

多模态交互在非遗美食体验中的应用设计

李怡雯, 刘敬博, 张羽菲, 刘雪

河南工业大学人工智能与大数据学院, 河南郑州, 中国

【摘要】随着数字技术的发展, 非遗美食传承存在形式单一、参与度低、技艺依靠口传身授等问题。本文结合增强现实、虚拟交互和人工智能等技术, 设计了一套 AR 交互系统, 并在用户体验的基础上提出多种模式(包括视觉、重力感应、语音交互)设计方案。通过 Vuforia 实现图像识别的功能, 使用 Unity 搭建虚拟厨房并将传统的烹饪流程转变为可以进行的操作过程, 结合 AI 语音助手以及智能反馈来完成对用户的烹饪指导、文化讲解、操作评价等功能, 使得用户从“观看式学习”到“参与式体验”。研究表明: 多模态交互能有效的减少非遗美食的学习认知障碍问题, 也给传统文化资源的数字化传承提供可行性的途径。

【关键词】多模态交互; 增强现实; 非遗美食; 交互设计; 数字化传承

1. 绪论

非物质文化遗产(以下简称“非遗”), 是某个民族在漫长的历史进程中形成的具有文化记忆的精神符号。非遗美食是最具生活气息和文化温度的品类之一。但目前非遗美食的传承面临许多困境, 当下非遗美食科普过程中, 用户没有机会动手实践, 对文化的认知仅限于表面, 并不能深入理解和产生感情共鸣。随着数字技术的发展, 让非遗文化的保护与传播有了新的方式。增强现实、虚拟现实、人工智能等技术的成熟落地使得沉浸式、交互式文化体验成为了现实。多模态大模型因其在感知、理解、生成等方面的优势, 可以为创新文化遗产叙事、营造沉浸体验提供新的技术路径[1]。AR 技术能够把虚拟的信息叠加到真实的场景上, 让使用者能够在虚实结合的世界里进行互动, 从而达到一种虚实交融的效果, 让用户可以在不离开现实生活的前提下与数字世界中的内容进行自然地交互。有学者认为增强现实技术所具备的沉浸感、想象性和交互性的特点恰好符合非遗不依赖物质载体而存在的活态特征。虚拟现实技术的迅速发展也让虚拟博物馆逐渐成为文化遗产数字化保护与传播的关键载体[2]。

多模态交互理论主张利用人的各种感官(如: 视觉、听觉、触觉、手势、语音等)进行人与系统的自然交互, 在人机交互、教育技术等方面已有比较成熟的研究成果; 但在文化传承方面尤其是非遗美食这种具体的场景中尚未有深入的应用研究, 并且现有的研究成果大多集中在交互技术本身的实现功

能上, 很少从用户的认知规律以及文化的习得机制角度来讨论多模态交互的设计逻辑。本文尝试以非遗美食体验作为切入口, 探索多模态交互应用于文化传承方面的设计范式, 试图回答“如何让技术服务于文化理解”的问题, 丰富了多模态交互理论在文化数字化领域的应用内涵。

本论文基于 AR 技术和虚拟交互及人工智能技术, 设计并搭建一套面向非遗美食体验的交互系统。把传统的烹饪流程变成可视化的、可以操作的过程, 让用户由“观看式的学习”变为“参与式的体验”。

2. 多模态交互理论

2.1 多模态交互的概念

“模态”(Modality)指的是人们感知并表达信息的不同通道或者方式, 包括: 视觉、听觉、触觉、手势、语音、姿态等等; 多模态交互(Multimodal Interaction)就是指一个系统能使用两种以上感知通道来跟人进行信息交换和互动。相比于传统单一的交互模式比如只用键盘鼠标或是触屏点击等方式, 多模态交互更加贴近人在自然状态下的与别人交流、与环境互动的状态。

2.2 AR 环境下的多模态交互优势

AR 的虚实融合特性与多模态交互的自然契合:

增强现实(AR)不同于虚拟现实(VR)的地方就在于它不是把用户封闭在一个完全虚拟的世界里, 在真实世界的基础之上加上了虚拟的信息, 让用户既能看到现实的物理空间也能感受到数字生成的东西。这使得 AR 天生就适合那些需要“在真实环境中完成具

体操作”的场景，而非遗美食的制作正是如此——用户需要在真实的物理空间中进行操作的学习，并且又需要虚拟信息的引导与辅助。

AR 环境是实现多模态交互的最佳场所，在 AR 环境中，虚拟的信息可以准确地锚定到现实空间中的某个点上（比如灶台上面、食材旁边），用户的目光、手势、身体移动都会和虚拟的内容产生空间上的对应关系，这样就可以让视觉、动作、认知的各种模态共同协作，提供一个统一的交互体验。

AR 场景中视觉、听觉、动作模态的协同机制：

视觉模态负责场景呈现以及文化信息传递的功能，在用户眼前由 AR 技术生成了 3D 虚拟厨房、悬浮食材库、文化图谱等等视觉内容；视觉信息除了包含场景之外还包括步骤高亮提示、操作结果反馈、文化知识卡片等等。优点在于信息量大，并且呈现比较直观，缺点则是在于如果信息过载的话会迅速地占据用户的认知资源。

听觉模态负责讲解引导及即时问答功能，伴身虚拟精灵以语音的形式进行操作提示、文化讲解与情感反馈，并且用户也可以通过语音提问获得个性化的回答。听觉模态的优点是不会占用视觉资源和双手，适合在用户专注于操作的时候给予辅助性信息，是对视觉模态的良好补充[3]。

动作模态则是用来进行操作执行以及具身体验的功能，用户通过触控拖拽、重力感应倾斜、旋钮旋转等的方式来进行烹饪过程中各种操作。动作模态可以把抽象的烹饪参数如克数、火候、时间转化为身体能感受到的动作。“理解”不再是单纯的发生在认知层面上的理解，也是发生在身体上的理解。

三种模态不是简单地相加，它们都是为了达成“完成烹饪流程、理解文化内涵”的目的而分工合作——视觉“告诉”，听觉“解释”，动作“实现”。

2.3 非遗烹饪场景的交互特殊性

非遗美食的制作过程具有鲜明的特殊性，这些特殊性对交互设计提出了不同于一般文化展示场景的要求。

烹饪过程中双手被占用的特殊性：

烹饪是一个典型的“双手作业”活动——切菜、搅拌、翻炒、调味都需要用到双手，在传统的数字产品中，用户需要用手指来操控手机或者平板，这就和烹饪的实操产生了矛盾，用户要么放下手中的“虚拟操作”去做“真

实的操作”，要么因为拿着手机/平板不能做真实的动作模拟。这个矛盾就决定了非遗美食的交互设计要考虑到“手”的解放的问题，通过语音交互的方式可以很好的解决这个问题，即用户可以通过语音指令完成步骤查询、文化讲解、视频生成等等操作而不必打断手中操作。

火候与用量难以量化的具身知识特性：

非遗美食的关键技艺往往是“只可意会、不可言传”的。“大火”“小火”“盐适量”“糖少许”这些用词在不同的地域和菜系里有着完全不同的意思，这样的知识本文称之为“具身知识”（embodied knowledge）——它们存在于身体的感觉和经验之中，并非存在于文字或者数字当中。

而具身知识很难被文字或语言所描述并传达出来，只能在操作实践中让学习者自己去“体会”。本研究所提到的重力感的应用量调节、旋钮式的火候控制等等都是把抽象的“适量”、“适当”转化成可以被身体感受的倾斜的角度以及旋转的幅度，让用户在反复地操作过程中形成自己的“手感”。

非物质文化遗产中的美食不仅仅只是做一道菜的方法，它包括了它的地理起源、历史演变、民俗仪式、节令文化、器具演变等等丰富的、深刻的内涵与文化信息。所以，如果只用一种方式来介绍和传播这些文化信息是不够全面的。比如仅仅依靠文字或者仅仅是视觉上的展示都不能完整的传达出非遗美食背后所蕴含的文化意义。

而采用多种渠道进行信息输入则可以解决这一问题。比如，在视觉上展示食物的外观以及工具的形状、步骤的画面；在听觉上讲述食物的起源故事、民俗背景及文化寓意；在动作上让使用者在动手的过程中感受到工艺的精妙与不易[1]。不同的通道承载了不同类型的信息，从而建立起对于非遗美食的认知。多通道的协同输入也符合人类认知的“多媒体原则”，即当人们看到各种形式的信息的时候，他们就容易把它们之间建立起来的知识关系，并且记忆更加长久。

3. 非遗美食 AR 体验需求分析

3.1 传统非遗美食传承与体验的局限性

以抖音为代表的各种数字化平台能够很快地扩大非遗内容的传播范围，不过这种优势首先解决的是“看得到”的问题，在烹饪技艺上还需要学习者去处理食材、调整火候，并且根据变化做出判断，仅仅通过观看是不

能够检验自己的操作的；而传统意义上的口传身授虽可结合学习者的实际操作来进行示范和纠正，但是由于传承人的时间和场地以及材料等条件所限，当学习者离开现场之后就不可能再对关键的动作反复观看，后续的学习过程中也不一定还能继续得到纠正。有学者认为数字技术可以支持非遗信息的保存、复原和传播，具身交互可以帮助把动作、感知与文化情境联系起来[4]。宣威火腿为例，其品质特征的形成同时受到原料、加工工艺与环境条件等因素影响[5]，仅呈现操作顺序难以完整说明这类经验知识的形成机制。所以对于非遗美食而言其数字化体验应该从单向展示转变为“示范——操作——反馈”这一连续的过程。

3.2 目标用户与需求分析

本文假设使用的场景有：景区体验、研学活动和亲子互动，在这些活动中参与者的

表 1.目标用户及需求分析

用户类型	典型场景	主要需求	设计响应
年轻游客	景区、文化街区、展馆短时体验	新颖、易上手、可记录分享	短任务引导；简化菜单；成果卡片
研学学生	课程教学、研学与劳动教育	文化知识、操作练习、重复学习	分步任务；过程纠错；学习复盘
亲子家庭	场馆亲子活动、家庭共同体验	轻量操作、共同参与、内容易理解	角色协作；语音讲解；合影分享
文旅场馆	展馆、景区和文化体验空间	低成本、稳定、内容易更新	移动端部署；配置化内容；维护接口

3.3 AR 多模态技术的适配性分析

从场馆部署实际来看，优先使用游客自带手机要比增加 AR 眼镜更容易实现且所需的输入设备已经包含在手机之中，在本项目中，摄像头负责卡片和平面识别，触摸屏进行食材的选择和参数调整，麦克风接受语音，而倾斜的动作则是通过惯性传感器来读取。已有研究证明 AR 可以用来把虚拟的信息叠加到现实场景上并用它来进行非遗以及文旅资源的交互展示[6]，同时也有学者对增强现实应用于博物馆进行了探索，证实了增强现实的空间定位、信息叠加和交互呈现是可行的。因此本项目的方案是以手机为主载体，以食品卡片识别和平面检测作为 AR 厨房的场景入口。

4.多模态交互系统设计

4.1 设计目标与原则

系统的设计目标在于把非遗美食的学习由单向的观看转变为一个“识别——操作——反馈——理解”连续的过程：用户在 AR 厨房里进行食材的选择、调料的添加、火候的调节，系统会根据当前工序返回结果，在出现跳步或者对象选择错误以及参数超出范围的时候给出提示，让用户的制作行为与工

进入系统的目的是不完全相同的，比如年轻的游客可能会只停留在很短时间内，而研学的学生则需要完成他们的学习任务，亲子的家庭更注重的是共同的操作等等，对于负责日常运营的文旅场馆来说，设备的成本、稳定性以及游客的停留情况也同样重要。因此本文暂分为年轻游客、研学学生、亲子家庭和文旅场馆 4 种类型来确定原型首先要覆盖的功能。

从表 1 可以看出，各类对象在体验时间、学习深度和成果形式上有所区别，但都希望操作不难理解、反馈足够清楚，文化信息也经得起核对。对系统而言，这些共同要求能够落实为三个问题：用户能否迅速看懂当前要做什么，操作结果能否解释相应的工艺目的，以及场馆能否在不过度依赖专用设备的情况下更新不同菜品内容。

艺解释在一个场景内产生关联。

设计上遵循三个原则：低认知负荷（界面只显示当前步骤所需要信息和控件）、文化准确性（工艺名称、步骤关系和文化条目的表述以审核的内容为准，大语言模型只调整表达方式）和具身参与（拖拽、旋钮和手机倾斜的动作尽量贴近日常的操作）。具身交互的研究表明，动作和感知可以参与到对非遗知识的理解过程中去[4]。

4.2 总体架构

系统的组成包括移动端 Unity 客户端、智能服务中台和可配置内容库三部分。其中客户端实现图像识别、平面检测、三维渲染、触控、姿态读取等功能；服务中台实现语音识别、知识检索、语言生成、语音合成和过程评价功能；内容库统一管理菜谱节点、文化条目、模型素材和评价规则。相关的研究从场域架构、关键技术与治理要求等方面讨论了数字非遗系统建设[7]，本文在此基础上进一步明确了端侧、服务与内容之间各自的职责。

如图 1 所示，在端侧内部按感知层、交互层、应用层和内容层进行组织。首先把感知的数据转化为一个统一的事件，并且由流

程的状态机来核对当前步骤和前置条件，经过验证之后才会更新 AR 场景以及写入过程记录。服务的能力是通过接口按需要调用，

菜品的不同是由内容包配置实现，减少了在增加新菜品的时候修改核心程序的工作量。

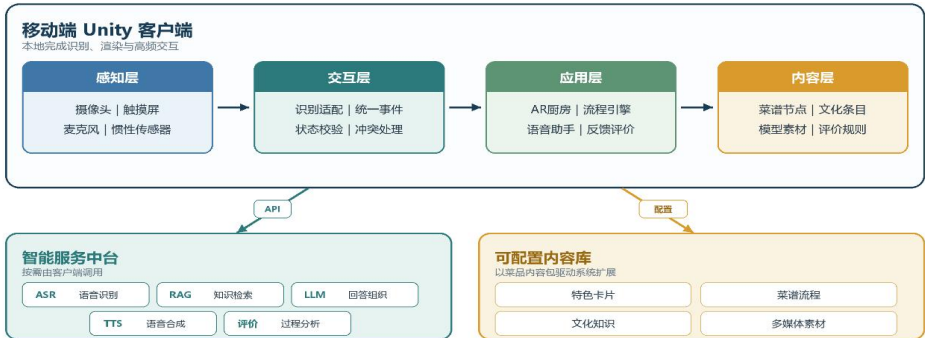


图 1.系统总体架构

4.3 视觉交互设计

视觉交互采取了“卡片确定内容、平面承载场景”这一条路子。当使用 Vuforia 识别到一张食品卡片以后，客户端会加载相应的菜品内容；然后利用 ARKit 或者 ARCore 检测出稳定的平面并且建立起空间锚点，接着就生成灶台、备菜区、器具和食材等。卡片暂时离开了镜头并不会让厨房消失，只有定位一直丢失的情况下才会重新选择平面。界面按照任务阶段分配信息，比如备菜时显示食材库和目标容器，加热时开放火候和计时控件，文化的内容则是随着工序以“起源——器具——流程——故事”的层级逐步展开。已经有研究者尝试运用 AR 用于非遗及文旅资源的空间化的交互展示[6]。所以本文的系统并不长久地堆叠标签，而是只会在相关的节点上显示必要的提示。图 2 中展示了 AR 厨房主要界面区域以及视觉、触控、姿态和语音入口。



图 2. AR 厨房交互界面原型

4.4 触控交互设计

触控交互用于食材的拖拽、火候调整和时间设置。当用户释放食材之后，系统会核对工序顺序、目标容器以及前置步骤是否正确，如果都符合则进行位置更新，不符合就

让食材回到原来的位置并且给出原因说明。火候使用文火、中火、旺火等旋钮档位，计时器是“选择——确认”的两个步骤。拖拽、旋钮和滑动都会记录开始、变化、结束和确认的状态，动画和声音只反馈已经生效的操作。

4.5 重力感应交互设计

重力感被用来模拟倒油、倒盐的动作，在进入倒料节点以后，客户端以当前握持的姿态作为参考，连续读取加速度计和陀螺仪的数据，只要倾斜保持稳定就开始增加虚拟用量，手机回正的时候停止。不同的倾斜幅度代表了不同的流速档位，并且用流线动画、用量刻度和振动来提示结果。这个功能只是表达了虚拟用量的变化而没有测定真实的重量，如果不稳定可以切换成触控滑块。

4.6 AI 语音交互设计

语音链路有：ASR 语音识别、内容检索、LLM 回答组织和 TTS 语音合成。根据当前菜品、工序和场景状态去检索审核的内容然后生成语音的回答和文字摘要。文化遗产的知识服务需要以可信数据为基础[8]，所以火候、时间和用量和步骤顺序都是以内容库记录为准，如果没有足够的资料就不生成具体的参数。

儿童、游客和研学模式共享同一个事实条目，只是改变信息长度、语言难度和讲解重点，“下一步”、“重复讲解”、“暂停计时”和“返回”，这些高频口令由客户端白名单处理，在场馆噪声导致识别失败的情况下在同一位置显示触控按钮，避免语音是唯一的操作入口。

4.7 智能反馈机制

反馈分为即时纠错和结束复盘：提前下料、容器错误、遗漏步骤或者参数越界，在

当前节点被拦截并且说明原因；停留在某个环节时间过长的时候只给出操作提示，误触和重复的动作会留到任务结束后进行汇总。判断条件根据菜品的内容包来设置，避免使用一个规则代替不同的技艺要求。

复盘从顺序、用量、火候、完成度和规范度 5 个方面展示过程证据，并标出需要重新练习的节点。文化解释与具体的操作绑定，按照“工艺用意——文化内涵——环境成因”的顺序展开，使得纠错的结果能够同时回答了“哪里不符合要求”以及“为什么要这样做。”

4.8 多模态融合协同机制

多模态协同依据任务的状态开放输入的方式，不是所有的模态都参与到每一个步骤



流程状态机持续校验前置条件 · 有效操作写入过程日志 · 识别失败时切换触控备用方式

图 3. 多模态协同体验流程

5. 系统实现

基于上一章节对于多模态交互的设计框架论述，在这一部分中，本文从开发环境的选择、识别模块搭建、核心交互功能和用户的界面呈现四个方面来介绍“与‘食’俱进——舌尖文脉 AR 交互平台”系统的具体实现过程。

5.1 开发环境和技术选型

本系统是围绕着非遗美食数字化体验的需求而展开的，是一个以增强现实技术和多模态交互技术以及人工智能技术为核心的“识别——呈现——交互——反馈”沉浸式的体验流程。整个系统的开发中使用了 Unity 为主要开发工具并辅之于 AR 引擎、语音识别接口和大语言模型的服务等来实现虚拟厨房场景构建、用户行为感知以及智能化交互反馈。

在开发环境方面选择了 Unity 来做客户端的开发，Unity 可以用来移动端 AR 应用开发并且可以方便地完成模型加载、动画控制以及交互逻辑编写等功能。相比于传统的应用开发方式而言，Unity 可以减少 AR 场景开发难度，提升虚拟场景与真实环境的融合程度[9]。

在增强现实识别上使用 Vuforia 对食品卡片的数据集进行识别，通过扫描非遗美食

中去。场景建立主要依靠视觉识别，食材和火候操作以触控为主，倒料节点读取姿态，知识查询由语音发起。系统首先检查工序条件，然后处理触控、语音或是传感器事件；当不同模态发生冲突时保持原状态并等待确认。

完整的体验流程如下图 3 所示，在扫描卡片产生 AR 厨房之后进行食材的拖拽、倾斜倒料以及火候调节，并且可以随时发出语音查询，对错误的操作可以在当前节点纠正，最后得到一个多维复盘。有效的操作会一直记录到过程日志中，当姿态或者语音识别失败的时候则使用触控作为备用方式，以保证任务能够继续完成。

卡片来匹配特征然后加载对应的虚拟厨房与文化内容；以特色卡片作为 AR 入口，触发 3D 厨房、文化图谱与烹饪流程，实体与数字相融合。

在空间定位与场景融合上，基于 ARKit 和 ARCore 利用平面检测和平面锚点在真实的桌面上稳定叠加虚拟厨房模型，可对用户进行食材的选择、烹饪的操作等实时交互。

在智能交互层中采用的是讯飞语音 API 和 DeepSeek 大模型结合的方式，讯飞负责语音识别 ASR 和语音合成 TTS 实现语音的输入和语音播报的功能；DeepSeek 负责自然语言的理解和生成，回答有关食材、烹饪、地域文化的问题，语音与大模型链路解耦形成一个完整的“听-理解-答”闭环。

5.2 识别模块与场景加载实现

系统识别模块是连接用户现实操作和虚拟内容的入口，本项目的识别方式选择的是图像的目标检测方案，以非遗美食实体卡片作为识别媒介，依靠特征匹配来唤醒数字内容。

具体做法：首先用 Vuforia Target Manager 解析出卡片纹理、边缘、特征点，并生成一个识别数据库，然后在移动端打开摄像头拍摄的画面与库内的特征进行比对，当匹配成功时，通过 Unity 识别接口读取食

品 ID 加载对应的数字资源。

扩展性的考虑：“卡片 ID——内容资源包”的映射关系，每一款非遗美食都单独配备一套历史、制作流程、3D 模型、文化故事等内容模块，降低了后期的内容迭代的成本，可以很快地增加全国各地的新美食资源。

AR 场景搭建：借助 ARKit/ARCore 的空间定位的能力，把虚拟厨房环境融入到现实空间当中去[10]，生成了虚拟厨房、食材区以及交互界面，通过食材拖拽、火候调节模拟真实的烹饪过程。

性能方面的考量：对三维模型减面、纹理压缩、按需加载，在不影响视觉效果的情况下减轻移动端的算力负担，提高 AR 场景运行的流畅性和稳定性。

5.3 核心交互功能实现

系统围绕着“重力感应调节、旋钮/拖拽操控、AI 智能反馈”这三种主要交互，分别实现了对应的交互技术方案，形成了一个多感官协同、多模态融合的全场景操作体系。

重力感应 API 调用和映射逻辑。调用移动设备的加速度传感器与陀螺仪 API (iOS 的 Core Motion 框架，Android SensorManager)，实时获取三轴倾斜、加速度数据；HTML5 和原生应用中重力感应模块是移动端各种交互得以实现的基础。“倒盐”交互方式就是通过倾斜手机来模拟倒盐的动作，不断监听俯仰、翻滚的姿态角，倾斜幅度、时长对应食盐投放量；重力感应也用来做“熬煮调温”，设备倾斜的方向和角度被映射到锅温的升高/降低，并在 UI 上以温度指针的实时摆动予以反馈。

旋钮/拖拽的事件监听与状态机驱动。旋钮控件以及拖拽交互都是基于 Unity 的 UGUI 事件系统：旋钮利用 IDragHandler 监听旋转拖拽，并把旋转的角度转化成火候和计时数值；食材拖拽则使用 IBeginDragHandler 和 IEndDragHandler 监听起止，根据坐标判断食材是否放入锅内。整个烹饪交互是由一个有限的状态机来驱动，在每一道菜品中都单独设置好「备菜—下油—入料—调味—出锅」流程节点，每个节点里存放着可以进行的操作、需要使用的食材、标准参数。状态机将繁琐复杂的流程逻辑分离开来，简化了代码也方便迭代。当用户的操作不符合流程的时候会触发状态跳转拦截，弹出习俗提示并且记录下该次操作的日志以供大模型后期复盘分析。

AI 接口封装和异步调用。DeepSeek 大模型、讯飞语音 API 都是独立的服务化，异步调用不会阻塞主线程。语音交互流程：用户语音经过 ASR 转换成文本，带上卡片 ID、烹饪节点、操作记录等上下文信息异步请求大模型，返回文本再由 TTS 语音播报出来同时显示字幕；AI 复盘是对操作时序、火候、用料的数据做综合评估得到一个多维度的评估向量，然后由大模型给出个性化的优点和不足总结。另外还搭配了超时重试、本地缓存来保证在弱网环境下核心功能依然可用。

5.4 用户界面和反馈呈现

系统的界面主要遵循两个原则：一是“沉浸式的厨房场景”，二是“轻量化的信息反馈”，下面按照功能模块进行介绍。

主界面、识别引导。应用打开之后会进入到一个相机预览界面，在中间位置有一个识别框以及提示文字：“请扫描非遗食品卡片。”识别框是一个半透明的环状动效来引导用户对准卡片，并且在最下方有“切换模式”的按钮（日常菜谱/自由创作），还有“语音助手”的快捷入口。

3D 虚拟厨房界面。识别到卡片之后，在画面之上叠加一个 3D 的虚拟厨房（包括灶台、备菜区、操作引导台），并在界面左上角显示菜品以及非遗产地，并且右上角有浮动的 3D 虚拟精灵随着待机、讲解、纠错的状态而进行切换动画。可以拖拽食材、旋转旋钮、倾斜设备来完成烹饪过程，所有的操作都配有实时视听反馈。

2D 悬浮的文化图谱。文化图谱是悬浮在厨房场景一边以 World Space Canvas 的形式出现的图文内容，分为“起源-器具-流程-故事”的 4 个 Tab 层级的内容，可以通过点击 Tab 来层层打开对应的文字图片信息，在图谱边缘带有一个辉光的效果和粒子流动的动效来增加用户的沉浸感。

AR 悬浮弹幕卡片和复盘清单。烹饪过程时弹出带有浮动动画的半透明 AR 弹幕，对操作进行实时点评，并且从工艺、文化、环境三方面科普；违规则会弹窗提醒纠正。烹饪结束之后弹出一个 2D 复盘面板，从操作顺序、用料、火候、工艺四个方面的评分并给出优化建议，可以本地存档、生成分享短视频。

语音交互界面。语音助手会以悬浮按钮的形式一直出现在界面右下方，在用户点击按钮之后开始聆听则会在界面底端显示出波

形动效和“请说话”字样，然后语音会被转化成文字并且展示在对话气泡里，而大模型生成的回答将会分文字+语音两个渠道进行输出，并且虚拟精灵也会做出相应的“讲解”的手势、口型等动画。同时对话历史将以一个比较紧凑的聊天列表形式收折在界面左边，可以随时展开查看。

6. 结语

部分以非遗美食传播所遇到的参与感低及隐性知识流失现象为出发点，在多模态交互与具身认知两种理论的指引下，对移动终端可用的AR交互系统作整体性的流程设计与开发；将非遗烹饪活动涉及的交互特征加以归纳，提出四层结构的技术方案，使触控操作、手机姿态识别、AI语音问答三者能同步运行，同时引入实时纠偏及多维判断类的反馈机制。经由用户实际操作的测试可知，新系统可令学习变得较具沉浸感、动作记忆较快、文化理解较深，因而所采用的多通道自然交互法在饮食类非遗从单纯存档走向实际传承方面是可行的、有效的。但目前研究仍有若干局限性：实验中用到的被试人数不多且都属于同地域群体、系统没有为头戴式装置预留接口、菜品信息不完全、有关姿态的识别算法尚不能精确处理高速动作等。接下来要做的工作是把AR的功能迁移到眼镜上（即解除手部占用），运用检索增强生成思路自动补充菜谱或知识内容，以多语言形式制作解说模块协助中华美食走向全球各地。

参考文献

[1] 魏立才.叙事、认同、沉浸：多模态大模

型赋能新时期文化遗产保护与传承的推进策略[J].云南民族大学学报（哲学社会科学版），2025，42（1）：31-39.

- [2] 杨雯晶，王俞璇.多模态视角下虚拟非遗博物馆的设计与实现[J].现代信息科技，2026，10（5）：95-100.3
- [3] 黄永林，谈国新.中国非物质文化遗产数字化保护与开发研究[J].华中师范大学学报（人文社会科学版），2012，51（2）：49-55.
- [4] 危田媛.具身交互设计在非物质文化遗产数字化传承中的应用[J].上海包装，2025（12）：58-60.
- [5] 乔发东，马长伟.传统宣威火腿品质特征及成因分析[J].食品科学，2004，25（8）：55-61.
- [6] 魏三强，王兵.AR技术在非物质文化遗产及文化旅游业中的应用[J].江汉大学学报（自然科学版），2016，44（4）：364-368.
- [7] 刘中华，焦基鹏.元宇宙赋能数字非遗的场域架构、关键技术与伦理反思[J].浙江大学学报（人文社会科学版），2023，53（1）：19-29.
- [8] 范炜，曾蕾.AI新时代面向文化遗产活化利用的智慧数据生成路径探析[J].中国图书馆学报，2024，50（2）：4-29.
- [9] 朱惠娟.基于Unity3D的虚拟漫游系统[J].计算机系统应用，2012，21（10）：36-39+65.
- [10] 张帅.分布式移动网络环境下大规模空间社交媒体数据快速增强实现可视化研究[D].南京大学2020. DOI:10.27235/d.cnki.gnjj.2020.002323.