

doi: 10.3969/j.issn.1672-4933.2026.03.005

人工智能技术在听觉言语康复中的应用

Applications of Artificial Intelligence in Auditory and Speech Rehabilitation

蒋春 薛静 赵倩 李瑶 张艳玲 孙静 于丽玫 邢亚静

JIANG Chun, XUE Jing, ZHAO Qian, LI Yao, ZHANG Yan-ling, SUN Jing, YU Li-mei, XING Ya-jing

【摘要】 听障人群听觉言语康复面临康复资源分布不均、康复费用高昂等问题,这种现状制约了其对于高质量康复需求的发展,尤其是传统的康复手段不能满足成人听障者对听觉言语的康复需求。人工智能(artificial intelligence, AI)技术为解决该困境提供了新方法。本文总结了AI在此领域的技术演进,从传统信号处理、机器学习、深度学习到大模型与临床转化阶段,介绍了其核心技术架构中的前端感知、中间处理及后端应用。目前AI在语音增强、效果预测、多模态融合及个性化康复方案生成等方面取得显著进展,但仍有数据稀缺与异质性、模型泛化能力不足、临床可解释性差及转化障碍等问题。未来需加强大规模多模态数据库建设,提高模型的分布与泛化能力,促进AI工具与临床的融合,以实现听障人群听觉言语康复的个性化、智能化。

【关键词】 人工智能;听力障碍;言语康复;语音增强;临床转化

【中图分类号】 R494

【文献标识码】 A

【文章编号】 1672-4933(2026)03-0267-05

【Abstract】 Auditory and speech rehabilitation for hearing-impaired population faces challenges such as uneven distribution of rehabilitation resources and high costs, which constrain the development of high-quality rehabilitation services. In particular, traditional rehabilitation methods fail to meet the auditory and speech rehabilitation needs of adults with hearing loss. The emergence of artificial intelligence (AI) provides a new way to solve these troubles in auditory and speech rehabilitation for the hearing-impaired population. This paper review the technological evolution of AI in this field, from traditional signal processing and machine learning to deep learning, large models, and clinical translation. Introduce the core technical architecture, including front-end perception, mid-level processing and back-end application. Now, AI has made significant progress in speech enhancement, outcome prediction, multimodal fusion and personalized rehabilitation plan generation. But, There are still difficulties remain, such as data scarcity and heterogeneity, insufficient model generalization capability, lack clinical interpretability and barriers to translation. Future works should focus on constructing large-scale multimodal databases, better out-of-distribution generalization of models and strengthen the integration of AI tools into clinical practice to make personalized and intelligent in auditory and speech rehabilitation for the hearing-impaired population.

【Key words】 Artificial intelligence; Hearing impaired; Speech rehabilitation; Speech enhancement; Clinical translation

1 研究背景

根据世界卫生组织(WHO)2023年发布的世界听力报告^[1],全球约有15亿人存在不同程度的听力损失,其中4.3亿人需要康复服务,听力障碍群体在语言表达、信息获取、社交互动等方面面临显著障碍,其教育机会、就业选择与生活质量均受到不同程度的影响。目前,听障人群听觉言语康复存在以下挑战:①专业人才结构性失衡和地域分布不均衡。我国听觉言语康复从业人员缺乏高学历、高职称、高年资人才,继续教育培训亦存在明显不足;虽然已覆盖全国大部分地区,但资源高度集中于省会

城市及中国听力语言康复研究中心等国家级/省级中心,基层及偏远地区可及性差^[2,3]。②听障群体的康复要求随年龄增长动态变化,受干预年龄、康复质量、经济状况等因素的影响。基层听障儿童成年后仍然存在口齿不清、言语可懂度差等问题。③目前国内康复服务体系主要聚焦于0~6岁听障儿童早期干预,成年听障者受家庭、工作等各种因素影响不适于采用长时间的康复训练。

随着人工智能(artificial intelligence, AI)技术在众多领域应用取得了显著进展,为解决以上挑战提供了全新的可行路径。本文就其在听觉言语康复中的应用予以系

基金项目:2024年中国残联课题“人工智能声音修复技术在听障人士病理腔调中的修复”(2024CDPFHS-10)

作者单位:中国听力语言康复科学杂志社 北京 100029

作者简介:蒋春 博士;研究方向:神经生物学

通讯作者:于丽玫, E-mail: limeiyu@vip.sina.com.

统梳理。

2 技术演进脉络

2.1 传统信号处理阶段(2010年前)

助听辅具基于谱减法、维纳滤波等算法仅对稳态噪声有效,对非稳态语音噪声效果有限^[4]。

2.2 机器学习阶段(2010~2018年)

传统的机器学习算法进入听力学领域,研究以小样本、单一任务的可行性验证为主,算法基本为随机森林(random forest)、决策树(decision trees)、支持向量机(support vector machine, SVM)等传统机器学习模型。随机森林和决策树算法在人工耳蜗术后效果预测准确率已达96%,贝叶斯线性回归达到96.2%,极限学习机(extreme learning machine, ELM)高达99%,但绝大多数研究都基于小样本数据集^[5]。

应用层面研究主要集中在2个方向:①以人工耳蜗植入术前临床变量(耳聋持续时间、植入年龄、残余听力水平)为基础的术后言语感知效果预测;②基础声学信号处理,即人工耳蜗编码策略的优化。该阶段重要、明确的发现是传统线性回归模型只能解释术后言语感知效果变异约22%,因此尚不能满足临床精准预测的实际需要。相关研究普遍存在样本量小(多数 $n < 200$)、无独立验证集、算法报告不充分等问题,故许多预测模型在数据集特征、模型构建、验证等方面报告不充分,过拟合程度不明确,对新数据的泛化能力存疑^[6]。

2.3 深度学习与多任务探索阶段(2018~2022年)

该时期深度学习技术的成熟,推动了听力学领域的加速发展。从传统机器学习转向深度学习范式,从单一预测任务转向多任务、多模态。

2.3.1 技术突破 卷积神经网络(convolutional neural network, CNN)和循环神经网络(recurrent neural network, RNN)开始应用于语音信号处理和手语识别。该时期手语识别、语音识别、文本到语音合成三大应用方向均取得重要进展^[7]。深度学习模型能自然、合理地学习声学信号的层次化特征,故在复杂听觉任务中表现优于传统方法。

2.3.2 数据集建设 研究者意识到数据规模与多样性对模型性能的关键作用。在机器学习应用中,大量数据、数据多样性及长训练时间是模型获得强泛化能力与高预测精度的重要因素^[5]。然而,听障者语音数据的采集成本高、标注困难,因此,大规模数据集建设仍受此限制。

2.3.3 应用场景拓展 除了术后效果预测,深度学习已在以下方向得到应用:基于自动语音识别(automatic speech recognition, ASR)的儿童构音障碍评估、噪声环境下的言语可懂度增强,以及AI驱动远程康复平台。AI-

Voice Therapy平台是该领域典型的创新尝试,其使用Open AI的Whisper模型自动、客观地检测口吃、失语症、构音障碍等,据此自动生成个性化康复训练方案,其移动端应用、表述性状态传递应用程序编程接口(representational state transfer application programming interface, REST API)、数据库3层架构设计自然,是技术从实验室走向实际应用示范^[8]。

2.4 大模型、系统综述与临床转化阶段(2022年~至今)

该阶段是该领域反思性深化阶段,研究者已不满足于单纯算法的准确率,有意识、有逻辑地考察技术转化临床的障碍、模型泛化的限制及技术与临床效果的落差。

2.4.1 系统综述的出现 这一时期出现了很多高质量系统综述对所涉领域的研究现状进行批判性考察。虽然机器学习预测模型的各项评估指标良好,但大部分研究在数据集特征、模型构造、方法验证等方面报告不充分,过拟合问题常见^[9,10]。Demirtaş^[5]分析了59篇符合纳入标准的研究,指出深度学习在处理噪声中言语理解等需要时间序列处理的复杂问题时仍需更多研究。

2.4.2 大模型的应用 预训练大模型(如OpenAI Whisper)开始被应用于听障人群的语音评估。Lachguer等^[8]研究展示Whisper模型在自动检测多种语音障碍类型(口吃、失语症、构音障碍、言语失用症)的潜力,并能够根据障碍类型和严重程度生成个性化训练方案。该研究提示通用ASR模型对听障/构音障碍儿童的识别表现显著下降。

2.4.3 关键转折 Shew等^[6]的大规模研究($n=1877$)是标志性成果。XGBoost模型用于预测人工耳蜗术后6个月言语感知效果时,统计优于逻辑回归,但其约1%~2%的性能提高临床上无实质意义。

2.4.4 技术演进与临床实际的脱节 目前,用深度学习设计语音增强算法多在训练集、测试集同源条件下评价性能^[4,11],而未见说话人或噪声类型时性能骤降,因此,大量算法至今仍停留在研发阶段,没有真正进入临床应用^[12,13]。

3 核心技术架构

人工智能技术在听障者听力和言语康复领域应用的核心技术架构分为前端感知层、中间处理层及后端应用层。

3.1 前端感知层

前端感知层是整个技术架构的基础,负责从听障者及其所处环境中采集原始信号,完成初步的特征提取与质量评估。

3.1.1 声学信号采集 传统助听器和人工耳蜗依赖单

麦克风或多麦克风阵列采集环境声学信号。单麦克风算法基于时均信号统计和噪声假设仅能有效处理可预测的平稳噪声,而对多人交谈等非平稳噪声无效。多麦克风阵列采集虽能利用空间信息,但需目标声源与噪声空间分离,且对混响及设备佩戴角度敏感^[4]。

3.1.2 视觉信号采集 为弥补声学信号不足,多模态架构引入了视觉通道。HI-TransPA 等系统采用两阶段唇部区域提取相关信息^[14]。

3.1.3 神经生理信号采集 非侵入式脑机接口技术利用脑电图(electroencephalogram, EEG)、功能性近红外光谱(functional near-infrared spectroscopy, fNIRS)和脑磁图(magnetoencephalography, MEG)捕获神经信号。然而,EEG信号与言语处理过程相关的成分常被噪声淹没,信噪比通常低于1:10,空间分辨率限制在2~3 cm,难以区分相邻皮层区域的神经活动^[15]。

3.1.4 数据质量评估与筛选 由于原始数据存在噪声、异质性问题,先进架构引入了数据质量评估机制,HI-TransPA设计了一套完整、严谨的预处理及筛选流程:先检测面部地标,再分离、稳定唇部区域,定量评价多模态样本质量,最后根据所得质量评分确定课程学习策略,即先用高质量样本训练,再适当引入难以拟合的样本^[14]。

3.2 中间处理层

中间处理层是技术架构的核心,负责将前端采集的多模态信号转化为可用的智能服务。

3.2.1 语音增强与分离 从循环神经网络(recurrent neural network, RNN)到Transformer深度神经网络在语音增强领域取得了突破性进展。Borjigin等^[4]实现了循环神经网络和基于Transformer的分离模型两种深度神经网络算法,在约30小时定制数据集上训练后,对13名人工耳蜗使用者进行测试。结果显示传统算法的局限、深度神经网络(deep neural network, DNN)算法的优势、RNN架构的局限性及Transformer架构的优点。

3.2.2 多模态融合 多模态融合结合声学、视觉信息,以提高系统的鲁棒性及泛化能力。目前该领域有代表性的是HI-TransPA,其将Omni-Model范式引入听障辅助技术,设计指令驱动的视听个人助手,模型通过融合不清晰语音与高帧率唇部动态,在统一多模态框架内解决了语音翻译、对话理解等问题^[14]。

3.2.3 效果预测模型 从线性回归到XGBoost机器学习在人工耳蜗术后效果预测中的应用,体现了技术架构从单一生物标志物向多维预测的演进。Shew等^[6]大规模研究(n=1877)评估了机器学习预测人工耳蜗术后6个月言语感知表现的能力。结果显示,使用XGBoost模型预测辅音-核心-辅音单词的错误率为17.4%,显著优于线性

回归的18.36%;亚利桑那生物医学研究所句子测试(arizona biomedical institute sentences, AzBio)的错误率为20.39%,优于线性回归的21.62%。在预测第20百分位水平(识别表现最差的20%人群)时,XGBoost的受试者工作特征曲线下面积(area under the receiver operating characteristic curve, AUROC)达0.708,显著优于逻辑回归的0.594,能够有效识别表现极端的类型。因此,今后宜增加解释效果的新变量(如神经生物学标志物、认知评估指标),并完善数据协作、标准化注册体系。

3.2.4 神经信号解码与闭环调控 人工智能正与非侵入性脑刺激技术融合,构建双向通信神经假体框架。将经颅直流电刺激(transcranial direct current stimulation, tDCS)和经颅交流电刺激(transcranial alternating current stimulation, tACS)与AI算法结合,实现对言语产生和听觉处理功能的个性化、自适应增强。然而,非侵入式脑机接口(brain-computer interface, BCI)目前存在明显的技术限制,导致其性能难以满足实际应用需求,与理论预期或临床需求之间存在较大差距。其信息传输率仅为0.5~3词/min,远低于自然语速120~180词/min;基于EEG的系统仅能解码50~100个词汇且准确率有限;跨被试迁移学习性能仅达到个体特异性模型的40%~60%^[15]。

3.3 后端应用层

后端应用层将中间处理层产生的智能结果转化为具体的临床应用和康复服务。

3.3.1 个性化康复方案生成 基于预测模型的输出,AI可为每位用户生成个性化的康复方案。AI-VoiceTherapy平台使用Whisper模型自动检测和分析语音障碍(口吃、失语症、构音障碍等),据此制订个性化康复训练方案^[14]。

3.3.2 临床决策支持 机器学习模型通过识别极端表现类型,为临床决策提供依据。

3.3.3 持续监测与反馈 基于AI技术的实时生物反馈系统打破了传统言语康复的机构中心化模式,实现了家庭场景下的自主化、高频次训练,促进了康复成果从受控实验环境向复杂日常交际情境的转化。

3.3.4 面向成人听障者声音修复技术应用 以华为小艺声音修复技术为代表的AI应用,是面向成年听障口语者的典型创新实践。该技术基于大规模成人听障者真实发音数据库,训练专属语音大模型,可在保留用户原有音色的前提下,对发音缺陷(如声调不准、辅音省略、语流中断)进行实时分析与修复,将语音可懂度平均提升80%以上。该技术已连续两年入选中国残联助残科技创新案例,为因错过最佳干预时期而导致口语不清的成年听障者提供了全新的沟通解决方案,显著增强了其社会参与自信。此类技术的核心价值在于其不是将听障者的声音

变成别人的,而是修复成自己本该有的样子。

4 挑战与局限

人工智能技术在听障康复领域已取得阶段性进展,但仍存在很多问题。

4.1 数据与算法的限制

听障者的异质性与数据采集的高成本,使构建大规模、高质量数据库非常困难,导致模型的分布脆弱性(distributional fragility),在受控训练环境下表现优异,但面对真实世界中复杂的噪声混叠、个体差异时,模型的分布外泛化能力(out of distribution generalization)显著降低^[6,16,17]。此外,当前研究以听障儿童为主要对象,针对成人听障者的大规模、高质量语音数据库仍然稀缺。成人的发音模式已固化、个体差异更大(病因、年龄、耳聋时长、助听设备类型等异质性强),对模型的分布外泛化能力提出更高要求。这导致许多在儿童数据上表现优秀的算法,迁移到成人场景时性能骤降,亟需建设面向成人听障者的专用语料库。

数据稀缺制约了技术发展及临床应用。深度学习的“黑箱”特性与临床决策对可解释性的要求存在冲突。临床医生难以理解和信任无法提供明确推理依据和可追责性的预测结果,限制了AI辅助诊断系统的应用^[6]。模型的预测能力受限于输入特征的信息含量。现有临床变量(如年龄、耳聋时长)对术后效果的解释力有限^[16-19]。

4.2 临床转化与整合的障碍

研究报告中模型性能的显著提升仅具有统计学意义,而非患者可感知的、有临床价值的实质性改善^[20]。许多算法在受控实验环境下有效,但难以适应临床的多因素耦合与时变非稳态特性^[21]。多数AI工具被设计为全自动系统,忽略了临床中多角色(治疗师、患者、家属)的合作^[17]。工具的可用性、专业人员的接受度与相关培训的缺乏,导致技术难以融入现有诊疗体系。

4.3 公平、伦理与技术发展的矛盾

训练数据若无法充分兼顾听障者内部的多样性,AI系统可能对少数群体产生系统性偏差,加剧服务的不平等^[22];康复数据的高敏感性与技术发展需要大规模数据协作间存在矛盾;技术本身的价值判断(如以正常语音为目标)需考虑其对社会多元性的影响^[22]。

5 小结

在科技助残大形势下,人工智能在听觉言语康复领域的应用已取得了一定进展。中国听力语言康复科学杂志社通过采集4926名听障口语者的声音和收集调查问卷,已经建立了有效声音时长1023小时的语料数据库及

成年听障者康复状况数据库。这些数据库为AI助残提供了宝贵的研究资料。

AI技术正在从辅助工具向智能康复伙伴演进,在降低康复成本、提升训练可及性、优化人工耳蜗效果预测等方面展现出巨大价值。随着大模型、多模态融合和边缘计算的发展,听障者的听觉言语康复将进入更加个性化、智能化和覆盖全生命周期,既持续巩固0~6岁儿童早期干预成果,也对错失黄金期的成年听障者提供修复型康复服务的新阶段。

参考文献

- [1] World Health Organization. World Report on Hearing[R]. Geneva: WHO, 2023.
- [2] 王霞,梁巍,卢晓月,等.我国听觉口语法发展及听觉口语师队伍建设现状[J].中国听力语言康复科学杂志,2026,24(2):215-218.
- [3] 陈仁吉.我国言语障碍康复治疗发展历程与现状[J].中国听力语言康复科学杂志,2025,23(4):348-351.
- [4] Borjigi A, Kokkinakis K, Bharadwaj HM, et al. Deep learning restores speech intelligibility in multi-talker interference for cochlear implant users[J]. Sci Rep, 2024, 14: 13241.
- [5] Demirtaş YM. Prediction of Auditory Performance in Cochlear Implants Using Machine Learning Methods: A Systematic Review[J]. Audiology Research, 2025, 15(3): 56-56.
- [6] Shew MA, Pavelchek C, Michelson A, et al. Machine Learning Feasibility in Cochlear Implant Speech Perception Outcomes-Moving Beyond Single Biomarkers for Cochlear Implant Performance Prediction[J]. Ear Hear, 2025, 46(5): 1266-1281.
- [7] Zaineldin H, GameL SA, Talaat FM, et al. Silent no more: A comprehensive review of artificial intelligence, deep learning, and machine learning in facilitating deaf and mute communication[J]. Artif Intell Rev, 2024, 57(7): 188-188.
- [8] Iachguer N, Azizi O, EL Mamoune S. AI-VoiceTherapy: An Automated Platform for Voice Rehabilitation Using Artificial Intelligence[J]. Int J Comput Eng Data Sci, 2025, 92(55): 463-463.
- [9] Mo JT, Chong DS, Sun C, et al. Machine-Learning Predictions of Cochlear Implant Functional Outcomes: A Systematic Review[J]. Ear Hear, 2025, 46(4): 952-962.
- [10] Nossier SA, Wall J, Moniri M, et al. An Experimental Analysis of Deep Learning Architectures for Supervised Speech Enhancement[J]. Electronics, 2021, 10(1): 17-17.
- [11] Pandey A, Wang D. On Cross-Corpus Generalization of Deep Learning Based Speech Enhancement[J]. IEEE/ACM Trans Audio Speech Lang Process, 2020, 28: 2489-2499.
- [12] Mawalim CO, Okada S, Unoki M. Are Recent Deep Learning-Based Speech Enhancement Methods Ready to Confront Real-World Noisy Environments[C]. Proceedings of Interspeech 2024. 2024.
- [13] Behera A, Easow RA, Parvathala V, et al. Test-Time Training for Speech Enhancement[C]. Proceedings of Interspeech 2025. 2025.
- [14] Zhang H, Wang Y, Liu J, et al. HI-TransPA: Hearing Impairments Translation Personal Assistant[C]. 2025 IEEE International Conference on Multimedia & Expo (ICME), 2025. 1-6.

(下转248页)