

沉浸式语言学习环境中的多模态融合与交互研究

周斯海 洪业浚 潘炬盛 梁译茗 林倩如*

广西职业师范学院

DOI:10.32629/jief.v7i9.18261

[摘要] 随着人工智能、5G通信、大数据分析等现代信息技术的快速发展,教育与技术的深度融合不断推动教学模式的创新与变革。在语言学习领域,传统的“单模态输入、被动接收”方式已难以满足学习者对场景化、个性化和互动性的需求,沉浸式语言学习环境应运而生。该模式以虚拟现实(VR)、增强现实(AR)、人工智能(AI)等技术为支撑,通过整合学习者的语言输出、行为轨迹、生理反馈等多模态数据,构建个性化学习场景与智能交互体系,形成“沉浸式体验、自适应学习、强互动协作”的语言学习新形态。本文系统分析了多模态数据采集的必然性、模态协同的非对称性以及“需求适配”机制在实际应用中的失灵问题,并从学习者认知特点与权利需求出发,深入探讨了模态冗余与缺失、交互反馈滞后、认知负荷失衡以及技术适配性不足等核心困境。在此基础上,提出构建“以学习者为中心”的融合框架、完善多模态学习数据治理法规、推动技术向善与适学化设计、强化教育机构与社会协同支持等多维优化路径,以促进沉浸式语言学习环境的良性发展,确保学习者在享受技术红利的同时,其学习自主权与数据权益得到充分保障。

[关键词] 沉浸式语言学习; 多模态融合; 人机交互; 认知负荷; 教育技术; 数据治理

中图分类号: G40 **文献标识码:** A

Research on Multimodal Integration and Interaction in Immersive Language Learning Environments

Sihai Zhou Yejun Hong Jusheng Pan Yiming Liang Qianru Lin*

Guangxi Vocational Normal University

[Abstract] With the rapid development of modern information technologies such as artificial intelligence, 5G communication, and big data analytics, the deep integration of education and technology continues to drive innovation and transformation in teaching models. In the field of language learning, the traditional "unimodal input and passive reception" approach can no longer meet learners' demands for contextualized, personalized, and interactive experiences, leading to the emergence of immersive language learning environments. Supported by technologies like virtual reality (VR), augmented reality (AR), and artificial intelligence (AI), this model integrates learners' multilingual outputs, behavioral trajectories, physiological feedback, and other multimodal data to create personalized learning scenarios and intelligent interaction systems, forming a new language learning paradigm characterized by "immersive experiences, adaptive learning, and strong interactive collaboration." This paper systematically analyzes the inevitability of multimodal data collection, the asymmetry of modal coordination, and the operational failures of the "demand adaptation" mechanism in practical applications. From the perspective of learners' cognitive characteristics and rights needs, it delves into core challenges such as modal redundancy and deficiency, delayed interaction feedback, cognitive load imbalance, and insufficient technical adaptability. Building on these findings, the study proposes multidimensional optimization strategies, including the construction of a "learner-centered" integration framework, the improvement of multimodal learning data governance regulations, the promotion of technology-for-good and learner-adapted design, and the strengthening of collaborative support between educational institutions and society. These measures aim to foster the healthy development of immersive language learning environments while ensuring learners' learning autonomy and data rights are fully protected as they enjoy the benefits of technological advancements.

[Key words] Immersive language learning; Multimodal integration; Human-computer interaction; Cognitive load; Educational technology; Data governance

引言

党的二十大报告明确提出,推进教育数字化,建设全民终身学习的学习型社会、学习型大国。总书记强调,教育数字化是塑造教育发展新优势的重要突破口,为建设教育强国指明了方向^[1]。在此背景下,沉浸式语言教学模式作为一种创新的教育方式,近年来受到广泛关注。该模式注重语言真实环境的模拟,旨在让学习者在近似真实的语境中获得自然、流畅的语言表达能力。随着人工智能、虚拟现实等技术的快速发展,语言学习正朝着“场景化、个性化、实践化”方向转型。《2024年中国语言教育技术发展报告》显示,全国语言学习者超过5亿人,其中72.3%的学习者认为“缺乏真实语境与多感官参与”是当前语言学习的主要瓶颈,65.8%的教育机构希望通过新技术解决个性化教学与规模化覆盖之间的矛盾。沉浸式语言学习环境融合VR、AR、AI等技术,通过智能设备采集学习者多维度数据,提供场景化练习与个性化反馈,然而在实际应用中,学习者往往被动接受模态融合,陷入学习效果与体验感双重困境^[2]。

国内外学者已围绕沉浸式语言学习开展了一系列研究。国外研究主要关注场景真实性与交互流畅性之间的张力、数据隐私风险以及数字鸿沟问题;国内研究则多从教育技术与认知科学视角出发,聚焦于多模态资源库构建与教学模式设计等方面。然而,现有研究大多将多模态融合视为静态技术组合,缺乏对“模态协同”动态过程及学习者个体差异的深入剖析。本文以多模态融合与交互为切入点,结合认知负荷理论与建构主义学习理论,采用文献研究方法,系统梳理沉浸式语言学习技术体系及数据流动路径,解析多模态融合引发的交互冲突,并结合典型产品分析,提出多主体协同优化框架,为沉浸式语言学习环境的规范发展提供理论参考与实践指导。

1 沉浸式语言学习环境的技术基础与多模态数据采集

沉浸式教学法起源于加拿大,其核心理念是从语言本质出发,通过营造真实的语言环境提升学习效果。为克服传统外语教学模式中语境缺失、参与度低等弊端,本文提出构建以虚拟现实、大数据等技术为支撑的沉浸式语言学习模式。该模式通过构建高度沉浸的虚拟场景,引导学习者在近似真实的环境中开展语言实践,从而加强语言输入与输出能力,培养自然语感,拓展文化视野,增强跨文化实践能力。在此基础上,通过智能化数据评估与知识转化机制,全面提升外语学习质量。沉浸式语言学习环境的技术体系主要包括虚拟现实(VR)、增强现实(AR)、人工智能(AI)、大数据与5G通信等^[3]。VR/AR技术通过构建虚拟场景或叠加虚实元素,为学习者提供“身临其境”的语言实践环境;大数据技术负责对海量异构多模态数据进行存储、清洗与分析,构建学习者能力画像与学习路径;人工智能则借助机器学习、自然语言处理等方法实现语音识别、错误纠正、智能推荐与交互响应;5G技术为数据传输与实时交互提供高速、低延迟的网络

支持。这些技术的综合应用,使语言学习从被动接收转向主动参与,从标准化教学迈向个性化定制,显著提升了学习效率与用户体验。

在沉浸式语言学习场景中,多模态数据的采集范围远超传统认知,涵盖语言表达、行为动作、生理反馈、环境交互等多个维度。具体而言,语言表达数据包括语音发音、语调节奏、文本输出等直接反映语言能力的信息;行为动作数据包括肢体姿态、手势动作、视线轨迹等刻画学习参与度的指标;生理反馈数据通过心率传感器、表情捕捉设备采集,间接反映学习者的注意力集中度与情绪状态;环境交互数据则通过场景操作记录、资源调用轨迹等,分析学习者的学习偏好与需求。这些数据通过交叉关联与深度挖掘,能够形成高度个人化的“学习者数字画像”,不仅揭示其显性的学习短板与能力优势,还可能预测潜在的学习风险与个性化需求。然而,数据价值的最大化也伴随着模态协同风险与认知负荷压力的指数级增长。一旦融合不当或交互设计失衡,将对学习者的学习效果与体验感造成严重且不可逆的影响。

2 多模态融合的必然性与实践困境

当前外语教学普遍采用信息化教学设备、PPT课件、原声音视频资源等多媒体手段配合完成教学任务,其主要目标仍集中在知识传授与基本语言技能训练。尽管信息化手段增强了课堂的交互性与内容的丰富性,引入了对话、小组讨论、演讲、同声传译等多种形式,但学生实际掌握程度仍值得质疑。究其原因,在于现有课堂往往缺乏真实语境,教学方法机械,教学内容枯燥,学习过程缺乏实质性参与与自然情感投入,极大限制了学生的学习兴趣与效率^[4]。因此,在沉浸式语言学习实践中,多模态融合具有必然性。学习者为了获取场景化练习、个性化反馈、协作式交互等服务,往往需要在“需求悖论”中作出权衡——尽管在观念上重视学习效率与体验感,却在具体情境下为换取技术赋能而被动接受模态组合与交互设计。

学习者的多模态融合体验具有鲜明的特殊性,主要体现在三个方面:一是需求适配的非对称性,学习者作为技术接受方,在模态选择、交互方式、反馈形式等方面的议价能力较弱,难以真正参与设计决策过程;二是“需求—适配”模型的形式化,面对复杂的技术参数与固定的模态组合,学习者的需求表达多流于表面,且在集体学习场景中,决策权常由教师或平台代为行使,进一步削弱其自主地位;三是个体脆弱性的加剧,青少年因认知能力尚未成熟,易受冗余模态的干扰;老年学习者与低数字素养群体因操作能力不足,难以适应复杂交互设计;甚至部分学习者因生理特征无法正常使用特定模态,导致学习权益受损。这些困境具体表现为模态冗余与缺失、交互反馈滞后、认知负荷失衡以及技术适配性不足等四大核心问题,严重制约了沉浸式语言学习环境的效果发挥与可持续发展。

3 应对实践困境的策略与路径

3.1 以学习者为中心, 构建动态适配的融合框架

确立“以学习者为中心”的多模态融合框架, 尊重学习者自主权与个性化需求: 提供简明的模态选择与交互设置, 允许根据学习目标(词汇记忆、口语练习等)与自身特征(年龄、数字素养等)调整模态组合, 设置“一键简化”“模态暂停”功能; 贯彻数据采集“最小必要原则”, 避免收集无关信息, 按学习阶段调整模态密度(基础阶段以“语音+文本+图像”为主, 高级阶段加入动作、环境感知模态); 推动技术从“技术主导”转向“需求适配”, 开发分群体版本(青少年版重趣味性、老年版重简易性、特殊需求版重无障碍设计)^[5]。

3.2 完善相关法律法规与标准体系

在《个人信息保护法》《教育法》基础上, 制定沉浸式语言学习多模态数据处理实施细则, 明确数据控制者责任边界, 加大侵权惩处力度; 建立行业标准与认证机制, 涵盖模态融合质量、交互反馈效果、适学化设计, 通过“模态协同度”“反馈精准度”等指标评估, 引导规范发展; 探索分层适配、动态调整的需求响应机制, 允许学习者分阶段、分场景选择模态与交互方式, 保留调整权利。

3.3 技术向善与适学化创新

多模态数据处理环节采用认知负荷调控技术与适配性设计, 如自适应算法调整模态密度、注意力引导技术突出核心信息; 推出分群体“专属模式”(青少年版强化游戏化交互、老年版用大字版与语音引导、特殊需求版提供手语交互), 简化操作, 通过多通道提示增强界面友好性; 提升系统透明度与算法可解释性, 以通俗语言说明数据收集与使用逻辑, 对AI反馈提供解释, 避免“黑箱”操作^[6]。

3.4 强化教育机构与社会支持网络

通过学校课程、社区讲座、公益培训, 开展数字素养与沉浸式学习能力培训(如为老年人开设设备使用课程、为青少年开展注意力管理培训); 引导教育机构与教师将多模态技术作为教学辅助工具, 结合系统数据调整教学策略, 在模态不匹配时补充传统教学; 加强服务提供者教育伦理教育, 将“以学为本、公平待学、个性化适配”融入设计与服务, 培育行业伦理文化。

从宏观上看, 社会环境、文化环境影响着语言学习者的动机和习得速度, 语言环境的属性直接决定语言运用的可能性与质量^[7]。沉浸式语言环境可以迫使语言学习者不断使用目标语言进行交流, 有助于促进目的语基础知识的掌握和交际能力的提升, 可以显著提高语言学习者的语言习得速度^[5]。

4 结论与展望

沉浸式语言学习作为应对教育数字化与语言学习需求升级的创新模式, 其健康发展必须直面多模态融合与交互设计的核心问题。学习者出于效率与体验考虑接受多模态融合的行为, 并非完全自主的选择, 而是深受技术逻辑与教育资源分配的影响, 易导致模态适配不足、反馈滞后、学习不公与体验感异化等

后果。破解这些困境, 不能仅依靠技术迭代或局部设计调整, 而需进行系统性的“需求转向”, 将学习者视为有差异、能自主、需适配的权利主体, 而非技术应用的被动接受者。

未来的沉浸式语言学习应实现“技术”与“教育”的深度融合。技术层面追求精准、适配与无感, 教育层面则坚守人本、公平与高效。在实践路径上, 需以“以学习者为中心”的融合框架为基石, 通过完善法律标准规范多模态数据处理与设计, 借助适学化技术降低使用门槛与认知风险, 依托教育机构与社会网络重建需求响应与支持体系。唯有在需求底座的支撑下, 通过技术创新、制度保障与文化培育的协同推进, 才能构建真正普惠、高效、可信的沉浸式语言学习生态, 使技术进步成果真正惠及每一位学习者, 实现学有所成、学有所乐、学有所适的语言教育愿景。

[项目信息]

本项目由自治区级大学生创新创业训练计划项目资助, 项目名称: 广西职业师范学院2025年大学生创新创业训练计划项目《虚实共生-基于人工智能的个性化沉浸式语言学习元宇宙》, 项目级别: 自治区级, 项目类别: 一般项目, 项目编号: S202514684078。

[参考文献]

- [1]梁俊斌, 谢贇. 生成式人工智能赋能教育数字化转型的机理、挑战与对策[J/OL]. 现代教育管理, 1-9[2025-10-26].
- [2]杜俊颖. 人工智能赋能高校外语教学研究[J]. 英语广场, 2025, (12): 73-76.
- [3]刘上. 多模态交互环境下的大学英语在线教学策略优化研究[J]. 佳木斯职业学院学报, 2025, 41(03): 139-141.
- [4]海月. 多模态视域下高校英语口语教学中学生语言输出能力分析[J]. 太原城市职业技术学院学报, 2025, (09): 137-139.
- [5]张文超, 袁磊, 闫若楠. 从游戏化学习到学习元宇宙: 沉浸式学习新框架与实践要义[J]. 远程教育杂志, 2022, 40(04): 3-13.
- [6]张轶骏, 周晶. VR与AI赋能的沉浸式情境口译教学模式研究[J]. 外语电化教学, 2021, (01): 78-84+13.
- [7]江琴. 论美剧对实施英语沉浸式学习的有效补充作用[J]. 重庆电子工程职业学院学报, 2017, 26(05): 129-131.

作者简介:

周斯海(2000--), 男, 汉族, 广西钦州人, 本科, 广西职业师范学院, 研究方向: 物联网工程。

洪业浚(2000--), 男, 汉族, 广西平南人, 本科, 单位: 广西职业师范学院, 研究方向: 物联网工程。

潘炬盛(2000--), 男, 壮族, 广西南宁人, 本科, 广西职业师范学院, 研究方向: 物联网工程。

梁译泓(2002--), 男, 汉族, 广西梧州人, 本科, 广西职业师范学院, 研究方向: 物联网工程。

*通讯作者:

林倩如(1997--), 女, 汉族, 广西南宁人, 研究生, 广西职业师范学院, 研究方向: 外国语言学、人工智能。