

清华大学苑克鑫、胡晓林团队实现混合语音分离技术 计算机学会像人脑一样“听话”了

● 本报记者 张思玮

在通勤的地铁上、嘈杂的餐厅里，广播声、音乐声、周围人说话声，似乎都不会妨碍你与同伴进行交流。而这就是大脑在处理声音信息时发挥的特殊优势——它可以将注意力集中在感兴趣的对话或声音上，忽略其他无关的声音或者噪声。

“尽管我们每天都在不知不觉地运用大脑这种优势，可以轻松地在混合的声音中识别特定的目标声音，但要让计算机做到这一点却非常困难。”近日，清华大学生物医学工程学院苑克鑫教授团队联合该校计算机系胡晓林教授团队，基于哺乳动物丘脑和皮层整合多模态感觉信息的工作原理，构建了一款新的脑启发 AI 模型（CTCNet），实现了混合语音分离技术突破，让计算机进一步学会像人脑一样“听话”。相关研究成果发表于《模式分析与机器智能 IEEE 汇刊》。

计算机尚不能 有效分离两个人声音

据了解，现有的多模态语音分离方法大多只是模拟了哺乳动物的皮层通路，即从较低功能区域（例如初级视觉、听觉皮层区）到较高功能区



图片来源：摄图网

域（例如高级视觉、听觉皮层区）。同时，在人工智能（AI）领域，迄今为止尚无有效的办法使计算机有效分离两个人的声音。

然而，实际上，听觉、视觉的信息整合中，以丘脑为代表的皮层下结构发挥了不可忽视的重要作用。

“高级听觉丘脑作为处理听觉信息的关键中枢节点，具有听觉、视觉双模态的特性。其腹内侧在介导听觉、视觉刺激触发的危险感知中发挥关键作用；其背侧既接收来自听觉皮层第5层的投射，也接收来自视觉皮层第5层的投射，且在整体上形成了皮层—丘脑—皮层（CTC）循环连接架构。”苑克鑫表示，这提示高级听觉丘脑可能通过特殊的联接模式整合听觉、视觉信息从而增强听觉感知。

在背侧高级听觉丘脑联接特点的启发下，苑克鑫团队与胡晓林团队合作提出了一种皮层—丘脑—皮层神经网络（CTCNet）来执行音频—视觉语音分离任务。

新模型打开脑启发范例

据介绍，CTCNet 包括三个模块——听觉子网络、视觉子网络和听—视融合子网络，分别模拟了听觉皮层、视觉皮层和背侧高级听觉丘脑。

其基本工作原理为：首先，听觉信息（语音）和视觉信息（唇部运动）以自下而上的方式分别在独立的听觉和视觉子网络中处理；然后，经过处理的听觉和视觉信息通过自上而下的连接在听—视融合子网络中进行多时间分辨率尺度的融合；最后，融合后的信息被

回传至听觉和视觉子网络。上述过程会重复数次，最终输出至听觉子网络。

研究显示，在三个语音分离基准数据集上的测试结果表明，在参数极少的情况下，CTCNet 能在视觉信息（唇部运动）的辅助下，高度准确地将混合在一起的语音分离。

苑克鑫表示，语音分离模型 CTCNet 是基于高级听觉丘脑的视、听融合能力及其皮层—丘脑—皮层循环联接架构构建的脑启发 AI 模型。通过多次融合和循环处理听觉和视觉信息，使 AI 能够更好地实现“鸡尾酒会效应”。

“新模型的语音分离性能大幅领先于现有方法，不仅为计算机感知信息处理提供了新的脑启发范例，而且在智能助手、自动驾驶等领域有潜力发挥重要作用。”该研究负责人之一苑克鑫说，未来研究团队将基于自主创新的组织光学透明化方法，在单神经元水平上深入解析高级听觉丘脑的输入—输出连接模式，进一步提升人工智能系统在自然场景中的感知能力。

相关论文信息：<https://doi.org/10.1109/TPAMI.2024.3384034>

顾兵团队揭示艰难梭菌感染新机制

本报讯 近日，广东省人民医院检验科教授顾兵团队在肠道共生原生物调控艰难梭菌感染作用机制上获得最新进展。研究团队发现了共生原生物能够调控宿主免疫反应和肠道稳态缓解艰难梭菌感染。进一步的研究表明，共生原生物 T.mu 通过调节宿主精氨酸—鸟氨酸代谢轴影响肠道免疫细胞分化，以减轻对艰难梭菌的易感性。上述研究成果发表于《自然—通讯》，为艰难梭菌感染的防控提供新方向。

艰难梭菌感染已成为世界性公共卫生问题之一，主要分为无症状感染、医源性腹泻和假膜性结肠炎等。无症状感染者是最重要的感染源，主要为新生儿和老年人；医源性腹泻主要表现为水

样便（24 小时内 > 3 次），另有下腹痛、低热、恶心、厌食等症状；假膜性结肠炎可出现血水样腹泻和腹痛伴发热等全身症状，严重者甚至危及生命。

肠道微生态失调是艰难梭菌感染的重要因素，但肠道真核微生物在其中有何作用尚不清楚。为此，该研究团队从共生原生物、艰难梭菌、宿主免疫三方互作入手进行研究并取得了上述成果。

据悉，顾兵为最后通讯作者，徐州医科大学教授王玉刚为共同通讯作者，徐州医科大学副教授杨欢、硕士研究生吴小小、讲师李晓为论文共同第一作者。

（陈祎琪）

相关论文信息：<https://doi.org/10.1038/s41467-024-47075-0>

本报讯 华南师范大学脑科学与康复医学研究院副研究员王本驰团队首次在人类大脑中追踪到显著干扰信息。日前，相关成果在线发表于《自然—人类行为》。

“我们首次在人类身上找到了加工显著干扰信息的神经信号，发现了异于动物的独特加工机制，揭示了人类注意捕获的神经机制。”王本驰说。

在日常生活中，对凸显物体的快速识别时刻干扰着人们的工作和学习。

“好比专心工作时电话铃声响起，人们不得不放下手头工作。”论文通讯作者王本驰表示，长久以来，许多研究在行为水平上证实了注意捕获现象。但是，对于人类如何加工凸显物体及其神经基础，却存在诸多争议和未解之谜。

该研究通过颅内脑电信号采集，在

人类大脑中追踪到了对于凸显物体（显著干扰）加工的神经信号。通过模型转换，研究发现，这一信号在刺激出现后的 200 至 300 毫秒就已经出现。

“人类对于凸显物体的加工更多的是源于前扣带回和颞叶，而非传统观点所认为的额顶脑区，这更新了对这一现象的认知。”王本驰说。

相关论文信息：<https://doi.org/10.1038/s41562-024-01852-5>

<https://doi.org/10.1038/s41562-024-01856-1>

科学家首次在大脑中
追踪到显著干扰信息

（朱汉斌）