

世界超高清视频产业联盟标准

T/UWA

基于电信网的智能家庭（AItoH）业务系统通用功能与 性能要求及测试方法

General Functions and Performance Requirements and Test Methods for AItoH (AI
To Home) Service based on Telecommunication Networks

（V1.0）

（初审稿）

2026-xx-xx 发布

2026-xx-xx 实施

世界超高清视频产业联盟

目 录

前 言	I
1. 范围	1
2. 规范性引用文件	1
3. 术语和定义	1
4. 缩略语	1
5. 概述	2
5.1. 业务流程	2
5.2. 应用场景	2
6. 通用功能要求	3
6.1. 多模态感知	3
6.2. 智能理解与决策	3
6.3. 交互反馈与媒体呈现	3
7. 通用性能要求	3
7.1. 网络适应性	3
7.2. 感知覆盖范围	3
7.3. AI 决策有效性	4
7.4. 交互响应质量	4
8. 测试方法	5
8.1. 通用测试条件	5
8.2. 通用功能要求测试方法	5
8.3. 通用性能要求测试方法	7
附 录 A	11
参考文献	15

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。

本文件起草单位：XXXXXX

本文件主要起草人：XXXXX

基于电信网的智能家庭（AItoH）业务系统通用功能与性能要求 及测试方法

1.范围

本文件规定了基于电信网的智能家庭（AItoH）业务系统的通用功能和性能要求及测试方法。
本文件适用于基于电信网部署的智能家庭业务系统的现网部署、验收和业务评测活动。

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。

本文件没有规范性引用文件。

3.术语和定义

下列术语和定义适用于本文件。

3.1 智能家庭业务系统 Artificial Intelligence to home service system

面向家庭用户提供娱乐、教育、康养等智能业务的技术系统。

3.2 真阳性 True Positive

实际为正类且被系统正确决策为正类的样本数。

3.3 真阴性 True Negative

实际为负类且被系统正确决策为负类的样本数。

3.4 假阳性 False Positive

实际为负类但被系统错误决策为正类的样本数。

3.5 假阴性 False Negative

实际为正类但被系统错误决策为负类的样本数。

4.缩略语

下列缩略语适用于本文件。

3D 三维（Three-Dimensional）

AI 人工智能（Artificial Intelligence）

AItoH 智能家庭（Artificial Intelligence To Home）

CPU 中央处理器（Central Processing Unit）

FN 假阴性（False Negative）

FP 假阳性（False Positive）

FPS 帧每秒（Frames Per Second）

NPC 非玩家角色（Non-Player Character）

OCR 光学字符识别（Optical Character Recognition）

RGB 红绿蓝色彩模型（Red Green Blue）

TN 真阴性（True Negative）

TP 真阳性（True Positive）

USB 通用串行总线（Universal Serial Bus）

Wi-Fi 无线局域网（Wireless Fidelity）

5.概述

5.1.业务流程

智能家庭（AI to Home, AItoH）业务系统的架构由终端设备（感知输入设备、家庭智能网关、反馈输出设备）、传输网络、业务平台和内容等环节组成，如图1所示。

智能家庭业务系统与用户交互间的核心流程包括多模态感知、智能理解与决策、交互反馈与媒体呈现，具体为：智能家庭业务系统通过感知输入设备，对用户行为、操作指令及环境状态等进行采集，并在家庭智能网关或云平台对所采集的感知数据进行智能理解与决策，形成交互反馈结果和媒体内容，最终在反馈输出设备上输出和呈现。本文件规定的智能家庭业务系统通用功能与性能要求及测试方法在用户与系统交互环节定义与完成。

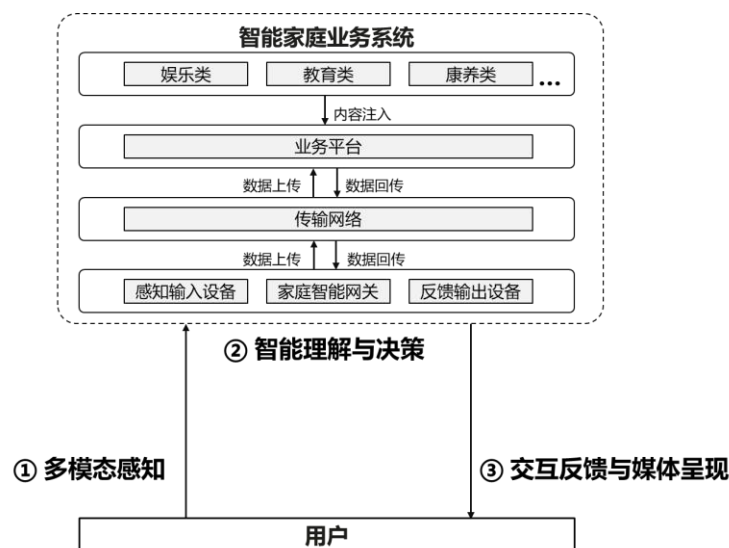


图 1 智能家庭业务系统架构及业务流程

5.2.应用场景

智能家庭业务系统承载了多样化的应用内容，通常包含娱乐、康养、教育等场景。当前典型应用场景包括 AI 互动健身、云游戏、3D 观影、康养看护跌倒检测、AI 伴学、AI 生活智能体。

a) AI 互动健身：通过对用户姿态、动作或操作的智能感知、分析和反馈，与用户进行实时、双向

交互的智能化健身业务。

- b) 云游戏：将游戏应用部署在云端服务器进行计算和渲染，用户通过终端实时操作，以视频流形式接收游戏画面的交互式娱乐业务，AI 常被应用于游戏内容生成、画质增强等。
- c) 3D 观影：通过 3D 同步信号实现 3D 眼镜左右眼画面交替显示，提供沉浸式观影业务，AI 常被用于 2D 转 3D 等。
- d) 康养看护跌倒检测：通过对用户行为的智能感知和分析，实时监测老人状态与位置，在发生跌倒或滞留异常时自动触发告警的居家安全看护业务。
- e) 教育 AI 伴学：通过语音交互、图像识别等 AI 能力，为用户提供实时问答、作业批改、学习辅导等教育业务。
- f) AI 生活智能体：通过环境感知、自主决策与主动服务能力，能够理解用户需求、协调应用和设备，为用户提供个性化、主动式的智慧生活服务。

以上典型应用场景应满足附录 A 中的要求。

6. 通用功能要求

6.1. 多模态感知

系统应支持一种或多种感知采集方式，包括但不限于：视觉感知（摄像头、深度相机等）、听觉感知（麦克风阵列等）、力学/接触感知（按键、触摸屏、手柄、鼠标、可穿戴设备等）、非接触式电磁感知（3D 光感知、毫米波雷达、红外传感器等）、环境物理感知（温湿度传感器、光照传感器等）。上述感知方式所对应的模块或设备，可通过内置、有线或无线（Wi-Fi、Zigbee、蓝牙、星闪等）的方式连接至家庭智能网关设备，以采集用户行为、操作指令及环境状态等一种或多种模态数据。当系统支持跨设备联动时，应支持通过家庭智能网关设备实现设备的互联和数据的互通。当系统涉及采集人脸、声纹、室内影像等敏感信息时，应取得用户同意或采取隐私保护措施。

6.2. 智能理解与决策

系统应对感知采集的多模态原始数据进行智能分析（如可通过 AI 推理），识别并提取用户状态、行为意图与环境信息，进而基于分析结果进行综合逻辑决策，输出业务决策结果。决策输出包括但不限于：识别结果、评价结果及生成的内容信息等。

6.3. 交互反馈与媒体呈现

系统应将业务决策结果转换为状态、信息等交互反馈，并通过一种或多种方式向用户呈现，包括但不限于视觉方式（如智能电视、家庭中控屏、手机等）、听觉方式（如智能音箱、耳机等）、触觉方式（如可穿戴设备、游戏手柄等）。

7. 通用性能要求

7.1. 网络适应性

在限定的网络条件下，系统应保持 6.1、6.2、6.3 所规定的功能正常运行，并满足 7.2、7.3、7.4 所规定的性能要求。

7.2. 感知覆盖范围

系统通过感知设备采集用户行为、操作指令及环境状态的有效空间范围，可通过感知距离和覆盖角度进行表征；对于面状区域监测场景，可同时以覆盖面积表征。系统应在其技术声明中明确感知覆盖范围对应的环境条件（如光照、噪声、安装位置、无遮挡等）。

感知距离、覆盖角度和覆盖面积的标定方法如下：

- a) 感知距离：感知设备能够稳定采集有效感知数据或正确识别目标的空间距离范围，以家庭智能网关或感知输入设备主轴方向上的最近和最远有效距离标定，单位为米（m）；
- b) 覆盖角度：感知设备能够稳定采集有效感知数据或正确识别目标的空间角度范围，应以家庭智能网关或感知输入设备主轴为基准，分别标定水平方向和垂直方向的有效覆盖角度，单位为度（°）；
- c) 覆盖面积：对于涉及面状区域监测的场景，应标定有效覆盖面积，单位为平方米（m²）。

7.3.AI 决策有效性

系统通过 AI 算法或模型对感知数据进行处理并输出结果的能力，可依据任务类型选用相应的评估指标进行表征。

系统对感知理解类任务（涉及包括但不限于姿态识别、行为检测、语音识别、目标检测等）的 AI 决策有效性，应以决策准确性作为主要评估指标。决策准确性的评估可采用真阳性（TP）、真阴性（TN）、假阳性（FP）和假阴性（FN）作为基础统计量，并依据应用场景从准确率、召回率、漏报率、误报率等指标中选用，各指标按公式（1）～公式（4）计算。

- a) 准确率：正确分类的样本数占样本总数的比例，反映系统的整体决策正确水平。

$$\text{准确率} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (1)$$

- b) 召回率：实际为正类的样本中被系统正确预测为正类的比例，反映系统对正类样本的检出能力。

$$\text{召回率} = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

- c) 漏报率：实际为正类样本中被系统错误预测为负类的比例，反映系统将目标事件漏判为正常事件的概率。

$$\text{漏报率} = \frac{FN}{TP + FN} \times 100\% \quad (3)$$

- d) 误报率：实际负类样本中被系统错误预测为正类的比例，反映系统将正常事件误判为目标事件的概率。

$$\text{误报率} = \frac{FP}{FP + TN} \times 100\% \quad (4)$$

系统对反馈优化类任务（包括但不限于 AI 超分辨率重构、AI 插帧、AI 2D 转 3D 等）的 AI 决策有效性，应以交互响应质量作为主要评估指标，交互响应质量的评估结果应满足 7.4 规定的性能要求。

7.4.交互响应质量

系统对用户输入或业务触发事件进行响应并完成相应交互呈现的时效性和内容质量，通过端到端响应时延和音视频呈现质量指标进行表征。

端到端响应时延是指从用户完成操作输入或业务触发事件发生时刻起，至系统向用户呈现相应反馈的时间间隔，单位为毫秒（ms）。

音视频呈现质量应结合相应应用场景输出的内容进行评估。当输出为视频内容时，应至少从画面分辨率和帧率进行评估；当系统输出为音视频内容时，还应从音画同步性进行主观体验质量要求，音画应无明显不同步；当涉及 3D 内容时，还应从入屏感、立体感、观看舒适度（不引起眩晕感）等维度进行主观体验质量评估。

8.测试方法

8.1.通用测试条件

8.1.1. 测试组网

智能家居业务系统测试组网由业务平台、网络、家庭智能网关、感知输入设备以及反馈输出设备组成，如图 2 所示。

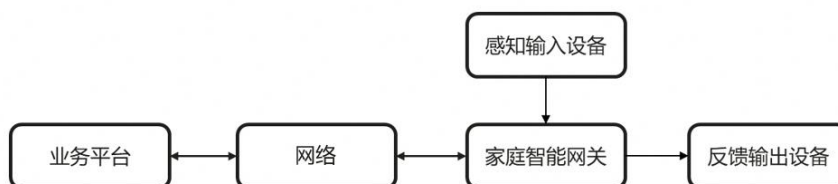


图 2 智能家居应用系统业务测试组网

8.1.2. 测试准备

在执行 8.2、8.3 规定的各项功能及性能测试前，应按以下要求完成测试准备：

- 应按被测系统技术声明或用户文档中的配置要求，将被测系统置于正常工作状态，确认各组网环节连接正常；
- 应确认所需的感知输入设备（如摄像头、麦克风等）与反馈输出设备（如显示器、音箱等）已正确连接至系统并处于待运行状态；
- 应根据被测系统声明的业务能力，准备相应的测试用例数据及预期结果判定标准。

8.2.通用功能要求测试方法

8.2.1. 多模态感知功能测试方法

8.2.1.1. 测试目的

验证系统能否支持一种或多种感知设备接入，并能正常采集环境信息、人员状态及操作等感知数据。

8.2.1.2. 测试步骤

在按 8.1 的规定完成测试组网搭建和测试准备后，应按以下步骤执行测试：

- 按照被测系统所声明支持的连接方式，将各类感知输入设备接入系统；
- 启动各感知设备的数据采集功能，并触发相应的感知事件（如移动、语音指令、环境变化等）；
- 检查系统能否正确识别已连接的设备类型及其工作状态；
- 检查系统能否实时获取并展示各设备采集的感知数据（如视频流、音频流、深度数据等），观察时间不应少于 30s。

8.2.1.3. 预期结果

测试结果应符合以下要求：

- 各类感知设备均应与系统成功建立连接；
- 系统应正确识别所有已连接设备的类型和工作状态；
- 系统应能实时显示各设备采集的感知数据，数据内容应与触发事件一致，且在观察期内不应出

现数据中断。

8.2.2. 智能理解与决策功能测试方法

8.2.2.1. 测试目的

验证系统智能理解与决策功能是否符合第 6.2 章的规定，包括对多模态感知数据的智能分析、用户状态与行为意图的识别提取、综合逻辑决策及业务决策结论的输出等能力。

8.2.2.2. 测试步骤

在按 8.1 的规定完成测试组网搭建和测试准备后，应按以下步骤执行测试：

- a) 根据被测系统所声明的智能理解与决策能力类型（如姿态识别、语音理解、图像识别、多模态融合决策等），输入相应的测试数据或触发指令；
- b) 观察并记录系统完成智能分析及综合决策后输出的结论（如识别结果、行为时间、评价结果、生成内容等）；
- c) 将系统输出的决策结论与预期决策结果进行比对，检查其一致性。

8.2.2.3. 预期结果

测试结果应符合以下要求：

- a) 系统应能完成多模态数据的智能分析，并输出相应的业务决策结论；
- b) 系统输出的决策结论应与预期决策结果一致；
- c) 测试过程中系统不应出现无法响应、决策中断或输出明显异常等情况。

8.2.3. 交互反馈与媒体呈现功能测试方法

8.2.3.1. 测试目的

验证系统交互反馈与媒体呈现功能是否符合第 6.3 章的规定，包括将业务决策结论转换为交互反馈内容，并通过视觉、听觉、触觉等一种或多种方式向用户正确呈现的能力。

8.2.3.2. 测试步骤

在按 8.1.2 的规定完成测试组网搭建和测试准备后，应按以下步骤执行测试：

- a) 根据被测系统所声明的交互方式（视觉、听觉、触觉中的一种或多种），触发相应的感知输入指令或业务场景；
- b) 观察并记录系统通过各反馈输出设备呈现的交互反馈内容：
 - 1) 视觉：观察显示终端呈现的画面内容（如实时动作反馈、训练数据可视化、游戏画面、立体视听内容、告警信息、学习内容展示等），验证画面内容是否与预期反馈一致；
 - 2) 听觉：听取音频输出终端呈现的语音反馈或提示音，验证语音内容是否与预期反馈一致，并判断其自然度和可理解性；
 - 3) 触觉：验证游戏手柄、可穿戴设备等设备的振动反馈模式是否与预期事件类型匹配。

8.2.3.3. 预期结果

测试结果应符合以下要求：

- a) 系统应能通过所声明的交互方式（视觉、听觉、触觉中的一种或多种）正确呈现交互反馈内容，其中：
 - 1) 视觉方式：呈现的画面内容应与预期反馈一致，不应出现卡顿、花屏或内容截断；
 - 2) 听觉方式：呈现的语音或提示音内容应与预期反馈一致，清晰可辨，不应出现杂音或异常

中断；

- 3) 触觉方式：呈现的振动反馈模式应与预期事件类型匹配，反馈启动延迟应符合第 7 章的规定。

- b) 测试过程中系统不应出现反馈内容错误、呈现中断或超时无响应等异常情况。

8.3.通用性能要求测试方法

8.3.1. 感知覆盖范围测试方法

8.3.1.1. 测试目的

验证系统在家庭环境下，各类感知输入设备在采集用户行为、操作指令和环境状态的过程中，其空间覆盖能力是否符合第 7.2 章（或附录 A）的规定。

8.3.1.2. 测试工具

测试所需的仪器设备及辅助工具如下：

- 角度测量工具：量角器或测角仪，精度不低于 $\pm 0.5^\circ$ ，用于标定水平和垂直方向的有效覆盖角度边界；
- 距离测量工具：激光测距仪或卷尺，精度不低于 $\pm 1\text{cm}$ ，用于标定空间距离；
- 视觉感知测试辅助设备：标准视觉测试目标物（如标准人体轮廓标靶、骨骼关键点标定假人等，可由经过培训的测试人员替代）；
- 听觉感知测试辅助设备：标准听觉测试目标物（如高保真音箱、标准语料播放设备等）；
- 环境条件测试工具：照度计、可调光源、分贝仪、噪声仪等；
- 辅助固定装置：用于固定标准测试目标物的支架或夹具。

8.3.1.3. 测试组网

感知覆盖范围测试组网应按图 3 所示连接。在智能家庭业务系统测试组网中加入标准测试目标（视觉、听觉感知测试辅助设备），并使用距离和角度测量工具来标定范围。

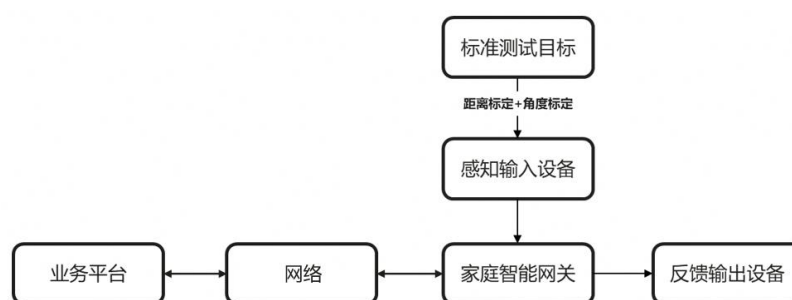


图 3 感知覆盖范围测试组网图

8.3.1.4. 测试步骤

在按 8.1.2 的规定测试准备后，应按以下步骤执行测试：

- 使用照度计和噪声仪测量测试区域的环境光照度和噪声值，并通过可调光源和噪声仪将环境条件调节至被测系统技术声明中规定的范围；
- 感知距离测试：使用距离测量工具标定标准测试目标物与感知输入设备之间的空间距离。将标准测试目标物沿感知输入设备主轴方向，依次放置于所规定的有效感知距离最近边界和最远

边界位置，在每个边界测试点持续观察不少于 10s，记录系统能否稳定采集感知数据或正确识别目标（如识别人体姿态、捕捉骨骼关键点、拾取语音信号、检测目标存在等）；

- c) 覆盖角度测试：将标准测试目标物固定于有效感知距离范围内的某一位置，使用角度测量工具标定水平和垂直方向的角度。沿水平方向分别移动所规定的水平覆盖角度左边界和右边界位置，在各边界点观察并记录系统能否稳定采集感知数据或正确识别目标；同法沿垂直方向分别移动至所规定的垂直覆盖角度上边界和下边界位置，在各边界点观察并记录结果。每个边界测试点应重复测试不少于 3 次；
- d) 覆盖面积测试：对于以覆盖面积表征的感知覆盖范围，以感知输入设备在墙面或顶面的安装位置为基准点，将标准测试目标物沿水平方向自基准点向外移动至所规定的覆盖面积边界位置，沿边界轨迹持续移动不少于 30s，记录系统能否稳定采集感知数据或正确识别目标。

8.3.1.5. 预期结果

测试结果应符合以下要求：

- a) 系统在有效感知距离最近边界和最远边界及范围内，应能稳定采集感知数据或正确识别目标，不应出现数据丢失或识别中断；
- b) 系统在水平及垂直覆盖角度边界及范围内，应能稳定采集感知数据或正确识别目标，各边界测试点重复测试不少于 3 次的结果应一致，且结果应满足 7.2 章或附录 A 要求；
- c) 对于以覆盖面积表征的场景，系统在覆盖面积边界及范围内，应能稳定采集感知数据或正确识别目标。

8.3.2. AI 决策有效性测试方法

8.3.2.1. 测试目的

验证系统 AI 决策有效性是否符合第 7.3 章（或附录 A）的规定，包括 AI 感知理解类任务（准确率、召回率、漏报率、误报率）和 AI 反馈优化类任务的各项性能指标。

8.3.2.2. 测试工具

测试所需的仪器设备及辅助工具如下：

标准测试用例集：用于评估系统 AI 决策有效性的测试用例集合，应覆盖被测系统所声称支持的各种典型场景及边界条件，包含预设的感知输入数据及对应的预期决策结论。其中：

- a) AI 感知理解类任务的数据集应明确声明总样本数、正例样本数、负例样本数及其占比，各类别样本数量及比例应满足测试结果的统计有效性，标注依据及标注结果的有效性应随数据集一并说明；
- b) AI 反馈优化类任务测试数据集应包含预设的输入源文件（如原始视频）和对应的预期输出效果要求。

8.3.2.3. 测试组网

AI 决策有效性测试组网应按图 4 所示连接。在智能家庭业务系统测试组网中加入标准测试用例（评估系统 AI 决策有效性的测试用例集合）作为感知输入设备的输入。

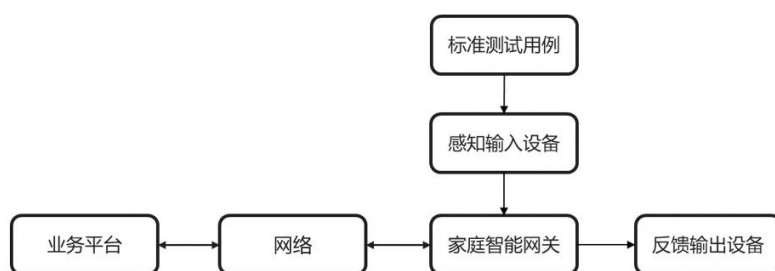


图 4 AI 决策有效性测试组网图

8.3.2.4. 测试步骤

在按 8.2.1 的规定完成测试准备后，应按以下步骤执行测试：

- 根据被测系统所声明的 AI 决策能力类型（如姿态识别、行为检测、语音识别、目标检测、内容生成等），选择对应的标准测试用例作为系统输入；
- 系统对输入数据进行推理分析后输出决策结论，逐项记录系统输出结果；
- 将系统输出结果与预期决策结论进行比对，验证决策结果的一致性。对于 AI 感知理解类任务，按照第 7.3 章规定的指标计算方法，统计各项性能指标的实测值；对于 AI 反馈优化类任务，按照 8.3.3 测试交互响应质量。

8.3.2.5. 预期结果

测试结果应符合如下要求：

- 系统在 AI 感知理解类任务中的准确率、召回率、漏报率和误报率等指标，其实测值应不低于附录 A 所规定的限值；系统在 AI 反馈优化类任务的相关性能要求，应满足 7.3 章（或附录 A）要求。
- 测试过程中系统不应出现无法输出决策结论、决策中断或输出明显异常等情况。

8.3.3. 交互响应质量测试方法

8.3.3.1. 测试目的

验证系统交互响应质量是否符合第 7.4 章（或附录 A）的规定，包括端到端响应时延、安全隐私保护能力和音视频呈现质量等指标。

8.3.3.2. 测试工具

测试所需的仪器设备及辅助工具如下：

- 高速摄像机：用于录制显示终端画面，拍摄帧率应不低于反馈输出终端（显示设备）刷新率的 2 倍；
- 逐帧视频播放器：具备逐帧播放和定位能力，支持的最高分辨率和帧率应不低于高速摄像机拍摄输出的分辨率和帧率；
- 评分记录表：用于记录主观体验评分结果。

8.3.3.3. 测试组网

交互响应质量测试组网应按图 5 所示连接。在智能家庭业务系统测试组网中加入高速摄像机，用来拍摄录制反馈输出设备画面。

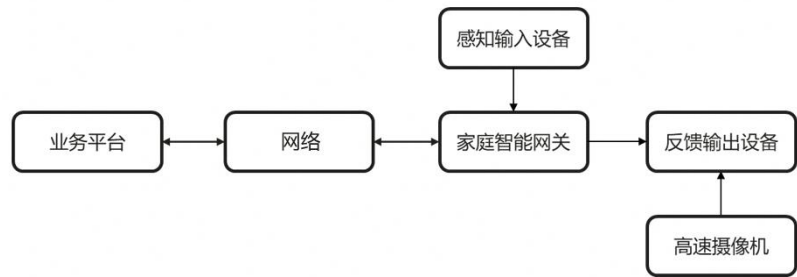


图 5 交互响应质量测试组网图

8.3.3.4. 测试步骤

在按 8.2.1 的规定完成测试准备后，应按以下步骤执行测试：

- a) 测试人员执行规定的标准操作输入（如体感动作、语音指令、遥控按键、游戏操控等），确认输入设备动作与显示终端画面均出现相应变化；
- b) 使用高速摄像机连续录制显示终端画面，每次录制时长应满足统计有效性，录制内容应覆盖被测系统所声明的典型交互场景；
- c) 将录制的测试视频导入逐帧视频播放器，分别定位到输入设备操作发生变化的起始画面帧和显示终端发生变化的起始画面帧，记录对应时刻 T_1 和 T_2 ；
- d) 按公式（5）计算每次响应的端到端时延：

$$\Delta t = T_2 - T_1 \tag{5}$$

式中：

Δt ：端到端响应时延，单位为毫秒（ms）；

T_1 ：输入设备操作发生变化的起始时刻，单位为毫秒（ms）；

T_2 ：显示终端发生变化的起始时刻，单位为毫秒（ms）。

- e) 将录制的测试视频导入逐帧视频播放器，分别定位到输入设备操作发生变化和显示终端发生变化的起始画面帧，记录对应时刻为 T_1 和 T_2 ；计算每次响应时延 $\Delta t = T_2 - T_1$ ，重复测量不少于 10 次，取算术平均值作为操作响应时延的测量值；
- f) 音视频呈现质量的测试应按以下要求执行：
 - 1) 视频内容呈现质量：当系统输出为视频内容时，应测试其画面分辨率和帧率。分辨率与帧率参数可从视频内容的生产方、传输流或家庭智能网关的媒体信息查询结果中获取，并以此作为评估依据；
 - 2) 音视频同步性：当系统输出同时包含音视频内容时，应对音画同步性进行主观评估；
 - 3) 3D 内容呈现质量：当系统输出涉及 3D 内容呈现时，应从入屏感、立体感、观看舒适度（不引起眩晕）等维度进行主观评估。：

8.3.3.5. 预期结果

测试结果应符合以下要求：

- a) 端到端响应时延的测量值应不低于第 7.4 章（或附录 A）所规定的限值；
- b) 画面分辨率和帧率应不低于第 7.4 章（或附录 A）所规定的限值；
- c) 音画同步性及 3D 内容呈现质量的主观评估结果应符合第 7.4 章（或附录 A）的规定；
- d) 测试过程中系统不应出现长时间无响应、画面凝固或周期性卡顿等异常情况。

附录 A

（规范性）

典型应用场景下智能家庭业务系统通用功能与性能要求

以下典型应用场景下智能家庭业务系统的通用功能与性能要求，适用于以下网络条件、环境条件及硬件规格：

- a) 网络条件：带宽 $\geq 1000\text{Mbps}$ ，边缘算力网络时延 $\leq 10\text{ms}$ ，中心算力网络时延 $\leq 30\text{ms}$ ，抖动 $\leq 30\text{ms}$ ，丢包率 $\leq 5 \times 10^{-5}$ ）；
- b) 环境条件：室内家居环境，安静环境噪声 $\leq 30\text{dBA}$ ，背景噪声 $\leq 60\text{dBA}$ ，光照度^[1] $\geq 75\text{lx}$ 且 $\leq 300\text{lx}$ ；
- c) 家庭智能网关硬件规格：CPU 核数 ≥ 4 核，内存（DDR）容量 $\geq 2\text{GB}$ 、存储（Flash）容量 $\geq 8\text{GB}$ 。

A.1 AI 互动健身

A.1.1 功能要求

AI 互动健身场景的功能应符合第 6 章的规定，并满足以下要求：

- a) 多模态感知：系统应通过有线网络、USB、Wi-Fi、星闪、红外、Zigbee 等至少一种连接方式，接入摄像头（含 RGB 摄像头、深度摄像头等）、激光雷达、外设（如电子球拍、健身环）等设备，实现对用户动作或操作的捕捉。当使用摄像头等设备采集环境信息时，应取得用户同意。
- b) 智能理解与决策：系统通过摄像头捕捉用户动作时，应支持 AI 骨骼识别；宜支持对用户动作进行实时规范度评估，识别关节角度偏差和运动轨迹错误。
- c) 交互反馈与呈现：系统应具备屏幕显示实时画面或虚拟人像的功能，宜支持实时动作提示和实时打分。

A.1.2 性能要求

AI 互动健身场景的性能应符合第 7 章的规定，并满足以下要求：

- a) 感知覆盖范围：单人模式下，系统对用户动作或操作的有效感知距离应覆盖 1.5 m 至 2 m；若系统支持双人模式，则有效感知距离应覆盖 1.5m 至 3m。在此范围内应能感知身高不超过 2 m 的站姿人员全身。
- b) AI 决策有效性：系统对人体骨骼关节识别的准确率应 $\geq 90\%$ ，按第 8.2 章规定的测试方法进行验证；动作库应至少包含标准站立、双手侧平举、深蹲（全蹲）和弓步蹲等动作。
- c) 交互响应质量：系统从用户开始执行动作到画面响应的端到端响应时延应 $\leq 500\text{ms}$ ；业务视频画面分辨率应 $\geq 1920 \times 1080$ ，帧率应 $\geq 30\text{FPS}$ ，应无明显音画不同步现象。

A.2 云游戏

A.2.1 功能要求

云游戏场景的功能应符合第 6 章的规定，并满足以下要求：

- a) 多模态感知：系统应支持对游戏手柄、键鼠、方向盘、触控屏等控制外设的操作信号采集。
- b) 智能理解与决策：系统可采用 AI 超分辨率、AI 插帧等技术提升画面质量与流畅度；系统可采用 AI 模型来生成游戏 NPC（非玩家角色）；
- c) 交互反馈与呈现：系统应支持游戏画面的实时呈现，交互以控制设备为核心，宜支持力反馈、振动等多维触觉反馈输出。

A. 2.2 性能要求

云游戏场景的性能应符合第 7 章的规定，并满足以下要求：

- a) 感知覆盖范围：系统外设的有效感知连接距离应 $\geq 2.5\text{ m}$ ；
- b) AI 决策有效性：系统采用 AI 超分辨率、AI 插帧等技术提升画面质量与流畅度时，其输出画面质量应符合“A2.2 c) 交互响应质量”的规定；
- c) 交互响应质量：系统从用户开始触发操作到画面响应的端到端交互响应时延应 $\leq 100\text{ms}$ ，游戏画面显示分辨率应 $\geq 1920 \times 1080$ ，宜支持 3840×2160 ，帧率应 $\geq 60\text{FPS}$ ，且无明显音画不同步现象。

A. 3 3D观影

A. 3.1 功能要求

3D 观影场景的功能应符合第 6 章的规定，并满足以下要求：

- a) 多模态感知：系统应具备与 3D 观影设备建立并维持无线连接的能力，以同步左右眼画面的交替或分配，支持在家庭环境中稳定使用；应支持通过红外、蓝牙、Wi-Fi 或星闪等无线方式连接的遥控器进行操控。
- b) 智能理解与决策：系统可在终端内置或通过云端调用 AI 算法，对标准 2D 视频源进行实时或离线分析处理，生成具有立体视觉效果的合成 3D 画面。
- c) 交互反馈与呈现：系统应具备解码并正确播放主流 3D 视频格式的能力，能支持 2D 与 3D 视频的手动或自动切换。

A. 3.2 性能要求

3D 观影场景的性能应符合第 7 章的规定，并满足以下要求：

- a) 感知覆盖范围：系统应保证用户所在位置的有效观看覆盖距离应 $\geq 3\text{m}$ ，有效观看覆盖角度水平视场角应 $\geq 60^\circ$ ，垂直视场角应 $\geq 60^\circ$ 。
- b) AI 决策有效性：当系统支持对标准 2D 视频源进行分析处理并生成具有立体视觉效果的合成 3D 画面时，其输出画面质量应满足“A.3.2 c) 交互响应质量”规定的要求。
- c) 交互响应质量：系统 3D 画面显示分辨率应 $\geq 1920 \times 1080$ ，宜支持 3840×2160 ，显示刷新率应 $\geq 60\text{Hz}$ ，宜支持 120Hz ，观看有入屏感和立体感，无眩晕感和明显音画不同步现象。

A. 4 康养看护跌倒检测

A. 4.1 功能要求

康养看护跌倒检测场景的功能应符合第 6 章的规定，并满足以下要求：

- a) 多模态感知：系统应支持通过有线网络、USB、Wi-Fi、星闪、红外、Zigbee 等至少一种连接方式，接入跌倒检测终端。当系统采集用户行为姿态信息时，应采取隐私保护措施。
- b) 智能理解与决策：系统应具备对处于有效探测范围内的人员姿态进行检测的能力，能识别跌倒、长时间滞留等异常状态，并在应用客户端或其他用户界面提示告警信息。
- c) 交互反馈与呈现：系统客户端应具备告警列表查看、告警画面回放功能；系统应具备隐私区域

监控功能，支持直播或回放呈现画面，画面应能清晰辨别老人位置，但不呈现人体轮廓细节。

A. 4. 2 性能要求

康养看护跌倒检测场景的性能应符合第 7 章的规定，并满足以下要求：

- a) 感知覆盖范围：系统终端采用壁挂方式安装，应满足安装方式对应要求：
 - 1) 在卫生间等天花板高度 $\leq 2.4\text{m}$ 房间，设备安装高度 2m，有效探测面积 $\geq 2\text{m} \times 2\text{m}$ 。
 - 2) 在卧室等天花板高度 $\leq 3\text{m}$ 的房间，设备安装高度在 2.5m，有效探测面积 $\geq 3\text{m} \times 3\text{m}$ 。
- b) AI 决策有效性：在所要求的感知覆盖范围内，检测跌倒、滞留，准确率应 $\geq 95\%$ ，漏报率应 $\leq 5\%$ 。跌倒动作应至少覆盖正例（快跌、慢跌、向前/向后跌倒、向斜/侧面跌倒、翻身滚落）和负例（扫地、拖地、擦地）。
- c) 交互响应质量：系统跌倒检测端到端告警时延应 $\leq 60\text{s}$ 。

A. 5 AI 伴学

A. 5. 1 功能要求

AI 伴学场景的功能应符合第 6 章的规定，并满足以下要求：

- a) 多模态感知：
 - 1) 语音感知：系统应支持对家庭学习环境中目标说话人（学生）语音的实时采集；采集用户语音等个人信息时，应取得用户同意。
 - 2) 视觉感知：系统应支持对书本页面、试卷、作业本等印刷文字与手写文字的实时图像采集与 OCR 识别，识别范围至少涵盖汉字、拼音、英文字母、阿拉伯数字及常见数学符号。应支持对数学公式、几何图形、化学结构式的结构化识别与解析；
 - 3) 交互行为感知：系统宜支持对学生电子白板手写轨迹、屏幕触控操作、键盘输入等交互行为的实时采集。
- b) 智能理解与决策：
 - 1) AI 作业批改：系统宜支持对客观题（选择、判断、填空）的自动批改与即时反馈，应支持对主观题（计算题、简答题、作文）的 AI 辅助批改。
 - 2) AI 实时问答：系统宜支持学生以语音或文字方式自由提问，覆盖学科知识问答、解题方法询问、概念解释等类型。
- c) 交互反馈与呈现：系统应具备教育内容播放和呈现功能；系统宜支持根据学生年龄和年级自动适配学习内容的难度等级、界面风格和交互方式。

A. 5. 2 性能要求

AI 伴学场景的性能应符合第 7 章的规定，并满足以下要求：

- a) 感知覆盖范围：系统应保证用户所在位置的有效拾音覆盖距离在安静环境下 $\geq 5\text{m}$ ，嘈杂环境下 $\geq 3\text{m}$ 。
- b) AI 决策有效性：对多年级学科范围全真模拟试卷（文本/图像）客观题 AI 批改准确率应 $\geq 95\%$ ，试卷题目评测集数量宜覆盖多年级（小/初/高）科目内容，测试题目宜 ≥ 100 道。
- c) 交互响应质量：系统进行 AI 作业批改（单题目）、AI 指读等 AI 交互类任务时，端到端响应时延应 $\leq 5\text{s}$ 。

A. 6 AI 生活智能体

A. 6. 1 功能要求

AI 生活智能体的功能应符合第 6 章的规定，并满足以下要求：

- a) 多模态感知：
 - 1) 语音感知：系统应支持在典型家庭环境中拾取用户语音指令，具备声源定位、回声消除、噪声抑制等功能；宜支持声纹识别，能够区分不同家庭成员身份。
 - 2) 视觉感知：系统宜支持通过摄像头采集家庭环境图像或视频，具备人脸表情识别、手势识别等能力。
 - 3) 交互行为感知：系统宜支持对屏幕触控操作、键盘输入等交互行为的实时采集。
- b) 智能理解与决策：
 - 1) 意图理解：系统应支持对用户输入的自然语言进行语义解析，识别用户的明确意图和隐含意图，覆盖简单指令、复杂指令及模糊指令等类型；
 - 2) 多设备协同：系统应具备家庭设备之间协同工作的能力，支持单一指令触发多个设备协同执行任务，提供无感切换的全屋一致体验，确保各设备协同完成预期结果。
- c) 交互反馈与呈现：
 - 1) 任务执行与反馈：系统应将理解后的意图转换为实际动作，并在任务执行后向用户反馈结果，确保信息闭环，完成从执行到反馈闭环；
 - 2) 鲁棒性：系统应能在不同输入表达、环境条件及多轮交互过程中，面对输入扰动、环境变化、系统异常、恶意攻击等不确定性因素时，保持一致的理解结果、决策逻辑和行为表现。

A. 6.2 性能要求

AI 生活智能体场景的性能应符合第 7 章的规定，并满足以下要求：

- a) 感知覆盖范围：系统应保证用户所在位置的有效语音拾音覆盖距离安静环境下 $\geq 5\text{m}$ ，嘈杂环境下 $\geq 3\text{m}$ 。
- b) AI 决策有效性：
 - 1) 语音指令识别准确率：在家庭典型噪声环境（包括安静至中度噪声、不同距离）下，系统正确识别语音指令准确率应 $\geq 85\%$ ，宜 $\geq 90\%$ 。测试语料应至少覆盖正例（短/中/长/超长指令、带轻微口音普通话、成年/老人/儿童等不停性别和年龄、语速快）和负例（家庭成员间正常交谈、电视/广播背景语音）。
 - 2) 声纹识别准确率：在所要求的典型家庭噪声环境下，系统正确识别已注册家庭成员身份，准确率应 $\geq 90\%$ 。声纹特征应至少覆盖正例（验证声纹特征多样性，包括成年男性/女性、老年男性/女性、儿童、带方言口音的家庭成员）和负例（未注册陌生人声音、高保真录音回放）。
 - 3) 任务完成成功率：在智能体所管理的设备和场景范围内，系统正确完成用户指令的成功率应符合以下要求：
 - 简单任务成功率应 $\geq 90\%$ ，复杂任务成功率应 $\geq 60\%$ 。简单任务应至少覆盖正例（单设备/网络控制、单设备调节、基础查询）和负例（设备不存在、设备离线、权限不足）。
 - 复杂任务应至少覆盖正例（设备场景联动、含定时与条件的复合指令、多轮对话上下文保持）和负例（部分设备离线、指令自相矛盾、超出能力范围请求）。
- c) 交互响应质量：
 - 1) 系统语音唤醒响应时延应 $\leq 300\text{ms}$ ，语音交互响应时延应 $\leq 1000\text{ms}$ 。
 - 2) 系统简单任务响应时延应 $\leq 500\text{ms}$ ，复杂任务响应时延应 $\leq 3000\text{ms}$ 。

参考文献

[1]中华人民共和国住房和城乡建设部. 住宅项目规范: GB 55038-2025[S]. 北京: 中国建筑工业出版社, 2025.