5G ToC 音视频体验需求分析及评测

（2022 年）

中国信息通信研究院泰尔系统实验室
中国移动通信有限公司研究院
华为技术有限公司

2022 年 2 月

版权声明

前　言

截至2021年12月，我国5G手机终端连接数达到了4.97亿户，5G基站累计开通142.5万个，所有地级市城区，超过97%的县城城区和50%的乡镇镇区，均实现了5G网络覆盖。随着网络、终端和应用的快速发展，视频类业务已经成为移动网络流量的主要来源。一方面，短视频、高清视频点播/直播、实时音视频类应用的使用量持续增长；另一方面，云游戏、手机全景视频、自由视角视频等新兴应用纷纷涌现。

音视频业务新趋势带来的影响是多方面的：从网络基础设施和应用的角度看，高码率、强交互的视频类业务在5G网络可以获得比4G网络更好的用户体验；从音视频应用体验评测的角度看，KQI需要增加交互响应维度才能更好地匹配视频类业务社交化和交互化的发展趋势。

为保障5G网络下音视频应用的用户体验，更好地推进音视频技术发展，本报告介绍了音视频发展状况及趋势，阐述了音视频应用的体验指标及其理想要求和基本要求，分析了音视频应用在运营商4G/5G 网络下的用户差异化体验。希望本报告可以为音视频应用优化与建网提供借鉴与参考，促进音视频应用为用户带来高质量体验。

目 录

一、音视频发展状况及趋势.....1
二、音视频应用的体验要求及现网评测.....2
（一）点播视频.....2
（二）实时音视频类业务.....7
（三）云渲染业务.....18
（四）空间视频.....23
（五）结论.....30
三、展望和建议.....31
（一）展望.....32
（二）建议.....33
四、参考文献.....36

图 目 录

图 1 视频类业务占全部移动流量的比例变化趋势 1
图 2 视频应用趋势 1
图 3 抖音 App 即点即开与用户退出率的关系 3
图 4 2021 中国网络视听报告关于拖拽/倍速交互行为的调研 4
图 5 国内某一线城市现网路测点播视频首屏时延 PDF 分布（ms） 5
图 6 国内某一线城市现网路测点播视频卡顿次数 PDF 分布 6
图 7 国内某一线城市现网路测点播视频卡顿时长占比 PDF 分布 6
图 8 实时音视频系统总体架构框图 7
图 9 实时音视频 E2E 交互时延演进趋势 8
图 10 连麦直播示意 8
图 11 视频会议示意 9
图 12 连麦直播/视频通话/会议架构示意 9
图 13 国内某省会城市现网路测视频通话的 E2E 时延 PDF 分布（ms） 11
图 14 实时合唱技术方案架构 12
图 15 实时合唱端到端时延构成 13
图 16 国内某一线城市现网路测在线 KTV 合唱的端到端时延分布（ms） 15
图 17 国内某一线城市现网路测在线 KTV 合唱的网络传输时延分布（ms） ..15
图 18 互动白板文件+信令方案架构图 16
图 19 互动白板 E2E 互动响应时延构成 17
图 20 国内某一线城市现网路测互动白板 E2E 互动响应时延 PDF 分布（ms）18
图 21 云游戏原理示意 19
图 22 国内某一线城市现网路测云游戏操作响应时延 PDF 分布（ms） 21
图 23 国内某一线城市现网路测云游戏卡顿次数 PDF 分布 21
图 24 国内某一线城市现网路测云游戏卡顿时长 PDF 分布（ms） 21
图 25 1080P60 云游戏基于 H.264 编码时良好和优秀画质对应的码率范围 22
图 26 1080P60 云游戏基于 H.265 编码时良好和优秀画质对应的码率范围 23
图 27 本测试系统的自由视角实现方案 24
图 28 国内某一线城市现网路测 5G/4G 下视角切换响应时延分布（ms） 26

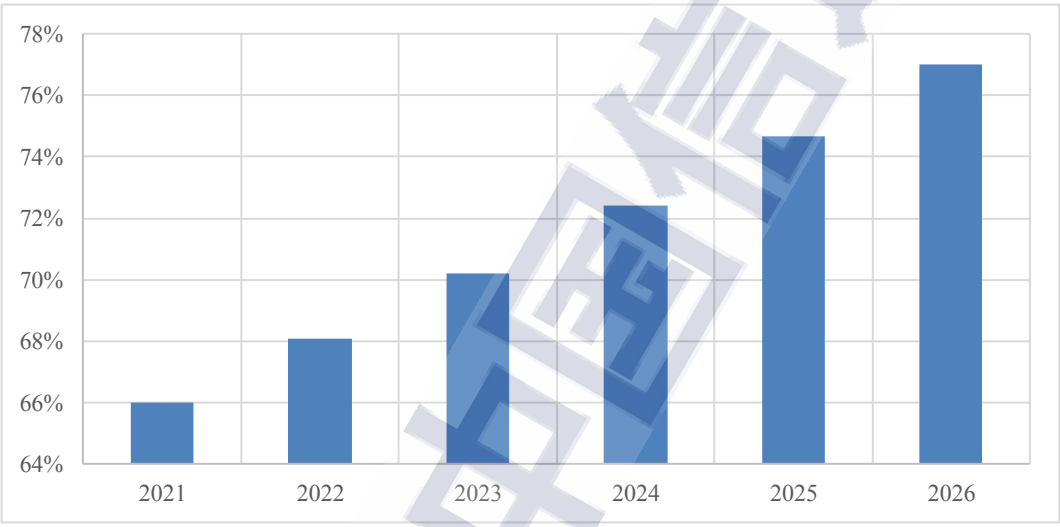
图 29 国内某一线城市现网路测 5G/4G 下首屏时延分布（ms）.....27
图 30 国内某一线城市现网路测手机全景视频（10M）首屏时延分布（ms）.29
图 31 视频技术发展趋势.....32
图 32 视频内容多元化.....33
图 33 5G2C 音视频应用演进路线图.....34
图 34 各类音视频业务与主要 KQI 的对应关系.....35

表 目 录

表 1 视频发展方向与典型应用的对应关系 2
表 2 交互响应容忍值分解 3
表 3 国内某一线城市现网路测点播视频 5G/4G 下 KQI 对比 4
表 4 国内某一线城市现网路测点播视频 5G/4G 下交互行为的 KQI 对比 6
表 5 视频通话/会议/连麦直播的主要 KQI 9
表 6 国内某省会城市运营商 5G/4G 下视频通话 KQI 对比 11
表 7 实时合唱端到端时延要求 13
表 8 实时合唱各个部分延迟当前业界水平 13
表 9 国内某一线城市运营商 5G/4G 下在线 KTV 合唱 KQI 对比 14
表 10 互动白板各个部分延迟当前水平 17
表 11 国内某一线城市现网路测互动白板 5G/4G 下 KQI 对比 18
表 12 国内某一线城市现网路测云游戏 5G/4G 下 KQI 对比 20
表 13 国内某一线城市现网路测自由视角视频 5G/4G 下 KQI 对比 26
表 14 全景视频内容的 PPD 及有效分辨率 27
表 15 国内某一线城市现网路测手机全景视频（10M）5G/4G 下 KQI 对比 28
表 16 单目 4K 全景视频不同体验等级对应的推荐码率 29
表 17 音视频应用的交互响应要求 30
表 18 典型音视频应用在 5G 网络下 KQI 相对于 4G 网络的增益 31

一、音视频发展状况及趋势

视频类业务已经成为现阶段移动网络流量的主要来源，根据 2021 年 6 月爱立信《移动市场报告》，视频类业务流量目前占全部移动流量的 66%，预计到 2026 年将增长到 77%（如图 1 所示）。在视频业态、网络、终端等多方面因素的影响下，视频类业务流量的组成和分布将出现明显变化。



来源：2021 年 6 月爱立信《移动市场报告》

图 1 视频类业务占全部移动流量的比例变化趋势

来源：公开信息整理

图 2 视频应用趋势

2020 年至今视频应用上线的主要新功能如图 2 所示，可见视频应用在向社交化、空间化、交互化、高清化、高帧率化发展，典型应用如表 1 所示。

表 1 视频发展方向与典型应用的对应关系

视频发展方向	典型应用
HDR、高清化、高帧率化	点播视频
社交化	视频通话/会议/连麦直播
	在线 KTV 合唱
	互动白板+连麦
	点播视频“一起看”功能
交互化	云游戏
空间化	自由视角视频
	手机全景视频

来源：公开信息整理

二、音视频应用的体验要求及现网评测

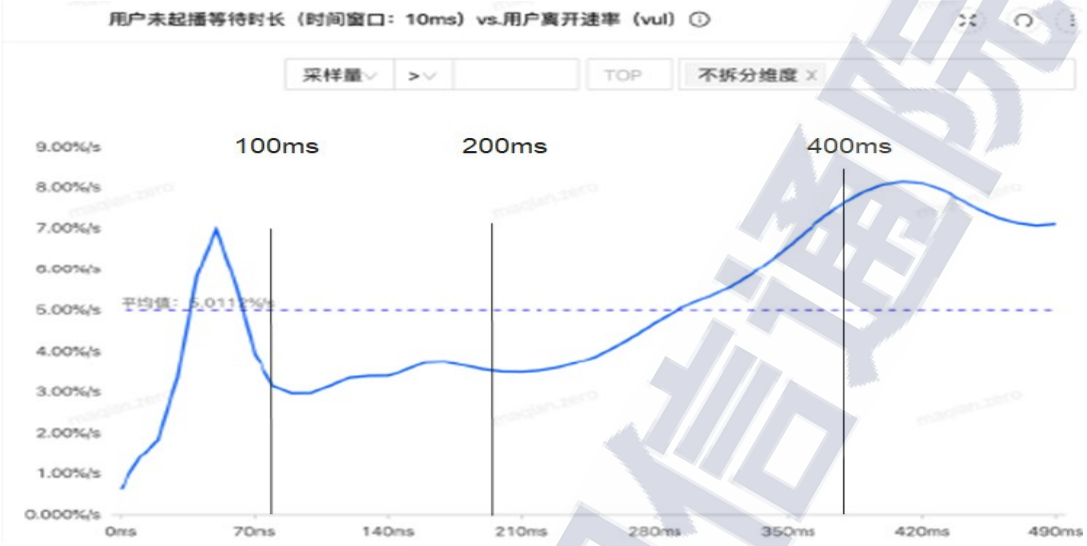
（一）点播视频

1. 应用机理及关键 KQI

根据 ITU-T P.1203[1]，并结合网络传输的角度，点播视频的主要 KQI 为：首屏时延（或初始缓冲时延，初始加载时长），卡顿次数，卡顿时长占比。

根据“多尔蒂阈值”[2]，系统响应时间应该低于 400 毫秒，该门限可以作为点播视频首屏时延的即点即开基本要求。此外，如图 3 所示，根据抖音 App 即点即开和用户退出率的关系[3]：100ms 以内，用户还没反应过来，快速滑动或误点；100~200ms，用户退出率最低

且平稳；200~400ms，用户退出率开始恶化；400ms 以上，用户退出率稳定在较大值。



来源：火山引擎 21 年 12 月产品发布会

图 3 抖音 App 即点即开与用户退出率的关系

不同的交互响应容忍值（即：首屏时延）下对应的时延分解如表 2 所示

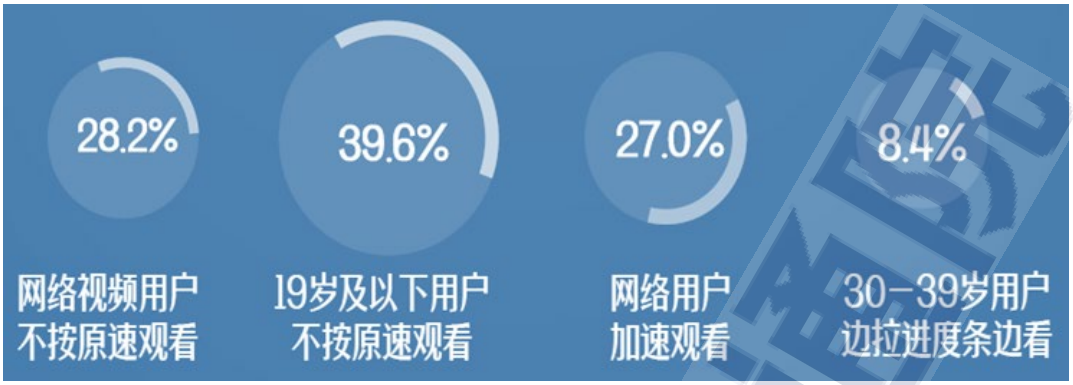
表 2 交互响应容忍值分解

交互响应容忍值（ms）	播放器延迟（ms）	网络传输延迟（ms）
100	70	30
200		130
400		330

来源：三方联合研究成果

如图 4 所示，根据 6 月份发布的《2021 中国网络视听发展研究报告》[4]显示，我国 9.44 亿网络视听用户里，28.2%会选择倍速观看视频，尤其是 00 后用户群体有近四成选择倍速观看方式。上述消费者调研数据表明，交互行为是点播视频的正常发展趋势，因此研究

交互情况下的视频 KQI 指标具有现实意义。



来源：《2021 中国网络视听发展研究报告》

图 4 2021 中国网络视听报告关于拖拽/倍速交互行为的调研

2. 测试结果分析

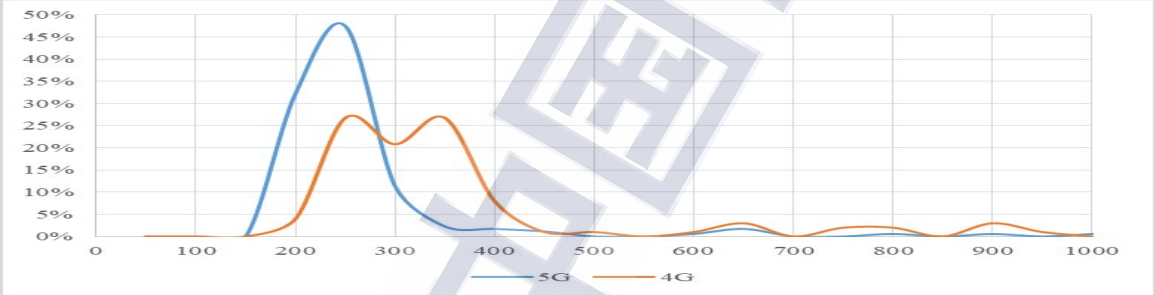
测试环境：运营商 5G/4G 精品网示范区（含交通枢纽、地铁、学校、写字楼等典型场景），测试时段涵盖网络忙闲时（7:00-21:00）。本次点播视频测试选取的内容源是《雪地奔驰》，视频传输协议是 HLS，编码方式是 H.265 Main@ Level 4，分辨率为 1080P，帧率为 30fps，编码码率约 3Mbps。

点播视频在 5G 与 4G 网络下的 KQI 统计数据如表 3 所示；至于点播视频的首屏时延、卡顿次数和卡顿时长占比的 PDF 分布则分别如图 6、图 7 和图 8 所示。综合来看，5G 网络下点播视频平均首屏时延比 4G 网络下低约 29%，5G 与 4G 网络下平均卡顿次数、平均卡顿时长占比分别相差约 88%和 60%; 从 PDF 分布看 5G 与 4G 网络下首屏时延即点即开 200ms 的满足度、即点即开 400ms 的满足度，以及无卡顿样本占比则分别相差约 713%、10%和 10%。

表 3 国内某一线城市现网路测点播视频 5G/4G 下 KQI 对比

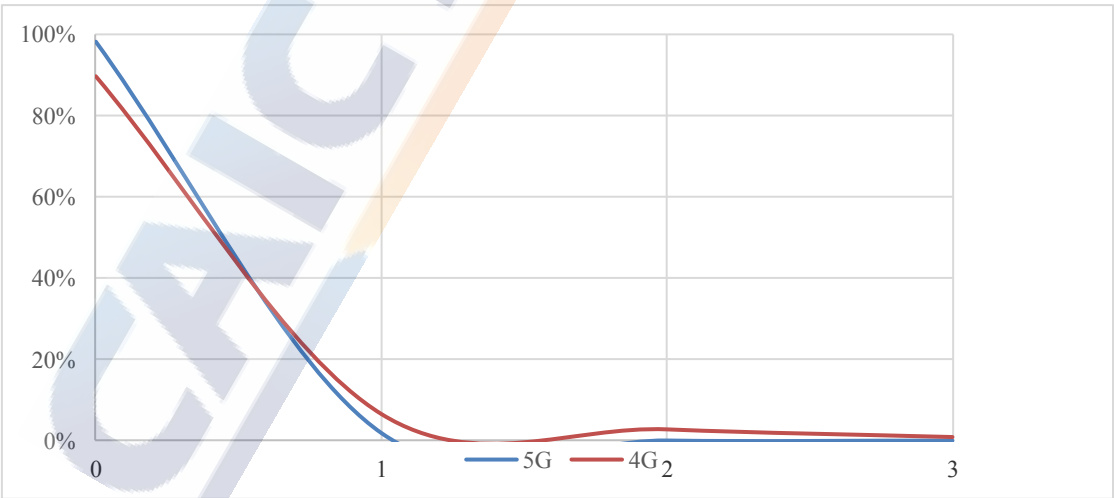
点播视频	首屏时延均值(ms)	200ms满足度	400ms满足度	10%分位值	90%分位值	卡顿次数均值	90%分位值	卡顿时长占比均值	90%分位值
5G网络	243	32.2%	94.9%	180	289	0.0185	0	1.5%	0
4G网络	340	4.0%	86.1%	220	610	0.1509	0	3.7%	0.03
5G相对4G增益	28.7%	713%	10.2%	18.4%	52.6%	87.7%	NA	60.4%	100%

来源：三方联合研究成果



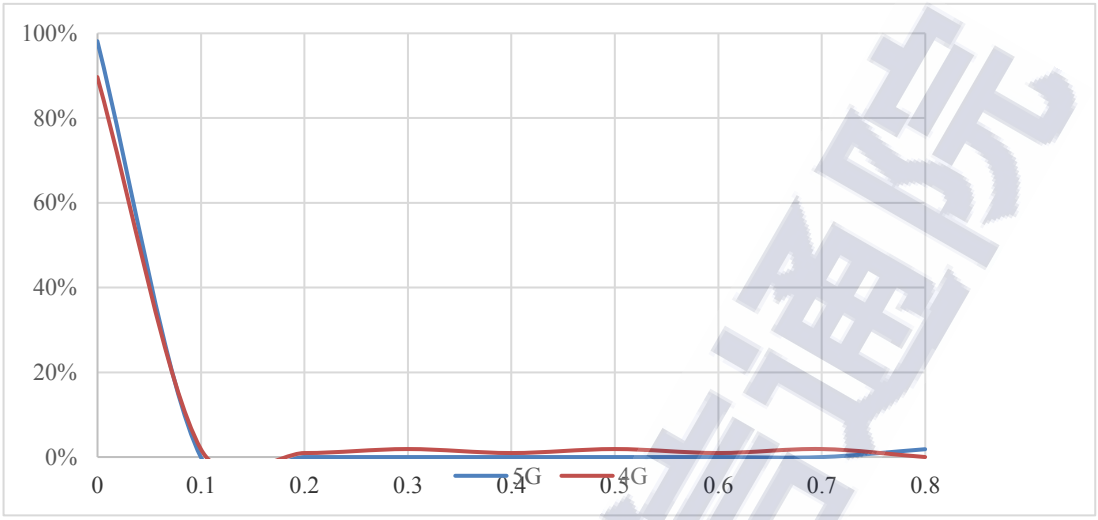
来源：三方联合研究成果

图 5 国内某一线城市现网路测点播视频首屏时延 PDF 分布（ms）



来源：三方联合研究成果

图 6 国内某一线城市现网路测点播视频卡顿次数 PDF 分布



来源：三方联合研究成果

图 7 国内某一线城市现网路测点播视频卡顿时长占比 PDF 分布

关于点播视频交互行为的统计结果如表 4 所示。综合来看，5G 网络下点播视频拖拽导致的卡顿时长占比均值比 4G 网络下低约 72%，5G 网络下拖拽响应时延均值比 4G 网络低约 35%；此外，5G 网络下点播视频以 1.5 倍速播放导致的卡顿时长占比均值比 4G 网络低 100%。

表 4 国内某一线城市现网路测点播视频 5G/4G 下交互行为的 KQI 对比

交互行为	卡顿时长占比均值			拖拽响应时延均值(ms)		
	5G 网络	4G 网络	5G 相对 4G 增益	5G 网络	4G 网络	5G 相对 4G 增益
20s 播放时长，拖拽 1 次（从第 10s 拖拽到第 70s）	1.18%	4.22%	72.1%	940	1440	34.8%
30s 播放时长，1.5 倍速播放	0	1.29%	100%	不涉及		

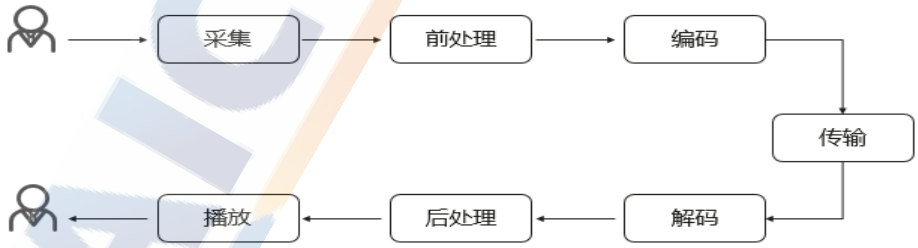
来源：三方联合研究成果

综上所述，5G 与 4G 网络相比，当无交互行为时，点播视频卡顿时长占比平均增益约为 60%；当 1.5 倍速播放时，点播视频卡顿时长占比平均增益为 100%；当拖拽 1 次播放时，点播视频卡顿时长占比平均增益约为 72%，拖拽响应时延平均增益约为 35%。

（二）实时音视频类业务

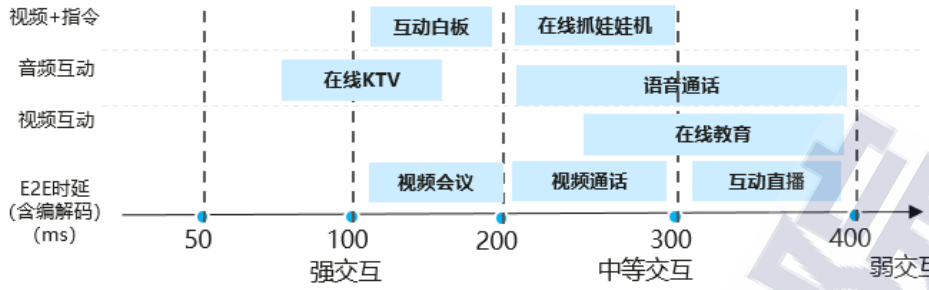
随着在线教育、电商直播、在线视频会议、远程医疗、泛娱乐社交等 App 的普及，实时音视频技术的应用需求也越来越多元化，所谓实时音视频（RTC：Real Time Communication）类业务，它的最大特点是业务互动性较强，对于低时延、无卡顿的要求更高。如图 8 所示，从功能和流程上来说，它包含采集、编码、前后处理、传输、解码、缓冲、渲染等诸多环节。

如图 9 所示，实时音视频的发展趋势为从中等交互（如视频通话、视频会议等）向强交互（如在线 KTV、互动白板等）演进。本文以视频通话/会议/连麦直播、在线 KTV 合唱和在线互动白板为例，重点研究 RTC 类音视频业务的用户体验。



来源：公开信息整理

图 8 实时音视频系统总体架构框图



来源：公开信息整理

图 9 实时音视频 E2E 交互时延演进趋势

1. 视频通话/会议/连麦直播

（1）应用机理及关键 KQI

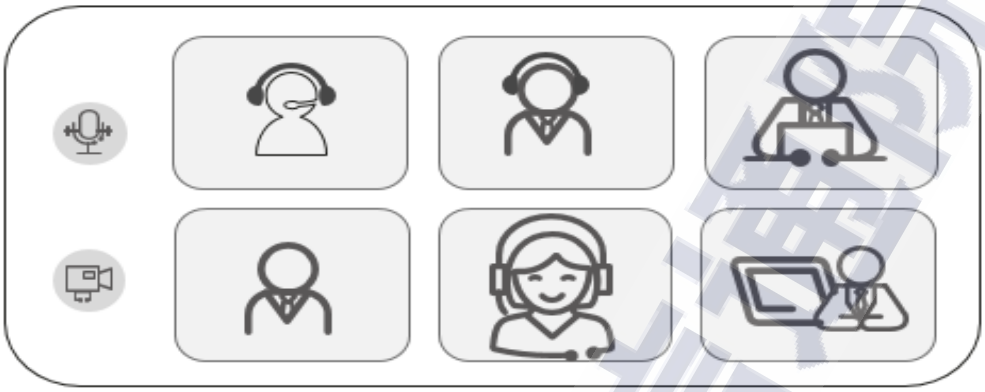
连麦直播主要指互动视频连麦，包括 1 对 1 或者多对多的场景。如图 10 所示，在互动连麦直播间里，为增进直播气氛、快速吸粉，房主可以邀请另一个直播间的房主进行连麦互动或在线 PK，连麦直播间内的观众可以同时观看两个房主互动，并根据房主和房主进行赠送礼物等互动。

来源：公开信息整理

图 10 连麦直播示意

视频通话也包括 1 对 1 和多对多的视频通话模式，适用于视频

聊天、视频会议、视频客服、远程医疗、金融双录、远程定损等场景，如图 11 所示。



来源：公开信息整理

图 11 视频会议示意

总体来说，视频通话、视频会议、连麦直播都是实时音视频类的典型应用，其原理及网络架构都比较类似，如图 12 所示。本文以视频通话为例，分析其特征及应用体验。

来源：公开信息整理

图 12 连麦直播/视频通话/会议架构示意

视频通话/会议/连麦直播的主要 KQI 如表 5 所示。

表 5 视频通话/会议/连麦直播的主要 KQI

KQI 名称	指标说明	极限体验指标	来源
视频卡顿率（%）	远端用户/主播卡顿累计时长占视频有效时长的百分比。其中：在通话过程中，视频帧率设置不低于 5 fps 时，连续渲染的两帧视频之间间隔超过 600 ms，则记为一次视频卡顿	单位时间（1 分钟）内 600ms 视频卡顿率大于 5% 时，记为视频卡顿率不达标	声网“体验等级协议 XLA”[5]
音频卡顿率（%）	远端用户/主播音频卡顿累计时长占音频总有效时长的百分比。其中：通话过程中音频无渲染持续时间超过 200ms 记为一次音频卡顿	单位时间（1 分钟）内 200ms 音频卡顿率大于 3% 时，记为音频卡顿率不达标	声网“体验等级协议 XLA”[5]
端到端时延（ms）	同一频道内用户 A 发出音视频报文到用户 B 接收到此音视频报文的时间	最佳 150ms 以内，最大不超过 400ms	ITU-T G.114 [6]
音画同步时延（ms）	播放器正在渲染的每一帧画面和正在播放的每一段声音（即肉眼所见和人耳所听）之间时间戳的差值	可接受范围：音频超前视频 ≤90ms 且音频落后视频 ≤185ms 之间	ITU-R BT.1359-1 [7]

来源：公开信息整理

（2）测试结果分析

测试环境：运营商 5G/4G 精品网示范区（含商业街、学校、旅游景点等典型场景），测试时段涵盖网络忙闲时（7:00-21:00）。本次视频通话测试，采集端帧率为 30fps，传输协议是 UDP+FEC，编码

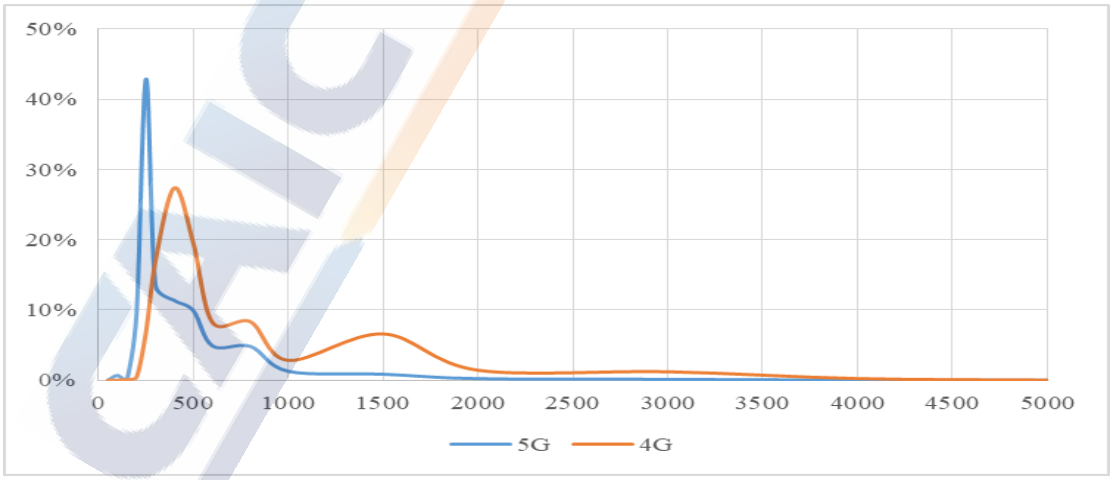
方式为 H.264，编码后分辨率为 1280*720，帧率为 15fps，编码目标码率为 2Mbps。

视频通话在 5G 与 4G 网络下的 KQI 统计数据如表 6 所示，E2E 时延的 PDF 分布如图 13 所示。综合来看，5G 网络下视频通话的音频卡顿达标率比 4G 网络高约 21%，5G 网络下视频通话的无卡顿样本 E2E 时延比 4G 网络低约 24%；从 PDF 分布看，5G 网络下视频通话的 E2E 时延 400ms 满足度比 4G 网络高约 50%。

表 6 国内某省会城市运营商 5G/4G 下视频通话 KQI 对比

视频通话	无卡顿样本 E2E 时延均值(ms)	E2E 时延 200ms 满足度	E2E 时延 300ms 满足度	E2E 时延 400ms 满足度	音频卡顿达标率	有卡顿样本卡顿时长均值(s)	有卡顿样本卡顿率均值
5G 网络	412.60	10.40%	66.72%	78.06%	85.05%	11.45	1.42%
4G 网络	541.81	0.54%	24.49%	51.89%	70.37%	19.71	3.26%
5G 相对 4G 增益	23.8%	1827.6%	172.5%	50.4%	20.9%	41.9%	56.4%

来源：三方联合研究成果



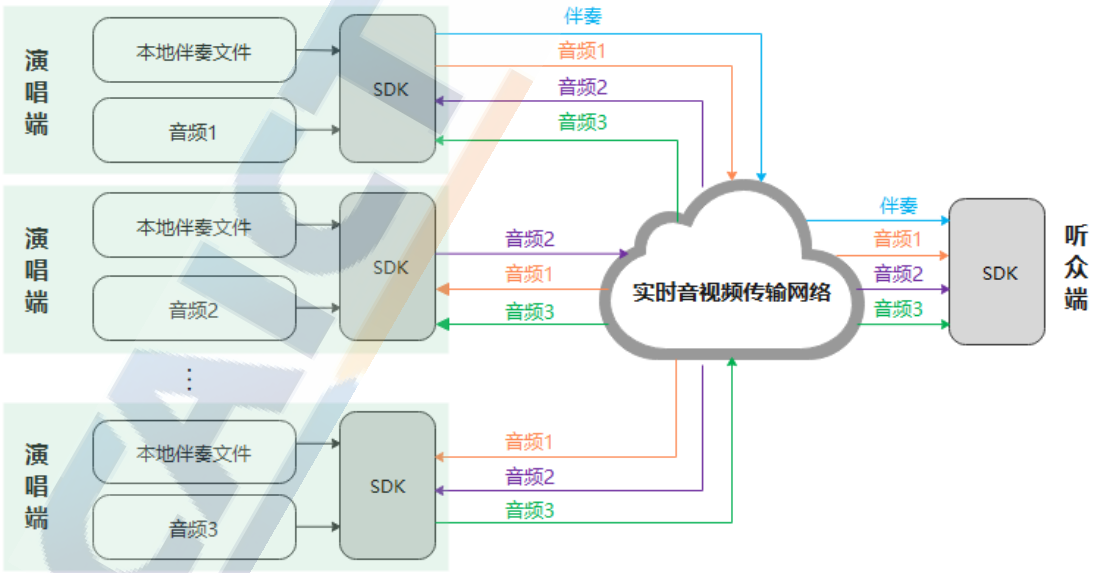
来源：三方联合研究成果

图 13 国内某省会城市现网路测视频通话的 E2E 时延 PDF 分布（ms）

2. 在线 KTV 合唱

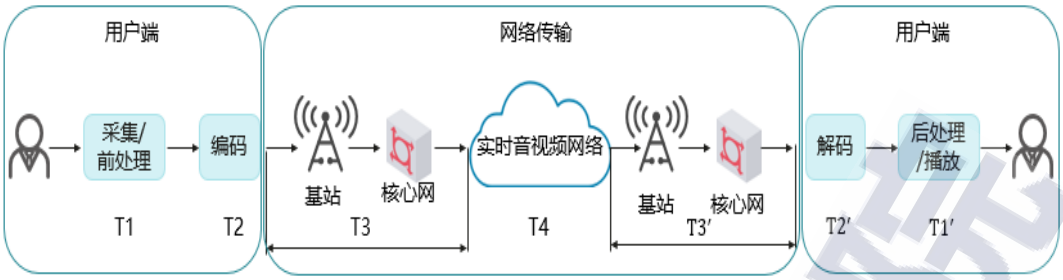
（1）应用机理及关键 KQI

随着互联网娱乐的发展，在线互动社交场景越来越受到人们的青睐，其中在线 KTV 实时合唱就是其中的一种在线娱乐强互动社交方式。实时合唱技术方案如图 14 所示，各个演唱者分别从本地获取 BGM（Background Music，伴奏），随着伴奏同时开启演唱；通过实时音视频网络传输，演唱者们可以实时听到其他人的歌声，达成合唱；同时观众可以享受到演唱者们“0 延时”的合唱效果。实时合唱的关键 KQI 为：端到端时延，其中端到端时延是指从演唱端系统采集声音到声音从听众播放端输出的单向时延，包括：音频在采集端和播放端的延时（即采集端的采集、前处理、编码，播放端的接收、解码、后处理过程产生的延时，以及两端（编码后和解码前）产生的网络延时），如图 15 所示。



来源：公开信息整理

图 14 实时合唱技术方案架构



来源：公开信息整理

图 15 实时合唱端到端时延构成

根据“声网实时合唱解决方案”[8]和《One World Together，线上实时合唱技术解析》[9]，实时合唱端到端时延追求的理想目标值是50ms；若实时合唱端到端时延控制在 150ms 以内，则合唱双方听到的对端歌声和自身感觉一致，不会影响到演唱。如表 7 所示。

表 7 实时合唱端到端时延要求

理想要求（ms）	基本要求（ms）
50	150

来源：声网

如表 8 所示，实时合唱当前采集、前处理、后处理、播放等几部分的延时可以做到 50ms 左右，编解码时延典型值为 15ms 左右；而实时音视频服务器网络时延优秀情况下可以做到 30ms 以内，大部分可以做到 50ms 以内。所以当实时音视频服务器网络时延为 30ms 时，若要达到实时合唱的效果，则运营商网络传输时延建议不高于 60ms；当实时音视频服务器网络时延为 50ms 时，若要达到实时合唱的效果，则运营商网络传输时延建议不高于 40ms。

表 8 实时合唱各个部分延迟当前业界水平

阶段	指标	延时典型值（ms）
T1、T1’	采集、前处理、后处理、播放	50
T2、T2’	编码、解码	15
T3、T3’、T4	网络传输时延（包含运营商网络传输时延 T3、T3’，以及实时音视频网络时延 T4）	实时音视频网络时延 T4 国内当前的水平： 优秀：＜30; 基本：＜50

来源：公开信息整理

（2）测试结果分析

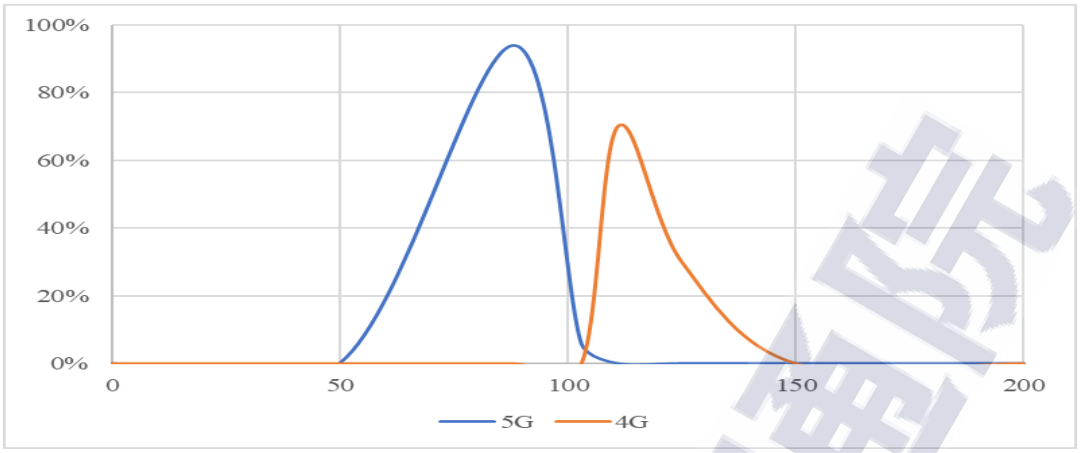
测试环境：运营商 5G/4G 精品网示范区（含交通枢纽、地铁、学校、写字楼等典型场景），测试时段涵盖网络忙闲时（7:00-21:00）。

在线 KTV 合唱在 5G 与 4G 网络下的 KQI 统计数据如表 9 所示，端到端时延的 PDF 分布如图 16 所示，运营商网络传输时延的 PDF 分布如图 17 所示。综合来看，5G 网络下在线 KTV 合唱的端到端时延均值比 4G 网络低约 24%；从 PDF 分布看，5G 网络下在线 KTV 中运营商网络传输时延 60ms 满足度比 4G 网络高约 19%，运营商网络传输时延 40ms 满足度比 4G 网络高约 86%（绝对值）。

表 9 国内某一线城市运营商 5G/4G 下在线 KTV 合唱 KQI 对比

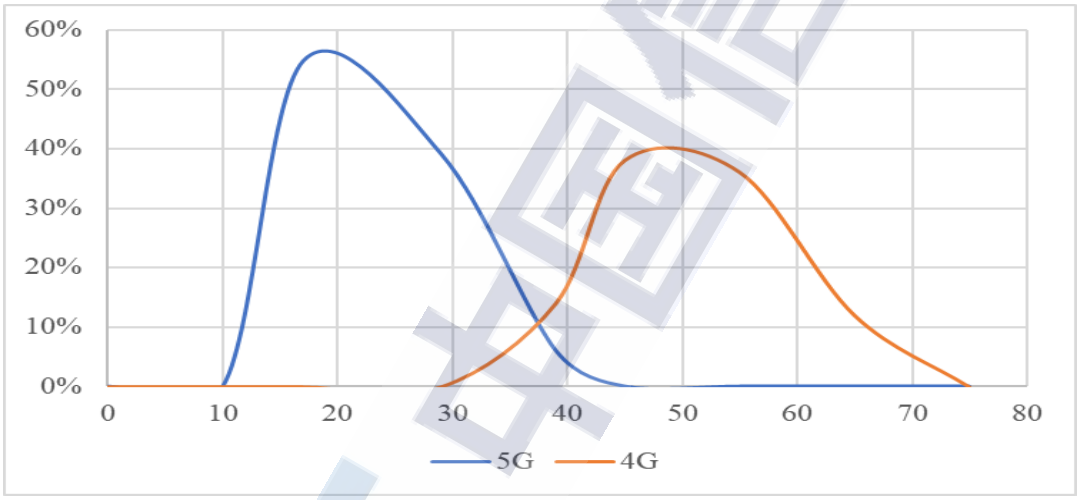
在线 KTV 合唱	E2E 时延均值(ms)	运营商网络传输时延 60ms 满足度	运营商网络传输时延 40ms 满足度
5G 网络	88.29	100%	100%
4G 网络	115.79	84%	14%
5G 相对 4G 增益	23.75%	19.05%	86%（绝对值）

来源：三方联合研究成果



来源：三方联合研究成果

图 16 国内某一线城市现网路测在线 KTV 合唱的端到端时延分布（ms）



来源：三方联合研究成果

图 17 国内某一线城市现网路测在线 KTV 合唱的网络传输时延分布（ms）

3. 互动白板

（1）应用机理及关键 KQI

互动白板提供丰富的多人实时白板互动服务，如白板涂鸦，实时轨迹同步，文档共享、文件转码、白板录制与回放、白板与实时音视频同步等，可应用于在线教育、会议协作、金融面签、游戏互动等多种场景。

多端实时互动是白板用户的业务体验的关键影响因素，从网络传输的角度，互动白板的主要 KQI 为：E2E 互动响应时延（用户 A 操作到用户 B 显示的时延）。

互动白板传输有两种方式：基于视频流和基于文件+信令两种方式。文件+信令在宽带占用、清晰度、互动性等方面具有明显的优势，所以该模式越来越成为业界的主流技术方式（如声网、即构等）。本文研究的传输模式为文件+信令传输模式。文件+信令的互动白板原理架构如图 18 所示。

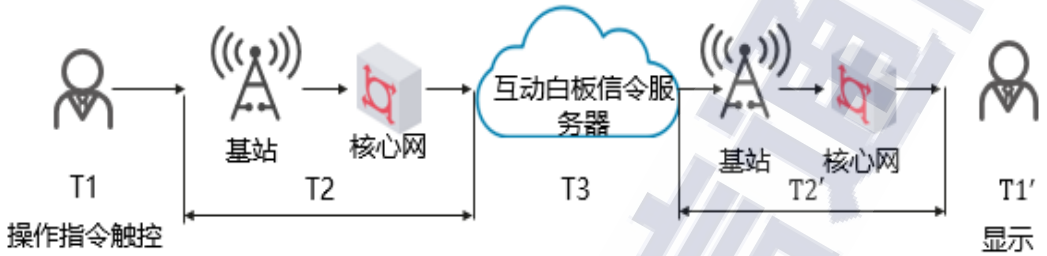
来源：公开信息整理

图 18 互动白板文件+信令方案架构图

通过互动白板信令服务器可以进行多端的实时互动，如标注，涂鸦等。根据《互动协作白板与音视频实时同步技术实践》[10]，目前主流厂家互动白板的 E2E 互动响应时延均可实现不超过 100ms，而如果时延超过 200ms，则多端实时互动会感受到不同步。

互动白板的 E2E 互动响应时延可以分为操作指令触控时延、渲染显示时延、互动白板信令服务器传输时延及运营商网络传输时延

几部分组成。如图 19 所示。如表 10 所示，当前操作指令触控及渲染显示时延可以做到 55ms 左右；互动白板信令服务器传输时延国内可以做到 100ms 以内。综合来看，为了满足互动白板实时互动用户体验，运营商网络传输时延建议控制在 50ms 以内。



来源：公开信息整理

图 19 互动白板 E2E 互动响应时延构成

表 10 互动白板各个部分延迟当前水平

阶段	指标	延时典型值（ms）
T1、T1’	操作指令触控时延、渲染显示时延	~55
T2、T2’、T3	网络传输时延（包含运营商网络传输时延 T2、T2’，以及互动白板信令服务器传输时延 T3）	互动白板信令服务器传输时延 T3，国内：＜100

来源：公开信息整理

（2）测试结果分析

测试环境：运营商 5G/4G 精品网示范区（含交通枢纽、地铁、学校、写字楼等典型场景），测试时段涵盖网络忙闲时（7:00-21:00）。

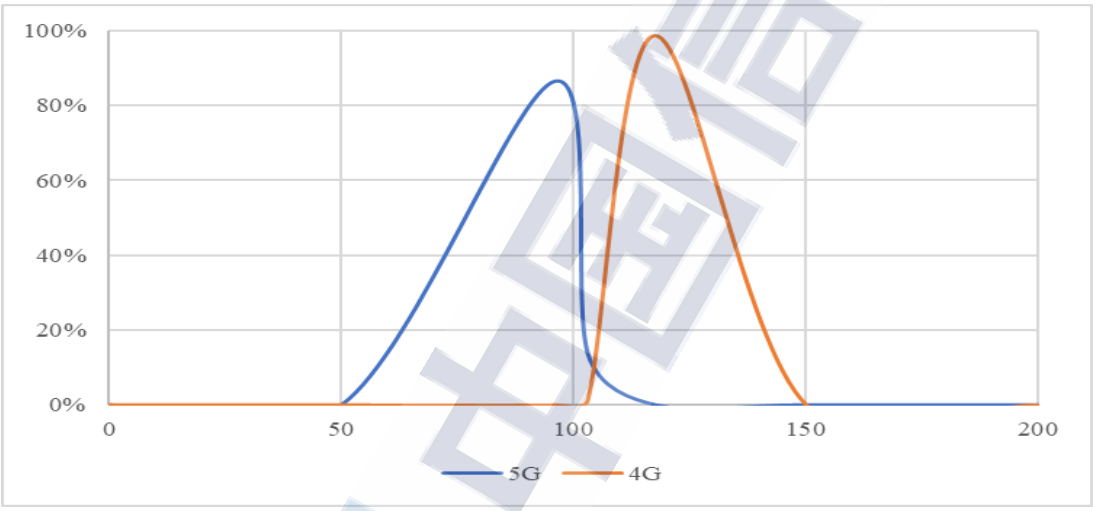
互动白板在 5G 与 4G 网络下的 KQI 统计数据如表 11 所示，E2E 互动响应时延 PDF 分布如图 20 所示。综合来看，5G 网络下互动白板的 E2E 互动响应时延比 4G 网络下低约 17%；从 PDF 分布看，5G 网络下互动白板的 E2E 互动响应时延 100ms 满足度比 4G 网络下高约 86%（绝对值），运营商网络传输时延 50ms 满足度 5G 比 4G 高约

95%（绝对值）。

表 11 国内某一线城市现网路测互动白板 5G/4G 下 KQI 对比

互动白板	E2E 互动响应时延均值(ms)	E2E 互动响应时延 100ms 满足度	运营商网络传输时延 50ms 满足度
5G 网络	96.98	86.43%	98.38%
4G 网络	117.29	0.47%	2.97%
5G 相对 4G 增益	17.32%	85.96%(绝对值)	95.41%(绝对值)

来源：三方联合研究成果



来源：三方联合研究成果

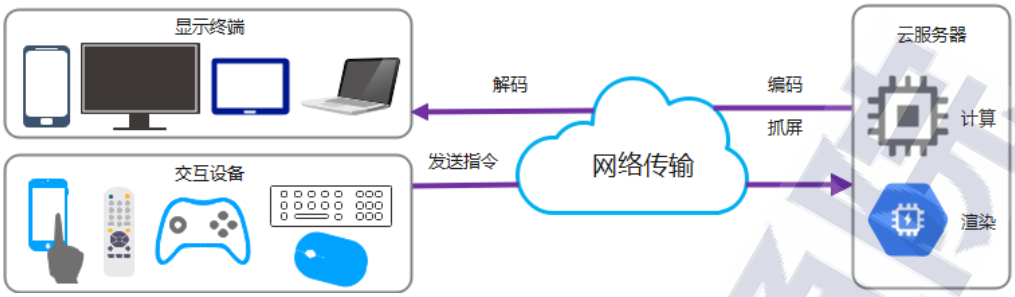
图 20 国内某一线城市现网路测互动白板 E2E 互动响应时延 PDF 分布（ms）

（三）云渲染业务

1. 应用机理及关键 KQI

云渲染类业务主要特征是，将需要强大计算和渲染能力的部分放到云服务器端，客户端只需要具备一定的流媒体解码和播放能力即可，如图 21 所示。以云游戏（Cloud Gaming）为例，它是以云计算为基础的游戏方式，本质上为交互性的在线视频流，在云游戏的运行模式下，游戏在云端服务器上运行，并将渲染完毕后的游戏画

面或指令压缩后通过网络传送给用户。



来源：公开信息整理

图 21 云游戏原理示意

根据《云游戏体验模型白皮书》[11]，并结合网络传输角度，云游戏的主要 KQI 为：卡顿次数，平均卡顿时长和操作响应时延。

根据“ETSI TR 126 928 V16.1.0 (2021-01)”[12]，强交互、中等交互和弱交互云游戏的操作响应时延门限分别为：100ms、500ms 和 1000ms。

2. 测试结果分析

测试环境：运营商 5G/4G 精品网示范区（含交通枢纽、地铁、学校、写字楼等典型场景），测试时段涵盖网络忙闲时（7:00-21:00）。本次云游戏测试选取的内容源是《世界汽车拉力锦标赛 8》（强交互），传输协议是 UDP+FEC，编码方式是 H.264，分辨率为 1080P，帧率为 60fps，编码目标码率为 10Mbps；实时码率根据网络状况自适应调整。

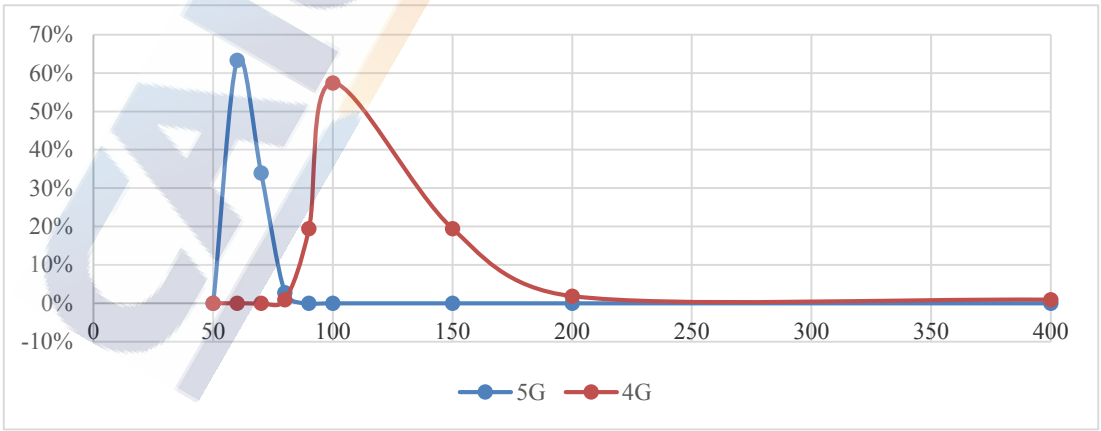
云游戏（强交互）在 5G 与 4G 网络下的 KQI 统计数据如表 12 所示；至于云游戏的操作响应时延、卡顿次数和卡顿时长的 PDF 分布则分别如图 22、图 23 和图 24 所示。综合来看，5G 网络下云游戏

平均操作响应时延比 4G 网络下低约 40%，5G 与 4G 网络下平均卡顿次数/卡顿时长相差约 75%；从 PDF 分布看 5G 与 4G 网络下操作响应时延 100ms 的满足度相差约 29%，无卡顿样本占比相差约 22%。

表 12 国内某一线城市现网路测云游戏 5G/4G 下 KQI 对比

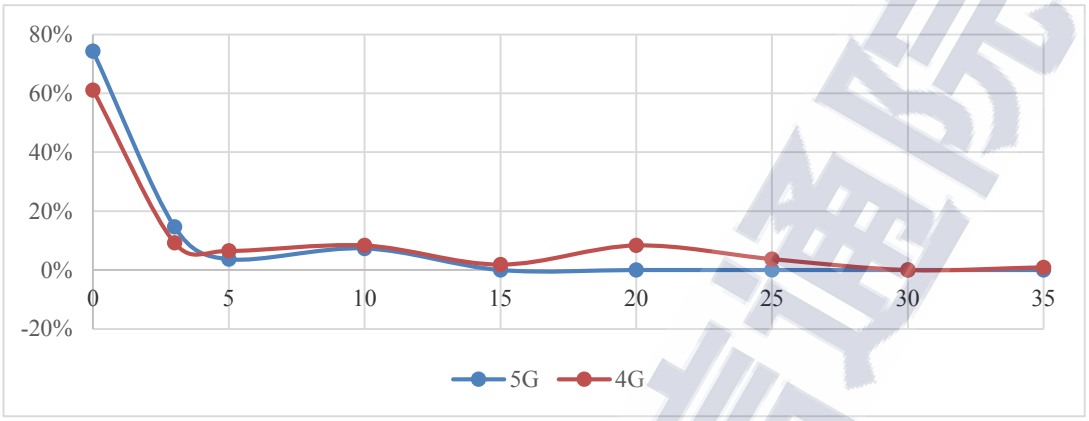
云游戏	操作响应时延均值(ms)	100ms 满足度	90%分位值	10%分位值	卡顿次数均值	90%分位值	卡顿时长均值(ms)	90%分位值
5G 网络	59.40	100%	63.74	55.68	1.02	4.00	203.67	800
4G 网络	98.39	77.8%	102.6	87.40	4.01	16.20	801.85	3240
5G 相对 4G 增益	39.6%	28.5%	37.9%	36.3%	74.6%	75.3%	74.6%	75.3%

来源：三方联合研究成果



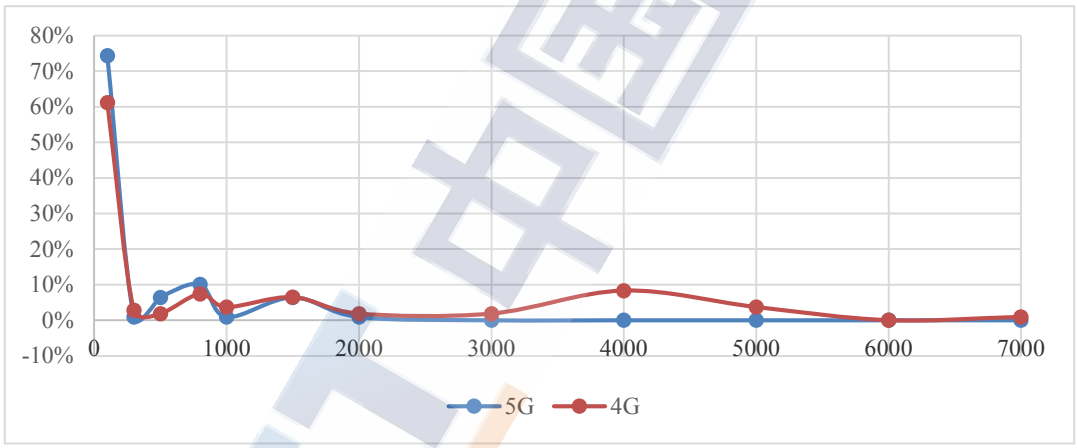
来源：三方联合研究成果

图 22 国内某一线城市现网路测云游戏操作响应时延 PDF 分布（ms）



来源：三方联合研究成果

图 23 国内某一线城市现网路测云游戏卡顿次数 PDF 分布

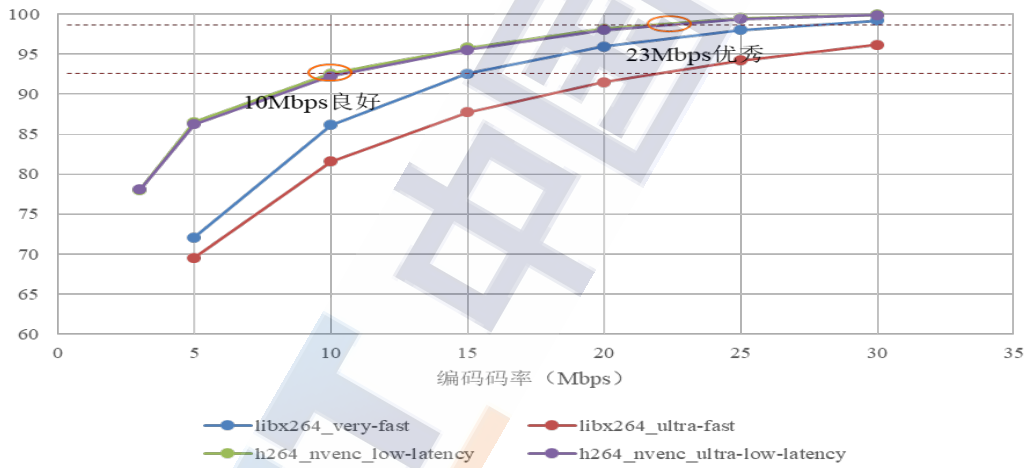


来源：三方联合研究成果

图 24 国内某一线城市现网路测云游戏卡顿时长 PDF 分布（ms）

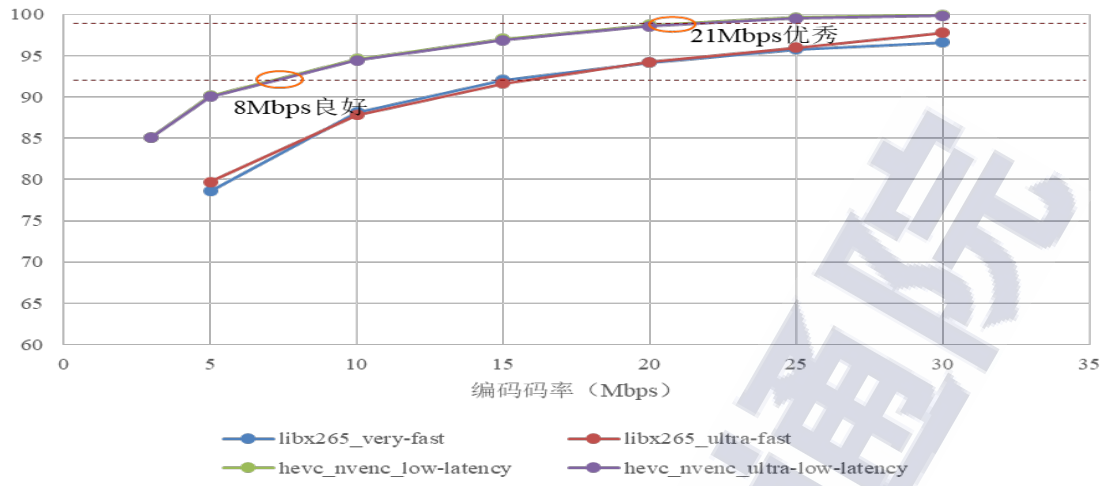
此外，为了厘清不同编解码器、编码方式和码率对云游戏画质的影响，从而在经济性和用户体验之间实现合理平衡，选取部分典型游戏的活泼明亮(Active & Bright)场景，利用 VMAF 工具[13](Video Multi-Method Assessment Fusion，视频多方法评估融合，系 Netflix 提供的感知视频质量评估算法开源项目）针对在一定编码器和编码方

式条件下不同编码码率的画质体验损伤（基于手机观看时）进行了对比测试，当基于 H.264 和 H.265 编码时其统计数据分别如图 25 和 26 所示。综合分析可知：1）在相同的编码码率下，基于 NVENC 硬件编码器的画质比基于软件编码器有明显提升；2）在相同的编码码率下，H.265 比 H.264 的画质略有提升；3）基于 NVENC 硬件编码器时，建议 1080P60 云游戏码率范围如下：不低于 10Mbps 为良好画质（手机端观看时对应 VMAF 为 93 分[14]）；不低于 20Mbps 为优秀画质（手机端观看时对应 VMAF 为 99 分，根据技术博客[15]，VMAF 6 分是一个最小可觉差 JND）。



来源：三方联合研究成果

图 25 1080P60 云游戏基于 H.264 编码时良好和优秀画质对应的码率范围



来源：三方联合研究成果

图 26 1080P60 云游戏基于 H.265 编码时良好和优秀画质对应的码率范围

根据上述研究结果，如果将 1080P60 云游戏的编码目标码率由 10Mbps 增加到 20Mbps，那么可以预期 5G 网络下云游戏的用户体验优势将更加凸显。

（四）空间视频

1. 自由视角视频

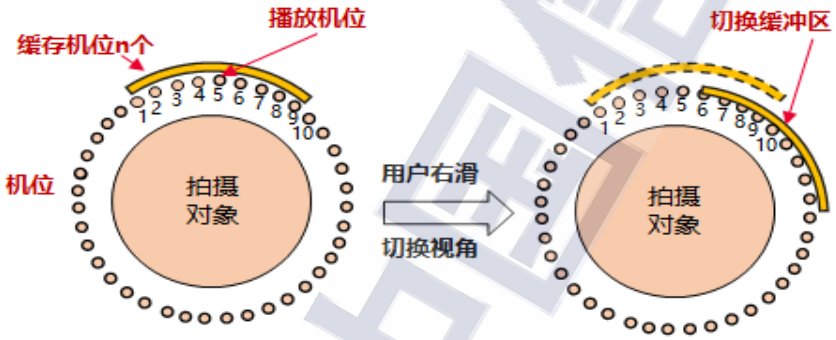
（1）应用机理及关键 KQI

自由视角是指以被拍摄物体/场景为中心，能够让用户通过操控界面，从任意角度自由旋转观看被拍摄物体/场景的业务形态。传统现场赛事直播因角度、机位限制，或受导播意志左右，观众往往不能随心所欲地选择观看自己最喜欢的角度或细节。自由视角完美解决了这一难题，将自主选择权交还用户，让用户随心而动、以交互方式自由旋转,“转”哪看哪，尽情窥探舞台中央的“隐私”。

自由视角方案一般通过增加视点的方法实现自由观看，采集端每个采集设备作为一个真实视点，用户滑动时通过切换视点的方式

选择自己喜欢的视角。为了保证用户旋转时的平滑与流畅，建议视点间的角度不大于 5 度，视点总度数与拍摄现场的部署相关。

如图 27 所示，在本测试系统实现方案中自由视角播放端使用动态缓存的方式，缓存的大小可以设置为 0%-100%。用户在观看视频的同时下载多路视频流，客户端根据用户观看的视点动态调整缓存的机位。这种动态调整的方式可以实现无死角观看，如果采集端使用 360 度拍摄，播放端即可实现 360 度观赏。



来源：公开信息整理

图 27 本测试系统的自由视角实现方案

当用户通过手机滑屏播放自由视角视频时存在两种情况。第 1 种情况，用户在滑屏操作过程中相关机位内容已提前下载到本地且解码完成，可以做到跟随旋转（即：视角切换响应时延在 100ms 以内。视角切换响应时延定义：从手指滑屏开始计时，到手机屏幕的视频画面开始旋转（即切换视角）结束，二者之间的时延）；第 2 种情况，用户在滑屏操作过程中相关机位内容未提前下载完成，需要实时向网络侧请求数据，则相关机位内容的下载、解码和渲染显示的时延，以及网络带宽需要满足特定的要求才能做到跟随旋转。

一方面，假设缓存机位数为 N，用户的单位视角切换（即相邻机位切换）时延为 TUVS，相关机位内容的下载时延为 TNw，视频解码和渲染显示的时延为 TDRD，则为了保障用户切换视角平滑的网络下载时延需要满足：$(N-1)/2 * T_{UVS} \geq T_{DRD} + T_{Nw}$；另一方面，假设网络带宽为 BW，单路视频码率为 Br，则为了保障用户切换视角平滑的网络下载带宽需要满足：$BW \geq N * 1.5 * Br$。结合上述两个条件可知用户切换视角平滑的带宽和时延要求为：

$$\frac{BW}{1.5 * Br} \geq 2 * \frac{T_{DRD} + T_{Nw}}{T_{UVS}} + 1$$

为了保障视角切换时的用户体验，建议 N>=9。

自由视角视频的主要 KQI 为：首屏时延，卡顿次数，卡顿时长，以及视角切换响应时延。

（2）测试结果分析

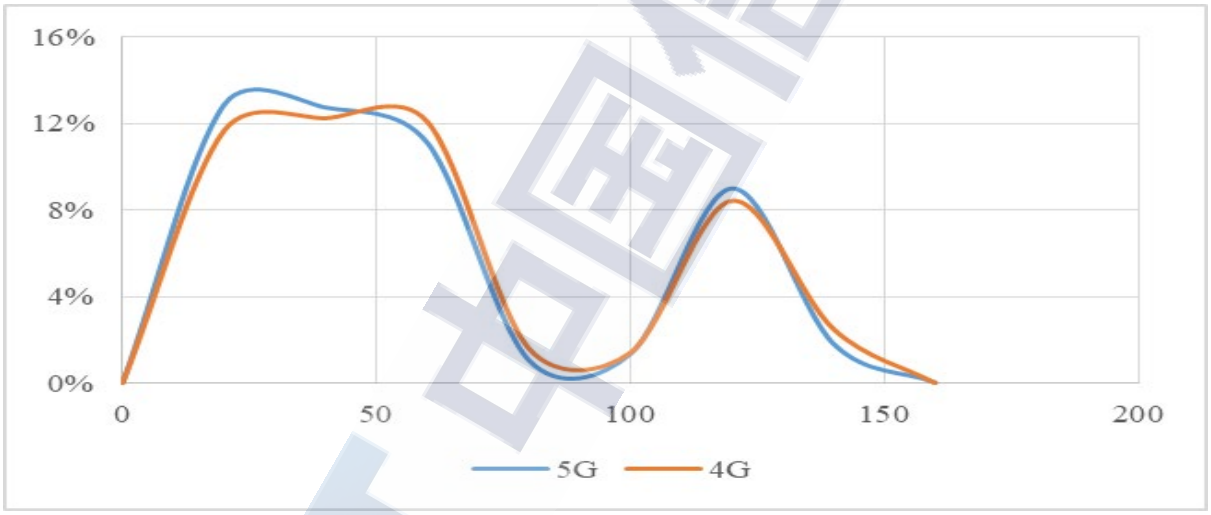
测试环境：运营商 5G/4G 精品网示范区（含交通枢纽、地铁、学校、写字楼等典型场景），测试时段涵盖网络忙闲时（7:00-21:00）。本次自由视角测试选取的内容源参数是：视频传输协议是 HLS，编码方式为 H.265，分辨率为 720P，帧率为 30fps，编码码率为 9*1Mbps。

自由视角视频在 5G 与 4G 网络下的 KQI 统计数据如表 13 所示；至于视角切换响应时延、首屏时延的 PDF 分布如图 28 和图 29 所示。综合来看，5G 网络下切换视角响应时延比 4G 网络下低约 19%，5G 与 4G 网络下首屏时延均值和平均卡顿次数分别相差约 60%和 14%；但从 PDF 分布看 5G 与 4G 网络下无卡顿样本占比相差约 2%。

表 13 国内某一线城市现网路测自由视角视频 5G/4G 下 KQI 对比

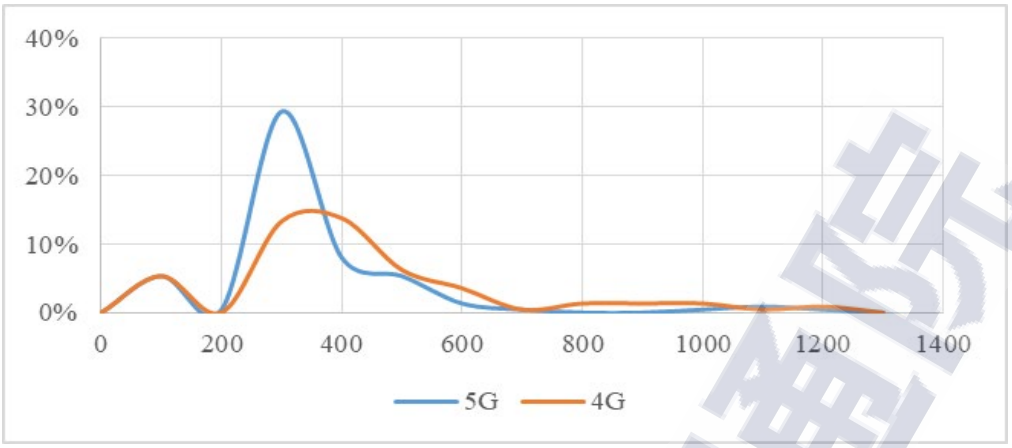
自由视角视频	切换视角响应时延均值(ms)	100ms 满足度	90%分位值	10%分位值	首屏时延均值(ms)	90%分位值	10%分位值	卡顿次数均值	90%分位值
5G 网络	49.19	86%	112	16	462	477	199	0.11	1
4G 网络	58.47	70%	118	16	738	1165	202	0.125	1
5G 相对 4G 增益	18.9%	23%	5.4%	0%	59.6%	144%	1.5%	13.6%	0%

来源：三方联合研究成果



来源：三方联合研究成果

图 28 国内某一线城市现网路测 5G/4G 下视角切换响应时延分布（ms）



来源：三方联合研究成果

图 29 国内某一线城市现网路测 5G/4G 下首屏时延分布（ms）

2. 手机全景视频

（1）应用机理及关键 KQI

根据《云游戏体验模型白皮书》[11]中的等效 PPD 计算方法，以市面主流的 6 英寸左右手机（如 Mate 40 Pro）为例，在 30cm 观看距离下，手机观看快手 App 和移动云 VR App 的全景视频内容时 PPD 及有效分辨率如表 14 所示。

表 14 全景视频内容的 PPD 及有效分辨率

平台	视频源分辨率	横向观看 FOV[1]	纵向观看 FOV[2]	横向 PPD[3]	纵向 PPD4[4]	有效分辨率
快手 App	2160*1080	30.38	14.34	17	37	540*540
移动	4096*2048	30.38	14.34	33	35	1024*512

[1] 横向 FOV：指当手机横屏观看时，眼睛与视频内容左右两侧中心位置连线的夹角（如果内容铺满全屏，则为手机高度方向显示屏左右两侧中心位置，因为此时视频内容与显示屏的左右两侧中心位置正好重合）。
[2] 纵向 FOV：指当手机竖屏观看时，眼睛与视频内容左右两侧中心位置连线的夹角（如果内容铺满全屏，则为手机宽度方向显示屏左右两侧中心位置，因为此时视频内容与显示屏的左右两侧中心位置正好重合）。
[3] 横向 PPD：指当手机横屏观看时，视频内容水平像素数（即手机宽度方向）与横向 FOV 的比值。
[4] 纵向 PPD：指当手机竖屏观看时，视频内容水平像素数（即手机高度方向）与纵向 FOV 的比值。

平台	视频源分辨率	横向观看 FOV[1]	纵向观看 FOV[2]	横向 PPD[3]	纵向 PPD4[4]	有效分辨率
云 VR App	（4K）					
	3840*3840（4K）	30.38	14.34	31	66	960*960

来源：公开信息整理

手机全景视频的主要 KQI 为：首屏时延，卡顿次数，卡顿时长。

（2）测试结果分析

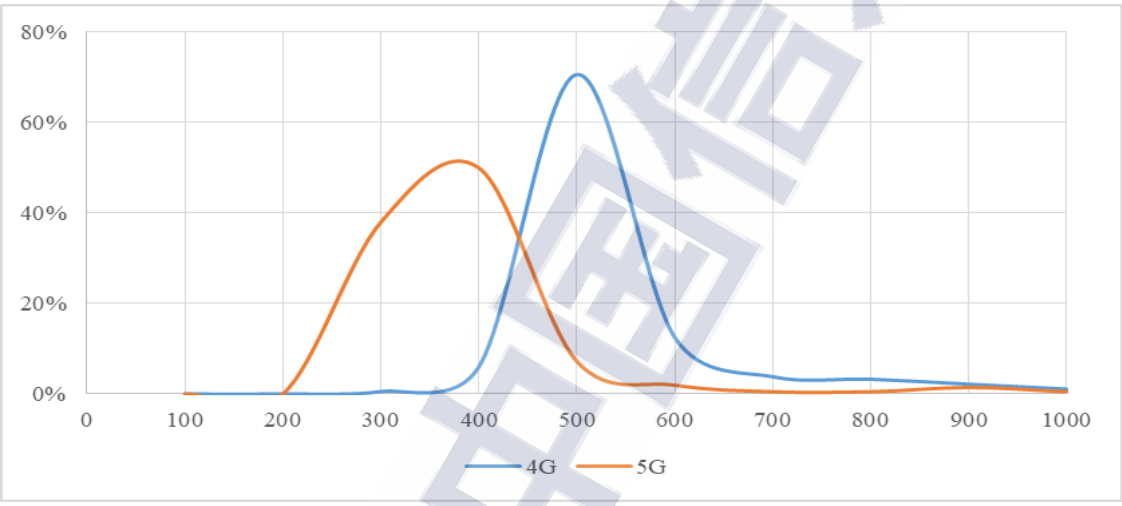
测试环境：运营商 5G/4G 精品网示范区（含交通枢纽、地铁、学校、写字楼等典型场景），测试时段涵盖网络忙闲时（7:00-21:00）。本次手机全景视频测试选取的内容源：《好身材厨房》，视频传输协议是 HLS，编码方式为 H.265，分辨率为 3840*1920，帧率为 25fps，编码码率为 10.6Mbps。

手机全景视频在 5G 与 4G 网络下的 KQI 统计数据如表 15 所示；至于首屏时延的分布则如图 30 所示。综合来看，5G 网络下手机全景视频（10M）的首屏时延比 4G 网络下低约 30%，5G 与 4G 网络下平均卡顿次数/卡顿时长相差约 75%；从 PDF 分布看 5G 与 4G 网络下首屏时延即点即开 400ms 的满足度，以及无卡顿样本占比分别相差约 1249%、8.3%。

表 15 国内某一线城市现网路测手机全景视频（10M）5G/4G 下 KQI 对比

全景视频(10M)	首屏时延均值(ms)	400ms满足度	90%分位值	10%分位值	卡顿次数均值	90%分位值	卡顿时长均值(ms)	90%分位值
5G 网络	336	88.0%	460	276	0.025	1	8297	14251
4G 网络	481	6.5%	649	404	0.1	2	9839	15003
5G 相对 4G 增益	30.1%	1249%	29.1%	31.7%	75%	50%	74.6%	5.3%

来源：三方联合研究成果



来源：三方联合研究成果

图 30 国内某一线城市现网路测手机全景视频（10M）首屏时延分布（ms）

此外，为了厘清一定编码方式（H.265）下码率对单目 4K 全景视频画质的影响，从而在经济性和用户体验之间实现合理平衡，选取 JVET 发布的全景视频测试序列中部分典型内容场景，针对不同编码码率的画质体验损伤（基于手机观看时）进行了主观体验测试，其统计数据如表 16 所示。

表 16 单目 4K 全景视频不同体验等级对应的推荐码率

MOS 范围	BPP(bit/pixel)	码率(Mbps)	单目 4K 全景视频体验等级
3.02 - 3.37	0.04	10	入门级

MOS 范围	BPP(bit/pixel)	码率(Mbps)	单目 4K 全景视频体验等级
3.60 - 3.81	0.1	25	良好
3.90 - 4.26	0.25	62	优秀

来源：三方联合研究成果

（五）结论

1. 音视频交互响应要求

视频应用的社交化与交互化发展过程中，时延是影响用户体验的重要指标。根据本章关于音视频应用交互响应的相关研究，不同音视频应用的交互响应要求归纳如表 17 所示。

表 17 音视频应用的交互响应要求

业务大类	交互类型	交互响应操作	交互响应要求（E2E 时延）	
			理想要求	基本要求
点播长视频	人机交互	即点即开/拖拽	100ms[3]	400ms[2]
		多视角切换	100ms[16]	500ms[16]
点播短视频	人机交互	即点即开/滑动到下一个	100ms[3]	200ms[3]
实时音视频	人人交互	视频通话交互	150ms[6]	400ms[6]
		白板交互	100ms[10]	200ms[10]
		实时合唱交互	50ms[8]	150ms[9]
业务大类	交互类型	交互响应操作	交互响应要求（E2E 时延）	
			理想要求	基本要求
云游戏	人机交互+人人交互	操控	100ms[17]	275ms[17]

来源：公开信息整理

2. 音视频 5G/4G 差异化体验

最后，针对典型音视频应用在 5G 网络下 KQI 相对于 4G 网络的增益进行总结，如表 18 所示。一方面，典型音视频应用在 5G 网络下相对于 4G 网络而言，卡顿时长占比平均降低约 70%，交互响应时延平均缩短约 30%。由此可见，交互响应是音视频应用 5G/4G 差异化体验的重要表现形式。另一方面，为了在经济性和用户体验之间实现合理平衡，新应用如云游戏和手机全景视频入门级体验的编码码率建议不低于 10Mbps。

表 18 典型音视频应用在 5G 网络下 KQI 相对于 4G 网络的增益

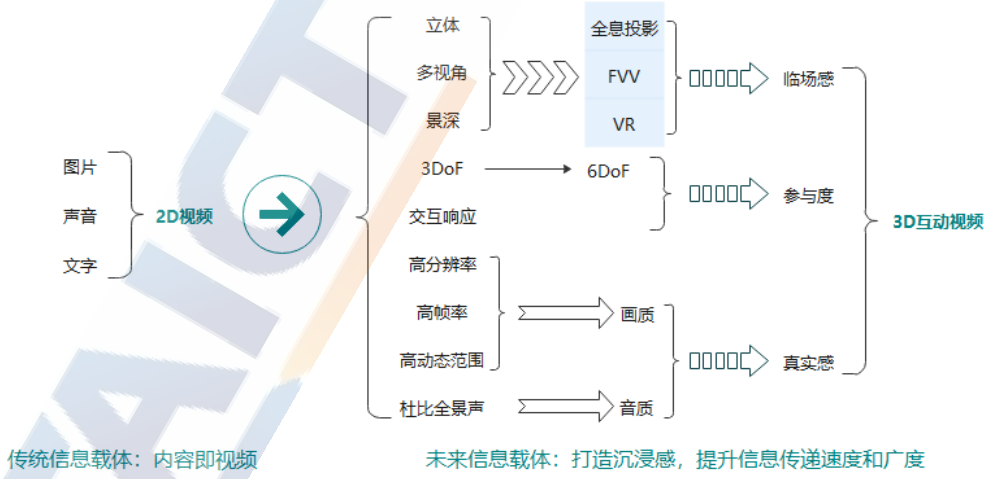
业务类型	首屏时延	卡顿时长占比	无卡顿样本占比/音频卡顿达标率	E2E 时延/操作响应时延/视角切换响应时延	400ms 满足度/100ms 满足度	码率（Mbps）
点播	29%	60%	10%	N/A	10%	3
点播 1.5 倍速播放	N/A	100%	2%	N/A	N/A	
点播拖拽	N/A	72%	6%	35%	N/A	
视频通话	N/A	56%	21%	24%	50%	2
实时合唱	N/A	N/A	N/A	24%	94%（绝对值）	N/A
互动白板	N/A	N/A	N/A	17%	86%（绝对值）	N/A
云游戏	N/A	75%	22%	40%	29%	10
自由视角视频	60%	N/A	2%	19%	23%	9*1
手机全景视频	30%	75%	8%	N/A	1249%	10.6

来源：三方联合研究成果

三、展望和建议

（一）展望

如图 31 所示，视频是由一种由图片、声音、文字组成的信息载体，是信息更加直观、接近真实世界的表达方式；为了贴近真实世界增加用户的沉浸感，视频会向空间化、高清化与交互化方向发展。在这个过程中，信息的交互速度与网络的能力息息相关。空间视频由体素组成的，与 2D 视频相比，空间视频多了一个维度的信息，在空间视频发展的同时未来视频分辨率与帧率的提升也会使视频码率翻倍。除了对于大带宽的需求，人机交互的理想时延在 100ms 内，除采集、编解码、渲染显示等相对固定的分段时延外，分解到网络时延典型只有 20ms 甚至更低，这种极低的时延需求对网络也是一个巨大的挑战。在大量应用共存的情况下，未来的网络会根据应用的需求为不同类型数据建立不同的保障等级，动态调度网络能力，以保障不同场景下的用户体验。

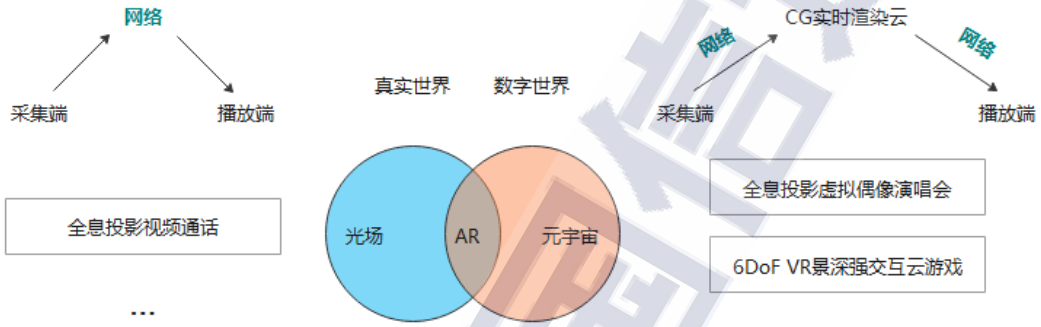


来源：三方联合研究成果

图 31 视频技术发展趋势

如图 32 所示，随着全息和 XR 技术的发展，对于客观世界的记

录与数字世界的建立都为视频带来源源不断的内容，新形态的视频将会给人们带来全新的体验。不久的将来有望成为主角的全息投影虚拟偶像演唱会、全息投影视频通话、6DoF VR 景深强交互云游戏，乃至跨越物理世界与虚拟世界的元宇宙都将成为可能，为了实现未来应用体验的跨越式发展，支持 RTBC 场景（大带宽+低时延+高可靠）能力的 5.5G 网络将大有可为。

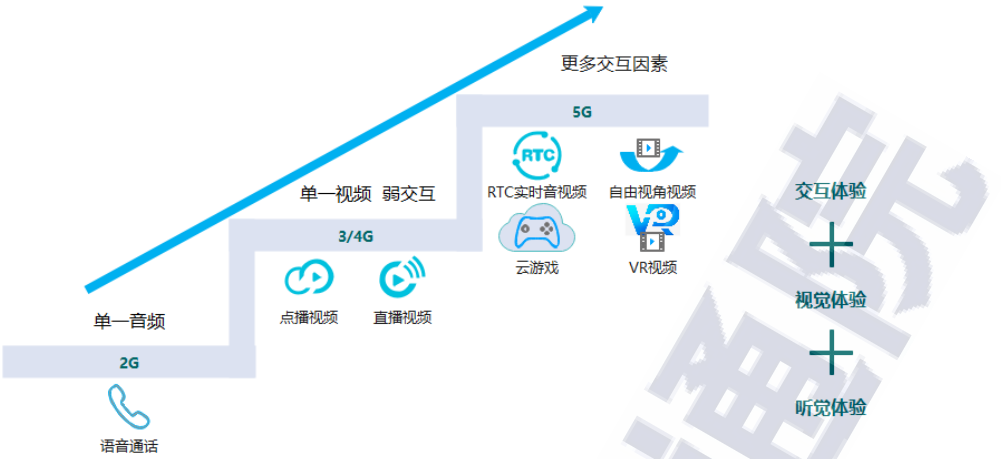


来源：三方联合研究成果

图 32 视频内容多元化

（二）建议

随着无线通信技术从 2G 开始、历经 3G/4G，并持续向 5G 演进，主要音视频业务形态也由语音通话、视频点播/直播，向交互性更强且形态更丰富（如 RTC 实时音视频、云游戏、自由视角视频、VR 视频等）的方向发展，如图 33 所示。



来源：三方联合研究成果

图 33 5G2C 音视频应用演进路线图

一方面，随着点播视频向高清化、交互化方向发展，HDR / HFR、X 倍速播放、拖拽等更多特性的影响需要纳入到点播视频体验评测 KQI 中。另一方面，与点播视频和普通直播视频相比，有 3 类视频因其社交/交互特征鲜明或对端到端时延敏感值得特别关注：1）实时音视频类应用（如连麦直播、视频通话、视频会议、实时合唱等）对音视频时延、音画同步等提出了更高要求；2）云渲染类视频（如强交互云游戏）对操作响应时延提出了更高要求，且游戏相对于点播视频对卡顿也更加敏感；3）用户可以自己做“导播”的自由视角视频（FVV，Free Viewpoint Video），新增了旋转切换视角、放大看细节、“子弹时间”等更丰富的交互维度。当然，由于影响各类音视频业务体验维度的侧重点不同，故相应地评价该类音视频业务的 KQI 也存在差异，图 34 给出了各类音视频业务与主要 KQI 的对应关系。

来源：三方联合研究成果

图 34 各类音视频业务与主要 KQI 的对应关系

综上所述，首先，音视频应用的所有这些发展趋势（交互化、社交化、高清化、高帧率化、空间化）召唤着音视频应用的体验评测 KQI 相应演进，尤其是 KQI 需要增加交互响应维度，这将有助于凸显音视频应用在运营商移动网络下的用户差异化体验；其次，音视频应用的众多 KQI 对其 QoE 提升的综合影响需要进一步研究；最后，孵化更多音视频新应用以繁荣 5G 生态需要广大 ICT 产业伙伴携手合作和努力。

四、参考文献

[1] ITU-T，“P.1203 Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport”，链接：https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-P.1203-201710-I!!PDF-E&type=items

[2] UCD 耍家，“UI&UE 实用方法论 | 做交互体验，你必须得知道的「多尔蒂阈值」”，链接：http://www.woshipm.com/ucd/5018256.html

[3] 马茜，““零耗时”首帧视频体验的优化实践”，链接：https://mp.weixin.qq.com/s/xrMTi57s5VAMH1BuYYVAwg

[4] 中国网络视听节目服务协会，“2021 中国网络视听发展研究报告”，链接：https://www.sohu.com/a/470596893_100065199

[5] 声网，“体验等级协议 XLA”，链接：
https://docs.agora.io/cn/Video/xla_call_video?platform=Android

[6] ITU-T，“G.114 One-way transmission time”，链接：
https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-G.114-200305-I!!PDF-E&type=items

[7] ITU-R，“BT.1359-1 Relative Timing of Sound and Vision for Broadcasting”，链接：
https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.1359-1-199811-I!!PDF-E.pdf

[8] 声网，“声网推出首个完整实时合唱解决方案即将上线“咪哒”全国线下K歌房”，链接：
https://www.cnblogs.com/Agora/p/15309611.html

[9] 声网，“One World Together，线上实时合唱技术解析”，链接：
https://blog.csdn.net/agora_cloud/article/details/105671914

[10] 陈晓聪，“互动协作白板与音视频实时同步技术实践”，链接：
https://blog.csdn.net/vn9PLgZvnPs1522s82g/article/details/108612893

[11] 华为 X Labs，“云游戏体验模型白皮书”，链接：
https://www.huawei.com/cn/technology-insights/industry-insights/outlook/mobile-broadband/xlabs/insights-whitepapers/cloud-gmos-white-paper

[12] ETSI，“TR 126 928 V16.1.0 (2021-01) Extended Reality (XR) in 5G”链接：
https://www.etsi.org/deliver/etsi_tr/126900_126999/126928/16.01.00_60

/tr_126928v160100p.pdf

[13] Netflix，“Perceptual Video Quality Assessment based on Multi-method Fusion”，链接：https://github.com/Netflix/vmaf

[14] RealNetworks，“VMAF Reproducibility: Validating a Perceptual Practical Video Quality Metric”， 链接：https://realnetworks.com/sites/default/files/vmaf_reproducibility_ieee.pdf

[15] Jan Ozer，“Finding the Just Noticeable Difference with Netflix VMAF”，链接：https://www.linkedin.com/pulse/finding-just-noticeable-difference-netflix-vmaf-jan-ozer

[16] Robert E. Kooij，“Perceived Quality of Channel Zapping”，链接：https://www.researchgate.net/publication/220726415_Perceived_Quality_of_Channel_Zapping

[17] Human benchmark，“Human Reaction Benchmark Test”，链接：https://humanbenchmark.com/tests/reactiontime

中国信息通信研究院 泰尔系统实验室

地址：北京市海淀区花园北路 52 号

邮编：100191

电话：010-68094140

传真：010-62304980

网址：www.caict.ac.cn