

基于英语教学的口语自动评分系统应用的文献综述

Application of Automatic Oral Scoring System Based on English Teaching:

A Literature Review

胡晓文^{1*}, 陶晶^{1,2}, 付慧琳¹¹ 北京邮电大学人文学院² 北京邮电大学智能科技与语言教育研究中心

* 2024111670@bupt.cn

【摘要】 随着全球一体化的深入发展以及英语教育的广泛普及,英语口语教学逐渐成为教育界关注的焦点。然而,现行的口语评估方法仍存在诸多不足之处。本研究聚焦于语言智能技术支持下的英语口语自动评分系统,运用文献综述法梳理国内外相关研究,探索英语口语评分系统在英语教学实践中的运用,探索智能技术在英语教育领域中的应用模式与实践路径,为外语教育政策的科学制定提供依据,为推动外语教育向数字化、智能化方向的高质量发展提供理论支持与实践参考,助力外语教育在新时代背景下实现创新与突破。

【关键词】 口语自动评分; 英语教学; 自动语音识别; 多维评分标准

Abstract: With the development of globalization and the expansion of English education, oral English has garnered increasing attention across various sectors. However, traditional oral assessment methods have numerous shortcomings. Using the literature analysis approach, this research focuses on the automatic scoring system based on teaching and explores its priorities and applications. By integrating existing achievements, it aims to provide support for future research. The study indicates that this system has shown promising results and can be integrated into all aspects of foreign language teaching. Nonetheless, some issues remain. Looking ahead, the future development of this system is expected to advance further, bringing about significant changes in foreign language education.

Keywords: Automatic oral scoring, English teaching, Automatic speech recognition, Multi-dimensional scoring criteria

1. 引言

在当前全球化语境下,英语口语能力的重要性愈发显著。口语教学在培养学生语言运用能力方面扮演着核心角色,并且是增强学生跨文化交流能力的关键路径。然而,传统的口语教学与评估手段存在诸多缺陷,例如主观性较强、效率低下以及缺乏统一的评估标准。随着人工智能技术的迅猛发展,自动评分系统应运而生,为英语口语教学与评估带来了创新性的解决策略。本研究采用文献分析法,全面审视了口语自动评分系统在英语教学中的应用现状、技术特点、优势与局限性,并对其在多样化教学环境中的适应性与可行性进行了深入探讨。本研究的目的在于为教育工作者提供高效运用该系统的理论依据,为技术开发者提出改进建议,促进智能语言评估技术与教学实践的深度融合,以支持大学英语教育的改革进程。

2. 语言智能基础技术支持口语评价的原理剖析

语音识别技术,亦即自动语音识别(Automatic Speech Recognition, ASR),隶属于模式识别领域。其核心原理涉及对语音信号的处理与分析,旨在将识别和理解过程转化为可读的电子文本或控制指令。具体而言,ASR可划分为基于特征提取的方法和基于深度学习的方法。

基于特征提取的语音识别技术通过提取语音信号的特征，将其转化为特征向量，并运用模式识别算法进行分类。而基于深度学习的语音识别技术则采用神经网络模型对语音信号进行特征提取，并通过大量语音数据的训练，赋予模型以语音识别的能力（梁姗，2024）。

3. 口语自动评分系统研究

随着语音识别算法的持续优化，多种先进方法被集成至口语评分体系中，显著增强了系统对口语的理解与适应能力。近年来，多种技术的融合促进了英语口语自动评分系统的发展，使其成为学术界和教育界的研究焦点。目前，该领域的研究在国内外均取得了显著的进展。

3.1. 国外口语自动评分研究

在国际学术领域，针对口语评价系统的研究较早启动。近年来，国外研究者在英语口语自动评分领域的研究重点逐渐集中于评分模型的构建、语音特征的提取以及评估效度的论证等方面。表 1 总结了四项国外研究者关于口语自动评分系统的研究成果。其中，前三项研究主要聚焦于评分模型的构建，而四项研究均涉及语音特征的提取。

表 1 国外近几年有关口语自动评分研究概览

研究者	技术模型	应用领域	评分特征参数	语料数据
Wang, Y. et al (2018)	基于深度学习的 语音自动识别系 统	剑桥大学商务语 言测试服务在线 口语考试	音频特征和流畅 性特征	994 份训练集数 据, 226 份测试集 数据
Wang, X. et al. (2021)	基于 Transformer 的评分系统	非母语儿童的口 语能力评估	词汇复杂度、流 利度和内容匹配 度等	8738 份数据（涵 盖 18 个提示和 14 种母语背景）
Almadhor et al. (2023)	时空构音障碍语 音识别系统	构音障碍患者的 辅助沟通	音素序列	2179 份训练集数 据, 1085 份测试 集数据
Kang et al. (2024)	端到端 ASR 技术 和高效自动能力 评估	在线英语辅导学 习（AI Peng Talk）	语调、重音、节 奏、语速、停顿 和分段特征	120 个学生的 7545 个语音数据

在构建评分模型的过程中，研究者们广泛采用多种前沿技术以实现口语水平的自动化评估。例如，Wang 等人(2018)将深度学习驱动自动语音识别(ASR)技术与高斯过程(Gaussian Process)分级器相结合，显著提升了评分系统的准确性和可靠性。而 Wang 等人(2021)提出基于 Transformer 架构的自动评分模型，该模型基于儿童非母语自然口语的转录文本，有效评估其口语熟练度。此外，Almadhor 等人(2023)开发了一种端到端深度学习的 Dysarthric 自动语音识别(E2E-DASR)方法，为语言障碍者的语音识别系统提供了新的思路。

在表 1 中所列的研究中，均对语音特征参数的提取进行了深入探讨，凸显了精确特征提取与高效处理的重要性。研究者们针对现存问题持续探索创新，例如，Kang 等(2024)则将端到端语音识别模型与熟练度评估模型相结合，为特征提取技术的发展奠定了坚实基础。此外，Emberi 等(2022)基于声学特征的自动语音可懂度评分系统，通过深度神经网络(DNN)技术显著提高了评分的精确度。

在口语能力评估过程中，评估效度成为核心考量指标。Ngo 和 Lai(2023)进行的元分析研究揭示，ASR 技术在英语作为第二语言(ESL)及外语发音评估的有效性方面取得了显著进展，显示出其在促进学习者语音技能提升方面的巨大潜力。此外，有学者通过量化分析，研究主流 ASR 系统(例如 Google、IBM)针对不同种族说话者识别错误率的差异研究，揭示

了算法在种族口音识别上存在的系统性偏差。这一发现表明，技术偏见可能加剧教育评估中的不平等现象，迫切需要进行针对性的优化措施，以确保评估过程的公平性。

3.2 国内口语自动评分研究

我国在英语口语智能评分系统领域的研究虽起步较晚，但发展迅速，取得了显著成就。近年来，研究呈现多元化趋势，广泛涉及技术应用创新、测评维度拓展和系统性能优化等方面。

表 2 国内近几年有关口语自动评分研究概览

研究者	技术模型	应用领域	评分特征参数	语料数据
骆雁雁 (2023)	科大讯飞语音识别引擎结合 C++ 语言	高校英语专业学生口语测试	相似度特征、句法特征、语音特征和情感语音信号	46 名英语专业学生的语音数据
苏琴和付瑞吉 (2023)	基于深度学习的端到端自动英语评分系统	在线英语测试	已识别单词中的词汇线索和声学线索	训练集 2930 份，验证集 731 份，测试集 1827 份数据
梁姗 (2024)	多模态注意力融合网络架构	高校英语口语水平面试	口语韵律特征、文本多模态分析	约 8000 份数据
张晓艺等 (2024)	非端到端方案和端到端方案相融合的技术框架	CET-SET	语言(语法)、语篇、语用、内容(概念)	400 份样本数据 3 个任务模块

相较于表 1，表 2 中的国内研究在评分特征的复杂度与精确度上实现了显著的提升。研究的范畴不再局限于发音的准确性，而是扩展至语音流畅度、词汇运用、情感表达等多维度评分标准。例如，骆雁雁（2023）提出了一种基于多元线性回归模型的智能评价方法，融合语音识别文本的相似度、句法和语音特征，并引入情感识别技术，以实现全面客观的口语评价。苏琴和付瑞吉（2023）同样强调了情绪识别在未来研究中的重要性。然而，孙海洋（2021）指出，当前系统特征提取与考试构念结合不足，智能评分系统在评价层面仍有改进的空间。

国内学者对英语口语评分系统的效度评估表现出极大的学术关注。研究者们构建了基于现代效度理论的验证框架，从多个维度论证系统的效度，为其广泛应用提供了理论支撑。谢雪梅（2018）提出的基于语音信号频谱检测的发音错误检测方法，在发音精准评分方面展现了显著优势。吴坚豪等（2024）从复杂性、精确度和流畅度三个维度探讨了生成式人工智能对口语学习者的影响，为效能评估提供了新的视角。然而，张晓艺等（2024）指出，基于任务的人机协同口语评分存在趋中现象，原因在于训练集样本数量和覆盖率不足。这凸显了数据处理和模型训练方面的挑战。孙海洋（2021）认为，自动评分系统的效度验证和应用研究仍显不足，将是未来研究的重要方向。

4. 教育应用

在实际教学场景中，口语自动评分系统为教师和学生提供了具体指导和高效支持。例如，Speakable-AI 语言教学助手和 EAP Talk 系统针对教学实践，提供便捷、高效的解决方案，显著推动了教学应用。EAP Talk 由西交利物浦大学开发，是一款学术英语口语练习自动评估系统，包含朗读、演讲和教学管理等功能模块。其即时反馈、多口音模拟和考试对标功能备受认可。Speakable 则是一款支持多语言的自动评分工具，专为语言教师设计，可快速评分口语作业并提供即时反馈，覆盖口语、听力、写作和阅读等多个维度。两者在提升学生口语能力、支持自主学习和减轻教师负担方面均表现出显著优势。

5. 思考与总结

本研究对英语口语自动评分系统进行了全面的文献综述，系统梳理了其在外语教学领域的技术发展与实践成效。技术层面，口语自动评分系统借助语言智能技术实现了显著突破，但未来仍需优化算法以弥补现有不足。研究层面，国内外学者虽存在时序差异，但均聚焦于多模态特征融合与评估指标创新（如多维效度验证），以提升评分的科学性与可靠性。未来研究应聚焦于自动评分系统在教育领域的应用，关注其对教学和学习过程的影响。通过收集和分析教师与学生的反馈，不断优化系统功能，使其更好地服务于教育实践，并探索如何将自动评分系统与传统教学评价方法相结合，实现评价方式的多元化与互补性。一方面，系统融合多维度特征（如语音韵律、语义连贯性）以增强评估精准度，开发个性化适配机制以满足多元学习者需求；另一方面，通过跨学科协作（教育学、语言学、心理学），深入挖掘口语学习的本质与教学规律，使系统功能设计更贴合外语教学需求。通过规范系统研发、测评与应用，为全球外语教育注入活力，助力培养更多具备卓越口语能力的国际化人才。

基金

北京邮电大学 2024 年教育教学改革项目（项目号：2024ZD24）立项资助

参考文献

- 梁珊 (2024)。基于多模态注意力融合网络的英语口语自动评分。兵工自动化(08),18-22。
- 骆雁雁 (2023)。基于多特征融合的英语口语智能评价方法研究。外语电化教学(2), 49-55。
- 苏琴和付瑞吉 (2023)。基于深度学习的英语口语自动评分研究。信息技术(02),97-101。
- 孙海洋 (2021)。国内外英语口语自动评分研究综述。外语教育研究前沿(02),28-36+89-90。
- 吴坚豪、周婉婷和曹超 (2024)。生成式人工智能技术赋能口语教学的实证研究。中国电化教育(04),105-111。
- 谢雪梅(2018)。英语口语测试系统发音错误智能检测技术研究。自动化与仪器仪表(12),58-61。
- 张晓艺、王伟和杨浩然 (2024)。基于任务的人机协同口语评分效度论证：以 CET-SET 为例。外语界(02),71-80。
- Almadhor, A., Irfan, R., Gao, J., Saleem, N., & Kadry, S. (2023). E2E-DASR: End-to-end deep learning-based dysarthric automatic speech recognition. *Expert Systems with Applications*, 222, 119797 – 119797.
- Emberi, N. B., Schnoor, T. T., Wright, R. A., & Tucker, B. V. (2022). Acoustic-based automatic speech intelligibility scoring using deep neural networks. *The Journal of the Acoustical Society of America*, 152(4_Supplement), A287-A287.
- Kang, B. O., Jeon, H. B., & Lee, Y. K. (2024). AI - based language tutoring systems with end-to-end automatic speech recognition and proficiency evaluation. *ETRI Journal*, 46(1), 48-58.
- Ngo, T., & Lai, K. (2023). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL*, 36(1), 4 – 21.
- Wang, X., Evanini, K., Qian, Y., & Mulholland, M. (2021). Automated scoring of spontaneous speech from young learners of english using transformers. In *2021 IEEE spoken language technology workshop (SLT)* (pp. 705-712). IEEE.
- Wang, Y., Gales, M. J., Knill, K. M., Kyriakopoulos, K., Malinin, A., van Dalen, R. C., & Rashid, M. (2018). Towards automatic assessment of spontaneous spoken English. *Speech Communication*, 104, 47-56.