

T/EJCCCSE

团 体 标 准

T/EJCCCSE XXXX-XXXX

智能音视频终端设计要求

Intelligent audio and video terminal design requirements

(征求意见稿)

20XX-XX-XX 发布

20XX-XX-XX 实施

中国商业股份制企业经济联合会 发布

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 基本规定 1

5 系统架构 2

6 音频性能 4

7 语音交互 4

8 安全性 5

9 智能化分级 7

10 测试方法 9

11 标志、包装、运输和贮存 12

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。

本文件由深圳市尊特数码有限公司提出。

本文件由中国商业股份制企业经济联合会归口。

本文件起草单位：深圳市尊特数码有限公司。

本文件主要起草人：×××

智能音视频终端设计要求

1 范围

本文件规定了智能音视频终端的基本规定、系统架构、音频性能、语音交互、安全性、智能化分级、测试方法以及标志、包装、运输和贮存等方面的要求。

本文件适用于具备蓝牙无线连接功能，可通过语音交互操作，且能提供音频播放服务的智能音视频终端产品及其系统的研发、设计和测试。

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 191-2008 包装储运图示标志
- GB 4943.1-2022 音视频、信息技术和通信技术设备 第1部分：安全要求
- GB/T 14277-2013 音频组合设备通用规范
- GB/T 33980-2017 电工产品使用说明书中包含电气安全信息的导则

3 术语和定义

下列术语和定义适用于本文件。

3.1

智能音视频终端 Intelligent audio and video terminal

具有多设备无线蓝牙互联功能的一种智能音视频音箱终端，兼具语音交互等功能，能够加载或播放音频以及视频中的音频。

4 基本规定

4.1 型号与规格

- 4.1.1 智能音视频终端的型号应具备唯一性与系统性，便于产品识别、管理及用户区分。
- 4.1.2 制造商内部应制定型号编制规则规范文件，明确型号各组成部分所承载的产品系列、功能特性、外观尺寸等信息。
- 4.1.3 产品规格应依据市场定位与用户需求进行设定。
- 4.1.4 便携式智能音视频终端长度应为 100 mm ~ 200 mm，宽度应为 60 mm ~ 120 mm，高度应为 50 mm ~ 100 mm，应符合单手握持的人体工程学要求，方便携带。
- 4.1.5 桌面式智能音视频终端长度应为 200 mm ~ 400 mm，宽度应为 100 mm ~ 200 mm，高度应为 100 mm ~ 150 mm，应适配桌面摆放空间。
- 4.1.6 便携式智能音视频终端应不超过 500 g，桌面式智能音视频终端应为 1 000 kg ~ 3 000 kg。

4.2 外观与结构

- 4.2.1 智能音视频终端外观设计应遵循简洁大方原则，契合人体工程学原理，兼顾美观性与实用性。
- 4.2.2 外壳材质宜优先选用高强度、耐磨损且具备一定阻燃性能的工程塑料，亦可采用铝合金等金属材料外壳或纺织品布面。
- 4.2.3 智能音视频终端经喷漆或阳极氧化处理后的表面应均匀平整，无气泡、流痕、砂眼等瑕疵，颜色应均匀且具备良好的耐久性，不易褪色。
- 4.2.4 智能音视频终端外观拼接缝隙应不明显，各部件装配应紧密，缝隙宽度应控制在 0.5 mm 以内。
- 4.2.5 智能音视频终端内部组件应安装稳固，具备良好的抗震性能。
- 4.2.6 扬声器、电路板、电池等核心部件应通过合理的机械固定方式安装于智能音视频终端内部，防止因震动致使部件松动、接触不良等问题。
- 4.2.7 智能音视频终端结构设计应便于拆卸维护，对于电池、传声器等可能需更换的零部件，应设计专用可拆卸结构，方便售后维修人员操作，且在拆卸过程中不得对其他部件造成损坏。

4.3 接口与按键

- 4.3.1 除特殊规定外，智能音视频终端充电接口应由制造商结合产品客观生产实际自行明确。
- 4.3.2 接口应具备高可靠性，插拔寿命应不低于 10 000 次，在插拔过程中应始终保持良好的电气连接，避免出现接触不良现象。
- 4.3.3 智能音视频终端宜配备标准的 MicroSD 卡槽，卡槽位置应便于用户插拔存储卡，并具备防尘、防异物侵入设计。
- 4.3.4 存储卡读写速度应满足音、视频文件快速读取要求。
- 4.3.5 按键设置宜包含但不限于下列各项：
 - a) 电源键；
 - b) 音量调节键；
 - c) 播放 / 暂停键；
 - d) 语音唤醒键。
- 4.3.6 按键布局应符合人体操作习惯，各按键间间距应不小于 5 mm，防止误操作。
- 4.3.7 按键应具备良好的触感反馈，按下与弹起动作应灵敏，按键操作力应控制在 50 克力 至 150 克力之间。
- 4.3.8 按键功能标识应清晰明确，宜采用丝印或激光雕刻等工艺在按键表面或周边进行标注，标识应具备高耐磨性，不易褪色。

4.4 注意事项

- 4.4.1 智能音视频终端正常运行时，表面温度应保持在合理范围，连续播放音频 2 h 以上，表面最高温度不得超过 50 ℃。
- 4.4.2 智能音视频终端指示灯光亮度应适度，避免对用户眼睛造成强光刺激。
- 4.4.3 智能音视频终端应具备优良的抗干扰性能，在存在其他电子设备干扰，如微波炉、无线路由器等环境中，应仍能稳定运行，不得出现蓝牙连接中断、语音识别错误、音频播放卡顿等现象。

5 系统架构

5.1 输入模块

5.1.1 语音采集

- 5.1.1.1 智能音视频终端需配备灵敏度适配的传声器或传声器阵列。
- 5.1.1.2 在安静室内环境，即环境噪声等效声级 ≤ 40 dB (A) 时，传声器应能清晰采集半径 3 m 范围内的用户语音。
- 5.1.1.3 在一般室内环境，即 40 dB (A) $<$ 环境噪声等效声级 ≤ 55 dB (A) 的场景下，传声器应能清晰采集半径 2 m 范围内的用户语音。
- 5.1.1.4 传声器的频率响应范围应在 20 Hz $\sim$ 20 kHz 内保持相对平坦，波动范围应控制在 ± 3 dB。
- 5.1.1.5 传声器应具备一定的抗噪性能，在 55 dB (A) 的噪声环境下，信噪比应不低于 40 dB。

5.1.2 蓝牙连接

- 5.1.2.1 智能音视频终端所支持的蓝牙版本应为最新版本。
- 5.1.2.2 在空旷环境中，蓝牙连接有效距离应不小于 10 m；在存在一定遮挡物的室内环境下，连接有效距离应不小于 8 m。
- 5.1.2.3 蓝牙连接建立时间应控制在 3 s 以内，且在连接过程中应具备自动识别并连接最近使用过设备的功能，连接切换时间应不超过 1 s。
- 5.1.2.4 数据传输过程中，蓝牙传输速率应不低于 2 Mbps，丢包率应低于 0.1%。

5.2 处理模块

5.2.1 语音识别

- 5.2.1.1 识别引擎须支持连续语音识别功能。
- 5.2.1.2 在低噪环境（环境噪声等效声级 ≤ 40 dB (A)）中，语音识别字准确率应不小于 85%（基础级）或 90%（增强级）；在高噪环境（ 40 dB (A) $<$ 环境噪声等效声级 ≤ 70 dB (A)）中，语音识别字准确率应不小于 80%（基础级）或 85%（增强级）。
- 5.2.1.3 从语音采集结束至给出识别结果的时间应不超过 1 s。
- 5.2.1.4 识别引擎应具备自适应学习能力，在用户持续使用过程中，可根据用户的语音习惯和常用指令，不断优化识别准确率，每 100 次交互后，识别准确率提升幅度应不低于 0.5%。

5.2.2 自然语言理解

- 5.2.2.1 系统应能精准理解用户自然语言表达的意图。
- 5.2.2.2 针对常见指令，如播放音乐、查询天气、设置闹钟等，理解准确率应不低于 85%（基础级）或 90%（增强级）。
- 5.2.2.3 对于模糊指令或相似指令，系统应借助语义分析算法、上下文理解技术等，具备一定的语义辨析能力，将误判率控制在 10% 以内（基础级）或 5% 以内（增强级）。

5.2.3 业务逻辑

- 5.2.3.1 依据自然语言理解的结果，业务逻辑模块需迅速且准确地映射到相应的业务线，并向相关应用下达指令以及提供反馈信息。
- 5.2.3.2 业务逻辑处理的平均响应时间应不超过 2 s（基础级）或 1.5 s（增强级）。
示例：在接收到播放音乐指令后，应在 1 s 内完成歌曲搜索范围的确定，2 s 内完成歌曲加载并开始播放。
- 5.2.3.3 业务逻辑模块应具备多任务处理能力，可同时响应多个用户指令，并按照指令的优先级和相关性进行合理排序与处理，在多任务处理时，响应时间延长应不超过 0.5 s。

5.3 输出模块

5.3.1 语音合成

- 5.3.1.1 智能音视频终端应支持清晰、流畅且自然的汉语普通话合成语音，宜支持多音色选择。
- 5.3.1.2 汉语普通话合成语音的 MOS 评分应不小于 3.5（基础级）或 4.2（增强级）（满分 5.0）。
- 5.3.1.3 除支持男声和女声合成外，宜包括部分特色音色，如儿童声、卡通声等。
- 5.3.1.4 应支持中英文混读，宜支持混合语种以及多语种合成。
- 5.3.1.5 合成语音的语速调节范围应在 -20% 至 $+20\%$ 之间，语调调节范围应能覆盖常见的情感表达，如欢快、严肃、悲伤等。

5.3.2 音频播放

- 5.3.2.1 应能将音、视频数据转换为高质量的声音输出。
- 5.3.2.2 应全面支持常见的音频格式，如 MP3、AAC、WAV、FLAC 等，对于不同格式音频的解码时间应控制在 0.5 s 以内。
- 5.3.2.3 在播放过程中，应能依据音频内容的特性，如音乐类型、动态范围等，自动调整音量、音效等参数，以营造最佳的听觉体验。
- 5.3.2.4 音量调节应具备至少 10 级的精细调节功能，相邻两级之间的音量变化应在 $1\text{ dB} \sim 3\text{ dB}$ 之间。

6 音频性能

6.1 声学性能参数

6.1.1 额定功率

- 6.1.1.1 在正常工作条件下，智能音视频终端的持续输出功率需依据产品定位进行精准设定。
- 6.1.1.2 对于小型便携式智能音视频终端，额定功率应为 $3\text{ W} \sim 10\text{ W}$ 。
- 6.1.1.3 对于中型桌面智能音视频终端，额定功率应为 $10\text{ W} \sim 30\text{ W}$ 。

6.1.2 频率响应

- 6.1.2.1 小型便携式智能音视频终端频率响应范围应为 $80\text{ Hz} \sim 18\text{ kHz}$ 。
- 6.1.2.2 中型桌面智能音视频终端频率响应范围应为 $60\text{ Hz} \sim 20\text{ kHz}$ 。
- 6.1.2.3 基础级产品的频率响应曲线波动应控制在 $\pm 3\text{ dB}$ 以内。
- 6.1.2.4 增强级产品的波动应控制在 $\pm 2\text{ dB}$ 以内。

6.1.3 信噪比

- 6.1.3.1 基础级智能音视频终端的信噪比应不低于 70 dB 。
- 6.1.3.2 增强级智能音视频终端的信噪比应不低于 80 dB 。

6.2 扬声器性能

- 6.2.1 智能音视频终端的扬声器性能应达到 GB/T 14277—2013 中的 B 类及以上。
- 6.2.2 基础级智能音视频终端的失真度应控制在 5% 以内。
- 6.2.3 增强级智能音视频终端的失真度应控制在 3% 以内。
- 6.2.4 智能音视频终端扬声器的灵敏度应在 $80\text{ dB} \sim 90\text{ dB}$ 之间。

7 语音交互

7.1 语音唤醒

7.1.1 智能音视频终端的语音唤醒功能需具备高可靠性与高效性。

7.1.2 智能音视频终端应支持特定唤醒词识别，唤醒词应具有唯一性与辨识度，避免与日常用语混淆；唤醒词的长度宜在 2 ~ 4 个音节。

7.1.3 在安静居住环境（环境噪声等效声级 ≤ 40 dB (A)）中，基础级产品的唤醒正确率应 $\geq 90\%$ ，增强级产品应 $\geq 95\%$ ，且应能在半径 5 m 范围内准确响应，响应时间应不超过 0.5 s。

7.1.4 在混合商业居住环境（ 40 dB (A) $<$ 环境噪声等效声级 ≤ 55 dB (A)），基础级产品唤醒正确率应 $\geq 85\%$ ，增强级产品应 $\geq 90\%$ ，且应能在半径 4 m 范围内准确响应，响应时间应不超过 0.8 s。

7.1.5 误唤醒频度应不大于 5 次 / 24 h。

7.1.6 智能音视频终端应具备自适应环境噪声调整功能，应根据环境噪声的变化实时优化唤醒灵敏度，以维持稳定的唤醒性能。

7.2 语音打断

7.2.1 在声源距离待测样品 1 m 距离处，基础级产品的语音打断成功率应不小于 85%，增强级产品应不小于 90%。

7.2.2 当用户发出打断指令时，智能音视频终端应能在 0.3 s 内暂停当前正在执行的语音输出任务，并迅速切换至接收新指令状态。

7.2.3 系统应实时监测语音输入，当检测到打断指令的语音特征时，应立即中断当前语音合成或音频播放进程，且确保打断过程不会对音频播放的音质产生明显影响。

7.2.4 语音打断功能应能与语音唤醒功能协同工作，在唤醒状态下及非唤醒状态下（若支持非唤醒状态下的语音打断），均能稳定运行。

7.3 交互拒识率

7.3.1 基础级产品的交互拒识率应小于 15%，增强级产品应小于 10%。

7.3.2 语音识别引擎应具备强大的抗噪能力与语言模型适应性，应能准确识别各种口音、语速及语言习惯下的语音指令。

7.3.3 自然语言理解模块应采用先进的语义分析算法，精确理解用户指令的意图，对于模糊或歧义指令，能够通过上下文推理及多轮交互引导用户明确意图。

7.3.4 业务逻辑处理模块应具备高效的任务调度与执行能力，在接收到明确指令后，应能够迅速准确地完成相应任务。

7.3.5 在实际应用中，应通过大数据分析 with 用户反馈，不断优化语音交互系统，持续降低交互拒识率。

7.3.6 在产品说明书中，应明确告知用户可能导致交互拒识的常见原因及解决方法，如网络状况不佳、环境噪声过大等。

8 安全性

8.1 信息安全

8.1.1 数据采集

8.1.1.1 智能音视频终端在数据采集环节，应仅收集与产品功能实现直接相关的用户数据。

示例：为实现语音交互功能，仅采集用户的语音指令数据，且在采集前需通过产品界面或相关提示明确告知用户数据采集的目的、范围及使用方式，并获得用户的明确授权。

8.1.1.2 采集的语音数据应进行匿名化处理，去除可识别用户身份的敏感信息。

8.1.2 数据存储

8.1.2.1 存储用户数据时，应采用安全可靠的存储技术。

8.1.2.2 本地存储设备应具备加密功能，对存储的用户语音数据、设置信息等进行加密存储，加密算法应符合相关规范。

8.1.2.3 对于存储在云端的数据，应选择具备高安全防护能力的云服务提供商，云平台应具备完善的数据备份与恢复机制。

8.1.2.4 应设置严格的访问控制策略，有且仅有经过授权的系统模块和人员可访问存储的数据，且访问记录应进行详细留存，以便审计和追溯。

8.1.3 数据传输

8.1.3.1 在数据传输过程中，须采用加密协议。

8.1.3.2 对于蓝牙传输的数据，应遵循蓝牙技术规范中的安全机制。

8.1.3.3 应建立网络安全通道，对传输的数据进行实时加密和解密，保障数据在不同网络环境下的安全传输。

8.1.3.4 应具备数据传输完整性校验机制，通过数字签名、哈希算法等技术，确保接收端接收到的数据与发送端发送的数据一致，未被中途篡改。

8.1.4 数据使用

8.1.4.1 智能音视频终端对用户数据的使用应严格遵循用户授权范围。

8.1.4.2 在利用用户数据进行功能优化、个性化推荐等操作时，应确保数据使用过程的透明度。

8.1.4.3 禁止将用户数据用于未经授权的第三方商业用途，防止用户数据被滥用。

8.1.4.4 若涉及数据共享，须事先获得用户的明确同意，并确保接收数据的第三方具备相应的信息安全保障能力。

8.1.5 用户隐私保护

8.1.5.1 产品应提供完善的用户隐私保护措施。

8.1.5.2 用户应能便捷查询、修改和删除自己的个人数据。

8.1.5.3 在产品界面中，应设置专门的隐私设置入口，允许用户自主选择是否开启数据采集功能，以及对已采集数据进行管理。

8.1.5.4 对于用户的隐私信息，如声纹、语音记录等，应进行严格保密，不得泄露给任何未经授权的第三方。

8.1.5.5 应建立用户隐私保护应急预案，若发生数据泄露事件，应迅速采取措施，通知用户并进行数据修复和安全加固，降低用户损失。

8.2 电气安全

8.2.1 硬件安全

8.2.1.1 智能音视频终端的硬件设计应符合 GB 4943.1-2022 中的规定。

8.2.1.2 智能音视频终端绝缘电阻应不低于 $2\text{M}\Omega$ ，接地电阻应不大于 0.1Ω 。

8.2.1.3 智能音视频终端外壳宜采用阻燃材料。

8.2.1.4 智能音视频终端外壳应具备良好的防护性能，应能有效防止外物侵入，避免内部电路受到损坏，影响电气安全。

8.2.2 充电安全

- 8.2.2.1 智能音视频终端充电接口应采用符合相关安全规范的设计。
- 8.2.2.2 在充电过程中，当检测到充电电压或电流异常时，应立即停止充电，防止电池过充、过热损坏甚至引发安全事故。
- 8.2.2.3 充电电路应具备过热保护功能，当充电过程中电路温度超过安全阈值时，自动启动散热措施或切断充电电源，待温度恢复正常后再恢复充电。
- 8.2.2.4 宜采用智能充电管理芯片，根据电池的充电状态自动调整充电电流和电压，延长电池使用寿命，保障充电过程的安全性。

8.2.3 电气兼容性

- 8.2.3.1 智能音视频终端应具备良好的电气兼容性，在与其他电子设备共同使用时，不应出现相互干扰，影响电气安全。
- 8.2.3.2 应通过电磁兼容性（EMC）测试，明确产品在正常工作过程中产生的电磁辐射符合相关规范，不应出现对周围的电子设备，如手机、电视、医疗设备等造成干扰。
- 8.2.3.3 产品自身应具备抗干扰能力，在受到外部电磁干扰时，仍能稳定运行，不出现功能异常、电气故障等问题，保障用户在各种复杂电磁环境下的安全使用。

9 智能化分级

9.1 基础级

9.1.1 语音交互功能

- 9.1.1.1 应具备基本的语音交互能力，能够识别并执行常见的简单指令，满足至少 50% 基础级语音交互指标要求。
- 9.1.1.2 语音识别准确率在安静环境（环境噪声等效声级 ≤ 40 dB (A)）下应不低于 80%，在一般环境（ 40 dB (A) $<$ 环境噪声等效声级 ≤ 55 dB (A)）下不低于 75%。

9.1.2 音频播放功能

- 9.1.2.1 应可通过蓝牙连接手机、平板等智能设备，流畅播放设备中的音频内容。
- 9.1.2.2 应支持常见音频格式，音频解码时间应不超过 0.5 s，播放过程中卡顿次数每小时不超过 2 次。
- 9.1.2.3 若具备连接云端声频媒体库功能，应能在 3 s 内完成连接并搜索到指定音乐资源，资源搜索准确率不低于 80%。

9.1.3 无线传输功能

- 9.1.3.1 应具备蓝牙无线传输功能。
- 9.1.3.2 在空旷环境下，蓝牙连接有效距离应不小于 8 m；在有一定遮挡物的室内环境下，连接有效距离应不小于 6 m。
- 9.1.3.3 蓝牙连接建立时间应不超过 4 s，连接稳定性应保证在持续播放音频 1 h 内，连接中断次数不超过 1 次。

9.1.4 扬声器性能

- 9.1.4.1 扬声器性能应符合 GB/T 14277—2013 中的 B 类标准。

9.1.4.2 在正常音量下，失真度应不超过 5%。

9.1.4.3 频率响应范围应能覆盖 100 Hz ~ 16 kHz，在该范围内频率响应曲线波动应控制在 ± 5 dB 以内。

9.2 增强级

9.2.1 全面基础级功能

9.2.1.1 应具备基础级所有功能，且在各项性能指标上进一步优化。

9.2.1.2 语音识别准确率在安静环境下应提升至 85%以上，一般环境下应达到 80%以上，指令响应时间应缩短至 1.5 s 以内。

9.2.1.3 音频播放应支持更多音频格式，解码时间应控制在 0.5 s 以内，播放卡顿次数每小时不超过 1 次。

9.2.1.4 蓝牙连接稳定性增强，在持续播放音频 2 h 内，连接中断次数不超过 1 次，连接有效距离在空旷环境下不小于 10 m，室内遮挡环境下不小于 8 m。

9.2.2 远距离识别功能

9.2.2.1 应具备远距离语音识别功能，在距离音箱不小于 5 m 的位置，用户发出的清晰语音指令应能被准确识别。

9.2.2.2 语音识别准确率在安静环境下应不低于 80%，一般环境下应不低于 75%。

9.2.3 降噪功能

9.2.3.1 在环境噪声等效声级为 55 dB (A) 的环境下，应能有效降低环境噪声对语音交互的影响。

9.2.3.2 语音识别准确率应保持在 75% 以上，语音唤醒正确率应达到 85%以上。

9.2.3.3 降噪处理不应影响语音的自然度和清晰度，合成语音的 MOS 评分应不低于 3.5。

9.2.4 智能家居控制功能（若有）

9.2.4.1 应支持常见的智能家居控制协议。

9.2.4.2 应能与智能家居设备进行稳定连接，连接建立时间应不超过 5 s。

9.2.4.3 应可通过语音指令实现对灯光、窗帘、智能插座等设备的控制，控制指令执行准确率不低于 90%。

9.2.5 扬声器性能提升

9.2.5.1 扬声器性能应符合 GB/T 14277—2013 中的 A 类标准。

9.2.5.2 失真度应在 3%以内，频率响应范围应为 80 Hz ~ 18 kHz，且在该范围内频率响应曲线波动应控制在 ± 3 dB 以内。

9.2.5.3 应能提供更丰富的声音细节和更宽广的音域，营造出更优质的音质效果。

9.3 高级

9.3.1 全面增强级功能

9.3.1.1 应具备增强级所有功能。

9.3.1.2 语音识别准确率在安静环境下应达到 90%以上，一般环境下应不低于 85%，指令响应时间应缩短至 1 s 以内。

9.3.1.3 音频播放的无损格式支持应更加完善，播放过程中应无卡顿现象，音频解码速度提升，对于高码率音频文件的解码时间不超过 0.3 s。

9.3.1.4 蓝牙连接在复杂环境下应依然稳定，持续播放音频 3 h 内，连接中断次数应不超过 1 次，连接有效距离在空旷环境下可达 15 m，室内遮挡环境下应不小于 12 m。

9.3.2 多轮语音交互功能

9.3.2.1 应具备强大的多轮语音交互能力，应能通过复杂的多轮对话准确完成用户任务，多轮交互成功率应不低于 90%。

9.3.2.2 应能理解用户的多轮对话中的上下文语义，应根据前序对话内容准确执行后续指令。

9.3.2.3 系统应具备智能引导功能，当用户指令不明确时，应能通过提问等方式引导用户明确需求，提升交互的流畅性和成功率。

9.3.3 个性化推荐功能

9.3.3.1 应根据用户的使用习惯和偏好，为用户推荐音乐、新闻、有声读物等内容，推荐准确率应不低于 80%。

9.3.3.2 应通过收集和分析用户的播放历史、收藏记录、收听时长、搜索记录等多维度数据，运用大数据分析和机器学习算法，构建用户兴趣模型。

9.3.3.3 推荐内容应具备多样性，避免单一重复推荐，同时应能根据用户反馈实时调整推荐策略，不断优化推荐效果。

9.3.4 智能场景联动功能

9.3.4.1 应可根据用户设置的场景条件，自动触发相关设备的操作，场景联动准确率应不低于 95%。

9.3.4.2 用户应能通过配套应用程序自定义场景，如“回家模式”“睡眠模式”“聚会模式”等。

示例：在“回家模式”下，当用户携带手机靠近音箱一定距离（通过蓝牙连接状态判断），音箱可自动联动打开家中灯光、调节空调温度、播放用户喜欢的音乐等一系列操作，且操作执行顺序和时间间隔可由用户自由设定。

9.3.4.3 场景联动功能应具备高可靠性，在网络延迟、设备故障等异常情况下，应能及时做出相应处理。

9.3.4.4 应能与更多类型的智能设备进行联动，拓展智能场景的应用范围。

10 测试方法

10.1 语音识别测试

10.1.1 测试环境搭建

10.1.1.1 应准备专门的消声室或半消声室作为测试场地，模拟安静室内环境（环境噪声等效声级 ≤ 40 dB (A)）；通过环境噪声发生器在室内制造噪声，营造一般室内环境（ 40 dB (A) $<$ 环境噪声等效声级 ≤ 55 dB (A)）以及高噪环境（ 55 dB (A) $<$ 环境噪声等效声级 ≤ 70 dB (A)）。

10.1.1.2 应在测试场地内合理布置智能音视频终端与用于播放语音样本的设备，明确智能音视频终端传声器与语音样本播放设备之间的距离符合实际使用场景，一般设置为 1 m。

10.1.2 测试样本准备

10.1.2.1 应收集并整理丰富的语音样本，一个涵盖常见的指令类型，如播放音乐、查询天气、设置闹钟、拨打电话（若具备此功能）等。

10.1.2.2 每种指令类型应准备不少于 50 个不同表述的样本，应包括不同口音（如普通话、粤语、英语等常见语种的不同口音）、语速（慢速、正常语速、快速）以及不同性别和年龄的发音。

10.1.2.3 语音样本应采用专业录音设备录制，音频质量应清晰、无失真。

10.1.3 测试流程

10.1.3.1 在不同噪声环境下，依次播放语音样本。

10.1.3.2 播放过程中，应记录智能音视频终端对每个语音样本的识别结果。

10.1.3.3 对于每个噪声环境，应重复播放所有语音样本至少 3 次。

10.1.3.4 在测试过程中，应明确智能音视频终端处于正常工作状态，连接稳定，且未进行其他干扰性操作。

10.1.4 数据记录与分析

10.1.4.1 应准确记录智能音视频终端正确识别的样本数量以及总样本数量。

10.1.4.2 应分别计算在不同噪声环境下的语音识别字准确率。

10.1.4.3 应对多次测试结果进行统计分析，计算平均值、标准差等统计量。

10.1.4.4 如条件允许，宜采用专业的数据统计分析软件进行数据分析。

10.2 语音唤醒测试

10.2.1 测试环境设定

10.2.1.1 应在消声室或半消声室中模拟 1 类安静居住环境（环境噪声等效声级 ≤ 40 dB (A)）和 2 类混合商业居住环境（ 40 dB (A) $<$ 环境噪声等效声级 ≤ 55 dB (A)）。

10.2.1.2 应在测试场地内设置多个测试点，距离智能音视频终端分别为 1 m、3 m、5 m（针对远距离唤醒功能测试）等不同距离，模拟用户在不同位置唤醒智能音视频终端的场景。

10.2.2 唤醒指令样本

10.2.2.1 应准备多个不同发音人、不同语速和语调的唤醒指令样本，每个发音人应录制不少于 20 次唤醒指令。

10.2.2.2 唤醒指令应涵盖智能音视频终端所支持的所有唤醒词或唤醒短语。

10.2.2.3 应将这些唤醒指令样本进行整理，形成测试指令集。

10.2.3 测试执行

10.2.3.1 在每个测试点和不同噪声环境下，按测试指令集依次播放唤醒指令。

10.2.3.2 记录每次播放唤醒指令后智能音视频终端的唤醒状态，即是否成功唤醒。

10.2.3.3 在测试过程中，应保持测试环境安静，避免其他无关声音干扰。

10.2.3.4 每个测试点和噪声环境组合下，应重复测试不少于 50 次。

10.2.3.5 应记录每次唤醒的响应时间，即从唤醒指令播放开始到智能音视频终端做出响应的时间间隔。

10.2.4 结果统计与评估

10.2.4.1 应统计不同测试点、不同噪声环境下智能音视频终端的正确唤醒次数和总唤醒次数。

10.2.4.2 对于误唤醒频度，在 24 h 连续测试过程中，应记录音箱出现误唤醒的次数，应明确误唤醒频度不大于 5 次 / 24 h。

10.2.4.3 应分析唤醒正确率与测试距离、噪声环境之间的关系，评估智能音视频终端语音唤醒功能在不同场景下的性能表现。

10.3 音频性能测试

10.3.1 测试设备准备

应采用专业的音频测试设备，应包括但不限于下列各项：

- a) 音频信号发生器；
- b) 功率放大器；
- c) 声级计；
- d) 频谱分析仪。

10.3.2 额定功率测试

- 10.3.2.1 应根据智能音视频终端的额定功率参数，通过音频信号发生器输入额定功率对应的音频信号，持续时间不少于 1 h。
- 10.3.2.2 在测试过程中，应采用功率计监测智能音视频终端的实际输入功率，明确其稳定在额定功率范围内。
- 10.3.2.3 应观察智能音视频终端的工作状态，检查是否有过热、异常噪音等现象。
- 10.3.2.4 测试结束后，应记录智能音视频终端在额定功率下的工作情况以及是否出现故障。

10.3.3 频率响应测试

- 10.3.3.1 应通过音频信号发生器产生频率范围为 20 Hz ~ 20 kHz 的扫频信号，以一定的频率间隔逐步改变信号频率，并将该扫频信号输入到智能音视频终端。
- 10.3.3.2 应采用频谱分析仪在智能音视频终端正前方 1 m 处接收输出的音频信号，分析其频率响应特性。
- 10.3.3.3 应记录在不同频率下智能音视频终端输出信号的幅度，绘制频率响应曲线。
- 10.3.3.4 应根据频率响应曲线，计算在整个频率范围内的波动情况，判断是否符合基础级（± 3dB 以内）或增强级（± 2dB 以内）的要求。

10.3.4 信噪比测试

- 10.3.4.1 应在智能音视频终端输入端口输入额定功率对应的音频信号，使用声级计测量智能音视频终端输出声音的声压级 L1。
- 10.3.4.2 关闭音频信号输入，测量智能音视频终端在无信号输入时的背景噪声声压级 L2。
- 10.3.4.3 应根据下列公式计算智能音视频终端的信噪比：

信噪比 = 20 log 10 (L1/L2) (1)

式中：
L1——输出声音的声压级；
L2——背景噪声声压级。

- 10.3.4.4 应重复测试 3 次，取平均值作为最终的信噪比测试结果。

10.3.5 扬声器失真度测试

- 10.3.5.1 应输入标准音频测试信号，调整输入信号幅度使智能音视频终端输出功率达到额定功率的 50%。
- 10.3.5.2 应采用失真度分析仪在音箱正前方 1 m 处测量智能音视频终端输出音频信号的失真度。
- 10.3.5.3 应分别测试基础级和增强级要求下的失真度，应明确在播放音乐或语音内容时，失真度控制在相应标准范围内（基础级 5% 以内，增强级 3% 以内）。

10.4 蓝牙连接测试

10.4.1 测试环境模拟

10.4.1.1 应搭建空旷环境测试场地，明确周围无明显的蓝牙信号干扰源。

10.4.1.2 应模拟有一定遮挡物的室内环境，在房间内放置桌椅、墙壁等遮挡物，以测试智能音视频终端在复杂环境下的蓝牙连接性能。

10.4.2 测试设备配对

10.4.2.1 应选择支持蓝牙功能的手机、平板等设备作为配对设备。

10.4.2.2 应将配对设备与智能音视频终端进行蓝牙配对，记录配对过程是否顺利，配对时间是否符合要求（不超过规定的蓝牙连接建立时间）。

10.4.3 连接距离测试

10.4.3.1 在空旷环境下，应将配对设备与智能音视频终端保持初始连接状态，然后逐渐增加两者之间的距离，同时播放音频文件以监测连接状态。

10.4.3.2 应记录智能音视频终端与配对设备之间能够保持稳定连接（无音频卡顿、中断现象）的最远距离，该距离应不小于规定的空旷环境蓝牙连接有效距离。

10.4.3.3 在有遮挡物的室内环境下，重复上述测试过程，记录室内环境下的有效连接距离。

10.4.4 连接稳定性测试

10.4.4.1 应在连接距离测试的基础上，在规定的连接有效距离内，让配对设备进行移动，模拟用户在实际使用过程中的移动场景。

10.4.4.2 在移动过程中，应持续播放音频文件，记录音频卡顿次数、连接中断次数以及中断后的自动重连时间。在多个蓝牙设备同时存在的环境下，测试智能音视频终端能否准确连接目标设备，避免出现连接错误或干扰其他设备的情况。

10.4.4.3 分析蓝牙连接稳定性与环境、设备移动等因素之间的关系，评估智能音视频终端蓝牙连接的可靠性。

11 标志、包装、运输和贮存

11.1 标志

11.1.1 智能音视频终端标志应包括但不限于下列各项内容：

- a) 产品名称；
- b) 产品型号与规格；
- c) 产品基本参数；
- d) 产品合格标识；
- e) 执行标准编号；
- f) 生产日期；
- g) 生产地址；
- h) 制造商商标。

11.1.2 包装箱上的包装储运图示标志应按 GB/T 191-2008 中规定的选择使用。

11.1.3 标志应清晰、牢固，不应因运输条件和自然条件而褪色、变色、脱落。

11.2 包装

- 11.2.1 包装箱内应有合格证、使用说明书及其他相关文件，使用说明书应符合 GB/T 33980-2017 中的规定。
- 11.2.2 包装箱应能保证智能音视频终端不受自然损坏。
- 11.2.3 包装箱内应有防尘、防震、防雨、防潮材料。
- 11.2.4 包装箱应有软性衬垫等，防止磕碰、划伤和污损。
- 11.2.5 运输包装的形式由制造厂商自行设计，但应保证智能音视频终端经过一般运输及正常装卸后完好无损。
- 11.2.6 包装箱上应有明显的注意标识和装箱方向等信息。
- 11.2.7 包装宜采用可降解材料或可回收材料。
- 11.2.8 包装箱与运输包装应符合 GB/T 191-2008 的规定。

11.3 运输

智能音视频终端在运输途中应平整堆放，应加遮盖物和进行必要的防护，避免冲击、局部重压、曝晒、雨淋及化学品的腐蚀。

11.4 贮存

- 11.4.1 智能音视频终端应贮存在干燥、清洁、通风的库房内。
 - 11.4.2 贮存温度应在 $-20\text{ }^{\circ}\text{C} \sim 50\text{ }^{\circ}\text{C}$ 之间，相对湿度应不大于 85%。
 - 11.4.3 智能音视频终端贮存环境中，空气中的有害物质的含量应不足以腐蚀蓝牙耳机。
 - 11.4.4 智能音视频终端应存放在平整的地面上。
 - 11.4.5 智能音视频终端堆放时应加衬垫物，以防挤压或变形。
-