442。

Agboma, F., & Liotta, A. (2012年). 移动内容交付系统中的体验质量管理。《电信系统》, *49*, 85–98。

Song, J., Yang, F., Zhou, Y., Wan, S., & Wu, H. R. (2016). 基于视听质量和用户兴趣的多媒体服务体验质量评估。*IEEE* 多媒体汇刊, *18*, 444–457。

Kobayashi, F., Masuda, M., & Hayashi, T. (2012年). 基于用户动机的移动视频服务体验质量评估方法。图像质量与系统性能 *IX*, *8293*, 第 301–309 页。

Marchewka, A., Żurawski, Ł., Jednoróg, K., & Grabowska, A.（2014年）。Nencki 情感图片系统（NAPS）：介绍一个新颖、标准化、覆盖范围广、质量高且逼真的图片数据库。《行为研究方法》，*46*，596–610。

Di Crosta, A., La Malva, P., Manna, C., Marin, A., Palumbo, R., Verrocchio, M. C., . . . Di Domenico, A. (2020). Chieti 情感动作视频数据库：一项用于心理学情绪研究的资源。《科学数据》，*7*，1–6。

Baveye, Y., Dellandrea, E., Chamaret, C., & Chen, L. (2015年). LIRIS-ACCEDE：一个用于情感内容分析的视频数据库。*IEEE*情感计算汇刊，*6*，43–55。

García, B., Gortázar, F., Gallego, M., & Hines, A. (2020). 使用全参考模型评估 WebRTC 应用中的视频和音频体验质量。*Electronics, 9*, 462.

García, B., López-Fernández, L., Gortázar, F., & Gallego, M. (2019). WebRTC 应用中 VMAF 感知视频质量的实践评估。*Electronics, 8*, 854.

De Moor, K., Quintero, M. R., Strohmeier, D., & Raake, A. (2013). 通过传统和替代性主观测量指标评估体验质量：一项关于 IPTV 的探索性“客厅实验室”研究。维也纳，奥地利.

Nam, H., Kim, K.-H., & Schulzrinne, H. (2016). 体验质量比服务质量更重要：人们为何停止观看猫咪视频。*IEEE INFOCOM 2016-*第*35*届 *IEEE* 国际计算机通信年度会议, (pp. 1–9).

Robitza, W., & Raake, A.（2016）。（重新）行动胜于言语？一种用于追踪网络视频服务中用户行为的新型测试方法。*2016* 第八届多媒体体验质量国际会议（*QoMEX*），（第 1–6 页）。

Robitza, W., Dethof, A. M., Göring, S., Raake, A., Beyer, A., & Polzehl, T. (2020). 你还在观看吗？人群中的流媒体视频质量与参与度评估。*2020* 第十二届多媒体体验质量国际会议 *(QoMEX)*，(第 1–6 页)。

Kenhove, P. V., & Desrumaux, P. (1997年)。零售环境中情绪状态与趋近或回避反应之间的关系。《零售、分销与消费者研究国际评论》，*7*，351–368。

Paulewicz, B., Siedlecka, M., & Koculak, M. (2020). 元认知研究中的混杂：初步因果分析框架。《心理学前沿》, *11*, 1933.

Banitalebi-Dehkordi, M., Ebrahimi-Moghadam, A., Khademi, M., & Hadizadeh, H. (2019). 基于视觉记忆建模的无参考视频质量评估。《*IEEE*广播汇刊》, *66*, 676–689.

Seufert, M., Wassermann, S., & Casas, P. (2019). 将用户行为纳入体验质量周期：迈向主动式体验质量感知流量管理。《*IEEE*通信快报》, *23*, 1145–1148.

Hu, M., Chen, J., Wu, D., Zhou, Y., Wang, Y., & Dai, H.-N. (2020). TVG流媒体：学习用户行为以优化体验质量的360度视频流媒体。《*IEEE*视频技术电路与系统汇刊》, *31*, 4107–4120.

Baraković, S., & Skorin-Kapov, L.（2017年）。网页浏览体验质量建模研究综述。《质量与用户体验》，*2*，1–31。

Ejembi, O., & Bhatti, S. N.（2015年）。你的 Netflix 使用习惯所产生的能源成本。*2015*年*ACM*第六届未来能源系统国际会议论文集，（第225–226页）。

Fogelberg, E.（2020年）。我们通过移动来感知：播放栏下卡顿的分支视频期间用户行为的鼠标追踪研究。我们通过移动来感知：播放栏下卡顿的分支视频期间用户行为的鼠标追踪研究。

Juluri, P., Tamarapalli, V., & Medhi, D.（2016）。视频点播服务的体验质量测量：一项综述。*IEEE Communications Surveys &Tutorials, 18*，401-418。doi:10.1109/COMST.2015.2401424

Perkis, A., Reichl, P., & Beker, S.（2014年）。体验质量的商业视角。载于 体验质量（第 97–108 页）。Springer。