计算机学会像人脑一样“听话”了

作者：writer 来源：科学网

本文原地址：https://www.iikx.com/news/progress/26708.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！

计算机学会像人脑一样“听话”了。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音。

其实，早在70多年前，神经科学家就注意到大脑的这种神奇能力，并将其称为鸡尾酒会效应。

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最新一期的《模式分析与机器智能IEEE汇刊》（IEEE Transactions on Pattern Analysis and Machine Intelligence）。

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，迄今为止也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。

采访中，记者了解到苑克鑫团队长期聚焦于脑的听觉处理机制，近年来的一系列工作逐渐揭示了高级听觉丘脑的联接、功能与工作机制。

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）循环联接架构。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模型包括三个模块——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开。

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。

据悉，清华大学硕士生李凯为论文第一作者，清华大学博士后谢凤华、博士生陈航分别为论文第二、第三作者，苑克鑫和胡晓林为该论文共同通讯作者。（来源：中国科学报 张思玮）

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

作者：苑克鑫等 来源：《模式分析与机器智能IEEE汇刊》

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