

生物物理基础期末汇报

黄杰 184511084146; 主题:生物与机器学习; 刊物:《Nature》; 时间:2019年4月25日;
文献: "Speech synthesis from neural decoding of spoken sentences"

摘要:

将神经活动转化为言语的技术对于因神经损伤而无法沟通的人来说将具有变革性。从神经活动中解码语音是具有挑战性的,因为说话需要对声道咬合架进行非常精确和快速的多维控制。在这里,我们设计了一种神经解码器,它明确地利用人类皮质活动中编码的运动和声音表示来合成可听语音。循环神经网络(RNN)络首先将记录的皮质活动直接解码为关节运动的表示,然后将这些表示转换为语音声学。在封闭的词汇测试中,听众可以很容易地识别和转录从皮层活动合成的语音。此外,当参与者静默地模仿句子时,解码器可以合成语音。这些发现提高了使用语音神经假体技术恢复口语交流的临床可行性。

主要内容:

2019年4月25日, Anumanchipalli等人在自然杂志上发表了一篇名为"Speech synthesis from neural decoding of spoken sentences" 的文章说的是他们可以讲脑电波转化为语音。这是一件多么神奇的事啊! 以前只是在科幻电影中看到过这样的画面,没想到这一次人类真的将其变成了现实。他们是怎么做到的呢? 没错,这一次也是用到了机器学习。这里主要用到的一种技术叫做RNN(循环神经网络)。他们没有直接将脑电波转换成语音,而是先将脑电波转化成模拟口腔发声的运动信号,再通过这种运动的信号还原出语音信号。下面是一些具体描述:

说话看起来似乎是一种毫不费力的活动,但它其实是我们执行的最复杂的行动之一。它需要精确,动态地协调肌肉,嘴唇,舌头,喉部等。当一些人由于中风,肌萎缩侧索硬化或其他神经障碍导致言语中断时,丧失交流能力对他们的打击是毁灭性的。

脑-机接口(Brain-Computer Interface:BCI)旨在通过直接从大脑“读取”他们的意图并使用该信息来控制外部设备或移动瘫痪的肢体来帮助瘫痪的人。用于交流的BCI的发展主要集中在脑控制打字上,允许瘫痪的人每分钟输入多达8个单词。尽管恢复这种功能水平可能会改变有严重沟通障碍的人的生活,但基于分型的BCI不太可能实现自然语音的流畅交流(平均每分钟约150字)。Anumanchipalli等人已经开发出一种方法,使用深度学习(Deep Learning)方法从脑信号中产生口语句子。

研究人员与五名志愿者一起工作,这些志愿者正在进行一项称为颅内监测的程序,其中电极用于监测大脑活动,作为治疗癫痫的一部分。作者使用一种称为高密度脑电图的技术来跟踪控制言语和发音运动的大脑区域的活动。为了重建语音,而不是将大脑信号直接转换为音频信号,Anumanchipalli等人使用两个阶段解码的方法,他们首先将神经信号转换为声道咬合架运动的表示,然后将解码的运动转换为口语句子。这两种转换都使用了递归神经网络(RNN)。这种人工神经网络,它在处理和转换具有复杂时间结构的数据时特别有效。

了解大脑信号如何与声道发音器的运动相关是具有挑战性的,因为在与患有癫痫的人一起在医院环境中工作时很难直接测量这些运动。相反,作者使用他们之前开发的模型中的信息,该模型使用人工神经网络将记录的语音转换为产生它的声道发音器的运动。该模型是使用从以前的研究参与者收集的大量数据库建立的。通过包括一个模型来估计录制语音的声道运动,作者可以将大脑活动映射到声道运动,而无需直接测量运动本身。

一些研究使用深度学习方法来重建来自脑信号的音频信号。这些包括令人兴奋的BCI方法,其中神经网络被用于直接从控制语音的大脑区域合成口语单词。相比之下,Anumanchipalli及其同事将他们的解码方法分为两个阶段(一个解码声道发音器的运动和一个合成语音的运动),基于他们之前的观察,语音相关的大脑区域的活动更接近于运动声音咬合器的声音,而不是说话时产生的声学信号。