|  |  |
| --- | --- |
| ICS | 35.240 |
| CCS | |  | | --- | | D:\000000部门项目\09标准化插件开发\程序源代码\StandardEditor_ShanDongKeXieYuan\团标首页面字母T.pngD:\000000部门项目\09标准化插件开发\程序源代码\StandardEditor_ShanDongKeXieYuan\团标首页面字母T后面的反斜杠.png XJBX |   L 67 |

西安市计量标准检测认证协会团体标准

T/XJBX 0066—2025

多模态数字媒体智能交互设计技术规范

Specification for intelligent interaction design of multimodal digital media

2025 - XX - XX发布

2025 - XX - XX实施

西安市计量标准检测认证协会  发布

目次

[前言 III](#_Toc205748472)

[引言 V](#_Toc205748473)

[1 范围 1](#_Toc205748474)

[2 规范性引用文件 1](#_Toc205748475)

[3 术语和定义 1](#_Toc205748476)

[4 总体设计原则 2](#_Toc205748477)

[5 功能结构与模块划分 3](#_Toc205748478)

[6 数据处理与融合方法 4](#_Toc205748479)

[7 交互逻辑设计 5](#_Toc205748480)

[8 用户体验优化 6](#_Toc205748481)

[9 安全与隐私保护 6](#_Toc205748482)

[10 性能测试与评价 7](#_Toc205748483)

[11 实施与维护 8](#_Toc205748484)

1. 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。

本文件由西安市计量标准检测认证协会提出并归口。

本文件起草单位：合肥财经职业学院。

本文件主要起草人：陈琳。

1. 引言

随着人工智能、虚拟现实、增强现实、物联网等技术的快速发展，数字媒体的交互方式逐渐从单一模态向多模态融合方向演进。多模态数字媒体智能交互技术通过整合视觉、听觉、触觉、语言及环境感知等多种信息通道，实现更自然、更高效的人机信息交换与情感互动，在文化创意、教育培训、医疗康复、智能制造、智慧城市等领域展现出广阔的应用前景。

然而，当前多模态交互系统在设计方法、性能评价、兼容性、可用性与安全性等方面尚缺乏统一的技术标准，导致系统开发过程存在接口不一致、设计规范不足、用户体验差异显著等问题。制定本技术规范旨在为多模态数字媒体智能交互系统的设计与实现提供系统性、可操作的技术要求，提升设计的科学性与一致性，促进产业链上下游的协同创新与跨平台兼容，保障用户的交互体验与数据安全。

本文件结合多模态交互技术的发展现状与前沿趋势，借鉴国内外相关研究成果与工程实践经验，明确了交互系统设计的总体原则、功能模块划分、数据处理与融合方法、交互逻辑设计、用户体验优化、安全与隐私保护等方面的技术要求，并提出测试与评价的方法框架，以指导相关单位在系统设计、研发、实施与验收过程中的标准化工作。

多模态数字媒体智能交互设计技术规范

* 1. 范围

本文件规定了多模态数字媒体智能交互系统在设计、开发、实施及评价过程中所需遵循的总体设计原则、功能结构与模块划分、数据处理与融合方法、交互逻辑设计、用户体验优化、安全与隐私保护、性能测试与评价及实施与维护等方面的技术要求。

本文件适用于涉及视觉、听觉、触觉、语言、姿态、环境感知等两种及以上模态信息的数字媒体交互系统的设计与实现，包括但不限于虚拟现实（VR）、增强现实（AR）、混合现实（MR）、全息影像、智能终端交互平台、智慧教室、沉浸式展览、智能客服等应用场景。

本文件可为多模态交互相关的研发单位、设计机构、集成商、测试与评价机构在产品设计、系统开发、工程实施及成果验收等环节提供参考，也可供行业管理部门、标准化组织和科研机构在相关标准制修订及技术研究中使用。

* 1. 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 2887—2011 计算机场地通用规范

GB/T 35273—2020 信息安全技术 个人信息安全规范

GB/T 38247—2019 信息技术 增强现实 术语

GB/T 25000.51 系统与软件工程 系统与软件质量要求和评价（SQuaRE） 第51部分：就绪可用软件产品（RUSP）的质量要求和测试细则

* 1. 术语和定义

下列术语和定义适用于本文件。

多模态交互 multimodal interaction

基于两种及以上感知模态（如视觉、听觉、触觉、语言、姿态等）的信息输入与输出，实现用户与系统之间信息交换与反馈的人机交互方式。

数字媒体 digital media

以数字化形式存储、处理和传播的文本、图像、音频、视频及其他多媒体内容的总称。

智能交互 intelligent interaction

利用人工智能技术对用户输入信息进行感知、理解、推理与决策，并生成符合语境和用户需求的交互反馈的过程。

模态 modality

人机交互中信息传递所依赖的感知通道或表现形式，如视觉模态、听觉模态、触觉模态、语言模态、嗅觉模态等。

多模态数据融合 multimodal data fusion

将来自不同模态的信息源进行同步处理、特征提取与优化整合，以提高交互系统感知精度、理解能力和响应质量的技术方法。

用户体验 user experience, UX

用户在使用产品或系统过程中所形成的全部主观感受与评价，包括可用性、易用性、沉浸感、满意度等方面。

沉浸感 immersion

用户在虚拟或增强环境中产生的深度参与和临场感受，通常由视觉、听觉、触觉等多模态感知的协调一致性和交互自然性决定。

* 1. 总体设计原则

多模态数字媒体智能交互系统的设计应遵循以下原则，以确保系统的功能性、兼容性、安全性及用户体验的整体优化。

* + 1. 用户中心原则

系统设计应以用户需求为导向，充分考虑目标用户群体的特征、使用习惯及交互偏好，通过需求分析和用户研究确保设计成果与用户期望高度契合。

* + 1. 多模态协同原则

应实现各模态信息的有机融合与互补优势发挥，确保视觉、听觉、触觉等信息在时间、空间及语义上的一致性，减少多模态冲突，提高交互的自然性与流畅度。

* + 1. 兼容性与可扩展性原则

系统设计应具备良好的软硬件兼容性和接口标准化能力，支持不同设备、平台与应用场景的无缝对接，同时便于后续功能扩展与升级。

* + 1. 安全与隐私保护原则

应在数据采集、传输、存储及处理的全过程中落实信息安全与隐私保护措施，防止数据泄露、滥用与非法访问，符合相关法律法规及行业标准要求。

* + 1. 高可用性与可访问性原则

系统应保持稳定性与可靠性，具备容错与冗余设计，保障在多种网络与硬件环境下的连续可用性。同时，应为不同年龄、能力与背景的用户提供无障碍交互支持。

* + 1. 低延迟与高性能原则

应优化交互链路与数据处理流程，降低响应延迟，提高渲染与计算性能，以提升实时交互的流畅度与沉浸体验。

* + 1. 可持续发展原则

应优先采用绿色节能的硬件方案与高效算法，减少能耗与碳排放，支持可持续的数字化建设与运营。

* 1. 功能结构与模块划分
     1. 感知与采集模块

负责对不同模态的信息进行实时感知与采集，包括视觉采集（摄像头、深度相机）、听觉采集（麦克风阵列）、触觉采集（力反馈装置、触摸传感器）、姿态采集（惯性测量单元）、环境感知（温湿度传感器、光照传感器等）。

该模块应具备多源同步采集能力和初步信号处理能力，确保数据质量与时序一致性。

* + 1. 数据处理与融合模块

对采集到的多模态数据进行特征提取、数据清洗、时间同步、格式转换及多模态融合处理，生成统一的特征表示。

融合方式可包括特征级融合、决策级融合及混合融合，应根据应用场景选择适合的融合策略，以提高感知精度与交互效率。

* + 1. 语义理解与决策模块

利用自然语言处理、计算机视觉、情感计算、知识图谱等人工智能技术，对多模态数据进行语义理解、情境推理与意图识别，并基于规则或机器学习模型进行决策生成。

* + 1. 输出与反馈模块

将系统决策结果通过多种模态输出方式反馈给用户，包括视觉输出（屏幕显示、全息影像）、听觉输出（语音播报、音效提示）、触觉输出（震动反馈、力反馈）等，确保反馈的及时性与可感知性。

* + 1. 用户交互与界面模块

负责构建直观、易用、可定制的用户界面和交互逻辑，包括交互流程管理、界面布局、自适应显示与多语言支持，提升用户体验与操作效率。

* + 1. 系统管理与安全模块

实现对硬件设备、软件服务及网络资源的统一管理，包含权限控制、数据加密、日志记录、异常监测与恢复机制，确保系统运行的稳定性与安全性。

* 1. 数据处理与融合方法
     1. 数据预处理

数据预处理应在采集后立即进行，包括噪声抑制、缺失值填补、畸变校正、信号增强、时间同步等步骤。处理包含以下内容：

1. 对视觉数据应进行去噪、色彩校正、畸变矫正及分辨率调整；
2. 对音频数据应进行滤波、语音增强、回声消除等处理；
3. 对触觉及环境传感数据应进行异常值检测与平滑处理。
   * 1. 特征提取与表示

应针对不同模态采用适配的特征提取方法，将原始数据转化为可用于融合的高质量特征表示：

1. 视觉模态可采用卷积神经网络（CNN）等深度学习模型提取空间特征；
2. 听觉模态可采用梅尔频率倒谱系数（MFCC）、声学特征向量等；
3. 语言模态可采用词向量、上下文嵌入（BERT等）；
4. 触觉模态可提取压力强度、震动频谱等参数。
   * 1. 融合策略

多模态融合应根据任务需求与系统架构选择合适的策略：

1. 特征级融合：在特征提取后对不同模态的特征进行拼接、加权或降维处理；
2. 决策级融合：在各模态独立处理后对输出结果进行加权投票、置信度融合等；
3. 混合融合：结合特征级与决策级融合方法，兼顾数据层与结果层的优势。
   * 1. 同步与对齐

应确保多模态数据在时间、空间及语义上的同步与对齐，避免因时序错位或语义不匹配导致交互性能下降。可采用时间戳同步、空间坐标映射、语义匹配等技术手段。

* + 1. 数据质量评估与优化

应建立多模态数据质量评估机制，包括数据完整性、精度、延迟、同步性等指标，并根据评估结果动态优化采集与融合策略。

* 1. 交互逻辑设计

多模态数字媒体智能交互系统的交互逻辑设计应在保证功能完整性与响应及时性的基础上，兼顾自然性、可理解性与用户体验的一致性。

* + 1. 交互流程规划

应根据应用场景和任务需求设计多模态交互流程，明确用户输入、系统处理、反馈输出的步骤与顺序，确保各环节信息传递流畅、衔接自然。

对任务驱动型交互，应优化任务分解与多模态信息互补的协作方式；对体验驱动型交互，应突出沉浸感与情感化反馈的连贯性。

* + 1. 多模态优先级与切换策略

在多模态并行输入或输出的情况下，应设定模态优先级与冲突处理机制。

优先级设定应结合模态的准确性、响应速度及用户习惯，切换策略应在主模态失效、环境变化或用户意图变化时实现平滑过渡，避免交互中断。

* + 1. 语义一致性与上下文管理

应保证不同模态输入输出在语义上的一致性，通过上下文管理机制维持交互状态的连续性与逻辑一致性。

上下文信息应涵盖用户历史操作、环境状态、任务进度等，系统应具备跨模态信息关联与语境理解能力。

* + 1. 用户反馈与确认机制

在关键交互节点，应提供多模态反馈确认机制，确保用户理解系统响应结果。反馈可包括视觉提示、语音播报、触觉震动等，并可结合任务重要性调整反馈强度与形式。

* + 1. 容错与自适应设计

应考虑用户输入错误、环境噪声干扰等情况，设计容错机制与自适应调整策略。

容错机制可包括冗余模态输入、错误提示与纠正建议，自适应策略应根据用户行为模式和环境变化动态优化交互逻辑。

* 1. 用户体验优化

多模态数字媒体智能交互系统的用户体验优化应在系统设计、开发与运行全过程中持续进行，以确保交互的自然性、舒适性和满意度。

* + 1. 可用性优化

应通过交互界面简化、功能布局合理化及交互路径优化，降低用户学习成本和操作负担。

优化界面信息层级，减少不必要的视觉干扰，设计直观的交互提示与引导，提升用户的操作信心。

* + 1. 沉浸感与临场感提升

应通过多模态信息的高同步性与高一致性来增强用户沉浸感。

保证视觉、听觉、触觉等输出的时间延迟在可接受范围内，利用空间音效、立体影像、触觉反馈等技术增强临场体验。

* + 1. 情感化交互设计

应在系统输出中引入情感计算与情绪识别机制，使交互更加人性化。

根据用户情绪状态动态调整交互语调、节奏与内容，在特定场景下提供情感激励或安抚性反馈。

* + 1. 个性化与自适应体验

应根据用户历史行为、偏好数据及当前环境进行个性化配置。

提供模态选择与布局自定义功能，实现基于AI的自适应优化，使系统行为与用户习惯逐步匹配。

* + 1. 可访问性支持

应保障不同能力水平、不同文化背景用户的平等使用权利。

提供辅助功能，如语音朗读、字幕、手势替代输入等，支持多语言、多文化内容适配，降低语言和文化障碍。

* + 1. 持续测试与反馈机制

应建立用户体验测试与反馈收集的长期机制，结合可用性测试、A/B测试及问卷调查等方法，持续优化交互体验。

* 1. 安全与隐私保护
     1. 数据安全管理

多模态数字媒体智能交互系统在数据采集、处理、传输、存储和输出等环节，应采取全面的安全与隐私保护措施，确保用户数据与系统运行的安全性。

系统应建立数据安全管理制度，对多模态数据全生命周期进行安全管控：

1. 数据采集环节应通过访问控制、授权管理等手段防止非法获取；
2. 数据存储应采用加密技术、防篡改机制和定期备份策略；
3. 数据传输应使用安全通信协议（如TLS/SSL）确保信息保密性和完整性。
   * 1. 用户隐私保护

系统应遵循最小化原则，仅采集完成交互功能所必需的数据，并在采集前明确告知用户数据用途、范围与保存期限。

应提供用户自主管理数据的功能，包括数据查询、导出与删除。

应支持匿名化和去标识化处理，降低隐私泄露风险。

* + 1. 身份认证与权限控制

系统应建立多层次身份认证机制，并对不同角色用户实行分级权限管理，防止越权访问。

* + 1. 安全防护与入侵检测

应部署防火墙、入侵检测系统（IDS）和恶意代码防护机制，实时监测系统运行状态，及时发现并响应安全威胁。

* + 1. 合规性与审计

系统应符合相关法律法规及行业标准的要求，建立安全审计与追溯机制，定期开展安全评估与渗透测试，确保持续符合安全与隐私保护规范。

* 1. 性能测试与评价

多模态数字媒体智能交互系统的性能测试与评价应在系统开发完成后及运行过程中定期进行，以验证系统是否满足设计要求，并为优化升级提供依据。

* + 1. 测试原则

性能测试应遵循全面性、客观性、可重复性原则，确保测试结果准确反映系统在实际应用环境中的表现。

* + 1. 测试环境

应在与目标应用场景高度一致的环境中进行测试，包括硬件配置、网络条件、环境噪声、光照条件等，确保测试结果的有效性。

* + 1. 测试指标

性能测试应包括但不限于以下指标：

1. 响应时间：从用户输入到系统输出的时间延迟；
2. 准确率：系统对多模态输入的识别与理解正确率；
3. 同步性：多模态输出在时间上的一致性；
4. 稳定性：系统在长时间运行下的故障率与恢复能力；
5. 资源占用：CPU、内存、网络带宽等资源的使用情况；
6. 可用性评分：基于用户可用性测试的综合得分。
   * 1. 测试方法

可采用自动化测试与人工测试结合的方法：

1. 自动化测试适用于响应时间、同步性、资源占用等指标；
2. 人工测试适用于准确率、可用性评分等需要主观评估的指标。
   * 1. 评价标准

应根据系统设计目标、行业标准及用户需求制定性能评价标准。对于不符合评价标准的指标，应进行针对性优化并复测，直至满足要求。

* + 1. 测试报告

测试完成后，应形成完整的测试报告，包括测试目的、方法、环境、数据、分析结果及改进建议，并作为系统验收与后续优化的重要依据。

* 1. 实施与维护
     1. 实施准备

多模态数字媒体智能交互系统的实施与维护应确保系统在目标场景中稳定运行、持续优化，并具备应对环境变化与技术更新的能力。

在系统部署前，应完成以下准备工作：

1. 根据应用场景进行软硬件环境的配置与验证；
2. 完成系统安装包、驱动程序、数据库等部署文件的完整性校验；
3. 制定实施计划，明确安装步骤、测试环节及验收标准。
   * 1. 系统部署

部署过程应遵循标准化和模块化原则，确保系统组件的正确安装与连接：

1. 对多模态采集设备应进行位置校准与性能检测；
2. 对数据处理与融合模块应进行功能验证与接口调试；
3. 对用户界面与交互逻辑应进行可用性测试与优化。
   * 1. 培训与交接

系统投入使用前，应对运维人员、终端用户进行培训，包括系统操作、常见故障处理、安全管理等内容，并完成相关文档与账户的交接。

* + 1. 日常运行维护

系统运行过程中，应定期进行设备巡检、软件更新、数据备份及安全监测，确保系统性能与安全性。

对硬件部分应开展防尘、防潮、防震等日常维护，对软件部分应及时更新补丁与版本，修复安全漏洞。

* + 1. 故障处理与应急响应

应建立故障处理流程和应急预案，在系统发生异常时快速定位问题、恢复运行，并记录完整的事件报告。

* + 1. 持续优化

应根据性能测试结果、用户反馈及新技术发展，持续优化系统功能、交互体验及安全性能，保持系统的先进性与适用性。

