Towards holistic neural encoding models for multi-modal naturalistic stimuli

Meenakshi Khosla[1], Gia H. Ngo[1], Keith Jamison[3], Amy Kuceyeski[3,4] and Mert R. Sabuncu[1,2,3]

[1]School of Electrical and Computer Engineering, Cornell University; [2]Nancy E. & Peter C. Meinig School of Biomedical Engineering, Cornell University; [3]Radiology, Weill Cornell Medicine; [4]Brain and Mind Research Institute, Weill Cornell Medicine

INTRODUCTION

- Neural encoding models predict cortical responses to novel stimuli. They:
  - Can shed light on functional organization
  - Can generate testable hypothesis
  - Are not paradigm bound
- There is dearth of machine learning models that exploit multi-sensory information to predict evoked neural response
- We propose an end-to-end deep neural network based encoding model that uses audio and visual stimuli sequences jointly to predict cortical responses.

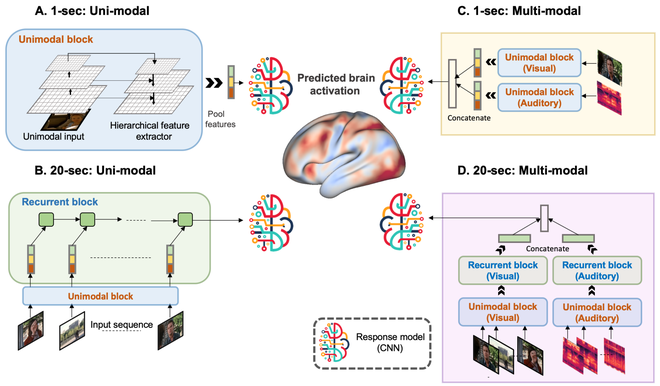
OBJECTIVES

GOAL: Build predictive models of cortical responses to arbitrary complex, dynamic stimuli

Our proposed model captures three critical inductive biases about information processing in the brain:

- *Hierarchical processing*
- *Influence of temporal history*
- *Multi-sensory assimilation*

MATERIALS & METHODS



MODEL INPUTS

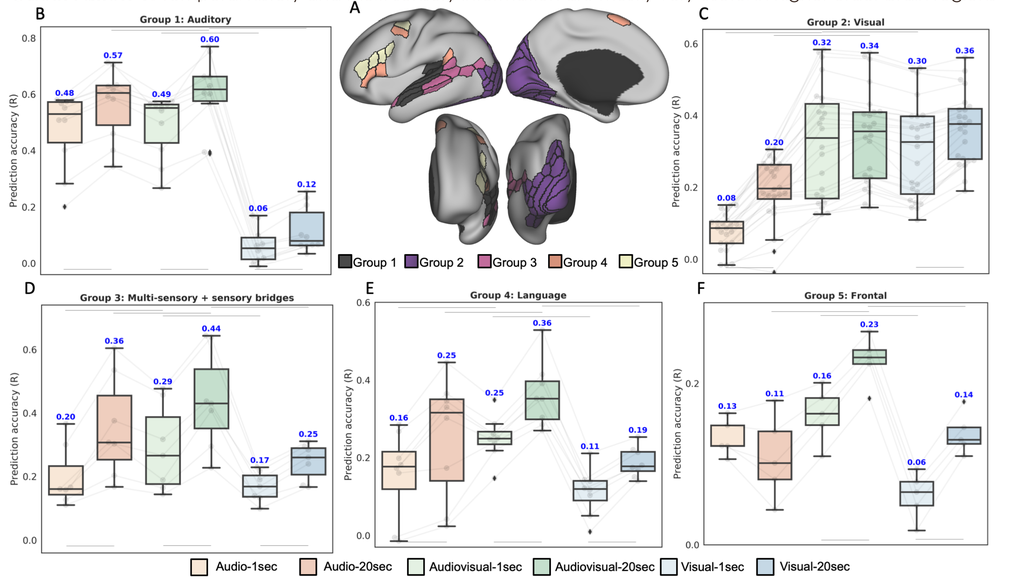
1. Audio-1sec: 1sec spectrogram
2. Visual-1sec: Single RGB frame, sampled every second
3. Audiovisual-1sec: 1sec spectrogram & 1 RGB frame
4. Audio-20sec: sequence of 20 consecutive spectrograms
5. Visual-20sec: sequence of 20 consecutive RGB frames
6. Audiovisual-20sec: sequences of 20 spectrograms and RGB frames

DATASET:

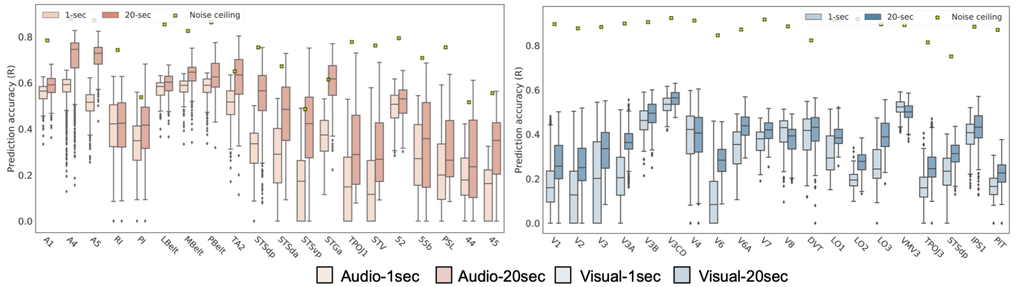
- 7T fMRI measurements from 158 subjects watching 4 audio-visual movies
- *Training/validation set*: 2265 stimulus-response pairs; *Testing set:* 699 stimulus-response pairs

RESULTS

Observation 1: Temporal history and multi-sensory information are mostly very useful in higher order brain regions

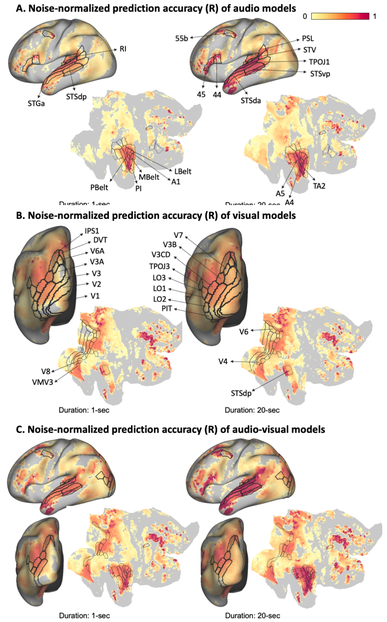


Observation 2: Among dominantly uni-sensory regions, longer time-scales improve performance in higher order auditory regions and dorsal visual stream



RESULTS (contd.)

Observation 3: Auditory and visual features jointly approach noise ceiling in multi-sensory regions



CONCLUSIONS

- Exploiting knowledge about stimulus along the axes of timescales and sensory modality substantially increases prediction accuracy of neural responses
- Encoding models can form an alternate framework to probe the preferred time-scales and sensory modality of different brain regions

REFERENCES

- Richards, B.A., et al. "A deep learning framework for neuroscience". *Nat Neurosci* 22 (2019)
- Yamins, D., et al. "Using goal-driven deep learning models to understand sensory cortex". *Nat Neurosci* 19 (2016)
- Kell, A.J. et al. "A task-optimized neural network replicates human auditory behavior, predicts brain responses and reveals a cortical processing hierarchy". *Neuron* (2018)
- Varoquaux, G. et al. "Predictive models avoid excessive reductionism in cognitive neuroimaging". *Curr Opin Neurol (2019)*

Presented at OHBM 2020. For more information, please feel free to reach out to Meenakshi Khosla at mk2299@cornell.edu