

这些非语言动作。美国加州大学旧金山分校 Edward Chang 团队利用单个脑表面植入阵列,同时解码语言与手势尝试。2026年9月14日,相关研究发表于《Nature Neuroscience》。

脑机接口旨在将脑活动转化为操控外部设备的指令,从而恢复丧失的功能。以往脑机接口研究多将说话、移动光标或做手势分别处理。然而,言语和上肢/头面运动在感觉运动皮层部分重叠,单独训练“只说话”或“只动手”的解码器并行时可能互相干扰。

此次,研究团队使用置于感觉运动皮层上方的脑植入装置,记录了3名声道和肢体瘫痪参与者的神经活动。他们首先证明,单个植入装置能采集与多种运动类型相关的信号,包括上肢运动和与言语相关的嘴部及面部运动,如耸肩、握拳挥动或点头;随后,团队在其中2名参与者中使用并行的言语和手势解码器,驱动一个全身虚拟形象,参与者被要求尝试做出特定手势、回应对话提示,或同时完成两者。

在对话任务中,该系统准确

解码了2名参与者的言语和手势:其中一名参与者在3个对话任务区块中,言语和手势解码的中位准确度均达到100%。

研究团队发现,同时使用言语和手势数据训练系统可提高性能并减少错误。他们指出,这项技术在投入应用之前,还需纳入更多参与者,并使用规模更大的词汇集和手势集进行测试。

(来源:《Nature Neuroscience》)

### 让机器人“想好再动”

叠杯子、抚平揉皱的毛巾、把东西放进抽屉——这些人类习以为常的日常琐事,对机器人来说却往往充满挑战。为了让机器人更聪明、更有条理地执行物理操作,美国卡内基梅隆大学(Carnegie Mellon University, CMU)的 Adam Hung、Deva Ramanan 等研究人员开发了一款名为 ModAR 的“世界动作模型”。它的核心逻辑很直观:在控制双臂真正动起来之前,先让机器人在脑海中“预演”场景的物理变化。2026年9月15日,相关研究已在 arXiv 预印本平台发布。

过去,赋予机器人“预测未来”能力的模型,往往习惯将未来场景生成为一张张逼真的彩色图像。但实际上,

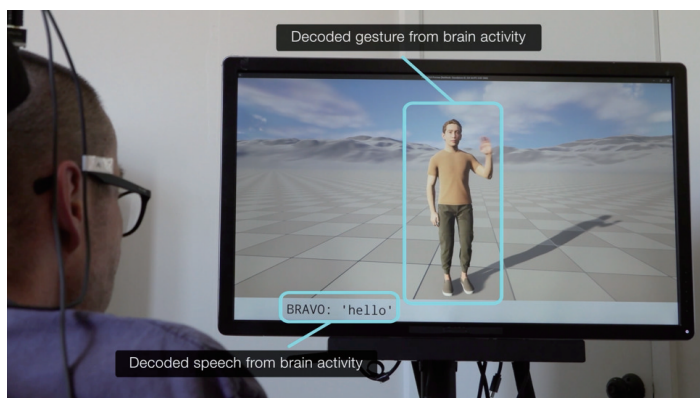
像素级地重建整个画面未必能提炼出操作真正需要的信息。ModAR 走了一条更务实的路线:它放弃了生成全彩画面,转而依次预测物体关键点的运动轨迹、反映场景语义的视觉特征,以及空间深度关系。前一步的预测结果直接作为下一步的推理依据,层层递进,最终推导并生成机器人的动作。这种机制相当于在动手前,先在系统中生成一份逐步细化、剥离了冗余视觉信息的“操作草图”。

这种“抽象预演”在真实的双臂机器人上展现出了威力。研究团队设计了叠杯子、折毛巾以及“放置物品并关闭抽屉”3项测试,每项执行30次。当仅使用100段通过遥操作录制的机器人示范数据进行训练时,系统的平均成功率为70.0%。

有趣的是,该模型能够有效吸收缺乏机器人动作标签的数据。当研究人员在训练集中加入200段同任务的纯人类操作示范后,机器人的成功率大幅跃升至81.1%;进一步混入1000段 Ego-Dex 数据集中的人类操作视频后,成功率继续小幅提升至83.3%(尽管这一微小增益的稳定性仍需更大规模的测试来确认)。

人类示范视频虽然没有提供关节电机的转动数据,但极大地丰富了模型对“物体被操作后应如何变化”的物理常识理解。这证明,模型完全可以利用海量且极易获取的人类操作视频来补充对物理世界的认知,而具体的执行控制再由少量的机器人示范来补足。

(来源:arXiv)



对瘫痪参与者的神经活动进行实时解码,以控制个性化的虚拟化身,实现言语和手势交流

(图片来源:作者团队)