

ICS号: ICS 33.050

中国标准文献分类号: M30

世界超高清视频产业联盟标准

T/UWA xxxx-202x

超高清远程交互系统音视频分级技术要求

Technical requirements of audio and video grading of ultra-high definition remote
interactive system

(征求意见稿 V1.0)

xxxx-xx-xx发布

xxxx-xx-xx实施

世界超高清视频产业联盟

目 次

前 言.....	III
1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
3.1 视频保真度 video fidelity.....	1
3.2 音频保真度 audio fidelity.....	2
3.3 音频质量评估 objective mean opinion score of audio.....	2
3.4 双讲期间发送信号衰减 attenuation of signal during double-talk.....	2
3.5 回声双讲控制特性 echo control characteristics during double-talk.....	2
3.6 终端耦合损耗 terminal coupling loss.....	2
3.7 通话建立后 AEC 收敛时间 initial convergence time at call start.....	2
3.8 调制传递函数 50%峰值的空间频率 50% peak of modulation transfer function.....	2
4 缩略语.....	3
5 系统架构.....	3
6 系统分级.....	4
6.1 分级说明.....	4
6.2 视频保真度分级.....	4
6.3 音频保真度分级.....	5
7 技术要求.....	5
7.1 视频保真度技术要求.....	5
7.2 音频保真度技术要求.....	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》给出的规则起草。

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这些专利的责任。

本文件由世界超高清视频产业联盟提出并归口。

本文件主要起草单位：华为技术有限公司、中兴通讯股份有限公司、中国信息通信研究院、海信视像科技股份有限公司、南京图格医疗科技有限公司、深圳创维-RGB电子有限公司、长春汽车职业技术大学、北京市博汇科技股份有限公司、北京市丰台区职业教育中心学校、浙江大华技术股份有限公司、四川国创新视超高清视频科技有限公司、西安诺瓦星云科技股份有限公司、苏州智聚芯联微电子有限公司。

本文件主要起草人：王勇军、徐海、刘智辉、耿东玉、杨友庆、阮冠春、陈显义、张鹏、黄成、丁元欣、王亚军、翟梦冉、张宏伟、朱宗花、李斌、王安、汪彦刚、黄新俊、张曼华、徐遥令、徐博强、郭忠武、洪太海、张晶、韩瑞雨、莫俊伟、章锋、刘征、唐晓宇、韩太初、张聪慧、关宇昕、赵春。

超高清远程交互系统音视频分级技术要求

1 范围

本文件针对超高清远程交互系统的音视频关键技术指标进行了分级定义，规定了从采集、编解码、传输到显示完整流程的技术要求。

本文件适用于超高清音视频远程交互系统的方案设计、系统检测、验收、应用以及与之相关的设备研发和生产。通过本文件的制定实施，推动超高清音视频远程交互系统相关技术的发展及规模化应用，为相关行业带来高品质的远程交互体验，促进经济社会的数字化转型。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；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。

GB/T 20090.2-2013	信息技术 先进音视频编码 第2部分：视频
GB/T 33475.2-2024	信息技术 高效多媒体编码 第2部分：视频
ITU-T H.323	基于分组的多媒体通信系统（Packet-based multimedia communications systems）
IETF RFC 3261	会话初始化协议（SIP: Session Initiation Protocol）
ITU-T G.722	7kHz音频编码（7 kHz audio-coding within 64 kbit/s）
IETF RFC 6716	音频编码格式（Opus: Definition of the Opus Audio Codec）
ISO/IEC 14496-3:2005	Coding of audio-visual objects Part 3: Audio
ITU-T H.264	用于通用音视频服务的高级视频编码（Advanced video coding for generic audiovisual services）
ITU-T H.265-2018	高效视频编码（2018 Conformance specification for ITU-T H.265 high efficiency video coding）
T/UWA 005.1-2022	高动态范围(HDR)视频技术 第 1 部分：元数据及适配
T/UWA 009.1-2023	三维声音技术技术规范 第1部分：编码分发与呈现
YD/T 4878.2-2024	智能视讯设备技术要求和测试方法 第2部分：音视频性能

3 术语和定义

下列术语以及定义适用于本文件。

3.1 视频保真度 video fidelity

是指远程交互系统在视频的采集、传输、处理和显示过程中，色彩、对比度和清晰度等视觉信息和交互体验都得到完整保留和精确传递的程度。

3.2 音频保真度 audio fidelity

是指远程交互系统在音频采集、传输、处理和播放过程中，音色、音调、响度和声场等原始音频信息和交互体验都得到完整保留和精确传递的程度。

3.3 音频质量评估 objective mean opinion score of audio

通过算法计算得出的平均意见分，用于评估音频质量，模拟主观听音测试的结果，包括：

a) MOS-LQO (Mean Opinion Score-Listening Quality Objective)

[参考：ITU-T P.863 POLQA]

b) 背景噪声下的语音和噪声评估：S-MOS (Speech Mean Opinion Score) 和N-MOS (Noise Mean Opinion Score)

[参考：ETSI TS 103 106]

3.4 双讲期间发送信号衰减 attenuation of signal during double-talk

用于衡量在双向通话中同时有发送和接收信号时发送信号的衰减程度。

[参考：ITU-T P.340]

3.5 回声双讲控制特性 echo control characteristics during double-talk

用于衡量终端设备在双讲情况下的近端信号衰减和漏回声的程度。

[参考：ETSI TS 126 132]

3.6 终端耦合损耗 terminal coupling loss

参考信号和回声抵消后的信号能量的比值，用于测量终端设备的声学回声抑制性能。

[参考：ITU-T Recommendation P.340]

3.7 通话建立后 AEC 收敛时间 initial convergence time at call start

指的是从通话建立开始后，回声衰减到有足够的抑制量所用的时间。

[参考 ITU-T P.340]

3.8 调制传递函数 50%峰值的空间频率 50% peak of modulation transfer function

MTF50P是当MTF为其峰值的一半时所对应的空间频率

[参考 YD/T 4878.2-2024 智能视讯设备技术要求和测试方法 第2部分：音视频性能]

4 缩略语

下列缩略语适用于本文件。

AEC 声学回声消除器 (Acoustic Echo Canceller)

AVC 高级视频编码 (Advanced Video Coding)

AVS 信源编码标准 (Audio Video coding Standard)

ASDT 双讲期间发送信号衰减 (Attenuation of Signal During Double-talk)

dB SPL 分贝声压级 (Decibel Sound Pressure Level)

ECC 回声控制特性 (Echo Control Characteristics)

FPS 帧率 (Frame Per Second)

HDR 高动态范围 (High Dynamic Range)

HEVC 高效视频编码 (High Efficiency Video Coding)

MCTS 运动约束片集 (Motion-Constrained Tile Sets)

MOS-LQO 平均意见分数-听觉质量客观评估 (Mean Opinion Score - Listening Quality Objective)

MTF 调制传递函数 (Modulation Transfer Function)

SDNR 信号失真噪声比 (Signal to Distortion and Noise Ratio)

SIP 会话初始协议 (Session Initiation Protocol)

TCL 终端耦合损耗 (Terminal Coupling Loss)

TMVP 时域运动矢量预测 (Temporal motion vector prediction)

5 系统架构

超高清远程交互系统总体架构分为业务层、会话控制层、媒体层和终端层，分别由会议管理软件、信令服务器、媒体服务器和终端等功能单元组成，如图1所示。

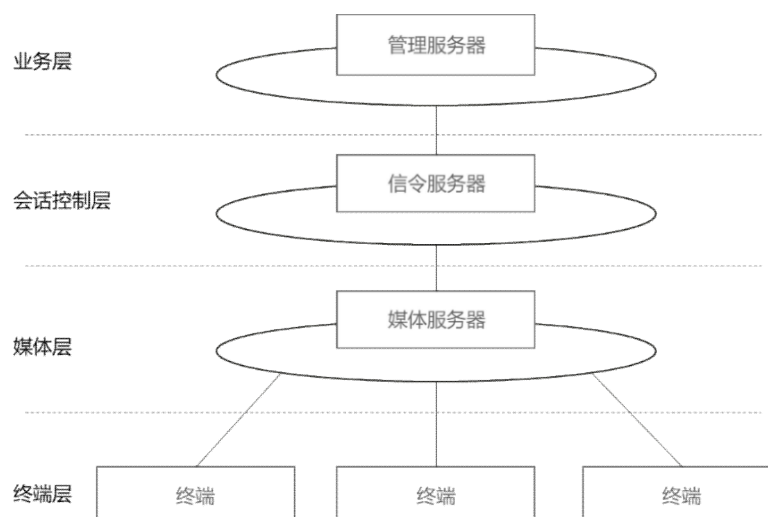


图 1 超高清远程交互系统架构图

超高清远程交互系统的业务功能，主要包括以下能力：

- a) 远程交互召集：支持点对点交互、支持多点交互、支持临时发起交互、支持预约多点交互。
- b) 远程交互控制：呼叫终端、挂断终端、终端闭音/取消闭音、终端静音/取消静音。
- c) 远程交互媒体：支持混音、支持多画面、支持多屏图像拼接、支持文档、白板等辅流共享。

6 系统分级

6.1 分级说明

超高清远程交互系统分级如图2所示，主要包括视频保真度分级和音频保真度分级，定义了采、编、传、解、显各个技术环节不同级别的参数指标，用于指导远程交互系统的音视频分级设计。

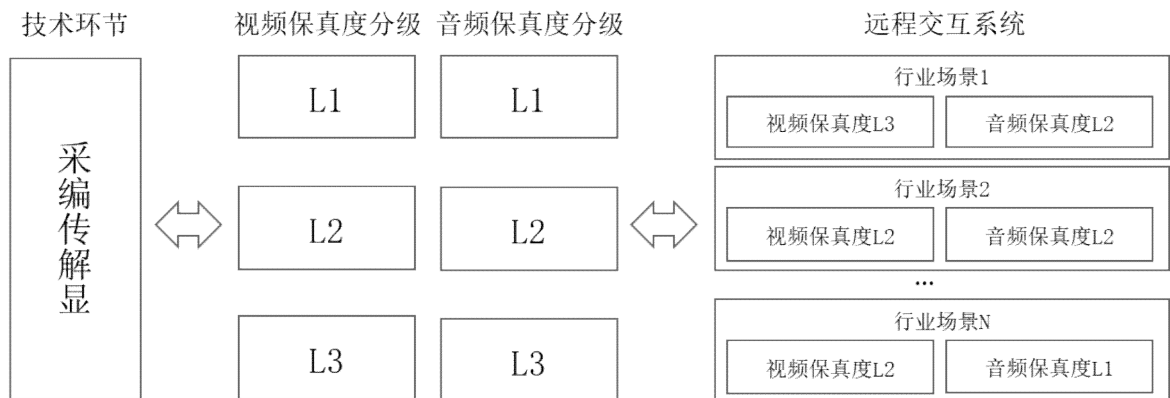


图 2 超高清远程交互系统分级

6.2 视频保真度分级

综合考虑到超高清远程交互系统的场景需求，技术实现难度和用户体验，方便管理和评估，将视频保真度分为三个级别：

- a) L1视频保真度：全高清视频，适度压缩，整体清晰流畅，细节有一定损失，色彩稍显简单；
- b) L2视频保真度：4K超高清视频，细微压缩，色彩还原度高，细节保留良好，流畅自然，纹理清晰；
- c) L3视频保真度：8K超高清视频，细节丰富逼真，色彩还原度极高，能还原原始影像的真实感受。

视频保真度分级指标框架如图3所示：

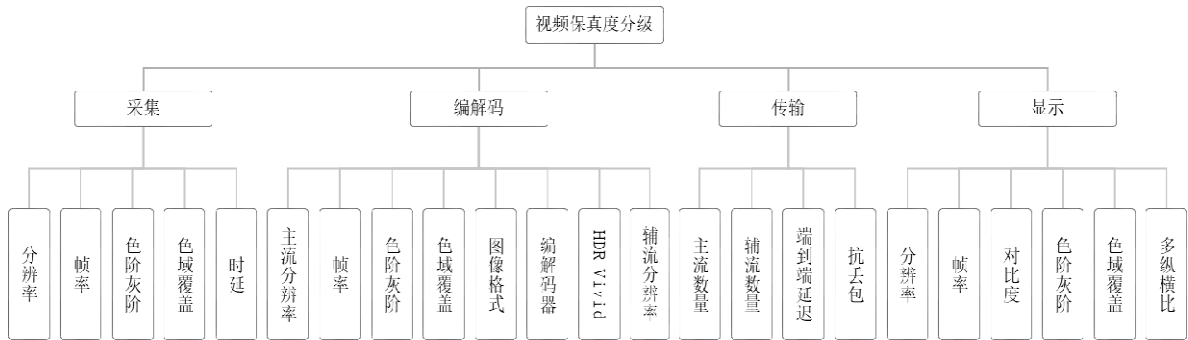


图 3 视频保真度分级指标框架

6.3 音频保真度分级

综合考虑到超高清远程交互系统的场景需求，技术实现难度和用户体验，方便管理和评估，将音频保真度分为三个级别：

- a) L1音频保真度：宽带音频，保留大部分原声，满足一般质量要求的通信需求，满足基本的交互体验；
- b) L2音频保真度：超宽带音频，接近原声质量，满足较高质量的通信需求，交互体验更加顺畅；
- c) L3音频保真度：全带音频，高保真原声音质，满足高质量的通信需求，交互体验质量高。

音频保真度分级指标框架如图4所示：

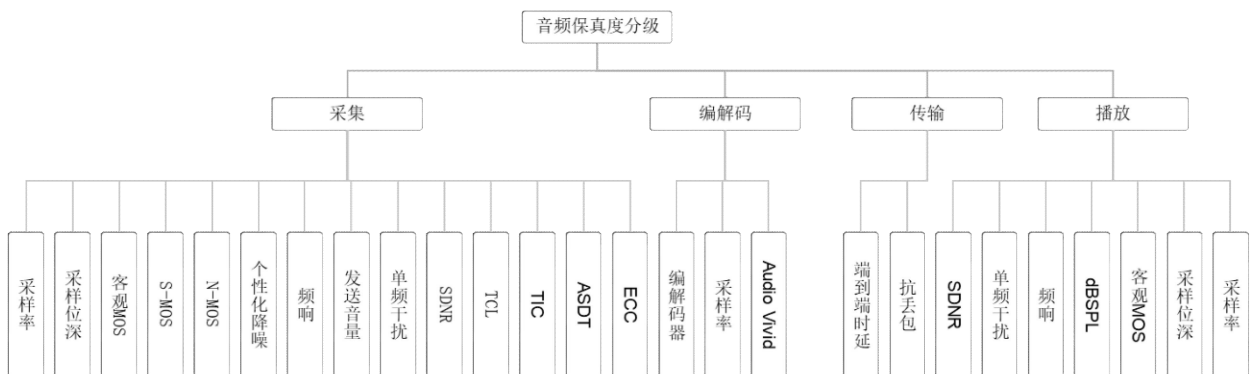


图 4 音频保真度分级指标框架

7 技术要求

7.1 视频保真度技术要求

7.1.1 L1 视频保真要求

超高清远程交互系统在采集、编解码、传输和显示技术环节，L1视频保真度应满足表1的指标要求。

表 1 L1视频保真度指标要求

环节	指标名称	指标要求
采集	分辨率	支持1920×1080, 图像中心10%视场MTF50P≥0.4
	帧率	支持30FPS
	色深	支持8bit
	色域覆盖	支持BT.709
	时延	a) 单目图像采集: 时延≤130ms b) 多目图像采集: 时延≤250ms
编解码	分辨率	支持1920×1080
	帧率	支持30FPS
	色深	支持8bit
	色域覆盖	支持BT.709
	图像格式	支持YUV420
	编解码器	a) 支持AVC main profile, 码率≥1.3M bps@30FPS b) 支持HEVC main profile, 码率≥1M bps@30FPS c) 宜支持AVS编解码
	辅流分辨率	支持1920×1080
传输	主流数量	≥1
	辅流数量	≥1
	端到端时延	≤400ms
	抗丢包	≥30%, 且最大连续丢包个数≥10
显示	分辨率	支持1920×1080
	帧率	支持60FPS
	对比度	≥1000:1 (测试方法参考SJT11348-2016, 5.2章节)
	色深	支持10bit
	色域覆盖	支持BT.709
	纵横比	支持16:9

7.1.2 L2 视频保真要求

超高清远程交互系统在采集、编解码、传输和显示技术环节，L2视频保真度应满足表2的指标要求。

表 2 L2视频保真度指标要求

环节	指标名称	指标要求
采集	分辨率	支持3840×2160， 图像中心10%视场MTF50P≥0.4
	帧率	支持30FPS， 宜支持60FPS
	色深	支持8bit， 宜支持10bit
	色域覆盖	支持BT.709， 宜支持DCI-P3
	时延	a) 单目图像采集： 延时应≤100ms b) 多目图像采集： 延时应≤200ms
编解码	分辨率	支持3840×2160
	帧率	支持30FPS， 宜支持60FPS
	色深	支持8bit， 宜支持10bit
	色域覆盖	支持BT.709， 宜支持DCI-P3
	图像格式	支持YUV420， 宜支持YUV422
	编解码器	a) 支持AVC main profile， 码率≥5M bps@30FPS b) 宜支持： AVC High10 profile， 码率≥6M bps@30FPS c) 支持HEVC main profile， 码率≥4M bps@30FPS d) 宜支持： HEVC main10 profile， 码率≥5M bps@30FPS e) 宜支持AVS编解码
传输	辅流分辨率	支持3840×2160
	主流数量	≥2
	辅流数量	≥2
	端到端时延	≤350ms
显示	抗丢包	≥40%， 且最大连续丢包个数≥12
	主流分辨率	支持3840×2160
	帧率	支持60FPS， 宜支持90FPS
	对比度	≥5000:1 (测试方法参考SJT11348-2016， 5.2章节)
	色深	支持10bit， 宜支持12bit

	色域覆盖	支持BT.709, 宜支持DCI-P3
	纵横比	支持16:9 宜支持21:9/32:9

7.1.3 L3 视频保真要求

超高清远程交互系统在采集、编解码、传输和显示技术环节, L3视频保真度应满足表3的指标要求。

表 3 L3视频保真度指标要求

环节	指标名称	指标要求
采集	分辨率	支持3840×2160, 宜支持7680×4320, 图像中心10%视场MTF50P≥0.4
	帧率	支持30FPS, 宜支持60FPS
	色深	支持10bit, 宜支持12bit
	色域覆盖	支持DCI-P3, 宜支持BT.2020
	时延	a) 单目图像采集: 时延≤70ms b) 多目图像采集: 时延≤150ms
编解码	分辨率	支持3840×2160, 宜支持7680×4320
	帧率	支持30FPS, 宜支持60FPS
	色深	支持10bit, 宜支持12bit
	色域覆盖	支持DCI-P3, 宜支持BT.2020
	图像格式	支持YUV444
	编解码器	a) 支持AVC High10 profile, 7680×4320分辨率下的码率≥20M bps@30FPS b) 支持HEVC main10 profile, 且应支持Tile及MCTS和TMVP参数, 7680×4320分辨率下的码率≥16M bps@30FPS c) 宜支持 AVC High4:4:4 profile, 7680×4320分辨率下的码率≥24M bps@30FPS d) 宜支持 HEVC main12 profile; 7680×4320分辨率下的码率≥20M bps@30FPS e) 支持兼容YUV420/YUV422编解码 f) 宜支持AVS
	HDR Vivid	宜支持

	辅流分辨率	支持3840×2160，宜支持7680×4320
传输	主流数量	≥3 应支持把多路码流合成一路码流
	辅流数量	≥3 应支持把多路码流合成一路码流
	端到端时延	≤300ms
	抗丢包	≥50%，且最大连续丢包个数≥14
显示	分辨率	支持3840×2160，宜支持7680×4320
	帧率	支持90FPS
	对比度	≥10000:1 (测试方法参考SJT11348-2016, 5.2章节)
	色深	支持12bit，宜支持16bit
	色域覆盖	支持DCI-P3，宜支持BT.2020
	纵横比	支持16:9/21:9/32:9，宜支持48:9

7.2 音频保真度技术要求

7.2.1 L1 音频保真要求

超高清远程交互系统在采集、编解码、传输和播放技术环节，L1音频保真度应满足表4的指标要求。

表 4 L1音频保真度指标要求

环节	指标名称	指标要求
采集	采样率	≥16KHz
	采样位深	支持16bit
	发送方向客观MOS分数 (MOS-LQO)	≥3.3
	噪声场景下的语音质量 及抑制噪声性能	混响房间下，平均S-MOS≥3.3，平均N-MOS≥3.0
	发送方向频率响应	a) 10dB带宽应满足100Hz~7KHz b) 频率响应在如下掩蔽曲线范围内： 100Hz（下限-30dB，上限6dB） 200Hz（下限-7dB，上限6dB）

	<p>2KHz (下限-7dB, 上限6dB)</p> <p>5KHz (下限-7dB, 上限6dB)</p> <p>6.3KHz (下限-11dB, 上限6dB)</p> <p>7.5KHz (下限-11dB, 上限6dB)</p> <p>8KHz (上限6dB)</p>
发送方向语音能量	<p>a) 在使用距离下, 正常语音能量满足[-26, -14]dBFS</p> <p>b) 在使用距离下, 低声语音能量\geq-37dBFS</p>
发送方向在安静环境下的单频干扰	<p>支持如下条件之一:</p> <p>a) 干扰峰值电平减去宽带噪声信号电平\leq10dB</p> <p>b) 干扰峰值\leq-80dBV</p>
发送方向总失真 (SDNR)	<p>各个频段SDNR满足:</p> <p>224-282 \geq30dB</p> <p>282-355 \geq30dB</p> <p>355-447 \geq30dB</p> <p>447-562 \geq30dB</p> <p>562-708 \geq30dB</p> <p>708-891 \geq30dB</p> <p>891-1122 \geq30dB</p> <p>1122-1413 \geq30dB</p> <p>1413-1778 \geq30dB</p> <p>1778-2239 \geq30dB</p> <p>2239-2818 \geq30dB</p> <p>2818-3548 \geq30dB</p> <p>3548-4467 \geq30dB</p> <p>4467-5623 \geq30dB</p>
回声耦合损耗TCL	混响房间下, 标称音量 \geq 50dB, 最大音量 \geq 50dB
双讲情况下的回声控制特性 (ECC)	<p>a) 混响房间下, 正常音量: A1+A2占比\geq50%; F占比\leq5%, G占比\leq5%</p> <p>b) 混响房间下, 最大音量: A1+A2占比\geq30%; F占比\leq8%, G占比\leq8%</p> <p>说明: A1情况: 全双工发送端电平变化$-4\text{dB} \leq \Delta L \leq 4\text{dB}$.</p> <p>A2情况: 发送端电平变化$-15\text{dB} \leq \Delta L < -4\text{dB}$, F情况: 突发漏回声, G情况: 持续漏回声</p>

	双讲情况下的发送语音衰减 (ASDT)	a) 混响房间下, 在使用距离下, 标称音量应 $\leq 12\text{dB}$ b) 混响房间下, 在使用距离下, 最大音量应 $\leq 15\text{dB}$
	语音通话建立后AEC收敛时间 (TIC)	混响房间下, 通话建立后, 发送方向的信号能量降到 -60dBFS 的时间 $t \leq 5\text{s}$
编解码	编解码器	支持G.722、Opus
	采样率	支持16KHz
传输	端到端时延	$\leq 400\text{ms}$
	抗丢包	$\geq 50\%$, 且最大连续丢包个数 ≥ 10
播放	采样率	$\geq 16\text{KHz}$
	采样位深	支持16bit
	接收方向客观MOS分数 (MOS-LQ0)	≥ 3.6
	接收方向输出声压级 (dBSPL)	不同使用距离下, 应各自满足以下要求 a) 耳机场景: 76 ± 2 b) 免提场景 (在以下使用距离, 距扬声器1m处测量收听音量值应分别满足): 1m: 64 ± 2 1.5m: 67 ± 2 2.5m: 69 ± 2 3.5m: 71 ± 2 4.5m: 73 ± 2 7.5m: 76 ± 2
接收方向频率响应	a) 10dB带宽满足100Hz~7KHz b) 频率响应如下掩蔽曲线范围内: 100Hz (上限6dB) 200Hz (下限-13dB, 上限6dB) 250Hz (下限-11dB, 上限6dB) 1KHz (下限-7dB, 上限6dB) 5KHz (下限-7dB, 上限6dB) 6.3KHz (下限-11dB, 上限6dB) 7.5KHz (下限-11dB, 上限6dB)	

		8KHz (上限7dB)
	接收方向单频干扰	支持如下条件之一： a) 干扰峰值电平减去宽带噪声信号电平 $\leq 10\text{dB}$ b) 干扰峰值 $\leq 24\text{dB SPL}$
	接收方向输出总失真 (SDNR)	各个频段在测试电平 -16dBFS 时，SDNR满足： a) 免提场景： 224-282 $\geq 20\text{dB}$ 282-355 $\geq 20\text{dB}$ 355-447 $\geq 22\text{dB}$ 447-562 $\geq 24\text{dB}$ 562-708 $\geq 24\text{dB}$ 708-891 $\geq 24\text{dB}$ 891-1122 $\geq 24\text{dB}$ 1122-1413 $\geq 24\text{dB}$ 1413-1778 $\geq 24\text{dB}$ 1778-2239 $\geq 24\text{dB}$ 2239-2818 $\geq 24\text{dB}$ 2818-3548 $\geq 24\text{dB}$ 3548-4467 $\geq 24\text{dB}$ 4467-5623 $\geq 24\text{dB}$ b) 耳机场景，在免提场景基础上加8dB

注1：以上指标测试方法可参考：[YD/T 4878.2-2024 智能视讯设备技术要求和测试方法 第2部分：音视频性能]。

注2：以上指标中的正常音量定义：在嘴参考点处根据ITU-T P. 56计算的信号活动语音电平为 -4.7dBPa ，

以上指标中的低声音量定义：在嘴参考点处根据ITU-T P. 56计算的信号活动语音电平为 -14.7dBPa 。

以上指标中的标称音量定义：为正常使用时的终端设置的音量，一般是指默认输出音量。

注3：以上指标中的使用距离定义：分为耳机、1m、1.5m、2.5m、3.5m、4.5m、7.5m几个距离，根据实际使用距离选取与前述距离最接近中的1个进行测试。如果实际使用距离在前述两个距离值的正中间，则选取较大值；最大距离不超过7.5m。

注4：注1~注3的测试方法和定义适用于L2和L3音频保真要求。

7.2.2 L2 音频保真要求

超高清远程交互系统在采集、编接码、传输和播放技术环节，L2音频保真度应满足表5的指标要

求。

表 5 L2音频保真度指标要求

环节	指标名称	指标要求
采集	采样率	$\geq 48\text{KHz}$
	采样位深	支持16bit, 宜支持24bit
	发送方向客观MOS分数 (MOS-LQO)	≥ 3.6
	噪声场景下的语音质量 及抑制噪声性能	混响房间下, 平均S-MOS ≥ 3.6 , 平均N-MOS应 ≥ 3.2
	发送方向频率响应	a) 10dB带宽应满足100Hz~14KHz b) 频率响应在如下掩蔽曲线范围内: 100Hz (下限-30dB, 上限5dB) 200Hz (下限-6dB, 上限5dB) 2KHz (下限-6dB, 上限5dB) 7.5KHz (下限-6dB, 上限5dB) 12.5KHz (下限-6dB, 上限5dB) 14KHz (下限-11dB, 上限5dB) 16KHz (上限5dB)
	发送方向语音能量	a) 在使用距离下, 正常语音能量满足[-23, -14]dBFS b) 在使用距离下, 低声语音能量 $\geq -34\text{dBFS}$
	发送方向在安静环境下的 单频干扰	支持如下条件之一: a) 干扰峰值电平减去宽带噪声信号电平 $\leq 10\text{dB}$ b) 干扰峰值 $\leq -80\text{dBV}$
	发送方向总失真 (SDNR)	各个频段SDNR满足: 224-282 $\geq 33\text{dB}$ 282-355 $\geq 33\text{dB}$ 355-447 $\geq 33\text{dB}$ 447-562 $\geq 33\text{dB}$ 562-708 $\geq 33\text{dB}$ 708-891 $\geq 33\text{dB}$ 891-1122 $\geq 33\text{dB}$ 1122-1413 $\geq 33\text{dB}$ 1413-1778 $\geq 33\text{dB}$

		<p>1778-2239 $\geq 33\text{dB}$</p> <p>2239-2818 $\geq 33\text{dB}$</p> <p>2818-3548 $\geq 33\text{dB}$</p> <p>3548-4467 $\geq 33\text{dB}$</p> <p>4467-5623 $\geq 33\text{dB}$</p>
	回声耦合损耗TCL	混响房间下，标称音量 $\geq 56\text{dB}$ ，最大音量 $\geq 56\text{dB}$
	双讲情况下的回声控制特性（ECC）	<p>a) 混响房间下，正常音量：A1+A2占比$\geq 60\%$；F占比$\leq 5\%$，G占比$\leq 5\%$</p> <p>b) 混响房间下，最大音量：A1+A2占比$\geq 40\%$；F占比$\leq 8\%$，G占比应$\leq 8\%$</p> <p>说明：A1情况：全双工发送端电平变化$-4\text{dB} \leq \Delta L \leq 4\text{dB}$；</p> <p>A2情况：发送端电平变化$-15\text{dB} \leq \Delta L < -4\text{dB}$，F情况：突发漏回声，G情况：持续漏回声</p>
	双讲情况下的发送语音衰减（ASDT）	<p>a) 混响房间下，在使用距离下，标称音量应$\leq 9\text{dB}$</p> <p>b) 混响房间下，在使用距离下，最大音量应$\leq 12\text{dB}$</p>
	语音通话建立后的AEC收敛时间（TIC）	混响房间下，发送方向的信号能量降到 -60dBFS 以下的时间 $t \leq 4\text{s}$
编解码	编解码器	应支持G. 722、Opus、AAC_LD
	采样率	支持16KHz、48KHz
传输	端到端时延	$\leq 350\text{ms}$
	抗丢包	$\geq 60\%$ ，且最大连续丢包个数 ≥ 12
播放	采样率	$\geq 48\text{KHz}$
	采样位深	支持16bit，宜支持24bit
	接收方向客观MOS分数（MOS-LQO）	≥ 3.8
	接收方向输出声压级（dB SPL）	<p>不同使用距离下，应各自满足以下要求</p> <p>a) 耳机场景：76 ± 2</p> <p>b) 免提场景（在以下使用距离，距扬声器1m处测量收听音量值应分别满足）：</p> <p>1m: 64 ± 2</p> <p>1.5m: 67 ± 2</p> <p>2.5m: 69 ± 2</p> <p>3.5m: 71 ± 2</p>

		<p>4.5m: 73 ± 2</p> <p>7.5m: 76 ± 2</p>
	接收方向频率响应	<p>1. 10dB带宽应满足100Hz~14KHz</p> <p>2. 频率响应在如下掩蔽曲线范围内:</p> <p>100Hz (上限5dB)</p> <p>180Hz (下限-12, 上限5dB)</p> <p>200Hz (下限-10dB, 上限5dB)</p> <p>250Hz (下限-6dB, 上限5dB)</p> <p>7.5KHz (下限-6dB, 上限5dB)</p> <p>12.5KHz (下限-6dB, 上限5dB)</p> <p>14KHz (下限-11dB, 上限5dB)</p> <p>16KHz (上限5dB)</p>
	接收方向单频干扰	<p>支持如下条件之一:</p> <p>a) 干扰峰值电平减去宽带噪声信号电平$\leq 13\text{dB}$</p> <p>b) 干扰峰值$\leq 24\text{dBSPL}$</p>
	接收方向输出总失真 (SDNR)	<p>各个频段在测试电平-16dBFS时, SDNR满足:</p> <p>a) 免提场景:</p> <p>224-282 $\geq 23\text{dB}$</p> <p>282-355 $\geq 23\text{dB}$</p> <p>355-447 $\geq 25\text{dB}$</p> <p>447-562 $\geq 27\text{dB}$</p> <p>562-708 $\geq 27\text{dB}$</p> <p>708-891 $\geq 27\text{dB}$</p> <p>891-1122 $\geq 27\text{dB}$</p> <p>1122-1413 $\geq 27\text{dB}$</p> <p>1413-1778 $\geq 27\text{dB}$</p> <p>1778-2239 $\geq 27\text{dB}$</p> <p>2239-2818 $\geq 27\text{dB}$</p> <p>2818-3548 $\geq 27\text{dB}$</p> <p>3548-4467 $\geq 27\text{dB}$</p> <p>4467-5623 $\geq 27\text{dB}$</p> <p>b) 耳机场景在免提场景基础上加7dB</p>

7.2.3 L3 音频保真要求

超高清远程交互系统在采集、编解码、传输和播放技术环节，L3音频保真度应满足表6的指标要求。

表 6 L3音频保真度指标要求

环节	指标名称	指标要求
采集	采样率	$\geq 48\text{KHz}$
	采样位深	支持16bit，宜支持24bit
	发送方向客观MOS分数 (MOS-LQ0)	≥ 3.9
	噪声场景下的语音质量 及抑制噪声性能	混响房间下，平均S-MOS ≥ 3.9 ，平均N-MOS ≥ 3.5
	发送方向频率响应	a) 10dB带宽应满足100Hz~20KHz b) 频率响应在如下掩蔽曲线范围内： 100Hz（下限-30dB，上限5dB） 200Hz（下限-6dB，上限5dB） 2KHz（下限-6dB，上限5dB） 12.5KHz（下限-6dB，上限5dB） 14KHz（下限-6dB，上限5dB） 16KHz（下限-11dB，上限5dB） 20KHz（上限5dB）
	发送方向语音能量	a) 在使用距离下，正常语音能量满足 $[-23, -14]\text{dBFS}$ b) 在使用距离下，低声语音能量 $\geq -34\text{dBFS}$
	发送方向在安静环境下的 单频干扰	支持如下条件之一： a) 干扰峰值电平减去宽带噪声信号电平 $\leq 10\text{dB}$ b) 干扰峰值 $\leq -80\text{dBV}$
	发送方向总失真（SDNR）	各个频段SDNR满足： 224-282 $\geq 36\text{dB}$ 282-355 $\geq 36\text{dB}$ 355-447 $\geq 36\text{dB}$ 447-562 $\geq 36\text{dB}$ 562-708 $\geq 36\text{dB}$ 708-891 $\geq 36\text{dB}$ 891-1122 $\geq 36\text{dB}$

		1122-1413 $\geq 36\text{dB}$ 1413-1778 $\geq 36\text{dB}$ 1778-2239 $\geq 36\text{dB}$ 2239-2818 $\geq 36\text{dB}$ 2818-3548 $\geq 36\text{dB}$ 3548-4467 $\geq 36\text{dB}$ 4467-5623 $\geq 36\text{dB}$
	回声耦合损耗TCL	混响房间下，标称音量 $\geq 56\text{dB}$ ，最大音量 $\geq 56\text{dB}$
	双讲情况下的回声控制特性（ECC）	a) 混响房间下，正常音量：A1+A2占比 $\geq 70\%$ ；F占比 $\leq 5\%$ ，G占比 $\leq 5\%$ b) 混响房间下，最大音量：A1+A2占比 $\geq 50\%$ ；F占比 $\leq 8\%$ ，G占比 $\leq 8\%$ 说明： A1情况：全双工发送端电平变化 $-4\text{dB} \leq \Delta L \leq 4\text{dB}$ ； A2情况：发送端电平变化 $-15\text{dB} \leq \Delta L < -4\text{dB}$ ，F情况：突发漏回声，G情况：持续漏回声
	双讲情况下的发送语音衰减（ASDT）	a) 混响房间下，使用距离下，标称音量 $\leq 6\text{dB}$ b) 混响房间下，使用距离下，最大音量 $\leq 9\text{dB}$
	语音通话建立后的回AEC收敛时间（TIC）	混响房间下，通话建立后，发送方向的信号能量降到 -60dBFS 以下的时间 $t \leq 3\text{s}$
	个性化降噪	对于个人使用设备宜支持，通过注册个人声纹，抑制声纹注册说话人以外的人声和噪声。
编解码	编解码器	支持G.722、Opus、AAC_LD 宜支持分层编解码Codec
	采样率	支持16KHz、48KHz
	Audio Vivid	宜支持
传输	端到端时延	$\leq 300\text{ms}$
	抗丢包	$\geq 70\%$ ，且最大连续丢包个数 ≥ 14
播放	采样率	$\geq 48\text{KHz}$
	采样位深	支持16bit，宜支持24bit
	接收方向客观MOS分数	≥ 4
	接收方向输出声压级（dBSPL）	不同使用距离下，应各自满足以下要求 a) 耳机场景： 76 ± 2

		<p>b) 免提场景（在以下使用距离，距扬声器1m处测量收听音量值应分别满足）：</p> <p>1m: 64 ± 2</p> <p>1.5m: 67 ± 2</p> <p>2.5m: 69 ± 2</p> <p>3.5m: 71 ± 2</p> <p>4.5m: 73 ± 2</p> <p>7.5m: 76 ± 2</p>
	接收方向频率响应	<p>a) 10dB带宽应满足 100Hz~20KHz</p> <p>b) 频率响应在如下掩蔽曲线范围内：</p> <p>100Hz（上限6dB）</p> <p>200Hz（下限-10dB，上限5dB）</p> <p>250Hz（下限-6dB，上限5dB）</p> <p>12.5KHz（下限-6dB，上限5dB）</p> <p>14KHz（下限-6dB，上限5dB）</p> <p>16KHz（下限-11dB，上限5dB）</p> <p>20KHz（上限5dB）</p>
	接收方向单频干扰	<p>支持如下条件之一：</p> <p>a) 干扰峰值电平减去宽带噪声信号电平≤ 16dB</p> <p>b) 干扰峰值≤ 24dB SPL</p>
	接收方向输出总失真（SDNR）	<p>各个频段在测试电平-16dBFS时，SDNR满足：</p> <p>a) 免提场景：</p> <p>224-282 ≥ 26dB</p> <p>282-355 ≥ 26dB</p> <p>355-447 ≥ 28dB</p> <p>447-562 ≥ 30dB</p> <p>562-708 ≥ 30dB</p> <p>708-891 ≥ 30dB</p> <p>891-1122 ≥ 30dB</p> <p>1122-1413 ≥ 30dB</p> <p>1413-1778 ≥ 30dB</p> <p>1778-2239 ≥ 30dB</p> <p>2239-2818 ≥ 30dB</p>

		2818-3548 $\geq 30\text{dB}$ 3548-4467 $\geq 30\text{dB}$ 4467-5623 $\geq 30\text{dB}$ b) 耳机场景在免提场景基础上加6dB
--	--	---