



UHD World Association
世界超高清视频产业联盟

ICS 33.050

CCS M30

世界超高清视频产业联盟标准

T/UWA 030-2025

超高清远程交互系统音视频分级技术要求

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

2025-01-20发布

2025-01-20实施

世界超高清视频产业联盟

目 次

前 言	II
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	3
5 系统架构	3
6 系统分级	4
6.1 概述	4
6.2 视频保真度分级	5
6.3 音频保真度分级	5
7 技术要求	6
7.1 视频保真度	6
7.2 音频保真度技术要求	9
8 测试方法	21
8.1 视频保真度测试方法	21
8.2 音频保真度测试方法	24

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、西安华为技术有限公司、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、中国电子技术标准化研究院。

本文件主要起草人：王勇军、徐海、刘智辉、张磊、耿东玉、阮冠春、方海鹏、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、张冉。

超高清远程交互系统音视频分级技术要求

1 范围

本文件规定了超高清远程交互系统音视频关键技术指标的分级要求，包括从采集、编解码、传输到显示/播放完整的流程。

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、验收及应用，可供相关设备研发和生产参考使用。

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 信息技术 先进音视频编码 第2部分：视频

GB/T 33475.2 信息技术 高效多媒体编码 第2部分：视频

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

T/AI 109.2 信息技术 智能媒体编码 第2部分：视频

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 三维声音技术规范 第1部分：编码分发与呈现

ITU-T G.722 64 kbit/s内的7 kHz音频编码（7 kHz audio-coding within 64 kbit/s）

ITU-T H.264 通用视听业务的先进的视频编码（Advanced video coding for generic audiovisual services）

ITU-T H.265 高效率视频编码（High efficiency video coding）

ITU-T P.56 活动语音电平测量（Objective measurement of active speech level）

ITU-T G.191 用于语音及音频编码标准的软件工具（Software tools for speech and audio coding standardization）

ITU-T P.501 用于电话和其它语音应用的测试信号（Test signals for use in telephony and other speech-based applications）

ITU-T P.863 感知客观聆听质量评估（Perceptual objective listening quality prediction）

IETF RFC 6716 音频编码格式（Opus: Definition of the Opus Audio Codec）

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

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得到完整保留和精确传递的程度。

3.2

音频保真度 audio fidelity

远程交互系统在音频的采集、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得到完整保留和精确传递的程度。

3.3

主流 video

视频会议通信中的第一视频流，一般是指和终端连接的摄像机的视频信号。

3.4

辅流 presentation

除主流外的辅助视频流，一般用于和终端连接的共享内容数字信号，如发言者的共享文档、共享白板等。

3.5

双讲 double talk

在会议中，近端本地说话人和远端说话人同时说话的场景。

3.6

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

发送语音信号在双讲和近端两种情况下的电平差异。

3.7

终端耦合损耗 terminal coupling loss

参考信号和回声抵消后的信号能量比值。

3.8

语音通话建立后的AEC收敛时间 AEC convergence time at call start

从通话建立开始到回声衰减至指定抑制量时所用的时间。

3.9

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

当MTF为其峰值的一半时所对应的空间频率。

[来源：YD/T 4878.2-2024, 8.1.1]

3.10

超宽带音频 Super Wide Band Audio

频宽大于等于16k Hz且对应采样率不低于32k Hz的音频信号。

3.11

全带音频 Full Band Audio

频宽大于等于20k Hz且对应采样率不低于44.1k Hz的音频信号。

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)

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

AVC 高级视频编码 (Advanced Video Coding)

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

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

DUT 被测设备 (Device Under Test)

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)

JVET 联合视频专家组 (Joint Video Experts Team)

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)

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分为业务层、会话控制层、媒体层和终端层，分别由管理服务器、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。

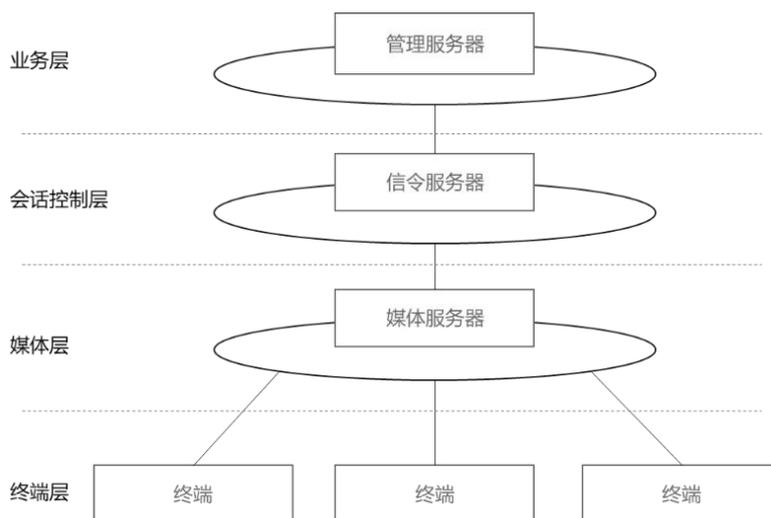


图 1 超高清远程交互系统架构图示例

超高清远程交互系统应具备支持以下业务的基本功能。

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

6 系统分级

6.1 概述

超高清远程交互系统分级示意如图2所示，包括视频保真度分级和音频保真度分级，定义了采集、编码、传输、解码及显示/播放各个技术环节在不同保真度分级的参数指标。

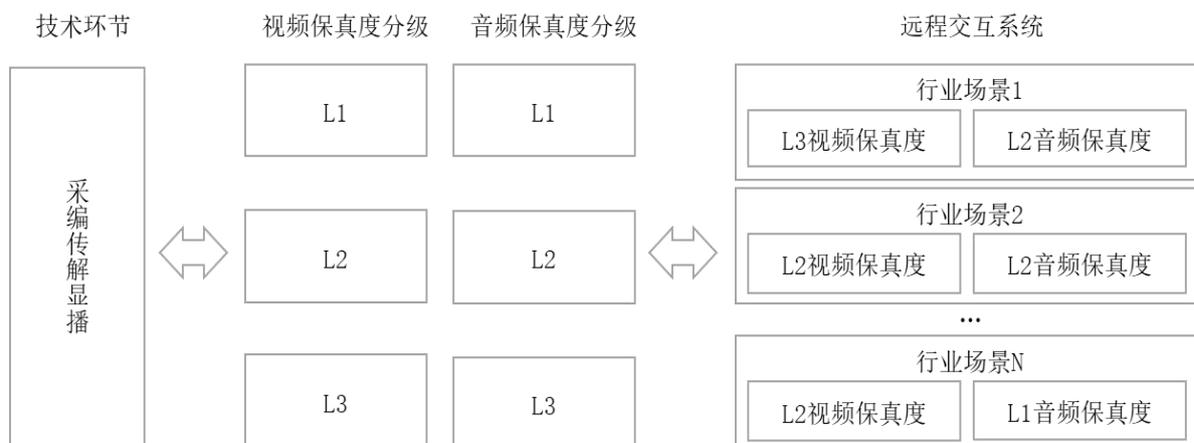


图 2 超高清远程交互系统分级

6.2 视频保真度分级

视频保真度分为三个级别：

- L1视频保真度：全高清视频，压缩度高，色彩还原度一般，细节有一定损失；
- L2视频保真度：超高清视频，压缩度中，色彩还原度高，细节保留良好；
- L3视频保真度：极高清视频，压缩度低，色彩还原度极高，细节丰富逼真。

视频保真度分级指标框架如图3所示。

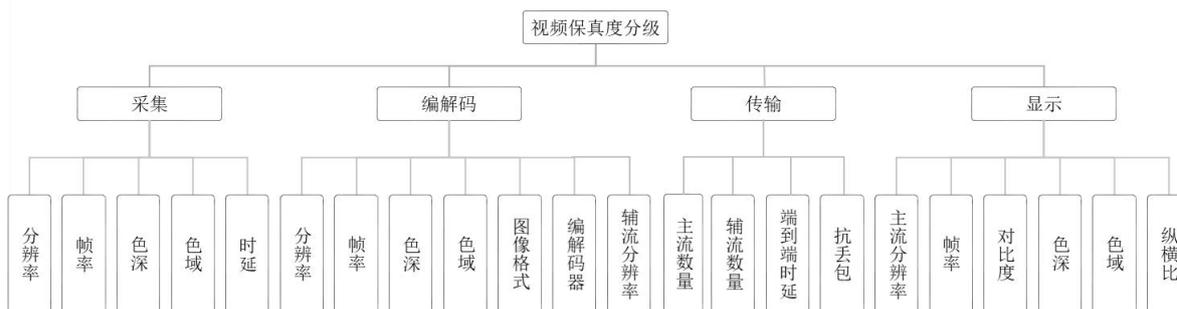


图 3 视频保真度分级指标框架

6.3 音频保真度分级

音频保真度分为三个级别：

- 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；
- 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；
- 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。

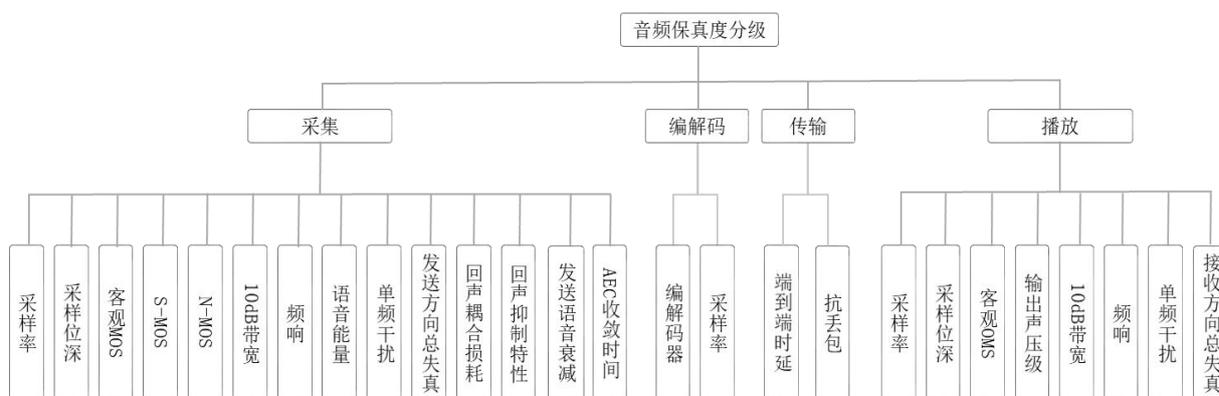


图 4 音频保真度分级指标框架

7 技术要求

7.1 视频保真度

7.1.1 L1 视频保真度

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

序号	环节	指标名称	指标要求
1	采集	分辨率	1920×1080, 图像中心10%视场MTF50P≥0.4
2		帧率	30 FPS
3		色深	8 bit
4		色域	BT.709
5		时延	a) 单目图像采集: 时延≤130 ms b) 多目图像采集: 时延≤250 ms
6	编解码	分辨率	1920×1080
7		帧率	30 FPS
8		色深	8 bit
9		色域	BT.709
10		图像格式	YUV420
11		编解码器	a) 应支持AVC main profile, 码率≥1.3M bps@30FPS b) 应支持HEVC main profile, 码率≥1M bps@30FPS c) 宜支持AVS编解码
12		辅流分辨率	1920×1080
13	传输	主流数量	≥1
14		辅流数量	≥1
15		端到端时延	≤400 ms
16		抗丢包	≥30%, 且最大连续丢包个数≥10
17	显示	分辨率	1920×1080
18		帧率	60 FPS

表 1 L1视频保真度指标要求（续）

序号	环节	指标名称	指标要求
19		对比度	≥1000:1
20		色深	10 bit
21		色域	BT.709
22		纵横比	16:9

7.1.2 L2 视频保真度

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

序号	环节	指标名称	指标要求
1	采集	分辨率	3840×2160， 图像中心10%视场MTF50P≥0.4
2		帧率	应支持30 FPS， 宜支持60 FPS
3		色深	应支持8 bit， 宜支持10 bit
4		色域	应支持BT.709， 宜支持DCI-P3
5		时延	a) 单目图像采集：时延应≤100 ms b) 多目图像采集：时延应≤200 ms
6	编解码	分辨率	3840×2160
7		帧率	应支持30 FPS， 宜支持60 FPS
8		色深	应支持8 bit， 宜支持10 bit
9		色域	应支持BT.709， 宜支持DCI-P3
10		图像格式	支持YUV420， 宜支持YUV422
11		编解码器	a) 应支持AVC main profile， 码率≥5M bps@30FPS b) 宜支持： AVC High10 profile， 码率≥6M bps@30FPS c) 应支持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
12		辅流分辨率	3840×2160

表 2 L2视频保真度指标要求（续）

序号	环节	指标名称	指标要求
13	传输	主流数量	≥2
14		辅流数量	≥2
15		端到端时延	≤350 ms
16		抗丢包	≥40%，且最大连续丢包个数≥12
17	显示	分辨率	3840×2160
18		帧率	应支持60 FPS，宜支持90 FPS
19		对比度	≥5000:1
20		色深	应支持10 bit，宜支持12 bit
21		色域	应支持BT.709，宜支持DCI-P3
22		纵横比	应支持16:9 宜支持21:9/32:9

7.1.3 L3 视频保真度

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

序号	环节	指标名称	指标要求
1	采集	分辨率	应支持3840×2160，宜支持7680×4320 图像中心10%视场MTF50P≥0.4
2		帧率	应支持30 FPS，宜支持60 FPS
3		色深	应支持10 bit，宜支持12 bit
4		色域	应支持DCI-P3，宜支持BT.2020
5		时延	a)单目图像采集：时延≤70 ms b)多目图像采集：时延≤150 ms
6	编解码	分辨率	应支持3840×2160，宜支持7680×4320
7		帧率	应支持30 FPS，宜支持60 FPS
8		色深	应支持10 bit，宜支持12 bit

表 3 L3视频保真度指标要求（续）

序号	环节	指标名称	指标要求
9	编解码	色域	应支持DCI-P3，宜支持BT.2020
10		图像格式	YUV444
11		编解码器	a) 应支持AVC High10 profile，7680×4320分辨率下的码率≥20M bps@30FPS
			b) 应支持HEVC main10 profile，且应支持Tile及MCTSTMVP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) 应支持兼容YUV420/YUV422编解码	
		f) 宜支持AVS	
12		HDR Vivid	宜支持
13		辅流分辨率	应支持3840×2160，宜支持7680×4320
14	传输	主流数量	≥3 应支持把多路码流合成一路码流
15		辅流数量	≥3 应支持把多路码流合成一路码流
16		端到端时延	≤300 ms
17		抗丢包	≥50%，且最大连续丢包个数≥14
18	显示	分辨率	应支持3840×2160，宜支持7680×4320
19		帧率	90 FPS
20		对比度	≥10000:1
21		色深	应支持12 bit，宜支持16 bit
22		色域	应支持DCI-P3，宜支持BT.2020
23		纵横比	应支持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 音频保真度

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

序号	环节	指标名称		指标要求
1	采集	采样率		≥16K Hz
2		采样位深		16 bit
3		发送方向客观MOS分数 ^a		≥3.3
4		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
5		发送方向10dB带宽		满足100 Hz~7K Hz
6	采集	发送方向 频率响应 (频响容 忍曲线)	100 Hz	-30 dB~6 dB
7			200 Hz	-7 dB~6 dB
8			2K Hz	-7 dB~6 dB
9			5K Hz	-7 dB~6 dB
10			6.3K Hz	-11 dB~6 dB
11			7.5K Hz	-11 dB~6 dB
12			8K Hz	上限6 dB
13			发送方向语音能量	
14	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≤10 dB b) 干扰峰值≤-80 dBV	

表 4 L1音频保真度指标要求 (续1)

序号	环节	指标名称		指标要求
15	采集	发送方向 总失真	224-282 Hz	≥30 dB
16			282-355 Hz	≥30 dB
17			355-447 Hz	≥30 dB
18			447-562 Hz	≥30 dB
19			562-708 Hz	≥30 dB
20			708-891 Hz	≥30 dB
21			891-1122 Hz	≥30 dB
22			1122-1413 Hz	≥30 dB
23			1413-1778 Hz	≥30 dB
24			采集	发送方向 总失真
25	2239-2818 Hz	≥30 dB		
26	2818-3548 Hz	≥30 dB		
27	3548-4467 Hz	≥30 dB		
28	4467-5623 Hz	≥28 dB		
29	采集	回声耦合损耗		混响房间下, 标称音量≥50 dB, 最大音量≥50 dB
30		双讲情况下的回声控制特性 ^b		a) 混响房间下, 正常音量: A1+A2占比≥50%; F占比≤5%, G占比≤5% b) 混响房间下, 最大音量: A1+A2占比≥30%; F占比≤8%, G占比≤8%
31		双讲情况下的发送语音衰减		a) 混响房间下, 在使用距离下, 标称音量应≤12 dB b) 混响房间下, 在使用距离下, 最大音量应≤15 dB

表 4 L1音频保真度指标要求（续2）

序号	环节	指标名称		指标要求	
32	采集	语音通话建立后的AEC收敛时间		混响房间，通话建立后发送方向的信号能量降到-60 dBFS的时间应≤5 s	
33	编解码	编解码器		支持G.722、Opus	
34		采样率		支持16K Hz	
35	传输	端到端时延		≤400 ms	
36		抗丢包		≥50%，且最大连续丢包个数≥10	
37	播放	采样率		≥16K Hz	
38		采样位深		16 bit	
39		接收方向客观 MOS 分数 ^o		≥3.6	
40		接收方向 输出声压 级	耳机场景		76±2 dB SPL
41			免提场 景（扬声 器距离）	1 m	64±2 dB SPL
42				1.5 m	67±2 dB SPL
43				2.5 m	69±2 dB SPL
44				3.5 m	71±2 dB SPL
45				4.5 m	73±2 dB SPL
46				7.5 m	76±2 dB SPL
47		接收方向 10dB 带宽		100 Hz~7K Hz	
48		接收方向 频率响应 （频响容 忍曲线）	100 Hz		-13 dB~6 dB
49			200 Hz		13 dB~6 dB
50			250 Hz		-11 dB~6 dB
51			1K Hz		-7 dB~6 dB
52			5K Hz		-7 dB~6 dB
53	6.3K Hz		-11 dB~6 dB		
54	7.5K Hz		-11 dB~6 dB		
55	8K Hz		上限 7 dB		

表 4 L1音频保真度指标要求 (续3)

序号	环节	指标名称		指标要求	
56	播放	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 ≤ 10 dB b) 干扰峰值 ≤ 24 dBSPL	
57		接收方向 输出总失真	免提场景	224-282 Hz	≥ 20 dB
58				282-355 Hz	≥ 20 dB
59				355-447 Hz	≥ 22 dB
60				447-562 Hz	≥ 24 dB
61				562-708 Hz	≥ 24 dB
62		接收方向 输出总失真	免提场景	708-891 Hz	≥ 24 dB
63				891-1122 Hz	≥ 24 dB
64				1122-1413 Hz	≥ 24 dB
65				1413-1778 Hz	≥ 24 dB
66				1778-2239 Hz	≥ 24 dB
67				2239-2818 Hz	≥ 24 dB
68				2818-3548 Hz	≥ 24 dB
69				3548-4467 Hz	≥ 24 dB
70				4467-5623 Hz	≥ 24 dB
71	耳机场景			在免提场景基础上加 8 dB	
<p>^a 按照ITU-T P.863的方法, 用MOS-LQO评分评估发送方向音频质量。</p> <p>^b 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。</p> <p>^c 按照ITU-T P.863的方法, 用MOS-LQO评分评估接收方向音频质量。</p>					

7.2.2 L2 音频保真度

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

序号	环节	指标名称		指标要求
1	采集	采样率		≥32K Hz
2		采样位深		应支持16 bit, 宜支持24 bit
3		发送方向客观MOS分数		≥3.6
4		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
5		发送方向10dB带宽		100 Hz~14K Hz
6		发送方向频率响应 (频响容忍曲线)	100 Hz	-30 dB~5 dB
7			200 Hz	-6 dB~5 dB
8			2K Hz	-6 dB~5 dB
9			7.5K Hz	-6 dB~5 dB
10			12.5K Hz	-6 dB~5 dB
11			14K Hz	-11 dB~5 dB
12			16K Hz	上限5 dB
13			发送方向语音能量	
14		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≤10 dB b) 干扰峰值≤-80 dBV
15		发送方向总失真	224-282 Hz	≥33 dB
16			282-355 Hz	≥33 dB
17			355-447 Hz	≥33 dB

表 5 L2音频保真度指标要求 (续1)

序号	环节	指标名称	指标要求	
18	采集	发送方向总失真	447-562 Hz	≥ 33 dB
19			562-708 Hz	≥ 33 dB
20			708-891 Hz	≥ 33 dB
21			891-1122 Hz	≥ 33 dB
22			1122-1413 Hz	≥ 33 dB
23			1413-1778 Hz	≥ 33 dB
24			1778-2239 Hz	≥ 33 dB
25			2239-2818 Hz	≥ 33 dB
26			2818-3548 Hz	≥ 33 dB
27			3548-4467 Hz	≥ 33 dB
28			4467-5623 Hz	≥ 28 dB
29			回声耦合损耗	混响房间下, 标称音量 ≥ 56 dB, 最大音量 ≥ 56 dB
30			双讲情况下的回声控制特性	<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 dB$\leq \Delta L \leq 4$ dB ;</p> <p>A2情况: 发送端电平变化-15 dB$\leq \Delta L < -4$ dB , F情况: 突发漏回声, G情况: 持续漏回声</p>
31	双讲情况下的发送语音衰减	<p>a) 混响房间下, 在使用距离下, 标称音量应≤ 9 dB</p> <p>b) 混响房间下, 在使用距离下, 最大音量应≤ 12 dB</p>		

表 5 L2音频保真度指标要求（续2）

序号	环节	指标名称		指标要求	
32	采集	语音通话建立后的AEC收敛时间		混响房间, 通话建立后发送方向的信号能量降到-60 dBFS以下的时间应 ≤ 4 s	
33	编解码	编解码器		应支持G.722、Opus、AAC_LD	
34		采样率		应支持16K Hz、48K Hz	
35	传输	端到端时延		≤ 350 ms	
36		抗丢包		$\geq 60\%$, 且最大连续丢包个数 ≥ 12	
37	播放	采样率		≥ 48 K Hz	
38		采样位深		应支持16 bit, 宜支持24 bit	
39		接收方向客观MOS分数		≥ 3.8	
40		接收方向 输出声压 级	耳机场景	76 ± 2 dB SPL	
41		接收方向 输出声压 级	免提 场景 (扬 声器 距离)	1 m	64 ± 2 dB SPL
42				1.5 m	67 ± 2 dB SPL
43				2.5 m	69 ± 2 dB SPL
44				3.5 m	71 ± 2 dB SPL
45				4.5 m	73 ± 2 dB SPL
46				7.5 m	76 ± 2 dB SPL
47	接收方向10dB带宽		100 Hz~14K Hz		
48	接收方向 频率响应 (频响容 忍曲线)	100 Hz		上限5 dB	
49		180 Hz		-12 dB~5 dB	
50		200 Hz		-10 dB~5 dB	
51		250 Hz		-6 dB~5 dB	
52		7.5K Hz		-6 dB~5 dB	
53		12.5K Hz		-6 dB~5 dB	
54		14K Hz		-11 dB~5 dB	
55		16K Hz		上限5 dB	

表 5 L2音频保真度指标要求 (续3)

序号	环节	指标名称		指标要求	
56	播放	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 ≤ 13 dB b) 干扰峰值 ≤ 24 dBSPL	
57		接收方向 输出总失真	免提 场景	224-282 Hz	≥ 23 dB
58				282-355 Hz	≥ 23 dB
59				355-447 Hz	≥ 25 dB
60				447-562 Hz	≥ 27 dB
61				562-708 Hz	≥ 27 dB
62				708-891 Hz	≥ 27 dB
63				891-1122 Hz	≥ 27 dB
64				1122-1413 Hz	≥ 27 dB
65				1413-1778 Hz	≥ 27 dB
66				1778-2239 Hz	≥ 27 dB
67				2239-2818 Hz	≥ 27 dB
68				2818-3548 Hz	≥ 27 dB
69				3548-4467 Hz	≥ 27 dB
70				4467-5623 Hz	≥ 27 dB
71				耳机场景	

7.2.3 L3 音频保真度

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

序号	环节	指标名称	指标要求	
1	采集	采样率	$\geq 48\text{K Hz}$	
2		采样位深	应支持16 bit, 宜支持24 bit	
3		发送方向客观MOS分数	≥ 3.9	
4		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	
5		发送方向10dB带宽	100 Hz~20K Hz	
6		发送方向 频率响应 (频响容忍曲线)	100 Hz	-30 dB~5 dB
7			200 Hz	-6 dB~5 dB
8			2K Hz	-6 dB~5 dB
9			12.5K Hz	-6 dB~5 dB
10			14K Hz	-6 dB~5 dB
11			16K Hz	-11 dB~5 dB
12			20K Hz	上限5 dB
13			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 ≥ -34 dBFS
14		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 ≤ 10 dB b) 干扰峰值 ≤ -80 dBV	
15		发送方向 总失真	224-282 Hz	≥ 36 dB
16			282-355 Hz	≥ 36 dB
17			355-447 Hz	≥ 36 dB
18			447-562 Hz	≥ 36 dB
19			562-708 Hz	≥ 36 dB
20			708-891 Hz	≥ 36 dB

表 6 L3音频保真度指标要求 (续1)

序号	环节	指标名称	指标要求	
21	采集	发送方向 总失真	891-1122 Hz	≥ 36 dB
22			1122-1413 Hz	≥ 36 dB
23			1413-1778 Hz	≥ 36 dB
24			1778-2239 Hz	≥ 36 dB
25			2239-2818 Hz	≥ 36 dB
26			2818-3548 Hz	≥ 36 dB
27			3548-4467 Hz	≥ 36 dB
28			4467-5623 Hz	≥ 30 dB
29		回声耦合损耗	混响房间下, 标称音量 ≥ 56 dB, 最大音量 ≥ 56 dB	
30		双讲情况下的回声控制特性	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 $\leq 8\%$	
31		双讲情况下的发送语音衰减	a) 混响房间下, 使用距离下, 标称音量 ≤ 6 dB b) 混响房间下, 使用距离下, 最大音量 ≤ 9 dB	
32		语音通话建立后的AEC收敛时间	混响房间, 通话建立后发送方向的信号能量降到-60 dBFS以下的时间应 ≤ 3 s	
33		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。	

表 6 L3音频保真度指标要求（续2）

序号	环节	指标名称		指标要求	
34	编解码	编解码器		应支持G.722、Opus、AAC_LD 宜支持分层编解码Codec	
35		采样率		应支持16K Hz、48K Hz	
36		Audio Vivid		宜支持	
37	传输	端到端时延		≤300 ms	
38		抗丢包		≥70%，且最大连续丢包个数≥14	
39	播放	采样率		≥48K Hz	
40		采样位深		支持16 bit，宜支持24 bit	
41		接收方向客观MOS分数		≥4	
42		接收方 向输出 声压级	耳机场景		76±2 dBSPL
43			免提场 景 (扬声 器距离)	1 m	64±2 dBSPL
44				1.5 m	67±2 dBSPL
45				2.5 m	69±2 dBSPL
46				3.5 m	71±2 dBSPL
47				4.5 m	73±2 dBSPL
48				7.5 m	76±2 dBSPL
49			接收方向10dB带宽		100 Hz~20K Hz
50		接收方 向频率 响应	100 Hz		上限6 dB
51			200 Hz		-10 dB~5 dB
52			250 Hz		-6 dB~5 dB
53			12.5K Hz		-6 dB~5 dB
54	14K Hz		-6 dB~5 dB		
55	16K Hz		-11 dB~5 dB		
56	20K Hz		上限5 dB		

表 6 L3音频保真度指标要求 (续3)

序号	环节	指标名称		指标要求		
57	播放	接收方向单频干扰		应支持如下条件之一： a) 干扰峰值电平减去宽带噪声信号电平 ≤ 16 dB b) 干扰峰值 ≤ 24 dB SPL		
58		接收方向输出总失真	免提场景	224-282 Hz	≥ 26 dB	
59				282-355 Hz	≥ 26 dB	
60				355-447 Hz	≥ 28 dB	
61				447-562 Hz	≥ 30 dB	
62				562-708 Hz	≥ 30 dB	
63				708-891 Hz	≥ 30 dB	
64				891-1122 Hz	≥ 30 dB	
65				1122-1413 Hz	≥ 30 dB	
66				1413-1778 Hz	≥ 30 dB	
67				1778-2239 Hz	≥ 30 dB	
68				2239-2818 Hz	≥ 30 dB	
69				2818-3548 Hz	≥ 30 dB	
70				3548-4467 Hz	≥ 30 dB	
71				4467-5623 Hz	≥ 30 dB	
72				耳机场景		在免提场景基础上加6 dB

8 测试方法

8.1 视频保真度测试方法

8.1.1 视频保真度

超高清远程交互系统在采集、编解码、传输和显示技术环节，针对视频保真度技术要求，测试方法如表7所示。

表 7 视频保真度技术要求对应测试方法

序号	环节	指标名称	测试方法
1	采集	分辨率	按照YD/T 4878.2-2024 中 9.1.1进行。
2		帧率	按照YD/T 4878.2-2024 中 9.1.18进行。
3		色深	—
4		色域	按照YD/T 4878.2-2024 中 9.1.7进行，将sRGB替换成视频保真度分级要求中的色域，计算该色域下色彩准确度。
5		时延	按照YD/T 4878.2-2024 中 9.1.20进行。
6	编解码	分辨率	—
7		帧率	—
8		色深	—
9		色域	—
10		图像格式	—
11		编解码器	—
12		HDR Vivid	—
13		辅流分辨率	—
14	传输	主流数量	—
15		辅流数量	—
16		端到端时延	按照8.1.2进行。
17		抗丢包	按照8.1.3进行。
18	显示	主流分辨率	按照YD/T 4878.2-2024 中 9.2.2.10进行。
19		帧率	—
20		对比度	按照SJ/T11348-2016 中 5.2进行。
21		色深	—
22		色域	按照YD/T 4878.2-2024 中 9.2.2.6进行。
23		纵横比	—

8.1.2 端到端时延

8.1.2.1 组网

测试设备和组网如图5所示，使用D65 300Lux光源进行测试。

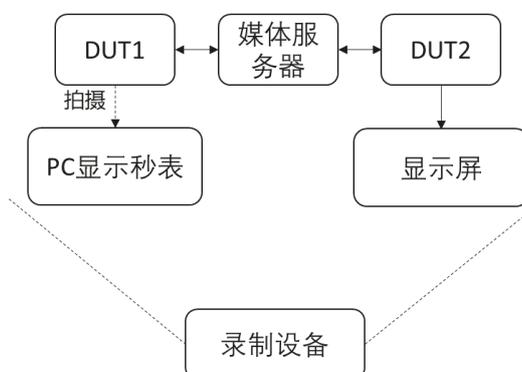


图 5 端到端时延测试环境和组网

8.1.2.2 测试方法

测试步骤如下：

- a) 在PC上播放秒表计时器，最小单位毫秒；
- b) DUT1拍摄PC屏幕；
- c) DUT2的显示输出接显示屏；
- d) 使用240 FPS的录像设备，同时拍摄PC显示秒表和DUT2的显示屏，连续录制10分钟；
- e) 计算录像文件所有帧时延的平均值，计算公式如下：

$$\text{端到端时延} = \Sigma (\text{DUT2显示屏的秒表时间} - \text{PC屏幕上秒表的时间}) / 10\text{分钟总帧数}$$

8.1.3 抗丢包

8.1.3.1 组网和参数配置

抗丢包测试组网如图6所示，按照JVET-J1010文档中定义的Class E测试序列，采用双线性插值方法，将测试序列分别缩放到1080P和2160P作为输入视频。



图 6 抗丢包测试组网

被测DUT建立两次视频交互，参数配置如表8所示。

表 8 参数配置

分辨率	帧率 (FPS)	编码格式	码率
1920×1080	30	HEVC main profile	1M bps
3840×2160	30	HEVC main profile	4M bps

8.1.3.2 测试方法

测试步骤如下：

a) 通过丢包仪，分别设置DUT1的上行丢包率和最大连续丢包个数：30%和10、40%和12、50%和14，其中最大连续丢包每分钟设置一次。

b) 满足以下两个测试条件，则满足要求：

- 1) 接收DUT录制10分钟视频数据，平均帧率 ≥ 25 FPS；
- 2) 在1080P配置下，分辨率 $\geq 720P$ ；在2160P配置下，分辨率 $\geq 1440P$ 。

8.2 音频保真度测试方法

8.2.1 音频保真度

超高清远程交互系统在采集、编解码、传输和播放技术环节，针对音频保真度技术要求，测试方法如表9所示。

表 9 音频保真度技术要求对应测试方法

序号	环节	指标名称	测试方法
1	采集	采样率	—
2		采样位深	—
3		发送方向客观MOS分数	按照YD/T 4878.2-2024 中 7.1.1进行。
4		噪声场景下的语音质量及抑制噪声性能	按照YD/T 4878.2-2024 中 7.6进行。
5		发送方向10dB带宽	按照YD/T 4878.2-2024 中 7.1.6进行。
6		发送方向频率响应	按照YD/T 4878.2-2024 中 7.1.6进行。
7		发送方向语音能量	按照YD/T 4878.2-2024 中 7.1.3进行。

表 9 音频保真度技术要求对应测试方法（续）

序号	环节	指标名称	测试方法
8	采集	发送方向在安静环境下的单频干扰	按照YD/T 4878.2-2024 中 7.1.5进行。
9		发送方向总失真	按照8.2.2 进行。
10		回声耦合损耗	按照YD/T 4878.2-2024 中 7.3.1进行。
11		双讲情况下的回声控制特性	按照YD/T 4878.2-2024 中 7.3.2进行。
12		双讲情况下的发送语音衰减	按照YD/T 4878.2-2024 中 7.3.3进行。
13		语音通话建立后的AEC收敛时间	按照YD/T 4878.2-2024 中 7.5.6进行。
14		个性化降噪	—
15	编解码	编解码器	—
16		采样率	—
17	传输	端到端时延	按照YD/T 4878.2-2024 中 7.2.2进行。
18		抗丢包	按照8.2.3和8.2.4进行。
19	播放	采样率	—
20		采样位深	—
21		接收方向客观MOS分数（	按照YD/T 4878.2-2024 中 7.2.1进行。
22		接收方向输出声压级	按照YD/T 4878.2-2024 中 7.2.3进行。
23		接收方向10dB带宽	按照YD/T 4878.2-2024 中 7.1.6进行。
24		接收方向频率响应	按照YD/T 4878.2-2024 中 7.2.8进行。
25		接收方向单频干扰	按照YD/T 4878.2-2024 中 7.2.6进行。
26		接收方向输出总失真	按照YD/T 4878.2-2024 中 7.2.7进行。

8.2.2 发送方向总失真

8.2.2.1 设备安装

消音室中所用的测试桌应为硬质结构，测试桌表面面积应大于0.8 平米，且各边长不小于0.8 米。被测设备应摆放如图7所示，设备的摆放朝向应使得位于头躯模拟器方位的用户面向屏幕。

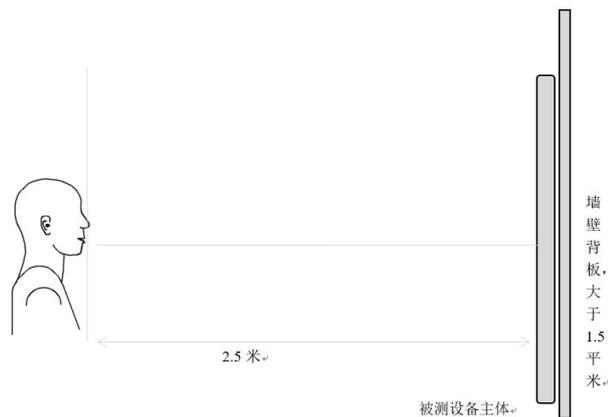


图 7 发送方向总失真测试设备摆放示意图

8.2.2.2 测试方法

测试步骤如下：

a) 在开始测量之前，播放89 dB SPL级别的语音信号，保持超过15 秒的稳定时间。测试信号使用ITU-T P.501中Annex C.2.3的中文真人语音信号（两男两女），按照ITU-T P.863的要求，使用48k Hz采样，按照ITU-T G.191的要求滤波至50 Hz~14k Hz频带宽度；

b) 测试刺激信号是带有频带限制的白噪声，限制在各自的1/3倍频程内，并具有250 ms的开启（活跃期）和150 ms的关闭（调节期）。该序列应重复10次，总共持续4 s。此测试信号应使被测设备在测量期间处于明确的、可重复的状态；

c) 在主参考点处应用测试信号，按照7.2节表中指定的级别和频率进行测试；

d) 计算信号功率与信号输出的总A加权失真噪声功率的比率；

e) 对每次测量的噪声爆发的最后200 毫秒使用汉宁窗进行处理，并进行5Hz分辨率的快速傅里叶变换，频率范围从100 Hz到8000 Hz。经过时间平均后，去除缺口频带内的数据；

d) 将每个频带的结果与要求进行比较。

8.2.3 最大抗随机丢包

8.2.3.1 组网

待测设备、网络路由器、丢包仪的组网方式如图 8 所示。

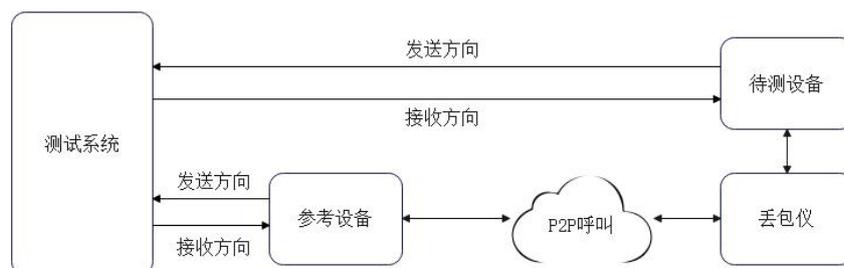


图 8 最大抗随机丢包测试组网

8.2.3.2 测试方法

测试步骤如下：

a) 测试信号使用ITU-T P.501中Annex C.2.3的中文真人语音信号（两男两女），按照ITU-T P.863的要求，使用48k Hz采样，按照ITU-T G.191的规定滤波至50 Hz~14k Hz频带宽度。该中文真人信号应按照8 s一个片段，组成长为32 s的测试语音。每个人的语音应位于8 s片段的中心位置。将测试音频复制1份，2组构成64 s的测试语音；

b) 向视频交互参考设备输入该真人语音信号，根据ITU-T P.56计算的信号活动语音电平为其信号能量大小为-16 dBm0。播放语音的同时在交互对端的视频交互测试设备处录制输出的语音信号，以测试信号为参考信号，录得的语音为受损信号；

c) 设置丢包仪带宽500k bps，设定不同随机丢包率，如：10%、20%、30%、40%、50%、60%、65%、70%等；

d) 按照ITU-T P.863标准定义的方法，计算后一组32s测试语音的平均MOS-LQO分数；

e) 记录平均MOS-LQO分数 ≥ 2.5 的最大随机丢包率。

8.2.4 最大抗连续丢包

按照 8.2.3.1 的要求组网。

测试步骤如下：

a) 测试信号使用ITU-T P.501中Annex C.2.3的中文真人语音信号（两男两女），按照ITU-T P.863的要求，使用48k Hz采样，按照ITU-T G.191的规定滤波至50 Hz~14k Hz频带宽度。该中文真人信号应按照8 s一个片段，组成长为32 s的测试语音。每个人的语音应位于8 s片段的中心位置；

b) 设置丢包仪带宽500k bps，设置5种循环丢包策略，分别为每间隔1 s丢包10包、每间隔1 s丢包11包、每间隔1 s丢包12包、每间隔1 s丢包13包、每间隔 1s丢包14包。每种丢包策略测试一次；

c) 人工听取录音，判断测试音频中最后一组8 s语音中是否存在丢包卡顿；

d) 记录无丢包卡顿的最大连续丢包数目。
