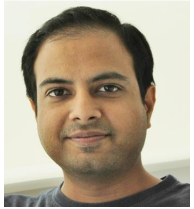
Deep Learning for Brain Encoding and Decoding

Subba Reddy Oota[1], Jashn Arora[2], Manish Gupta[2,3], Raju S. Bapi[2], Mariya Toneva[4,5]

[1]Inria Bordeaux, France; [2]IIIT Hyderabad, India; [3]Microsoft, India; [4]Princeton Neuroscience Institute, USA; [5]MPI for Software Systems, Germany

subba-reddy.oota@inria.fr, jashn.arora@research.iiit.ac.in, gmanish@microsoft.com, raju.bapi@iiit.ac.in, mtoneva@mpi-sws.org





Agenda

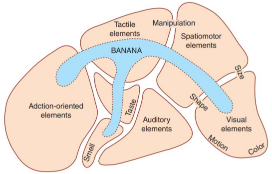
- Introduction to Brain encoding and decoding [30 min]
- Stimulus Representations (Theory + Hands-on) [1 hour 30 min]
- Coffee break [15 min]
- **Deep Learning for Brain Encoding (Theory + Hands-on) [1 hour 30 min]**
- Lunch break [1 hour 15 min]
- Deep Learning for Brain Decoding (Theory + Hands-on) [1 hour 30 min]
- Coffee break [15 min]
- Advanced Methods [1 hour]
- Summary and Future Trends [15 min]

Agenda

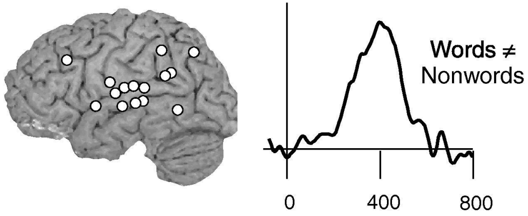
- Introduction to Brain encoding and decoding [30 min]
- Stimulus Representations (Theory + Hands-on) [1 hour 30 min]
- Coffee break [15 min]
- **Deep Learning for Brain Encoding (Theory + Hands-on) [1 hour 30 min]**
    - **Classic findings & common approaches**
    - More recent findings utilizing deep learning
    - Hands-on with multi-modal fMRI data [40 min]
- Lunch break [1 hour 15 min]
- Deep Learning for Brain Decoding (Theory + Hands-on) [1 hour 30 min]
- Coffee break [15 min]
- Advanced Methods [1 hour]
- Summary and Future Trends [15 min]

Mechanistic understanding of information processing in the brain: 4 big questions

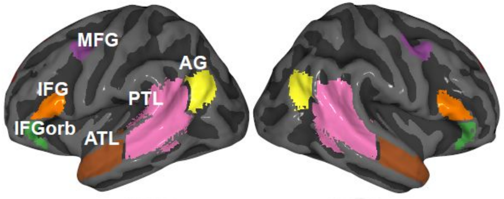
What



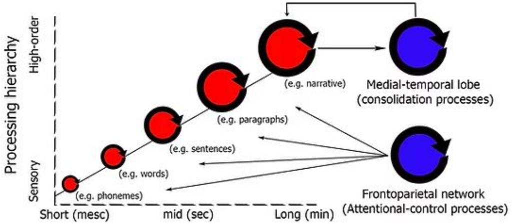
When



Where



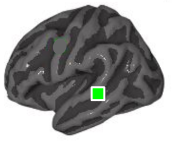
How



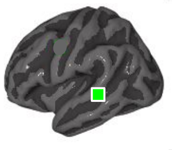
Encoding models have a causal interpretation

Reveal which brain areas are affected by stimulus properties [Weichwald et al. 2015]

"The problem is when the capsule moves from an elliptical orbit to a parabolic orbit."



y_{train}



Train: $f(\langle 0,1,\ldots 0\rangle) \approx$

Evaluate: $\mathrm{corr}(Y_{test}, \hat{Y}_{test})$

Part of Speech: Noun

<0, 1, … 0>

stimulus properties

m. representation

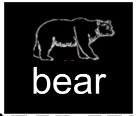
stimulus representation = hypothesis for latent brain-relevant stimulus properties

Classic findings using encoding models

- Using representations of stimuli not from deep learning
- Language:
  - Mitchell et al. 2008, Science
- Vision:
  - Kay et al. 2008, Nature
- Audio:
  - Santoro et al. 2014, PLoS Comp Bio

Classic encoding model finding: Language

- Stimuli: concrete nouns + line drawings



- Stimulus representation: corpus co-occurrence counts with 25 sensory-motor verbs (e.g. see, hear, taste, smell)

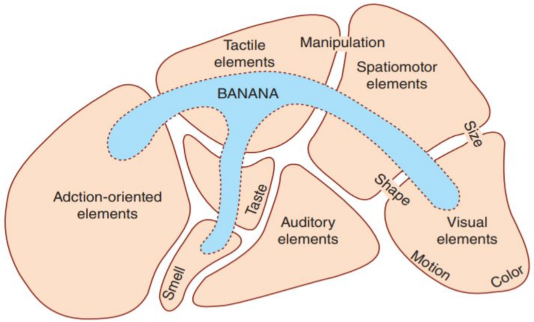


figure from Kemmerer, 2014; adapted from Thompson-Schill et al. 2006

[Barsalou, 1999; Barsalou, 2008; Pecher et al., 2005]

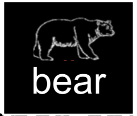
Empirical evidence for distributed organization for attributes related to:

- audition [Kiefer et al., 2008]
- color [Simmons et al., 2007]
- shape [Chao et al., 1999]
- motion [Damasio et al., 1996]
- olfaction and taste [Goldberg, Perfetti, et al., 2006a; Goldberg, Perfetti, et al., 2006b]

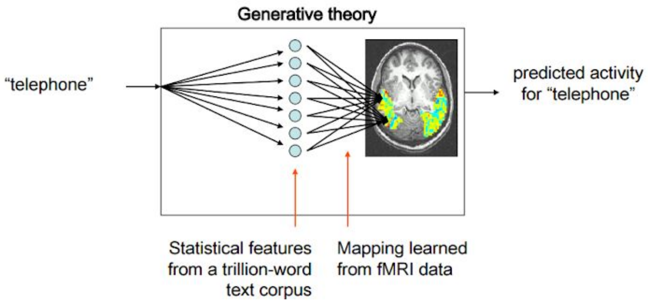
Mitchell, Tom M., Svetlana V. Shinkareva, Andrew Carlson, Kai-Min Chang, Vicente L. Malave, Robert A. Mason, and Marcel Adam Just. "Predicting human brain activity associated with the meanings of nouns." science 320, no. 5880 (2008): 1191-1195.

Classic encoding model finding: Language

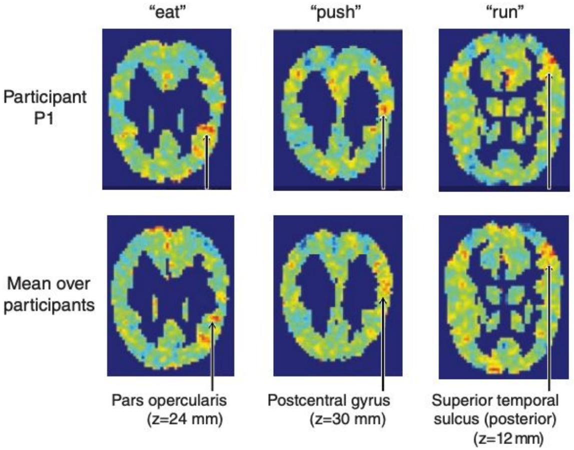
- Stimuli: concrete nouns + line drawings



- Stimulus representation: corpus co-occurrence counts with 25 sensory-motor verbs (e.g. see, hear, taste, smell)
- Brain recording: fMRI



Accurately predicts fMRI recordings for a novel word



Correspondences between a semantic property ("push") and the function of the cortical regions where the fMRI recordings are well predicted

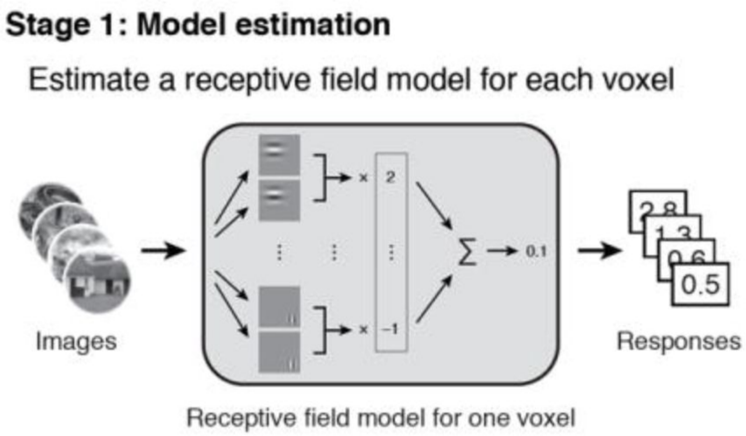
Mitchell, Tom M., Svetlana V. Shinkareva, Andrew Carlson, Kai-Min Chang, Vicente L. Malave, Robert A. Mason, and Marcel Adam Just. "Predicting human brain activity associated with the meanings of nouns." science 320, no. 5880 (2008): 1191-1195.

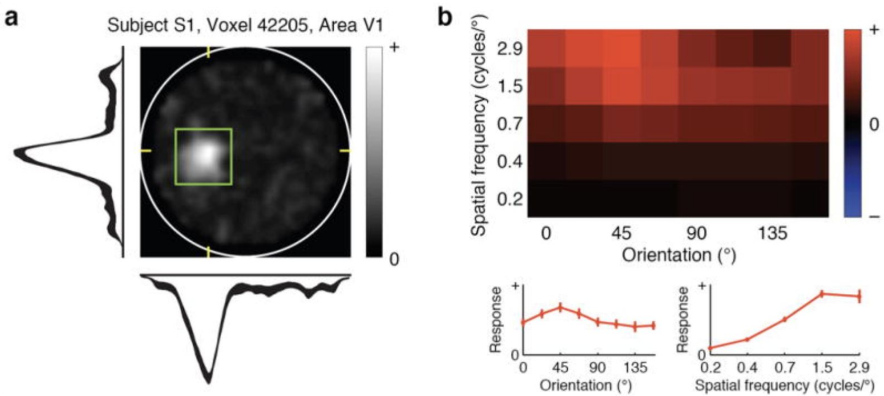
Classic encoding model finding: Vision

- Stimuli: natural images
- Stimulus representation: mixtures of Gabor wavelets
- Brain recording & modality: fMRI, viewing

Encoding models estimated quantitative receptive fields for V1-V3 voxels

Identified which of a set of candidate natural image was viewed by a participant

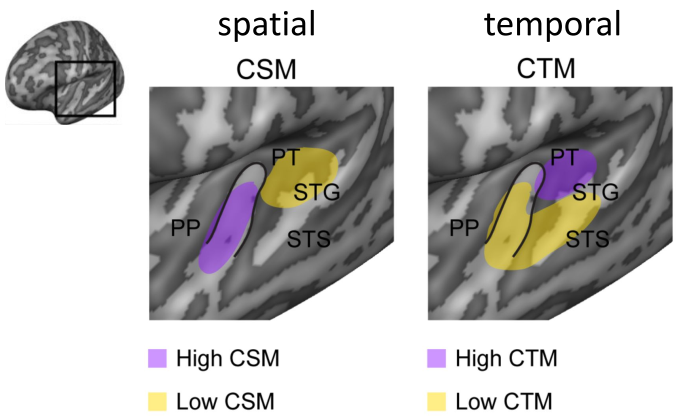




Kay, Kendrick N., Thomas Naselaris, Ryan J. Prenger, and Jack L. Gallant. "Identifying natural images from human brain activity." *Nature* 452, no. 7185 (2008): 352-355.

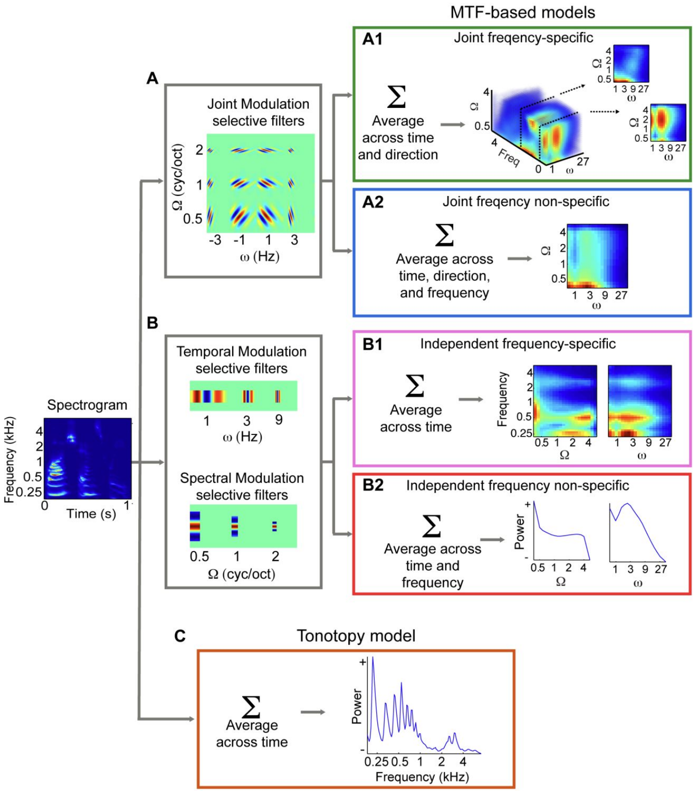
Classic encoding model finding: Audio

- Stimuli: natural sounds (speech, music, nature, tools)
- Stimulus representation: spectro-temporal filters that are selective for modulations along space and/or time
- Brain recording & modality: fMRI, listening



posterior/dorsal auditory: coarse spectral info & high temporal precision

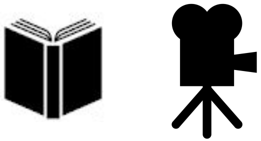
anterior/ventral auditory: fine-grained spectral & low temporal precision



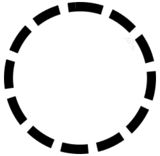
[Santoro, Roberta, Michelle Moerel, Federico De Martino, Rainer Goebel, Kamil Ugurbil, Essa Yacoub, and Elia Formisano. "Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex." PLoS computational biology 10, no. 1 (2014): e1003412.]

Deep learning models enable data-driven encoding models for naturalistic stimuli

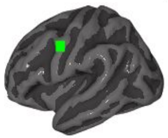
more naturalistic stimuli



more stimulus properties that affect brain activity



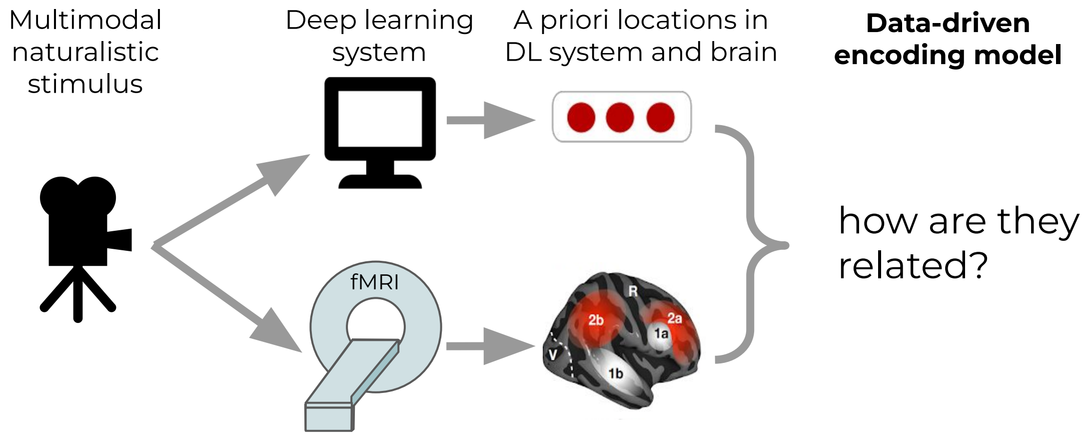
simple stim. representations explain less variance in brain activity

$f(\langle 0,1,\ldots 0 \rangle) \approx$ 

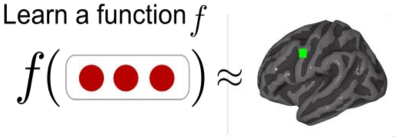
DeepMind’s New AI Taught Itself to Be the World’s Greatest Go Player
Singularity Hub

Meet GPT-3. It Has Learned to Code (and Blog and Argue)
The New York Times

Data-driven encoding models evaluate the relationships between brains and deep learning models



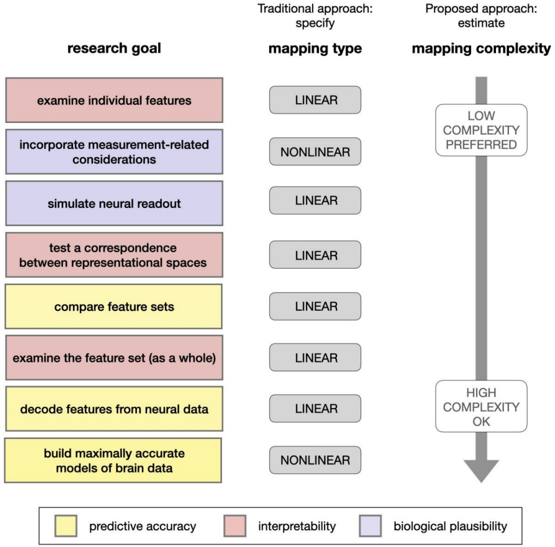
Encoding: **training and evaluation**



function f often modeled as linear

[Mitchell et al. 2008, Nishimoto et al., 2011; Sudre et al., 2012; Wehbe et al., 2014]

Considerations for Linear vs non-linear f

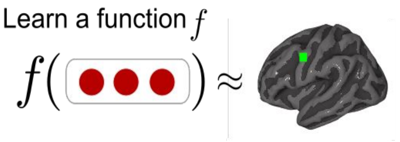


Ivanova, Anna A., Martin Schrimpf, Stefano Anzellotti, Noga Zaslavsky, Evelina Fedorenko, and Leyla Isik. "Is it that simple? Linear mapping models in cognitive neuroscience." bioRxiv (2021).

Encoding: **training and evaluation**

function f often modeled as linear

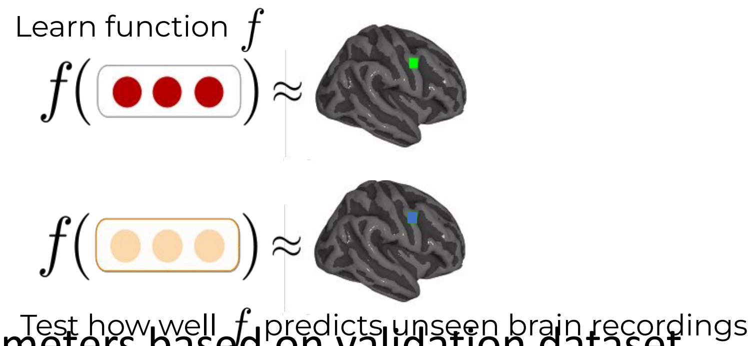
[Mitchell et al. 2008, Nishimoto et al., 2011; Sudre et al., 2012; Wehbe et al., 2014]



Training: cross validation (CV), regularization parameter chosen via nested CV

Evaluation: 1) make predictions for heldout data
2) compare predictions with true brain data
3) stringent statistical testing

Encoding: training **setup**

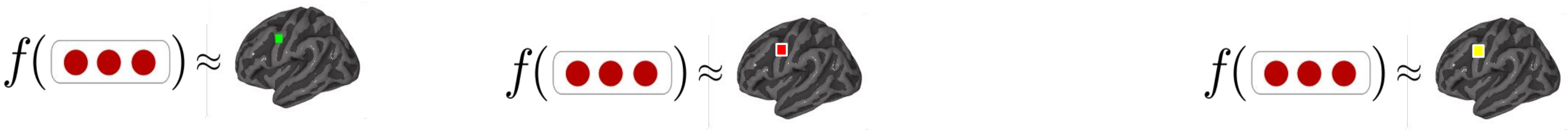


- Goal: find a mapping from stimulus representation to brain data that **generalizes** to new brain data
- Method:
  - Split dataset into train, validation, and test
  - Employ cross-validation to select model parameters based on validation dataset
  - Reduce overfitting by using regularization
    - Ridge regularization

Encoding: training **independent** models

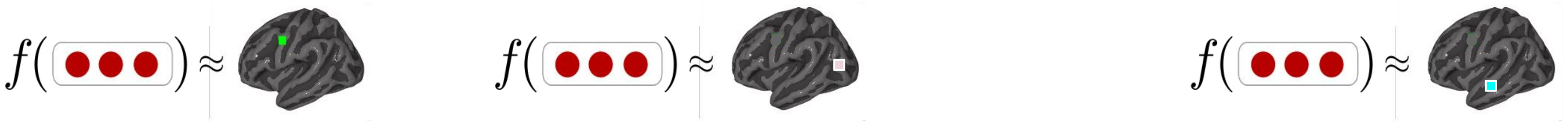
- Independent model per participant

P1 P2 … PN



- Independent model per voxel / sensor-timepoint

P1, v1 P1, v2 … P1, vm



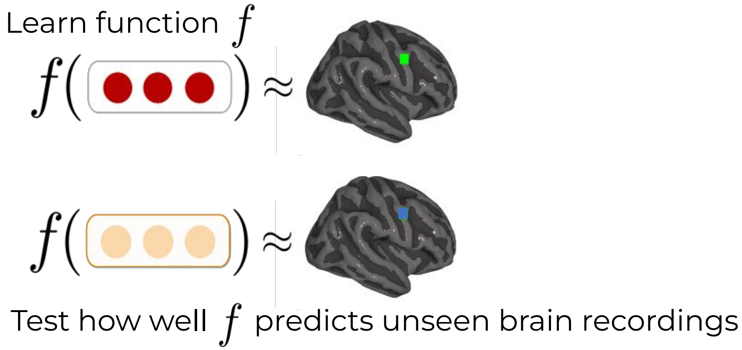
Encoding: fMRI specifics

Jain, Shailee, Vy Vo, Shivangi Mahto, Amanda LeBel, Javier S. Turek, and Alexander Huth. "Interpretable multi-timescale models for predicting fMRI responses to continuous natural speech." Advances in Neural Information Processing Systems 33 (2020): 13738-13749.

Encoding: evaluation **setup**

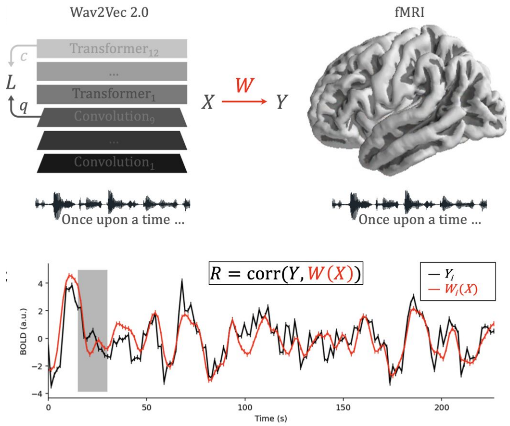
- Predict data heldout from training by applying learned function to corresponding stimulus representations

- Compare predictions of brain data to true brain data:
    - Evaluation metrics



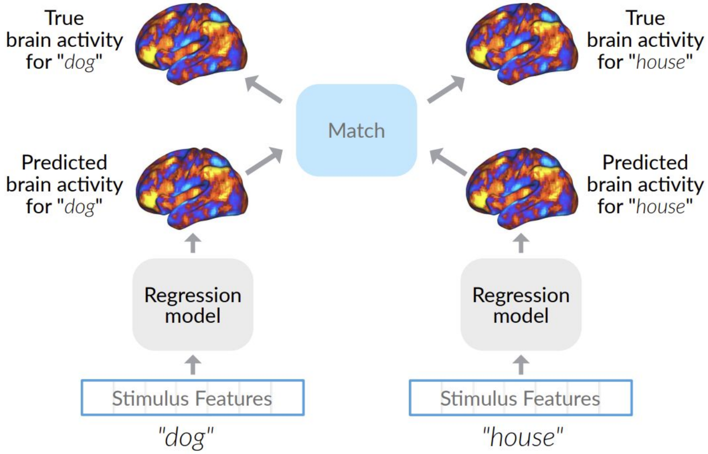
Encoding: evaluation **metrics**

Pearson correlation



Millet, Juliette, Charlotte Caucheteux, Pierre Orhan, Yves Boubenec, Alexandre Gramfort, Ewan Dunbar, Christophe Pallier, and Jean-Remi King. "Toward a realistic model of speech processing in the brain with self-supervised learning." arXiv preprint arXiv:2206.01685 (2022).

2v2 accuracy



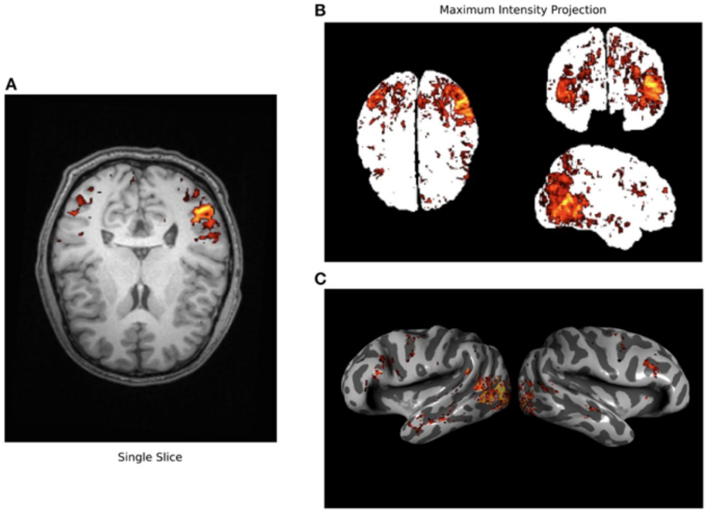
Toneva, Mariya, Otilia Stretcu, Barnabás Póczos, Leila Wehbe, and Tom M. Mitchell. "Modeling task effects on meaning representation in the brain via zero-shot meg prediction." Advances in Neural Information Processing Systems 33 (2020): 5284-5295.

Encoding: statistical **significance**

- Goal: determine whether the estimated similarity between the DL representations and the brain recordings is significant
- Simple method that makes no assumptions about underlying data:
    - Permutation test
        - Break input-to-output correspondence by permuting output labels
        - Estimate similarity
        - Repeat 1000s times to estimate null distribution
        - P-value = proportion of times the similarity metric from permuted labels >= sim. metric from original labels
    - Specifically for fMRI:
        - Permute labels in blocks to preserve the autoregressive structure
- Correct for multiple comparisons
    - FDR, FWER, etc.

Encoding: performance **visualization**

fMRI

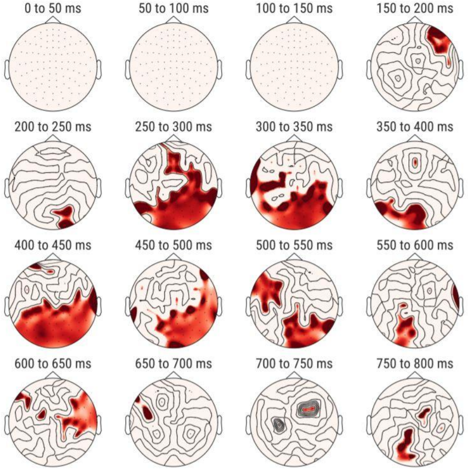


Single Slice

Inflated Surface

Gao, James S., Alexander G. Huth, Mark D. Lescroart, and Jack L. Gallant. "Pycortex: an interactive surface visualizer for fMRI." Frontiers in neuroinformatics (2015): 23.

MEG/EEG



Gramfort, Alexandre, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj et al. "MEG and EEG data analysis with MNE-Python." Frontiers in neuroscience (2013): 267.

Agenda

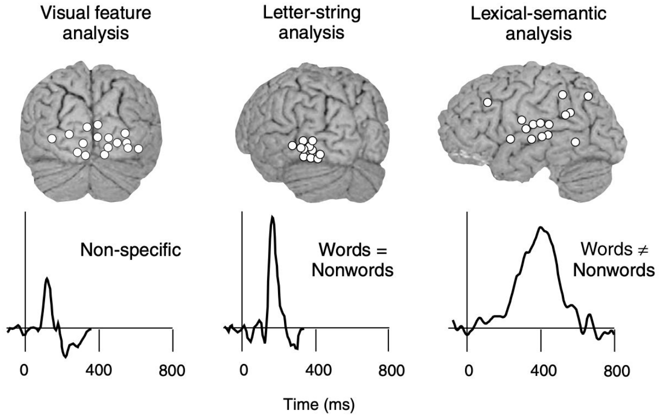
- Introduction to Brain encoding and decoding [30 min]
- Stimulus Representations (Theory + Hands-on) [1 hour 30 min]
- Coffee break [15 min]
- **Deep Learning for Brain Encoding (Theory + Hands-on) [1 hour 30 min]**
  - Classic findings & common approaches
  - **More recent findings utilizing deep learning**
  - Hands-on with multi-modal fMRI data [40 min]
- Lunch break [1 hour 15 min]
- Deep Learning for Brain Decoding (Theory + Hands-on) [1 hour 30 min]
- Coffee break [15 min]
- Advanced Methods [1 hour]
- Summary and Future Trends [15 min]

Recent work utilizing progress in DL for encoding

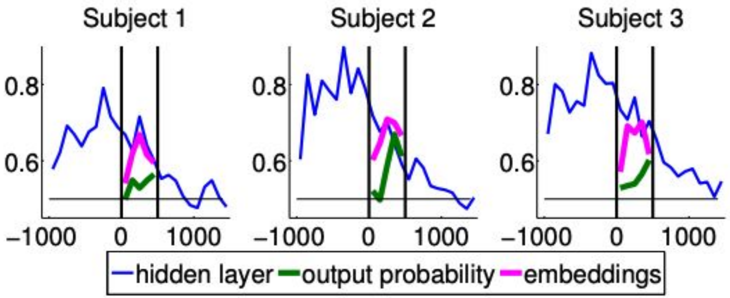
- Using representations of stimuli from deep learning systems
- **Language:**
  - Wehbe et al. 2014; Jain and Huth, 2018; Toneva and Wehbe, 2019; Caucheteux and King, 2020/2022; Schrimpf et al. 2020/2021; Goldstein et al. 2021/2022
- Vision:
  - Yamins et al. 2014; Cichy et al. 2016; Konkle and Alvarez, 2020/2022; Zhuang et al. 2022
- Audio:
  - Kell et al. 2018; Vaidya, Jain, and Huth 2022; Millet et al. 2022

Language: work utilizing DL progress

- Stimuli: one chapter of Harry Potter
- Stimulus representation: derived from an NLP system (RNN) trained on Harry Potter fan fiction
- Brain recording: MEG, reading



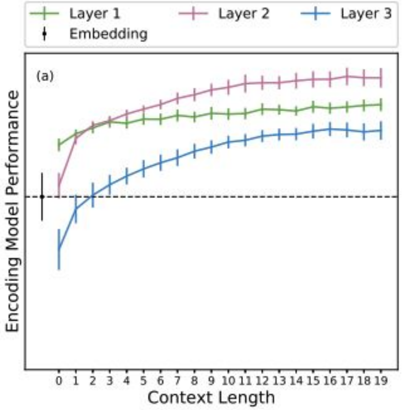
significant word-by-word alignment between MEG & representations of words and context from recurrent NLP systems

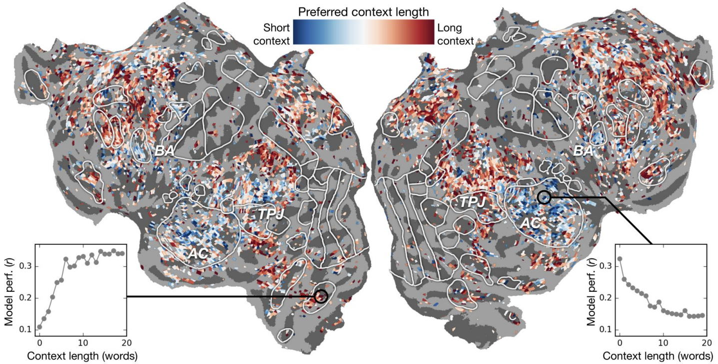


[Wehbe, Leila, Ashish Vaswani, Kevin Knight, and Tom Mitchell. "Aligning context-based statistical models of language with brain activity during reading." In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 233-243. 2014.]

Audio: work utilizing DL progress

- Stimuli: Moth Radio Hour
- Stimulus representation: derived from **self-supervised text language model** trained to predict upcoming word in other radio stories
- Brain recording & modality: fMRI, listening



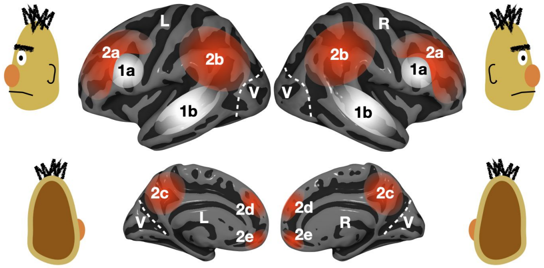


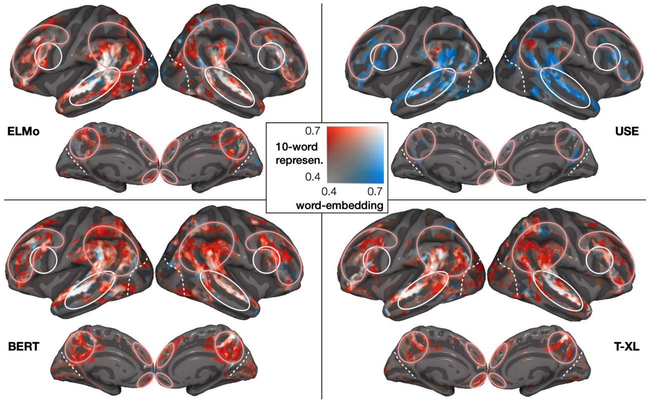
alignment between fMRI & recurrent NLP representations w/ varying context;
best alignment with middle layer

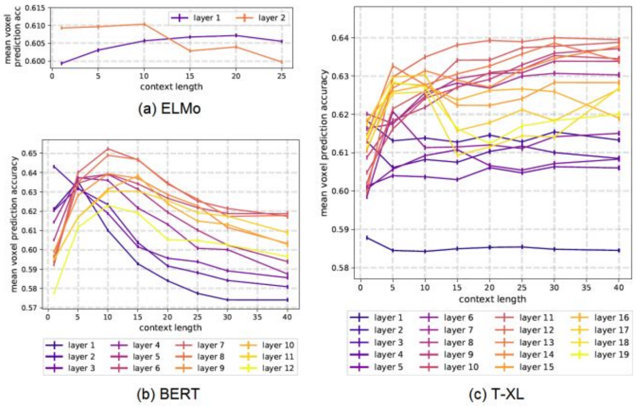
[Jain, Shailee, and Alexander Huth. "Incorporating context into language encoding models for fMRI." Advances in neural information processing systems 31 (2018).]

Language: work utilizing DL progress

- Stimuli: one chapter of Harry Potter
- Stimulus representation: derived from **pretrained** NLP systems
- Brain recording & modality: fMRI, reading







across several types of large NLP systems, best alignment with fMRI in middle layers

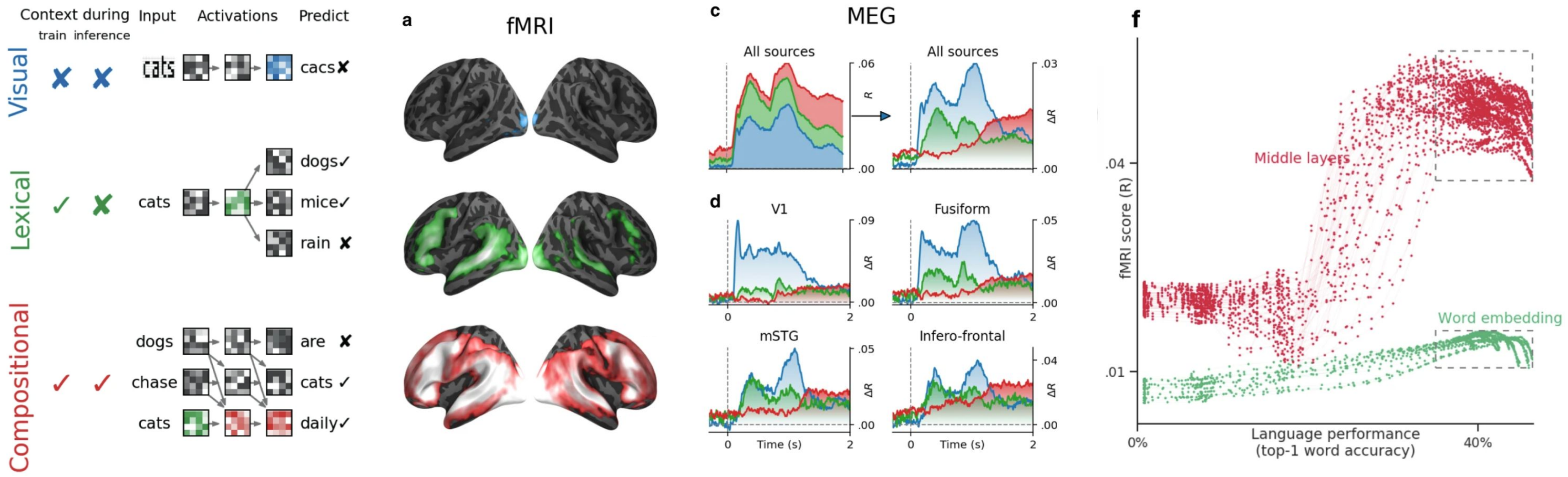
Toneva, M., & Wehbe, L. (2019). Interpreting and improving natural-language processing (in machines) with natural language-processing (in the brain). Advances in Neural Information Processing Systems, 32.

Language: work utilizing DL progress

- Stimuli: sentences
- Stimulus representation: derived from pretrained NLP systems
- Brain recording & modality: MEG & fMRI, reading

best alignment with fMRI & MEG in middle layers

better performance at predicting next word -> better prediction of fMRI & MEg

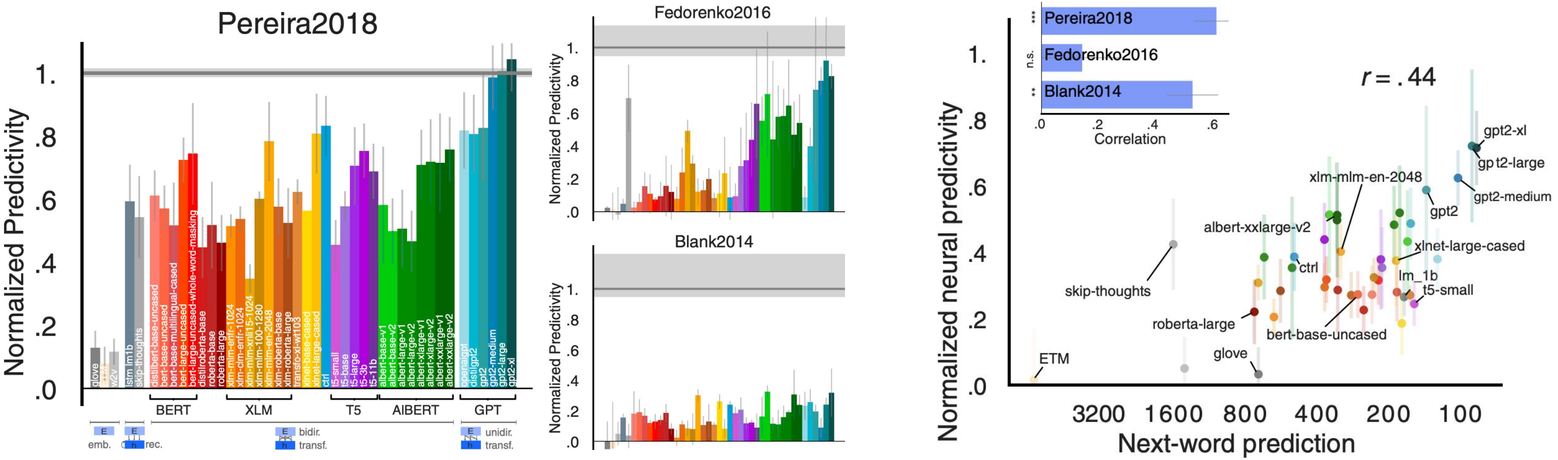


Caucheteux, Charlotte, and Jean-Rémi King. "Brains and algorithms partially converge in natural language processing." Communications biology 5, no. 1 (2022): 1-10.

Language: work utilizing DL progress

- Stimuli: sentences, passages, short story
- Stimulus representation: derived from pretrained NLP systems
- Brain recording & modality: fMRI & ECoG, reading & listening

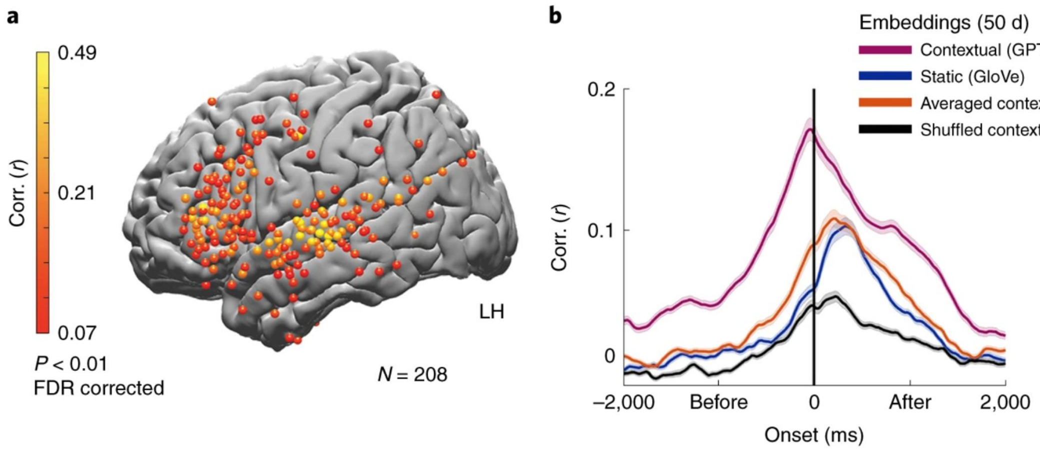
some NLP systems can predict fMRI and ECoG up to 100% of estimated noise ceiling



Schrimpf, Martin, Idan Asher Blank, Greta Tuckute, Carina Kauf, Eghbal A. Hosseini, Nancy Kanwisher, Joshua B. Tenenbaum, and Evelina Fedorenko. "The neural architecture of language: Integrative modeling converges on predictive processing." Proceedings of the National Academy of Sciences 118, no. 45 (2021): e2105646118.

Language: work utilizing DL progress

- Stimuli: story
- Stimulus representation: derived from pretrained NLP systems
- Brain recording & modality: ECoG, listening



NLP word representations predict ECoG recordings for upcoming words

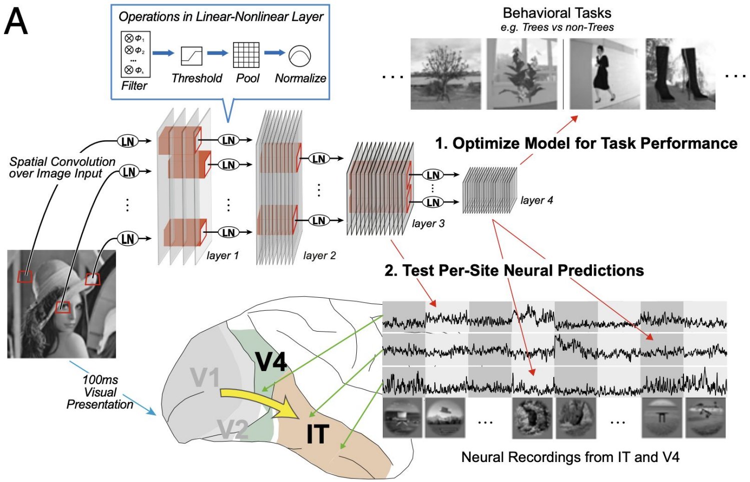
Goldstein, Ariel, Zaid Zada, Eliav Buchnik, Mariano Schain, Amy Price, Bobbi Aubrey, Samuel A. Nastase et al. "Shared computational principles for language processing in humans and deep language models." Nature neuroscience 25, no. 3 (2022): 369-380.

Recent work utilizing progress in DL for encoding

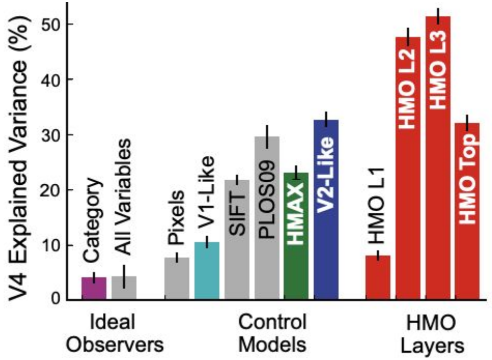
- Using representations of stimuli from deep learning systems
  - Data-driven
- Language:
  - Wehbe et al. 2014; Jain and Huth, 2018; Toneva and Wehbe, 2019; Caucheteux and King, 2020/2022; Schrimpf et al. 2020/2021; Goldstein et al. 2021/2022
- **Vision:**
  - Yamins et al. 2014; Cichy et al. 2016; Konkle and Alvarez, 2020/2022; Zhuang et al. 2022
- Audio:
  - Kell et al. 2018; Vaidya, Jain, and Huth 2022; Millet et al. 2022

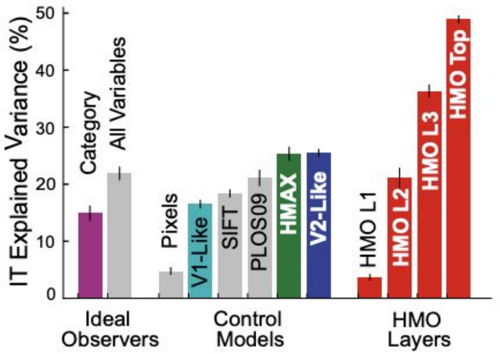
Vision: work utilizing DL progress

- Stimuli: images of natural objects
- Stimulus representation: layers in pretrained CNNs
- Brain recording & modality: multiarray recordings in rhesus macaques, vision



Highest layer in CNN model most predictive of IT; intermediate layers most predictive of V4

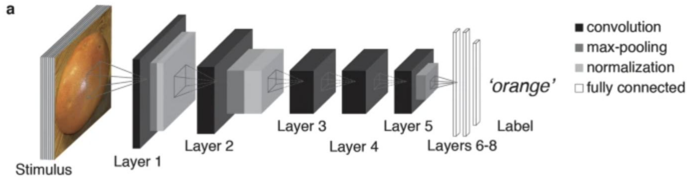


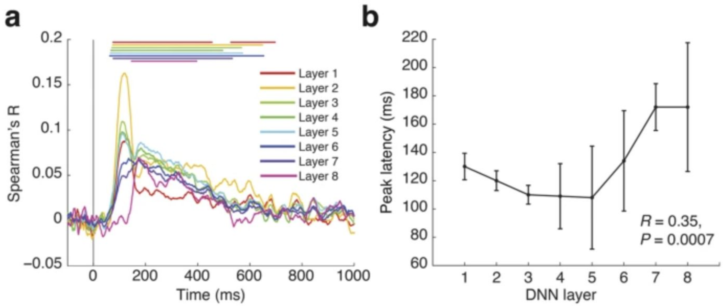


Yamins, Daniel LK, Ha Hong, Charles F. Cadieu, Ethan A. Solomon, Darren Seibert, and James J. DiCarlo. "Performance-optimized hierarchical models predict neural responses in higher visual cortex." *Proceedings of the national academy of sciences* 111, no. 23 (2014): 8619-8624.

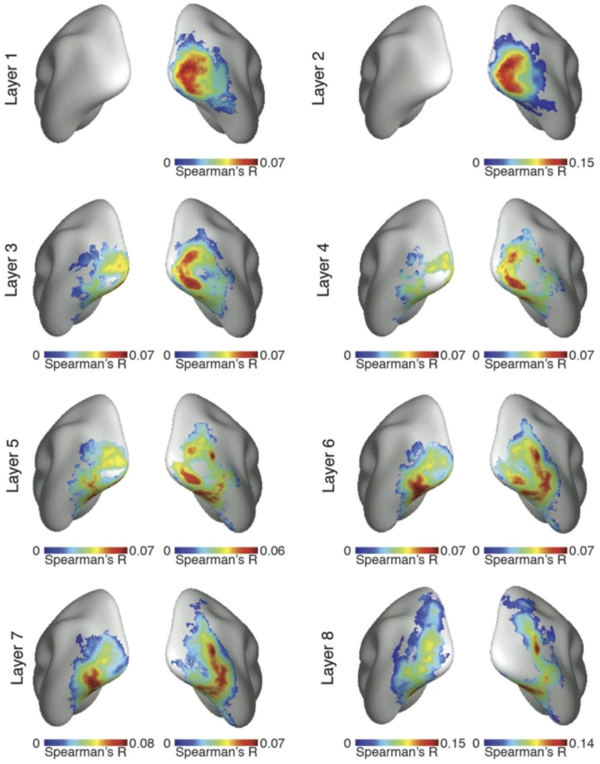
Vision: work utilizing DL progress

- Stimuli: images of natural objects
- Stimulus representation: layers of CNN tuned for object classification
- Brain recording: fMRI & MEG, vision





A CNN tuned for object classification captures stages of human visual processing in both space and time

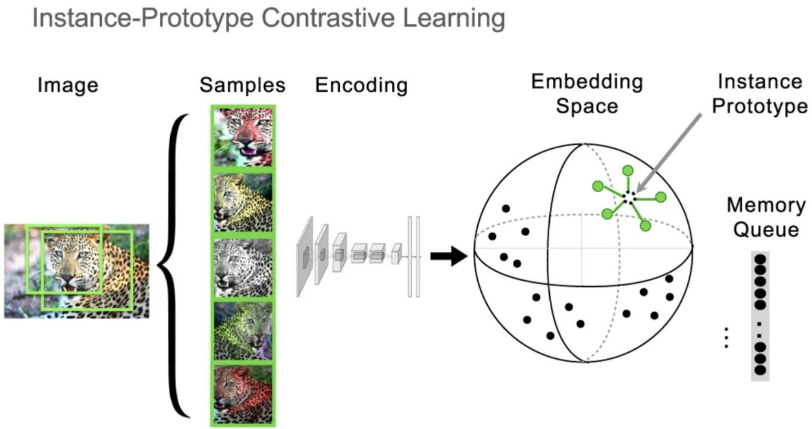


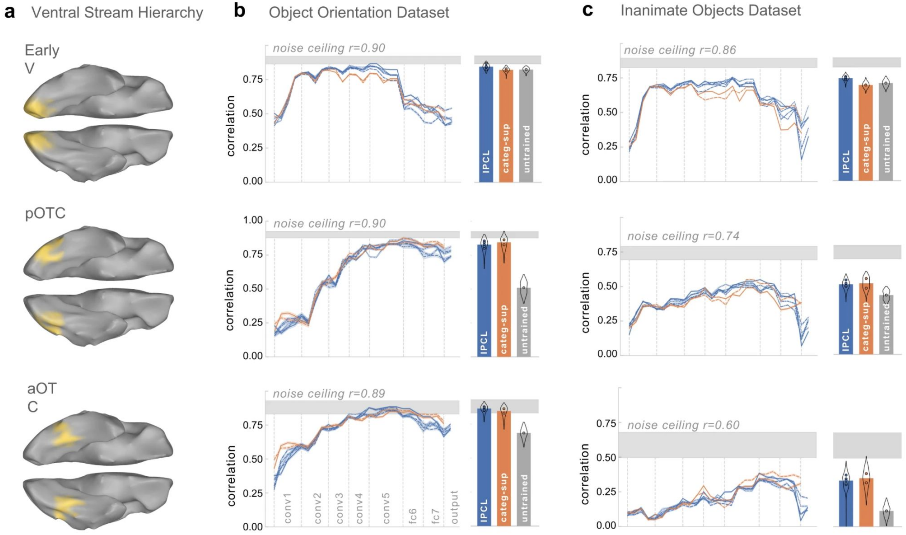
Cichy, Radoslaw Martin, Aditya Khosla, Dimitrios Pantazis, Antonio Torralba, and Aude Oliva. "Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence." Scientific reports 6, no. 1 (2016): 1-13.

Vision: work utilizing DL progress

- Stimuli: images of objects
- Stimulus representation: layers in **self-supervised** deep model
- Brain recording: fMRI, vision

Self-supervised deep models achieve parity with category-supervised models in predicting fMRI responses along visual hierarchy



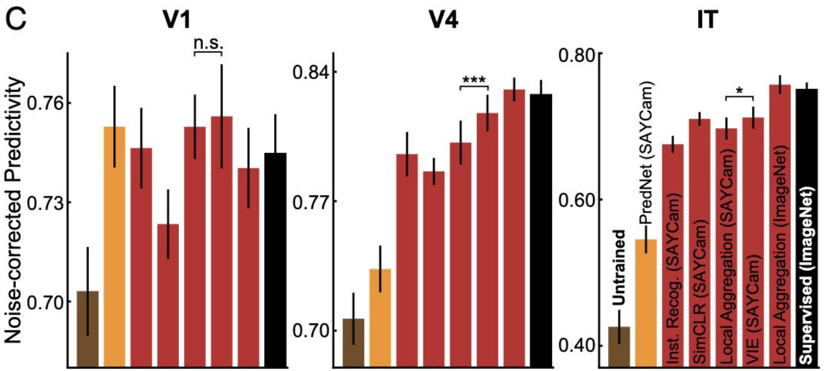


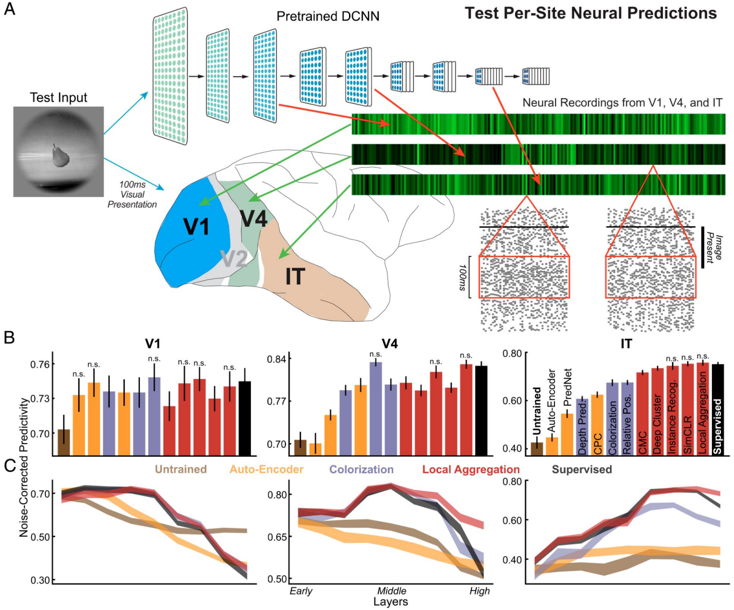
Konkle, Talia, and George A. Alvarez. "A self-supervised domain-general learning framework for human ventral stream representation." Nature communications 13, no. 1 (2022): 1-12.

Vision: work utilizing DL progress

Self-supervised deep models produce brain-like representations even when trained solely with noisy data from child head-mounted cameras

- Stimuli: images of objects
- Stimulus representation: layers in self-supervised deep model
- Brain recording: multiarray recordings in rhesus macaques, vision





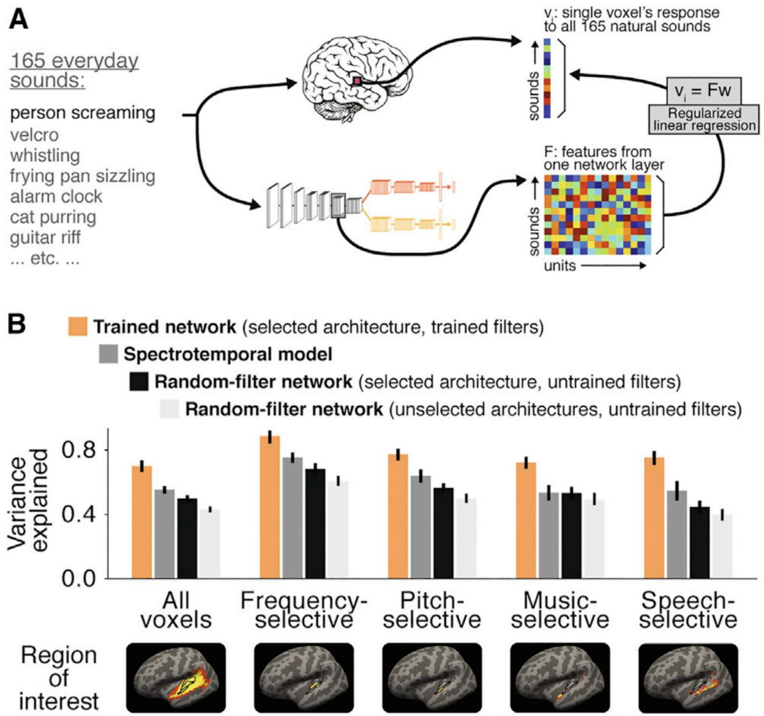
[Zhuang, Chengxu, Siming Yan, Aran Nayebi, Martin Schrimpf, Michael C. Frank, James J. DiCarlo, and Daniel LK Yamins. "Unsupervised neural network models of the ventral visual stream." Proceedings of the National Academy of Sciences 118, no. 3 (2021): e2014196118.]

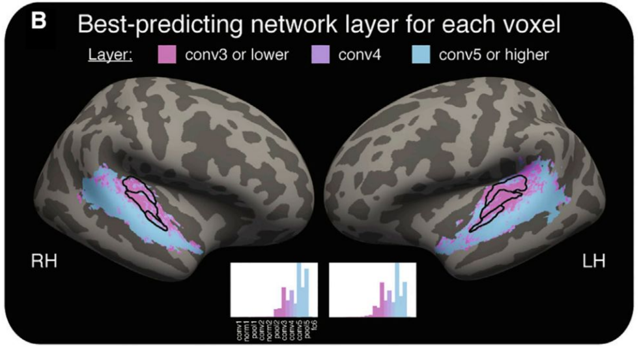
Recent work utilizing progress in DL for encoding

- Using representations of stimuli from deep learning systems
  - Data-driven
- Language:
  - Wehbe et al. 2014; Jain and Huth, 2018; Toneva and Wehbe, 2019; Caucheteux and King, 2020/2022; Schrimpf et al. 2020/2021; Goldstein et al. 2021/2022
- Vision:
  - Yamins et al. 2014; Cichy et al. 2016; Konkle and Alvarez, 2020/2022; Zhuang et al. 2022
- **Audio:**
  - Kell et al. 2018; Vaidya, Jain, and Huth 2022; Millet et al. 2022

Audio: work utilizing DL progress

- Stimuli: natural sounds
- Stimulus representation: deep model optimized for speech and music recognition
- Brain recording & modality: fMRI, listening





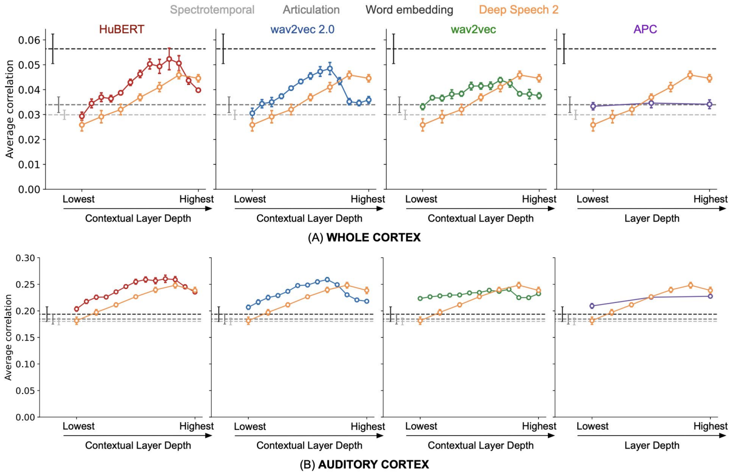
Primary auditory responses predicted best by intermediate layers of task-optimized model; non-primary responses predicted best by late layers

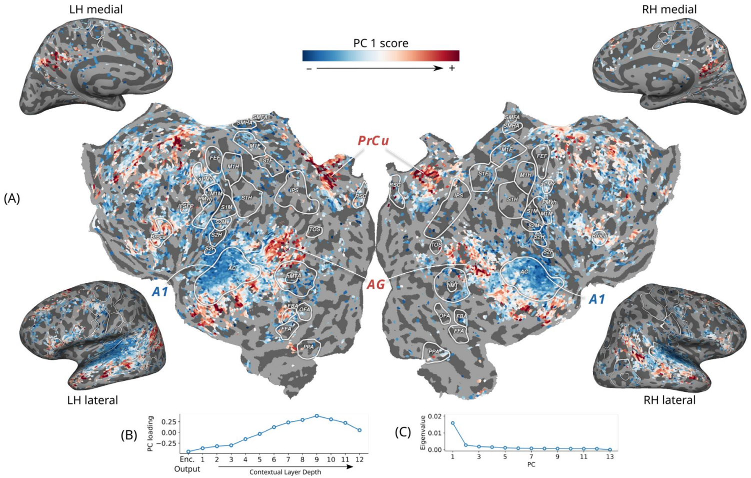
Kell, Alexander JE, Daniel LK Yamins, Erica N. Shook, Sam V. Norman-Haignere, and Josh H. McDermott. "A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy." Neuron 98, no. 3 (2018): 630-644.

Audio: work utilizing DL progress

- Stimuli: Moth Radio Hour
- Stimulus representation: derived from pretrained **self-supervised speech models**
- Brain recording & modality: fMRI, listening

Middle layers of self-supervised speech models predict auditory cortex the best

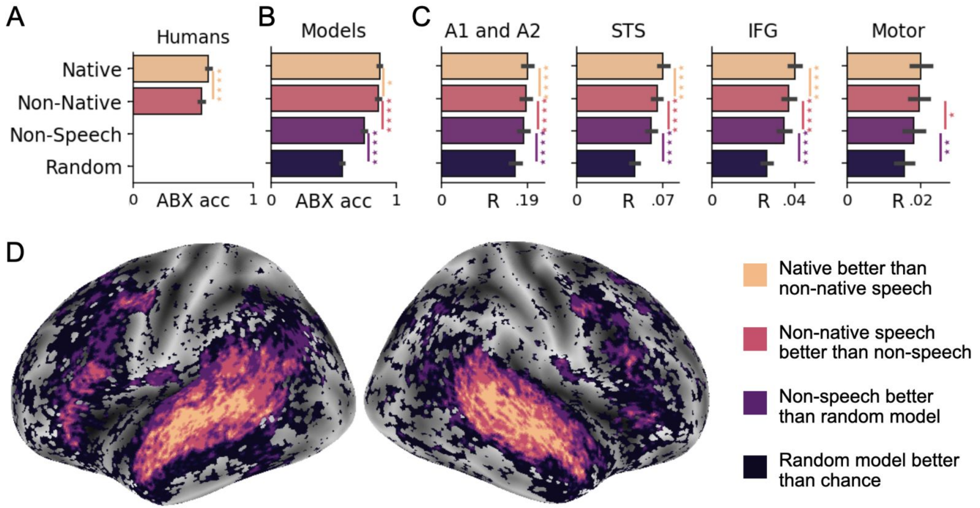




Vaidya, Aditya R., Shailee Jain, and Alexander G. Huth. "Self-supervised models of audio effectively explain human cortical responses to speech." ICML (2022).

Audio: work utilizing DL progress

- Stimuli: audio books
- Stimulus representation: derived from pretrained self-supervised speech model
- Brain recording & modality: fMRI, listening in 3 languages (Eng, Fr, Mandarin)



Self-supervised speech models exhibit specialization for native sounds in the STS and MTG; IFG and AG show more general specialization for speech rather than native-language

Millet, Juliette, Charlotte Caucheteux, Pierre Orhan, Yves Boubenec, Alexandre Gramfort, Ewan Dunbar, Christophe Pallier, and Jean-Remi King. "Toward a realistic model of speech processing in the brain with self-supervised learning." arXiv preprint arXiv:2206.01685 (2022).

Agenda

- Introduction to Brain encoding and decoding [30 min]
- Stimulus Representations (Theory + Hands-on) [1 hour 30 min]
- Coffee break [15 min]
- **Deep Learning for Brain Encoding (Theory + Hands-on) [1 hour 30 min]**
    - Classic findings & common approaches
    - **More recent findings utilizing deep learning**
    - Hands-on with multi-modal fMRI data [40 min]
- Lunch break [1 hour 15 min]
- Deep Learning for Brain Decoding (Theory + Hands-on) [1 hour 30 min]
- Coffee break [15 min]
- Advanced Methods [1 hour]
- Summary and Future Trends [15 min]