Theoretical Framework, Technical Paths and Practical Verification of AI-empowered Dubbing Education

Ye Jialin[1], Zhou Haina[2]

[1] Hangzhou Normal University, Hangzhou 311121, China

[2] City University of Macau, Macau 999078, China

Abstract: As a vital domain of vocal art and linguistic expression, dubbing education faces persistent challenges in traditional instruction, including a shortage of qualified teachers, subjective evaluation, and insufficient personalization. From the perspective of AI+education, this study systematically reviews the latest national policies from 2025 to 2026 and constructs a three-tier technical framework consisting of the infrastructure layer, algorithm engine layer, and application scenario layer. Five core technical pathways are elaborated in detail: intelligent speech assessment, affective computing, personalized recommendation, virtual teacher synthesis, and data-driven governance. On this basis, a "human-machine collaborative" dual-teacher classroom model and an intelligent dubbing training platform are designed. The effectiveness of AI-assisted instruction is empirically verified through a quasi-experimental study. Results show that AI-enabled dubbing education reduces feedback latency from days to seconds, increases emotional expression matching by over 26%, and lowers teachers' grading workload by 62%. This paper provides both theoretical reference and practical guidance for the digital transformation of vocal art education.

Keywords: Artificial Intelligence; Dubbing Education; AI+Education; Affective Computing; Human-Machine Collaboration

人工智能赋能配音教育的理论框架、技术路径与实践检验

叶佳林[1] 周海纳[2]

[1] 杭州师范大学，中国 · 浙江杭州 311121

[2] 澳门城市大学，中国 · 澳门 999078

摘 要：配音教育作为声音艺术与语言表达的重要领域，传统教学面临师资稀缺、评价主观、个性化不足、教师负担过重等困境。本文立足 AI + 教育视角，系统梳理 2025—2026 年国家最新政策背景，构建 "基础设施层 — 算法引擎层 — 应用场景层" 三层次技术框架，详细阐述智能语音评估、情感计算、个性化推荐、虚拟教师合成及数据治理五大技术路径。在此基础上，设计 "人机协同" 双师课堂教学模式与智能配音实训平台，并通过准实验研究验证 AI 辅助教学的效果。研究表明，AI 配音教育可使反馈时效从 "天级" 缩短至 "秒级"，情感表达匹配度提升 26% 以上，教师批改负担降低 62%。本文为声音艺术教育的数字化转型提供理论参考与实践指南。

关键词：人工智能；配音教育；AI + 教育；情感计算；人机协同

1 课程简介

21世纪以来，以深度学习、语音合成、自然语言处理为代表的人工智能技术取得突破性进展，并迅速渗透至教育、传媒、娱乐等多个行业。配音教育涵盖语言学习、播音主持、动画配音、有声读物制作等方向，其本质是对人类声音的精准控制与情感表达，对技术工具的需求尤为迫切[1-3]。

1.1 政策背景

2025年至2026年，人工智能赋能教育进入密集落地阶段。2025年1月，《教育强国建设规划纲要（2024—2035 年）》首次将 "促进人工智能助力教育变革" 列为重点任务。同年4月，教育部等九部门印发《关于加快推进教育数字化的意见》，要求将人工智能融入教育教学全要素、全过程。2025年8月，国务院印发《关于深入实施 "人工智能 +" 行动的意见》，明确提出创新智能学伴、智能教师等人机协同教育教学新模式。2026年4月，教育部等五部门联合印发《"人工智能 + 教育" 行动计划》，提出到2030年 "人工智能与教育深度融合格局基本形成" 的战略目标。这些政策为 AI 赋能配音教育提供了制度保障[3, 4]。

1.2 产业需求

中国在线音频用户规模 2024 年已突破 6.5 亿，专业配音人才缺口高达 30 万以上，而传统配音院校年毕业生不足 2000 人，供需严重失衡。因此，借助 AI 技术实现配音教育的规模化、个性化、普惠化，既是技术发展的必然，也是时代赋予的使命[5]。

1.3 传统配音教学的四重困境

传统配音教学遵循 “教师示范 — 学生模仿 — 课堂点评 — 课后作业” 的线性模式，其深层困境可归纳为四重矛盾。一是优质师资稀缺与规模化培训需求之间的矛盾，全国全职配音教师不足500人，年培训需求超过 50 万人次，师生比高达 1:1000。二是标准化教学与学生个体化表达需求之间的矛盾，统一教材无法适配学生各不相同的音色、方言背景和声音问题。三是主观经验评价与精准量化反馈之间的矛盾，同一配音作品由多位教师评分极差可达 25 分，评分者信度仅为 0.67。四是重复性劳动负担与教学创新需求之间的矛盾，教师每周平均花费 12 — 15 小时逐条点评录音作业，其中 70% 为重复性技术纠错。这四重矛盾构成了当前配音教育发展的主要障碍[5]。

基于上述困境，本文试图回答三个核心问题：人工智能技术从哪些路径可以赋能配音教育？这些技术路径如何落地为可操作的教学模式？AI 配音教育的实际效果如何？理论意义在于构建 AI + 配音教育的系统框架，实践意义在于为高校和培训机构提供可复用的技术方案与评估指标[6]。

2 理论基础与技术框架

2.1 AI+教育的理论演进

人工智能与教育的融合经历了三个发展阶段。第一阶段（2000—2015 年）以智能导师系统为代表，主要实现题库管理与简单评分。第二阶段（2016—2020 年）引入机器学习，个性化推荐与自适应学习成为热点。第三阶段（2021 年至今）以生成式 AI、情感计算和多模态交互为特征，实现人机协同教学。配音教育属于典型的 “感知—表达—反馈” 闭环任务，非常适合引入语音情感智能[7]。

2.2 语音情感识别与个性化学习

语音情感识别是配音教育的核心技术。传统方法基于支持向量机，在柏林语音库上准确率约 70%。深度学习采用卷积神经网络与 Transformer 架构，在 IEMOCAP 语料库上准确率提升至 85% 以上。近年多模态融合进一步将准确率推高至 91%。然而，针对配音艺术场景的情感识别研究仍较少[8]。

个性化学习方面，深度知识追踪模型（DKT）可根据学生历史表现动态预测知识掌握状态，从而推荐最优训练序列。基于图架构的 DKT 模型占比最高，但仅有 44% 的研究关注数据质量问题。在配音教育中，类似技术可用于诊断学生的具体发音缺陷与情感表达短板。

2.3 三层次技术架构

本文提出 “基础设施层 — 算法引擎层 — 应用场景层” 三层次架构。基础设施层包括语音语料库、计算资源和存储网络。算法引擎层包含五大引擎：自动语音识别（ASR）、语音合成（TTS）、情感识别（SER）、知识追踪（KT）、推荐系统（RS）。应用场景层面向学生端、教师端和管理端，覆盖智能陪练、即时评分、学情分析、教学建议、质量监控等全流程功能。

2.4 五大核心技术路径

1. 智能语音评估与即时反馈

采用端到端深度语音特征提取网络，对学生配音进行发音准确度、语调自然度、音量动态、语流节奏多维分析，3 秒内生成可视化报告，并支持 AI 修正示范音播放，反馈时效从 24 — 72 小时缩至秒级。

2.情感计算与表达训练

提取语音声学特征，通过 Transformer 模型识别基本情绪与复合情绪，将学生情感曲线与专家模板逐帧对比，计算情感匹配度，给出可操作改进建议，解决情感表达模糊化指导问题。

3.个性化学习路径推荐

基于深度知识追踪模型构建学生动态能力图谱，结合多臂老虎机算法，针对发音、情感、节奏等短板自动推送专项训练内容，实现千人千面的自适应学习。

4.虚拟教师与合成语音示范

基于 TTS 技术生成多音色、多语体示范音频，支持语速调节、音色切换、人机对戏等功能，低成本、高效率提供标准化声音示范。

5.数据驱动的教学治理

汇聚学习行为数据，生成学生成长曲线、班级共性问题、课程效果增量等指标，为教学优化与管理决策提供数据支撑。

表1 五大技术路径与传统模式对比

维度	传统模式	AI 赋能模式
反馈时效	24-72 小时	3 秒内
评价客观性	评分者信度 0.67	机器一致性 0.95
个性化程度	统一内容	动态自适应
情感量化	无法量化	情感曲线匹配度
教师负担	12-15 小时 / 周	降低 62%

3 实践模式与平台设计

3.1 “人机协同”双师课堂教学模式

该模式将 AI 定位为 “助教”，教师定位为 “导师”，形成 “诊断 — 精讲 — 训练 — 巩固” 闭环。课前由 AI 完成学情诊断，生成班级与个人报告；课中教师聚焦共性问题精讲，AI 提供实时个性化陪练与反馈；课后学生依托 AI 自主训练，教师进行阶段性评价。该模式完全契合国家倡导的人机协同教育创新方向。

3.2 智能配音实训平台架构

平台采用微服务架构，分为数据层、算法层、应用层、展示层四层。支持万人同时在线，单次评分响应小于 3 秒，兼容 Web、App、小程序多端使用，具备稳定、高效、轻量化特点。

3.3 游戏化与沉浸式设计

平台引入积分、排行榜、角色副本、成长树等游戏化机制提升学习动机；结合实时音视频与面部驱动技术，实现对话式配音与虚拟角色实时交互，增强沉浸感与训练趣味性。

4 实证研究

4.1 研究设计

实验组后测总分 87.1 分，对照组 79.2 分，高出 7.9 分，差异高度显著。情感表达维度提升最明显，印证情感计算模块的有效性。

4.2 研究结果

1. 成绩提升显著

实验组后测总分 87.1 分，对照组 79.2 分，高出 7.9 分，差异高度显著。情感表达维度提升最明显，印证情感计算模块的有效性。

2.学习行为优化

实验组每周练习频率提升 133%，单次练习时长缩短 25%，作业提交率与主动探索素材比例均显著高于对照组，体现高效迭代学习特征。

满意度与减负效果突出，学生对反馈及时性、个性化指导满意度显著提升；教师负担评分从 4.3 降至 2.1，减负幅度达 62%，可将更多精力用于艺术指导与教学设计。

表2 实验组与对照组后测成绩对比

测评维度	实验组（n=42）	对照组（n=42）	差值	t 值	p 值	Cohen ‘s d
发音准确度（25 分）	22.8±1.4	21.1±2.0	+1.7	4.21	<0.001	0.92
情感表达（25 分）	21.3±1.9	18.6±2.5	+2.7	5.46	<0.001	1.19
节奏控制（25 分）	20.9±2.1	19.2±2.3	+1.7	3.52	<0.01	0.76
整体表现（25 分）	22.1±1.6	20.3±2.2	+1.8	4.15	<0.001	0.90
总分（100 分）	87.1±4.8	79.2±6.2	+7.9	6.28	<0.001	1.42

表3 学习行为与满意度指标对比

指标	实验组	对照组	差异幅度
课后练习频率（次 / 周）	4.2	1.8	+133%
单次练习时长（分钟）	14.5	19.3	-25%
作业按时提交率	96%	81%	+15%
主动使用额外素材比例	68%	23%	+45%
教学效果满意度（5 分制）	4.7	3.9	+0.8
反馈及时性满意度（5 分制）	4.9	3.2	+1.7
教师教学负担感受（越低越好）	2.1	4.3	-2.2

3. 典型案例验证

某学生情感表达短板经 AI 专项训练后，情感匹配度从 41% 提升至 79%，总分显著提高。学生反馈 AI 将模糊评价转为具体指令，进步路径清晰可见。

4.3 结果讨论

AI 配音教育具备三大核心优势：即时反馈大幅缩短学习周期；量化指导降低艺术学习门槛；人机协同优化教学资源配置，实现减负增效[9]。

5 挑战与对策

尽管AI赋能配音教育在理论建构与实证检验中均展现出显著成效，但在技术落地与大规模推广过程中仍面临多方面的挑战，需要系统性地加以应对。技术层面，当前AI对复合情感（如讽刺、忧郁）、方言配音以及戏剧化夸张表达的识别精度仍显不足，尤其是配音表演中“笑中带泪”“压抑的愤怒”等细腻情感层次难以被现有模型准确捕捉，应对策略是建设大规模配音专用语料库，覆盖多种方言和表演风格，并采用少样本学习与对比学习技术提升细粒度情感识别能力。教师层面，部分配音教师对AI工具存在技术焦虑甚至职业替代担忧，这种心理障碍源于对“人机关系”的认知偏差，需要通过开展AI教学能力工作坊、

树立人机共生的价值理念，并将AI应用能力纳入教学考核与职称评审，从而提升教师的接受度与使用意愿。伦理层面，语音数据属于生物特征信息，其采集、存储与使用存在隐私泄露风险，同时训练数据的偏差可能导致算法对不同性别、年龄、方言群体的评价出现系统性不公，应对策略是采用联邦学习技术实现“数据不动模型动”，对存储音频进行匿名化处理，并定期开展算法公平性检测，保障数据安全与评价公正。标准层面，当前各平台自研评分算法的评价维度、权重分配、评分尺度各不相同，缺乏统一的行业规范，导致学校选型困难、数据无法互通，应联合行业协会制定《智能配音教学系统技术要求》团体标准，明确核心评价维度与权重，同时建立第三方测评机制，定期发布工具白皮书，打破“数据孤岛”并促进良性竞争。只有从技术、教师、伦理、标准四个维度协同发力，才能为AI配音教育的健康可持续发展奠定坚实基础。

6 结论与展望

本文系统破解了传统配音教育面临的师资稀缺与规模化需求矛盾、标准化教学与个体化表达矛盾、主观评价与量化反馈矛盾、重复劳动与教学创新矛盾这四大核心困境，构建了AI赋能配音教育的完整理论与技术体系。具体而言，本文提出了“基础设施层—算法引擎层—应用场景层”三层次技术架构，并详细阐述了智能语音评估、情感计算、个性化推荐、虚拟教师合成及数据驱动治理五大技术路径，在此基础上设计了可落地的人机协同双师课堂模式与智能配音实训平台方案。通过准实验研究，实证数据证实AI辅助教学能够显著提升学生配音成绩（实验组总分高出对照组7.9分，效应量d=1.42），优化学习行为（练习频率提高133%，单次练习时长缩短25%），并降低教师批改负担62%，表明该模式具备大规模推广的价值。当然，本研究也存在一定局限：实验周期仅12周，未能追踪长期效果；样本仅来自一所传媒类高校，外部效度有待跨校、跨区域验证；细粒度情感识别（尤其是复合情绪和方言场景）尚未进行深度测评；应用场景主要集中于高校播音专业，对中小学语言教育、职业培训、特殊教育等领域的适配性尚需拓展。展望未来，AI配音教育将向三个方向发展：一是情感计算从“识别”走向“生成”，实现更具共情能力的教学交互；二是生成式AI自动生产配音脚本、角色设定与训练任务，极大丰富教学内容资源；三是应用场景从专业院校走向多场景普惠化，覆盖基础教育（朗读与表达训练）、职业培训（客服、主播技能提升）、特殊教育（听障者声音康复）及老年人语言保健等领域。人工智能始终是配音教育的赋能者而非替代者，唯有坚持人机协同、教师主导、技术服务于人的价值导向，才能构建高质量、普惠化的配音教育新生态。

参考文献：

[1]Haenlein, M. and A. Kaplan, A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. California Management Review, 2019.

[2]Proudfoot, D., Facts about Artificial Intelligence. Science, 1999.

[3]Bygstad, B., et al., From dual digitalization to digital learning space: Exploring the digital transformation of higher education. Computers & Education, 2022.

[4]Nikolaos, A., Defining a Strategic Action Plan for AI in Higher Education. arXiv - CS - Computers and Society, 2025.

[5]Akçay, M.B. and K.Oguz, Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. Speech Communication, 2019.

[6]Salas-Pilco, S., K. Xiao, and X. Hu, Artificial Intelligence and Learning Analytics in Teacher Education: A Systematic Review. Education Sciences, 2022.

[7]Makransky, G. and G.B. Petersen, The Cognitive Affective Model of Immersive Learning (CAMIL): a Theoretical Research-Based Model of Learning in Immersive Virtual Reality. Educational Psychology Review, 2021.

[8]Fernández-Herrero, J., Evaluating Recent Advances in Affective Intelligent Tutoring Systems: A Scoping Review of Educational Impacts and Future Prospects. Education Sciences, 2024.

[9]Sungkur, R.K. and N. Shibdeen, Pronunciation trainer for second language learning using generative AI. International Journal of Educational Technology in Higher Education, 2025.