

基于多模态信息交互的智能家居设计研究

马晓娜¹, 张雨欣¹, 于茜²

(北京航空航天大学 a.新媒体艺术与设计学院 b.软件学院, 北京 100191)

摘要: **目的** 针对界面、语音等单一智能家居交互模式阻碍用户自然体验的瓶颈问题, 探索基于多模态信息交互的智能家居设计方法, 实现智能家居更自然、友好的用户体验。 **方法** 基于用户日常生活多模态的自然交互特点, 探索一种以多模态为特征的智能家居信息交互模式, 构建了面向智能家居的多模态信息交互设计框架, 并对该框架中的构成要素、交互流程、模态组织方式及多模态信息的注意度、模态密度、模态结构配置进行分析讨论。 **结果** 智能家居多模态信息交互实现了智能家居系统的主动性、感知化信息获取和情境化的信息分析与反馈, 对提升智能系统自主性, 降低用户在交互过程中的认知负荷, 以及实现更加自然、友好的用户体验有积极价值。 **结论** 智能家居环境需要改变传统相对单一的人机系统信息交互方式, 以多模态信息交互促进智能家居使用体验的最优化。

关键词: 智能家居; 多模态; 信息交互; 交互设计

中图分类号: TB472 **文献标识码:** A **文章编号:** 1001-3563(2022)16-0059-09

DOI: 10.19554/j.cnki.1001-3563.2022.16.006

Smart Home Design Based on Multimodal Information Interaction

MA Xiao-na¹, ZHANG Yu-xin¹, YU Qian²

(a.School of New Media Art and Design b. School of Software, Beihang University, Beijing 100191, China)

ABSTRACT: Aiming at the problem that a single smart home interaction mode based on interface and voice hinders the user's natural experience, this paper explores a smart home design method based on multimodal information interaction to achieve a more natural and friendly user experiences of smart homes. Based on the multi-modal natural interaction characteristics of user daily life, this paper explores a smart home information interaction mode characterized by multimodality, constructs a multimodal information interaction design framework for smart home, and analyzes the elements, interaction process, modal organization, attention of multimodal information, modal density, modal structure configuration, etc. The multimodal information interaction of smart home realizes the initiative, sympathetic information acquisition and contextualized information analysis and feedback of the smart home system, which can improve the autonomy of the intelligent system, reduce the cognitive load of the user interaction process, and achieve a more natural, friendly user experience. In general, the smart home environment needs to change the traditional single human-machine system information interaction mode, and promote the optimization of the smart home use experience with multimodal information interaction.

KEY WORDS: smart home; multimodal; information interaction; interaction design

20 世纪 90 年代初, 美国联合科技公司将建筑设备信息化、整合化概念应用于首栋“智能型建筑”, 由此揭开了智能家居设计的序幕。以家居环境为平

台, 利用控制技术、自动化技术和信息交互技术的智能家居成为当今家居生活的发展趋势。

用户、环境与智能家居之间的有效信息交互是智

收稿日期: 2022-03-08

基金项目: 北京航空航天大学 2022 年度人文社科青年骨干支持项目

作者简介: 马晓娜 (1985—), 女, 博士, 助理教授, 主要研究方向为交叉学科创新设计。

通信作者: 于茜 (1991—), 女, 博士, 副研究员, 主要研究方向为计算机视觉。

能家居设计的关键。智能家居通过相应信息采集设备（如各类传感器、摄像头等）获取用户与环境信息，实时控制反馈以满足用户需求和应对环境改变。当前的智能家居主要依赖用户通过触控界面、语音或者手势等单一控制模式的主动输出、系统被动接收的方式获取指令。这种交互方式限制了用户在人机交互中的自然性需求，同时局限了智能家居系统在智能性上的潜力。因此，改变当前智能家居被动接收用户单一模态信息输入的旧模式，探索能够主动获取、加工及反馈多模态信息的新模式，将为创造更加自然的智能家居交互体验提供可能。

1 多模态信息交互

1.1 多模态研究

现代脑科学研究中以“模态”通指感官及其相应的神经系统^[1]，感官模态包括视觉（眼睛）、听觉（耳朵）、嗅觉（鼻子）、味觉（舌头）、触觉（皮肤）等“五觉”，以及其他如体觉、痛觉、痒觉、温觉、内脏觉、平衡觉等^[2]。“模态”被认为是人类通过以上感官系统跟外部环境（人、机器、物件、动物等）之间的互动方式，上述“五觉”分别对应视觉模态、听觉模态、触觉模态、嗅觉模态和味觉模态等互动模态^[3]。用单个感官进行互动叫单模态，用2个感官进行互动叫双模态，用3个及以上叫多模态。一种感官对应一种模态，以视觉为例，眼睛和处理视觉信号的神经系统组成一个视觉模态^[4]。相应的，多模态是指多个感觉器官和处理各自信号的神经系统。

从信息交互的角度看，模态是交流的渠道和媒介。社会符号学派认为模态是在社会文化中形成的创造意义的符号资源，是一种可被感知和阐释的符号系统，通常包括语言、图像、声音、空间和身体动作等不同形式^[5]，主要以符号系统的数量来界定多模态话语。在信息传递中，混合使用声音、图像、颜色、动作等不同形式的符号资源，这种多元意义呈现方式被称为多模态^[4]。语言学家 Norris^[6]指出，人类互动具有多模态性。人际互动通过人的语言、表情、身体姿势、动作及与人相关的空间距离、场景布局、音乐、物体等多元化模态信息的协同使用来进行。

不同视角下，多模态研究的对象与方法各有不同，但其理论假设较为一致：一是多模态资源协同运作以表达意义和完成动作；二是每种模态资源或符号系统都有自己的组织系统。这为多模态人机交互的研究提供了基础。

1.2 人机交互的多模态转向

话语意义是由多种模态构成、传递和解释的，因此，人类交际中使用一种模态会同时触动使用另外一种或几种模态，例如使用口语交流时，通常还涉及注

视、手势、身体姿势、身体距离等。人机交互也是基于相同的思维方式，用户向他人（包括计算机）发出信息，准确的话语意义往往是通过多种模态构成和传递的，计算机系统需要能够对用户的多模态信息分解识别，并通过融合分析进行用户意图的准确把握。

单一模态的交互方式信息通量低，交互过程死板，难以满足用户自然交互的需求。增加人机互动模态是目前人机交互的主要研究内容之一。人机交互过程是一种多模态信息耦合交互的过程，涉及文字、图像、语音、动作等多种交互信息^[7]。多模态人机交互就是人们使用多种感官模态，通过多种物理媒介，与计算机等机器进行多渠道、多形式的信息交互。例如仿真机器人让机器像人一样通过5种感官模态获取信息，以人的认知逻辑进行信息采集^[8]。对智能系统来说，多模态人机交互的根本是依靠智能系统类人脑的多模态功能，在多种智能感官系统的支撑下，通过多模态协同机制实现多元信息的采集、分析、理解与意义的构建。

2 智能家居信息交互模式发展

2.1 智能家居信息交互传统模式

智能家居系统的感知与人的感官感知遵循相同的逻辑，给机器配上感官系统（即信息采集设备）从而模拟人的视觉、听觉、触觉等功能，通过手势识别、语音识别、触控界面等途径进行信息获取。

2.1.1 智能系统与用户交互

目前，智能家居系统与用户的信息交互中，主要依赖用户发出指令。用户基于智能系统的信息采集方式选择单一模态的输入（如语音或者手势等）向智能系统发出指令，智能系统通过麦克风、摄像头和不同类型的传感器被动地接收用户指令，进行分析与反馈。

以语音辅助界面交互为例，基于语音交互的智能系统包括3个部分，即输入单元、处理单元和输出单元，见图1。第1部分，即智能交互输入单元，当用户发出语音指令，输入单元中的电子收音设备采集到语音信息，该信息将在语音模块中转化成文本信息。如果用户同步通过软键盘输入信息，输入单元的电子触摸模块通过其传感器会采集相关文本信息；第2部分，智能交互处理单元的人工智能模块接收到文本信息进行分析，根据原有数据库合成反馈信息；第3部分，智能交互输出单元中的语音合成模块接收到处理单元输出的反馈信息，将文本数据转化为语音数据，再通过音频输出设备向用户播放。

智能语音助手是智能家居最直接的信息交互载体之一。亚马逊的Echo系列是家庭智能语音助手典型产品。在新一代产品升级中，亚马逊在语音交互之外增加了触摸屏幕与摄像头，成为一个加入了图像识

别(摄像头)和触控交互(触摸屏)的智能语音助手, 让用户拥有了自由度更高的交互选择, 亚马逊 Echo 多模态信息交互见图 2。

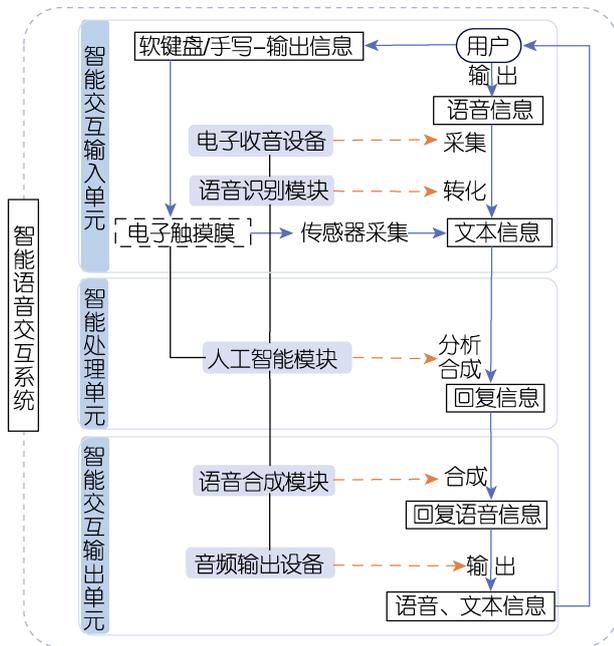


图 1 智能系统与用户语音模态信息(辅以界面)交互流程
Fig.1 The interaction process between smart home and user voice modal information (supplemented by interface)



图 2 亚马逊 Echo 系列家庭智能语音助手多模态信息交互
Fig.2 Amazon Echo series home intelligent voice assistant multimodal information interaction

人脸识别也是目前智能家居中信息交互的重要模态, 人脸识别基于专用摄像机进行主动探测, 基于

动态视频流识别用户身份, 不需要用户主动配合, 能快速精准地识别人脸信息。

随着互动技术的发展, 通过身体交互实现姿态智能控制, 逐渐成为智能家居系统与用户的主要交互方式。手势的手型、运动、位置、方位等协同言语和其他身体行为在日常会话互动中有效传递信息, 表达意义。Alanwar 等^[9]面向智能家居环境, 提出手势控制智能手表 SeleCon, 该设备使用惯性传感器进行指向手势检测, 可以实现 84.5% 的设备选择准确率和 97% 的手势识别准确率。此外, 通过脑电、眼动等信号读取用户意图成为一种重要的智能系统信息交互手段。Alrajhi 等^[10]开发基于脑机接口 (BCI) 的智能家居系统, 通过 Emotiv EPOC+ 检测用户的大脑信号, 让四肢瘫痪的人仅使用大脑信号打开/关闭门, 以减少对护理人员的需求。

2.1.2 智能系统与环境交互

目前, 智能家居系统通过对环境中人的语音、体态、人脸信息, 以及环境中的温度、湿度、光照度、压力等信息的主动采集与分析进行自动化反馈。胡旭央等^[11]研究智能家居场景中的睡眠场景, 认为对智能家居场景中的光线、温度、湿度、气味等氛围要素, 空间布局、规模、功能等空间要素, 用户社交、服务关系等社交要素, 以及数据的输入、输出等数据要素的关联设计十分重要, 智能家居场景需要关注以上各方面的影响与相互作用, 才能为用户提供良好的体验。

智能家居系统通过各类传感器和执行器监测、采集环境信息, 通过分析环境数据的变化实现特定功能的控制。例如通过温湿度传感器、烟雾传感器、光照传感器等感知模块进行数据信息自动采集。在基于 AVR 的智能家居系统中, 当温湿度传感器采集的温度超过了预设的阈值后, 系统将向执行层的空调发送打开空调的命令以启动空调; 当烟雾传感器探测发现烟雾环境异常时, 系统将向通信模块发出报警信息。系统的主控层对传感层上传的数据进行分析及处理, 然后向执行层发送控制指令。各模块的主要流程图 3^[12]。

2.2 智能家居自主交互与多模态交互需求

2.2.1 智能家居自主交互

家庭环境有特殊的信息交互需求, 用户希望在家庭环境中获得舒适、自然、安逸的生活状态, 这需要智能家居设计中强化智能系统对用户与环境信息的主动获取, 以提供更自然、更友好的人机交互体验。

Rose 等^[13]提出物体具有“感知的自主性”, 这是一种“部分来自人们认为机器拥有和做的东西, 部分来自被设计成机器的东西的属性”。Karthik 等^[14]优化智能系统任务分配方法, 提出了新的任务分配技术, 智能系统主动寻求任务分配来降低用户负担, 实现人机平等的伙伴型无缝协作。Philipp^[15]等将光纤传感器

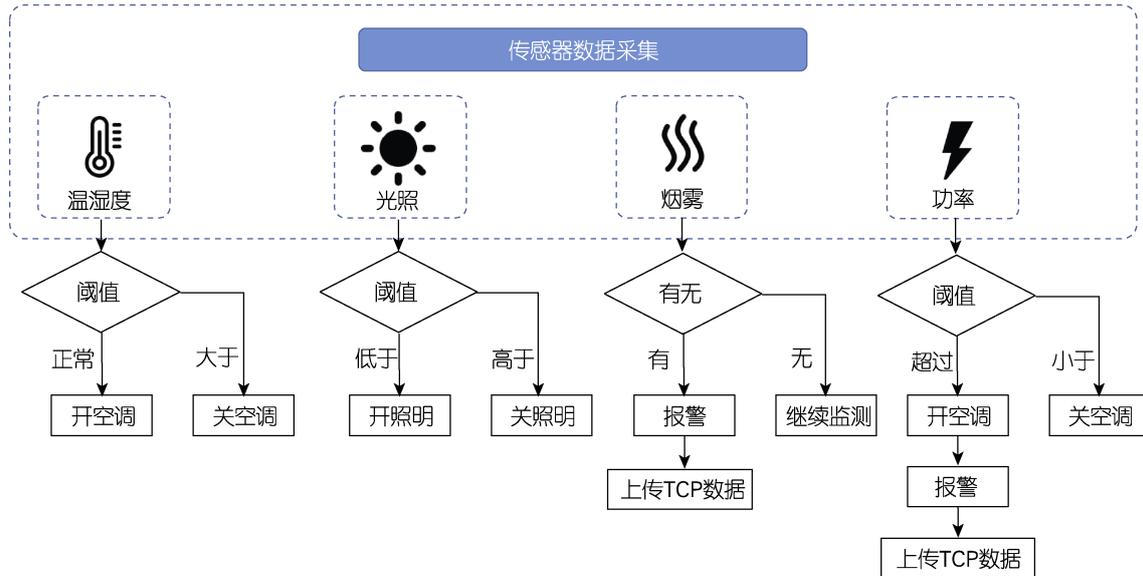


图3 智能家居系统对环境信息的数据采集流程

Fig.3 Smart home system environmental information data collection process

及压电传感器集成到家庭环境的木地板中,通过室内定位、步数检测、步态分析、运动分析等监测住户位置及步数信息,让用户在自然行走中无意识输出信息,紧急事件发生时系统将自动启动救援程序。Wu^[16]等提出基于异构传感器数据的智能家居非参数活动识别系统,系统通过集成环境传感器和体感传感器的异构数据来识别有意义的日常活动,无需用户标记参数,减少了用户负荷,实现适老服务。窦金花等^[17]针对老年用户的任务、时间、环境情境提出了主动交互式语音用户界面设计,为老年用户提供了更加自然、舒适的智能家居产品使用体验。缪珂等^[18]对新中产人群生活情境特征进行了分析,提出面向新中产人群生活情境的智能家居产品设计对策,认为智能产品应具备“学习”能力,能更好地自动捕捉、记录、适应个体的行为习惯。

通过对相关工作的整理,不难看出,智能系统交互的主动性是将来智能家居的发展方向之一。智能系统的主动性本身又涵纳多元问题,如前文所言,人的信息交互中,话语者需要借助多种模态信息来充分表达意图。同样,在智能家居的信息交互中,智能家居系统需要通过多元信息的获取与分析,来实现对用户意图和环境变化的准确判断。

2.2.2 智能家居多模态信息交互需求

多模态是信息交互的理想状态,用户与智能家居的交互,需要参考并利用大脑的多模态协同机制和人与外界的多模态交互规律,开发服务于人的“理想状态”的相关技术。随着智能技术的发展,智能家居与用户及环境的信息交互必然走向“人(人工智能系统)与人(用户)”的多模态会话交互逻辑。

一般而言,智能家居环境中包含丰富的信息模态,如语音、图像、影像、手势、体感等。智能家居

对用户与环境的多模态信息获取,需要给机器配上感官系统(即匹配各种传感设备)以模拟人的视觉、听觉、触觉等功能,通过语音模块、界面控制、动作捕捉,以及温度、湿度等传感器等进行数据采集。智能系统作为居家会话中的参与方,需要协调多种传感装置、控制部件和相应的信息处理系统同时工作,处理用户和环境中的多模态信息。但是,现有的多模态信息交互系统,仅通过不同模态信号识别特定的操作指令,将不同模态的指令进行简单的串联与叠加。这种简单的“信号时序叠加”方式,并未考虑多模态信号之间的信息互补及增强特性,相对复杂的系统设置需要用户记忆大量新的操作规则,增加了用户的认知负荷^[16]。

因此,智能家居对用户需求的满足与环境变化的回应,需要对现场话语状况进行尽可能全面的、真实的数据采集,通过多媒体介质与多模态信息,力争捕捉用户与环境实时、完整的信息交互。在这方面,王江涛等^[19]通过研究用户的操作行为和认知行为,获取用户对智能家居产品的功能和操作需求,提出基于用户行为的设计研究方法,对提高智能家居产品可用性、降低用户的认知负荷有积极价值。

3 基于多模态信息交互的智能家居设计框架

智能家居多模态信息交互基于多模态理论,将语言、图像、声音、动作等多模态信息整合成为最有效的意义表达,实现信息交互的准确传达和有效反馈。本研究提出一种具有多模态特性的智能家居信息交互新模式,并建立多模态智能家居信息交互框架。

3.1 智能家居多模态信息交互模型

真实世界中,多模态强调人作为主体通过身体和大脑调动多个感官、通过多种媒体协同参与信息交互

过程。而在智能家居信息交互中, 智能家居系统被视为“人”, 通过多元传感设备调动智能系统“感官潜能”, 获取用户及环境的多模态信息, 促进多模态信息理解、意义构建与控制反馈。

在智能家居多模态信息交互中, 包含对象、感官模态与解释 3 个要素, 即以用户与环境为对象, 以各种传感设备和信息获取硬件为感官模态, 对获取对象多元化信息进行智能融合与分析; 按照 3 个步骤开展

工作, 即多模态信息获取、多模态信息融合与推理、多模态信息反馈; 这一流程性工作中, 多模态交互技术按照互补、非互补的组合形式, 共时、顺时的时序方式组合 2 个及以上的输入模式, 借助多种非侵入式的传感设备, 识别天然形成的人类语言和行为, 最后通过多模态智联终端对用户需求与环境变化主动做出反馈。基于上述思路, 结合多模态研究构建了智能家居多模态信息交互模型, 模型见图 4。

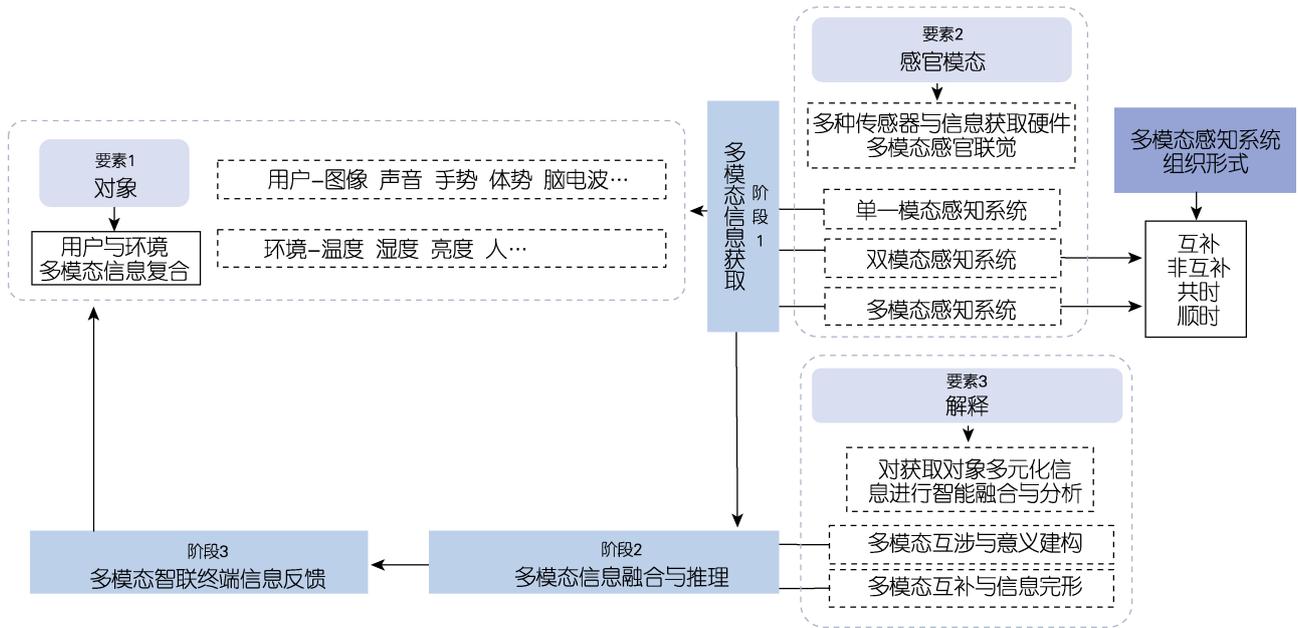


图 4 智能家居多模态信息交互模型
Fig.4 Smart home multimodal information interaction model

3.1.1 3 个要素：对象、感官模态与解释

1) 智能家居的信息交互对象。在家庭环境中, 智能系统的交互对象是家庭用户与家庭环境的多符号信息输出。用户信息包括身体与非身体信息, 身体信息是通过语音、手势、面部识别、体感等方式输出, 非身体信息是通过用户主动的界面操作等方式输出。环境信息方面包括可感知、可量化的变化, 如温度、湿度、亮度和人等因素。在智能家居信息交互中, 提高用户体验的路径是使智能系统主动寻求任务分配来降低用户负担, 也就是强化智能系统的信息获取主动性。在此条件下, 用户在遵循自然交互的身体逻辑下发出信息, 相对随机地选择与组织多模态身体信息。用户可能会发出无意识的指令, 智能系统需主动识别、采集, 以此减少用户信息交互负荷, 实现用户在智能家居控制方式上的自洽。

2) 智能家居的感官模态。人可以同时运用多种符号资源构成多模态信息展开交互, 环境也同步具有温度、湿度等多维度数据变化。智能家居的多模态信息交互通过多感官感知, 以及各种感觉系统间的联觉来开展工作。通过特定感官模态系统获取对应感官信

息, 以视觉为例, 视觉器官(摄像头或视觉传感器)加上处理视觉信号的分析系统组成一个视觉模态。智能家居与用户及环境的一次信息交互过程中的感官模态可能是单一模态、或者是双模态、多模态。例如用户通过语音控制智能家居设备, 但同时智能设备会获取语音交互的用户面部图像信息, 如用户发出“播放歌曲”指令, 智能设备同步识别该用户面部信息, 结合系统数据库内置用户画像, 精准推送该用户喜爱的歌曲。在对这些多模态信息处理的过程中, 智能家居系统通过多感官接收信息, 多元传感与分析模块同步工作, 以多模态的感官系统形成多模态的信息交互。

3) 智能系统对用户和环境多模态信息的解释。用户的语音、手势、姿态、表情等多模态信息相互关联, 以模态复合构成意指。智能系统对用户潜在意图或显性意图的准确反馈基于对用户需求的准确掌握。

一方面, 用户信息模态互涉, 即不同模态信息之间意义的相互阐释与呈现。智能系统需要根据模态互涉机制对获取信息进行解释。对多模态信息的互涉解释是目前智能交互技术的难点所在。目前能够对语音、文字和手势信息进行相对准确的解释。例如徐云平^[21]通过语音识别硬件 LD3320 的语音识别特征库直

接、快速实现语音识别、声控、人机对话功能；同步利用罗技 B525 摄像头获取手势信息，应用基于深度学习的手势识别算法，将语音和手势识别的智能系统应用在智能分类垃圾桶设计上。其他模态信息的互涉以建构相对完整的信息意义还有待持续研究。

另一方面，智能系统通过多模态信息的获取实现对部分模态信息不完备的补充。用户在智能家居使用中，日常生活自然行为存在语音、手势、姿态等多模态信息不完备的情况。用户日常行为中的多模态信息实际是高度整合的格式塔。在智能系统的多模态信息交互中语音信息句法结构不完备的情况下，用户身体的视觉信息作为替代与补充成分，协作构建相对完整的信息。对于不同模态信息之间的补充与建构关系需要借鉴语言学等学科成果。

3.1.2 智能家居多模态信息交互技术流程

对于面向用户信息和环境信息的智能家居系统，其多模态信息交互流程主要包括多模态信息获取、多模态信息融合与推理、多模态信息反馈等 3 个阶段，具体流程见图 5。

1) 多模态信息获取。智能家居系统通过传感器、触控屏、摄像头、麦克风等多种终端收集不同模态信息；其中，触控屏以被动形式获取用户输入，传感器、摄像头和麦克风则可以在用户预先给予权限的前提

下主动采集用户的姿态、手势、表情、动作以及语音等多种模态的信息。

2) 多模态信息融合与推理。不同终端采集的信息可经过不同的方式进行处理。目前，在机器学习尤其是深度学习的推动下，计算机视觉和自然语言处理领域都得到了空前的发展。因此，可应用深度学习方法对摄像头采集的图像、视频信息，以及麦克风采集的语音信号进行特征提取。若想让不同模态的信息互相补充，则需要将不同模态的信息进行融合。当前基于 Transformer 的跨模态大模型可以有效地结合视觉、文本和语音信息，并处理跨模态任务，如基于文本查询的视频中的动作定位、自动的视频字幕生成等。因此可以应用该模型对不同模态的信息进行融合和基于多模态信息的推理。此部分任务均在智能分析模块完成。智能分析模块对多模态信息进行整合处理，结合语料库和知识库的资源，将其转变为感知、记忆和行为，进一步输入智能控制模块。

3) 多模态信息反馈。智能控制模块根据智能分析模块的输出，对用户意图及环境变化做出反馈。完成转换的多模态信息将被编码为智能家居系统的行为潜势，进入智能家居系统的控制模块。在做出反馈时智能控制模块会结合用户的输入模态，以及用户画像决定信息反馈的模态和形式，实现符合用户需求和家居环境变化的功能操作。

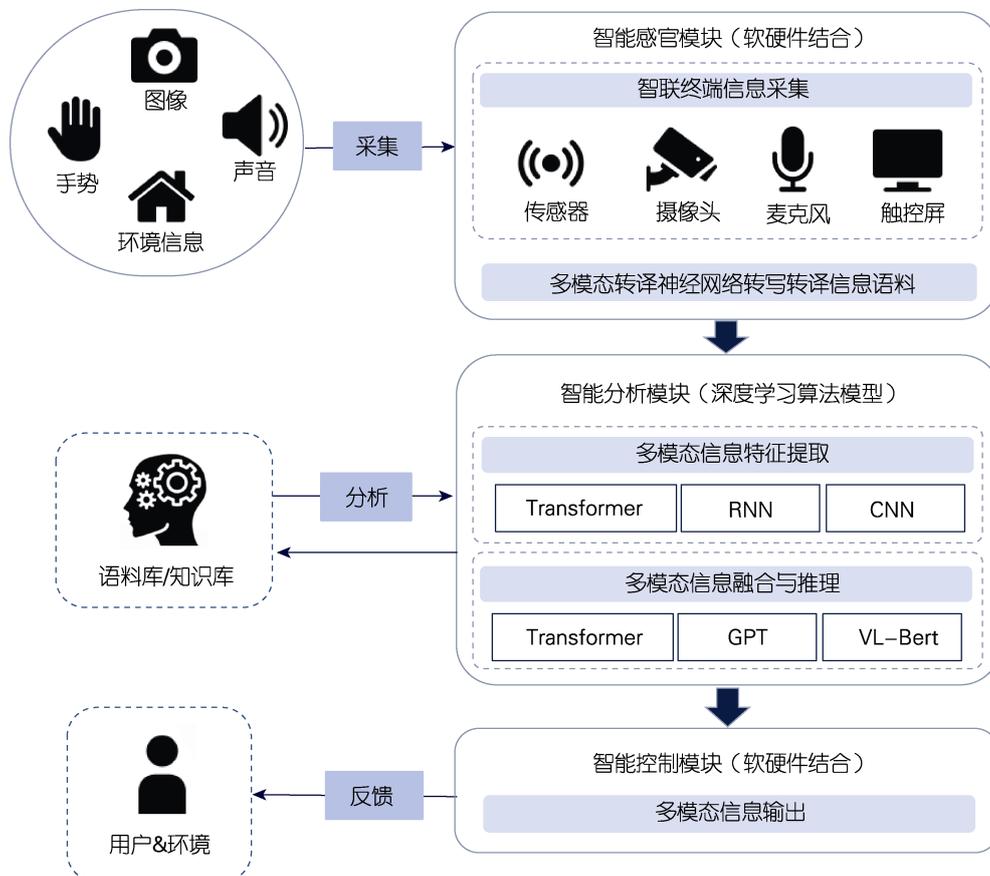


图 5 智能家居多模态信息交互技术流程

Fig.5 Smart home multimodal information interaction technology process

总体来说,智能家居系统的智能感官模块各司其职,将各自获取到的信息进行分拣、过滤、转换等处理,智能分析处理模块对获取的多模态信息进行特征提取、整合和分析,智能控制模块根据处理结果实施相应操作。这一过程类似于人类的中枢神经系统把感官系统接收的多模态信息整合起来,经过分析然后输出对用户意图及环境变化的行为反馈。

多模态智能家居系统的信息交互方式涉及基于多模态智联终端的信息采集、多模态信息资源整合、多模态信息特征提取、多模态行为互动、多模态评估等领域。利用多媒体技术,接入传感器、触控屏、摄像头、麦克风等多种模态的智联终端,实现多种软硬件设备间的无缝互联、协同工作,构建智能家居环境中多模态信息交互、多通道融合互补的多模态反馈运行机制。例如,针对用户的语言逻辑做出准确有效的回答,并且还能给予用户在视觉、听觉、体感等多维度的反馈,使人机交互更加自然,交互信息更加丰富,交互结果更加准确。

3.1.3 智能家居多模态感知系统的组合形式

在智能家居多模态信息交互中,不同模态感官系统以互补、非互补的结合形式,共时、顺时等时序方式进行组合,借助多种非侵入式的传感器,识别人类语言、行为和环境变量。

从用户交际活动的多模态方式来说,需要借助另一种模态来对主要模态信息进行补充,以实现意义的充分表达,这是模态之间的“互补关系”,而其他的则是非互补关系^[22]。互补的模态可以实现信息传达中的强化,一种模态是主要的交际形式,而另一种或者多种形式是对它的强化、补充和扩展。基于强化目的的感官模态主次组合形式,智能家居以一种模态的感知系统获取主要模态信息,同时借助另一种模态感知系统获取次要模态信息,消除单一模态信息带来的不确定性,增强对信息判断的准确性。

非互补关系表示智能家居的第 2 种模态感官系统对第 1 种在意义的获取上并无特殊贡献,但仍作为一种模态出现。例如,通过语音就达到信息交互目的,但同时通过界面显示交互信息,界面的补充并没有在信息内容上进行增减,没有对语音信息做出明显补充或强化,也没有对语音信息造成干扰。但一定程度上可以向用户进行语音信息交互有效性的可视化确认。

此外,从模态出现顺序的层面,Norris^[23]提出,在进行多模态信息交互时,信息模态结合形式有多模态共时、多模态顺时等方式。应对用户意图及环境变化信息的多种模态,智能家居也以共时和顺时的时序方式组织多模态感官系统。几种模态的感官系统可以同时开启信息获取与分析,例如界面交互时可以包括手势、压力、图形等几种不同的模态感知;也可以按顺序进行,一种模态感官系统的信息获取与分析工作结束则另一种模态的感官系统开始工作,例如语音开

机后,通过界面进行细化控制。

3.2 多模态交互中的用户状态与智能系统

3.2.1 智能家居系统状态

通过多模态感官系统从多种信息通道收集用户和环境多方面的行为数据和结果数据,实现了智能家居系统的主动性、交感化信息获取和情境化的信息分析与反馈。

3.2.1.1 信息的主动采集与交感化获取

一方面,在传统的智能家居使用中,用户只能通过语音、界面等单向度的信息输出方式向智能系统发出指令,在多模态智能交互场景下,智能家居系统对用户和环境的行為及结果数据主动采集,融合分析,并提供多元反馈,实现智能系统对用户意图与环境变化的主动感知。智能家居系统中的多模态感官系统高度分类细化与自动化,依托专门系统执行对不同模态信息的处理。

另一方面,智能家居通过交感化方式获取用户与环境信息。智能家居系统在接收任务后,会将其分配至某个任务模块进行处理。当面临复杂问题或任务时,某个交互任务涉及多个模态的信息,需要系统选取多个路径,在处理加工时在多个模态中开展协同,将任务分配给多个子任务处理模块同时处理、计算,从而共同完成这一交互任务。例如在生活场景中,用户的听觉(语音)信号和视觉(身势)信号同时启动,智能系统就需要同步感知、获取用户的语音与姿态信息,进行输入、分析,在这一过程中系统要关注不同模态信息之间的相互关联性和互涉性,感官形态交融,使对用户与环境信息的解读更加全面与准确。

3.2.1.2 信息的情境化分析

多模态智能家居信息交互把语言、图像、手势、颜色、环境等符号系统结合起来,从整体的角度分析各类信息所组成的话语意义,对用户和环境的理解借助对不同模态信息的元功能和组合形式等进行采集与分析,解读各种模态信息及其变化构成的意义。通过多模态信息的融合分析,理解用户情境,根据情境进行需求判断,提供低误差、一体化的服务反馈控制。

未来,5G 技术将促进全息技术、高清晰视频、VR/AR、物联网技术、人机交互多样形态的融合,为多模态融合智能家居信息交互创造条件,使意义表征和信息交流更加多模态化,使多模态信息融合交互更加便捷,为多终端互通互联提供传输条件,推进更加智能化的多模态家居服务。

3.2.2 用户与环境状态

在多模态智能家居信息交互中,与传统交互模式相比较,智能系统自主性提升,用户投入适宜家庭环境的交互行为,信息交互更加自然,交互过程认知负荷降低。

1) 多维度信息。家庭环境中,用户基于日常生

活习惯开展活动,多模态的智能家居信息交互满足用户自然、无意识的信息输出。智能家居在交互过程中积极、持续地投入多模态感官状态,通过隐式传感设备的一体性联动与多模态信息融合分析,使智能系统使用多种模态感官设备对用户和环境数据进行持续监测和跟踪,对人们日常生活中的自然行为进行多维度采集、分析与融合,并提供反馈,更符合人的自然交互特征。

2) 多用户交互。智能家居交互涉及多名用户。在智能家居系统使用中,多设备、多模态、多用户交互的复杂关系涉及行为参与、社会关系等维度,需要综合用户的外显行为和内隐心理等多维数据来表征其状态。此时单一模态信息不足以使智能系统充分解读特定用户需求,多模态满足智能家居系统对特定用户多维信息分析基础上的准确反馈。

3) 时序性投入。智能家居使用过程中,用户的投入状态会随时间发生变化。智能家居系统通过共时及顺时的时序方式,通过多模态感官系统全时关注用户在使用过程中的活动顺序、行为特征、心理和生理数据,在时间序列上研究其动态变化,分析出用户和环境轨迹化、时序化特征,基于智能分析技术,为用户及环境提供基于数据动态变化的持续反馈。

3.2.3 智能家居多模态数据库、语料库构建

智能家居系统在用户使用中建立基于特定用户、特定环境的多模态语料库,对信息交互中采集到的语音、姿态、人脸等用户信息,以及环境空间、温度、湿度、亮度、声音分贝等环境信息进行分析与集成,并建立以多模态方式加工、检索和统计的多模态信息语料库,便于后续持续性地调用与语料处理,对应特定用户需求的数据匹配,满足与特定用户及环境持续的、准确的信息交互。在这方面,Sandra等^[24]提出了一种用于家庭环境中设备管理的自动化系统,能够管理所有电器,服务器接收和发送的所有数据都存储在数据库中,通过系统配备的智能算法检查、比较、分析相关数据,辅助系统进行设备管理。在多模态语料库构建上,国内外学者把人与外界的多模态互动作为重要的数据来源,在语料采集、加工、标注、分析框架和工具研发等方面形成了诸多成果^[22],这些成果对于视听识别系统、人机智能对话等都具有重要意义,可作为智能家居设计中多模态信息交互的数据包使用。

4 智能家居中多模态交互信息的形式及其关系

智能家居多模态信息交互在相对复杂、动态的家庭环境活动中开展,智能家居设计中,通过分析用户与环境多模态信息的注意力、模态密度和模态结构的配置,能够厘清多模态信息交互中模态的数量、重要性和模态之间关系等问题。

4.1 智能系统对模态信息的注意力

在实际的家居环境中,存在多样复杂的事件和行为,智能家居系统需要对用户需求指令精准识别,这就需要对复杂环境中不同模态信息给予不同程度的注意。智能家居系统内置传感设备运行中要对环境中不同模态信息进行前景、中景和背景的分层感知,并给予不同的注意程度,“注意的焦点是前景高层活动,对中景、背景活动分配的注意力依次递减”^[22]。例如在复杂的环境音中迅速过滤环境音,甄别出用户指令语音;能够在多个用户中迅速识别发出指令的用户,并对用户需求做出即时反馈。

此外,在家庭环境中,转头、手指或手掌敲击、注视转移等动作常被设定为预置动作,参与到交互过程。这些预置动作常出现在正式的交互行为之前,对正式指令发出起着组织和调节的作用,是标示正式指令模态信号的转换信号,正如在人和人的交流中,这些转换信号往往能够影响互动的对方,引起“听话者”的注意。在用户与智能家居的交互中,这些信号与用户指令内容无关,没有具体指向,但能控制智能系统注意或感知的焦点,因此智能家居系统对用户行为指令之前的预置动作进行习得性认知也是必要的。

4.2 模态密度

模态密度可以体现为模态强度和模态复杂度2种形式^[22]。模态强度是指一种模态在互动过程中的重要性或权重,在互动中发挥主要作用的模态就是高强度模态。当高强度模态被停止或改变时,由其映射的用户需求信息或环境变化信息也随之停止或改变。此外,在互动过程中一种模态的强度、权重或重要性取决于使用情境变化和用户活动环境变化等诸多因素。例如白天用户对智能家居的控制通过语音进行,夜间为了避免语音干扰,停止使用语音,转而使用界面控制。模态强度与智能系统对复杂环境中不同模态信息的注意程度呈正相关,智能家居信息交互设计中需要甄别不同情境、用户群的高强度模态信息,据此采用适应性的多模态智能交互技术。

模态复杂度是指共同完成交互的信息模态数量和组织方式。智能系统的多模态感知系统需以共时、顺时等形式进行组织,模态复杂度与智能家居系统的传感设备数量及布局成正相关。多种模态信息共同完成交互时,一种模态的变化不会引起交互活动的突然改变,例如借助体感、语音等多模态信息进行的家庭游戏系统,用户使用语音、注视、手势、物体操作、姿态等多种模态,系统本身对用户多模态信息有综合识别的设置,停止或改变使用其中任何一种模态,不会直接改变游戏路径。

4.3 模态结构配置

由于不同模态信息在交互中共存的复杂性,智能家居设计中要结合用户与产品类型进行模态结构配置。

在多模态交互中, 每种模态作用不同。结合智能家居产品应用场景、目标用户与产品类型进行判断, 标注在此产品多模态交互中的不同模态强度, 界定每种模态在互动过程中的重要程度, 据此进行相应的多模态信息交互系统设计。

Norris^[25]认为可以通过回答以下 3 点来确定信息交互中的模态配置: 哪些模态是必需的; 哪种模态比其他模态更重要; 哪些模态不是必需的但活动者仍然在使用。智能家居多模态信息交互中的模态结构配置需要结合用户及环境变化, 据此分析模态的组成和模态之间关系, 确定由哪些模态参与相关的信息交互, 以及相对应的智能家居多模态感知系统以何种组合方式布局。

5 结语

多模态的研究在语言学、计算机科学中处于热潮, 其方法和理论在智能家居多模态交互方面有重要的参考和应用价值。在多模态传感及交互技术发展下, 智能家居多模态信息交互基于多模态感知系统, 对用户与环境数据信息进行实时动态采集, 为实现多模态的用户需求反馈提供数据支持, 在多模态信息融合分析基础上开展精准决策和调控。随着人工智能技术与 5G 等移动通信网络的进一步发展, 智能家居多模态信息交互将会变得更为人性化, 更好地服务于家庭应用场景。

参考文献:

- [1] KOLB B, WHISHAW I Q. Fundamentals of Human Neuropsychology[M]. New York: Worth Publishers, 2005:135.
- [2] KANDEL E, JAMES H, THOMAS M, STEVEN A. Principles of Neural Science[M]. New York: The McGraw-Hill Companies, Inc, 2013: 449.
- [3] 朱永生. 多模态话语分析的理论基础与研究方法[J]. 外语学刊, 2007(5): 82-86.
ZHU Yong-sheng. Theory and Methodology of Multimodal Discourse Analysis[J]. Foreign Language Research, 2007(5): 82-86.
- [4] 顾日国. 多模态感官系统与语言研究[J]. 当代语言学, 2015, 17(4): 448-469.
GU Ri-guo. Multimodal Sensory System and Language Research[J]. Contemporary Linguistics, 2015, 17(4): 448-469.
- [5] KRESS G, LEEUWEN V T. Reading Images: The Grammar of Visual Design[M]. London: Routledge, 2006: 22.
- [6] NORRIS S. Analyzing Multimodal Interaction[M]. New London: Routledge, 2004.
- [7] 王党校, 郑一磊, 李腾, 等. 面向人类智能增强的多模态人机交互[J]. 中国科学: 信息科学, 2018, 48(4): 449-465.
WANG Dang-xiao, ZHENG Yi-lei, LI Teng, et al. Multi-Modal Human-Machine Interaction for Human Intelligence Augmentation[J]. Scientia Sinica (Informationis), 2018, 48(4): 449-465.
- [8] 顾日国. 多模态感官系统: 天官、仿人机器人、“修辞即做人”新释[J]. 当代修辞学, 2019(5): 48-67.
GU Yue-guo. Multimodality and Rhetoric: From Xunzi, Aristotle, TCM, Empiricism, Humanoid to Man[J]. Contemporary Rhetoric, 2019(5): 48-67.
- [9] ALANWAR A, ALZANTOT M, HO B J. SeleCon: Scalable IoT Device Selection and Control Using Hand Gestures[C]// Proceedings of the Second International Conference on Internet-of-Things Design and Implementation. Pittsburgh PA: ACM, 2017: 47-58.
- [10] Alrajhi W, Alaloola D, Albarqawi A. Smart Home: Toward Daily Use of Bci-based Systems[C]// 2017 International Conference on Informatics, Health & Technology (ICIHT). Riyadh: IEEE, 2017: 1-5.
- [11] 胡旭央, 张寒凝. 以睡眠场景为例的智能家居场景模型构建研究[J]. 包装工程, 2021, 42(10): 124-129.
HU Xu-yang, ZHANG Han-ning. Construction of Smart Home Scene Model Taking Sleep Scene as a Case[J]. Packaging Engineering, 2021, 42(10): 124-129.
- [12] 张伟, 王宜怀. 基于 AVR 的智能家居系统设计与实现[J]. 计算机技术与发展, 2022, 32(3): 209-213.
ZHANG Wei, WANG Yi-huai. Design and Implementation of Smart Home System Based on AVR[J]. Computer Technology and Development, 2022, 32(3): 209-213.
- [13] ROSE J, TRUEX D. Machine Agency as Perceived Autonomy: An Action Perspective[M]. Boston: Springer, 2000: 371-388.
- [14] MAHADEVAN K, SOUSA M, TANG A. "Grip-that-there": An Investigation of Explicit and Implicit Task Allocation Techniques for Human-Robot Collaboration[C]// Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. Yokohama: ACM, 2021.
- [15] LEUSMANN P, MOLLERING C, KLACK L. Your Floor Knows Where You Are: Sensing and Acquisition of Movement Data[C]// 2011 IEEE 12th International Conference on Mobile Data Management. Lulea: IEEE, 2011.
- [16] WU Chao-lin, CHEN Y H, CHIEN Yi-wei, et al. Non-parametric Activity Recognition System in Smart Homes Based on Heterogeneous Sensor Data[J]. IEEE Transactions on Automation Science and Engineering, 2019, 16(2): 678-690.
- [17] 窦金花, 齐若璇. 基于情境分析的适老化智能家居产品语音用户界面设计策略研究[J]. 包装工程, 2021, 42(16): 202-210.
DOU Jin-hua, QI Ruo-xuan. Elderly-Adaptability Voice User Interface Design Strategy of Smart Home Products Based on Context Analysis[J]. Packaging Engineering, 2021, 42(16): 202-210.