

多模态 AI 驱动的本科英语演讲能力训练模型构建与验证

刘青 李爱琴

昆明城市学院，云南昆明，650101；

摘要：随着 AIGC 与多模态技术融合，本科英语演讲教学正转向“认知—社会—表达协同发展”新范式。为突破传统教学模式资源单一、情境失真、反馈滞后等局限，研究构建了一个 AI 驱动的多模态训练模型。该模型以“认知建构—情境模拟—多模态反馈”为核心路径，整合 AI 在内容生成、虚拟仿真、实时评估与情感计算等技术，通过“人机辩证互动”、“情境再语境化”及“元认知分阶段发展”三大机制，应对了演讲中“内容浅表化、情境脱节、反馈滞后”等核心问题，旨在系统提升学生的演讲内容深度、逻辑性、表达力及跨模态协调能力。

关键词：多模态 AI；英语演讲；认知发展；再语境化；元认知；人机协同

Construction and Verification of a Multimodal AI-driven Training Model for Undergraduate English Public Speaking Skills

Liu Qing Li Aiqin

Kunming City University, Kunming, Yunnan, 650101;

Abstract: With the integration of AIGC and multimodal technology, undergraduate English speech teaching is shifting towards a new paradigm of "cognitive-social-expression coordinated development". To overcome the limitations of traditional teaching models such as single resources, distorted scenarios, and delayed feedback, this study constructs an AI-driven multimodal training model. This model takes "cognitive construction - scenario simulation - multimodal feedback" as its core path, integrating AI technologies in content generation, virtual simulation, real-time assessment, and emotion computing. Through three mechanisms: "dialectical human-machine interaction", "recontextualization of scenarios", and "phased development of metacognition", it addresses core issues in speeches such as "superficial content, disconnection from scenarios, and delayed feedback", aiming to systematically enhance students' speech content depth, logic, expressiveness, and cross-modal coordination abilities.

Keywords: Multimodal AI; English speech; Cognitive development; Recontextualization; Metacognition; Human-machine collaboration

DOI: 10.64216/3104-9702.25.06.019

引言

在当今全球化和高等教育国际化的环境下，英语学术演讲能力是本科学生重要的学术素养。这种能力既有语言表达的能力，又兼有逻辑思辨、知识整合以及跨文化交流等多方面的高阶技能^[1]。但是，传统的英语演讲教学仍然存在诸多问题，教学内容容易模板化，缺少思想深度；教学情境容易脱离真实学术领域，造成知行脱节；教学反馈由于容量小、精力有限而无法得到及时性和个性化的改善^[2]等。传统英语演讲教学难以满足新时代创新人才培养的需求，亦难以应对人工智能对传统认知模式带来的深刻变革。

以大型语言模型（LLM）、生成式多模态人工智能（AIGC）为代表的大规模语言模型，给解决该类教学

难题带来了前所未有的机会。AI 不仅能生成结构严谨、视角多元的演讲内容，模拟高度仿真的学术演讲场景，还能通过语音识别、情感计算、姿态捕捉等技术，提供实时、多维、精准的评估与反馈^[3]。已有研究已经开始探讨 AI 在语言教学中的应用，如口语训练时用 AI 提高学习者流利度和词汇复杂度^[4]；写作教学时用 AI 促进批判性思维、体裁知识的发展^[5]；听说一体化教学时用 AI 构建起教学、评的闭环^[6]等。但是，这些研究大多只关注于某一单项技能（发音、写作）或者某一种模态（语音），缺少一个可以将内容、情境和表达有机地融合起来，贯穿到认知发展全过程的教学模型。因此，在英语演讲这一综合性、社会性极强的能力培养方面，怎样把多模态 AI 技术融入到教学当中去，既保证技术赋能的

同时又保证认知自主发展，是目前理论探索与实践创新的瓶颈。

鉴于此，本文建构起一个多模态AI驱动的训练模型。该模型既关注技术工具的作用，又重视AI怎样被理解成一个认知促进者、情境创建者的和元认知支持者，促进学生在“内容构建—情境实践—反思调节”三个方面发展自己的学术演讲能力。

1 多模态AI驱动模型的核心维度与运作逻辑

研究提出的模型是以学生认知发展为基础，依靠多模态AI技术来构建的动态循环系统。其核心是三个互相联系而且层层递进的方面，即认知建构层、情境模拟层和多模态反馈层。

1.1 认知建构层：AI作为思想深化与逻辑组织的深度协作者

学术演讲主要是内容的深度、新颖性和逻辑自洽这三个方面的任务，AI在这三个方面起着深度协作者的作用。第一，AI可通过深度内容生成技术和批判性对话搭建理论框架。学生可以使用大模型形成初步的观点，但是需要借助指令模型扮演“反方”或者给出跨学科的角度，进行批判性互动来推动自己的论述发展。第二，AI通过逻辑结构可视化优化讲稿。利用工具将论点、论据及结论的逻辑关系以思维导图等形式呈现，可以促使学生分析并修补逻辑断层，从而加强论证的严谨性。第三，AI借助学科知识图谱来加强专业性。系统可以快速的连接核心的概念、经典的理论以及文献，创建起一个主题的知识图谱，让学生把演讲的内容融入到学科的话语体系里，打牢学术基础。

1.2 情境模拟层：多模态虚拟环境中的社会化再语境化实践

学术演讲是特定社会文化语境中的规约性交际行为。为避免脱离语境的“纸上谈兵”，本维度融合VR/AR、AIGC与智能对话体技术，构建高保真、可交互的虚拟演讲场景，实现“再语境化”训练。

平台依靠自身智能的方法创建出一套动态的虚拟情境，能依据相应主题自动创建对应虚拟会场，之后模仿听众即时反应（如赞同或困惑），大幅度加深表演的真实性以及沉浸感。其次，依靠智能虚拟听众达成深层次的交互。这样的“数字人”，在问答环节提出挑战性问题，推动学生即时组织语言、调整策略，将独白转化

为对话，从而很好地发展学生的临场应变能力和说服能力。最后，平台聚焦跨文化语用意识的培养。通过文化参数的调整来改变系统所模拟的不同学术语境下的语言正式度、体态语和互动方式，从而呈现不同的期望。学生经过多次实践之后，不断变化策略去应对不同的交际环境，并且提高了对全球化学术交流的认识和使用能力。

1.3 多模态反馈层：实时、精准、多维的评估与元认知能力发展

及时、有效的反馈是演讲能力发展的催化剂。本维度整合了语音识别（ASR）、计算机视觉、情感计算等多模态感知技术，提供超越传统语言范畴的综合反馈，以促进学生元认知能力发展。

首先对语音进行细致的分析，从发音准确度、语速变化、重音分布、停顿模式和语调韵律等各方面提供多维度数据对比与改进建议。同时通过摄像头捕捉并分析演讲者的面部表情、眼神接触、手势运用及身体姿态等副语言行为，评估其在信息传递、建立连接与展现自信方面的效果，并给予针对性指导。每一次训练之后，系统把多维的数据融合成可视化《演讲能力发展诊断报告》，用雷达图、趋势线等方式来直观呈现学生在逻辑严密性、语言流畅度、体态表现力这三个方面的表现，与以往数据进行对比。根据诊断结果，AI就会给学生推荐个性化的学习任务包，引导学生从关注评分转向自我监控与调节，进而逐渐形成起持续的元认知能力。

2 核心机制

该模型的有效性不仅源于技术的集成，更根植于其背后契合认知发展规律的三大核心运作机制。

2.1 AI赋能与认知自主的辩证协同机制

演讲训练时，人工智能的作用是辅助而非替代思考。初期系统借助内容生成、情境模拟以及详细的反馈，为学生搭建“支架”，来帮忙学生跨越能力门槛，培育起自信。随着能力的提升，模型才开始“支架撤离”，逐步增加任务复杂性与自主性。比如减少直接的内容生成，把关键词或者矛盾的观点交给学生自己组织；在虚拟场景中，降低提示明确度，增加互动性的难度。该机制使学生在技术赋能过程中拓宽认识疆域的同时，保持认知主体的自主性，最终目的就是培养出可以独立开展深入思考和富有创造性的“适应型专长”。

2.2 多模态情境驱动的再语境化实践机制

借助多模态AI技术来完成再语境化的认知过程，使之可以被操作起来。学生的各个主题、文化环境以及互动模式下不断演练，实际上就是密集的、反复的语言情境化演练。他们要根据实时变化的语境（如虚拟听众的反应），主动调用已经有的知识、规范以及策略来实现动态改变，达成有效的意义协商。这种高频率、低风险、多样化的情境实践，促使学生将陈述性知识（知道是什么）转化为程序性知识（知道怎么做）与条件性知识（知道何时何地为何做），从而深化和重塑其认知结构，实现知识的社会化运用。

2.3 循证反馈支持下的元认知分阶段发展机制

这是模型实现能力内化与迁移的保障机制。元认知能力，就是对自己认知过程的自我调控能力，是区别高效学习者和普通学习者的关键。模型通过两阶段设计，促进元认知能力从“AI辅助监控”到“自主监控”的内化与迁移。第一阶段，AI给学生提供的即时反馈和全面评价就是一种“认知镜子”，促使学生建立起“优秀演讲”的认识标准，明白自身表现与标准的差距。第二阶段，经过了练习和反馈报告的积累之后，教师鼓励学生由依赖外界的评价转而依靠自己的分析去寻找数据背后的意义，并且开始有目的地去规划改进措施。学生能自行分析停顿源于紧张还是思路不畅，然后根据不同的情况去采取相应的练习方式。最后，学生即使离开AI环境，在稳定的自我反思、评价和调节的过程中也能形成稳定的内部循环，从而达成真正的自主学习和终身发展。

3 结论

在AI浪潮之下，对于高等教育来说，外语教育者要探索出新的技术驱动方式。研究提出的多模态AI驱动的教学模型，对AI是怎样用内容协同、情境模拟和循证反馈去支撑教学全过程做了细致论述，采用辩证协同、再语境化实践和元认知发展的机制，给解决传统教学痛点、提高人才培养质量赋予了可以操作的方法。始终坚守育人为本原则，在拥抱新技术的同时保障学习的深度和人的主体性，是AI时代下英语教学改革中的根本方向。

参考文献

- [1]徐昉. AI赋能英语读写认知发展的教学与研究框架构建[J]. 外语教育研究前沿, 2025, 8(6):3848.
- [2]刘邦奇, 胡健, 刘梦蝶, 等. AI赋能英语听说教一学一评一体化：模式、策略及效果验证[J]. 中国考试, 2025(1):4254.
- [3]焦建利, 陈婷. 大型语言模型赋能英语教学：四个场景[J]. 外语电化教学, 2023(2):1217.
- [4]吴坚豪, 周婉婷, 曹超. 生成式人工智能技术赋能口语教学的实证研究[J]. 中国电化教育, 2024(4):105 111.
- [5]庄雪茜. 跨学科视域下生成式人工智能赋能音乐生英语写作路径探究[J]. 现代英语, 2025(20):6971.
- [6]万子琪. 人工智能生成内容在构建个性化英语口语练习环境中的应用与挑战[J]. 现代英语, 2025(20):78 80.

基金项目：2025年云南省教育厅科学研究基金项目《大学英语AI+OBE教学新模式探索与实践》（立项编号：2025J1206）