一种语音驱动舌超声成像的统一生成框架发布

作者：writer 来源：科学网

本文原地址：https://www.iikx.com/news/progress/36679.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

一种语音驱动舌超声成像的统一生成框架发布。近日，中国科学院深圳先进技术研究院团队提出了一种从语音到舌超声生成的统一扩散模型框架，能够根据输入的未知语音信号如健康或病理性语音，自适应地生成高保真度的舌体超声数据，为解决言语障碍评估与康复中的相关技术难题提供了新路径。相关研究发表于《信息融合》。

在言语治疗与康复领域，准确、直观观察发音时舌头的内部运动对诊断评估与康复训练具有关键意义。然而，传统超声舌成像设备因操作专业要求高、成本昂贵，在临床普及上面临较多限制。基于AI技术实现的语音到舌超声生成方法，通过算法仅凭语音信号即可推算出对应的舌体运动数据，有望给临床提供了一种低成本、非侵入式的可视化解决方案。

然而，该技术长期面临两大挑战：一是配对的病理性语音-舌动数据极其稀缺，导致模型泛化性能差；二是健康与病理性言语在发音机制和声学特性上存在显著差异，导致模型在处理未知类型的语音时容易失败。

针对这些问题，研究团队提出了 Uni-UTIDiff统一框架。该框架具有统一建模、智能辨声、自适应生成等优势。研究结果表明，该框架不仅能够分别达到针对健康与病理语音的专家模型性能水平，还能在生成的舌超声图像中展现出优异的清晰度与自然度，充分验证了其在跨发音模式下的鲁棒生成能力。

研究人员介绍，基于该技术，未来可以仅凭用户语音，即可生成对应的发音器官内部运动图像，在多个领域具有应用前景。

例如，在精准康复与远程医疗平台，可为不同地区的构音障碍患者提供专业言语治疗，并通过实时可视化的舌位运动生物反馈，辅助患者高效开展家庭康复训练；在临床辅助诊断中，该技术可作为筛查工具，帮助医生快速评估患者言语运动功能，生成客观影像报告，从而提升诊断效率；在语言教学与保护方面，它既能帮助学习者直观理解发音要领，也可用于记录或推断特定罕见语言的发音生理特征，助力语言资源的留存与研究。（来源：中国科学报 刁雯蕙）

相关论文信息：https://doi.org/10.1016/j.inffus.2025.103896

作者：王岚等 来源：《信息融合》

更多 科学进展 请访问 https://www.iikx.com/news/progress/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发