

语音伪造与检测发展及合成样本分析研究——基于 CiteSpace 的可视化分析

蒋奕可 何枳岑 葛梓旭 戈子阳 丁誉洋

江苏警官学院，江苏南京，210031；

摘要：本文聚焦于 1997 年至 2025 年语音伪造与检测技术研究，使用可视化工具对中国知网数据库数据研究。研究表明，该研究经历了三个具有显著特征的阶段，呈现出快速增长趋势；文献发表时间跨度长，但集中于短时间内，形成明显的“短期爆发”特征；研究领域内研究网络复杂，但存在缺乏多学科交叉、理论滞后问题。展望未来，需强化理论研究与理论转实践进程，深化作者网络构建，推动语音检测持续发展。

关键词：CiteSpace；语言识别；可视化分析；理论

DOI：10.69979/3041-0673.26.04.009

1 绪论

近年来，随着通讯技术的高速发展，电信网络诈骗案件呈高发态势，深度伪造语音技术被滥用呈现规模化、精准化趋势。此类技术滥用已严重威胁社会信任体系，2023 年国务院《关于加强打击治理电信网络诈骗违法犯罪工作的意见》明确将“遏制 AI 技术非法滥用”列为关键攻坚方向，明确提出“强化人工智能安全技术研发，遏制技术滥用”，亟需通过技术创新实现检测能力与伪造技术的动态对抗。2025 年 1 月 24 日中共中央书记处书记、国务委员王小洪在全国打击治理电信网络诈骗工作视频会上强调：以更大的决心和力度推动打击治理电信网络诈骗工作向纵深发展。

当前，语音伪造技术呈现“低成本+高逼真”双重特征。开源工具（如 So-VITS-SVC、VITS）大幅降低伪造门槛，普通用户仅需 5 分钟语音即可生成高相似度合成语音；扩散模型（Diffusion Model）生成的伪造语音在 ASVspoof 2021 测试中欺骗率高达 41.2%，且能规避传统频域特征检测。扩散模型生成语音的 MOS 分已达 4.2（接近真人 4.5）。（ICASSP 2024）反观检测技术，发展速度相对较慢，且难以具有法律效应，难以满足当下反诈需求：现有模型在跨数据集（如 ASVspoof 电话录音）测试中误报率提升超 30%；司法取证场景要求难以满足当下反诈需求分钟级鉴定，而主流模型（如

ResNet-ATS）单条语音检测耗时超 15 秒；检测结果缺乏符合《公安机关电子数据取证规则》的证据链支撑，制约司法采信。同时，存在学术研究断层。当前 95% 文献聚焦单一数据集优化（Web of Science 统计），缺乏对 TTS 技术代际演进（传统参数合成→WaveNet→Diffusion Model）的跨代检测框架研究。

同时，当前的语音检验技术主要为以下三大类：基于声学特征的分析，识别梅尔频率倒谱系数（MFCC）与线性预测编码（LPC）。基于深层特征和 AI 模型的方法，利用深度学习模型对语音数据进行分析学习，构建针对语音伪造技术的模型。基于生物力学和生理学约束的方法，对视频的人物唇形、面部肌肉动作是否与说话内容匹配进行对比。

2 数据来源

本文选取中国知网（CNKI）数据库中 1997 年至 2025 年间相关文献作为研究对象。使用关键词“语言识别”、“语音识别”、“语音合成”、“语音检测”检索，选取学术期刊与学位论文，并关闭中英文拓展仅保留中文材料。经人工筛选后最终共收集到有效文献共 332 篇，并选取其中 58 篇强相关性论文标记为重要论文。重要论文以 pdf 格式下载，其余所有材料以 Refwork 格式下载。

3 结果分析

3.1 发文量年度分布

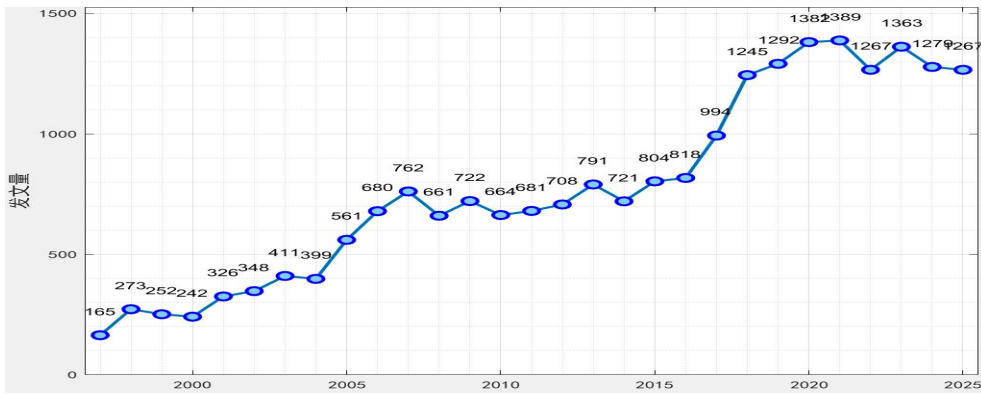


图 1 1997-2025 发文章量与年份关系示意图

图 1 展示了语言合成、识别领域文献数量的年度变化趋势。从图中可以清晰地观察到，该领域的发文章量变化轨迹呈现出 J 形曲线特征。总体而言，该领域自 20 世纪 60 年代该领域初步萌芽以来，大致经历了三个具有显著特征的阶段，即从初期的缓慢探索、中期的稳定积累，到后期的爆发式增长，形成一条完整的演化路径。

在第一阶段（约 1964 年至 20 世纪 90 年代中期），年度发文章量长期维持在极低水平，表明其在当时尚处于一个概念引入与理论初步探索的萌芽期，相关研究较为浅显，应用场景有限，并未引起学界的广泛关注。

进入第二阶段（20 世纪 90 年代中期至 2010 年左右），发文章量开始摆脱停滞状态，呈现出稳定的线性增长趋势，并在 2007 后进入较小的波动期。这标志着随着网络技

术与相关硬件的研发升级，语音识别逐渐得到重视，在部分领域初步应用并取得初步发展，为后续的爆发式增长奠定了基础。

自 2010 年左右起，尤其是 2016 年后，发展进入了第三阶段，即爆炸性增长期。图表数据显示，此阶段的发文章量曲线陡然攀升，并长期稳定在较高水平。这一现象与通讯及人工智能的发展息息相关，一方面，更便捷的通讯平台为其提供了极大的应用平台，配音产业、有声阅读都需要便宜迅速的音频来源，人工智能也降低了语音合成、识别的成本与难度，并提供了较大的需求。另一方面，电信网络诈骗大量使用相关技术制作虚假音视频，针对其的反制成为了当前公安工作的一大重点。

3.2 关键词共现与聚类分析

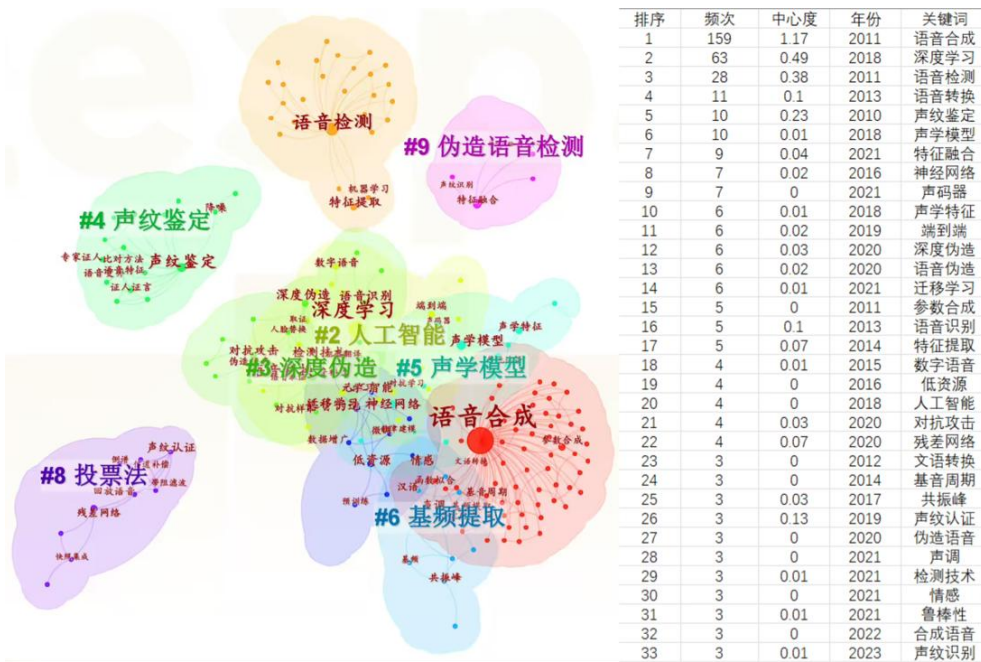


图 2 关键词共现与聚类图

如图所示，关键词共现网络中，节点大小与关键词频次呈正相关，节点间连线数量表示共现关系紧密程度，多节点高密度连线形成聚类模块。节点数量与连线密度共同反映研究主题的受关注程度及内部关联性——节点越多、连线越密集，表明该聚类所代表的研究方向聚合了更丰富的子主题，且关键词间关联性更强。

从网络结构来看，“语音合成”、“深度学习”等关键词节点显著大于其他节点，在图2中也展示出了较高的中心度，表明二者是“语音伪造与检验”领域的核心主题：一方面，其高频出现印证了技术基础（深度学习）与应用场景（语音合成）在当前研究中的主导地位；另一方面，复杂的网络连接结构也揭示了其具有较强的主题辐射性，表明该领域关键词之间的紧密联系与相互

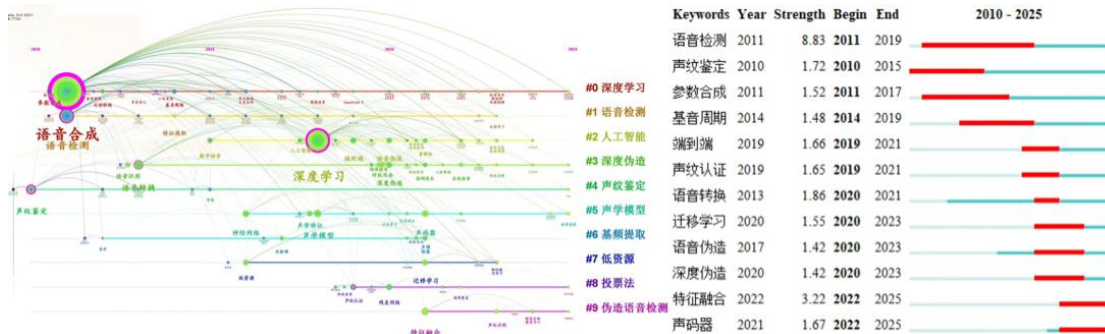


图3 关键词时间线谱图

图3可以看出，“深度学习”“语音检测”等节点较大，表明它们是该领域的核心研究主题。“深度学习”作为当今人工智能的关键技术，在语音合成与检测中发挥着基石作用，其强大的特征学习和模式识别能力为语音相关研究提供了新的方法和思路。“语音检测”则是保障语音信息真实性和安全性的重要研究方向，涉及对语音信号的准确识别和判断。

同时，可以看出“语音检测”具有显著的影响力，在研究期间具有最高的引用强度（8.83），表明这一时期（2011-2019）语音检测是该领域的研究热点。随着语音技术在各个领域的广泛应用，与日益加剧的网络诈骗，对语音信号的真实性和安全性检测需求日益增长，推动了相关研究的深入开展。“声纹鉴定”起步最早，2010年开始有引用爆发，持续到2015年，尽管强度时间比相对较低（0.344），低于其余关键词，但其对后续的发展起到了领导与开拓作用，代表其在早期是重要研究方向，吸引了大量研究者的关注和引用。同时，在2019

作用。

生成的知识图谱网络中节点数量为263，边的数量为491，网络密度为0.0143，相对较低，表明该网络中各个关键词之间联系相对分散，联系主要聚焦于语音合成，其次为语音检测与深度学习、声纹鉴定，其余关键词中心度均较低，表明该领域可能与其他成熟研究方向的关联不够紧密，关键词与其他相关领域关键词暂未形成较好的结合与关联。同时聚类后，可得模块度（Q）值=0.6474，平均轮廓值（S）值=0.9631，调和平均数为0.7743。Q值较高，表明该网络具有一定社区结构。S值为0.9631，表明该网络内社区结构紧密，划分质量高。

3.3 研究前沿与热点演变

年至2023年期间出现了较多短期集中爆发的关键词，如“端到端”“声纹认证”“语音转换”等，表明这些是当前新兴的研究方向，但部分可能随着技术迭代被逐步替代或是转化为其他先进技术。其中“特征融合”，作为2022年爆发引用至今的关键词，以3.22的高引用强度与1.073的高强度时间比成为研究关键词中的新兴热点，这显示了研究领域的扩展和深化，也展示了后续研究中对轻量化、综合化的发证导向。

4 主要研究结果

4.1 已经呈现的发展趋势

(1) 技术融合加速，应用场景多元化拓展。当前，语音合成与识别领域呈现出技术深度融合的显著趋势。深度学习作为核心技术基石，持续推动着语音相关研究的创新发展，其强大的特征学习和模式识别能力为语音合成、语音检测、声纹鉴定等多个方向提供了关键支撑。同时，语音技术的应用场景不断多元化，从传统的配音

产业、有声阅读, 延伸至公安反诈、智能安防、智能客服等众多领域。

(2) 合作网络松散, 跨领域关联有待加强。通过对关键词共现网络的分析发现, 当前该领域的合作网络紧密程度不高。其中, 虽然“语音合成”“语音检测”“深度学习”“声纹鉴定”等关键词之间联系相对紧密, 但与其他成熟研究方向的关联不够紧密, 关键词与其他相关领域关键词暂未形成较好的结合与关联。这在一定程度上限制了该领域的创新发展, 不利于整合多方资源开展跨学科研究。

(3) 研究热点持续演变, 新兴方向不断涌现。从关键词时间线谱图可以看出, 该领域的研究热点处于动态演变之中。“语音检测”在2011-2019年期间凭借较高的引用强度成为研究热点, 反映出当时对语音信息安全的迫切需求。而近年来, “端到端”“声纹认证”“语音转换”等新兴关键词短期集中爆发, 表明这些方向成为当前研究的新热点。特别是“特征融合”自2022年爆发引用至今, 以高引用强度和高强度时间比成为新兴热点, 体现了研究领域向轻量化、综合化方向的发展导向, 研究者们更加注重多种技术手段的综合运用, 以提升语音处理的效果和效率。

4.2 未来研究展望

(1) 技术创新持续深化, 推动性能提升与成本降低。未来, 随着人工智能技术的不断发展, 深度学习算法将进一步优化, 语音合成与识别技术的性能将得到显著提升。同时, 技术的创新也将促使相关成本进一步降低, 使得语音技术能够在更多领域得到广泛应用, 对于中小企业和民用领域的发展可以起到较好的推动作用。

(2) 跨领域合作加强, 拓展应用边界与深度。鉴于当前合作网络松散、跨领域关联不足的现状, 未来该领域将更加注重跨学科、跨领域的合作研究。通过与多学科的深度融合, 语音合成与识别技术将在医疗、教育、

交通、执法等更多领域发挥重要作用。跨领域合作将为语音技术带来新的研究思路和应用模式, 拓展其应用边界和深度。

(3) 关注伦理与安全问题, 完善技术规范与监管体系。随着语音技术的广泛应用, 伦理和安全问题日益凸显。因此, 未来需要加强对语音技术伦理和安全问题的研究, 制定完善的技术规范和监管体系。一方面, 要加强对语音伪造技术的检测和防范, 在合成语音中加入电子水印、研发更加先进的语音检测算法和工具, 提高对虚假语音的识别能力; 另一方面, 要建立健全相关法律法规, 明确语音技术的使用边界和责任主体, 加强对违法违规行为的打击力度, 保障语音技术的健康、有序发展。

参考文献

- [1] 厉鹏, 李晔, 王铭欣. 刍议语音识别在公安工作中的应用. 中国人民公安大学; 山东省公安厅出入境管理局. 2018.
- [2] 魏方达, 刘淼, 孙毅, 王晶, 赵胜辉. 音视频深度伪造与鉴伪综述. 北京理工大学. 2025.
- [3] 李旭嵘, 纪守领, 吴春明, 刘振广, 邓水光, 程鹏, 杨珉, 孔祥维. 深度伪造与检测技术综述. 软件学报, 2021, 32(2): 496-518.
- [4] 梁瑞刚, 吕培卓, 赵月, 陈鹏, 邢豪, 张颖君, 韩冀中, 赫然, 赵险峰, 李明, 陈恺. 视听觉深度伪造检测技术研究综述. 信息安全学报. 2020.

作者简介: 蒋奕可(2005—), 男, 汉, 江苏常州人, 本科在读, 单位: 江苏警官学院, 研究方向: 刑事科学技术。

基金项目: 江苏省高等学校大学生创新创业训练计划项目《语音伪造与检测发展及合成样本分析研究》; 项目编号: XJ202510329022。