人工智能在音频信号处理中的应用问题与解决策略

曹一凡
中国农业大学，北京 100083

DOI: 10.61369/TACS.2026070024

摘　　要： 人工智能通过智能算法和模型处理音频信号，能够实现音频增强、识别及转换等功能。当前，AI 语音合成、语音识别、音频增强、声纹识别等技术，在文娱创作、新闻传媒、实时通信、工业检测等已经广泛应用，可以有效降低行业成本、提升工作效率。然而，在技术快速普及的同时，技术性能缺陷、版权侵权、用户语音隐私泄露、虚假语音传播等问题也逐渐显现。本文结合当下行业实际情况，梳理人工智能在音频信号处理领域的主流应用，剖析各类问题的具体表现，并探讨具体的解决策略，期望能够规范市场秩序，推动人工智能音频信号处理技术的进步。

关 键 词： 人工智能；音频信号处理；语音合成；声纹识别；版权保护；隐私安全

Problems and Solutions in the Application of Artificial Intelligence in Audio Signal Processing

Cao Yifan
China Agricultural University, Beijing 100083

Abstract: Artificial intelligence (AI) processes audio signals through intelligent algorithms and models, enabling functions such as audio enhancement, recognition, and conversion. Currently, AI technologies including speech synthesis, speech recognition, audio enhancement, and voiceprint recognition have been widely applied in fields such as entertainment creation, news media, real-time communication, and industrial testing, which can effectively reduce industry costs and improve work efficiency. However, with the rapid popularization of the technology, problems such as technical performance defects, copyright infringement, leakage of users' voice privacy, and dissemination of fake voices have gradually emerged. Based on the current actual situation of the industry, this paper sorts out the mainstream applications of AI in the field of audio signal processing, analyzes the specific manifestations of various problems, and discusses specific solutions. It is expected to regulate the market order and promote the progress of AI-based audio signal processing technology.

Keywords: artificial intelligence (AI); audio signal processing; speech synthesis; voiceprint recognition; copyright protection; privacy security

引言

数字化时代下，音频技术已经广泛应用于有声读物、网络音乐、短视频、新闻播报、智能家居等场景。随着人工智能、深度学习与 AIGC 技术不断发展，音频信号处理迎来全面变革。文本转语音、声纹识别、音频增强等 AI 技术，可以简化音频制作、语音交互、噪声处理等工作流程，丰富音频内容的形式与应用场景 [1]。但在技术快速落地的过程中，各类隐患也逐步暴露，这不仅扰乱行业正常秩序，还会损害普通用户的合法权益。本文结合行业发展现状，梳理人工智能在音频信号处理领域的核心应用，分析现存各类问题，并提出针对性解决策略，为该行业良性发展提供参考。

一、人工智能在音频信号处理中的主流应用

（一）语音合成，赋能多元音频内容生产

语音合成是 AI 音频领域应用最广泛的技术，依托文本转语音（TTS）模型可以将文字转化为人声音频。如今市面上主流的网络阅读软件，如番茄小说、七猫、掌阅等软件都上线了 AI 朗读功能。用户开启听书模式后，可以根据小说题材挑选人声，搭配背景音乐，自由调节朗读速度。部分产品还支持多角色对话功能，能够优化用户听书体验 [2]。

不少头部音频平台开始尝试 AI 声线复刻，喜马拉雅推出“单

 Copyright © This Work is Licensed under A Commons Attribution-Non Commercial 4.0 International License.

田芳原音重现”项目，用 AI 模拟评书大师的嗓音，重新演绎大师生前没能完成的作品。平台还根据特色声线打造多款新的有声内容。这项技术让有声内容的创作门槛明显降低，普通创作者也能制作专业有声内容。数字音乐行业也广泛使用声线克隆技术，网络热门的“AI 孙燕姿”就是通过技术复刻歌手原本的音色，演绎了大量不同风格的翻唱曲目；酷狗也曾借助 AI 还原邓丽君声线，推出全新单曲。人工智能走进音乐制作领域后，改变了传统的制作流程，创作者的创意空间进一步被打开，整个音乐行业的创作形式也更加多样。

短视频、影视和游戏领域也将 AI 配音作为内容制作的常用方式。抖音“一键成片”功能可自动生成 AI 画外音，降低短视频创作难度。纪录片、动画可以用 AI 录制解说人声、现场环境音效，部分海外动画的中文角色台词也由 AI 录制。手游行业遇到真人配音无法继续工作的情况时，制作团队就会启用 AI 语音承接后续的配音工作，保障项目正常推进。

（二）语音识别与智能播报，革新传媒与人机交互

语音识别可将语音转化为文字，与语音合成相互结合后，可打造出智能播报系统，重塑传媒行业运作模式。腾讯新闻运用语音合成技术自动生成播报音频，投放到多个平台。受众可以随时随地收听新闻内容，团队也不需要安排专人长期负责语音播报工作，节约人工播报成本。国内多家广播电视平台也布局虚拟 AI 主播赛道，如北京卫视数字主播“时间小妮”高度还原真人主持人的语音与形象，安徽广播电视台虚拟主持人“小安”也多次亮相大型活动，成为全新传播形式。除传媒领域，语音识别也在融入大众日常生活。手机语音助手、智能音箱、车载设备都依靠该技术解读用户指令。

（三）音频增强与声纹识别，服务通信与安全领域

音频增强技术主要用于过滤环境噪声，提升语音清晰度。基于深度学习的 WaveCRN 等模型，能有效抑制街道、咖啡厅、商场等场景的杂音，在复杂环境中提升语音信噪比 [3]。目前这项技术广泛应用于手机通话、线上会议、网络直播等场景，改善实时音频的传输质量与听觉体验。声纹识别根据人体独有声音特征完成身份核验，常搭配语音唤醒功能使用 [4]。智能家居利用声纹区分家庭成员，分配设备使用权限。广播电视领域依靠声纹识别开展跨频点节目监测，排查非法转播行为。经过优化的深度学习声纹模型，在噪声环境下依旧保持高识别率。语音唤醒功能则可以简化智能设备的操作方式，提升人机交互的便捷性。

（四）环境声检测与声学仿真，拓展工业应用场景

AI 可以对环境声音进行检测和分类。城市治理中，系统能自动识别交通、施工噪音，辅助监管开展降噪工作。智能安防设备可捕捉玻璃破碎、紧急警报等异常声响，及时发出预警。工业场景下，工作人员通过 AI 分析机械运行声音，判断设备故障，提前规避生产风险。

传统声学结构设计需要反复仿真测试，效率偏低。AI 声学仿真技术可以快速模拟声学效果，优化建筑隔音材料、声学超材料、音响设备的结构。如今该技术已应用在建筑工程、交通降噪、音响制造等行业，可以有效提升声学产品的设计效率。

二、人工智能在音频信号处理中存在的突出问题

（一）技术层面存在短板，实际体验大打折扣

当前，AI 音频技术还不够完善，存在一些技术缺陷。多数文本转语音模型容易出现多音字读错、语句断句生硬等问题，遇到专业术语、生僻人名地名时，出错的情况可能会更多。机器语言也很难还原人类丰富的情感，表达抒情内容、剧情对白时语音语调僵硬失真，甚至会误导儿童、老年人等群体。AI 音频模型均基于特定场景的数据完成训练，一旦遇到方言、强噪声等陌生环境，语音识别、音频增强等功能的性能会下降 [5]。而且，高精度深度学习模型对运算能力有较高要求，手机、智能音箱等小型终端设备无法承载。若使用轻量化模型，音频处理质量会下降；若强行运行复杂模型，又会出现响应延迟，无法满足实时通话、语音唤醒等场景的要求 [6]。

（二）声音权与版权侵权泛滥，监管难度较大

AI 语音克隆的门槛不断降低，声音侵权问题日益严重。声音属于自然人的人格权益，但我国暂时还没有针对 AI 声音克隆的专项法律条款。市场中有些个人与结构，在没有取得当事人许可的前提下，克隆歌手、主播、公众人物的声线，制作 AI 翻唱、虚假配音并在网络传播 [7]。短视频领域的 AI 配音也普遍存在声线来源不明、授权缺失的问题。

AIGC 音频的著作权界定也十分模糊。目前，法律层面和行业内部，都没有针对这类内容的版权归属形成统一标准。部分平台利用 AI 复刻经验音频内容并二次传播，原版版权方维权困难。加上，网络上存在大量免费的语音克隆工具，侵权成本低。监管部门想要追查侵权发起发者也是困难重重，各类侵权乱象很难得到彻底整治。

（三）语音隐私泄露风险突出，威胁用户安全

AI 模型训练需要手机大量语音数据，很多智能音箱、手机 APP、智能家居设备，会在用户不知情时收录日常对话、语音指令。语音数据包中收录用户生活习惯、私密谈话等隐私，而声纹属于独一无二的生物识别信息，泄露后危害极大 [8]。

一般的中小型企业的数据安全防护能力弱，数据库容易遭到黑客攻击，造成用户语音、声纹数据被盗。还有不法分子在网络倒卖语音数据牟利。不法分子拿到声纹后，会伪造语音实施电信诈骗、冒用身份，威胁用户的财产与人身安全。

（四）虚假语音与对抗攻击，引发伦理和公共风险

AI 克隆技术可以制作高度逼真的虚假语音，如有人伪造公众人物在网络散布谣言、诈骗财物，这不仅损害当事人名誉，还会扰乱网络秩序。对抗性攻击也是一大隐患。攻击者在正常音频中加入人耳无法感知的微小对抗扰动，干扰语音识别、智能播报系统。新闻、安防等关键系统一旦被攻击，会输出错误信息，造成不良社会影响。

三、人工智能在音频信号处理领域现存问题的解决策略

（一）迭代核心技术，补齐技术短板

企业与科研机构需要联合优化 AI 音频模型。扩充高质量的语

Copyright © This Work is Licensed under A Commons Attribution-Non Commercial 4.0 International License.

音数据库，完善多音字、专业词汇的发音规则，结合情感相关技术调整人声的语调与节奏，让机器生成的语言更具准确，同时提升情感表现力。模型训练阶段可以融入不同地域的口音素材、生活中常见的各类环境杂音，让模型逐步适应真实环境里的复杂声音条件[9]。不同终端设备的运算条件存在明显区别，行业可以转向轻量化的深度学习框架，推广边缘计算技术，将一部分音频处理工作放在终端完成，以此改善数据传输过程中出现的延迟问题。行业组织可以牵头整合业内资源，搭建开源标准化音频数据集，降低中小企业的标注成本。同时加大抗攻击算法研发，提升音频系统抵御对抗性攻击的能力。

（二）完善法律法规，健全版权保护体系

立法部门需要补充声音权、AIGC 音频著作权的相关法律条款，明确 AI 语音克隆、AI 创作内容的使用边界、版权归属和侵权处罚标准。监管部门出台细则，要求使用他人声线、原创音频素材必须获得正式授权。

音频行业应用技术手段强化版权保护，普及音频数字水印技术，在 AI 音频中嵌入隐形标识，记录作品和声线来源。可将区块链技术应用在版权认定工作中，这样作品从创作到网络流转的每个环节都会留下真实记录，可以追溯全程。各大音频、短视频平台履行审核义务，定期排查侵权内容并及时下架，遏制侵权传播。

（三）搭建防护体系，守护用户语音隐私

监管部门要明确数据采集规则，要求企业收集语言、声纹数据前，必须获得用户明确授权，同时告知数据用途与存储时长，禁止设备私自收录用户对话[10]。企业采用加密技术存储语音数据，运用差分隐私、联邦学习等技术，在模型训练中保护原始数据安全。

执法部门加大打击力度，严查非法采集、倒卖语音数据的行为，建立数据泄露追责制度。同时面向普通用户开展科普，引导用户关闭设备不必要的语音收录功能，主动规避隐私风险。

（四）落实行业规范，强化内容与伦理监管

行业协会制定统一准则，规定所有 AI 生成音频必须标注“AI 制作”，方便公众辨别。各大平台升级审核系统，搭配专业检测工具，识别虚假语音、违法音频，及时处置违规内容与账号。技术机构继续优化伪造音频检测工具，向监管部门和平台开放使用。行业建立自律机制，企业自觉抵制低俗、违法音频制作。公安、网信、文旅等部门开展联合整治，重点打击 AI 语音诈骗、传播虚假信息等违法行为，从源头管控风险。

四、结语

人工智能改变了音频信号处理的发展模式，在传媒、文娱、通信、安防、工业等多个领域发挥重要作用，催生了虚拟主播、AI 音乐、智能声纹安防等新业态。技术快速发展的同时，技术缺陷、版权侵权、隐私泄露、虚假语音等问题接踵而至，这是行业发展的必经阶段。想要推动行业行稳致远，不能只侧重技术研发，还要同步完善法律法规、行业准则与监管机制。

参考文献

[1] 罗潇 . 人工智能驱动的计算机语音识别技术及应用 [J]. 数字技术与应用 ,2025,43 (8):45-47.

[2] 赵远杰 , 宋霄钰 , 秦振宇 . 从文本到声音：AI 语音合成技术对有声书产业的影响 [J]. 传播与版权 ,2025, (4):37-40.DOI:10.16852/j.cnki.45-1390/g2.2025.04.025.

[3] 姚晓龙 , 张雪 . 语音增强技术在实时音频处理中的应用研究 [J]. 电声技术 ,2024,48 (11):73-75.DOI:10.16311/j.audioe.2024.11.022.

[4] 钱兴维 , 张祥 . 基于深度学习的声纹识别语音唤醒技术优化研究 [J]. 电声技术 ,2024,48 (6):53-55.DOI:10.16311/j.audioe.2024.06.016.

[5] 武堂颖 , 杨璐 , 徐丽丽 . 人工智能在音频信号处理中的应用与挑战 [J]. 电声技术 ,2024,48 (5):31-34.DOI:10.16311/j.audioe.2024.05.009.

[6] 李粟 . 人工智能音频信号处理中的隐私保护与安全性问题研究 [J]. 电声技术 ,2023,47 (9):20-22+26.DOI:10.16311/j.audioe.2023.09.006.

[7] 李雅筝 , 刘宇星 . AIGC 技术赋能数字音频内容生产：应用场景、存在问题与应对策略 [J]. 数字出版研究 ,2023,2 (3):13-20.

[8] 李粟 . 人工智能在音频信号处理中的应用与挑战 [J]. 电声技术 ,2023,47 (8):45-47+51.DOI:10.16311/j.audioe.2023.08.011.

[9] 田华 . 智能化技术在音频制作中的应用与展望 [J]. 电视技术 ,2023,47 (5):89-92.DOI:10.16280/j.videoe.2023.05.023.

[10] 李晓坤 , 郑永亮 , 袁烺 , 等 . 基于深度学习的声纹识别方法研究 [J]. 黑龙江大学工程学报 ,2018,9 (1):64-70.DOI:10.13524/j.2095-008x.2018.01.012.

 Copyright © This Work is Licensed under A Commons Attribution-Non Commercial 4.0 International License.