

ICS 33.050

CCS M 30

团 体 标 准

T/TAF 307—2025

会议系统及终端音视频质量及分级 评价技术规范

Technical specifications for audio and video quality and grading
evaluation of conference systems and terminals

2025-08-11 发布

2025-08-11 实施

电信终端产业协会 发布

版权声明

本文件的版权属于电信终端产业协会，任何单位和个人未经许可，不得进行技术文件的纸质和电子等任何形式的复制、印刷、出版、翻译、传播、发行、合订和宣贯等，也不得未经允许采用其具体内容编制本团体以外各类标准和技术文件。如有以上需要请与本团体联系。

邮箱：tafrb@taf.org.cn

电话：010-82052809



目 次

前言	III
引言	IV
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 缩略语	2
5 系统架构	2
5.1 系统组网	2
5.2 终端设备类型	3
6 评价指标体系框架	3
6.1 评价对象	3
6.2 评价指标体系框架	3
7 评价指标技术要求	4
7.1 会议终端评价指标	4
7.1.1 音频指标	4
7.1.2 视频指标	6
7.2 会议系统评价指标	9
7.2.1 音频指标	9
7.2.2 视频指标	11
8 评价指标测试方法	14
8.1 会议终端测评	14
8.1.1 测试环境	14
8.1.2 音频测评	18
8.1.3 视频测评	21
8.2 会议系统测评	25
8.2.1 测试环境	25
8.2.2 音频测评	31
8.2.3 视频测评	34
9 分级评价方法	37
9.1 指标权重	37
9.2 得分判定	37
9.3 计算公式	37
9.4 等级判定	38

附录 A（规范性）	会议系统及终端音视频质量评价指标权重.....	39
附录 B（规范性）	音视频主观等级评分标准.....	41



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》的规定起草。

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专利的责任。

本文件由电信终端产业协会（TAF）提出并归口。

本文件起草单位：中国信息通信研究院、华为技术有限公司、中兴通讯股份有限公司、北京邮电大学、广州视源电子科技股份有限公司、维沃移动通信有限公司、深圳高新兴瑞联科技有限公司、泰尔认证中心有限公司、博鼎实华（北京）技术有限公司。

本文件主要起草人：王亚军、翟梦冉、赵兴龙、陈显义、张磊、徐海、耿东玉、彭灿辉、庞文全、周广荣、张宏伟、闫石、李勤、牛永耀、高立发、高纪、薛刚、李杰强、马凡。



引 言

随着计算机和网络技术的发展，视频会议越来越普及。视频会议广泛应用于办公、政务、教育等场景，逐渐成为政府、企业、办公的必备基础设施。

现有视频通信框架中，为了降低数据传输带宽，通过对音频数据和视频图像的压缩，来实现音视频交互。视频会议系统支持音视频的采集、编解码和呈现，虽然对会议的音视频质量有直观感受，但是当前业界缺少对真实应用场景的音视频质量的客观评价标准。

本文件旨在引导真实应用场景下的视频会议系统，制定针对会议系统及会议终端的音视频质量评价以及分级的规范，更好的推进视频会议终端及系统的技术发展和规模化应用，为相关行业带来高品质的视频会议体验，促进经济社会的数字化转型。



会议系统及终端音视频质量及分级评价技术规范

1 范围

本文件规定了会议系统及终端音视频质量评估技术要求和测试方法,并给出了相应的评价方法及判定依据。

本文件适用于面积不小于10m²且不大于100m²的会议室,适用于专门支持视频会议及相关业务的视频会议系统和终端产品,终端产品仅包括一体式终端和分体式终端两种形态。

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。

YD/T 2309—2011 音频质量主观测试方法

ITU-T G.122-1993 国内系统对国际连接的稳定度、发话回声的影响(Influence of National Systems on Stability and Talker Echo in International Connections)

ITU-T G.191 语音及音频编码标准软件库(Software tools for speech and audio coding standardization)

ITU-T G.711 脉冲编码调制(Pulse code modulation (PCM) of voice frequencies)

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

ITU-T P.501 电话测量用测试信号(Test signals for use in telephony)

ITU-T P.502 对于语音通信系统使用综合测试信号的客观测试方法(Objective test methods for speech communication systems using complex test signals)

ITU-T P.570 实验室条件下的人工噪声场(Artificial noise fields under laboratory conditions)

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

ISO/CIE 11664-6 色度学 第6部分: CIEDE2000 色差公式(Colorimetry Part 6: CIEDE2000 colour-difference formula)

ISO 12233 数码相机-分辨率和空间频率响应(Digital cameras — Resolution and spatial frequency responses)

ISO 17850 摄影-数码相机-几何畸变(GD)测量(Photography — Digital cameras — Geometric distortion (GD) measurements)

ETSI ES 103 106 在噪声环境下的语音质量性能(Speech quality performance in the presence of background noise: Background noise transmission for mobile terminals-objective test methods)

3 术语和定义

下列术语和定义适用于本文件。

3.1

主流 main video channel

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。

3.2

辅流 second video/presentation channel

视频会议通信中的第二视频流，用于共享屏幕内容、PPT、文档和白板等内容，辅流也称为共享演示。辅流与主流一起，为视频会议提供丰富的信息共享功能。

3.3

双讲 double talk

是指通话的两端出现同时讲话、同时插话、打断对方讲话的情况，或者一端语音与另一端噪声同时存在的情况。

3.4

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

在两端同时讲话时，回声算法引入的发送方向信号的衰减。

3.5

语音平均意见得分 speech-mean opinion score

对抑制后的语音能量进行评分，为主观听音质量评分方式的一种，常用S-MOS表述。

3.6

噪声平均意见得分 noise-mean opinion score

对抑噪后的残留噪声质量进行评分，为主观听音质量评分方式的一种，常用N-MOS表述。

3.7

MTF50P

在图像经过过度锐化处理，MTF 数值的最大值下降到50%时所对应的空间频率。

4 缩略语

下列缩略语适用于本文件。

BMP: 位图文件 (Bitmap-File)

dBFS: 分贝全幅波形 (Decibels Relative to Full Scale)

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

DUT: 被测设备 (Device Under Test)

FPS: 帧率 (Frame Per Second)

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

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

MJPEG: 运动静止图像压缩技术 (Motion Joint Photographic Experts Group)

MTF: 调制传递函数 (Modulation Transfer Function)

PNG: 便携式网络图形 (Portable Network Graphics)

PSNR: 峰值信噪比 (Peak Signal to Noise Ratio)

SSIM: 结构相似性 (Structural Similarity)

5 系统架构

5.1 系统组网

视频会议系统框架见图1。

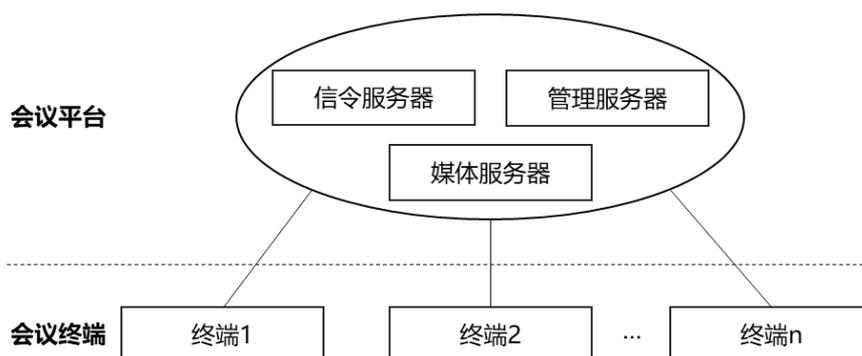


图1 视频会议系统框架图

视频会议系统框架包括会议平台和会议终端，其中媒体服务器和会议终端与音视频质量相关，管理服务器和信令服务器为会议系统的辅助部件。

——管理服务器：

完成视频会议业务的管理功能，实现会议的预约、召开、调度和控制等。

——信令服务器：

完成视频会议的注册或鉴权、呼叫或路由等信令处理，实现会议过程中的呼叫控制等功能。

——媒体服务器：

完成视频会议系统的媒体处理，实现音频、视频及数据的处理，包括：媒体流转发、音视频转码、混音以及画面组合等功能。

——会议终端：

作为会议系统中和用户直接交互的设备，实现音视频的采集、编解码和显示等功能。

5.2 终端设备类型

会议室类型终端主要为以下两种：

——一体式终端：

一体式终端是指集成图像采集能力、音频采集能力、音视频编解码能力、音视频解码能力、辅流采集能力以及音视频输出和呈现能力；

——分体式终端：

分体式终端是指终端的显示屏、摄像头及麦克风等功能可分成多个独立的组件，通过连接组合成完整的会议终端系统。

6 评价指标体系框架

6.1 评价对象

结合视频会议系统的实际应用场景，测评对象包括视频会议终端和视频会议系统两类。其中视频会议终端包括一体式终端和分体式终端。

6.2 评价指标体系框架

视频会议终端和视频会议系统的评价指标体系框架各由两大部分组成，其中，视频会议终端评价指标体系框架见图2，视频会议系统评价指标体系框架见图3。

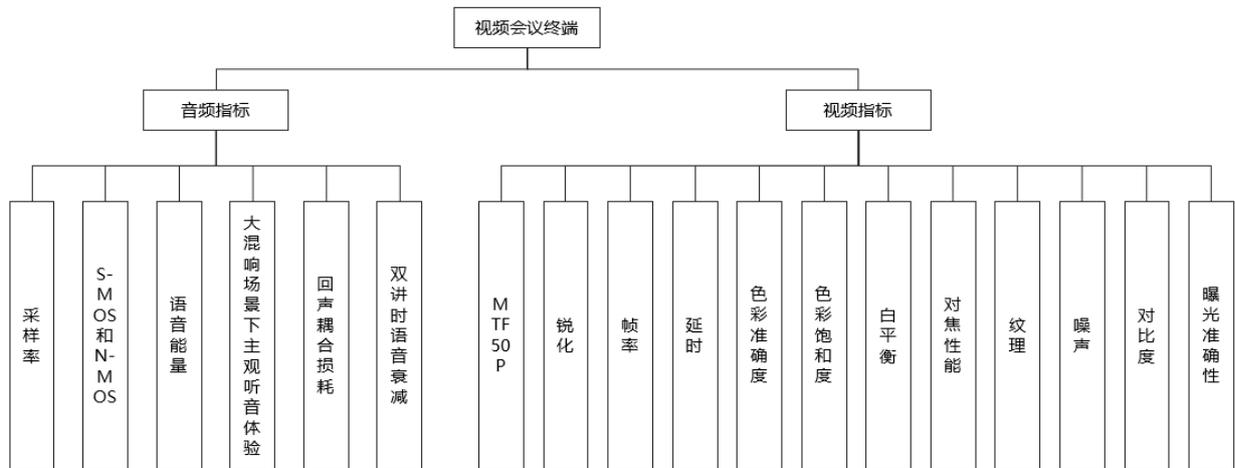


图2 视频会议终端评价指标体系框架

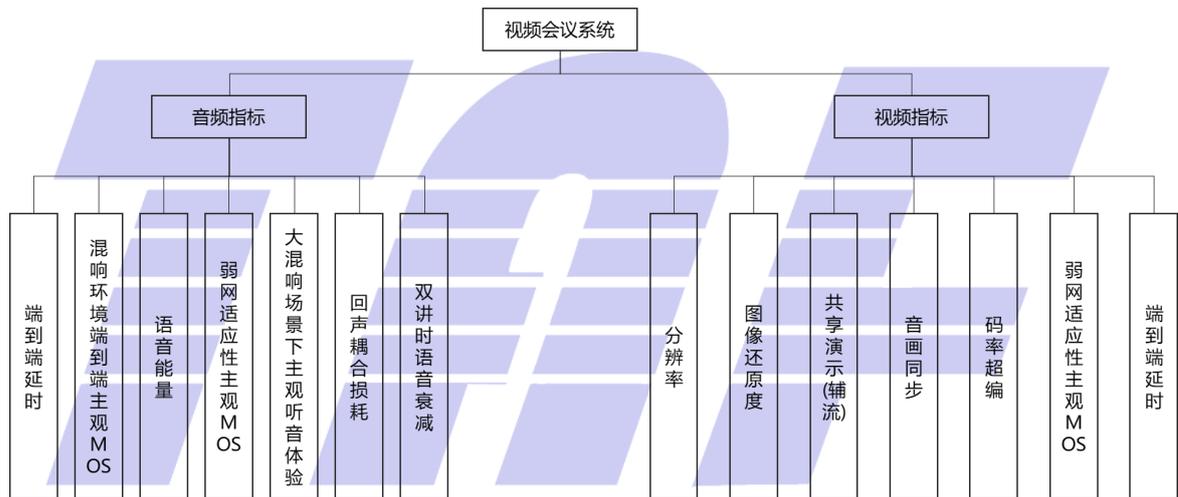


图3 视频会议系统评价指标体系框架

7 评价指标技术要求

7.1 会议终端评价指标

7.1.1 音频指标

7.1.1.1 采样率

采样率决定了语音采集的频率范围和精度，其直接影响语音信号的质量和可用性，会议终端的采样率需要满足如下三种要求的一种，采样率技术要求见表1。

表1 采样率技术要求

指标	优秀	良好	一般
采样率	≥48k	≥32k	≥16k

7.1.1.2 S-MOS 和 N-MOS

在视频会议实际应用中，会议室常规拾音距离包括6m和10m，因此会议终端的S-MOS和N-MOS需要满足如下三种要求的一种，S-MOS和N-MOS技术要求见表2。

表2 S-MOS和N-MOS技术要求

指标	拾音距离	优秀	良好	一般
S-MOS	6 m	≥ 3.8	≥ 3.5	≥ 3.2
N-MOS		≥ 3.8	≥ 3.5	≥ 3.2
S-MOS	10 m	≥ 3.5	≥ 3.2	≥ 2.9
N-MOS		≥ 3.8	≥ 3.5	≥ 3.2

7.1.1.3 语音能量

声音能量过小，将会加大用户聆听语音的困难程度；同时能量过大又可能导致语音质量损伤。因此会议终端的语音能量需要满足如下三种要求的一种，语音能量技术要求见表3。

表3 语音能量技术要求

指标	说话人音量	优秀	良好	一般
		发送方向语音平均能量（100Hz-14kHz频带范围内）		
发送方向语音能量	正常音量	[-23, -14] dBFS	[-26, -16] dBFS	[-31, -18] dBFS
	较低音量	≥ -34 dBFS	≥ -37 dBFS	≥ -40 dBFS
接收方向语音能量	正常音量	≥ -68 dBSPL	≥ -65 dBSPL	≥ -62 dBSPL

7.1.1.4 大混响场景下主观听音体验

对于写字楼等办公场景，会议室装修墙面多采用玻璃材质，其房间混响较大，因此会议终端对大混响场景下主观听音体验需要满足如下三种要求的一种，大混响场景下主观听音体验技术要求见表4。

表4 大混响场景下主观听音体验技术要求

指标	优秀	良好	一般
拾音主观 MOS	≥ 3.8	≥ 3.3	≥ 2.8
放音主观 MOS	≥ 3.8	≥ 3.3	≥ 2.8

7.1.1.5 回声耦合损耗

如果设备对回声的抑制不足，那么远端用户在其说话时将持续听到自身语音，从而打扰语音对话。因此会议终端对回声耦合损耗需要满足如下三种要求的一种，回声耦合损耗技术要求见表5。

表5 回声耦合损耗技术要求

指标	优秀	良好	一般
标称音量耦合损耗	≥ 60 dB	≥ 55 dB	≥ 50 dB

表5 回声耦合损耗技术要求（续）

指标	优秀	良好	一般
最大音量耦合损耗	≥60 dB	≥55 dB	≥50 dB

7.1.1.6 双讲时语音衰减

双讲是视频会议的常见使用方式，会议终端对双讲时语音衰减要满足如下三种要求的一种，双讲时语音衰减技术要求见表6。

表6 双讲时语音衰减技术要求

指标	扬声器音量	优秀	良好	一般
双讲时的语音衰减	标称音量	≤-6 dB	≤-9 dB	≤-12 dB
	最大音量	≤-9 dB	≤-12 dB	≤-15 dB

7.1.2 视频指标

7.1.2.1 MTF50P

视频清晰度直接影响视频会议效果，因此会议终端对调制传递函数在中心10%视场处需要满足如下三种要求的一种，MTF50P技术要求见表7。

表7 MTF50P技术要求

指标	分辨率	光源条件	优秀	良好	一般
MTF50P	4K或1080P	D65 300 lx	MTF50P ≥ 0.4 cycles/pixel	MTF50P ≥ 0.35 cycles/pixel	MTF50P ≥ 0.3 cycles/pixel
		CWF 80 lx	MTF50P ≥ 0.35 cycles/pixel	MTF50P ≥ 0.32 cycles/pixel	MTF50P ≥ 0.29 cycles/pixel

7.1.2.2 锐化

过度锐化（或不足）会导致图像生硬及边缘振铃等伪像（或模糊图像），因此会议终端对锐化的要求如下，应满足三种要求的一种，其中指标是指中心方块所有边的过度锐化或锐化不足的百分比，锐化技术要求见表8。

表8 锐化技术要求

指标	分辨率	光源条件	优秀	良好	一般
锐化	4K或1080P	D65 300 lx CWF 80 lx	≤10%	≤12%	≤15%

7.1.2.3 帧率

为了保障整个图像画面内的流畅性，会议终端的帧率需满足如下三种要求的一种，帧率技术要求见表9。

表9 帧率技术要求

指标	分辨率	光源条件	优秀	良好	一般
帧率	4K或1080P	300 lx	Frame rate ≥ 30 fps	Frame rate ≥ 28 fps	Frame rate ≥ 25 fps
		80 lx	Frame rate ≥ 30 fps	Frame rate ≥ 25 fps	Frame rate ≥ 20 fps

7.1.2.4 延时

为了保障端到端会议延时足够小，会议终端摄像头不可引入过大的延时，延时需满足如下三种要求的一种，延时技术要求见表10。

表10 延时技术要求

指标	分辨率	光源条件	优秀	良好	一般
延时	4K或1080P	300 lx	≤ 150 ms	≤ 180 ms	≤ 210 ms

7.1.2.5 色彩准确度

会议终端设备对还原真实的色彩和肤色至关重要，为保证摄像系统的颜色真实性，需通过色卡进行色彩准确度测试，对色卡的色差要求如下，应满足三种要求的一种，色彩准确度技术要求见表11。

表11 色彩准确度技术要求

指标	分辨率	光源条件	优秀	良好	一般
色彩准确度	4K或1080P	H 80 lx, A300 lx	Max $\Delta C \leq 10$ 且Mean $\Delta C \leq 8$	Max $\Delta C \leq 12$ 且Mean $\Delta C \leq 10$	Max $\Delta C \leq 14$ 或Mean $\Delta C \leq 12$
		CWF 80 lx, CWF 300 lx, D50 80 lx, D50 300 lx, D65 300 lx, D65 700 lx	Max $\Delta C \leq 8$ 且Mean $\Delta C \leq 5$	Max $\Delta C \leq 10$ 且Mean $\Delta C \leq 7$	Max $\Delta C \leq 12$ 或Mean $\Delta C \leq 9$

7.1.2.6 色彩饱和度

会议终端对色彩饱和度的要求如下，应满足三种要求的一种，色彩饱和度技术要求见表12。

表12 色彩饱和度技术要求

指标	分辨率	光源条件	优秀	良好	一般
色彩饱和度	4K或1080P	H 80 lx, A 300 lx, CWF 80 lx, CWF 300 lx, D50 80 lx, D50 300 lx, D65 300 lx, D65 700 lx	Saturation \geq 95%, 且 Saturation \leq 120%	Saturation \geq 90%, 且 Saturation \leq 125%	Saturation \geq 85%, 且 Saturation \leq 130%

7.1.2.7 白平衡

会议终端对白平衡的要求如下，应满足三种要求的一种，白平衡技术要求见表13。

表13 白平衡技术要求

指标	分辨率	光源条件	优秀	良好	一般
白平衡	4K或1080P	H 80 lx, A 300 lx	Max $\Delta C \leq 10$	Max $\Delta C \leq 12$	Max $\Delta C \leq 14$
		CWF 80 lx, CWF 300 lx, D50 80 lx, D50 300 lx, D65 300 lx, D65 700 lx	Max $\Delta C \leq 5$	Max $\Delta C \leq 6$	Max $\Delta C \leq 7$

7.1.2.8 对焦性能

定焦镜头的会议终端设备：测试时，其实时拍摄的画面和视频，没有出现虚焦和图像画面不清晰的情况判为优秀；出现虚焦或图像画面不清晰的情况判为不合格。

自动对焦的会议终端设备：其对焦收敛时间和稳定性直接影响视频效果，会议终端对自动对焦的要求如下，应满足三种要求的一种。测试时，需要在真实场景中持续5min测试，记录对焦时间所占整体时间的百分比S和失焦后快速对焦的时间t，自动对焦技术要求见表14。

表14 自动对焦技术要求

指标	分辨率	光源条件	优秀	良好	一般
自动对焦	4K或1080P	300 lx	S \geq 99%, 且t \leq 1 s	S \geq 99%, 且t \leq 2 s	S \geq 99%, 且t \leq 3 s

7.1.2.9 纹理

会议终端对纹理的要求如下，应满足三种要求的一种，纹理技术要求见表15。

表15 纹理技术要求

指标	分辨率	光源条件	优秀	良好	一般
纹理	4K或 1080P	D65 300 lx	$Acutance \geq 0.8$	$Acutance \geq 0.78$	$Acutance \geq 0.76$
		CWF 80 lx	$Acutance \geq 0.7$	$Acutance \geq 0.68$	$Acutance \geq 0.66$

7.1.2.10 噪声

会议终端对噪声的要求如下，应满足三种要求的一种，噪声技术要求见表16。

表16 噪声技术要求

指标	分辨率	光源条件	优秀	良好	一般
噪声	4K或 1080P	D65 300 lx CWF 80 lx	$SNR \geq 40$ dB	$SNR \geq 38$ dB	$SNR \geq 36$ dB

7.1.2.11 对比度

会议终端对视频的对比度要求如下，应满足三种要求的一种，对比度技术要求见表17。

表17 对比度技术要求

指标	分辨率	光源条件	优秀	良好	一般
对比度	4K或 1080P	D65 300 lx CWF 80 lx D65 700 lx	$62\% \leq Contrast \leq 68\%$	$60\% \leq Contrast \leq 70\%$	$55\% \leq Contrast \leq 75\%$

7.1.2.12 曝光准确性

为保证在各亮度条件下不会出现过曝光和欠曝光，会议终端对曝光准确性的要求如下，应满足三种要求的一种，曝光准确技术要求见表18。

表18 曝光准确性技术要求

指标	分辨率	光源条件	优秀	良好	一般
曝光准确度	4K或 1080P	D65 300 lx, CWF 80 lx	$100 \leq Y \leq 142$	$100 \leq Y \leq 155$	$100 \leq Y \leq 160$

7.2 会议系统评价指标

7.2.1 音频指标

7.2.1.1 端到端延时

会议系统端到端延时需要满足如下三种要求的一种，端到端延时技术要求见表19。

表19 端到端延时技术要求

指标	优秀	良好	一般
接收全链路延时	≤350 ms	≤400 ms	≤450 ms

7.2.1.2 混响环境端到端主观 MOS

端到端音频效果不仅要关注发送方向拾音质量，还要关注接收方向的处理算法及音箱对声音的还原能力。因此会议系统的主观MOS需要满足如下三种要求的一种，混响环境端到端主观MOS技术要求见表20。

表20 混响环境端到端主观MOS技术要求

指标	测试设备	拾音距离	放音距离	优秀	良好	一般
主观 MOS	一体式端到端系统	6 m	6 m	≥ 3.8	≥ 3.5	≥ 3.0
	分体式端到端系统	10 m	6 m	≥ 3.5	≥ 3.0	≥ 2.5
	分体式端到端系统	1.5 m	6 m	≥ 4	≥ 3.5	≥ 3.0

7.2.1.3 语音能量

声音能量过小，将会加大用户聆听语音的困难程度；同时能量过大又可能导致语音质量损伤。因此会议系统的语音能量需要满足如下三种要求的一种，语音能量技术要求见表21。

表21 语音能量技术要求

指标	说话人音量	优秀	良好	一般
语音能量	正常音量	62±3 dB SPL	59±3 dB SPL	56±3 dB SPL
	较低音量	59±3 dB SPL	56±3 dB SPL	53±3 dB SPL

7.2.1.4 弱网适应性主观 MOS

为保证会议体验，端到端会议系统要具备良好的网络适应性，可以抵抗一定范围的延时、网络抖动及丢包，使得音频的连续性及其可懂度不应有明显影响，会议系统需要满足如下要求，应满足三种要求的一种，弱网适应性主观MOS技术要求见表22。

表22 弱网适应性主观MOS技术要求

指标	测试环境	优秀	良好	一般
弱网适应性主观 MOS	50ms 延时 + 30ms 抖动 + 10% 丢包率（包含 8 个连续丢包）	≥ 4	≥ 3.5	≥ 3.0
	100ms 延时 + 30ms 抖动 + 20% 丢包率（包含 10 个连续丢包）	≥ 3.8	≥ 3.3	≥ 2.8

表 22 弱网适应性主观 MOS 技术要求（续）

指标	测试环境	优秀	良好	一般
弱网适应性主观 MOS	150ms 延时 + 50ms 抖动 + 30% 丢包率（包含 12 个连续丢包）	≥ 3.5	≥ 3.0	≥ 2.5

7.2.1.5 大混响场景下主观听音体验

现在越来越多公司在写字楼办公，其中会议室装修墙面多采用玻璃材质，相比于传统经过专业声学设计的会议室，其房间混响会大很多，因此会议系统对大混响场景下主观听音体验需要满足如下三种要求的一种，大混响场景下主观听音体验技术要求见表23。

表 23 大混响场景下主观听音体验技术要求

指标	优秀	良好	一般
拾音主观 MOS	≥ 3.8	≥ 3.3	≥ 2.8
放音主观 MOS	≥ 3.8	≥ 3.3	≥ 2.8

7.2.1.6 回声耦合损耗

如果系统对回声的抑制不足，那么远端用户将在其说话时持续听到自身语音，从而打扰语音对话。因此会议系统对回声耦合损耗需要满足如下三种要求的一种，回声耦合损耗技术要求见表24。

表 24 回声耦合损耗技术要求

指标	近端会场/远端会场扬声器播放音量	优秀	良好	一般
回声耦合损耗	标称音量	≥ 55 dB	≥ 50 dB	≥ 45 dB
	最大音量	≥ 55 dB	≥ 50 dB	≥ 45 dB

7.2.1.7 双讲时语音衰减

双讲是视频会议的常见使用方式，因此会议系统双讲时语音衰减需要满足如下三种要求的一种，双讲时语音衰减技术要求见表25。

表 25 双讲时语音衰减技术要求

指标	扬声器音量	优秀	良好	一般
双讲时语音衰减	标称音量	≤ -6 dB	≤ -9 dB	≤ -12 dB
	最大音量	≤ -9 dB	≤ -12 dB	≤ -15 dB

7.2.2 视频指标

7.2.2.1 分辨率

视频会议系统端到端的最高分辨率要求如下，应满足三种要求的一种，分辨率技术要求见表26。

表26 分辨率技术要求

指标	优秀	良好	一般
分辨率	4K	1080P	720P

7.2.2.2 图像还原度

视频会议系统的图像还原度需要满足如下三种要求的一种，图像还原度技术要求见表27。

表27 图像还原度技术要求

指标	分辨率@带宽	优秀	良好	一般
图像还原度	4K30@4Mbps	PSNR \geq 30	PSNR \geq 28	PSNR \geq 25
	1080P30@1.5Mbps	PSNR \geq 30	PSNR \geq 28	PSNR \geq 25
	720P30@768Kbps	PSNR \geq 30	PSNR \geq 28	PSNR \geq 25

7.2.2.3 共享演示（辅流）

视频会议系统的共享演示（辅流）需要满足如下三种要求的一种，共享演示（辅流）技术要求见表28。

表28 共享演示（辅流）技术要求

指标		分辨率@会议带宽	优秀	良好	一般
共享演示 （辅流）	清晰度(2S 收敛后， 时间按测试刷新)	4K30@4Mbps 或 1080P30@1.5Mbps 或 720P30@768Kbps	PSNR \geq 18	PSNR \geq 16	PSNR \geq 14
	切换响应时间		t \leq 1.2 s	t \leq 1.5 s	t \leq 1.8 s

7.2.2.4 音画同步

视频会议系统的音画同步需要满足如下三种要求的一种，音画同步技术要求见表29。

表29 音画同步技术要求

指标	分辨率	优秀	良好	一般
音画同步	4K或1080P或720P	声音落后 80ms 或超前 45ms	声音落后 90ms 或超前 60ms	声音落后 100ms 或超 前 80ms

7.2.2.5 码率超编

结合实际应用场景，为防止码流突然大幅度变化对网络产生冲击，按使用HEVC编码，视频会议系统的编码码率需要满足如下三种要求的一种，码率超编技术要求见表30。

表30 码率超编技术要求

指标	分辨率@带宽	优秀	良好	一般
码率超编	4K30@4Mbps	≤ 100%	≤ 110%	≤ 120%
	1080P30@1.5Mbps	≤ 120%	≤ 130%	≤ 140%
	720P30@768Kbps	≤ 150%	≤ 160%	≤ 170%

7.2.2.6 弱网适应性主观 MOS

为保证会议体验，端到端会议系统要具备良好的网络适应性，可以抵抗一定范围的延时、网络抖动及丢包，视频弱网适应性主观MOS需要满足如下三种要求的一种，弱网适应性主观MOS技术要求见表31。

表31 弱网适应性主观MOS技术要求

指标	分辨率	测试条件	优秀	良好	一般
弱网适应性 主观 MOS	4K30@4Mbps	50ms 延时 + 30ms 抖动 + 10% 丢包率 (包含 8 个连续丢包)	MOS ≥ 4.0	MOS ≥ 3.5	MOS ≥ 3.0
		100ms 延时 + 30ms 抖动 + 20% 丢包率 (包含 10 个连续丢包)	MOS ≥ 3.5	MOS ≥ 3.0	MOS ≥ 2.8
		150ms 延时 + 50ms 抖动 + 30% 丢包率 (包含 12 个连续丢包)	MOS ≥ 3.0	MOS ≥ 2.8	MOS ≥ 2.5
	1080P30@1.5Mbps	50ms 延时 + 30ms 抖动 + 10% 丢包率 (包含 8 个连续丢包)	MOS ≥ 3.5	MOS ≥ 3.0	MOS ≥ 2.8
		100ms 延时 + 30ms 抖动 + 20% 丢包率 (包含 10 个连续丢包)	MOS ≥ 3.0	MOS ≥ 2.8	MOS ≥ 2.5
		150ms 延时 + 50ms 抖动 + 30% 丢包率 (包含 12 个连续丢包)	MOS ≥ 2.8	MOS ≥ 2.5	MOS ≥ 2.2
	720P30@768Kbps	50ms 延时 + 30ms 抖动 + 10% 丢包率 (包含 8 个连续丢包)	MOS ≥ 3.0	MOS ≥ 2.8	MOS ≥ 2.5
		100ms 延时 + 30ms 抖动 + 20% 丢包率 (包含 10 个连续丢包)	MOS ≥ 2.8	MOS ≥ 2.5	MOS ≥ 2.2
		150ms 延时 + 50ms 抖动 + 30% 丢包率 (包含 12 个连续丢包)	MOS ≥ 2.5	MOS ≥ 2.2	MOS ≥ 2.0

7.2.2.7 端到端延时

会议系统端到端延时需要满足如下三种要求的一种，端到端延时技术要求见表32。

表32 端到端延时技术要求

指标	分辨率	优秀	良好	一般
端到端延时	4K或1080P或720P	Delay ≤ 350 ms	Delay ≤ 400 ms	Delay ≤ 450 ms

8 评价指标测试方法

8.1 会议终端测评

8.1.1 测试环境

8.1.1.1 音频

8.1.1.1.1 测试环境

除特殊规定，所有测试均应在下列条件下进行。

- 环境温度：15℃~35℃。
- 相对湿度：5% RH~85% RH。
- 大气压力：86k Pa~106k Pa。
- 测试环境1：

为匹配实际使用环境，音频测试在混响室进行；为满足测试要求，用于测试房间的长度应大于10 m，宽度应大于4.5 m，高度在2.8 m到3.25 m之间，房间的混响时间应在0.4 s到0.7 s之间。具体各频段混响时间应符合下表，房间内的噪声声级应控制在35 dB SPL (A) 以内，测试环境1各频段混响时间见表33。

表33 测试环境1各频段混响时间

倍频程中心频率	下限(RT60, ms)	上限 (RT60, ms)
125	350	950
250	350	850
500	350	700
1000	350	600
2000	350	600
4000	350	600
8000	300	700
房间平均值	400	700

- 测试环境2：

为匹配玻璃房会议室等大混响环境，还需提供玻璃房会议室测试环境，用于测试房间的长度应大于6 m，宽度应大于3.5 m，高度在2.8 m到3.25 m之间，最终房间的混响时间应在0.9 s到1.1 s之间。

房间具体各频段混响时间应符合下表，房间内的噪声声级应控制在35 dB SPL (A) 以内，测试环境2各频段混响时间见表34。

表34 测试环境2各频段混响时间

倍频程中心频率	下限(RT60, ms)	上限 (RT60, ms)
125	900	1430
250	900	1430
500	900	1100
1000	900	1100

表 34 测试环境 2 各频段混响时间 (续)

倍频程中心频率	下限(RT60, ms)	上限 (RT60, ms)
2000	810	990
4000	810	990
8000	720	880
房间平均值	900	1100

8.1.1.1.2 测试链路

测试链路网络采用有线网络, 应确保测试网络未引入额外损伤, 网络无明显延时抖动和丢包。测试链路图见图4。

链路中参考客户端是与待测终端互通的软件终端, 并安装在测试电脑上, 也可以是与待测终端同样或同类型的终端。测试系统与参考客户端之间可以使用数字音频或者模拟音频方式传输语音信号, 整个链路的电平/音量设置应确保0 dBFS正弦信号数字音频的所对应的电平为3.14 dBm0。

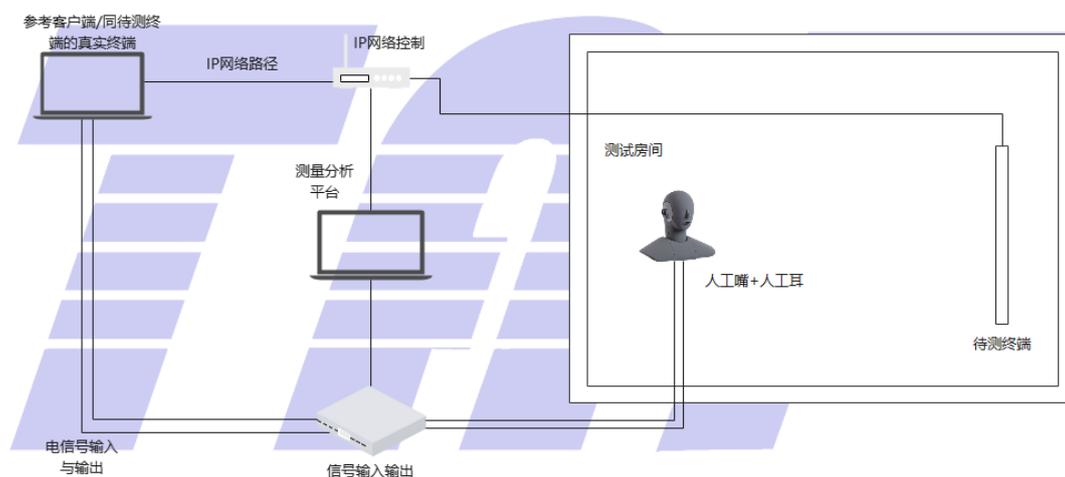


图4 会议终端音频测试组网示意图

8.1.1.1.3 测试终端和系统设置

参考终端的扬声器播放和麦克风增益除特殊说明外采用默认配置。若使用软件参考客户端, 其扬声器播放和麦克风录制音量调到100%。参考端系统扬声器音量调到100%, 不开启任何播放效果器; 系统麦克风录制增益调到100%, 不开启任何麦克风增益效果器。

8.1.1.1.4 测试设备安放

对于一体式会议终端, 其安放示意图见图5。若大屏自身具备支架, 可以将设备安放在支架上进行测试, 否则需要将待测终端安装在墙体上进行测试。测试距离可根据测试项要求对人工嘴进行布置。

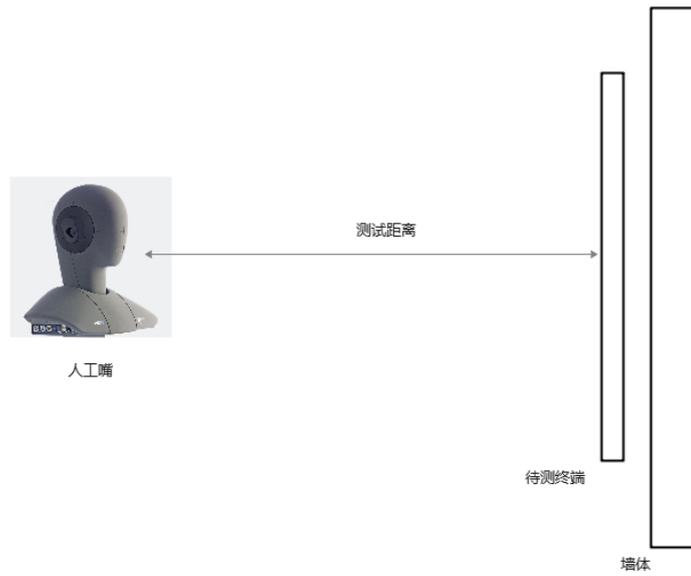


图5 会议终端音频测试设备摆放示意图

8.1.1.2 视频

8.1.1.2.1 测试组网

视频测试组网如图6和图7所示。

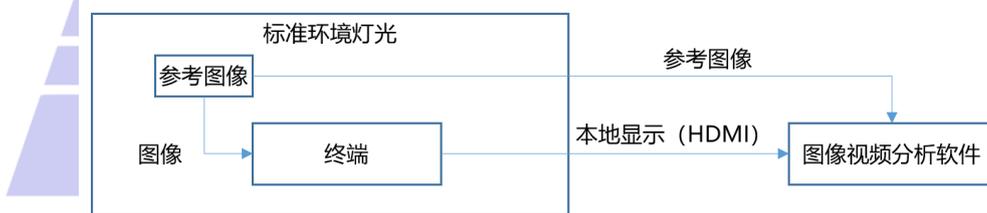


图6 会议终端视频测试实验室图卡、光源、DUT摆放示意图

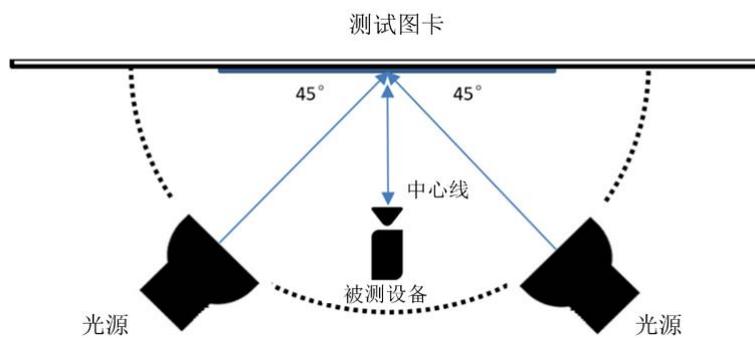


图7 会议终端视频测试设备摆放示意图

8.1.1.2.2 测试设备

测试设备包含如下内容：

- 反射式光源箱；
- 透射式光源箱；
- 照度计；
- 色温计；
- 三脚架；
- 计时器：精度10ms以内；
- 直尺：精度1mm以内；
- 图像捕捉软件：静态图像采集格式设置为PNG或BMP。

8.1.1.2.3 测试环境

摄像头相关的客观测试应在如下测试环境进行：

- 环境温度：15℃~35℃；
- 相对湿度：5% RH~85% RH；
- 大气压力：86k Pa~106k Pa；
- 暗室要求：环境照度应小于1lx；
- 光源要求：测试光源应采取遮光措施，防止直射图卡或镜头。实测光源色温标准偏差范围为 $\pm 200\text{K}$ ；
- 测试图卡表面照度要求：详见第8.1.3各小节；同时图卡表面不应存在任何被测设备或其器具的阴影。若有需要，可实时调节光源照度使图卡表面平均照度达到相应要求；
- 测试图卡照度均匀性要求：在指定光源下，测试图卡中心照度与指定照度之差应不大于10%，图卡表面任意一点的照度与图卡中心照度之差应不大于10%；
- 测试图卡背景要求：测试图卡背景（包括背板及周围墙面）采用反射率18%的中性灰（N5）涂料，测试图卡周围（包括图卡器具）应为低反射率哑光材质；
- 其它要求：若被测设备存在高亮表面、或具备工作时点亮的指示灯等干扰源，测试前需进行遮盖处理。

8.1.1.2.4 标准光源

本文件中使用的反射式照明系统的光源如下：

- D65：色温6500K，标准荧光日光灯；
- D50：色温5000K，标准荧光日光灯；
- CWF：色温4150K，暖白荧光灯；
- A：色温2856K，暖色调钨丝灯；
- Horizon：H光，色温2300K，暖黄色调钨丝灯；
- 透射式均匀光源箱：D65/D50；
- 荧光灯光源（D65/D50/CWF）在测试工作开始前需要预热10 min，保证光源亮度和色温的稳定性。

实际测试环境的光源色温与标准色温的偏差范围为 $\pm 200\text{K}$ ；光源应采取必要遮光措施，防止直射被测设备的镜头。

8.1.1.2.5 测试图卡

本文件中测试所需的图卡包括：标准斜边测试卡、24色卡、点阵图、扇形星测试图、彩色枯叶图、36阶透射式动态范围测试卡、均匀光源箱。

- 标准斜边测试卡：应满足ISO 12233对于斜边测试图样的要求，斜边对比度为4:1，图卡有效区域包含上下两条黑边；
- 24色卡：使用ColorChecker® 24 Patch Classic测试图卡；
- 点阵图：应满足ISO 17850的要求；
- 扇形星测试图：每个扇形星由多个顶点指向同一圆心的扇形区域组成，从内到外空间频率逐渐降低（推荐测试图卡TE161）；
- 彩色枯叶图：包括枯叶图纹理细节图样和灰阶块图样，枯叶区域最大对比度范围为3:1，包含2:1和4:1对比度的斜边、定位标记和16个灰阶块。创建灰阶块的线性灰度级为0~255，步长为17；
- 36阶透射式动态范围测试卡：背景色为中灰，包括36个不同光密度的低噪声密度灰度色块；
- 均匀光源箱：发出均匀的漫射光，表面照度均匀度不低于93%；

测试图卡选取原则：图卡大小应根据测试取景距离选择，在相应的取景距离下，所选的图卡应在被测设备上得到合适的图像（按照相应小节的取景视场要求）。

8.1.1.2.6 测试分辨率、帧率及码流格式

为确保视讯设备的摄像头输入质量，设备最低需支持1920×1080@30fps。

测试码流格式的选择遵循以下原则：在分辨率及帧率同时满足以上要求的前提下，优先测试YUY2/NV12格式；若不能同时满足分辨率、帧率要求，则测试MJPEG格式。

8.1.2 音频测评

8.1.2.1 采样率

测试链路按照8.1.1.1.2的描述进行组网连接。

测试步骤如下：

- a) 测试信号使用ITU-T P.501标准Annex C.2.2的中文真人语音信号（两男两女），该信号需按照ITU-T P.863的要求，使用48k Hz采样。通过Line-in方式将信号输入给待测终端；
- b) 在参考客户端/同待测终端同样的真实终端侧录制音频；
- c) 通过专业音频分析软件分析录制音频的高频截止点，以此判断待测设备的真实采样率。

8.1.2.2 S-MOS 和 N-MOS

测试环境布局如下：

- 测试环境为8.1.1.1.1所规定的测试环境1；
- 测试位置按照8.1.1.1.4 的描述进行设置；
- 测试链路按照8.1.1.1.2 的描述进行组网连接。

测试步骤如下：

- a) 测试信号使用ITU-T P.501标准Annex C.2.2的中文真人语音信号（两男两女），该信号需按照ITU-T P.863的要求，使用48k Hz采样，并按照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32s的测试语音。每个人的语音应位于8 s片段的中心位置。嘴参考点处根据ITU-T P.56计算的信号活动语音电平为-4.7 dBPa；
- b) 将人工嘴放置在距离设备中心6 m及10 m位置；

- c) 在被测设备的麦克风位置附近摆放一个测量麦克风,使用测量麦克风录制原始的在被测设备位置的噪声和语音混合信号,同时在视频会议参考客户端处录制由被测设备处理后的含噪语音信号,参考ETSI ES 103 106标准,计算该噪声场景下每句话的S-MOS-LQO及N-MOS-LQO,并平均得到该噪声场景下的S-MOS和N-MOS;

注:本文所指0 dBFS是幅度为满量程的方波的能量。对于根据ITU-T G.711规定的0 dBFS正弦波对应能量为3.14 dBm0 (A-law)和3.17 dBm0 (μ -law)的编码信号基准,0 dBFS方波对应于约6.1 dBm0。

8.1.2.3 语音能量

测试环境布局如下:

- 测试环境为8.1.1.1.1所规定的测试环境1;
- 测试位置按照8.1.1.1.4的描述进行设置;
- 测试链路按照8.1.1.1.2的描述进行组网连接。

测试步骤如下:

a) 发送方向语音能量测试:

- 1) 将人工嘴放置在距离设备中心6 m位置;
- 2) 测试信号使用ITU-T P.501标准Annex C.2.2中文真人语音信号(两男两女),该信号需按照ITU-T P.863的要求,使用48k Hz采样,并按照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。嘴参考点处根据ITU-T P.56计算的信号活动语音电平为-4.7 dBPa(正常语音能量)和-14.7 dBPa(较低语音能量);
- 3) 分别在参考客户端处录制输出的语音信号,并计算录制得到的语音的平均能量。该发送方向语音平均能量为正常和较低语音能量下该设备的测试结果。

b) 接收方向语音能量测试:

- 1) 将待测设备的播放音量调至最大;
- 2) 将人工耳放置在距离设备中心6 m位置;
- 3) 测试信号使用符合ITU-T P.501第7.4节规定的真人语音信号;
- 4) 向会议参考客户端注入该真人语音信号,根据ITU-T P.56计算的信号活动语音电平大小为-16 dBm0。被测设备输出语音时,使用人工耳录制输出的语音信号,计算100 Hz至14k Hz频带范围内录得语音信号的平均声压。

8.1.2.4 大混响场景下主观听音体验

测试环境布局如下:

- 设备测试环境为8.1.1.1.1规定的测试环境2;
- 测试位置按照8.1.1.1.4的描述进行设置;
- 测试链路按照8.1.1.1.2的描述进行组网连接。

测试包括拾音主观MOS和放音主观MOS,具体测试步骤如下。

a) 拾音主观MOS测试

- 1) 将人工嘴放置在距离待测设备中心6 m位置;
- 2) 测试信号使用ITU-T P.501标准Annex C.2.2中文真人语音信号(两男两女),该信号需按照ITU-T P.863的要求,使用48k Hz采样,并按照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。嘴参考点处根据ITU-T P.56计算的信号活动语音电平为-4.7 dBPa;

- 3) 主观听音人员在测试环境1进行听音测试;
 - 4) 主观听音人员不得少于5人, 包含声学专家和非专业人员(需通过听力筛查);
 - 5) 原始音频播放, 将测试信号直接通过测试环境1的PC+监听音箱进行播放, 播放声压级控制在65~70 dB SPL@1 m;
 - 6) 终端音频播放, 将参考终端接收到的信号通过监听音箱进行播放, 播放声压级控制在65~70 dB SPL@1 m;
 - 7) 5) 和6) 步骤重复3次, 主观听音人员根据附录B音频质量主观等级评分标准进行;
 - 8) 所有人员测试完成, 将评分结果取平均值即为最终得分。
- b) 放音主观MOS测试
- 1) 主观听音人员在测试环境2进行听音测试, 距离待测设备中心6 m位置;
 - 2) 测试信号使用符合ITU-T P.501第7.4节规定的真人语音信号;
 - 3) 向会议参考客户端注入该真人语音信号, 根据ITU-T P.56计算的信号活动语音电平大小为-16 dBm₀;
 - 4) 主观听音人员不得少于5人, 包含声学专家和非专业人员(需通过听力筛查);
 - 5) 原始音频播放, 将测试信号直接通过测试环境2的PC+监听音箱进行播放, 播放声压级控制在65~70 dB SPL@1 m;
 - 6) 终端音频播放, 将参考终端发送3) 步骤信号给待测终端, 待测终端进行信号播放, 播放声压级控制在65~70 dB SPL@1 m;
 - 7) 5) 和6) 步骤重复3次, 主观听音人员评分根据附录B音频质量主观等级评分标准进行;
 - 8) 所有人员测试完成, 将评分结果取平均值即为最终得分。

8.1.2.5 回声耦合损耗

测试环境布局如下:

- 设备测试环境为8.1.1.1.1中规定的测试环境1;
- 测试位置按照8.1.1.1.4的描述进行设置。

测试步骤如下:

- a) 测试信号使用符合ITU-T P.501第7.3.3节的压缩真人语音信号;
- b) 视频会议参考客户端注入该真人语音信号, 根据ITU-T P.56计算的信号活动语音电平大小为-10 dBm₀。同时在视频会议参考客户端录制传回来的语音信号;
- c) 传回来语音信号的能量, 并与注入信号的能量比较, 根据ITU-T G.122 Annex B, 在100 Hz-11000 Hz频带范围内, 计算出回声耦合损耗。

8.1.2.6 双讲时语音衰减

测试环境布局如下:

- 设备测试环境为8.1.1.1.1中规定的测试环境1;
- 测试位置按照8.1.1.1.4的描述进行设置。

测试步骤如下:

- a) 测试信号使用符合ITU-T P.501第7.3.5节和第7.3.7节的真人语音信号。整个测试信号由一个ITU-T P.501第7.3.7节的初始训练语音和符合第7.3.5节的双向通话语音组成。双向通话序列被分为两部分, 第一部分双向通话序列近端为单个词语(0-20 s), 第二部分为连续的双向通话序列(20-35 s)。测试需要在有双讲发生和仅近端语音两种情况下, 分别从参考客户端记录对应的输出语音。之后再对两个录音的时域信号进行分析。两种情况下, 使用的测试信号按下述方式组织。

1) 有双讲发生情况时, 参考表35。

表35 双讲测试信号

双讲测试信号	训练语音	词语(第一部分) 和整句 (第二部分) 双向通话
注入参考客户端的信号	第7.3.7节信号女声单讲部分	第7.3.5节信号男女声单讲部分
人工嘴信号	第7.3.7节信号男声单讲部分	第7.3.5节信号男女声双讲部分

2) 仅近端语音情况时, 参考表36。

表36 近端语音测试信号

近端语音测试信号	训练语音	词语(第一部分) 和整句 (第二部分) 双向通话
远端信号	第7.3.7节信号女声单讲部分	静音
人工嘴信号	第7.3.7节信号男声单讲部分	第7.3.5节信号男女声双讲部分

- 3) 向视频会议参考客户端和人工嘴注入测试信号。测试信号嘴参考点处根据ITU-T P.56计算的信号活动语音电平为-4.7 dBPa, 注入参考客户端的语音信号电平为-16 dBm0。为保证时间对齐同步, 人工嘴信号需进行延时, 延时量为该测试过程中接收方向全链路延时的时间。
- 4) 按ITU-T P.502定义的计算方法, 根据在仅近端语音和有双讲发生两种情况下的录音, 分析双讲情况下上行语音的衰减。

8.1.3 视频测评

8.1.3.1 MTF50P

按7.1.2.1节和8.1.1.2节的要求, 分别在CWF 80 lx、6500K 300 lx光源下拍摄标准斜边测试卡。调节被测设备位置, 使其水平视场内刚好填满完整图卡, 且图卡的上下黑边完整显示。对于定焦摄像头, 如果上述取景距离不满足设备的景深范围, 则可调节设备与图卡之间的距离, 但是最近距离下需保证VFOV内完整显示图卡的上下边缘。对于具有自动对焦功能的设备, 每个分辨率及光源下拍3-5张, 每张之间重新触发自动对焦, 选取各分辨率及光源下最清晰的图片。

取图卡中心方块的四条斜边, 通过对单边线扩散函数(LSF)进行标准离散傅里叶变换(DFT), 分别推导四条斜边的MTF 50P, 记录其中的最小值。计算公式如下:

$$e - \text{SFR}(k) = D(j) \left| \frac{\sum_{j=1}^N LSD'_W(j) e^{-2\pi k j / N}}{\sum_{j=1}^N LSD'_W(j)} \right| \dots \dots \dots (1)$$

式中:

$k=0,1,2, \dots, 2/N$, 若N为奇数则为 $(N+1)/2$, k为空间频率的指数;

$LSD'_W(j)$ —— 由图中所选区域构成的窗化的、平均的、中心的、超采样线扩散函数;

$D(j) = \min \left[\frac{1}{\sin(\pi k / N)}, 10 \right]$, 是从边缘扩散函数推导点扩散函数时所用离散导数的频率响应修正。

8.1.3.2 锐化

按8.1.1.2节的要求，取7.1.2.2节中拍摄的标准斜边测试卡图片，选取中心方块的四条斜边，分别按以下公式计算锐化强度。计算结果为正时则为过度锐化Oversharpening，结果为负时则为锐化不足Undersharpening，记录其中绝对值最大的数值。

$$\text{Sharpening} = 100\%[MTF(f_{eq1}) - 1] \dots\dots\dots (2)$$

式中：

$MTF(f_{eq1})$ ——当 $f_{eq1} = 0.3f_N$ 、即空间频率为0.3倍的理论极限分辨率(Nyquist frequency)时的MTF。此时空间频率较低，容易受锐化算法的影响；空间频率更高时的MTF则更取决于镜头及图像质量。

8.1.3.3 帧率

按7.1.2.3节和8.1.1.2节的要求，使用被测设备分别在80 lx、300 lx下录制一段时间t(长度不小于60 s)内的视频片段。输出并分析该视频片段，计算产生的完整图像的画面帧数(frame)，此画面帧数除以t，即为视频的平均帧率(fps)。

8.1.3.4 延时

按7.1.2.4节和8.1.1.2节的要求，在300 lx或更亮的光源下执行以下测试。被测设备有多种软硬件接口协议，选择被测设备延时最小的接口作为测试结果。

当被测设备为包含屏幕的一体式设备时，使用被测设备拍摄计时器画面，通过相机软件调节摄像头至相应的分辨率并开启视频录制。用另一台照相机或手机相机同时拍摄计时器画面与被测设备屏幕上显示的采集画面，拍照至少50次。

当被测设备为不含显示设备的一体式时设备，将被测设备连接至辅助显示屏。通过相机应用调节摄像头至相应的分辨率并开启视频录制，拍摄计时器画面。用另一台照相机或手机相机同时拍摄计时器画面与被辅助显示屏上显示的采集画面，拍照至少50次。

当被测设备为不含显示设备的外接USB摄像机时，将被测设备连接至测试电脑并调节摄像头对准测试电脑上的计时器画面，使测试电脑上同时显示计时器画面和被测设备采集的画面。通过相机软件调节摄像头至相应的分辨率并开启视频录制，用截屏工具截取测试电脑的显示画面，截屏至少50次。对于厂家单独提供的摄像头单体组件，也可按此方法测试。

查看上述步骤拍摄的照片或截屏，记录计时器的原始时刻 t_{12} 和被测设备采集画面中的显示时刻 t_{11} ，则单次延时 $d_i = t_{12} - t_{11}$ 。计算所有单次延时的平均值，得到平均延时 d 。

有条件的情况下，也可利用专业延时测试设备或软件工具进行此项测试。

8.1.3.5 色彩准度

按7.1.2.5节和8.1.1.2节的要求，分别在各光源下拍摄24色卡，图卡在画面中居中且充满2/3的VFOV。对于搭配定焦摄像头的被测设备，如果由于其景深范围限制导致画面明显模糊，可换用更大尺寸的色卡。

选取24个色块，在每个色块中部截取面积不小于50%的区域，分别测出对应色块的R、G、B值，根据sRGB色域换算到CIE L*a*b*色彩空间，计算出各自相应的L*、a*、b*测试值。根据色卡的参考文件，查询得到24个色块的L*、a*、b*标准值。采用CIEDE2000色差公式(ΔE_{00})，计算每个色块经饱和度和校正后的色度差 ΔC_{00_corr} ，记录18个彩色块的平均值、最大值以及色块1、色块2的值。主要计算公式如下：

$$\Delta E_{00} = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T \frac{\Delta C'}{k_C S_C} \frac{\Delta H'}{k_H S_H} \dots\dots\dots (3)}$$

以上总色差包含明度差，相机的曝光误差会显著影响计算结果。为了（尽可能）不受曝光影响，独立地计算色彩误差，选择色度差 ΔC_{00} 作为测试项，去掉明度差的影响，如下面公式所示：

$$\Delta C_{00} = \sqrt{\left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2} + R_T \frac{\Delta C'}{k_C S_C} \frac{\Delta H'}{k_H S_H} \quad \dots\dots\dots (4)$$

式中， k_L 、 k_C 和 k_H 通常是一致的，默认设置为1。

由于颜色的饱和度会影响色度，进而影响上述 ΔE_{00} 和 ΔC_{00} 的值。为了消除饱和度的影响，在上述计算前，需对所测得的色度值进行饱和度校正，过程如下面公式所示：

$$a_{i_corr}^* = \frac{100a_{i_meas}^*}{Sat}, b_{i_corr}^* = \frac{100b_{i_meas}^*}{Sat} \quad \dots\dots\dots (5)$$

式中：

- $a_{i_corr}^*$ 、 $b_{i_corr}^*$ ——校正后的色度值；
- $a_{i_meas}^*$ 、 $b_{i_meas}^*$ ——测试所得色度值；
- Sat ——色彩饱和度（ $\times 100$ ）。

详细计算过程见ISO/CIE 11664-6中的CIEDE2000色差公式。

8.1.3.6 色彩饱和度

按8.1.1.2节的要求，取7.1.2.6节中拍摄的色卡图片，选取第1-18个色块，在每个色块中部截取面积不小于50%的区域，分别测出对应色块的R、G、B值，根据sRGB色域换算到CIE L*a*b*色彩空间，计算出各自相应的a*、b*测试值。根据色卡的参考文件，查询得到18个色块的a*、b*标准值，代入以下公式计算18个色块的平均色彩饱和度值。

$$Sat = 100\% \frac{\sum_{i=1}^n \sqrt{a_{i-meas}^{*2} + b_{i-meas}^{*2}}}{\sum_{i=1}^n \sqrt{a_{i-ideal}^{*2} + b_{i-ideal}^{*2}}} \quad \dots\dots\dots (6)$$

式中：

- $a_{i_meas}^*$ 、 $b_{i_meas}^*$ ——测试所得色度值；
- $a_{i_ideal}^*$ 、 $b_{i_ideal}^*$ ——图卡的标准色度值；
- $1 \leq i \leq 18, n=18$ 。

8.1.3.7 白平衡

按8.1.1.2节的要求，取7.1.2.7节中拍摄的色卡图片，选取第19-24灰块，在每个灰块中部截取面积不小于50%的区域，分别测出对应色块的R、G、B值，根据sRGB色域换算到CIE L*a*b*色彩空间，计算出各自相应的L*、a*、b*测试值。根据色卡的参考文件，查询得到第19-24灰块的L*、a*、b*标准值。代入第9.1.7节中的CIEDE2000 ΔC_{00} 色差公式，计算第19-24灰块的色度差（无需进行饱和度校正），即为白平衡误差 ΔC_{00} ，记录其中的最大值。

8.1.3.8 对焦性能

对焦性能适用于定焦镜头的会议终端设备或自动对焦的会议终端设备。

按7.1.2.8节和8.1.1.2节的要求，在300 lx或更亮的光源下执行以下测试。对含多个摄像头组合的设备，需要测试每个摄像头的对焦情况，防止出现对不上焦，或者出现镜头会来回移动以尝试找到焦点，画面反复出现模糊不清楚的情形发生。

定焦镜头的会议终端设备：按以上规定的测试环境，在摄像机景深范围内有人像来回移动，检查实时拍摄的画面和视频，是否有出现虚焦或图像画面不清晰的情况发生。

自动对焦的会议终端设备：评测DUT自动对焦的能力，包括快速的对焦收敛时间和对焦的稳定性。按以上规定的测试环境，在真实场景持续5 min测试，对焦时间所占整体时间的百分比S和失焦后快速对焦的时间t。通过计算标准图卡的梯度判断是否对焦成功。

测试时，在纸片人的位置旁边，距离摄像头相同的位置放置测试图卡，使用导轨使纸片人在距离摄像头1 m~5 m范围内来回移动，当纸片人静止时，实时计算全图的梯度，最后计算梯度总和，梯度最大时对焦成功。测试图卡采用扇形星测试图卡，具体计算方法如下面公式所示。

水平梯度（Sobel 算子）的计算如公式（7）所示。

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} * I \quad \dots\dots\dots (7)$$

式中， G_x 是水平方向的梯度， I 是输入的图像。

垂直梯度（Sobel 算子）的计算如公式（8）所示。

$$G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} * I \quad \dots\dots\dots (8)$$

式中， G_y 是垂直方向的梯度， I 是输入的图像。

梯度幅值为： $|G_x| + |G_y|$ 。

8.1.3.9 纹理

按7.1.2.9节和8.1.1.2节的要求，分别在CWF 80 lx、6500K 300 lx光源下拍摄枯叶图，图卡平面需与被测设备镜头平行且在画面中居中，且图卡充满2/3的VFOV。对于搭配定焦摄像头的被测设备，如果该取景下的拍摄距离不满足其景深范围，则可调节被测设备与图卡间的距离，但是最近距离下需保证VFOV内完整显示整张图卡，最远距离不超过1m。对于具有自动对焦功能的被测设备，每个光源下拍摄彩色枯叶图3-5张，每张之间重新触发自动对焦，选取各光源下最清晰的图片。

为贴近会议设备的实际使用场景，按照CPIQ定义的30英寸4K UHDTV观测条件（显示高度37.5cm，像素密度146PPI，观测距离50cm），计算纹理锐度数值。计算过程如下：

$$\text{Texture Acutance} = \frac{A}{A_r} \quad \dots\dots\dots (9)$$

式中：

$$A = \int_0^{v_c} MTF(v)M(v)CSF(v)dv$$

$$A_r = \int_0^{\infty} CSF(v)dv$$

其中，MTF的计算公式如下：

$$MTF = \sqrt{\frac{PSD(image)-PSD(noise)}{PSD(target)}} \quad \dots\dots\dots (10)$$

式中：

PSD——功率谱密度（power spectral density）；

CSF——对比敏感度函数（contrast sensitivity function），描述人眼感知，其公式为：

$$CSF(v) = a \cdot v^c \cdot e^{-bv} \quad \dots\dots\dots (11)$$

式中:

$a=1$, $b=0.2$, $c=0.8$;

v ——空间频率, 单位是cycles/degree。

$M(v)$ 表示显示介质所限定的能力, 由观测条件决定。当在显示器上观看时, $M(v)$ 的计算公式如下:

$$M(v) = \left| \frac{\sin(\pi kv)}{\pi kv} \right| \quad \dots\dots\dots (12)$$

式中:

k 值取决于观测距离, $k = 5 \cdot distance \cdot \pi/180$, 单位是degrees。

积分区间 v_c 为截止频率, 应设定为显示器的最大分辨率, 即奈奎斯特频率, $v_c = 0.5/k$ 。

8.1.3.10 噪声

噪声由空间噪声和时域噪声构成, 两者的测试方法如下, 最后取整体噪声为最后结果。

空间噪声计算, 按7.1.2.10节和8.1.1.2节的要求, 分别在CWF 80 lx、6500K 300 lx光源下拍摄24色卡, 图卡在画面中居中且充满2/3的VFOV。对于搭配定焦摄像头的被测设备, 如果由于其景深范围限制导致画面明显模糊, 则换用更大尺寸的色卡。按以下公式计算第22灰块Y通道的SNR值, 单位为dB:

$$SNR = 20\log_{10}(S/N) \quad \dots\dots\dots (13)$$

式中:

S ——即Signal, 为色块的平均亮度 Y ;

N ——即Noise, 为色块亮度值的标准差。

时域噪声计算, 按7.1.2.10节和8.1.1.2节的要求, 分别在CWF 80 lx、6500K 300 lx光源下拍摄24色卡, 捕捉连续两帧静态图像。对于搭配定焦摄像头的被测设备, 如果由于其景深范围限制导致画面明显模糊, 则换用更大尺寸的色卡。同时读取两张图像文件 I_1 、 I_2 , 按以下公式计算第22灰块Y通道的时域噪声:

$$\sigma_{temporal} = \frac{\sigma(I_1 - I_2)}{\sqrt{2}} \quad \dots\dots\dots (14)$$

式中:

$\sigma(I_1 - I_2)$ 是两张图片间的噪声差, 空间噪声(即固定噪声)相抵消。

再将Temporal Noise结果代入空间噪声的SNR公式, 计算得到第22灰块Y通道的 $SNR_{Temp}(dB)$ 。

最后平均加权计算整体噪声, 空间噪声和时域噪声权重各占0.5。

8.1.3.11 对比度

按8.1.1.2节的要求, 取7.1.2.11节中拍摄的色卡图片, 测量第19白块及24黑块的亮度值, 分别记为 Y_{19} 和 Y_{24} , 对比度 C 计算公式如下:

$$C = \frac{Y_{19} - Y_{24}}{Y_{19} + Y_{24}} \quad \dots\dots\dots (15)$$

8.1.3.12 曝光准确性

按8.1.1.2节的要求, 取7.1.2.12节中拍摄的色卡图片, 测量第22灰块的R、G、B值, 按以下公式计算其亮度值 Y 。

$$Y = 0.3R + 0.59G + 0.11B \quad \dots\dots\dots (16)$$

8.2 会议系统测评

8.2.1 测试环境

8.2.1.1 音频

8.2.1.1.1 测试环境

除特殊规定，所有测试均应在下列正常条件下进行：

——环境温度：15 °C~35 °C；

——相对湿度：5% RH~85% RH；

——大气压力：86k Pa~106k Pa；

——测试环境1：端到端主要测试项均在带混响的测试房间进行，为满足测试距离要求：

- 1) 发送方可用于测试房间的长度应大于10 m，宽度应大于4.5 m，高度在2.2 m到3.25 m之间，最终房间的混响时间应在0.4 s到0.7 s之间；
- 2) 接收方可用于测试房间的长度应大于6 m，宽度应大于3 m，高度在2.2 m到3.25 m之间，最终房间的混响时间应在0.4 s到0.7 s之间。

两个房间具体各频段混响时间应符合表37，房间内的噪声声级应控制在30~34 dB SPL (A)。

表37 测试环境1各频段混响时间

倍频程中心频率	下限(RT60,ms)	上限 (RT60, ms)
125	400	950
250	500	850
500	350	700
1000	350	600
2000	350	600
4000	350	600
8000	400	700
房间平均值	400	700

——测试环境2：为测试大混响场景下的拾放音体验，发送和接收方向的测试房间应满足：

- 1) 发送方可用于测试房间的长度应大于6 m，宽度应大于3.5 m，高度在2.8 m到3.25 m之间，最终房间的混响时间应在0.9 s到1.1 s之间；
- 2) 接收方可用于测试房间的长度应大于6 m，宽度应大于3.5 m，高度在2.8 m到3.25 m之间，最终房间的混响时间应在0.9 s到1.1 s之间。

房间具体各频段混响时间应符合表38，房间内的噪声声级应控制在30~34 dB SPL (A)。

表38 测试环境2各频段混响时间

倍频程中心频率	下限(RT60,ms)	上限 (RT60,ms)
125	900	1430
250	900	1430
500	900	1100
1000	900	1100
2000	810	990
4000	810	990
8000	720	880
房间平均值	900	1100

8.2.1.1.2 测试链路

测试链路网络采用有线网络，应确保测试网络未引入额外损伤，网络无明显延时抖动和丢包。测试链路图如图8所示。当测试项需要模拟背景噪声的情况下，可按照ITU-T P.570标准第8.2.6节来均衡和回放噪声。

测试链路中的待测发送终端和待测接收终端均为同一型号，可为一体式或分体式会议终端。

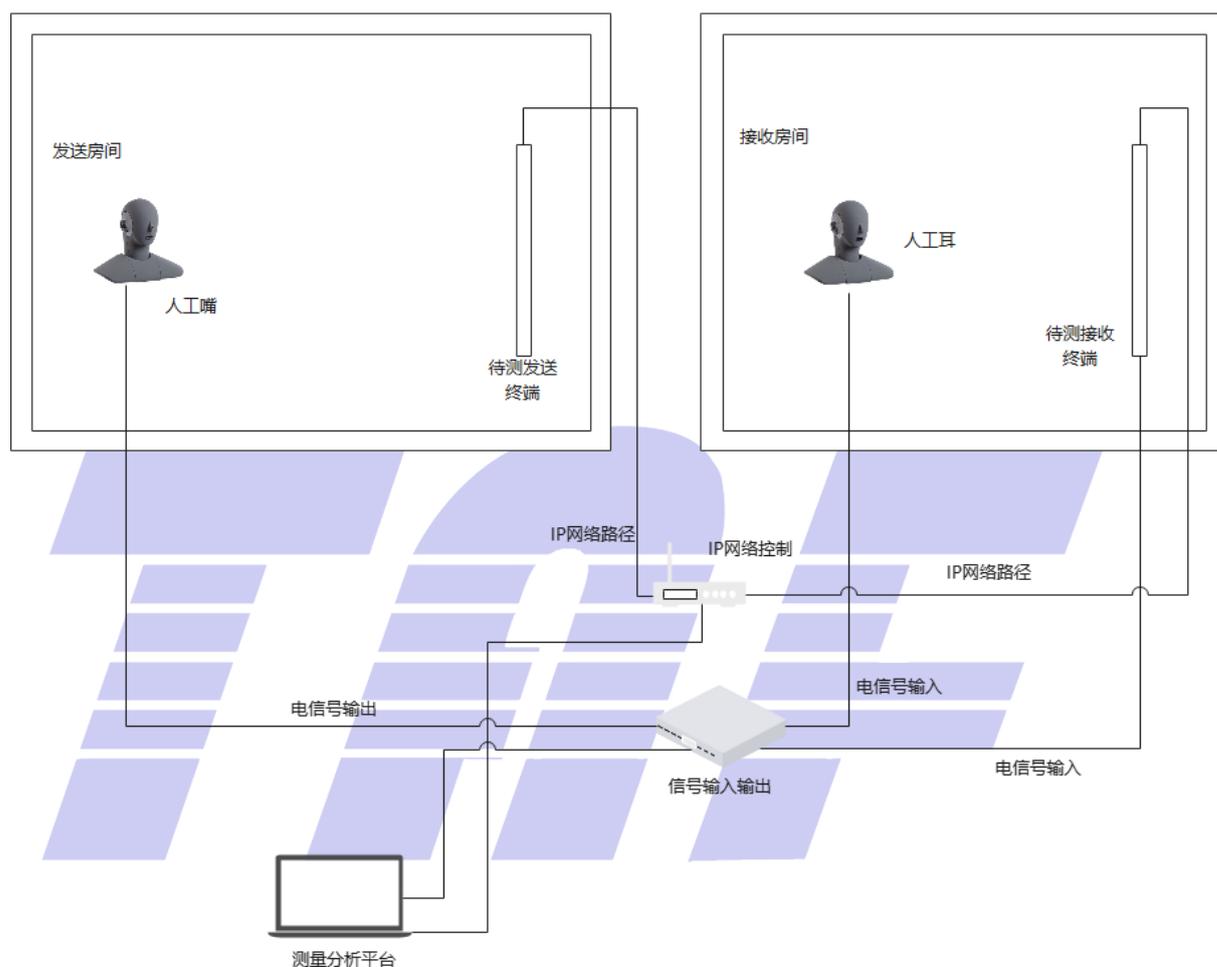


图8 会议系统音频测试组网示意图

8.2.1.1.3 测试终端和系统设置

待测终端的扬声器播放和麦克风增益除特殊说明，否则采用默认配置。

8.2.1.1.4 测试设备安放

一体式会议终端安放方式如图9和图10所示。

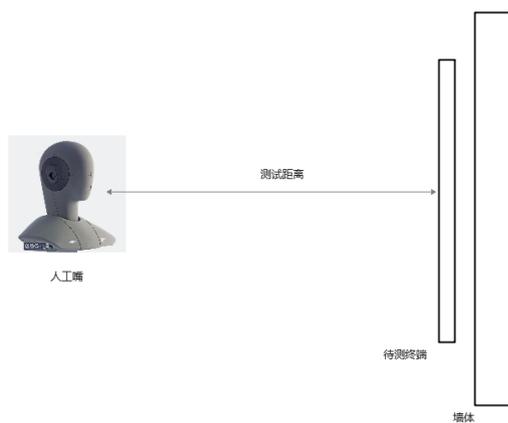


图9 会议系统音频测试一体式终端发送侧安装示意图

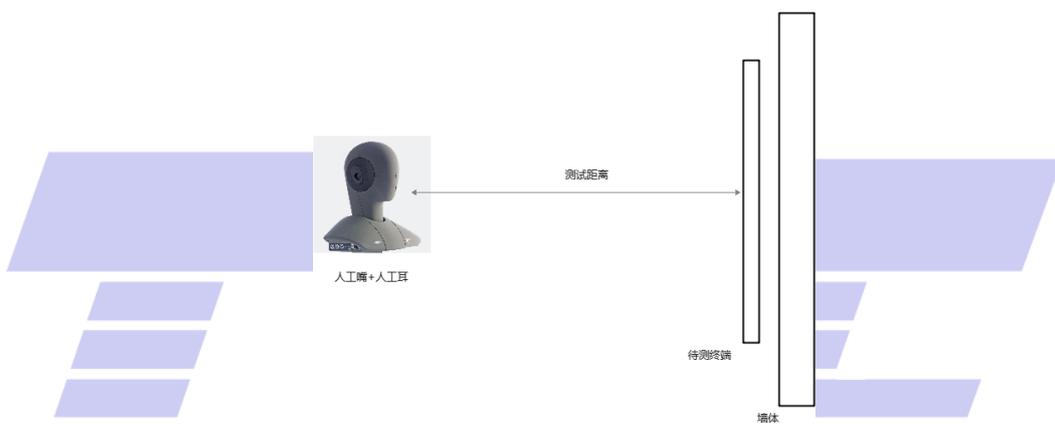


图10 会议系统音频测试一体式终端接收侧安装示意图

对于一体式会议终端，其发送侧和接收侧安放示意图如上图所示。若大屏自身具备支架，可以将设备安放在支架上进行测试，否则需要将待测终端的安装要求安装在墙体上进行测试。测试距离可根据测试项要求对人工嘴进行布置。

分体式会议终端按照方式如图11和图12所示。

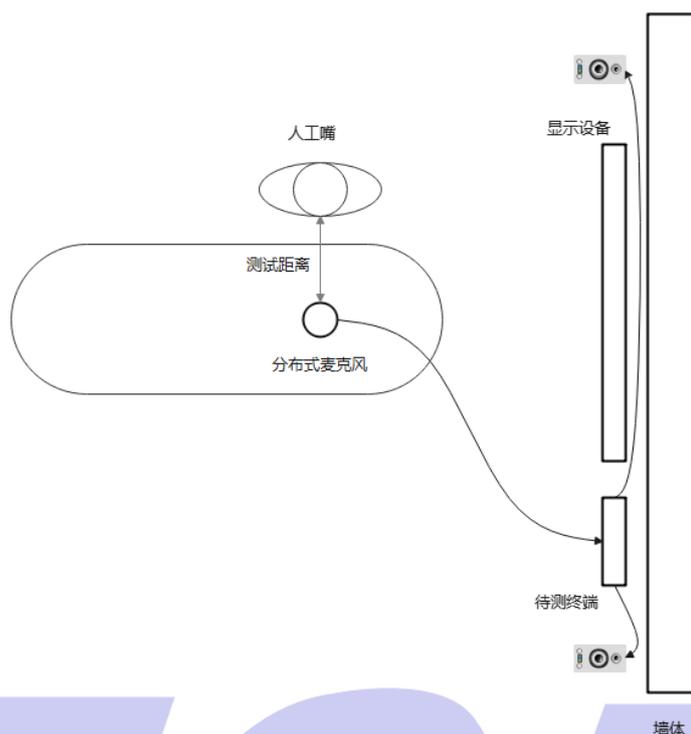


图11 会议系统音频测试分体式终端发送侧安装示意图

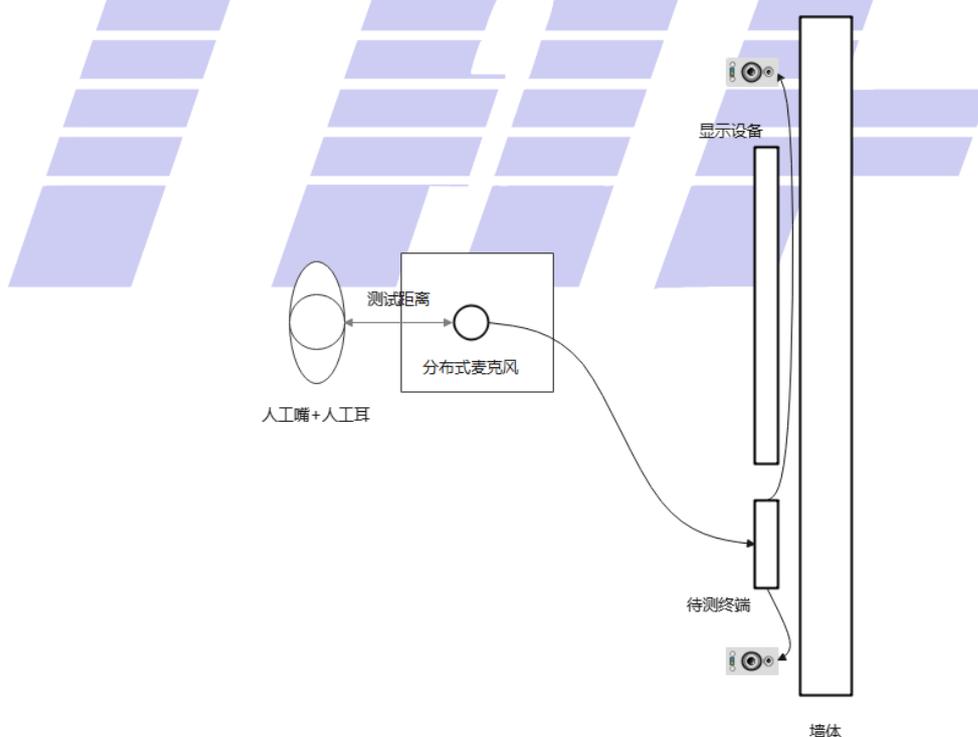


图12 会议系统音频测试分体式终端接收侧安装示意图

对于分体式会议终端，其发送侧和接收侧安放示意图如上图所示。分体式会议终端要完成端到端会议，需搭配分布式麦克风和外置音箱。外置音箱型号可按照厂家推荐进行搭配，若无特殊要求，外置音箱采用一对组合方式，放置在靠近墙体的位置。分布式麦克风的摆放位置可根据测试项进行调整。

8.2.1.2 视频

8.2.1.2.1 测试组网

测试组网参考图13所示。

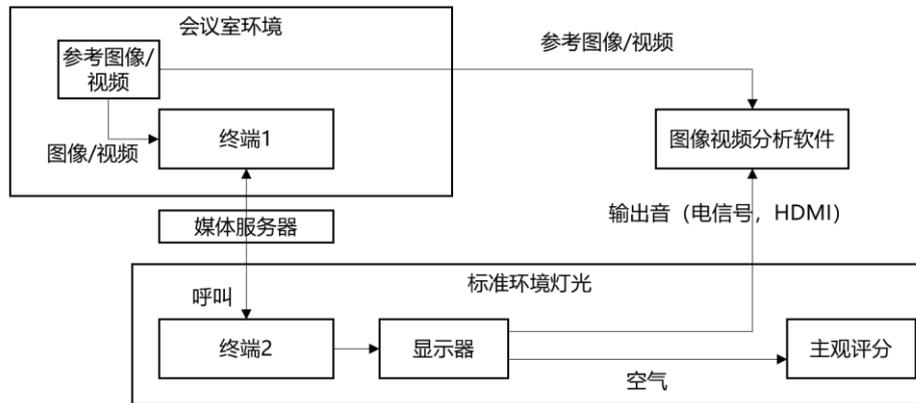


图13 会议系统视频测试组网示意图

8.2.1.2.2 测试环境

本文件中摄像头相关的客观测试应在如下测试环境进行。

- 环境温度：15℃~35℃。
- 相对湿度：5% RH~85% RH。
- 大气压力：86k Pa~106k Pa。

本文件中摄像头相关的视频主观评分测试应在如下测试环境进行。

- 标准环境灯光：色温在 5000 K-6500 K 之间。
- 照度：会议室平均照度 300 lx~400 lx。
- 主观评测人数：5 人。
- 测试距离：显示屏幕对角线距离的 2 倍。

8.2.1.2.3 测试序列

主流测试序列为 ClassE 测试序列和导播视频。测试视频序列的播放选用标准 4K 显示器，频率为 60Hz，覆盖采集设备全画面。

辅流测试为 PPT，Word 和 Excel 常用办公文档。

8.2.1.2.4 测试设备

包括如下设备：

- 网损仪：设置网络延时、丢包、抖动；
- 网络流量统计软件：统计视频网络流量；
- 测试视频序列：模拟会议室人员运动，构造码率超编场景；
- 计时器：精度10ms以内，计算帧率和延时；
- 图像捕捉软件：静态图像采集格式设置为PNG或BMP，计算图像PSNR和SSIM；
- 音画同步分析仪；
- 高帧率摄像机（建议至少240 fps）；

——三脚架。

8.2.2 音频测评

8.2.2.1 端到端延时

测试环境布局如下：

——测试环境为8.2.1.1.1所规定的测试环境1；

——测试位置：一体式设备按照8.2.1.1.4的描述进行设置，发送端人工嘴距离设备中心小于1 m，接收端人工耳距离设备中心小于1 m；分体式设备按照8.2.1.1.4的描述进行设置，发送端人工嘴距离分布式MIC中心小于1 m，接收端人工耳距离扬声器发射平面小于1m。

测试步骤如下：

- a) 测试信号使用ITU-T P.501标准Annex C.2.3的中文真人语音信号（两男两女），该信号需按照ITU-T P.863的要求，使用48k Hz采样，并按照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) 使用人工嘴播放该真人语音信号，嘴参考点处根据ITU-T P.56计算的信号活动语音电平为-4.7 dBPa；
- c) 播放语音的同时在会议对端使用人工耳录制输出的语音信号。按照ITU-T P.863标准定义的方法，计算录得信号与输入信号之间的时间延时。

8.2.2.2 混响环境端到端主观 MOS

测试环境布局如下：

——测试环境为8.2.1.1.1所规定的测试环境1；

——测试位置：一体式终端设备按照8.2.1.1.4的描述进行设置，发送端人工嘴距离设备中心6 m或10 m，接收端主观听音座椅中心距离设备中心6 m；分体式终端设备按照8.2.1.1.4的描述进行设置，发送端人工嘴距离分布式MIC中心1.5 m，接收端主观听音座椅中心距离扬声器发射平面最远不超过6 m。

测试步骤如下：

- a) 测试信号使用ITU-T P.501标准Annex C.2.2的中文真人语音信号（两男两女），该信号需按照ITU-T P.863的要求，使用48k Hz采样，并按照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。嘴参考点处根据ITU-T P.56计算的信号活动语音电平为-4.7 dBPa；
- b) 主观听音人员在测试环境1进行听音测试；
- c) 主观听音人员不得少于5人，包含声学专家和非专业人员（需通过听力筛查）；
- d) 原始音频播放，将测试信号直接通过测试环境1的PC+监听音箱进行播放，播放声压级控制在65~70 dB SPL@1 m；
- e) 终端音频播放，将接收终端接收到的信号通过待测系统进行播放，播放声压级控制在65~70 dB SPL@1 m；
- f) d) 和e) 步骤重复3次，主观听音人员MOS评分根据附录B音频质量主观等级评分标准进行；
- g) 所有人员测试完成，将评分结果取平均值即为最终得分。

8.2.2.3 语音能量

测试环境布局如下：

——测试环境为8.2.1.1.1所规定的测试环境1；

——测试位置：一体式终端设备按照8.2.1.1.4的描述进行设置，发送端人工嘴距离设备中心6 m，接收端人工耳距离设备中心6 m；分体式终端设备按照8.2.1.1.4的描述进行设置，发送端人工嘴距离分布式MIC中心1.5 m，接收端人工耳距离扬声器发射平面最远不超过6 m。

测试步骤如下：

- a) 测试信号使用ITU-T P.501标准Annex C.2.2中文真人语音信号(两男两女)，该信号需按照ITU-T P.863的要求，使用48k Hz采样，并按照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。嘴参考点处根据ITU-T P.56计算的信号活动语音电平为-4.7 dBPa；
- b) 将接收端设备的播放音量调至最大；
- c) 用人工耳录制输出的语音信号，计算100 Hz至14k Hz频带范围内录得语音信号的平均声压。

8.2.2.4 弱网适应性主观 MOS

测试环境布局如下：

——测试环境：为8.2.1.1.1所规定的测试环境1；

——测试位置：一体式终端设备按照8.2.1.1.4的描述进行设置，发送端人工嘴距离设备中心6 m，接收端主观听音座椅中心距离设备中心6 m；分体式终端设备按照8.2.1.1.4的描述进行设置，发送端人工嘴距离分布式MIC中心1.5 m，接收端主观听音座椅中心距离扬声器发射平面最远不超过6 m。

测试步骤如下：

- a) 测试信号使用ITU-T P.501标准Annex C.2.2的中文真人语音信号(两男两女)，该信号需按照ITU-T P.863的要求，使用48k Hz采样，并按照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。嘴参考点处根据ITU-T P.56计算的信号活动语音电平为-4.7 dBPa；
- b) 在中间传输网络上设置不同网络环境，如6.2.1.4描述的测试条件；
- c) 主观听音人员在测试环境1进行听音测试；
- d) 主观听音人员不得少于5人，包含声学专家和非专业人员(需通过听力筛查)；
- e) 原始音频播放，无网损条件，将接收终端接收到的信号通过待测系统进行播放，播放声压级控制在65~70 dB SPL@1 m；
- f) 终端音频播放，增加网络延时、抖动及丢包条件，播放2段以上音频后，将接收终端接收到的信号通过待测系统进行播放，播放声压级控制在65~70 dB SPL@1 m；
- g) e) 和f) 步骤重复3次，主观听音人员MOS评分根据附录B音频质量主观等级评分标准进行。
- h) 所有人员测试完成，将评分结果取平均值即为最终得分。

8.2.2.5 大混响场景下主观听音体验

测试环境布局如下。

——测试环境为8.2.1.1.1所规定的测试环境1和测试环境2。

——拾音主观体验测试：一体式终端设备按照8.2.1.1.4的描述进行设置，发送测试终端在测试环境2，接收测试终端在测试环境1。发送端人工嘴距离设备中心6 m，接收端主观听音座椅距离设备中心6 m；分体式终端设备按照8.2.1.1.4的描述进行设置，发送测试终端在测试环境2，接收测试终端在测试环境1。发送端人工嘴距离分布式MIC中心1.5 m，接收端主观听音座椅距离扬声器发射平面最远不超过6 m。

测试步骤如下：

- 1) 测试信号使用ITU-T P.501标准Annex C.2.2中文真人语音信号（两男两女），该信号需按照ITU-T P.863的要求，使用48k Hz采样，并按照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。嘴参考点处根据ITU-T P.56计算的信号活动语音电平为-4.7 dBPa；
- 2) 主观听音人员在测试环境2进行听音测试；
- 3) 主观听音人员不得少于5人，包含声学专家和非专业人员（需通过听力筛查）；
- 4) 原始音频播放，将测试信号直接通过测试环境2的PC+监听音箱进行播放，播放声压级控制在65~70 dBSPL@1 m；
- 5) 终端音频播放，将接收终端接收到的信号通过待测系统进行播放，播放声压级控制在65~70 dBSPL@1 m；
- 6) d) 和e) 步骤重复3次，主观听音人员听音体验根据附录B音频质量主观等级评分标准进行；
- 7) 所有人员测试完成，将评分结果取平均值即为最终得分。

——放音主观体验测试：一体式终端设备按照8.2.1.1.4的描述进行设置，发送测试终端在测试环境1，接收测试终端在测试环境2。发送端人工嘴距离设备中心6 m，接收端主观听音座椅距离设备中心6 m；分体式终端设备按照8.2.1.1.4的描述进行设置，发送测试终端在测试环境1，接收测试终端在测试环境2。发送端人工嘴距离分布式MIC中心1.5 m，接收端主观听音座椅距离扬声器发射平面最远不超过6 m。

测试步骤如下：

- 1) 测试信号使用ITU-T P.501标准Annex C.2.2中文真人语音信号（两男两女），该信号需按照ITU-T P.863的要求，使用48k Hz采样，并按照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。人工嘴参考点处根据ITU-T P.56计算的信号活动语音电平为-4.7 dBPa；
- 2) 主观听音人员在测试环境2进行听音测试；
- 3) 主观听音人员不得少于5人，包含声学专家和非专业人员（需通过听力筛查）；
- 4) 原始音频播放，将测试信号直接通过测试环境2的PC+监听音箱进行播放，播放声压级控制在65~70 dBSPL@1 m；
- 5) 终端音频播放，将接收终端接收到的信号通过待测系统进行播放，播放声压级控制在65~70 dBSPL@1 m；
- 6) d) 和e) 步骤重复3次，主观听音人员听音体验根据附录B音频质量主观等级评分标准进行；
- 7) 所有人员测试完成，将评分结果取平均值即为最终得分。

8.2.2.6 回声耦合损耗

测试环境布局如下：

——设备测试环境为8.2.1.1.1所规定的测试环境1；

——测试位置按照8.2.1.1.4的描述进行设置。

测试步骤如下：

- a) 测试信号使用符合ITU-T P.501第7.3.3节的压缩真人语音信号。人工嘴参考点处根据ITU-T P.56计算的信号活动语音电平为-4.7 dBPa；
- b) 在接收端侧，使用人工嘴播放测试信号；

- c) 并在接收端测试设备Line Out接口录制传回来语音信号的能量，并与注入信号的能量比较，根据ITU-T G.122 Annex B，在100Hz-11kHz频带范围内，计算出回声耦合损耗。

8.2.2.7 双讲时语音衰减

测试环境布局如下：

- 设备测试环境为8.2.1.1.1所规定的测试环境1；
- 测试位置：将待测终端按照8.2.1.1.4的描述进行设置。

测试步骤如下：

- a) 测试信号使用符合ITU-T P.501第7.3.5节和第7.3.7节的真人语音信号。整个测试信号由一个ITU-T P.501第7.3.7节的初始训练语音和符合第7.3.5节的双向通话语音组成。双向通话序列被分为两部分，第一部分双向通话序列近端为单个词语（0-20 s），第二部分为连续的双向通话序列（20-35 s）。测试需要在有双讲发生和仅近端语音两种情况下，分别从接收端测试终端的Line Out记录对应的输出语音。之后再对两个录音的时域信号进行分析。两种情况下，使用的测试信号按下述方式组织。
- 1) 有双讲发生情况时，见表39。

表39 双讲测试信号

双讲测试信号	训练语音	词语（第一部分）和整句（第二部分）双向通话
注入参考客户端的信号	第7.3.7节信号女声单讲部分	第7.3.5节信号男女声单讲部分
人工嘴信号	第7.3.7节信号男声单讲部分	第7.3.5节信号男女声双讲部分

- 2) 仅近端语音情况时，见表40。

表40 近端语音测试信号

近端语音测试信号	训练语音	词语（第一部分）和整句（第二部分）双向通话
远端信号	第7.3.7节信号女声单讲部分	静音
人工嘴信号	第7.3.7节信号男声单讲部分	第7.3.5节信号男女声双讲部分

- b) 向接收端和发送端人工嘴注入测试信号。测试信号嘴参考点处根据ITU-T P.56计算的信号活动语音电平为-4.7 dBPa。为保证时间对齐同步，人工嘴信号需进行延时，延时量为该测试过程中接收方向全链路延时的时间。
- c) 根据ITU-T P.502定义的计算方法，根据在仅近端语音和有双讲发生两种情况下的录音，分析双讲情况下上行语音的衰减。

8.2.3 视频测评

8.2.3.1 分辨率

按照 8.2.1.2.1 视频测试组网，将解码后图像输出到标准显示器，通过显示器查看支持的最大图像分辨率大小。

8.2.3.2 图像还原度

按照 8.2.1.2.1 视频测试组网，会议系统有 4K、1080P 和 720P 的分辨率，系统所支持的分辨率均需要测试。

发送端编码前的图像和接收端解码后的图像进行 PSNR 对比，图像格式均采用 YUV420。在无叠加网损的环境下，取 3 张图进行后续对比。

网络环境设置如下：

- 分辨率@帧率 4K@30 fps，带宽 4M bps；
- 分辨率@帧率 1080P@30 fps，带宽 1.5M bps；
- 分辨率@帧率 720P@30 fps，带宽 768K bps。

图像还原度 PSNR 计算方法如下：

——首先，计算均方误差（Mean Squared Error, MSE）：

设原始图像为 I ，处理后的图像为 K ，图像的尺寸为 $M \times N$ （ M 为图像的高度， N 为图像的宽度），则 MSE 的计算公式为：

$$MSE = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [I(i,j) - K(i,j)]^2 \dots \dots \dots (17)$$

式中：

$I(i,j)$ ——原始图像在坐标 (i,j) 处的像素值；

$K(i,j)$ ——处理后图像在相同坐标处的像素值。

——然后，按照公式（18）计算 PSNR 值。

$$PSNR = 10 \times \log_{10}(MAX_I^2 / MSE) \dots \dots \dots (18)$$

式中：

MAX_I ——图像像素值的最大可能取值。对于 8 位图像，像素值的范围是 0 到 255，因此 $MAX_I = 255$ 。

8.2.3.3 共享演示（辅流）

按照 8.2.1.2.1 视频测试组网，分别完成 4Mbps、1.5Mbps 和 768Kbps 会议呼叫，呼叫时长为 30 min，网络环境设置为 50 ms~150 ms 随机延时 + 30 ms~50 ms 随机抖动 + 0~30% 随机丢包，通过网损仪限制呼叫带宽分别为 4Mbps、1.5Mbps 和 768Kbps，然后测试如下指标。

a) 共享演示的 PSNR：

视频会议共享桌面一个标准 word 文档，excel 文档和 ppt 文档，每个文档测试 10 min，每 5 s 翻页一次，翻页后经过 5 s 计算收敛后的 PSNR 值，总共测试 30min。远端接收到辅流视频后，记录并保存，使用视频分析软件分析 PSNR 值。发送端采用 YUV444 格式图像，接收端采用 YUV420 格式图像，共享演示的 PSNR 计算公式按照 8.2.3.2 计算。

b) 共享演示的切换响应时间

共享演示翻页场景，例如 PPT 翻页，接收端接收翻页后图像收敛后（图像质量稳定）的时间与发送端翻页的时间差 T ，如公式（19）所示。

$$T = T_{\text{ReceiveConvergence}} - T_{\text{send A}} \dots \dots \dots (19)$$

计算翻页切换 10 次的平均时间。

8.2.3.4 音画同步

按照 8.2.1.2.1 视频测试组网，分别完成 4K30@4Mbps、1080P30@1.5Mbps 和 720P30@768Kbps 会议呼叫，呼叫时长为 10 min，网络环境设置为 50 ms 随机延时 + 30 ms 随机抖动 + 10% 随机丢包，然后统计不同呼叫带宽下的音画同步情况。

对于不带屏幕显示的分体式终端，将支持音画同步测试仪放置于测试环境中，音画同步测试仪采集

源端音视频码流及被测系统音视频码流，并分析出音画不同步数值。

对于其他类型终端，包括一体式终端，将音画同步测试仪放置于测试环境中，音画同步测试仪发出同步光信号和声信号，被测系统采集音画同步测试仪发出的光源信号和声音信号。音画同步测试仪采集接收端相应的视频和音频信号，并分析采集的光信号和声信号的时间差即音画同步。

8.2.3.5 码率超编

按照 8.2.1.2.1 视频测试组网，分别完成 4K30@4Mbps、1080P301.5Mbps 和 720P30768Kbps 会议呼叫，呼叫时长为 30 min，网络环境设置为 50 ms 随机延时 + 30 ms 随机抖动 + 10%随机丢包，然后统计不同呼叫带宽下每 s 的最大码率。

采用JVET-J1010文档中定义的Class E测试序列，同时加入导播切换场景，使用双线性插值方法，将测试序列分别缩放到720P、1080P和4K作为输入视频。

表41 测试序列

测试序列名称	分辨率	帧数	帧率	位深
FourPeople	720P	600	60	8
Johnny	720P	600	60	8
KristenAndSara	720P	600	60	8
导播	720P	600	60	8

循环播放以上视频 10 min，播放 10 min 后统计码率超编的情况，包括秒级超编。超编比例的公式为：

$$\text{OverEncodingRate} = \text{RATE}_{\text{real}} / \text{RATE}_{\text{expect}} \dots\dots\dots (20)$$

式中：

RATE_{real} ——每秒钟实际的码率；

RATE_{expect} ——期望的码率。

8.2.3.6 弱网适应性主观 MOS

按照 8.2.1.2.1 视频测试组网完成抗丢包测试，采用 JVET-J1010 文档中定义的 Class E 测试序列，使用双线性插值方法，将测试序列分别缩放到 720P、1080P 和 4K 作为输入视频。

被测 DUT 建立三次视频交互，参数配置如表 42 所示。

表42 参数配置

分辨率	帧率 (fps)	编码格式	码率
1280×720	30	HEVC main profile	768Kbps
1920×1080	30	HEVC main profile	1.5Mbps
3840×2160	30	HEVC main profile	4Mbps

测试步骤如下：

- a) 通过丢包仪，分别设置 DUT1 的上行延时、抖动、丢包率和最大连续丢包个数：
 - 50 ms 延时 + 30 ms 抖动 + 10%丢包率（包含 8 个连续丢包）；
 - 100 ms 延时 + 30 ms 抖动 + 20%丢包率（包含 10 个连续丢包）；
 - 150 ms 延时 + 50 ms 抖动 + 30%丢包率（包含 12 个连续丢包）。

其中最大连续丢包每秒钟设置一次。

- b) 按以上配置和丢包分别进行测试，并按 8.2.1.2.2 视频主观评分环境要求和附录 B 组织评分，给出 MOS 值。

8.2.3.7 端到端延时

按照图 14 视频测试组网，分别完成 4K30@4Mbps、1080P30@1.5Mbps 和 720P30@768Kbps 会议呼叫，通话时长至少为 10 min，然后统计不同呼叫带宽下的延时。

注：对于一体化终端，端到端延时包括采集和显示的延时；对于分体式终端，端到端延时不包括外设的采集或显示的延时。

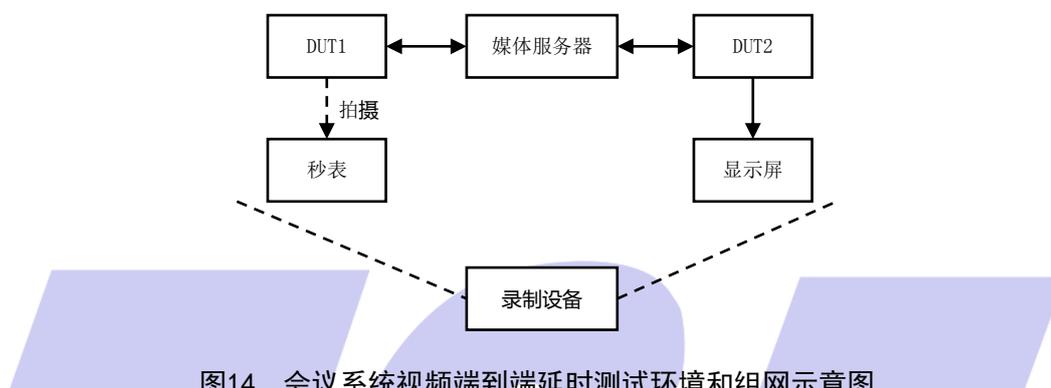


图14 会议系统视频端到端延时测试环境和组网示意图

测试步骤如下：

- 播放秒表计时器，最小单位毫秒；
- DUT1 拍摄秒表；
- 观察 DUT2 输出图像；
- 使用高帧率摄像机，同时拍摄秒表和 DUT2 的显示屏，连续录制 5 min；
- 计算录像文件所有帧延时的平均值，计算公式如下：

$$\text{端到端延时} = \sum(\text{DUT2 显示屏的秒表时间} - \text{秒表的时间}) / \text{录制 5 min 内的总帧数} \dots\dots\dots (21)$$

9 分级评价方法

9.1 指标权重

根据会议系统及终端评价指标框架特点，分别给出会议系统和终端评价指标权重，见附录A。

9.2 得分判定

应对照评价指标按照符合情况对各项评价条款进行得分判定，符合要求的条款项得分，不符合要求的条款项不得分。每个指标项分为优秀、良好、一般和不合格，其中优秀得分为100，良好得分为80，一般得分为60，小于60分为不及格。

9.3 计算公式

二级指标得分C为该域下所有三级指标得分D 的加权求和，得分按式 (22) 计算：

$$C = \sum D \times \gamma \dots\dots\dots (22)$$

式中：

C ——二级指标得分；

γ ——三级指标权重。

一级指标的得分 B 为该域下二级指标得分的加权求和，得分按式（23）计算：

$$B = \sum C \times \beta \quad \dots\dots\dots (23)$$

式中：

B ——一级指标得分；

β ——二级指标权重。

总得分 A 为各个一级指标得分的加权求和，整体得分按式（24）计算：

$$A = \sum B \times \alpha \quad \dots\dots\dots (24)$$

式中：

A ——总得分；

α ——一级指标权重。

9.4 等级判定

会议系统和会议终端的评定等级与总得分的关系如表43所示。

表43 等级判定方法

总得分	评定等级
$90 \leq A \leq 100$	优秀
$80 \leq A < 90$	良好
$60 \leq A < 80$	一般
$A < 60$	不合格

附录 A

(规范性)

会议系统及终端音视频质量评价指标权重

A.1 指标权重

表A.1和表A.2分别给出了会议终端和会议系统的指标权重。

表A.1 会议终端音视频质量评价指标权重

一级指标		二级指标		三级指标	
音频	50%	采样率	5%	-	-
		S-MOS和N-MOS	25%	S-MOS (6 m)	25%
				N-MOS (6 m)	25%
				S-MOS (10 m)	25%
				N-MOS (10 m)	25%
		语音能量	10%	发送方向语音能量 (正常音量)	33%
				发送方向语音能量 (较低音量)	33%
				接收方向语音能量	34%
		大混响场景下主观听音体验	20%	拾音主观MOS	50%
				放音主观MOS	50%
		回声耦合损耗	15%	标称音量	50%
				最大音量	50%
双讲时语音衰减	25%	标称音量	50%		
		最大音量	50%		
视频	50%	MTF50P	10%	-	-
		锐化	5%	-	-
		帧率	5%	-	-
		延时	10%	-	-
		色彩准确度	10%	-	-
		色彩饱和度	10%	-	-
		白平衡	10%	-	-
		对焦性能	10%	-	-
		纹理	5%	-	-
		噪声	5%	-	-
		对比度	10%	-	-
		曝光准确性	10%	-	-

表A.2 会议系统音视频质量评价指标权重

一级指标		二级指标		三级指标	
音频	50%	端到端延时	10%	-	-
		混响环境端到端主观MOS	25%	一体式终端（拾音距离6 m）	33%
				一体式终端（拾音距离10 m）	33%
				分体式终端（拾音距离1.5 m）	34%
		语音能量	5%	正常音量	50%
				较低音量	50%
		弱网适应性主观MOS	15%	测试环境1	33%
				测试环境2	33%
				测试环境3	34%
		大混响场景下主观听音体验	15%	拾音主观MOS	50%
放音主观MOS	50%				
回声耦合损耗	10%	标称音量	50%		
		最大音量	50%		
双讲时语音衰减	20%	标称音量	50%		
		最大音量	50%		
视频	50%	分辨率	10%	-	-
		图像还原度	15%	-	-
		共享演示（辅流）	15%	清晰度	50%
				切换响应时间	50%
		音画同步	15%	-	-
		码率超编	15%	-	-
		弱网适应性主观MOS	15%	测试环境1	33%
				测试环境2	33%
测试环境3	34%				
端到端延时	15%	-	-		

附录 B
(规范性)
音视频主观等级评分标准

B.1 评分方法

音频质量主观等级评分标准参考YD/T 2309—2011中规定的方法，具体如表B.1所示。

表B.1 音频质量主观等级评分标准

等级评分	听音质量
5	感知不到失真
4	能感知到失真但不恼人
3	能感知到失真且稍微恼人
2	感知到失真且恼人
1	感知到失真且非常恼人

主观评测中的声学专家，要求人员至少从事声学研究、音频设计、音质评价等相关工作至少2年，且熟悉主观声学评价方法（如 ITU-T P.800、ISO 3382 等标准）。

主观评测中的非专业人员，为确保人员听觉系统正常，避免影响测试结果，要求人员至少无耳疾、耳道清洁、无噪声过度暴露史、未使用耳疾相关药物且无家族听力损失史。

视频主观评价过程中关注视频会议的视频体验效果，主要是评价清晰度和卡顿情况，视频质量主观等级评分标准具体如表B.2所示。

表B.2 视频质量主观等级评分标准

等级评分	视频质量
5	很好，无明显可见损伤，卡顿不明显
4	稍差，损伤轻微可见，卡顿不明显，不影响观看体验
3	还可以，损伤明显，卡顿稍微明显，可能引起注意但尚可接受
2	勉强，损伤严重干扰观看，严重卡顿，显著降低体验
1	极差，损伤极其严重，极其卡顿，内容几乎无法理解

电信终端产业协会团体标准

会议系统及会议终端音视频质量及分级评价技术规范

T/TAF 307—2025

*

版权所有 侵权必究

电信终端产业协会发布

地址：北京市西城区新街口外大街 28 号

电话：010-82052809

电子版发行网址：www.taf.org.cn