nature neuroscience

Article

https://doi.org/10.1038/s41593-023-01468-4

Dissecting neural computations in the human auditory pathway using deep neural networks for speech

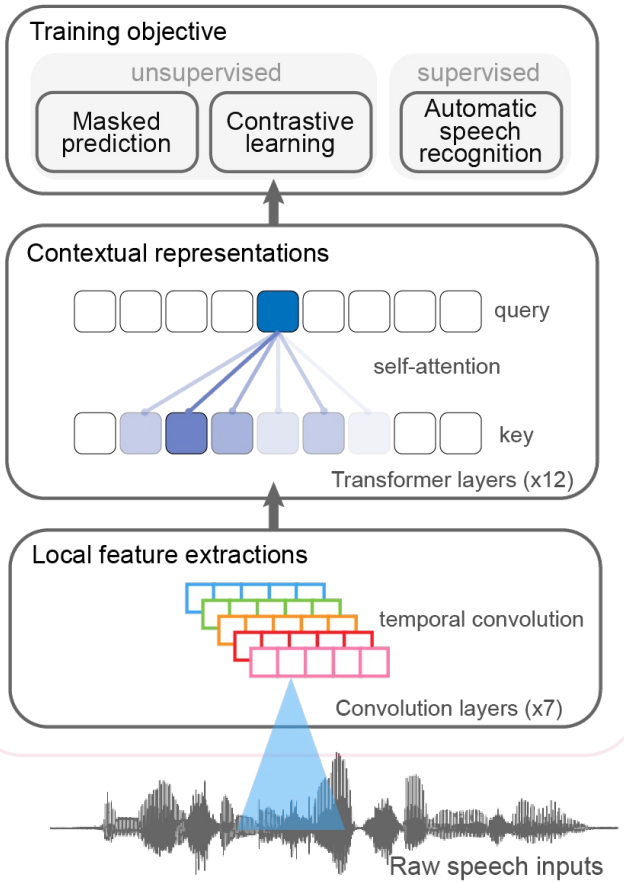
Received: 14 April 2022

Accepted: 13 September 2023

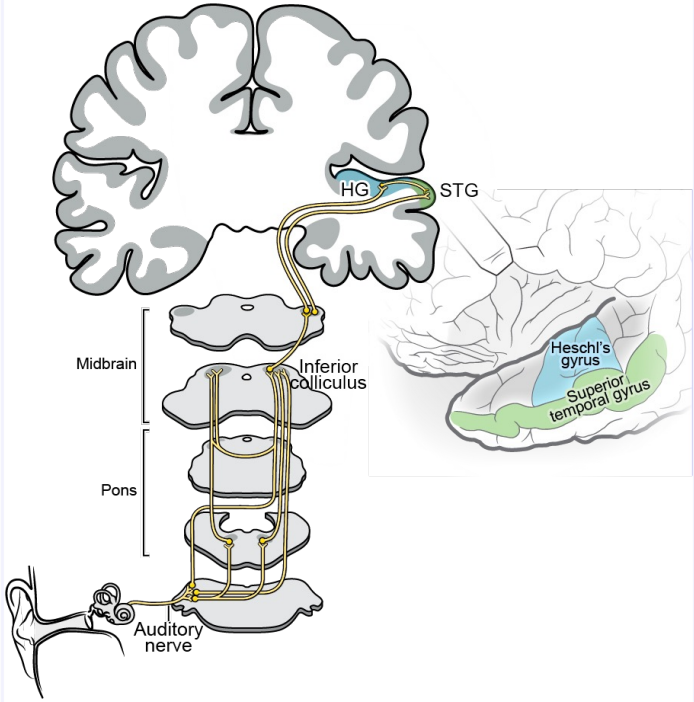
Yuanning Li[1,9], Gopala K. Anumanchipalli[2,3], Abdelrahman Mohamed[4], Peili Chen[5], Laurel H. Carney[6], Junfeng Lu[7,8], Jinsong Wu[7,8] & Edward F. Chang[1,2]

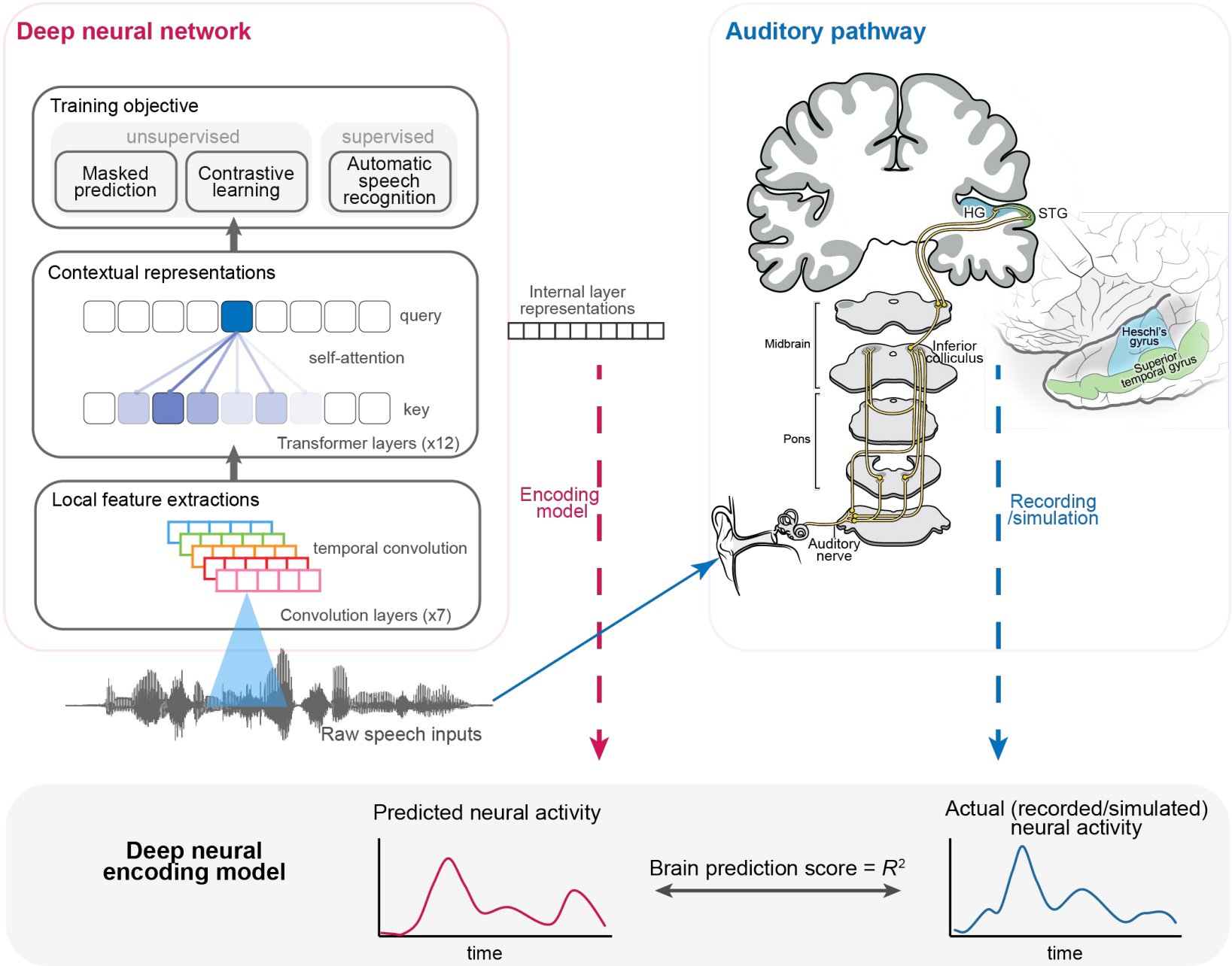
UCSF
University of California
San Francisco

Deep neural network



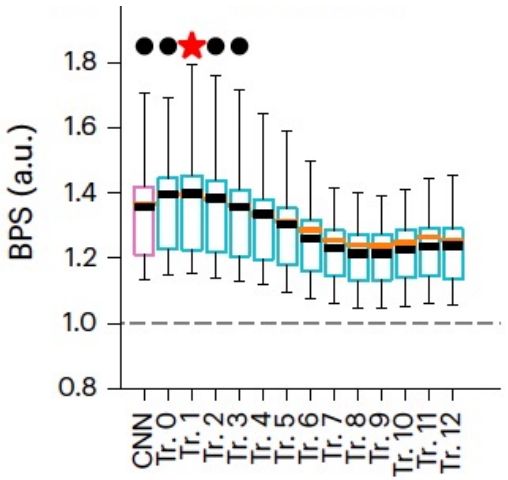
Auditory pathway



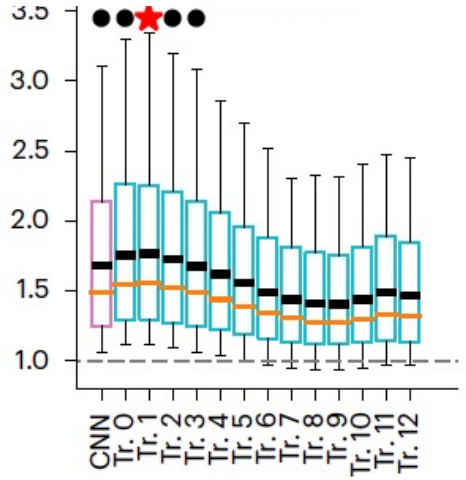


Inspired by DiCarlo, McDermott, et al

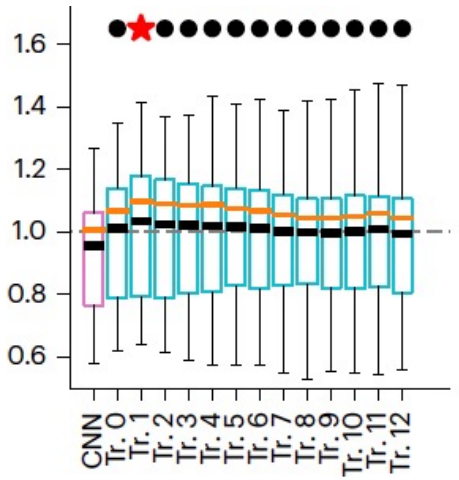
DNN and biology converge on similar hierarchy



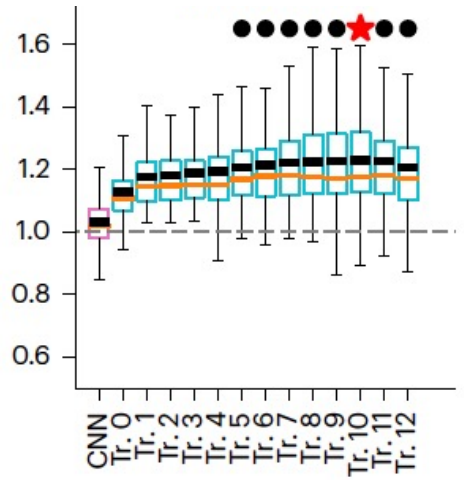
Auditory nerve



Inferior Colliculus



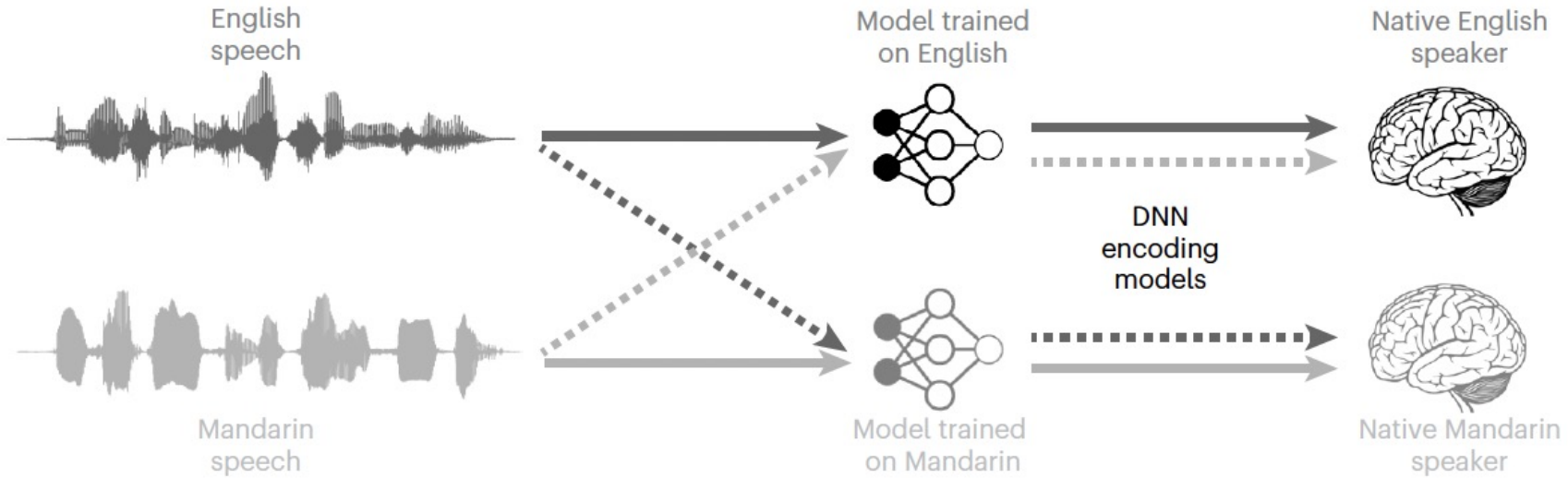
Primary Auditory Cortex (A1)



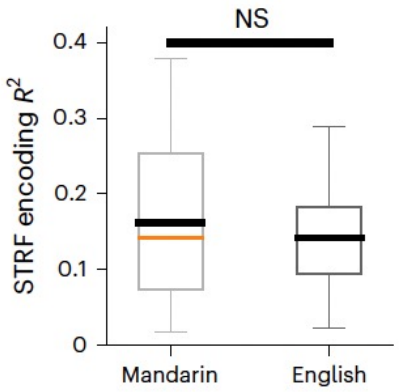
Superior Temporal Gyrus (STG)

- except A1 performs similarly to simple spectrogram model, confirming recent demonstration of parallel not serial processing across cortex (Hamilton 2021)

Cross-linguistic comparisons: DNN and STG

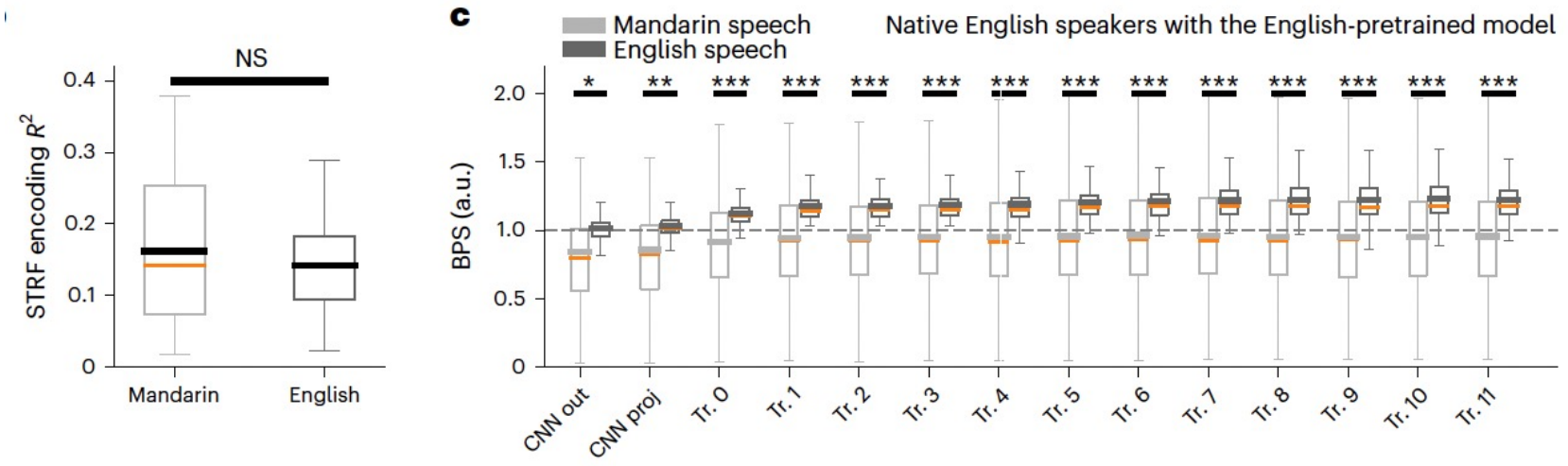


Language-specific representations in STG



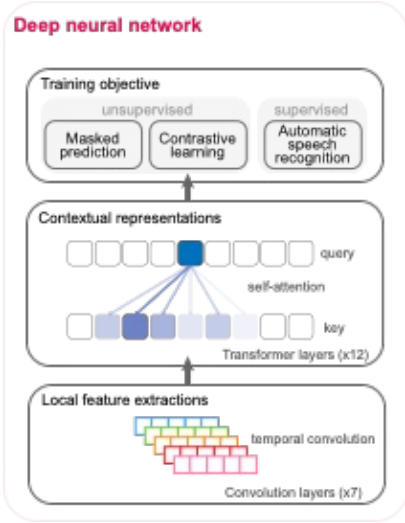
- Spectrotemporal representation is not language specific

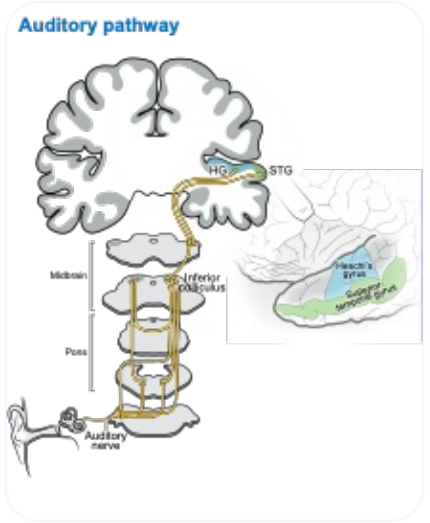
Language-specific representations in STG



- Spectrotemporal representation is not language specific
- English pre-trained model aligned to English speech better than Mandarin speech for native English speaker
  - …and vice versa for Mandarin

Convergence of DNN and biological processing for speech





1. The hierarchy in DNNs correlates with ascending auditory pathway
2. Unsupervised models without explicit linguistic knowledge can learn similar feature representations as the human auditory pathway
3. Deeper layers in speech DNNs correlate with human STG, driven by specific computations aligned with critical linguistically relevant temporal structures, such as phonemic and syllabic contexts
   - Static nonlinear filters (CNN) is good for subcortical and A1, Contextual models (transformer) consistently outperform for STG
4. DNN-based models, unlike traditional linear encoding models, can reveal language-specific properties in cross-language speech perception

nature neuroscience

Article

https://doi.org/10.1038/s41593-023-01468-4

Dissecting neural computations in the human auditory pathway using deep neural networks for speech

Received: 14 April 2022

Accepted: 13 September 2023

Yuanning Li[1,9], Gopala K. Anumanchipalli[2,3], Abdelrahman Mohamed[4], Peili Chen[5], Laurel H. Carney[6], Junfeng Lu[7,8], Jinsong Wu[7,8] & Edward F. Chang[1,2]

UCSF
University of California
San Francisco