

# 神经网络如何模拟人耳听觉机制

人的耳朵是一个非常精巧的声音处理系统。总体来说，声音引起鼓膜振动，经过听小骨传递到耳蜗后，形成神经脉冲传到大脑中形成声音感知。在这一过程中，耳蜗是物理振动转换成神经电信号的核心部件，很多人耳感知的特性就是在这里形成的。例如，当一个声音已经存在且音量较高时，我们将很难感知另一个音高差不多的声音。类似的，当两个强音之间夹杂一个短暂的低音或噪音，我们的也很难察觉这些短暂变化的存在。这一现象称为人耳的掩蔽效应。这些感知特性和声音信号在耳蜗中的处理过程相关。声音在耳蜗中的传导是一个“频率选择”的过程，当传导到耳蜗的不同位置时对某一特定频率产生反应，这些反应互相叠加组合，从而形成了复杂的神经信号。

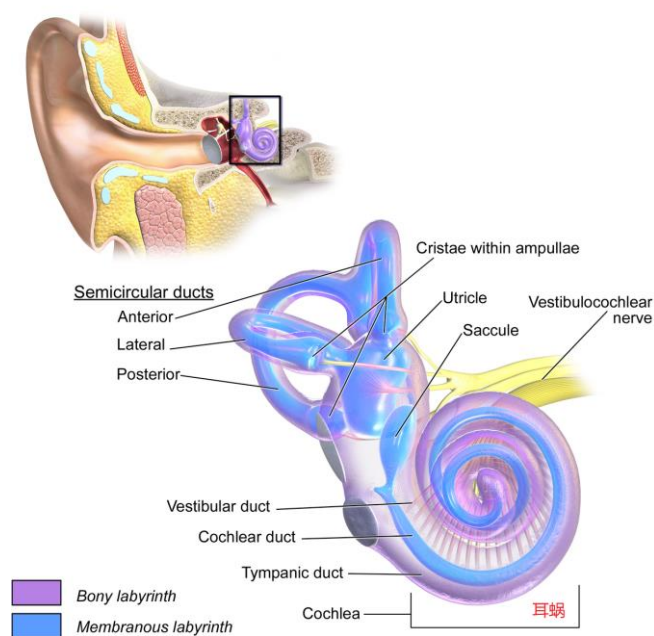


图 1: 人的内耳[1]

为描述这一感知过程，科学家们提出了一种称为传输线（Transmission Line, TL）的计算模型。这一模型将耳蜗分成若干段，每一段描述为一个差分方程。这一模型可以很精确地描述耳蜗的生理特性，可惜计算量太大，很难用到实际系统中。

近日，自然-机器智能杂志刊载了一篇论文[2]，用机器学习方法解决了 TL 模型的计算问题。在这篇论文中，作者将耳蜗分成 210 段建立 TL 模型，然后利用一个称为 CoNNear 的卷积神

神经网络来近似 TL 模型的输出（图 2）。实验表明，CoNNear 可以对 TL 模型做很了的近似，而计算效率提高了 2000 倍。

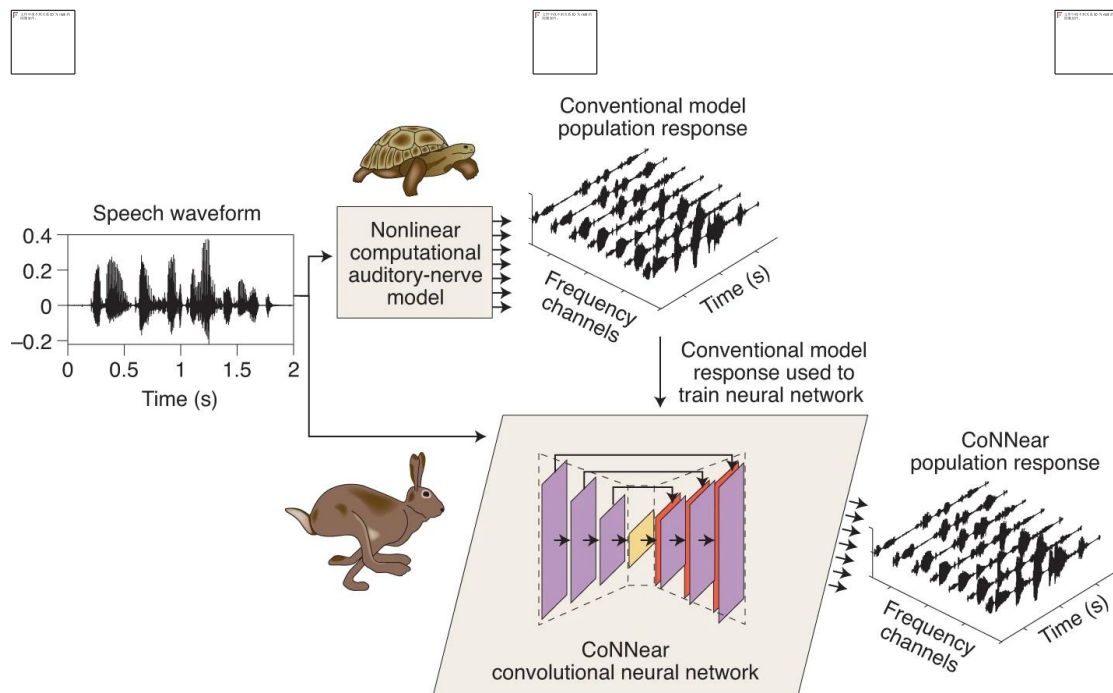


图 2: CoNNear 训练示意图。给定一段语音，利用 TL 模型（龟速运行）生成耳蜗响应，这些响应作为目标来训练 CoNNear（兔子模拟），使得网络可以模拟 TL 的输出。

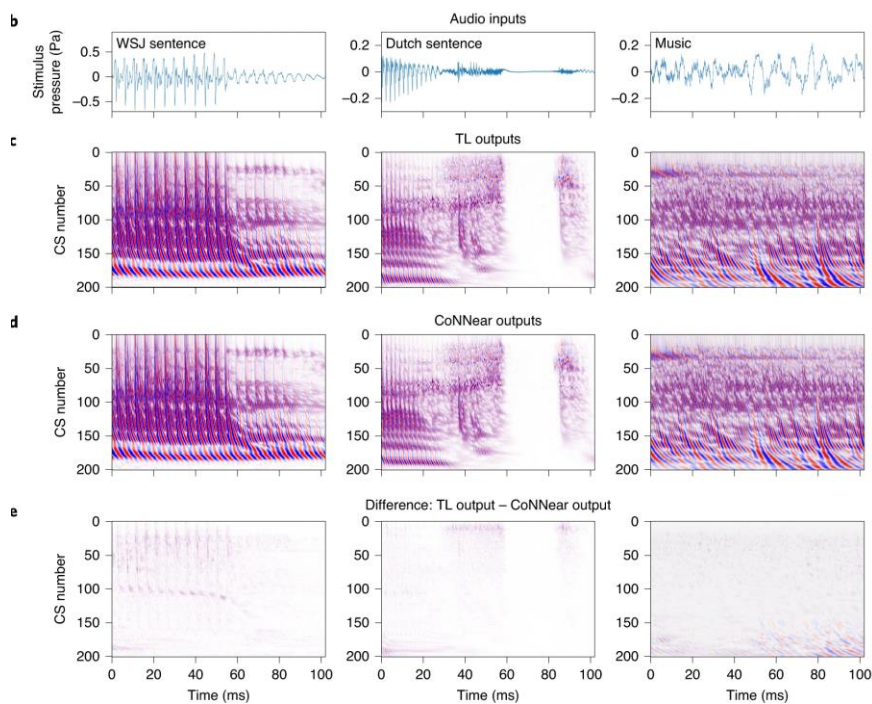


图 3: CoNNear 对 TL 的近似结果。

图 3 给出了 CoNNear 对 TL 的近似结果，其中第一行为原始声音，第二行为 TL 的输出（类似人耳的真实响应），第三行为 CoNNear 的近似结果，最后一行为 TL 和 CoNNear 预测的误差。可以看到，速度提高了 2000 倍的 CoNNear 在预测结果上和 TL 非常要似，验证了这一方法的有效性。

CoNNear 带来的一个启发是利用机器学习，一些计算很复杂的传统方法有可能用神经网络进行合理的近似，从而极大提高传统方法的应用范围。同时，传统方法的本身理论基础又为神经网络学习提供了约束，从而提高神经网络的可信性。

1. [https://en.wikipedia.org/wiki/Inner\\_ear](https://en.wikipedia.org/wiki/Inner_ear)
2. Baby, D., Van Den Broucke, A. & Verhulst, S. A convolutional neural-network model of human cochlear mechanics and filter tuning for real-time applications. *Nat Mach Intell* 3, 134–143 (2021).
3. Carney, L.H. Speeding up machine hearing. *Nat Mach Intell* 3, 190–191 (2021).