Deep Neural Networks and Brain Alignment: Brain Encoding and Decoding

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

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

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




Agenda

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

Agenda

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

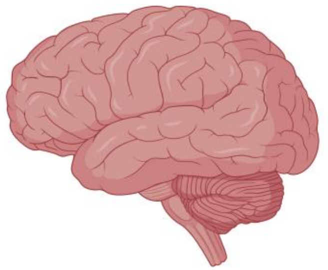
Agenda

- Introduction to Brain encoding and decoding [30 min]
  - Brain Encoding/Decoding: Techniques and Research Goals
  - Introduction to popular datasets
    - Text, Visual, Audio, Multi-modal
- Stimulus Representations [1 hour]
- Coffee break [30 min]
- Deep Learning for Brain Decoding [1 hour 30 min]
- Lunch break [1 hour 30 min]
- Deep Learning for Brain Encoding [1 hour 30 min]
- Coffee break [30 min]
- Advanced Methods [1 hour 15 min]
- Summary and Future Trends [15 min]

Agenda

- Introduction to Brain encoding and decoding [30 min]
  - **Brain Encoding/Decoding: Techniques and Research Goals**
  - Introduction to popular datasets
    - Text, Visual, Audio, Multi-modal
- Stimulus Representations [1 hour]
- Coffee break [30 min]
- Deep Learning for Brain Decoding [1 hour 30 min]
- Lunch break [1 hour 30 min]
- Deep Learning for Brain Encoding [1 hour 30 min]
- Coffee break [30 min]
- Advanced Methods [1 hour 15 min]
- Summary and Future Trends [15 min]

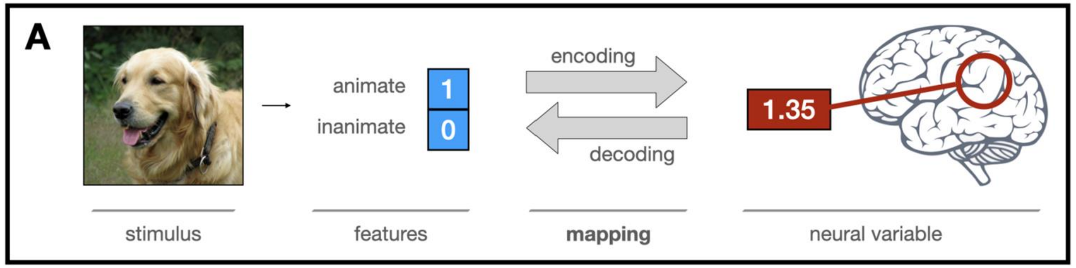
Neuroscience

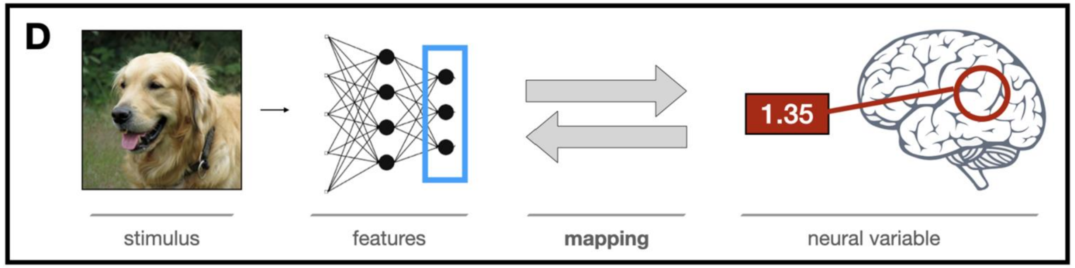


- Field of science that studies the structure and function of the nervous system of different species.
- Involves answering interesting questions
  - How learning occurs during adolescence, and how it differs from the way adults learn and form memories.
  - Which specific cells in the brain (and what connections they form with other cells), have a role in how memories are formed.
  - How animals cancel out irrelevant information arriving from the senses and focus only on information that matters.
  - How do humans make decisions.
  - How humans develop speech and learn languages.
- Neuroscientists study diverse topics that help us understand how the brain and nervous system work.

Brain encoding and decoding in cognitive neuroscience

- Encoding is the process of learning the mapping e from the stimuli S to the neural activation F.
  - Using feature engg (A) or deep learning (D)
- Decoding constitutes learning mapping d, which predicts stimuli S back from the brain activation F.
  - Oftentimes, we predict a stimulus representation R rather than actually reconstructing S.
- Other forms of encoding/decoding
  - (B): Map participants' behaviour to neural variables.
  - (C): Mapping between activity in different brain regions.



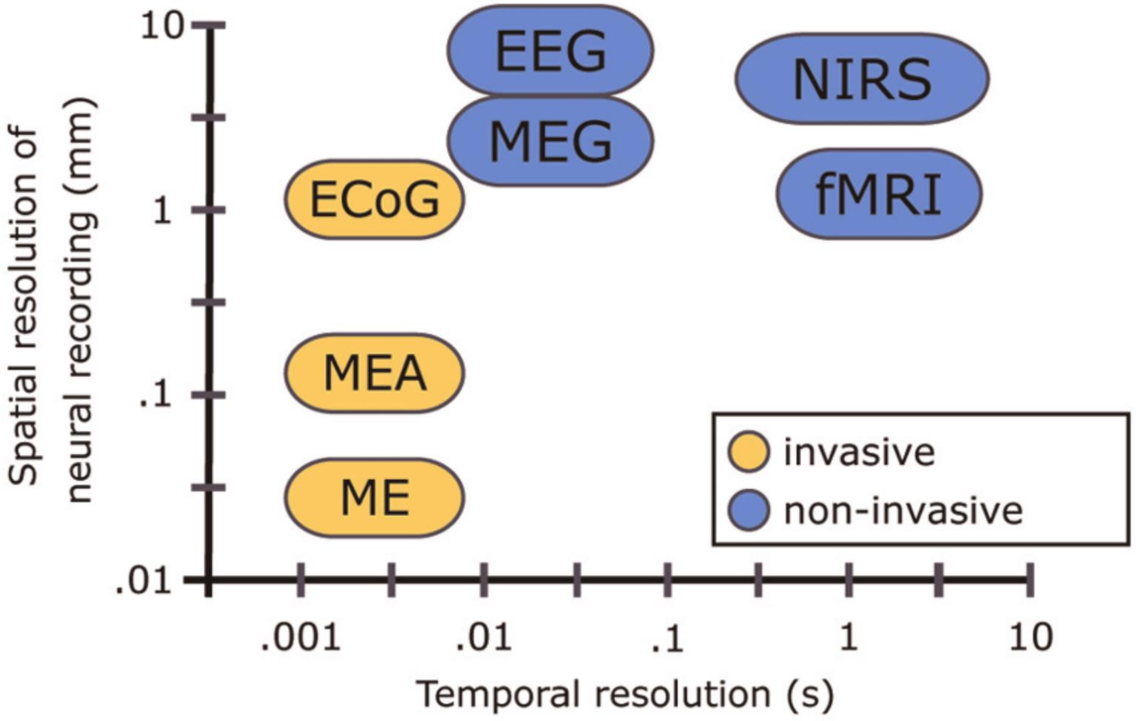


Ivanova, Anna A., Martin Schrimpf, Stefano Anzellotti, Noga Zaslavsky, Evelina Fedorenko, and Leyla Isik. "Is it that simple? Linear mapping models in cognitive neuroscience." *bioRxiv* (2021).

Brain encoding and decoding

- For both encoding and decoding, the first step is to learn a stimulus representation R of the stimuli S at the train time.
- F is the brain response.
- Next
  - For encoding, a regression function $e: R \rightarrow F$ is trained.
  - For decoding, a function $d: F \rightarrow R$ is trained.
- These functions e and d can then be used at test time to process new stimuli and brain activations, respectively.

Techniques for studying the brain function

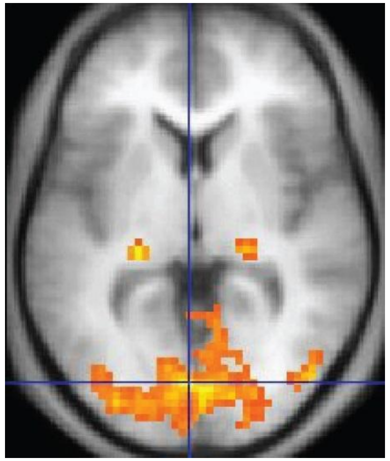


Single Micro-Electrode (ME), Micro-Electrode array (MEA), Electro-Cortico Graphy (ECoG), Positron emission tomography (PET), functional MRI (fMRI), Magneto-encephalography (MEG), Electro-encephalography (EEG), Near-Infrared Spectroscopy (NIRS)

- fMRI: high spatial but low time resolution.
  - Good to study a specific location in the brain
  - Unsuitable for sentence-level analysis. fMRI takes about two seconds to complete a scan. This is far lower than the speed at which humans can process language.
  - Cannot capture syntactic information (Gauthier and Levy, 2019)
- EEG: high time but low spatial resolution.
  - Can preserve rich syntactic information (Hale et al., 2018)
  - But cannot use for source analysis.
- fNIRS: compromise option
  - Time resolution better than fMRI
  - Spatial resolution better than EEG
  - Balance of spatial and temporal resolution may not be enough to compensate for the loss in both.

Vogel, Jörn, Sami Haddadin, Beata Jarosiewicz, John D. Simeral, Daniel Bacher, Leigh R. Hochberg, John P. Donoghue, and Patrick van der Smagt. "An assistive decision-and-control architecture for force-sensitive hand–arm systems driven by human–machine interfaces." *The International Journal of Robotics Research* 34, no. 6 (2015): 763-780.

fMRI

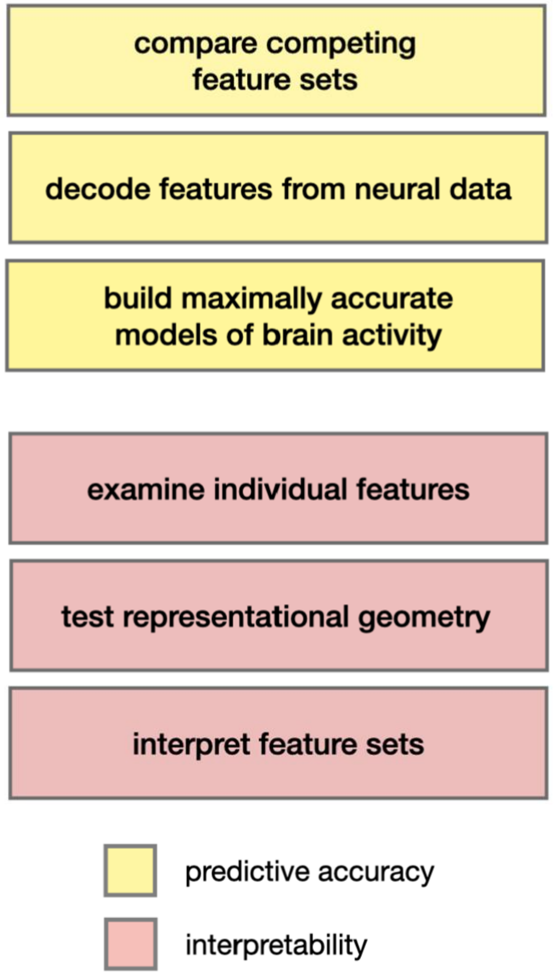


An fMRI image with yellow areas showing increased activity compared with a control condition

- No injections, surgery, the ingestion of substances, or exposure to ionizing radiation.
- The primary form of fMRI uses the blood-oxygen-level dependent (BOLD) contrast, discovered by Seiji Ogawa in 1990.
  - Measures brain activity by detecting changes associated with blood flow.
  - When an area of the brain is in use, blood flow to that region also increases.
- Hemodynamic response (HRF)
  - It takes a while for the vascular system to respond to the brain's need for glucose.
  - Blood flow lags the neuronal events triggering it by about 5 seconds.

Computational Cognitive Science Research goals

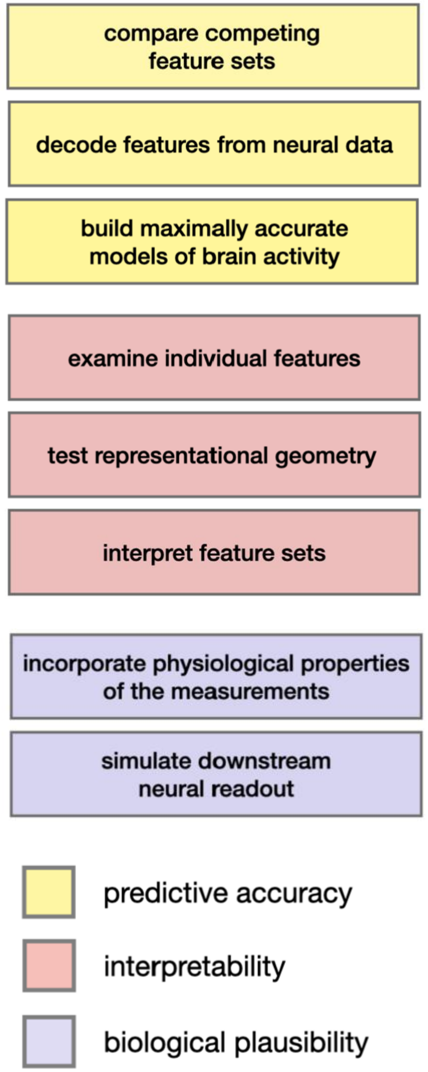
- Predictive Accuracy
  - Compare feature sets: Which feature set provides the most faithful reflection of the neural representational space?
  - Test feature decodability: "Does neural data Y contain information about features X?"
  - Build accurate models of brain data: Aim is to enable simulations of neuroscience experiments.
- Interpretability
  - Examine individual features: Which features contribute the most to neural activity?
  - Test correspondences between representational spaces
    - "CNNs vs ventral visual stream" or "Two text representations"
  - Interpret feature sets
    - Do features X, generated by a known process, accurately describe the space of neural responses Y?
    - Do voxels respond to a single feature or exhibit mixed selectivity?
  - How does the mapping relate to other models or theories of brain function?



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).

Computational Cognitive Science Research goals

- Biological plausibility
  - Simulate linear readout
    - If the features can be extracted with a linear mapping model, it means that they require few additional computations in order to be used downstream.
  - Incorporate measurement-related considerations
    - Rather than assuming a fixed HRF across voxels and/or conditions, what are better ways?



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).

Agenda

- Introduction to Brain encoding and decoding [30 min]
  - Introduction to Brain Encoding/Decoding and applications
  - **Introduction to popular datasets**
    - **Text, Visual, Audio, Multi-modal**
- Stimulus Representations [1 hour]
- Coffee break [30 min]
- Deep Learning for Brain Decoding [1 hour 30 min]
- Lunch break [1 hour 30 min]
- Deep Learning for Brain Encoding [1 hour 30 min]
- Coffee break [30 min]
- Advanced Methods [1 hour 15 min]
- Summary and Future Trends [15 min]

Types of stimuli and popular datasets

- Text (Words, Sentences, Paragraphs): Harry Potter Story, ZUCO EEG, Question-Answering MEG.
- Visual: Binary visual patterns, Natural Images (Vim-1), BOLD5000, Algonauts and SS-fMRI.
- Audio: Alice’s Adventures in Wonderland, Narratives, The Moth Radio Hour, Audio stories.
- Videos: BBC’s Doctor Who, Japanese Ads, Pippi Langkous, Algonauts.
- Other Multimodal Stimuli: Words + line drawing of concept named by each word, Pereira.

Forms of stimulus presentation and data collection

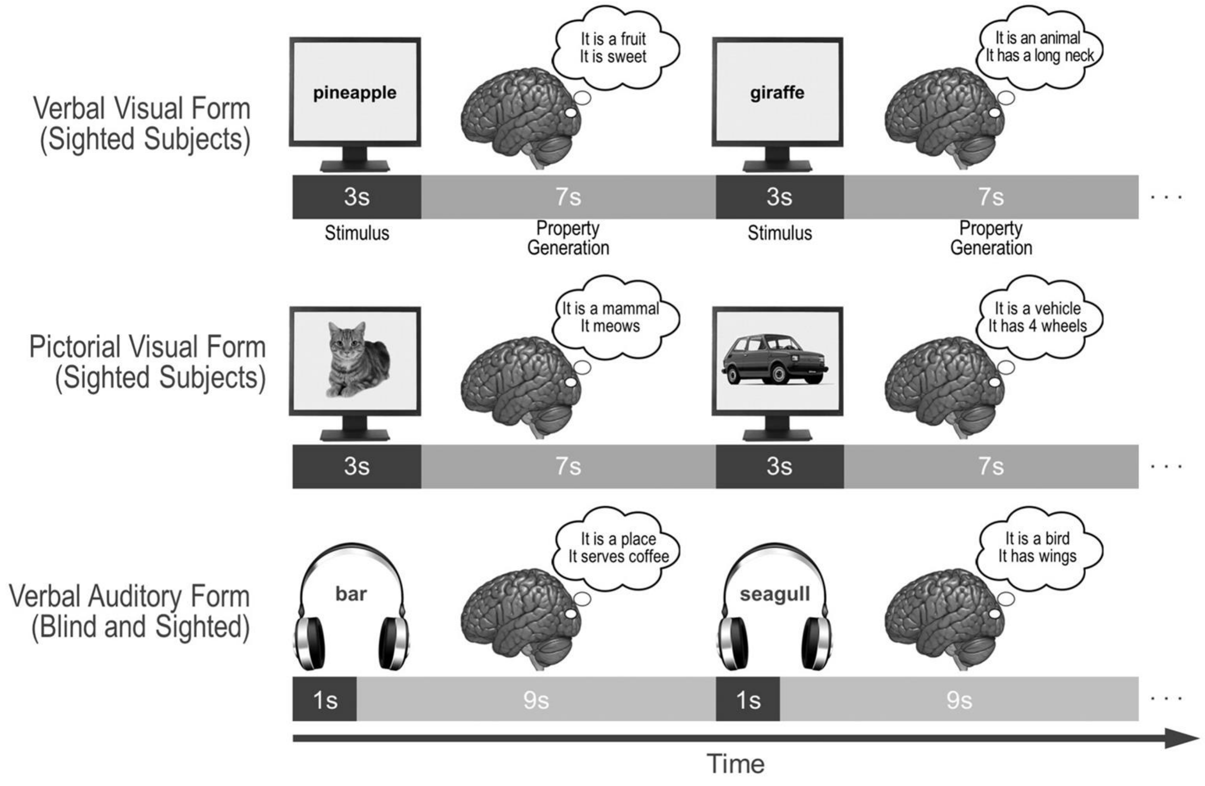
- Type: fMRI, EEG, MEG, …
- TR: Sampling time.
- Fixation points: location, color, shape.
- Form of stimuli presentation: text, video, audio, images.
- Task: question answering, property generation, understanding, …
- Time given to participants: 1 minute to list properties, …
- Type of participants: males/females, sighted/blind, …
- Number of times the response to stimuli was recorded.
- Language

Text Stimulus Datasets

Dataset	Type	Language	Stimulus	#Subjects	Paradigm	Size	Task
Wehbe et al., 2014	fMRI	English	Chapter 9 of *Harry Potter and the Sorcerer's Stone*	9	Reading stories	5000 word chapter was presented in 45 minutes.	Story understanding
Handjaras et al., 2016	fMRI	Italian	Verbal, pictorial or auditory presentation of 40 concrete nouns	20	Reading, viewing or listening	40 nouns * 4 times.	Property Generation
Anderson et al., 2017	fMRI	Italian	70 concrete and abstract nouns from law/music.	7	Reading	70 nouns * 5 times.	Imagine a situation that they personally associate with the noun
Zurich Cognitive Language Processing Corpus (ZuCo): Hollenstein et al., 2018	EEG and eye-tracking	English	Sentences from movie reviews or Wikipedia	12	Reading natural sentences	21,629 words in 1107 sentences and 154,173 fixations	Rate movie quality, answer control questions, check for existence of a relation
Anderson et al., 2019	fMRI	English	240 active voice sentences describing everyday situations	14	Reading	240 sentences seen 12 times (by 10 subjects) and 6 times (by 4 subjects)	Passive reading
BCCWJ-EEG: Oseki and Asahara, 2020	EEG	Japanese	20 newspaper articles	40	Reading	1 time reading for ~30-40 minutes	Passive reading
Deniz et al., 2019	fMRI	English	Subset of Moth Radio Hour. 11 stories	9	Reading	11 10- to 15 min stories presented twice word by word	Passive reading and Listening

Data for concrete nouns from sighted/blind subjects

- Participants were asked to verbally enumerate in one minute the properties (features) that describe the entities the words refer to.
- 4 groups of participants
  - 5 sighted individuals were presented with a pictorial form of the nouns
  - 5 sighted individuals with a verbal visual (i.e., written Italian words) form
  - 5 sighted individuals with a verbal auditory (i.e., spoken Italian words) form
  - 5 congenitally blind with a verbal auditory form.



Handjaras, Giacomo, Emiliano Ricciardi, Andrea Leo, Alessandro Lenci, Luca Cecchetti, Mirco Cosottini, Giovanna Marotta, and Pietro Pietrini. "How concepts are encoded in the human brain: a modality independent, category-based cortical organization of semantic knowledge." *Neuroimage* 135 (2016): 232-242.

70 - Italian word stimuli fMRI data

- Taxonomic categories in law and music domain
  - Ur-abstract: that are classified as abstract in WordNet
  - Attribute: A construct whereby objects or individuals can be distinguished
  - Communication: Something that is communicated by, to or between groups
  - Event/action: Something that happens at a given place and time
  - Person/Social role: Individual, someone, somebody, mortal
  - Location: Points or extents in space
  - Object/Tool: A class of unambiguously concrete nouns

	LAW		MUSIC	
Ur-abstracts	giustizia	justice	musica	music
	liberta'	liberty	blues	blues
	legge	law	jazz	jazz
	corruzione	corruption	canto	singing
	refurtiva	**loot**	punk	punk
Attribute	giurisdizione	jurisdiction	sonorita'	sonority
	cittadinanza	citizenship	ritmo	rhythm
	impunita'	impunity	~~melodia~~	~~melody~~
	legalita'	legality	tonality'	tonality
	illegalita	illegality	intonazione	pitch
Communication	divieto	prohibition	canzone	song
	verdetto	verdict	**pentagramma**	**stave**
	ordinanza	decree	ballata	ballad
	addebito	accusation	ritornello	refrain
	ingiunzione	injunction	sinfonia	symphony
Event/action	arresto	arrest	**concerto**	**concert**
	processo	trial	recital	recital
	reato	crime	assolo	solo
	furto	theft	**festival**	**festival**
	assoluzione	acquital	**spettacolo**	**show**
Person/Social-role	**giudice**	**judge**	**musicista**	**musician**
	ladro	**thief**	**cantante**	**singer**
	imputato	**defendant**	**compositore**	**composer**
	testimone	**witness**	**chitarrista**	**guitarist**
	avvocato	**lawyer**	**tenore**	**tenor**
Location	**tribunale**	**court/tribunal**	**palco**	**stage**
	carcere	**prison**	**auditorium**	**auditorium**
	~~**questura**~~	~~**police-station**~~	**discoteca**	**disco**
	penitenziario	**penitentiary**	**conservatorio**	**conservatory**
	patibolo	**gallows**	**teatro**	**theatre**
Object/Tool	**manette**	**handcuffs**	**violino**	**violin**
	toga	**robe**	**tamburo**	**drum**
	manganello	**truncheon**	**tromba**	**trumpet**
	cappio	**noose**	**metronomo**	**metronome**
	~~**grimaldello**~~	~~**skeleton-key**~~	**radio**	**radio**

Anderson, Andrew J., Douwe Kiela, Stephen Clark, and Massimo Poesio. "Visually grounded and textual semantic models differentially decode brain activity associated with concrete and abstract nouns." *Transactions of the Association for Computational Linguistics* 5 (2017): 17-30.

Zurich Cognitive Language Processing Corpus (ZuCo)

	Task 1 Normal reading (Sentiment)	Task 2 Normal reading (Wikipedia)	Task 3 Task-specific reading (Wikipedia)
Material	Positive, negative or neutral sentences from movie reviews	Wikipedia sentences containing specific relations	Wikipedia sentences containing specific relations
Example	*"The film often achieves a mesmerizing poetry."* (positive)	*"Talia Shire (born April 25, 1946) is an American actress of Italian descent."* (relations: *nationality, job title*)	*"Lincoln was the first Republican president."* (relation: *political affiliation*)
Task	Read the sentences, rating the quality of the movie based on the sentence read	Read the sentences, answer control questions	Mark whether a specific relation occurs in the given sentence or not
Control question	*"Based on the previous sentence, how would you rate this movie from 1 (very bad) to 5 (very good)?"*	*"Talia Shire was a ... 1) singer 2) actress 3) director"*	*"Does this sentence contain the* political affiliation *relation? 1) Yes 2) No"*

- Personal reading speed.
  - Sentences were presented to the subjects in a naturalistic reading scenario
  - Complete sentence is presented on the screen
  - Subjects read each sentence at their own speed, i.e., the reader determines for how long each word is fixated and which word to fixate next.

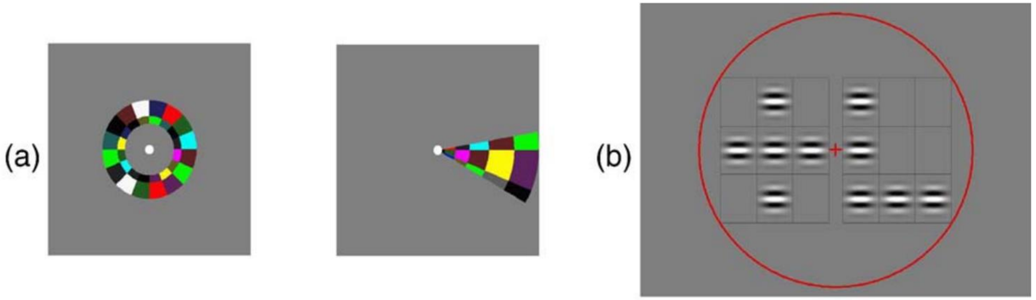
Hollenstein, Nora, Jonathan Rotsztejn, Marius Troendle, Andreas Pedroni, Ce Zhang, and Nicolas Langer. "ZuCo, a simultaneous EEG and eye-tracking resource for natural sentence reading." *Scientific data* 5, no. 1 (2018): 1-13.

Visual Stimulus Datasets

Dataset	Type	Stimulus	#S	Paradigm	Size	Task
Thirion et al., 2006	fMRI	Rotating wedges, expanding/contracting rings, rotating Gabor filters, grid	9	Viewing visual patterns	Wedges/rings for 8 times, 36 Gabor filters for 4 times, grid 36 times	Passive viewing, imagine one of the 6 domino stimuli when prompted to.
Vim-1: Kay et al., 2008	fMRI	Sequences of natural photos	2	Viewing natural images	Each subject viewed 1750 (Stage 1)+ 120 (Stage 2) novel natural images	Passive viewing
Horikawa et al., 2017	fMRI	Object images	5	Viewing and Reading	Each subject: (1) Image presentation: 1,200 images from 150 object categories and 50 images from 50 object categories; (2) Imagery: 10 times.	One-back repetition detection task, imagine object images pertaining to the category
BOLD5000: Chang et al., 2019	fMRI	5254 images depicting real-world scenes	4	Viewing natural images	~20 hours of MRI scans per each of four participants	Passive viewing
Algonauts: Cichy et al., 2019	fMRI (EVC and IT)/MEG (early and late in time)	Object images	15	Viewing object images	92 silhouette object images and 118 images of objects on natural background	Passive viewing
Natural Scenes Dataset: Allen et al., 2022	fMRI	73000 natural scenes	8	Viewing natural scenes	~73000 distinct natural scene images from MSCOCO.	Passive viewing
THINGS: Hebart et al., 2023	fMRI/EEG	31188 natural images across 1,854 object concepts.	8	Viewing natural images	fMRI: 3 Participants. 8,740 unique images. 720 objects. MEG: 4 Participants. 22,448 unique images. 1,854 objects	oddball detection task (synthetic image).

Visual Binary Patterns

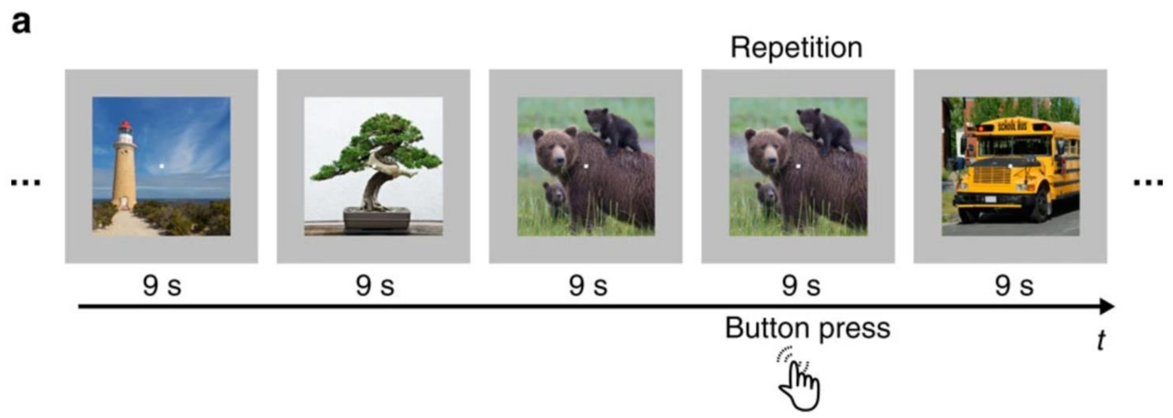
a) Retinotopic mapping experiment: flickering rotating wedges and expanding/contracting rings.
b) Domino experiment: groups of quickly rotating Gabor filters in an event-related design. Disks appeared simultaneously on the left and right side of the visual field.
c) 6 different patterns in each hemifield.
d) Subject was presented with the same grid. When the central fixation cross (left) became a left arrow (middle) or a right arrow (right), the subject had to imagine one of the 6 patterns presented previously, either in the left or right hemifield.



Thirion, Bertrand, Edouard Duchesnay, Edward Hubbard, Jessica Dubois, Jean-Baptiste Poline, Denis Lebihan, and Stanislas Dehaene. "Inverse retinotopy: inferring the visual content of images from brain activation patterns." *Neuroimage* 33, no. 4 (2006): 1104-1116.

Seen and imagined objects

- Two fMRI experiments: An image presentation experiment, and an imagery experiment.
- Image presentation experiment
  - Subjects performed a one-back repetition detection task on the images, responding with a button press for each repetition.
- Imagery experiment
  - Cue stimuli composed of an array of object names were visually presented.
  - The onset and the end of the imagery periods were signalled by auditory beeps.
  - After the first beep, the subjects were instructed to imagine as many object images as possible pertaining to the category indicated by red letters.
  - They continued imagining with their eyes closed (15 s) until the second beep.
  - Subjects were then instructed to evaluate the vividness of their mental imagery (3 s).



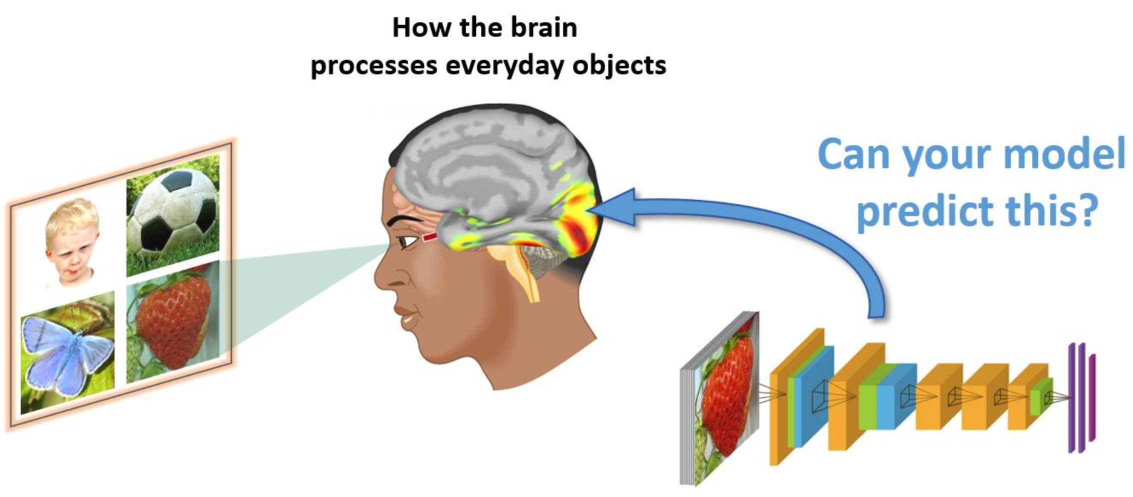
Horikawa, Tomoyasu, and Yukiyasu Kamitani. "Generic decoding of seen and imagined objects using hierarchical visual features." *Nature communications* 8, no. 1 (2017): 1-15.

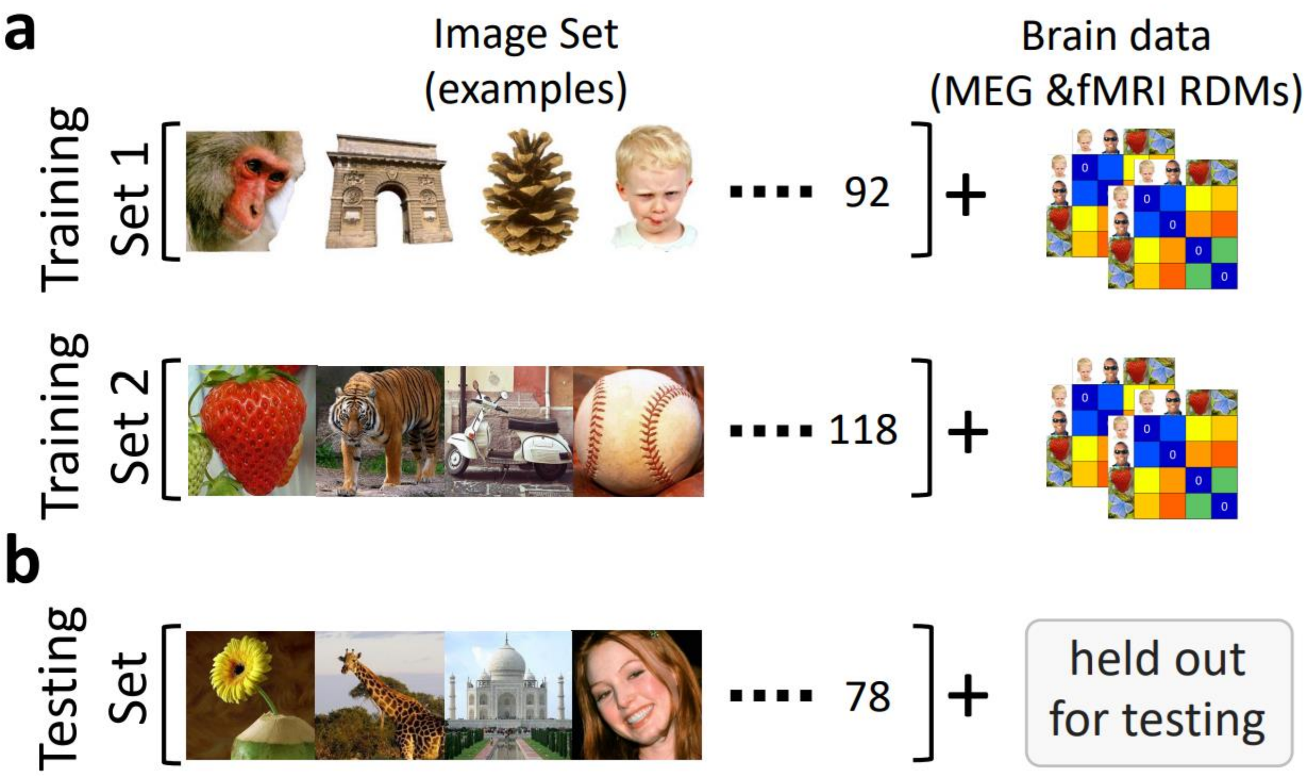
BOLD5000

- ~20 hours of MRI scans per each of the four participants.
- 4,916 unique images were used as stimuli from 3 image sources

Chang, Nadine, John A. Pyles, Austin Marcus, Abhinav Gupta, Michael J. Tarr, and Elissa M. Aminoff. "BOLD5000, a public fMRI dataset while viewing 5000 visual images." *Scientific data* 6, no. 1 (2019): 1-18.

Algonauts





Training and Testing Material.

a) **There are two sets of training data, each consisting of an image set and brain activity in RDM format (for fMRI and MEG). Training set 1 has 92 silhouette object images, and training set 2 has 118 object images with natural backgrounds.**

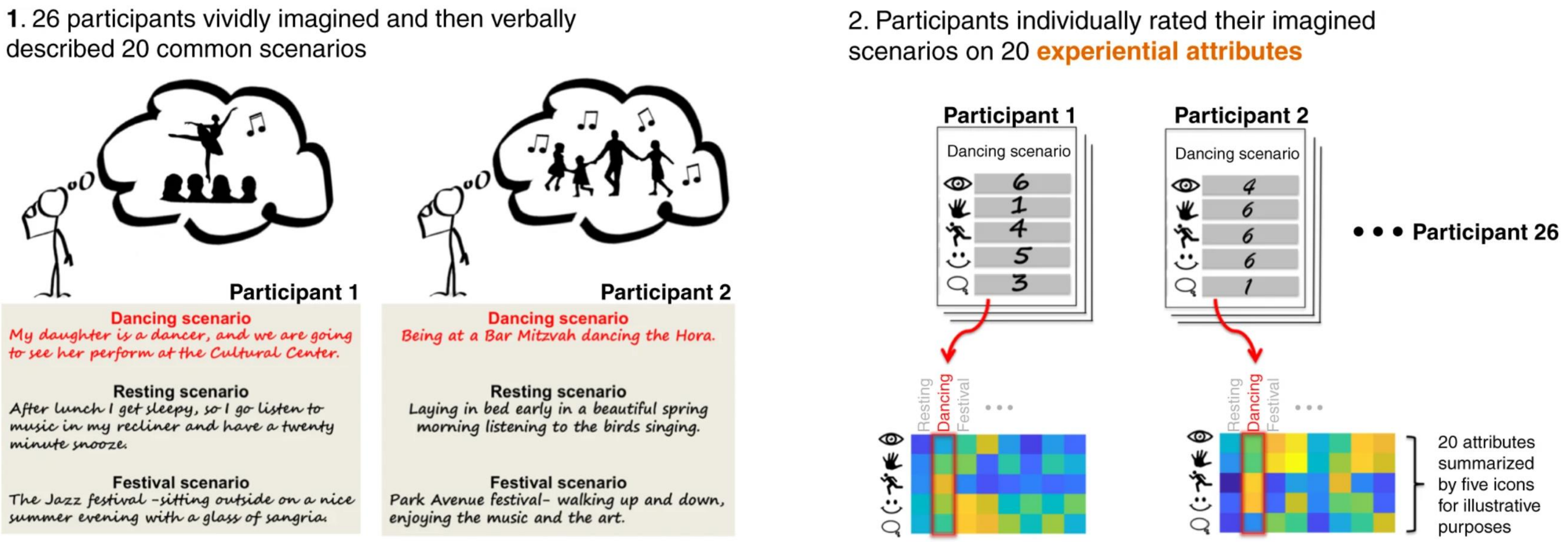
b) **Testing data consists of 78 images of objects on natural backgrounds.**

Cichy, Radoslaw Martin, Gemma Roig, Alex Andonian, Kshitij Dwivedi, Benjamin Lahner, Alex Lascelles, Yalda Mohsenzadeh, Kandan Ramakrishnan, and Aude Oliva. "The algonauts project: A platform for communication between the sciences of biological and artificial intelligence." *arXiv preprint arXiv:1905.05675* (2019).

Audio Stimulus Datasets

Dataset	Type	Languag e	Stimulus	#S	Paradigm	Size	Task
Handjaras et al., 2016	fMRI	Italian	Verbal, pictorial or auditory presentation of 40 concrete nouns	20	Reading, viewing or listening	40 nouns * 4 times.	Property Generation
Huth et al., 2016	fMRI	English	Eleven 10-minute stories	7	Listening	2 hours of stories from The Moth Radio Hour	Passive Listening
Brennan and Hale, 2019	EEG	English	Chapter one of Alice's Adventures in Wonderland as read by Kristen McQuillan	33	Listening	2,129 words in 84 sentences. The entire experimental session lasted 1–1.5 h (including QA)	8 MCQ Question answering concerning the contents of the story
Anderson et al., 2020	fMRI	English	One of 20 scenario names	26	Listening scenario name	20 scenario prompts displayed 5 times.	Imagine themselves personally experiencing common scenarios
Narratives: Nastase et al., 2021	fMRI	English	27 diverse naturalistic spoken stories	345	Listening	891 functional scans, totaling ~4.6 hours of unique stimuli (~43,000 words)	Passive Listening
Natural Stories: Zhang et al., 2020	fMRI	English	Moth-Radio-Hour naturalistic spoken stories	19	Listening	5 h 33 m (repeated twice). Each story is 6 m 48 s avg or 2492 words.	Passive Listening
The Little Prince: Li et al., 2021	fMRI	English, Chinese, French	Audiobook	112	Listening	English audiobook is 94 minutes long. Chinese: 99min. French: 97 min.	Passive Listening. 4 quiz questions.
MEG-MASC: Gwilliams et al., 2022	MEG	English	4 English fictional stories: Cable spool boy, LW1, Black willow, Easy money.	27	Listening	Two hours of naturalistic stories. 208 MEG sensors.	Passive Listening

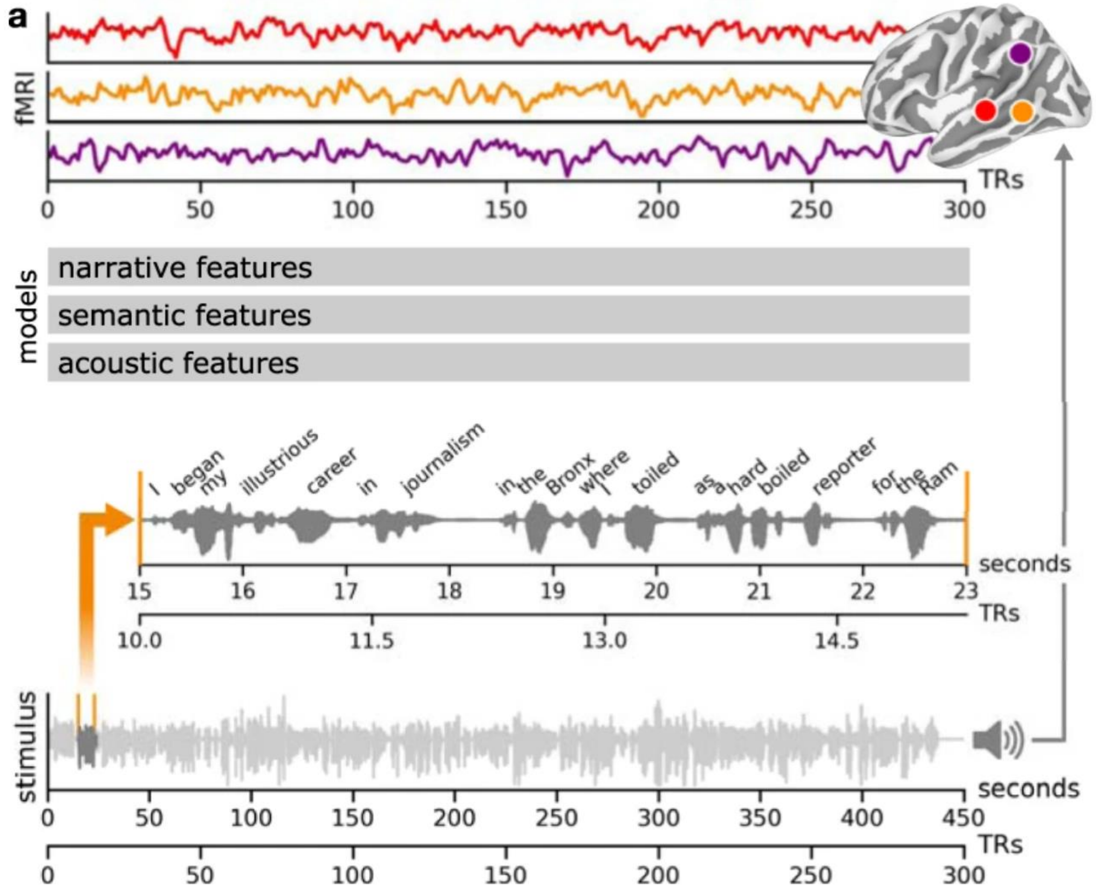
Imagining common scenarios



- Participants underwent fMRI as they reimagined the scenarios when prompted by standardized cues.
- 20 Scenarios: resting, reading, writing, bathing, cooking, housework, exercising, internet, telephoning, driving, shopping, movie, museum, restaurant, barbecue, party, dancing, wedding, funeral, festival.
- 20 attributes: bright, color, motion, touch, audition, music, speech, taste, head, upperlimb, lowerlimb, body, path, landmark, time, social, communication, cognition, pleasant, unpleasant.

Anderson, Andrew James, Kelsey McDermott, Brian Rooks, Kathi L. Heffner, David Dodell-Feder, and Feng V. Lin. "Decoding individual identity from brain activity elicited in imagining common experiences." *Nature communications* 11, no. 1 (2020): 1-14.

Narratives



Story	Duration	TRs	Words	Subjects
"Pie Man"	07:02	282	957	82
"Tunnel Under the World"	25:34	1,023	3,435	23
"Lucy"	09:02	362	1,607	16
"Pretty Mouth and Green My Eyes"	11:16	451	1,970	40
"Milky Way"	06:44	270	1,058	53
"Slumlord"	15:03	602	2,715	18
"Reach for the Stars One Small Step at a Time"	13:45	550	2,629	18
"It's Not the Fall That Gets You"	09:07	365	1,601	56
"Merlin"	14:46	591	2,245	36
"Sherlock"	17:32	702	2,681	36
"Schema"	23:12	928	3,788	31
"Shapes"	06:45	270	910	59
"The 21st Year"	55:38	2,226	8,267	25
"Pie Man (PNI)"	06:40	267	992	40
"Running from the Bronx (PNI)"	08:56	358	1,379	40
"I Knew You Were Black"	13:20	534	1,544	40
"The Man Who Forgot Ray Bradbury"	13:57	558	2,135	40
Total:	4.6 hours	11,149 TRs	42,989 words	
Total across subjects:	6.4 days	369,496 TRs	1,399,655 words	

Nastase, Samuel A., Yun-Fei Liu, Hanna Hillman, Asieh Zadbood, Liat Hasenfratz, Neggin Keshavarzian, Janice Chen et al. "The "Narratives" fMRI dataset for evaluating models of naturalistic language comprehension." *Scientific data* 8, no. 1 (2021): 1-22.

Video Stimulus Datasets

Dataset	Type	Language	Stimulus	#Subjects	Paradigm	Size	Task
BBC's Doctor Who: Seeliger et al., 2019	fMRI	English	Spatiotemporal visual and auditory naturalistic stimuli (30 episodes of BBC's Doctor Who)	1	Viewing episode videos	120.830 whole-brain volumes (approx. 23 h) of single-presentation data, and 1.178 volumes (11 min) of repeated narrative short episodes (22 repetitions)	Passive viewing
Japanese Ads: Nishida et al., 2020	fMRI	Japanese	368 web and 2452 TV Japanese ad movies (15-30s)	40 and 28 for web and TV ads. 16 were overlapped	Viewing Ads	7200 train and 1200 test fMRIs for web; fMRIs from 420 ads.	Passive viewing
Pippi Langkous: Berezutskaya et al., 2020	ECoG	The movie was originally in Swedish but dubbed in Dutch	30 s excerpts of a feature film (in total, 6.5 min long), edited together for a coherent story	37 patients	Viewing	6.5 min movie.	Passive viewing
Algonauts: Cichy et al., 2021	fMRI	English	1000 short video clips	10	Viewing video clips	1000 short video clips (3 sec each)	Passive viewing
Natural Short Clips: Huth et al., 2022	fMRI	English	Natural short movie clips	5	Watching natural short movie clips	3870 responses per subject.	Passive viewing

Japanese Ads

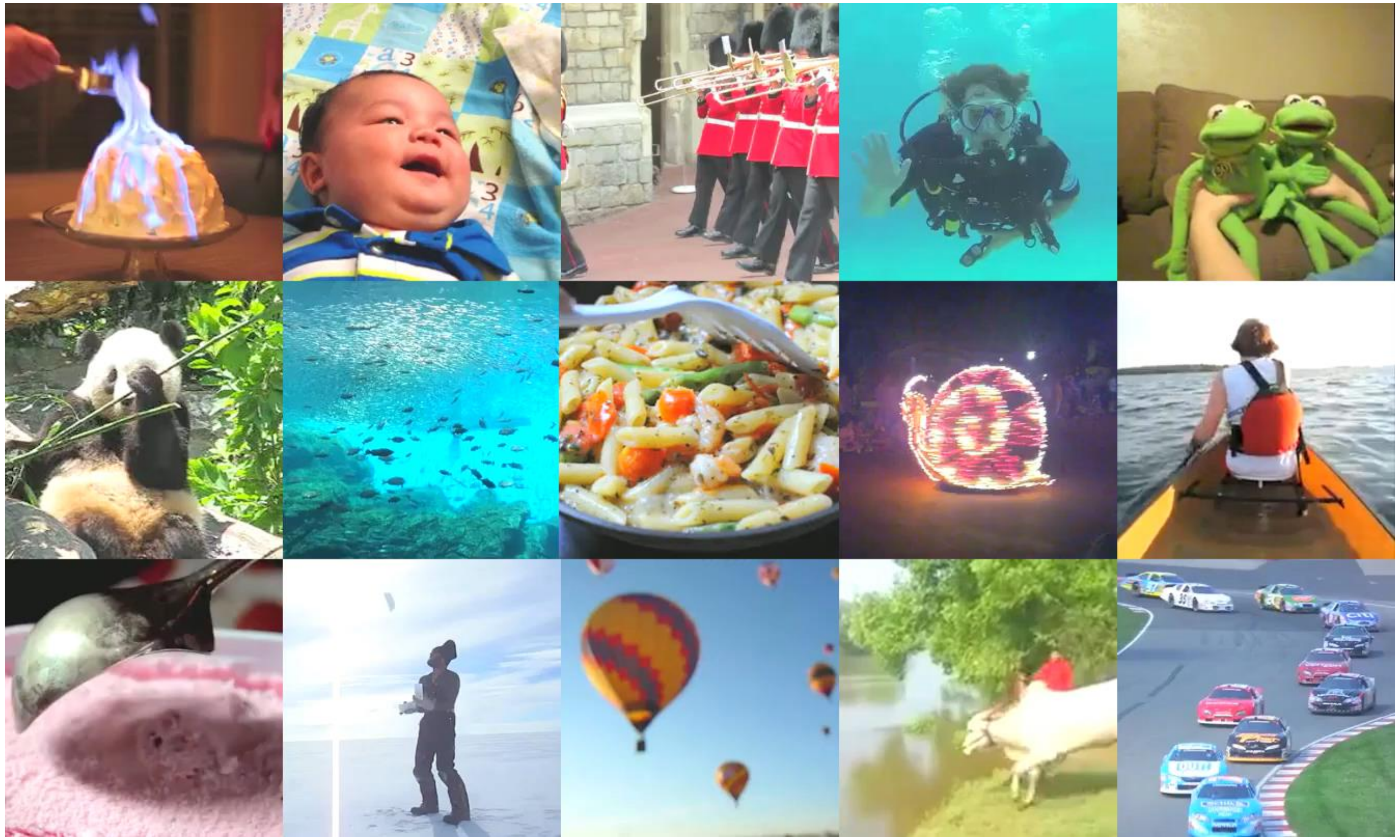
- Two sets of movies were provided by NTT DATA Corp: web and TV ads.
- Four types of cognitive labels associated with the movie datasets
  - Scene descriptions
    - Human judges create scene descriptions with 50+ words per 1s scene.
  - Impression ratings
    - Human rating on 30 factors for every 2s clip on a scale of 0-4.
  - Ad effectiveness indices
    - Click rate: fraction of viewers who clicked the frame of a movie and jumped to a linked web page
    - View completion rate: fraction of viewers who continued to watch an ad movie until the end without choosing a skip option.
  - Ad preference votes
    - Each tester was asked to freely recall a small number of favorite TV ads from among the ads recently broadcasted.
    - The total number of recalls of an ad was regarded as its preference value.

Categories	Web ad movies	TV ad movies
Electronic & Precision	4	50
Audiovisual	5	6
Appliance	16	23
Car	31	145
Food & Confectionery	7	369
Beverage & Alcoholic drink	20	236
Medical & Health	35	156
Cosmetics	49	85
Sundries & Home equipment	10	254
Garment/apparel	9	43
Entertainment	42	237
Media & Education	41	82
Distribution & Retailer	12	112
Communication & Service	35	328
House & Construction	9	90
Finance	9	145
Enterprise, Public service, & Others	34	91

Nishida, Satoshi, Yusuke Nakano, Antoine Blanc, Naoya Maeda, Masataka Kado, and Shinji Nishimoto. "Brain-mediated transfer learning of convolutional neural networks." In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, pp. 5281-5288. 2020.

Algonauts 2021

- fMRI from 10 human subjects that watched over 1,000 short (3 sec) video clips.



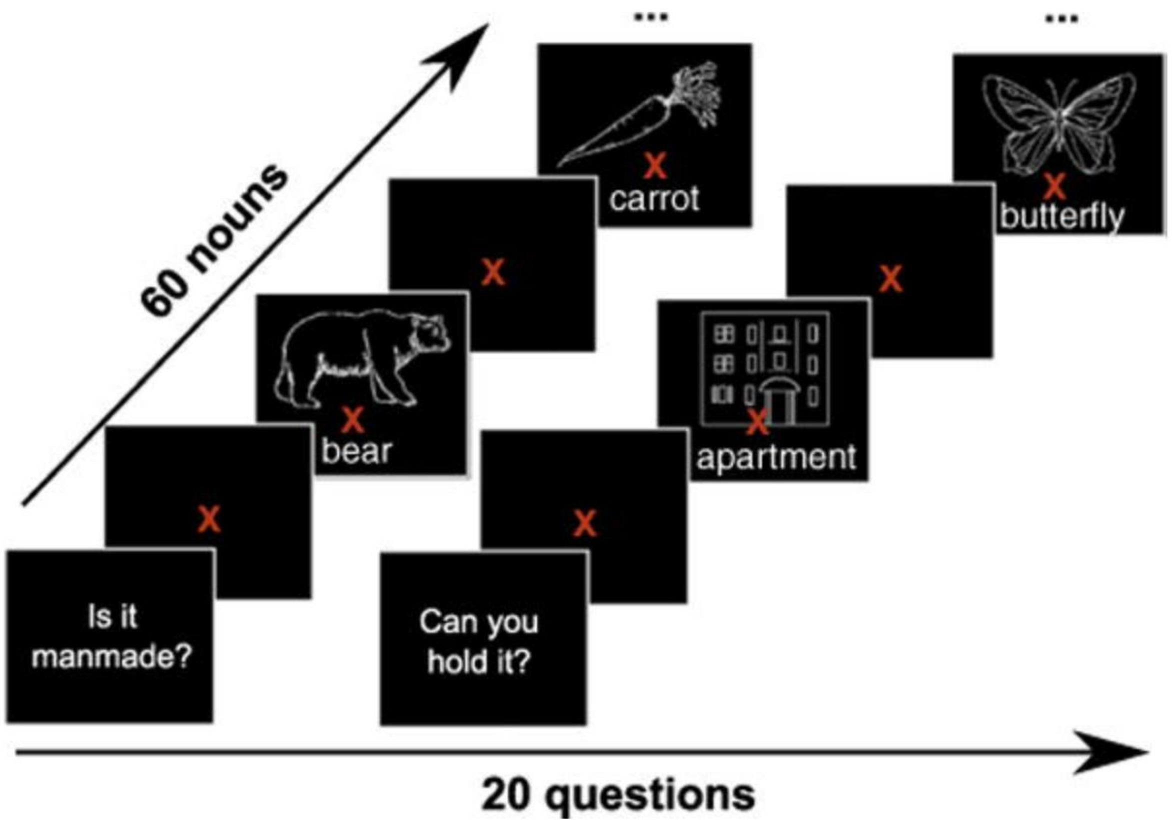
Cichy, Radoslaw Martin, Kshitij Dwivedi, Benjamin Lahner, Alex Lascelles, Polina Iamshchinina, M. Graumann, A. Andonian et al. "The Algonauts Project 2021 Challenge: How the Human Brain Makes Sense of a World in Motion." *arXiv preprint arXiv:2104.13714* (2021).

Other Multimodal Stimulus Datasets

Dataset	Type	Language	Stimulus	#Subjects	Paradigm	Size	Task
Mitchell et al., 2008	fMRI	English	60 different word-picture pairs from 12 categories.	9	Viewing word-picture pairs	60 different word-picture pairs presented six times each	Passive viewing
Sudre et al., 2012	MEG	English	60 concrete nouns along with line drawings	9	Reading	60 stimuli × 20 questions = 1200 examples	Question answering
Zinszer et al., 2017	fNIRS	English	8 concrete nouns (audiovisual word and picture stimuli): bunny, bear, kitty, dog, mouth, foot, hand, and nose	24	Viewing and listening	12 blocks with the 8 stimuli per subject.	Passive viewing and listening
Pereira et al., 2018	fMRI	English	180 Words with Picture, Sentences, word clouds; 96 text passages; 72 passages	16	Viewing WP, sentences or word clouds	180 WP, S and WC per subject; 96+72 passages shown 3 times	Passive viewing
Cao et al., 2021	fNIRS	Chinese	50 concrete nouns from 10 semantic categories	7	Viewing and listening	Each stimulus is presented 7 times.	Passive viewing and listening
Courtois Neuromod	fMRI	full-length movies and TV show	6	Viewing and Listening	~100 hours of data per participant	Passive viewing	

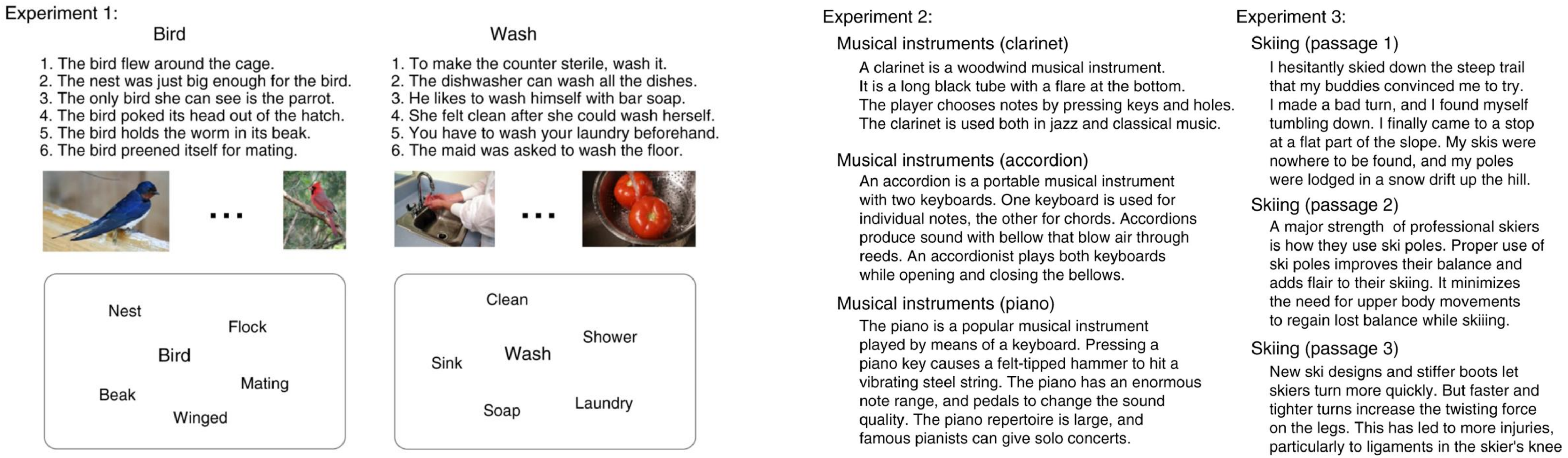
Concrete nouns with line drawings

- Subjects were asked to perform a QA task, while their brain activity was recorded using MEG.
- Subjects were first presented with a question (e.g., “Is it manmade?”), followed by 60 concrete nouns, along with their line drawings, in a random order.
- Each stimulus was presented until the subject pressed a button to respond “yes” or “no” to the initial question.
- Once all 60 stimuli are presented, a new question is shown for a total of 20 questions.



Sudre, Gustavo, Dean Pomerleau, Mark Palatucci, Leila Wehbe, Alona Fyshe, Riitta Salmelin, and Tom Mitchell. "Tracking neural coding of perceptual and semantic features of concrete nouns." NeuroImage 62, no. 1 (2012): 451-463.

Word+Picture, Sentences, Word Clouds, Passages

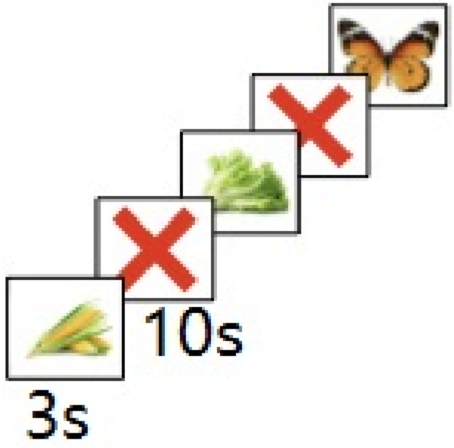


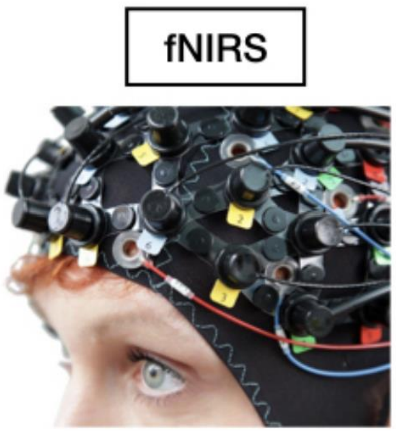
- Experiment 1: 180 words (128 nouns, 22 verbs, 29 adjectives and adverbs, and 1 function word). 3 paradigms.
- Experiment 2: 96 text passages, each with 4 sentences from 24 broad topics (e.g., professions, clothing, birds, musical instruments, natural disasters, crimes, etc.)
- Experiment 3: 72 passages, each with 3-4 sentences from another 24 topics.

Pereira, Francisco, Bin Lou, Brianna Pritchett, Samuel Ritter, Samuel J. Gershman, Nancy Kanwisher, Matthew Botvinick, and Evelina Fedorenko. "Toward a universal decoder of linguistic meaning from brain activation." *Nature communications* 9, no. 1 (2018): 1-13.

fNIRS with audio-visual stimuli

- Stimuli are pictures and audios of 50 objects from 10 categories.
- Visual presentation lasts for 3s, with audio presented immediately at the onset, followed by a 10s rest period.
- During rest period, participants are instructed to fixate on an X displayed in the center of the screen.





Category	Exemplar
tool	pliers, saw, screwdriver, scissor, hammer
vegetable	celery, corn, carrot, tomato, lettuce
building	bird's nest, tiananmen, oriental pearl TV tower, pyramid, water cube
insect	bee, butterfly, dragonfly, ant, fly
transportation	car, train, truck, airplane, bicycle
furniture	sofa, chair, desk, bed, bookshelf
cloth	sweater, jeans, shirt, skirt, dress
animal	panda, cat, dog, horse, cow
body-part	arm, eye, foot, palm, leg
kitchen	knife, pan, spoon, glass, chopsticks

Cao, Lu, Dandan Huang, Yue Zhang, Xiaowei Jiang, and Yanan Chen. "Brain decoding using fnirs." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 14, pp. 12602-12611. 2021.

Agenda

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

Agenda

- Introduction to Brain encoding and decoding [30 min]
- Stimulus Representations [1 hour]
  - Text Stimulus Representations
  - Visual Stimulus Representations
  - Audio Stimulus Representations
  - Multimodal Stimulus Representations
- Coffee break [30 min]
- Deep Learning for Brain Decoding [1 hour 30 min]
- Lunch break [1 hour 30 min]
- Deep Learning for Brain Encoding [1 hour 30 min]
- Coffee break [30 min]
- Advanced Methods [1 hour 15 min]
- Summary and Future Trends [15 min]

Stimulus Representations

- Text Stimuli
  - Basic NLP Representations: Corpus co-occurrence counts, topic models, Linguistic (POS, dependencies, roles)
  - Discourse features.
  - Semantic: word embedding methods, sentence representation models, recurrent neural networks and Transformer methods.
  - Experiential attributes: Rated on 0-6 scale or binary.
- Visual Stimuli
  - Visual field filter banks
  - Gabor wavelet pyramid
  - HMAX model
  - Convolutional neural networks
- Audio Stimuli
  - Phoneme rate and presence of phonemes.
- Multimodal Stimuli

Agenda

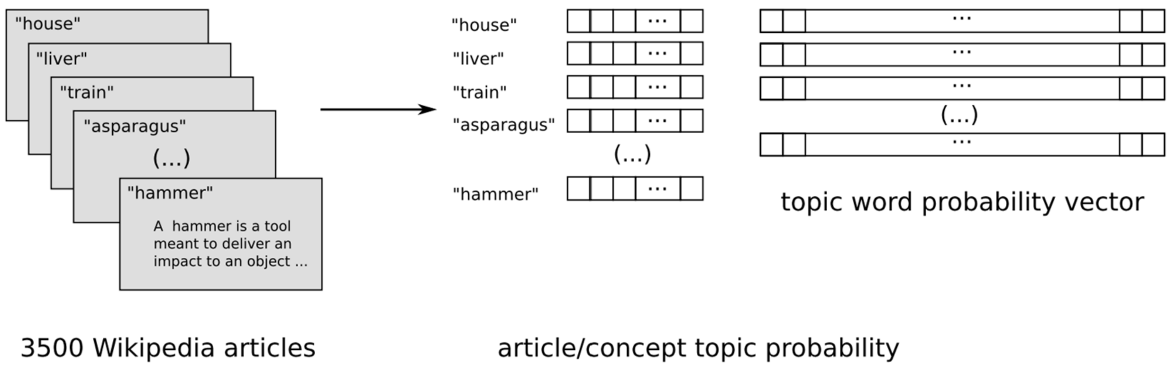
- Introduction to Brain encoding and decoding [30 min]
- Stimulus Representations [1 hour]
    - **Text Stimulus Representations**
    - Visual Stimulus Representations
    - Audio Stimulus Representations
    - Multimodal Stimulus Representations
- Coffee break [30 min]
- Deep Learning for Brain Decoding [1 hour 30 min]
- Lunch break [1 hour 30 min]
- Deep Learning for Brain Encoding [1 hour 30 min]
- Coffee break [30 min]
- Advanced Methods [1 hour 15 min]
- Summary and Future Trends [15 min]

Text Stimulus Representations

- Basic NLP Representations
  - Corpus co-occurrence counts
  - Topic models
  - Linguistic: POS, dependencies, roles.
- Discourse
  - Characters, motion, speech, emotions, non-motion verbs
- Deep Learning based Representations
  - Embeddings
  - Longer context using LSTMs
  - Transformers
- Experiential attributes
  - Rated on 0-6 scale
  - Binary

Basic NLP Representations for Word Stimuli

- Corpus co-occurrence counts
  - 25 verbs (Mitchell et al., 2008; Pereira et al., 2013)
    - Verbs: see, hear, listen, taste, smell, eat, touch, nib, lift, manipulate, run, push, fill, move, ride, say, fear, open, approach, near, enter, drive, wear, break, and clean.
    - These verbs generally correspond to basic sensory and motor activities, actions per formed on objects, and actions involving changes to spatial relationships.
    - For each (verb, stimulus word w), feature value = normalized co-occurrence count of w with any of three forms of the verb (e.g., taste, tastes, or tasted) over the text corpus.
  - 985 common English words (such as above, worry, and mother) in (Huth et al., 2016).
- Topic models (Pereira et al., 2013)
  - Get relevant Wiki pages (e.g., “airplane” is “Fixed-Wing Aircraft”) and other linked pages (e.g. “Aircraft cabin”)
  - LDA topic modelling on 3500 pages with #topics from 10 to 100, in increments of 5, setting the α parameter to 25/#topics.
  - LSA topic modelling (Wang et al., 2017)



Basic NLP Representations for Word Stimuli

- Word length
- Is the word related to one of the 28 unique parts of speech and 17 unique dependency relationships?
- Position of word in the sentence
- Roles
  - Main verb
  - Agent or experiencer
  - Patient or recipient
  - Predicate of a sentence (The window was dusty)
  - Modifier (The angry activist broke the chair)
  - Complement in adjunct and propositional phrase, including direction, location, and time (The restaurant was loud at night).

Wehbe, Leila, Brian Murphy, Partha Talukdar, Alona Fyshe, Aaditya Ramdas, and Tom Mitchell. "Simultaneously uncovering the patterns of brain regions involved in different story reading subprocesses." *PloS one* 9, no. 11 (2014): e112575.

Wang, Jing, Vladimir L. Cherkassky, and Marcel Adam Just. "Predicting the brain activation pattern associated with the propositional content of a sentence: modeling neural representations of events and states." *Human brain mapping* 38, no. 10 (2017): 4865-4881.

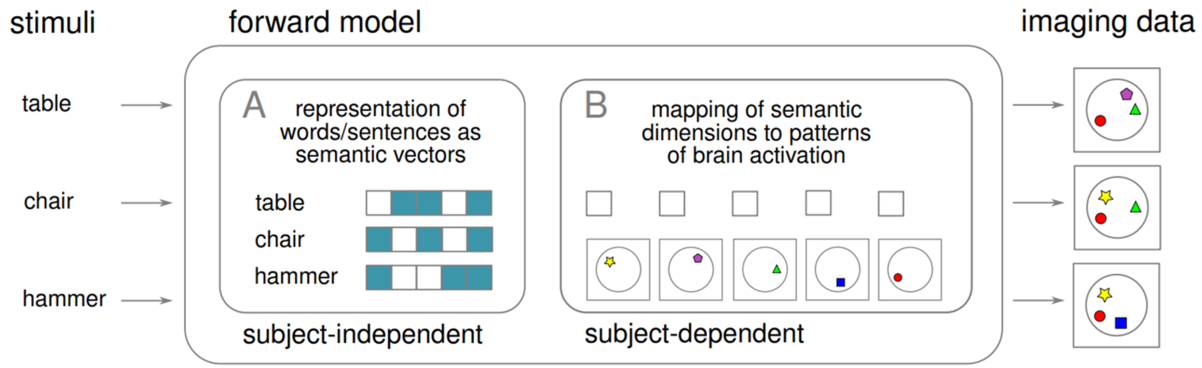
Discourse features (for Harry Potter dataset)

- Characters: Resolve all pronouns to the character to whom they refer, and make binary features to signal which of the 10 characters are mentioned.
- Motions: Identify a set of motions that occurred frequently in the chapter (e.g. fly, manipulate, collide physically, etc.).
- Speech: Indicate the parts of the story that correspond to direct speech between the characters. Used the presence of dialog as a feature.
- Emotions: Identified a set of emotions that were felt by the characters in the chapter (e.g. annoyance, nervousness, pride, etc.).
- Verbs: Identified a set of actions that occurred frequently in the chapter that were distinct from motion (e.g. hear, know, see, etc.).

Wehbe, Leila, Brian Murphy, Partha Talukdar, Alona Fyshe, Aaditya Ramdas, and Tom Mitchell. "Simultaneously uncovering the patterns of brain regions involved in different story reading subprocesses." *PloS one* 9, no. 11 (2014): e112575.

Wang, Jing, Vladimir L. Cherkassky, and Marcel Adam Just. "Predicting the brain activation pattern associated with the propositional content of a sentence: modeling neural representations of events and states." *Human brain mapping* 38, no. 10 (2017): 4865-4881.

DL Representations: Using embeddings for word stimuli

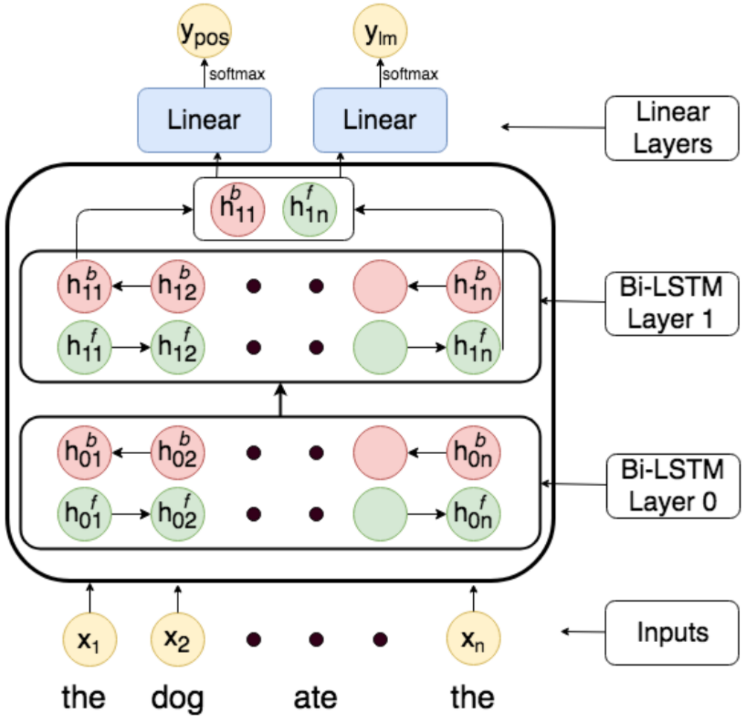


	Noun	Verb	Adjective
GloVe	**0.8768**(0.0792)	0.8544(0.0713)	0.8337(0.1081)
Word2Vec	**0.8386**(0.0942)	0.8309(0.0636)	0.8210(0.1028)
Fasttext	**0.8407**(0.0676)	0.8235(0.0766)	0.8077(0.0996)
RWSGwn	**0.8123**(0.0886)	0.7453(0.0771)	0.7425(0.1032)
ELMo	**0.9088**(0.0632)	0.8520(0.0797)	0.7993(0.1244)
ConceptNet	0.8646(0.0875)	**0.8702**(0.0695)	0.8249(0.0925)
Dependency	**0.8554**(0.0731)	0.8137(0.0755)	0.7891(0.0808)

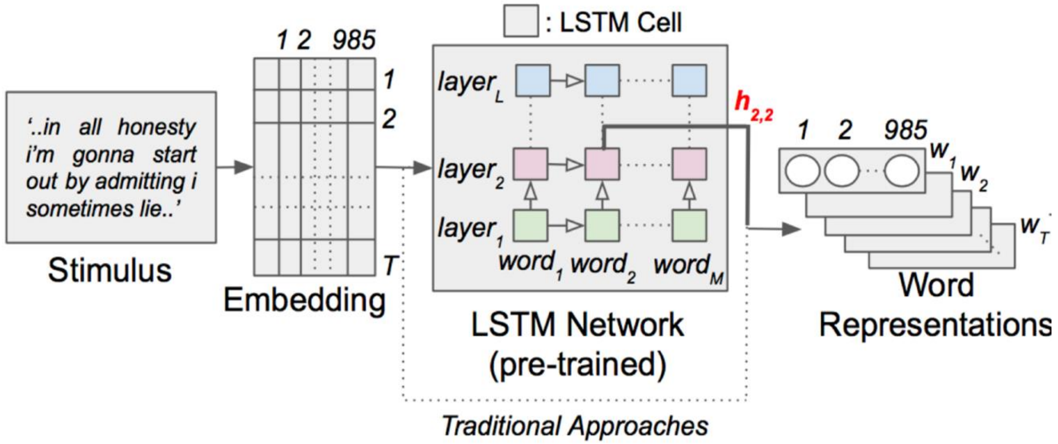
- GloVe 300D vectors (Pereira et al., 2016; Wang et al., 2017; Pereira et al., 2018; Anderson et al., 2019)
- 1000D Non-negative sparse embeddings (Wehbe et al., 2014).
- 300D embeddings by training a skip-gram model using negative sampling (SGNS) on Italian and English Wikipedia dumps using Gensim. (Anderson et al., 2017a)
- FastText (Berezutskaya et al., 2020)
- Comparison across multiple embedding methods
  - GloVe, word2vec, WordNet2Vec, FastText, ELMo (Hollenstein et al., 2019)
  - word2Vec, fastText, GloVe, Dependency-based word2vec, RWSGwn, ConceptNet, ELMo, averaged and concatenated combinations (Wang et al., 2020)

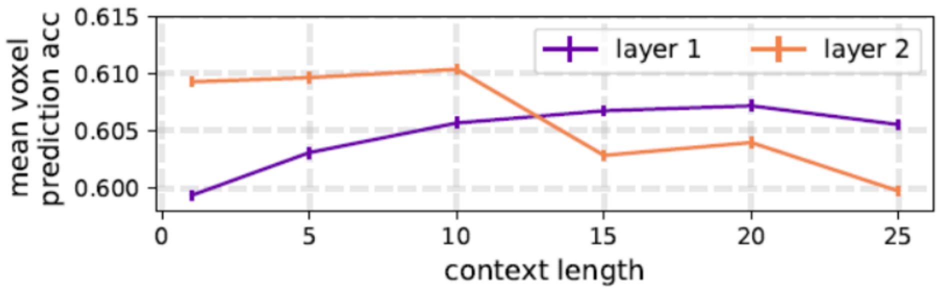
DL Representations: Using longer context for word stimuli

- Multi-task LSTMs
  - Predict next word and POS of next word.



- ELMo embeddings: LSTM based pretrained language model





(a) ELMo

Toneva, Mariya, and Leila Wehbe. "Interpreting and improving natural-language processing (in machines) with natural language-processing (in the brain)." *Advances in Neural Information Processing Systems* 32 (2019).

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

Jat, Sharmistha, Hao Tang, Partha Talukdar, and Tom Mitchell. "Relating simple sentence representations in deep neural networks and the brain." *arXiv preprint arXiv:1906.11861* (2019).

DL Representations: Using sentence embeddings

- Unstructured Models: Ignore sentence structure
  - Simple Pooling Methods
    - Average/max/concat(max, avg) pooling over word embeddings.
  - Advanced Pooling Methods
    - FastSent (Hill, Cho, and Korhonen 2016) sums word embeddings in a sentence as its representation to predict the surrounding sentences.
    - SIF (Arora, Liang, and Ma 2016) adapts the naïve averaging of word embeddings to weighted averaging.
- Structured Models
  - Unsupervised Methods: Skip-thought, QuickThought.
  - Supervised Methods: InferSent, GenSen (Subramanian et al. 2018), Universal Sentence Encoder

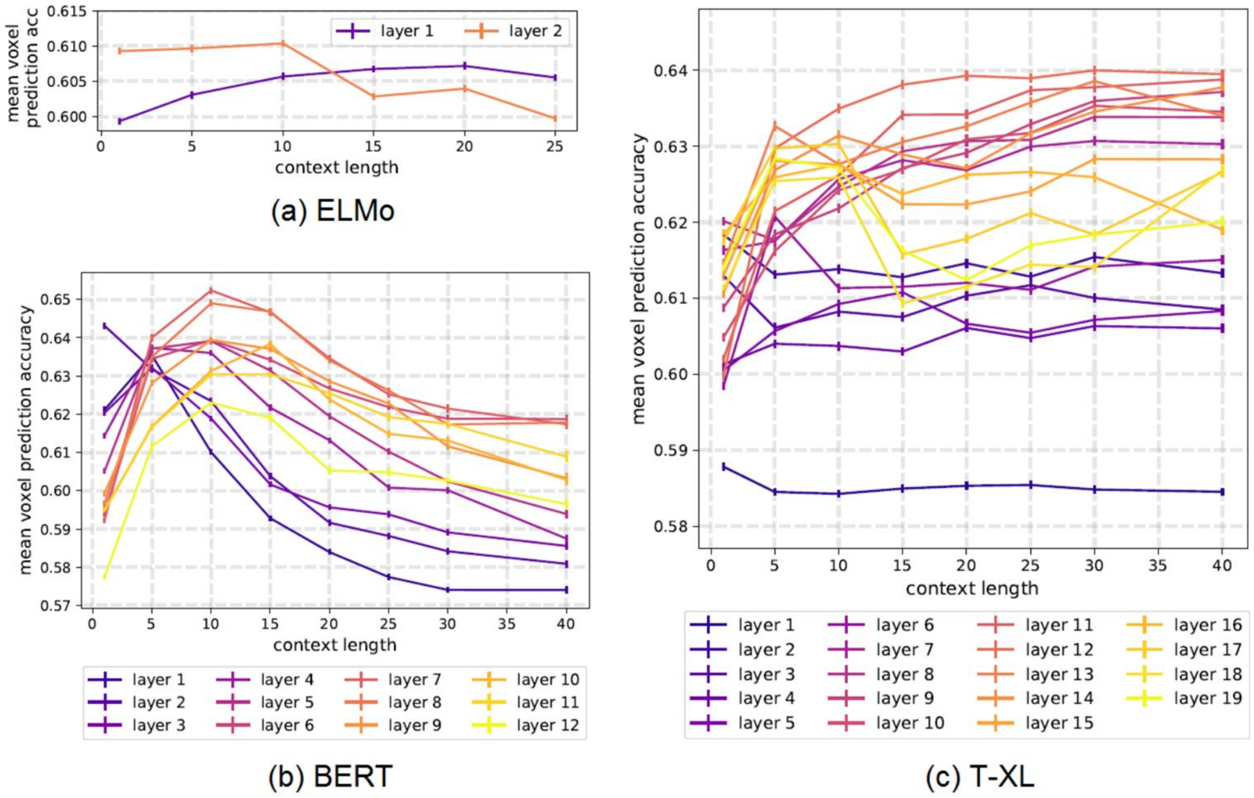
Topic	Passage	Sentence [b]
Musical Instruments	Piano	1. The piano is a popular musical instrument… 2. Pressing a piano key causes a felt-tipped hammer... 3. The piano has an enormous note range.
	Accordion	1. A clarinet is a woodwind musical instrument… 2. It is a long black tube with a flare at the bottom 3. The player chooses notes by pressing keys and holes.
	Clarinet	1. An accordion is a portable musical instrument. 2. One keyboard is used for individual notes 3. Accordions produce sound with bellow that blow air
…	…	…

[a]	Ridge			Lasso			MLP		
	topic	passa.	sente.	topic	passa.	sente.	topic	passa.	sente.
Max	0.88	0.76	0.65	0.88	0.75	0.70	0.83	0.70	0.63
Avg	0.90	0.83	0.73	0.92	0.81	0.78	0.89	0.78	0.67
Cat	0.92	0.83	0.74	0.90	0.81	0.80	0.86	0.74	0.66
Sif	0.89	0.84	0.69	0.91	0.77	0.72	0.84	0.73	0.65
Fast	0.92	0.81	0.74	0.90	0.79	0.77	0.88	0.76	0.67
Skip	0.90	0.82	0.75	0.91	0.80	0.79	0.86	0.81	0.73
Quik	0.91	0.84	0.75	0.91	0.81	0.79	0.90	0.82	0.77
Gen	0.91	0.84	0.78	0.92	0.84	**0.84**	0.91	**0.84**	**0.80**
Inf	**0.94**	**0.90**	**0.83**	**0.93**	**0.86**	**0.84**	**0.92**	**0.84**	0.79

Toneva, Mariya, and Leila Wehbe. "Interpreting and improving natural-language processing (in machines) with natural language-processing (in the brain)." *Advances in Neural Information Processing Systems* 32 (2019).

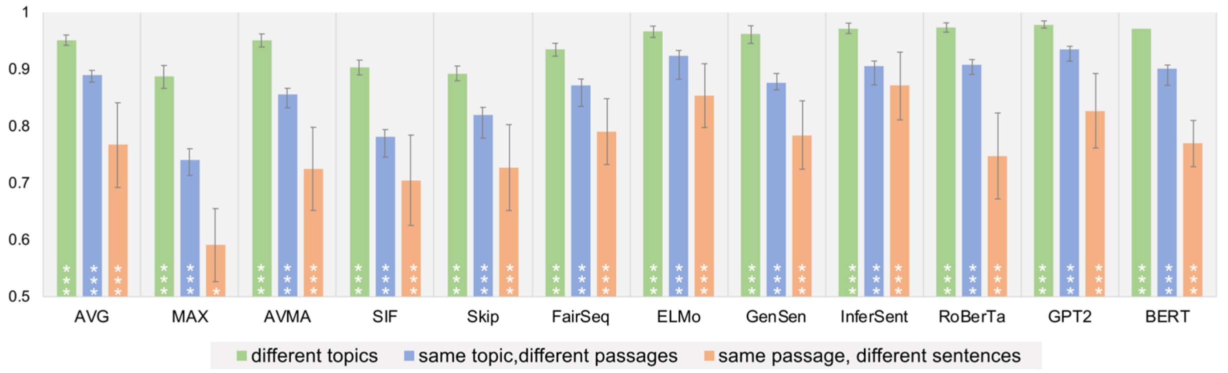
Sun, Jingyuan, Shaonan Wang, Jiajun Zhang, and Chengqing Zong. "Towards sentence-level brain decoding with distributed representations." In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 7047-7054. 2019.

DL Representations: Transformer-based methods for text stimuli (Layer #, context length, architecture)



DSM	Name	Structure and Training Task
Unstructured	AVG	Average Pooling
	MAX	Max Pooling
	AVMA	Concatenation of AVG and Max
	SIF	Weighted Average Pooling
Structured	FairSeq	CNN (language model)
	Skip	LSTM (language model)
	GenSen	BiLSTM (multi-task learning)
	InferSent	CNN-BiLSTM (natural language inference)
	ELMo	CNN-BiLSTM (language model)
	BERT	Transformer (language model)
	RoBerTa	
	GPT2	

Transformer-XL is the only model that continues to increase performance as the context length is increased. In all networks, the middle layers perform the best for contexts longer than 15 words. The deepest layers across all networks show a sharp increase in performance at short-range context (fewer than 10 words), followed by a decrease in performance. [Toneva and Wehbe, 2019]



Toneva, Mariya, and Leila Wehbe. "Interpreting and improving natural-language processing (in machines) with natural language-processing (in the brain)." *Advances in Neural Information Processing Systems* 32 (2019).

Sun, Jingyuan, Shaonan Wang, Jiajun Zhang, and Chengqing Zong. "Neural encoding and decoding with distributed sentence representations." *IEEE Transactions on Neural Networks and Learning Systems* 32, no. 2 (2020): 589-603.

DL Representations: Transformer-based methods for text stimuli (NLP task finetuning and scrambled LM)

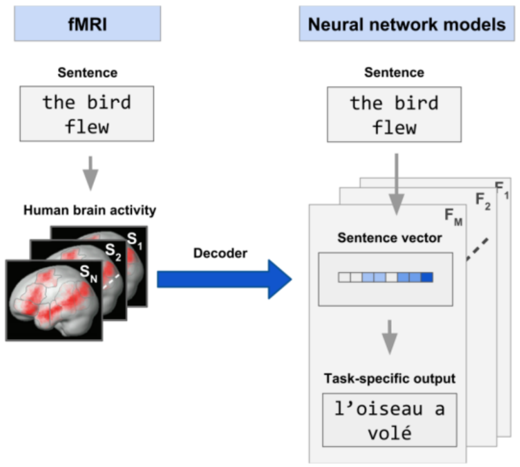
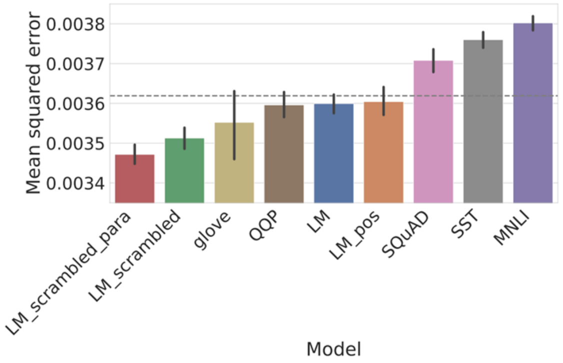


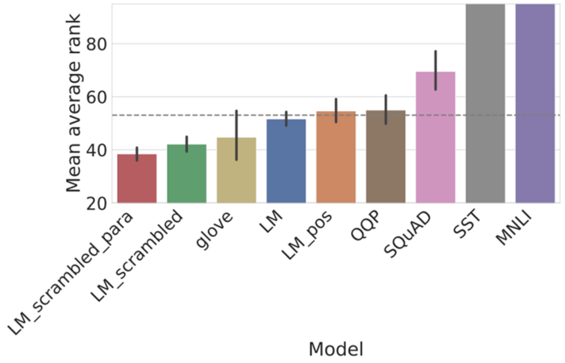
Figure 1: Brain decoding methodology. We use human brain activations in response to sentences to predict how neural networks represent those same sentences.

Task	Dataset	Domain	# train	Avg sent len.	# types
Paraphrase classification	Quora Question Pairs	social QA	364k	22.1	103k
Question answering	SQuAD 2.0 (Rajpurkar et al., 2018)	wiki	130k	11.2	43.9k
Natural language inference	MNLI (Williams et al., 2017)	mixed	393k	16.8	83.3k
Sentiment analysis	SST-2 (Socher et al., 2013)	movie reviews	67.3k	9.41	14.8k

Table 2: Details of the tasks used for fine-tuning.



(a) Mean squared error metric



(b) Average rank metric

- Scrambled LM
  - Randomly shuffle words from the corpus samples, to remove all first order cues to syntactic structure.
  - LM-scrambled: words are shuffled within sentences
  - LM-scrambled-para: words are shuffled within their containing paragraphs in the corpus.
- LM_pos: predict only the part of speech of a masked word, rather than the word itself.
- **Scrambled LMs work best!**

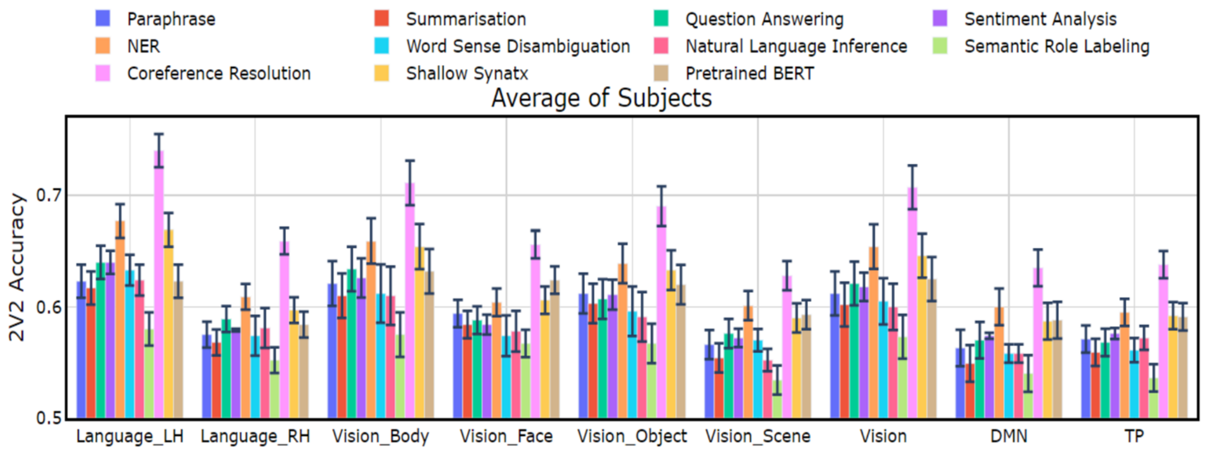
Gauthier, Jon, and Roger Levy. "Linking artificial and human neural representations of language." *arXiv preprint arXiv:1910.01244* (2019).

DL Representations: Transformer-based methods for text stimuli (NLP task finetuning)

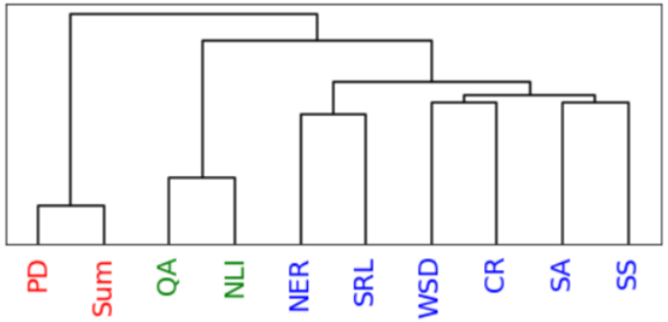
Task	HuggingFace Model Name	Dataset
NLI	bert-base-nli-mean-tokens	Stanford Natural Language Inference (SNLI), MultiNLI
PD	bert-base-cased-finetuned-mrpc	Microsoft Research Paraphrase Corpus (MRPC)
SS	bert-base-chunl	CoNLL-2003
Sum	bart-base-samsum	SAMSum
WSD	bert-base-baseline	English all-words
CR	bert_coreference_base	OntoNotes and GAP
NER	bert-base-NER	CoNLL-2003
QA	bert-base-qa	SQUAD
SA	bert-base-sst	Stanford Sentiment Treebank (SST)
SRL	bert-base-srl	English PropBank SRL

Tasks

Paraphrase, Summarization, Question Answering, Sentiment Analysis, NER, Word Sense Disambiguation, Natural Language Inference, Semantic Role Labeling, Coreference Resolution, Shallow Syntax Parsing



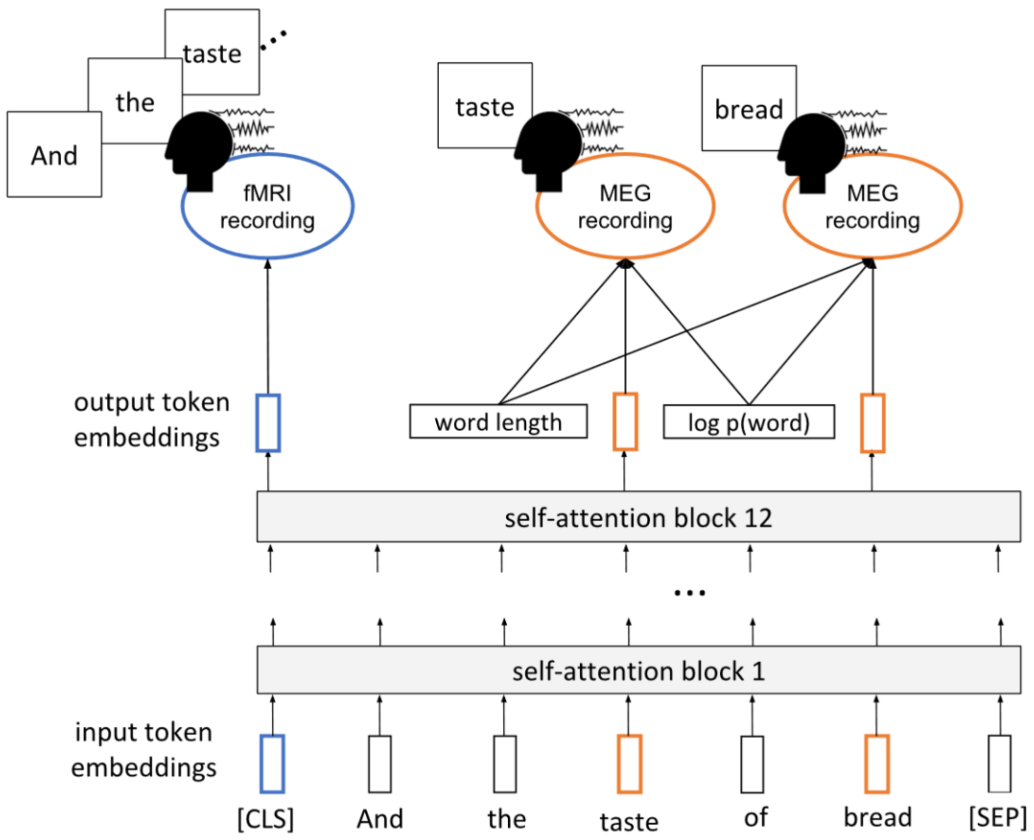
Pereira dataset: CR, NER, and SS perform the best.



Dendrogram constructed using similarity on representations from task-specific Transformer encoder models with stimuli from the dataset passed as input.

Oota, Subba Reddy, Jashn Arora, Veeral Agarwal, Mounika Marreddy, Manish Gupta, and Bapi Raju Surampudi. "Neural Language Taskonomy: Which NLP Tasks are the most Predictive of fMRI Brain Activity?." *arXiv preprint arXiv:2205.01404* (2022).

DL Representations: Transformer-based methods for text stimuli (Multi-task setup)

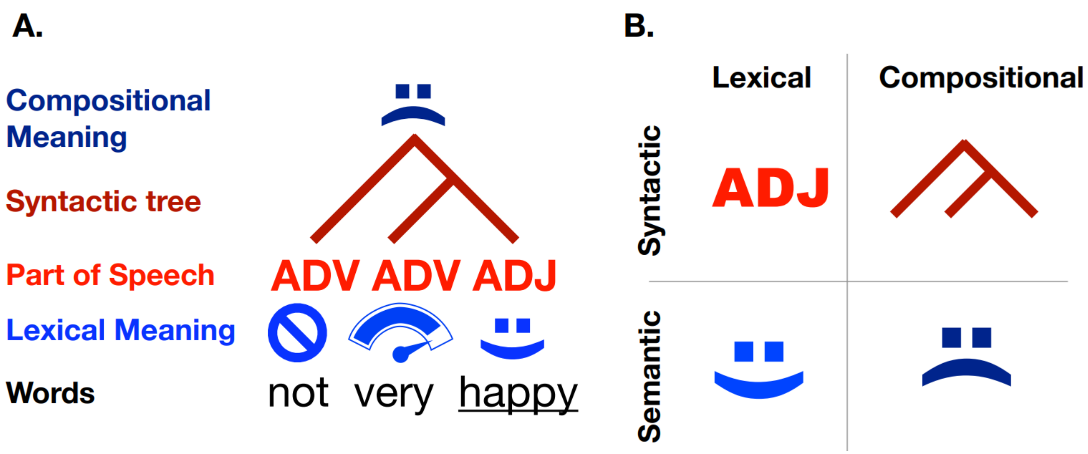


- Settings
  - Finetune BERT vs not
  - Finetune BERT using one representative subject and train dense layer for each subject, vs finetune BERT for each subject.
  - Finetune BERT on MEG for all subjects, then finetune BERT on fMRI.
  - Multi-task finetune BERT for fMRI+MEG prediction task
- Results
  - Fine-tuned models predict fMRI data better than vanilla BERT
  - Relationships between text and brain activity generalize across experiment participants.
  - Using MEG data can improve fMRI predictions.
  - A single model can be used to predict fMRI activity across multiple experiment participants.

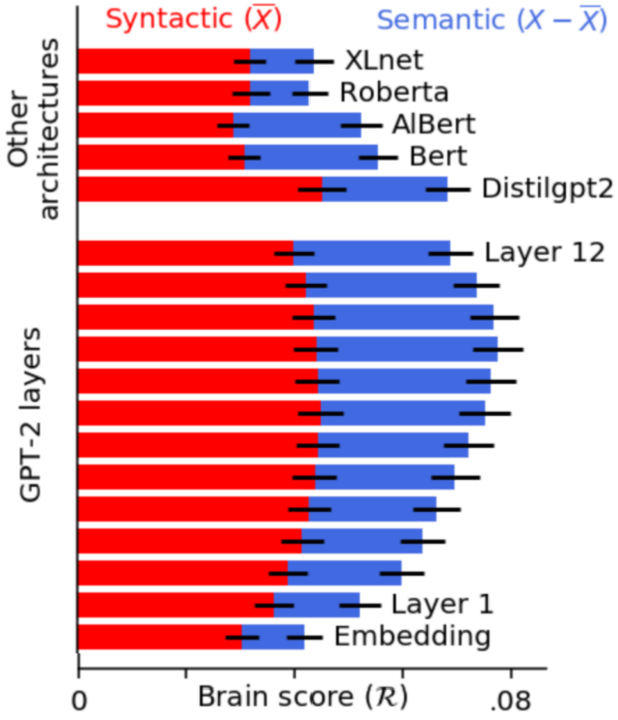
Schwartz, Dan, Mariya Toneva, and Leila Wehbe. "Inducing brain-relevant bias in natural language processing models." *Advances in neural information processing systems* 32 (2019).

DL Representations: Comparing Transformers and extracting syntax vs semantics

- Representations:
  - Lexical: representation that is context-invariant. E.g., word embeddings.
  - Compositional: "contextualized" representation generated by a system combining multiples words. E.g., parse trees
  - Syntax: representation associated with the structure of sentences independently of their meaning
  - Semantics: representation of a language system that are not syntactic.



- If $X^{(l)}$ is activation of l^{th} layer, $\overline{X^{(l)}}$ is average activation across similar syntax inputs
  - Lexical: $X^{(0)}$
  - Compositional: $X^{(l)}; l > 0$
  - Syntax: $\overline{X^{(l)}}$, $l \geq 0$
  - Semantic: $X^{(l)}$- $\overline{X^{(l)}}$



Caucheteux, Charlotte, Alexandre Gramfort, and Jean-Remi King. "Disentangling syntax and semantics in the brain with deep networks." In *International Conference on Machine Learning*, pp. 1336-1348. PMLR, 2021.

Experiential attributes model for text stimuli

- Represents words in terms of human (Amazon Mechanical Turk) ratings of their degree of association with different attributes of experience
    - “On a scale of 0 to 6, to what degree do you think of a banana as having a characteristic or defining color?”
    - Anderson et al., 2019: 65 attributes spanning sensory, motor, affective, spatial, temporal, causal, social, and abstract cognitive experiences.
- Value-add on top of text models: a lot of experiential information goes unstated in natural verbal communication.
    - E.g., it is rarely useful to communicate the color of bananas because it is obvious to all those with experience of bananas.
    - E.g., it would be unusual to specify that dropping things involves movement.
- Nishida et al., 2020 use a subset of 20 attributes.

Table 1 List of attributes first arranged by modality, and then subdivided into individual attributes

Dominant modality	Attribute
Vision	vision, bright, dark, color, pattern, large, small, motion, biomotion, fast, slow, shape, complexity, face, body.
Auditory	audition, loud, low, high, sound, music, speech.
Somatosensory	touch, temperature, texture, weight, pain.
Gustatory +Smell	taste, smell.
Motor	head, upper limb, lower limb, practice.
Attention	attention, arousal.
Event	duration, long, short, caused, consequential, social, time.
Evaluation	benefit, harm, pleasant, unpleasant.
Cognition	human, communication, self, cognition, number.
Emotion	happy, sad, angry, disgusted, fearful, surprised.
Drive	drive, needs.
Spatial	landmark, path, scene, near, toward, away.

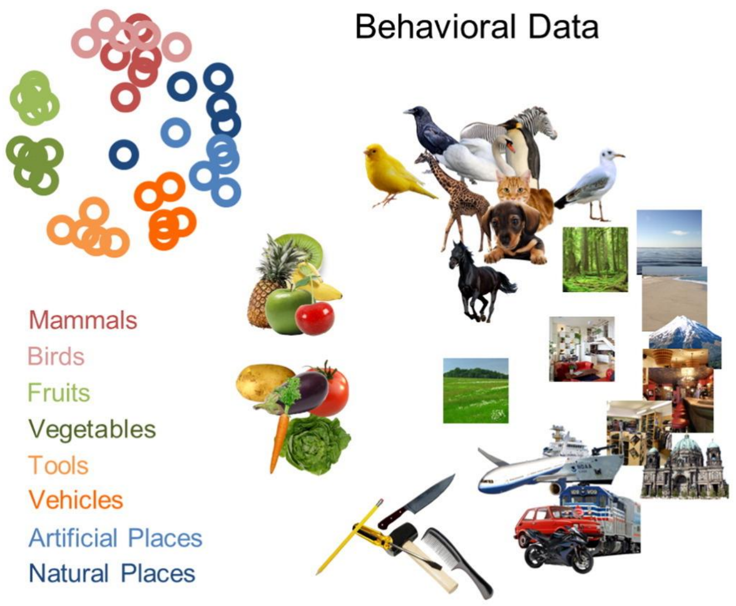
Anderson, Andrew James, Jeffrey R. Binder, Leonardo Fernandino, Colin J. Humphries, Lisa L. Conant, Rajeev DS Raizada, Feng Lin, and Edmund C. Lalor. "An integrated neural decoder of linguistic and experiential meaning." *Journal of Neuroscience* 39, no. 45 (2019): 8969-8987.

Anderson, Andrew James, Jeffrey R. Binder, Leonardo Fernandino, Colin J. Humphries, Lisa L. Conant, Mario Aguilar, Xixi Wang, Donias Doko, and Rajeev DS Raizada. "Predicting neural activity patterns associated with sentences using a neurobiologically motivated model of semantic representation." *Cerebral Cortex* 27, no. 9 (2017): 4379-4395.

Anderson, Andrew James, Kelsey McDermott, Brian Rooks, Kathi L. Heffner, David Dodell-Feder, and Feng V. Lin. "Decoding individual identity from brain activity elicited in imagining common experiences." *Nature communications* 11, no. 1 (2020): 1-14.

Binary attribute representations

- Each stimulus is represented using a binary vector capturing membership to one of the eight semantic categories.



- 42 neurally plausible semantic features (NPSFs)
  - Perceptual and affective characteristics of an entity (10 NPSFs coded such features, such as man-made, size, color, temperature, positive affective valence, high affective arousal), animate beings (person, human-group, animal), and time and space properties (e.g. unenclosed setting, change of location)

Word	NPSF features
Interview	Social, Mental action, Knowledge, Communication, Abstraction
Walk	Physical action, Change of location
Hurricane	Event, Change of physical state, Health, Natural, Negative affective valence, High affective arousal
Cellphone	Social action, Communication, Man-made, Inanimate
Judge	Social norms, Knowledge, Communication, Person
Clever	Attribute, Mental action, Knowledge, Positive affective valence, Abstraction

Handjaras, Giacomo, Emiliano Ricciardi, Andrea Leo, Alessandro Lenci, Luca Cecchetti, Mirco Cosottini, Giovanna Marotta, and Pietro Pietrini. "How concepts are encoded in the human brain: a modality independent, category-based cortical organization of semantic knowledge." *Neuroimage* 135 (2016): 232-242.

Wang, Jing, Vladimir L. Cherkassky, and Marcel Adam Just. "Predicting the brain activation pattern associated with the propositional content of a sentence: modeling neural representations of events and states." *Human brain mapping* 38, no. 10 (2017): 4865-4881.

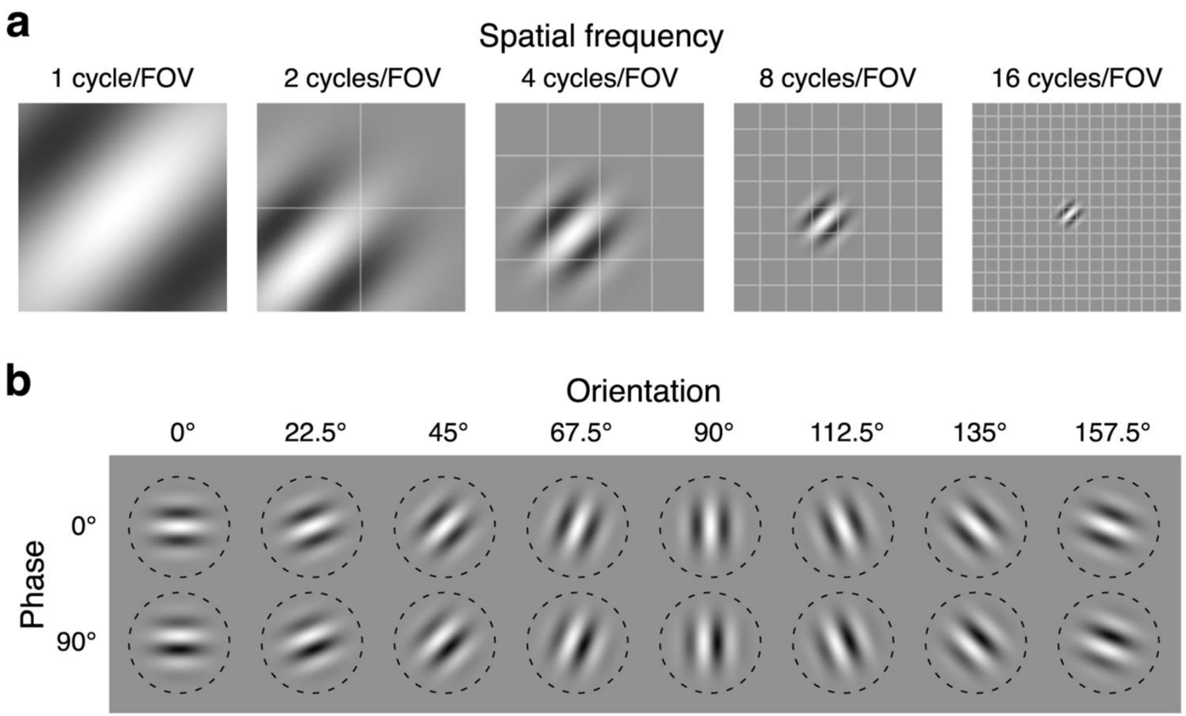
Agenda

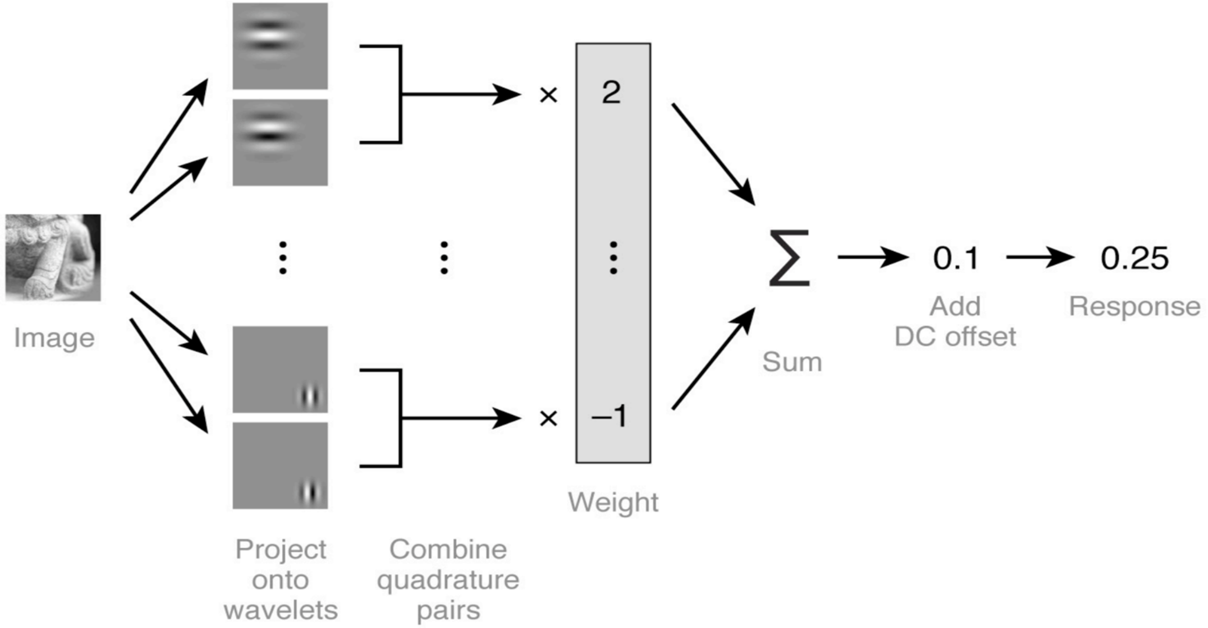
- Introduction to Brain encoding and decoding [30 min]
- Stimulus Representations [1 hour]
    - Text Stimulus Representations
    - **Visual Stimulus Representations**
    - Audio Stimulus Representations
    - Multimodal Stimulus Representations
- Coffee break [30 min]
- Deep Learning for Brain Decoding [1 hour 30 min]
- Lunch break [1 hour 30 min]
- Deep Learning for Brain Encoding [1 hour 30 min]
- Coffee break [30 min]
- Advanced Methods [1 hour 15 min]
- Summary and Future Trends [15 min]

Visual Stimuli

- Visual field filter banks (Thirion et al., 2006; Nishimoto et al., 2011).
- Gabor wavelet pyramid (Kay et al., 2008).
- HMAX model (Horikawa et al., 2017).
- Convolutional neural networks (Yamins et al., 2014; Anderson et al., 2017a; Beliy et al., 2019; Du et al., 2020; Nishida et al., 2020).

Visual Stimuli: Gabor wavelet pyramid





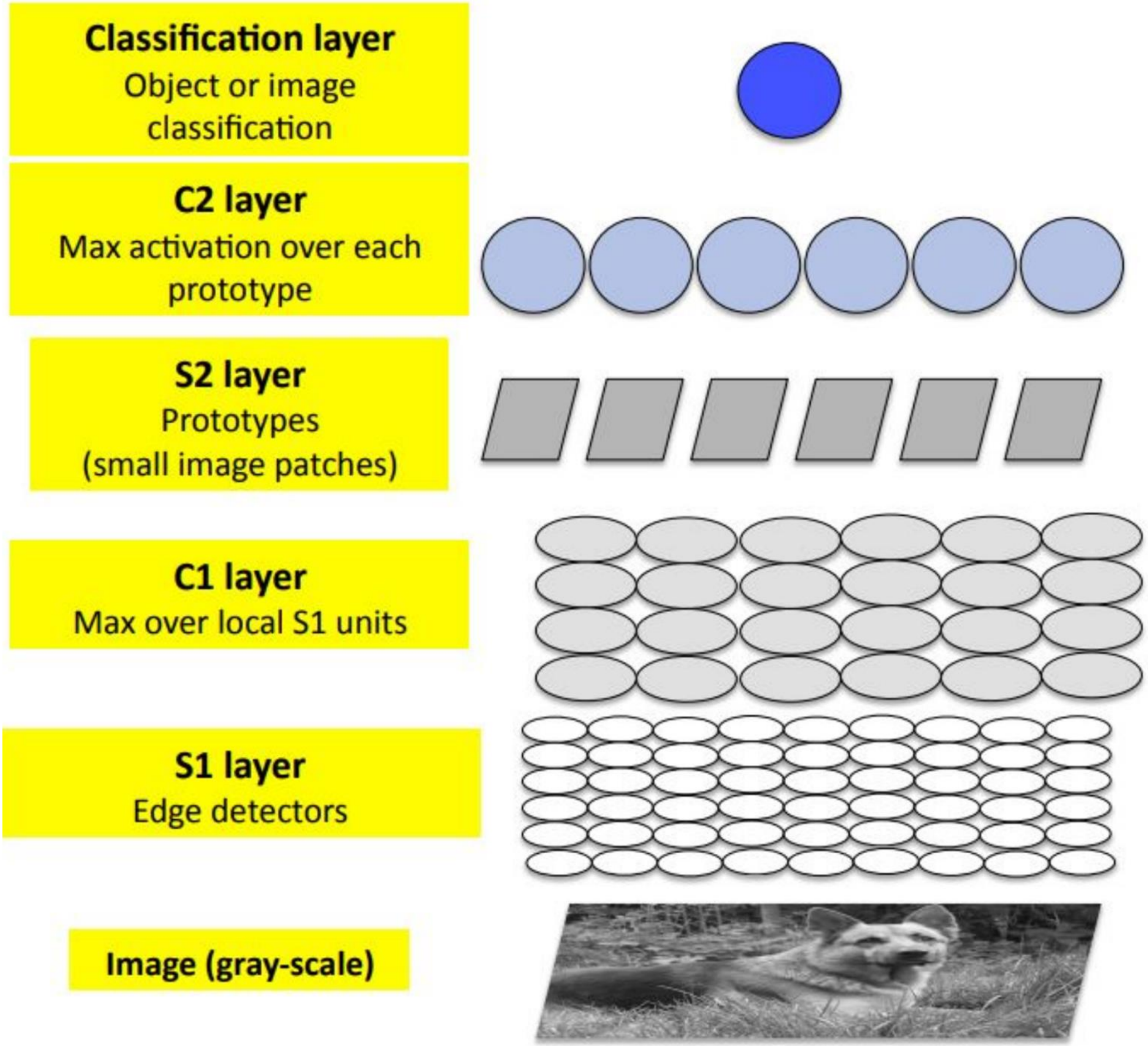
a, Spatial frequency and position. Wavelets occur at five spatial frequencies. This panel depicts one wavelet at each of the first five spatial frequencies. At each spatial frequency f cycles/field-of-view (FOV), wavelets are positioned on an f × f grid, as indicated by the translucent lines.
b, Orientation and phase. At each grid position, wavelets occur at eight orientations and two phases. This panel depicts a complete set of wavelets for a single grid position. Dashed lines indicate the bounds of the mask associated with each wavelet.

Gabor wavelet pyramid model. Each image is projected onto the individual Gabor wavelets comprising the Gabor wavelet pyramid. Gabor wavelets differ in size, position, orientation, spatial frequency, and phase. The projections for each quadrature pair of wavelets are squared, summed, and square-rooted, yielding a measure of contrast energy. The contrast energies for different quadrature wavelet pairs are weighted and then summed. Finally, a DC offset is added. The weights are determined by gradient descent with early stopping.

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.

Visual Stimuli: HMAX model

- Simple Cells S1
  - Input images are densely sampled by arrays of two-dimensional filters.
  - Output: -1 to 1
- Complex Cells C1: max pooling
- Simple Cells S2
  - Gaussian with mean 1 and standard deviation 1.
- Complex Cells C2: max pooling
- View Tuned Units (VTUs)
  - C2 units provide input to VTUs
  - C2 → VTU connections are the only stage of the HMAX model where learning occurs.



Riesenhuber, Maximilian, and Tomaso Poggio. "Hierarchical models of object recognition in cortex." *Nature neuroscience* 2, no. 11 (1999): 1019-1025.

Horikawa, Tomoyasu, and Yukiyasu Kamitani. "Generic decoding of seen and imagined objects using hierarchical visual features." *Nature communications* 8, no. 1 (2017): 1-15.

Visual Stimuli: Convolutional Neural Networks (CNNs)

- For word stimuli, gather 20 most relevant images using Google search, then get CNN representation (Anderson et al., 2017).
- AlexNet, VGG-16 (Nishida et al., 2020; Berezutskaya et al., 2020), Inception, ResNet, DenseNet.

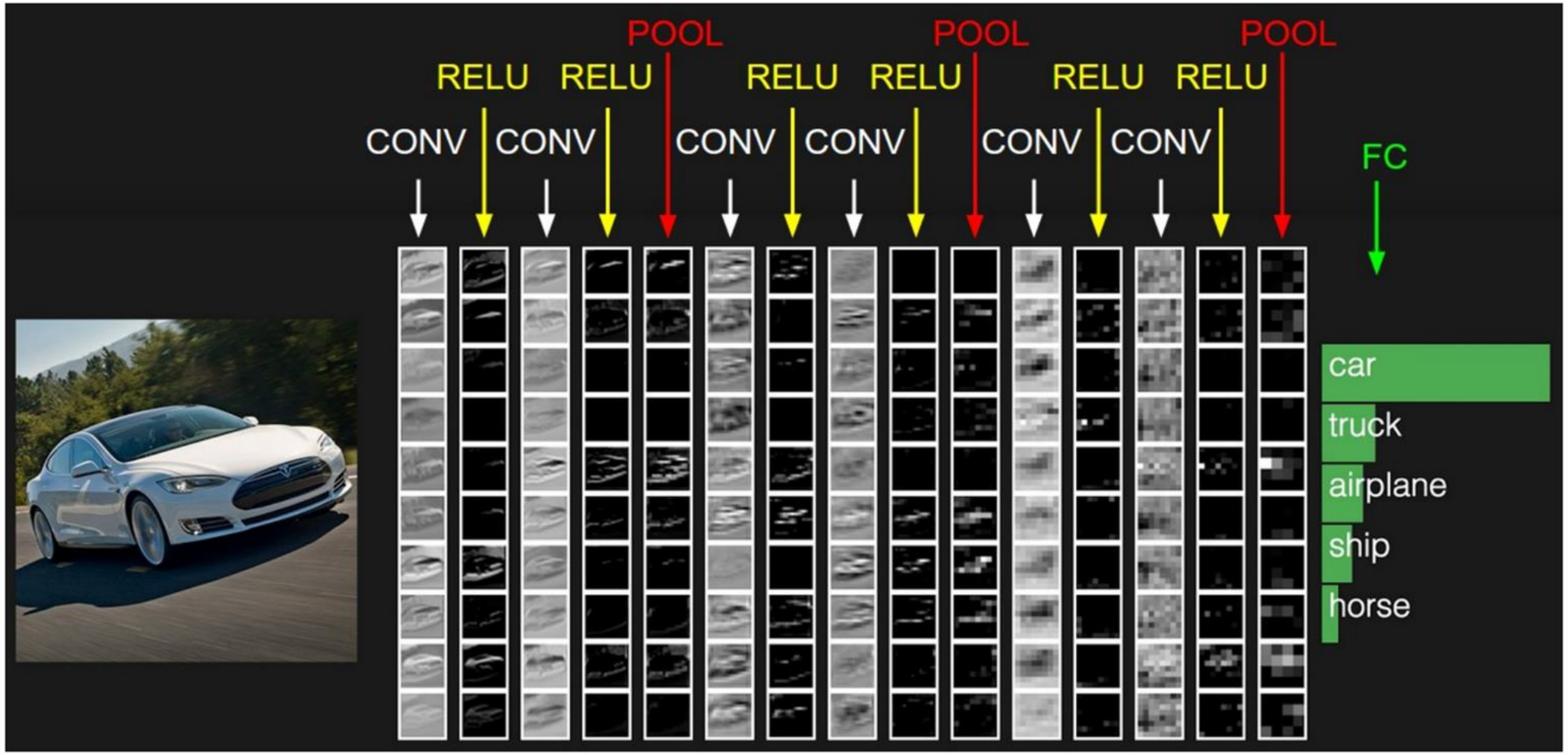
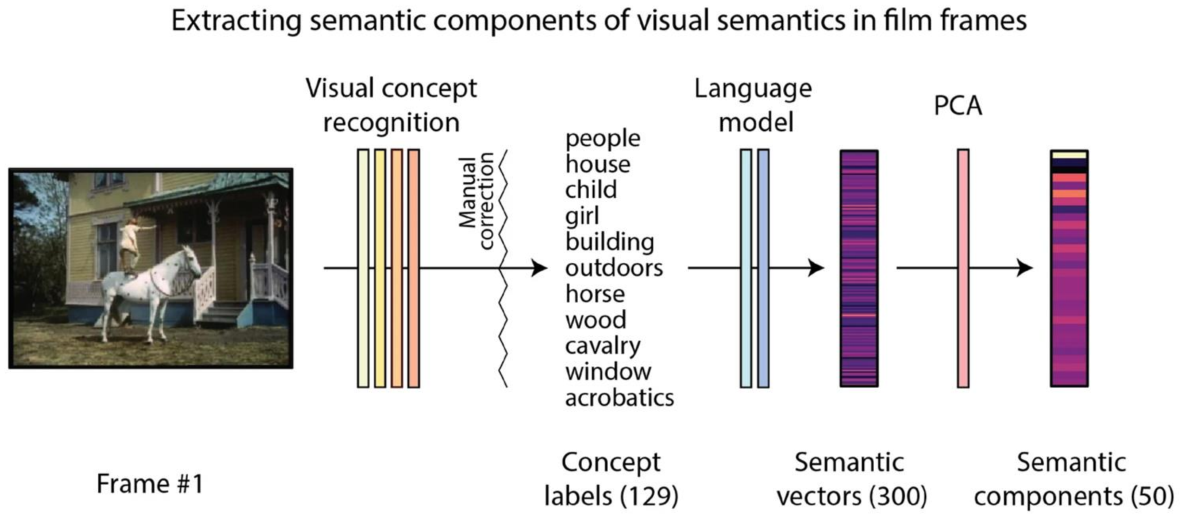
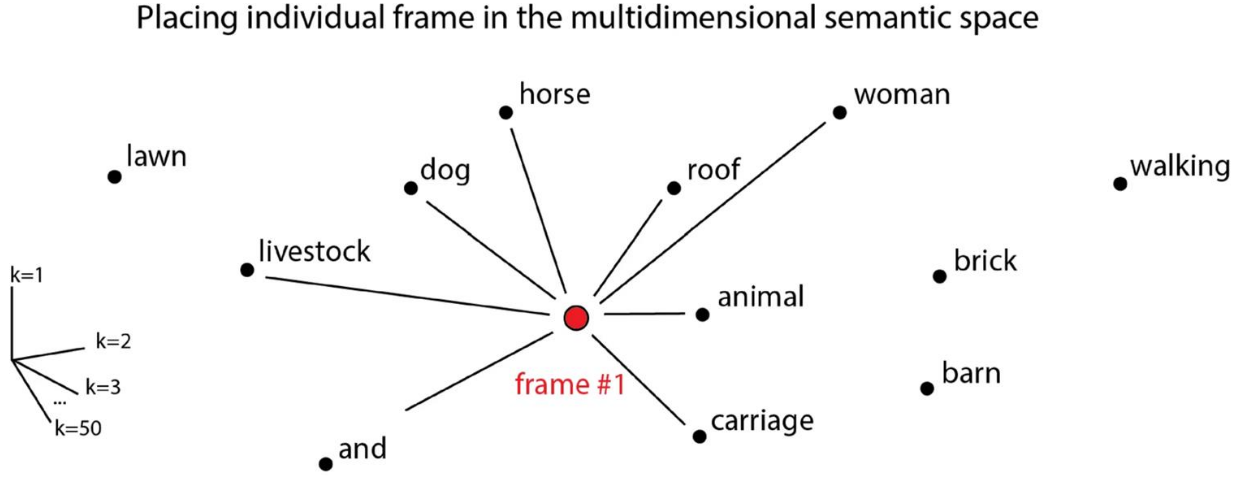


Figure source: A. Karpathy

Visual Stimuli: Object Recognition with Word embeddings

- Step 1: Pass film frames through concept recognition module to get up to 20 concept labels per frame.
  - Used Clarifai.
- Step 2: Get fastText embeddings for each concept label. Frame embedding is average of word embeddings.
- Step 3: PCA for dimensionality reduction.

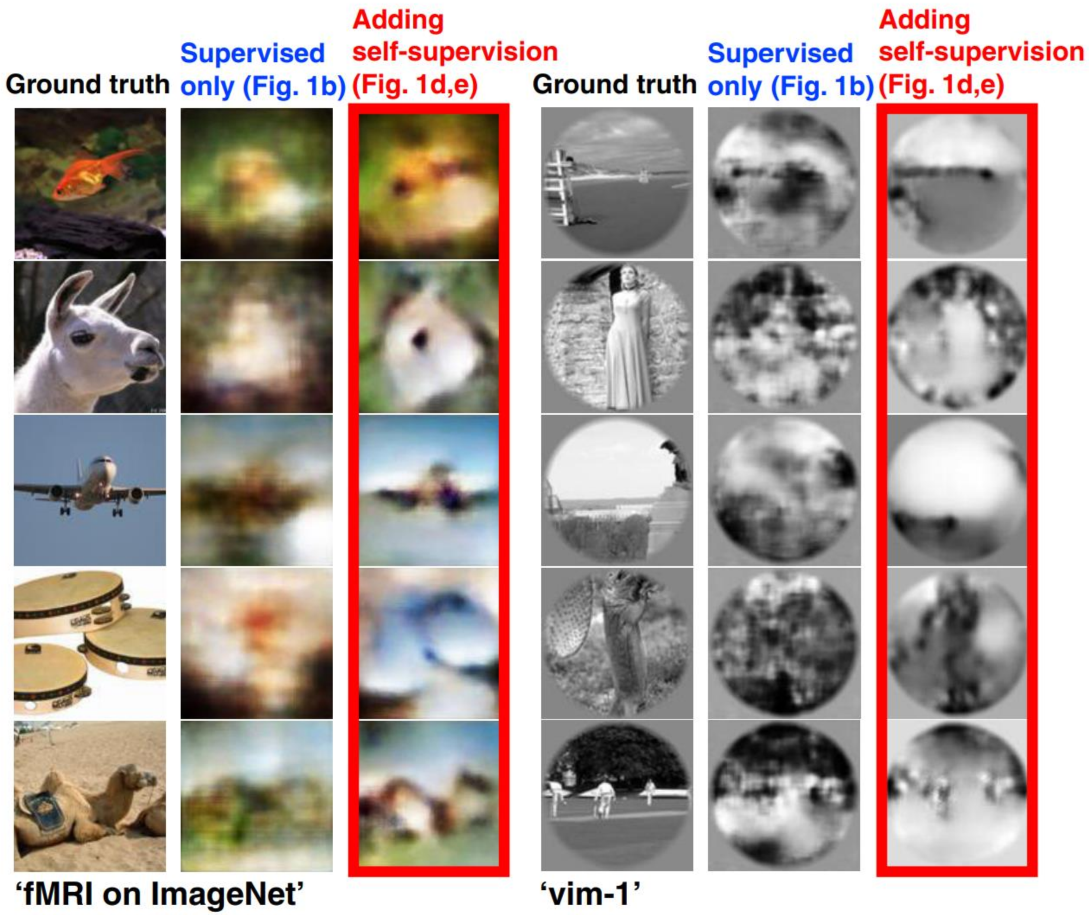


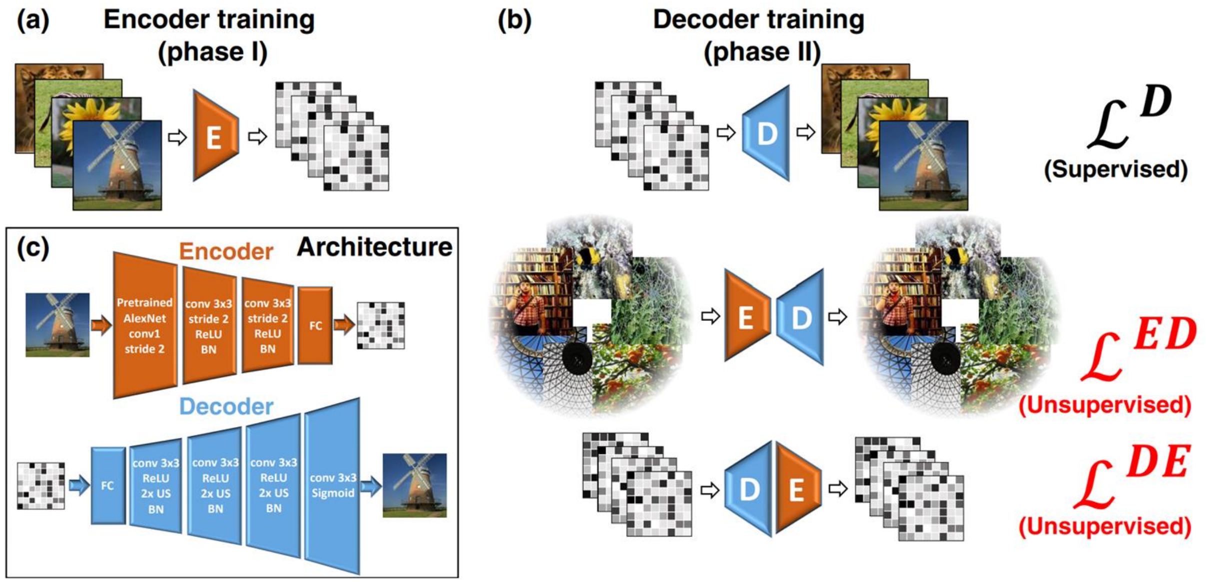


Berezutskaya, Julia, Zachary V. Freudenburg, Luca Ambrogioni, Umut Güçlü, Marcel AJ van Gerven, and Nick F. Ramsey. "Cortical network responses map onto data-driven features that capture visual semantics of movie fragments." *Scientific reports* 10, no. 1 (2020): 1-21.

Visual Stimuli: Semi-supervised CNNs

- Problem: Scarce labeled data.





Training phases & Architecture. (a) The first training phase: Supervised training of the Encoder with {Image, fMRI} pairs. (b) Second phase: Training the Decoder simultaneously with 3 types of data: {Image, fMRI} pairs (supervised examples), unlabeled natural images (self-supervision), and unlabeled test-fMRI (self-supervision). Note that the test-images are never used for training. The pretrained Encoder from the first training phase is kept fixed in the second phase. (c) Encoder and Decoder architectures. BN, US, and ReLU stand for batch normalization, up-sampling, and rectified linear unit, respectively.

Beliy, Roman, Guy Gaziv, Assaf Hoogi, Francesca Strappini, Tal Golan, and Michal Irani. "From voxels to pixels and back: Self-supervision in natural-image reconstruction from fMRI." *Advances in Neural Information Processing Systems* 32 (2019).

Visual Stimuli: Convolutional LSTM Autoencoder

StepEncog, a convolutional LSTM autoencoder model trained on fMRI voxels.

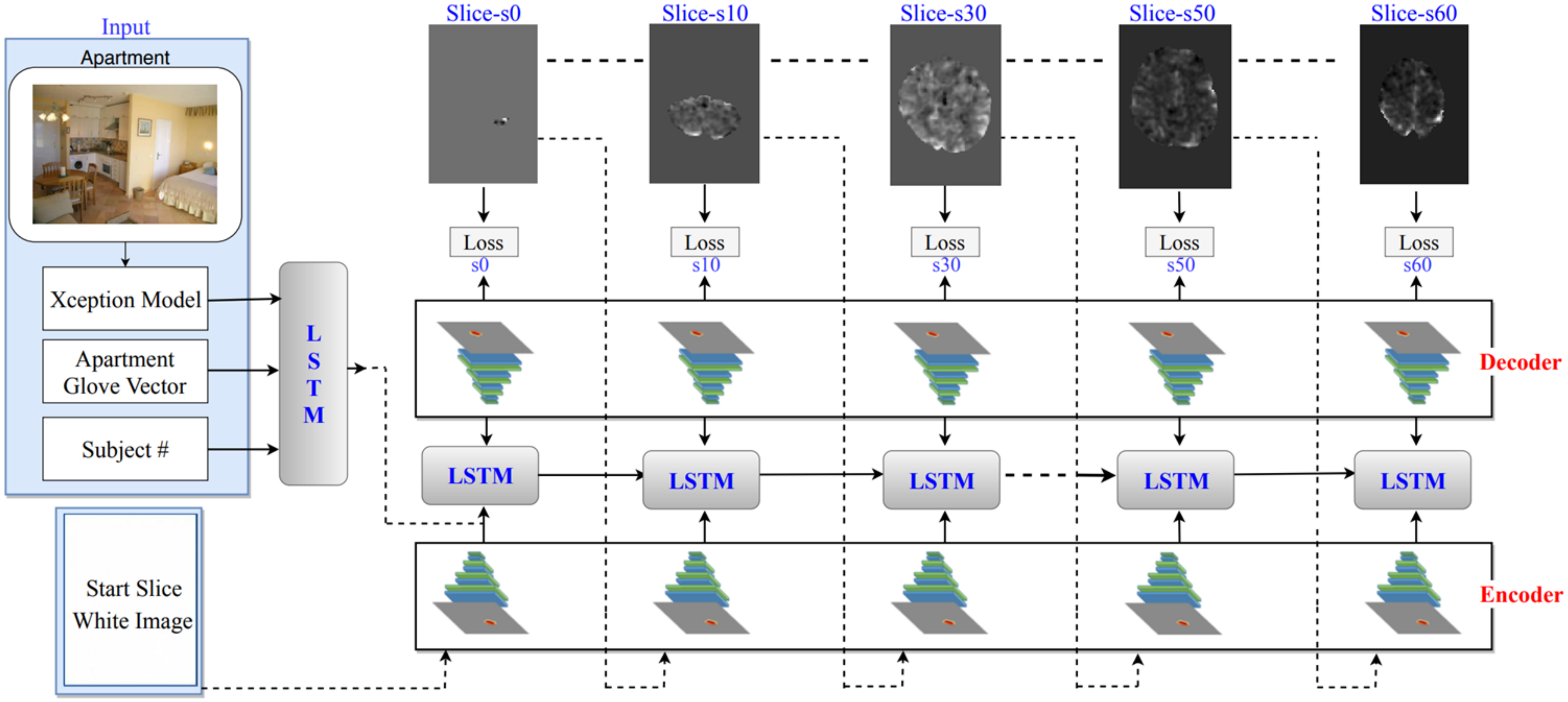
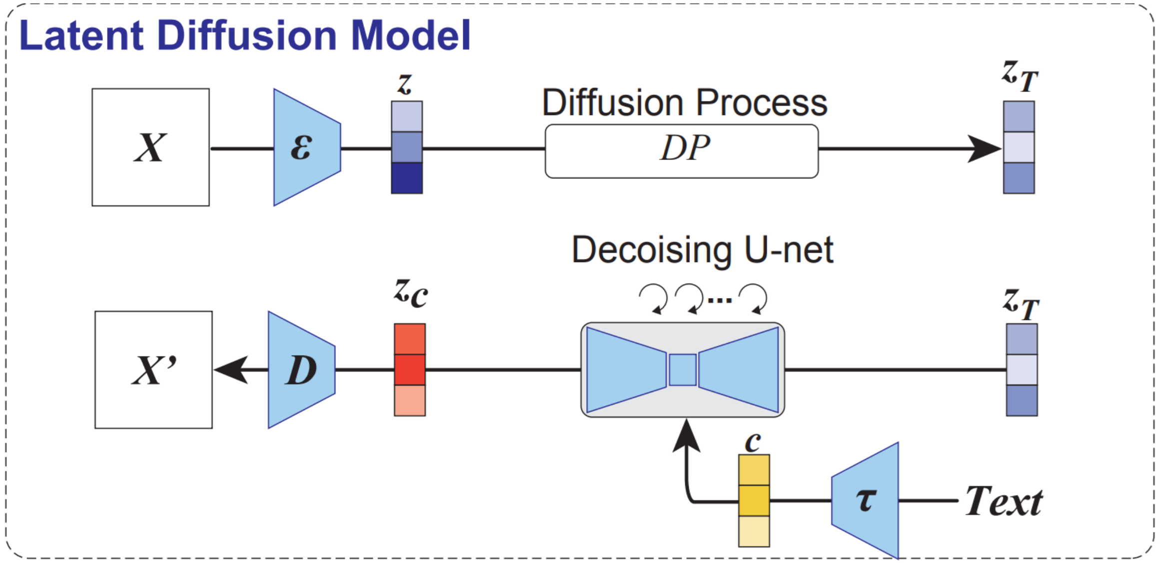


Fig. 2. Architecture of the StepEncog: the Convolutional LSTM autoencoder model used for our experiments. We used multi-modal embedding along with fMRI slices as input, and "step-ahead" fMRI slices as output.

Oota, Subba Reddy, Vijay Rowtula, Manish Gupta, and Raju S. Bapi. "StepEncog: A convolutional LSTM autoencoder for near-perfect fMRI encoding." In *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-8. IEEE, 2019.

Latent Diffusion Models



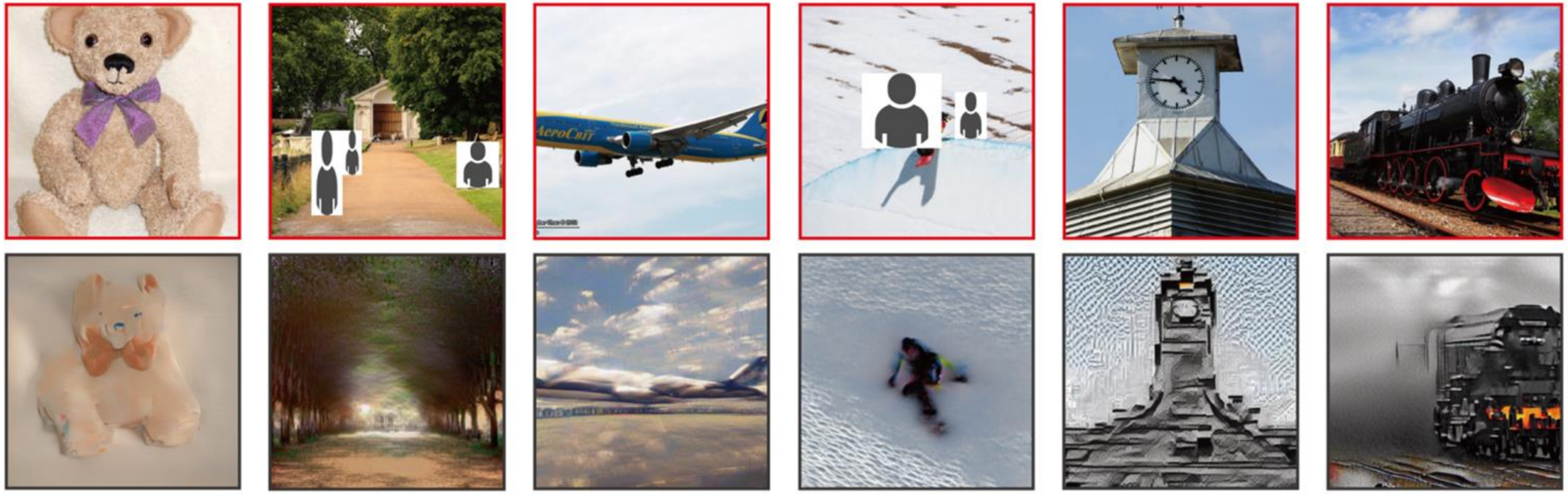


Figure 1. Presented images (red box, top row) and images reconstructed from fMRI signals (gray box, bottom row) for one subject (subj01).

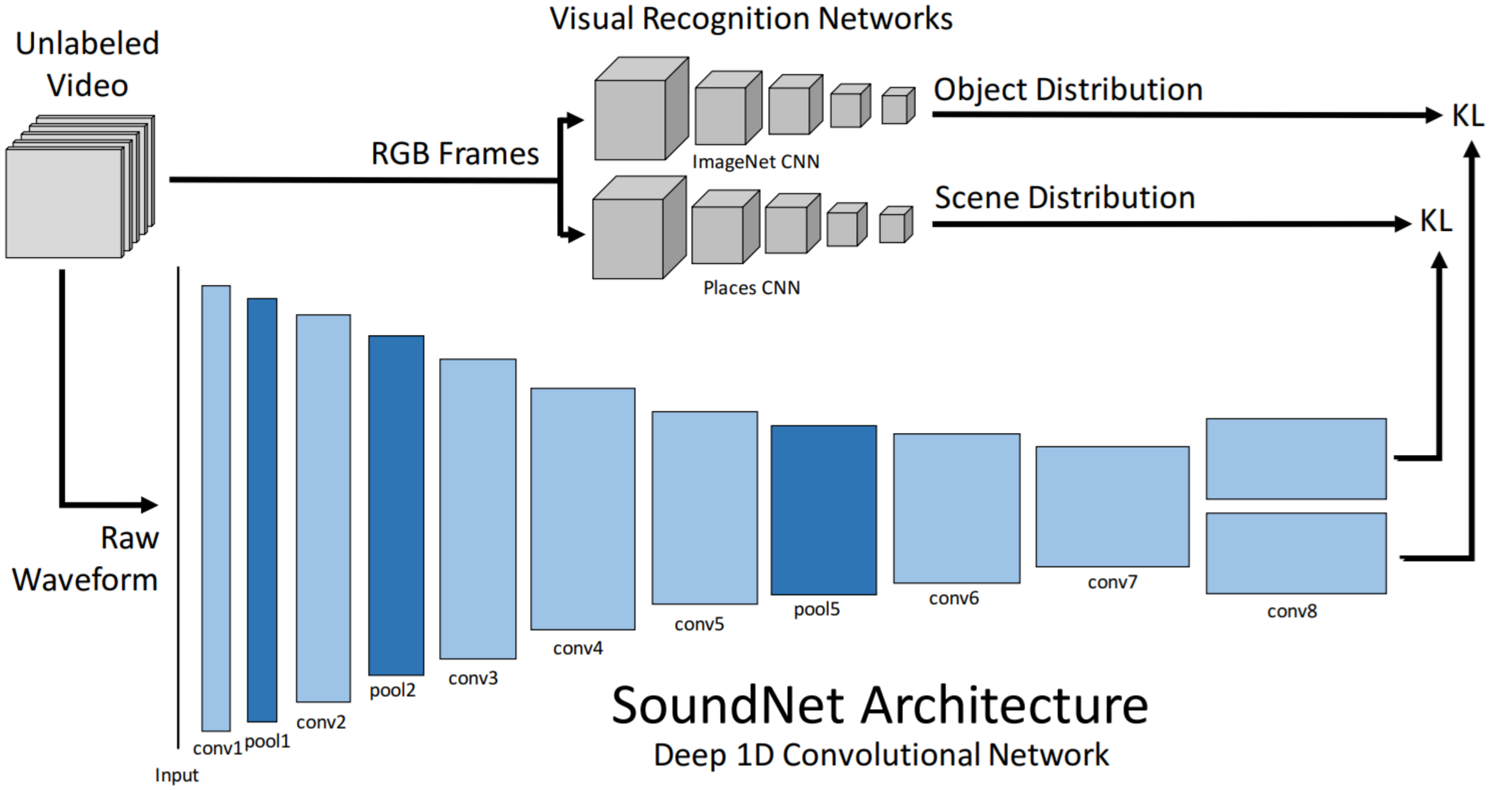
Takagi, Yu, and Shinji Nishimoto. "High-resolution image reconstruction with latent diffusion models from human brain activity." In *CVPR*, pp. 14453-14463. 2023.

Agenda

- Introduction to Brain encoding and decoding [30 min]
- Stimulus Representations [1 hour]
  - Text Stimulus Representations
  - Visual Stimulus Representations
  - **Audio Stimulus Representations**
  - Multimodal Stimulus Representations
- Coffee break [30 min]
- Deep Learning for Brain Decoding [1 hour 30 min]
- Lunch break [1 hour 30 min]
- Deep Learning for Brain Encoding [1 hour 30 min]
- Coffee break [30 min]
- Advanced Methods [1 hour 15 min]
- Summary and Future Trends [15 min]

Audio Stimuli

- Word rate, Phoneme rate, Presence of phonemes (Huth et al., 2016).
- SoundNet (Aytar, Vondrick, and Torralba 2016) features (Nishida et al., 2020)



Huth, Alexander G., Wendy A. De Heer, Thomas L. Griffiths, Frédéric E. Theunissen, and Jack L. Gallant. "Natural speech reveals the semantic maps that tile human cerebral cortex." *Nature* 532, no. 7600 (2016): 453-458.

Nishida, Satoshi, Yusuke Nakano, Antoine Blanc, Naoya Maeda, Masataka Kado, and Shinji Nishimoto. "Brain-mediated transfer learning of convolutional neural networks." In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, pp. 5281-5288. 2020.

Agenda

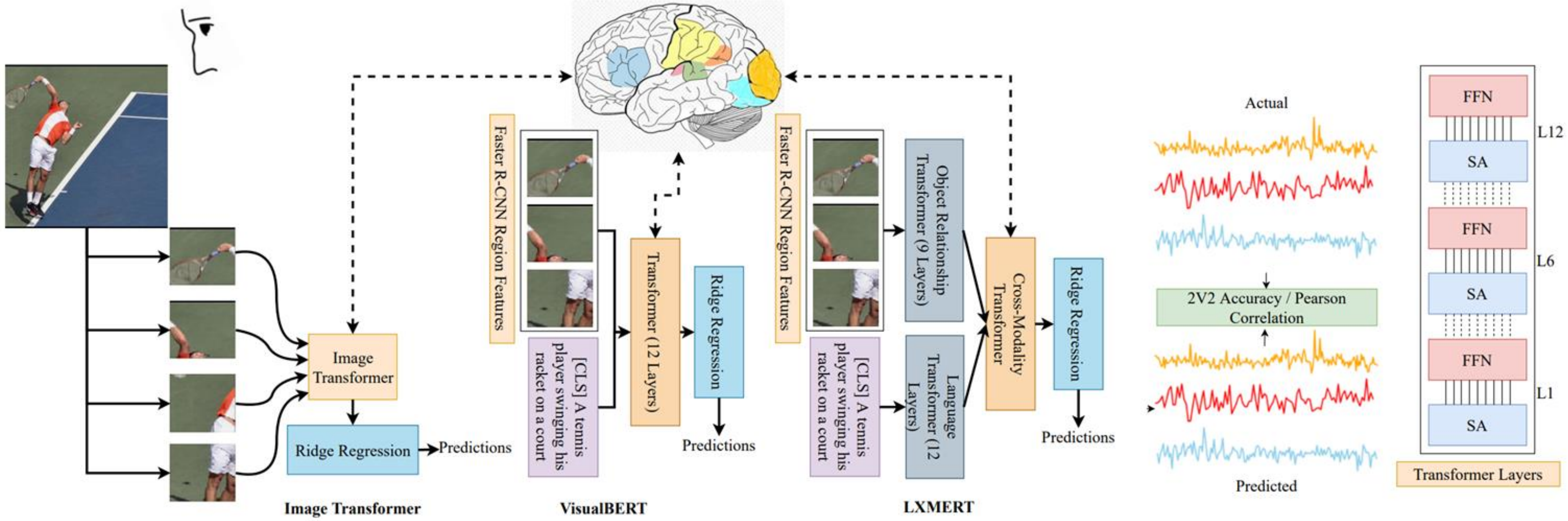
- Introduction to Brain encoding and decoding [30 min]
- Stimulus Representations [1 hour]
  - Text Stimulus Representations
  - Visual Stimulus Representations
  - Audio Stimulus Representations
  - **Multimodal Stimulus Representations**
- Coffee break [30 min]
- Deep Learning for Brain Decoding [1 hour 30 min]
- Lunch break [1 hour 30 min]
- Deep Learning for Brain Encoding [1 hour 30 min]
- Coffee break [30 min]
- Advanced Methods [1 hour 15 min]
- Summary and Future Trends [15 min]

Multimodal Stimulus Representations

- Processing videos required audio+image representations
  - E.g., VGG+SoundNet (Nishida et al., 2020)
- Image+text combination models (Wang et al., 2020)
  - GloVe+VGG, and ELMo+VGG
  - Averaging or concatenation

Wang, Shaonan, Jiajun Zhang, Haiyan Wang, Nan Lin, and Chengqing Zong. "Fine-grained neural decoding with distributed word representations." *Information Sciences* 507 (2020): 256-272.

Multimodal Stimuli: Visio-linguistic representations



- Pretrained CNNs: VGGNet19, ResNet50, InceptionV2ResNet and EfficientNetB5
- Pretrained text Transformers: RoBERTa
- Image Transformers: Vision Transformer (ViT), Data Efficient Image Transformer (DEiT), and Bidirectional Encoder representation from Image Transformer (BEiT).
- Late-fusion models: VGGNet19+RoBERTa, ResNet50+RoBERTa, InceptionV2ResNet+RoBERTa and EfficientNetB5+RoBERTa.
- Multi-modal Transformers: Contrastive Language-Image Pre-training (CLIP), Learning Cross-Modality Encoder Representations from Transformers (LXMERT), and VisualBERT.
  - VisualBERT performs the best for brain encoding!

Oota, Subba Reddy, Jashn Arora, Vijay Rowtula, Manish Gupta, and Raju S. Bapi. "Visio-Linguistic Brain Encoding." *arXiv preprint arXiv:2204.08261* (2022).

Agenda

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

References

[1] Nicolas Affolter, Beni Egressy, Damian Pascual, and Roger Wattenhofer. Brain2word: Decoding brain activity for language generation. arXiv preprint arXiv:2009.04765, 2020.

[2] Andrew J Anderson, Douwe Kiela, Stephen Clark, and Massimo Poesio. Visually grounded and textual semantic models differentially decode brain activity associated with concrete and abstract nouns. Transactions of the Association for Computational Linguistics, 5:17–30, 2017.

[3] Andrew James Anderson, Jeffrey R Binder, Leonardo Fernandino, Colin J Humphries, Lisa L Conant, Mario Aguilar, Xixi Wang, Donias Doko, and Rajeev DS Raizada. Predicting neural activity patterns associated with sentences using a neurobiologically motivated model of semantic representation. Cerebral Cortex, 27(9):4379–4395, 2017.

[4] Andrew James Anderson, Jeffrey R Binder, Leonardo Fernandino, Colin J Humphries, Lisa L Conant, Rajeev DS Raizada, Feng Lin, and Edmund C Lalor. An integrated neural decoder of linguistic and experiential meaning. Journal of Neuroscience, 39(45):8969–8987, 2019.

[5] Andrew James Anderson, Kelsey McDermott, Brian Rooks, Kathi L Heffner, David Dodell-Feder, and Feng V Lin. Decoding individual identity from brain activity elicited in imagining common experiences. Nature communications, 11(1):1–14, 2020.

[6] Richard Antonello, Javier Turek, Vy Vo, and Alexander Huth. Low-dimensional structure in the space of language representations is reflected in brain responses. arXiv preprint arXiv:2106.05426, 2021.

[7] Roman Beliy, Guy Gaziv, Assaf Hoogi, Francesca Strappini, Tal Golan, and Michal Irani. From voxels to pixels and back: Self-supervision in naturalimage reconstruction from fmri. arXiv preprint arXiv:1907.02431, 2019.

[8] Julia Berezutskaya, Zachary V Freudenburg, Luca Ambrogioni, Umut Güçlü, Marcel AJ van Gerven, and Nick F Ramsey. Cortical network responses map onto data-driven features that capture visual semantics of movie fragments. Scientific reports, 10(1):1–21, 2020.

[9] Charlotte Caucheteux, Alexandre Gramfort, and Jean-Remi King. Disentangling syntax and semantics in the brain with deep networks. In International Conference on Machine Learning, pages 1336–1348. PMLR, 2021.

[10] Charlotte Caucheteux and Jean-Rémi King. Language processing in brains and deep neural networks: computational convergence and its limits. BioRxiv, 2020.

[11] Joshua S Cetron, Andrew C Connolly, Solomon G Diamond, Vicki V May, James V Haxby, and David JM Kraemer. Decoding individual differences in stem learning from functional mri data. Nature communications, 10(1):1–10, 2019.

References

[12] Nadine Chang, John A Pyles, Austin Marcus, Abhinav Gupta, Michael J Tarr, and Elissa M Aminoff. Bold5000, a public fmri dataset while viewing 5000 visual images. Scientific data, 6(1):1–18, 2019.

[13] Radoslaw Martin Cichy, Kshitij Dwivedi, Benjamin Lahner, Alex Lascelles, Polina Iamshchinina, M Graumann, A Andonian, NAR Murty, K Kay, Gemma Roig, et al. The algonauts project 2021 challenge: How the human brain makes sense of a world in motion. arXiv preprint arXiv:2104.13714, 2021.

[14] Radoslaw Martin Cichy, Gemma Roig, Alex Andonian, Kshitij Dwivedi, Benjamin Lahner, Alex Lascelles, Yalda Mohsenzadeh, Kandan Ramakrishnan, and Aude Oliva. The algonauts project: A platform for communication between the sciences of biological and artificial intelligence. arXiv e-prints, pages arXiv–1905, 2019.

[15] Changde Du, Changying Du, Lijie Huang, and Huiguang He. Conditional generative neural decoding with structured cnn feature prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 2629–2636, 2020.

[16] Michael Eickenberg, Alexandre Gramfort, Gaël Varoquaux, and Bertrand Thirion. Seeing it all: Convolutional network layers map the function of the human visual system. NeuroImage, 152:184–194, 2017.

[17] Jack Gallant. Human brain mapping and brain decoding, 2017.

[18] Jon Gauthier and Roger Levy. Linking artificial and human neural representations of language. arXiv preprint arXiv:1910.01244, 2019.

[19] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.

[20] Giacomo Handjaras, Emiliano Ricciardi, Andrea Leo, Alessandro Lenci, Luca Cecchetti, Mirco Cosottini, Giovanna Marotta, and Pietro Pietrini. How concepts are encoded in the human brain: a modality independent, category-based cortical organization of semantic knowledge. Neuroimage, 135:232–242, 2016.

[21] Nora Hollenstein, Antonio de la Torre, Nicolas Langer, and Ce Zhang. Cognival: A framework for cognitive word embedding evaluation. In Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL), pages 538–549, 2019.

[22] Nora Hollenstein, Jonathan Rotsztejn, Marius Troendle, Andreas Pedroni, Ce Zhang, and Nicolas Langer. Zuco, a simultaneous eeg and eye-tracking resource for natural sentence reading. Scientific data, 5(1):1–13, 2018.

References

[23] Alexander G Huth, Wendy A De Heer, Thomas L Griffiths, Frédéric E Theunissen, and Jack L Gallant. Natural speech reveals the semantic maps that tile human cerebral cortex. Nature, 532(7600):453–458, 2016.

[24] Shailee Jain and Alexander G Huth. Incorporating context into language encoding models for fmri. In Proceedings of the 32nd International Conference on Neural Information Processing Systems, pages 6629–6638, 2018.

[25] S Jat, H Tang, P Talukdar, and T Mitchel. Relating simple sentence representations in deep neural networks and the brain. In ACL 2019-57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, pages 5137–5154. Association for Computational Linguistics (ACL), 2020.

[26] Marcel Adam Just, Vladimir L Cherkassky, Sandesh Aryal, and Tom M Mitchell. A neurosemantic theory of concrete noun representation based on the underlying brain codes. PloS one, 5(1):e8622, 2010.

[27] Kendrick N Kay, Thomas Naselaris, Ryan J Prenger, and Jack L Gallant. Identifying natural images from human brain activity. Nature, 452(7185):352–355, 2008.

[28] Jonas Kubilius, Martin Schrimpf, Kohitij Kar, Rishi Rajalingham, Ha Hong, Najib Majaj, Elias Issa, Pouya Bashivan, Jonathan Prescott-Roy, Kailyn Schmidt, et al. Brain-like object recognition with high-performing shallow recurrent anns. Advances in Neural Information Processing Systems, 32:12805–12816, 2019.

[29] Tom Mitchell. Neural representations of language meaning, 2014.

[30] Tom M Mitchell, 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(5880):1191–1195, 2008.

[31] Thomas Naselaris, Ryan J Prenger, Kendrick N Kay, Michael Oliver, and Jack L Gallant. Bayesian reconstruction of natural images from human brain activity. Neuron, 63(6):902–915, 2009.

[32] Samuel A Nastase, Yun-Fei Liu, Hanna Hillman, Asieh Zadbood, Liat Hasenfratz, Neggin Keshavarzian, Janice Chen, Christopher J Honey, Yaara Yeshurun, Mor Regev, et al. Narratives: fmri data for evaluating models of naturalistic language comprehension. bioRxiv, pages 2020–12, 2021.

[33] Satoshi Nishida, Yusuke Nakano, Antoine Blanc, Naoya Maeda, Masataka Kado, and Shinji Nishimoto. Brain-mediated transfer learning of convolutional neural networks. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 5281–5288, 2020.

References

[34] Shinji Nishimoto, An T Vu, Thomas Naselaris, Yuval Benjamini, Bin Yu, and Jack L Gallant. Reconstructing visual experiences from brain activity evoked by natural movies. Current biology, 21(19):1641–1646, 2011.

[35] Subba Reddy Oota, Vijay Rowtula, Manish Gupta, and Raju S Bapi. Stepencog: A convolutional lstm autoencoder for near-perfect fmri encoding. In 2019 International Joint Conference on Neural Networks (IJCNN), pages 1–8. IEEE, 2019.

[36] Francisco Pereira, Matthew Botvinick, and Greg Detre. Using wikipedia to learn semantic feature representations of concrete concepts in neuroimaging experiments. Artificial intelligence, 194:240– 252, 2013.

[37] Francisco Pereira, Bin Lou, Brianna Pritchett, Nancy Kanwisher, Matthew Botvinick, and Evelina Fedorenko. Decoding of generic mental representations from functional mri data using word embeddings. bioRxiv, page 057216, 2016.

[38] Francisco Pereira, Bin Lou, Brianna Pritchett, Samuel Ritter, Samuel J Gershman, Nancy Kanwisher, Matthew Botvinick, and Evelina Fedorenko. Toward a universal decoder of linguistic meaning from brain activation. Nature communications, 9(1):1–13, 2018.

[39] Martin Schrimpf, Idan Blank, Greta Tuckute, Carina Kauf, Eghbal A Hosseini, Nancy Kanwisher, Joshua Tenenbaum, and Evelina Fedorenko. The neural architecture of language: Integrative reverseengineering converges on a model for predictive processing. PNAS, Vol:To appear, 2021.

[40] Martin Schrimpf, Jonas Kubilius, Ha Hong, Najib J Majaj, Rishi Rajalingham, Elias B Issa, Kohitij Kar, Pouya Bashivan, Jonathan Prescott-Roy, Franziska Geiger, et al. Brain-score: Which artificial neural network for object recognition is most brain-like? BioRxiv, page 407007, 2020.

[41] Dan Schwartz, Mariya Toneva, and Leila Wehbe. Inducing brain-relevant bias in natural language processing models. Advances in Neural Information Processing Systems, 32:14123–14133, 2019.

[42] K Seeliger, RP Sommers, Umut Güçlü, Sander E Bosch, and MAJ Van Gerven. A large singleparticipant fmri dataset for probing brain responses to naturalistic stimuli in space and time. bioRxiv, page 687681, 2019.

[43] Vishwajeet Singh, Krishna P. Miyapuram, and Raju S. Bapi. Detection of cognitive states from fmri data using machine learning techniques. In Manuela M. Veloso, editor, IJCAI, pages 587–592, 2007.

[44] Jonathan Smallwood and Jonathan W Schooler. The science of mind wandering: empirically navigating the stream of consciousness. Annual review of psychology, 66:487–518, 2015.

References

[45] Jingyuan Sun, Shaonan Wang, Jiajun Zhang, and Chengqing Zong. Towards sentence-level brain decoding with distributed representations. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, pages 7047–7054, 2019.

[46] Jingyuan Sun, Shaonan Wang, Jiajun Zhang, and Chengqing Zong. Neural encoding and decoding with distributed sentence representations. IEEE Transactions on Neural Networks and Learning Systems, 32(2):589–603, 2020.

[47] Bertrand Thirion. Statistical inference in highdimension and application to brain imaging, 2019.

[48] Bertrand Thirion, Edouard Duchesnay, Edward Hubbard, Jessica Dubois, Jean-Baptiste Poline, Denis Lebihan, and Stanislas Dehaene. Inverse retinotopy: inferring the visual content of images from brain activation patterns. Neuroimage, 33(4):1104– 1116, 2006.

[49] Mariya Toneva, 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.

[50] Mariya Toneva and Leila Wehbe. Interpreting and improving natural-language processing (in machines) with natural language-processing (in the brain). arXiv preprint arXiv:1905.11833, 2019.

[51] Aria Wang, Michael Tarr, and Leila Wehbe. Neural taskonomy: Inferring the similarity of task-derived representations from brain activity. Advances in Neural Information Processing Systems, 32:15501– 15511, 2019.

[52] Jing Wang, Vladimir L Cherkassky, and Marcel Adam Just. Predicting the brain activation pattern associated with the propositional content of a sentence: Modeling neural representations of events and states. Human brain mapping, 38(10):4865– 4881, 2017.

[53] Shaonan Wang, Jiajun Zhang, Haiyan Wang, Nan Lin, and Chengqing Zong. Fine-grained neural decoding with distributed word representations. Information Sciences, 507:256–272, 2020.

[54] Leila Wehbe, Brian Murphy, Partha Talukdar, Alona Fyshe, Aaditya Ramdas, and Tom Mitchell. Simultaneously uncovering the patterns of brain regions involved in different story reading subprocesses. in press, 2014.

[55] Daniel LK Yamins, 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(23):8619–8624, 2014.

[56] Boyle, Julie A., Basile Pinsard, A. Boukhdhir, S. Belleville, S. Bram-batti, J. Chen, J. Cohen-Adad et al. "The Courtois project on neuronal modelling: 2020 data release." In Presented at the 26th annual meeting of the Organization for Human Brain Mapping. 2020.

Deep Neural Networks and Brain Alignment: Brain Encoding and Decoding

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

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

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

Agenda

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

Agenda

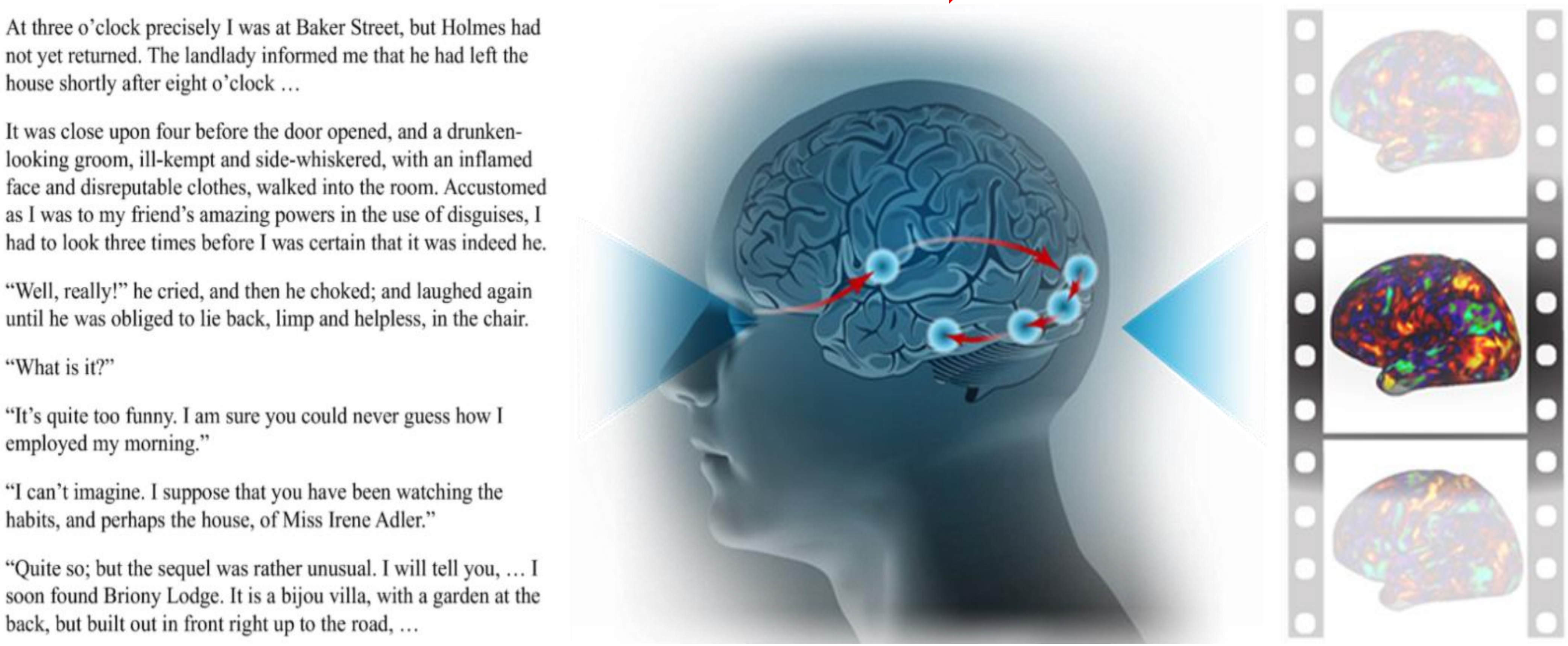
- Introduction to Brain encoding and decoding [30 min]
- Stimulus Representations [1 hour]
- Coffee break [30 min]
- **Deep Learning for Brain Decoding [1 hour 30 min]**
- Lunch break [1 hour 30 min]
- Deep Learning for Brain Encoding [1 hour 30 min]
- Coffee break [30 min]
- Advanced Methods [1 hour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

Outline

- Introduction to Brain Decoding
- Decoding models
  - Linear Models
  - Non-Linear Models (including DNNs)
- Language
  - Periera et al. 2018, Gauthier et al. 2019, Huth et al. 2023, Oota et al. 2022

Encoding vs. Decoding

Stimulus Representation → Encoding → fMRI



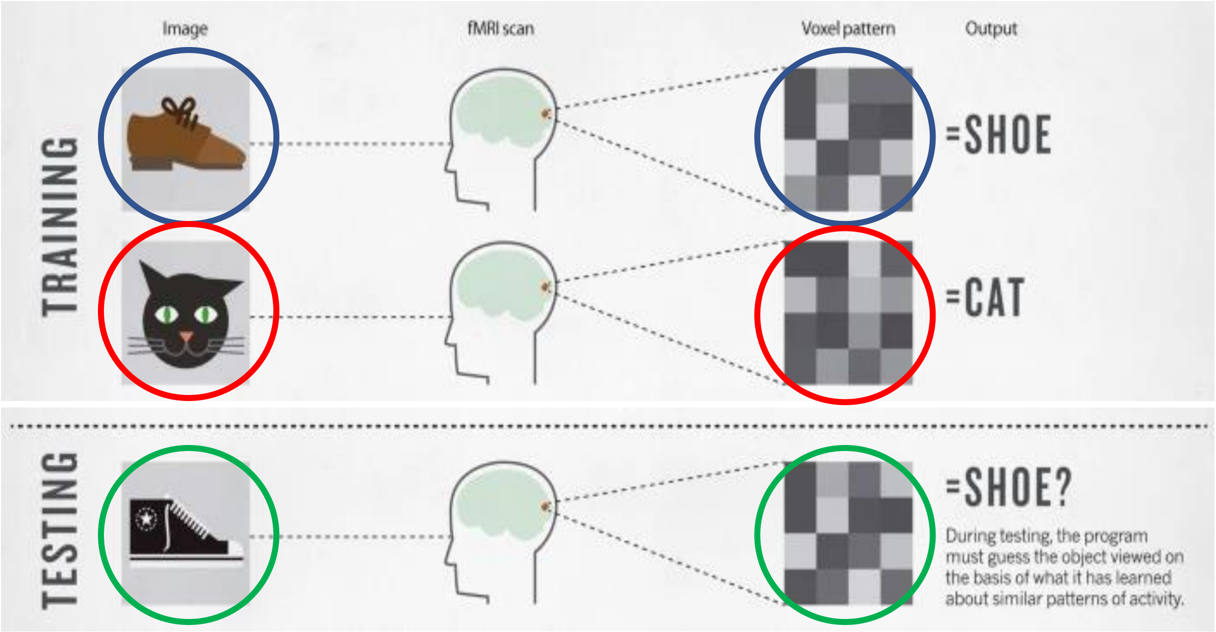
Stimulus Representation ← Decoding ← fMRI

Haiguang Wen et al, 2017

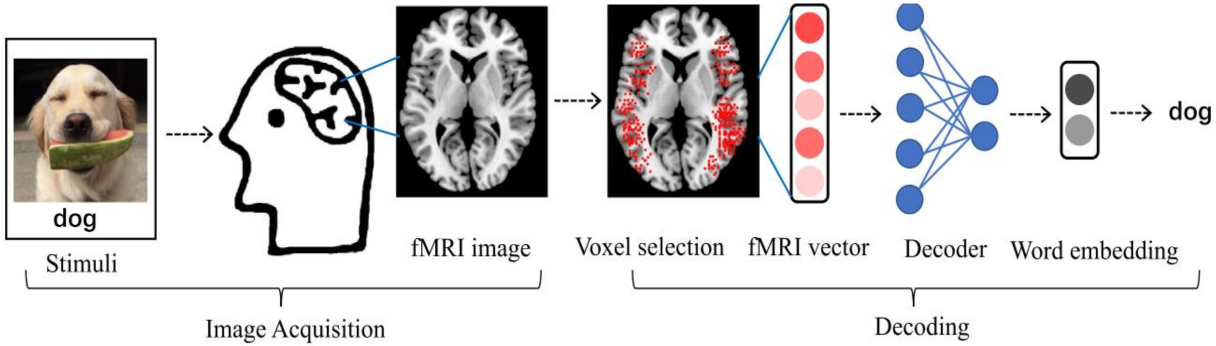
What is Brain Decoding?

- Can we reconstruct the stimulus, given the brain response?
- Can you read the mind with fMRI?
- Or at least tell what the person saw?

Visual Task

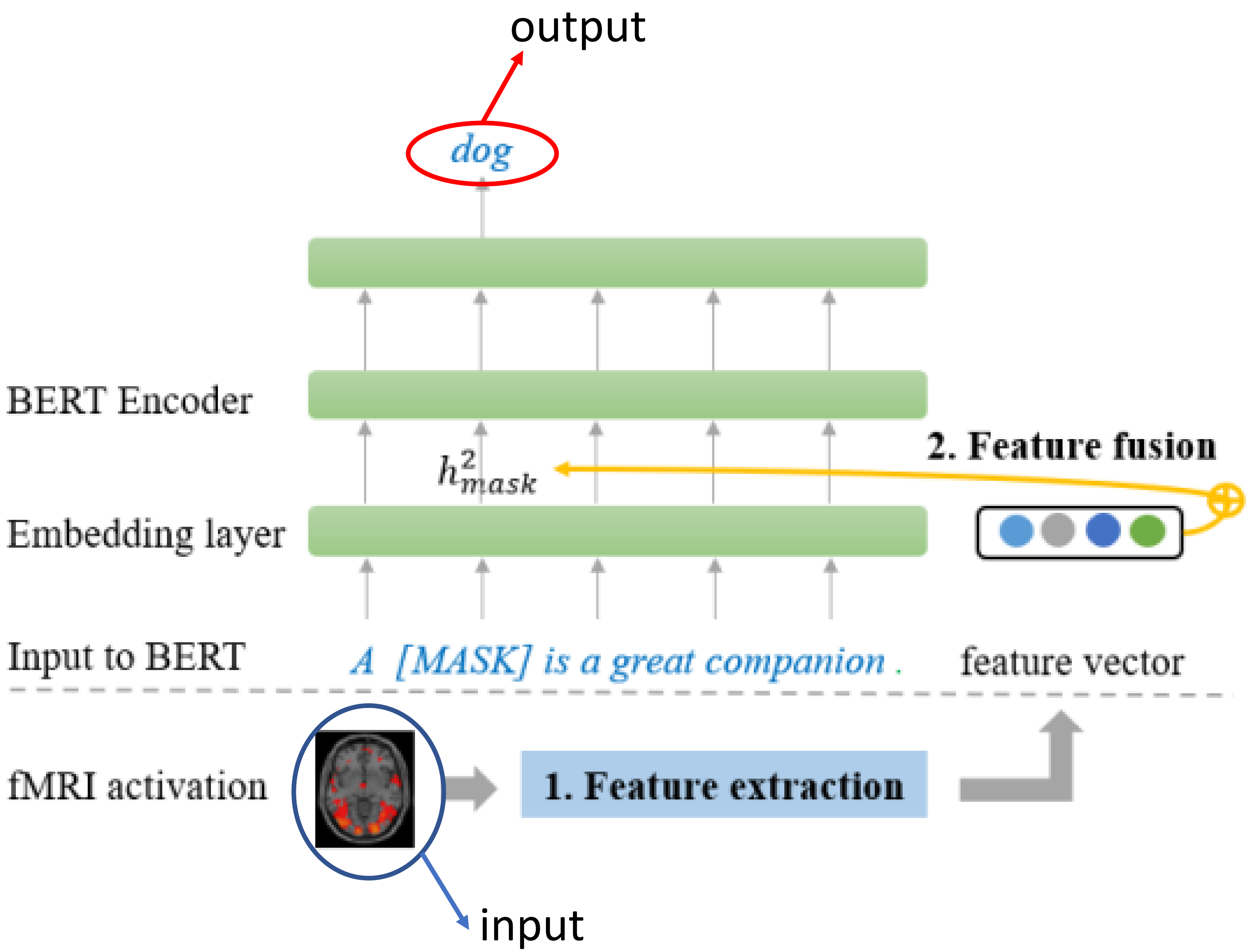


Language Task



Smith et al., 2011, Wang et al. 2019

Linguistic Decoding



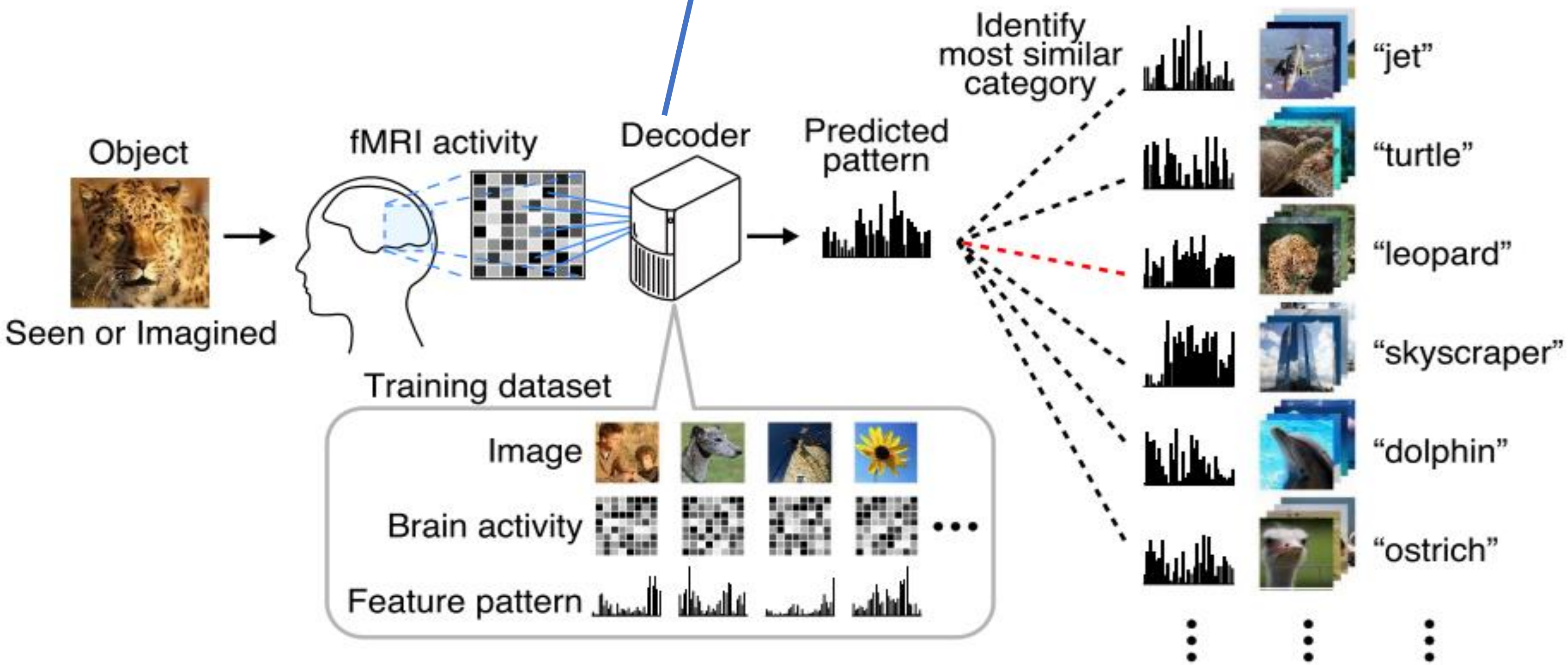
Zou et al., 2022

Outline

- Introduction to Brain Decoding
- Decoding models
  - Linear Models
  - Non-Linear Models (including DNNs)
  - Evaluation Metrics
- Language
  - Periera et al. 2018, Gauthier et al. 2019, Huth et al. 2023, Oota et al. 2022

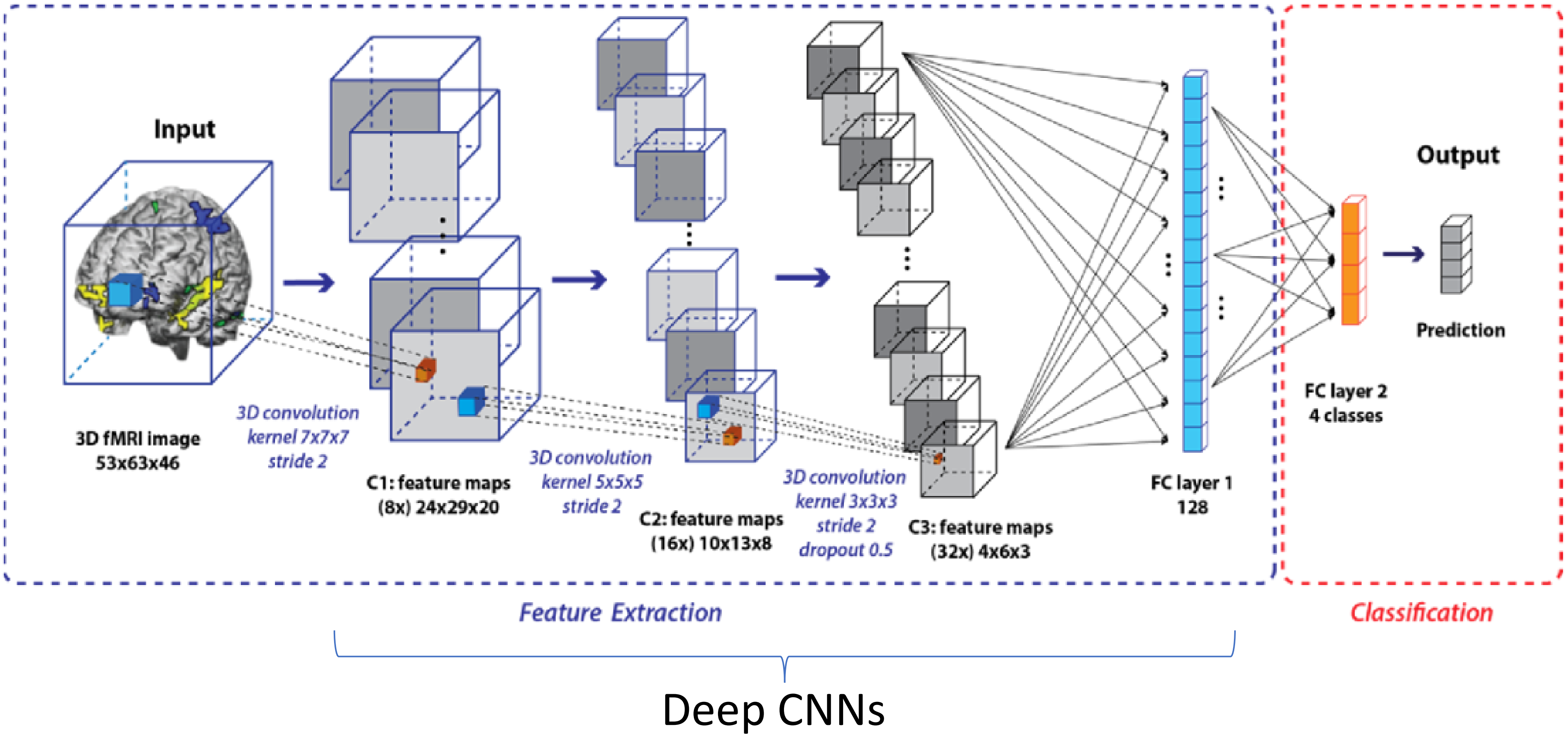
Linear Decoder Models

Stimulus Representation ← Ridge / Logistic Regression → Stimulus Classification



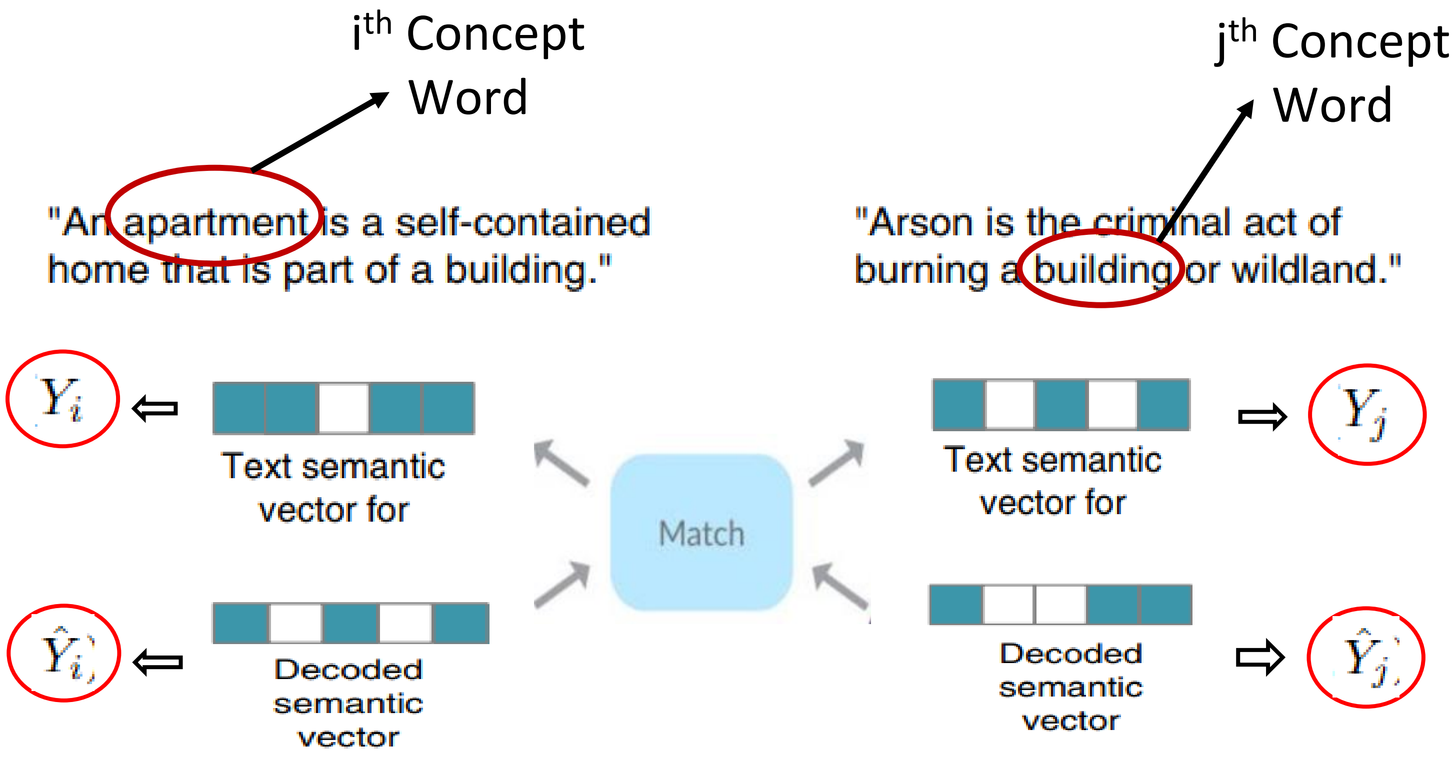
Horikawa et al. 2018

Non-Linear Decoder



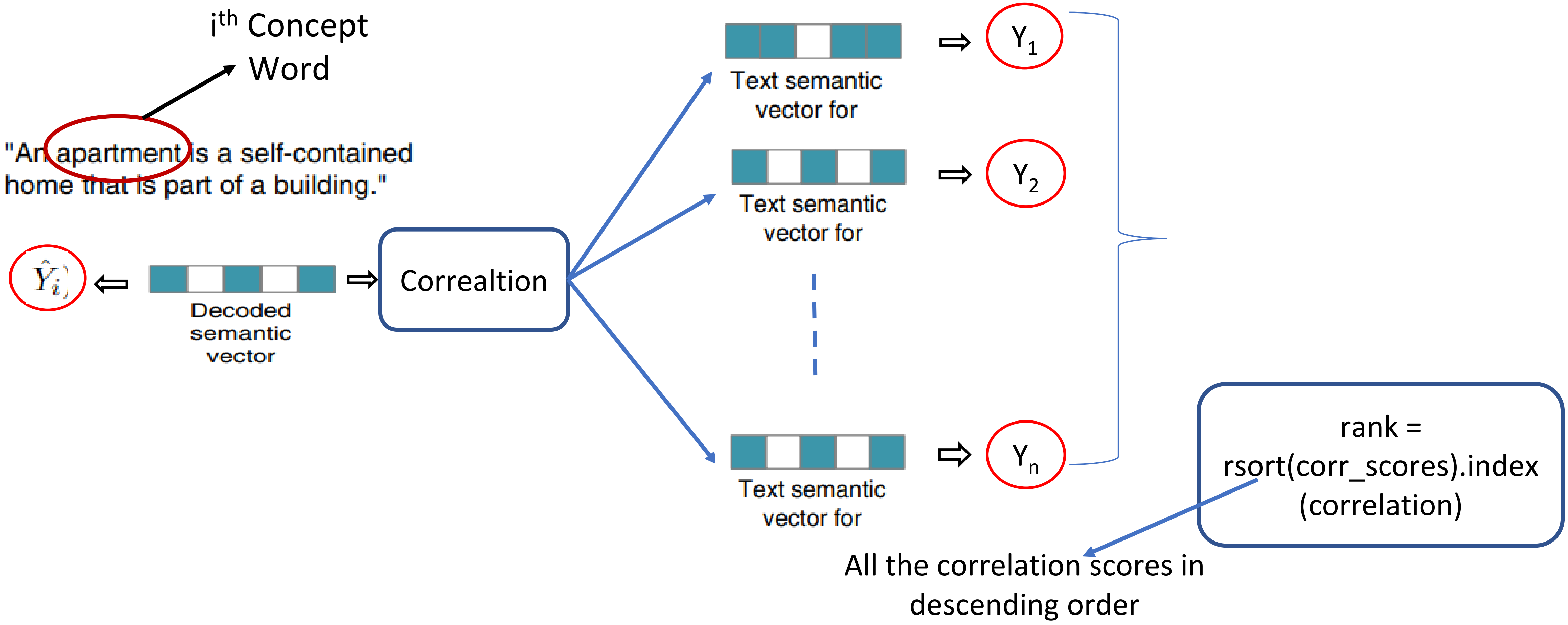
Vu et al. 2018

Evaluating Decoding Models: Pairwise Accuracy



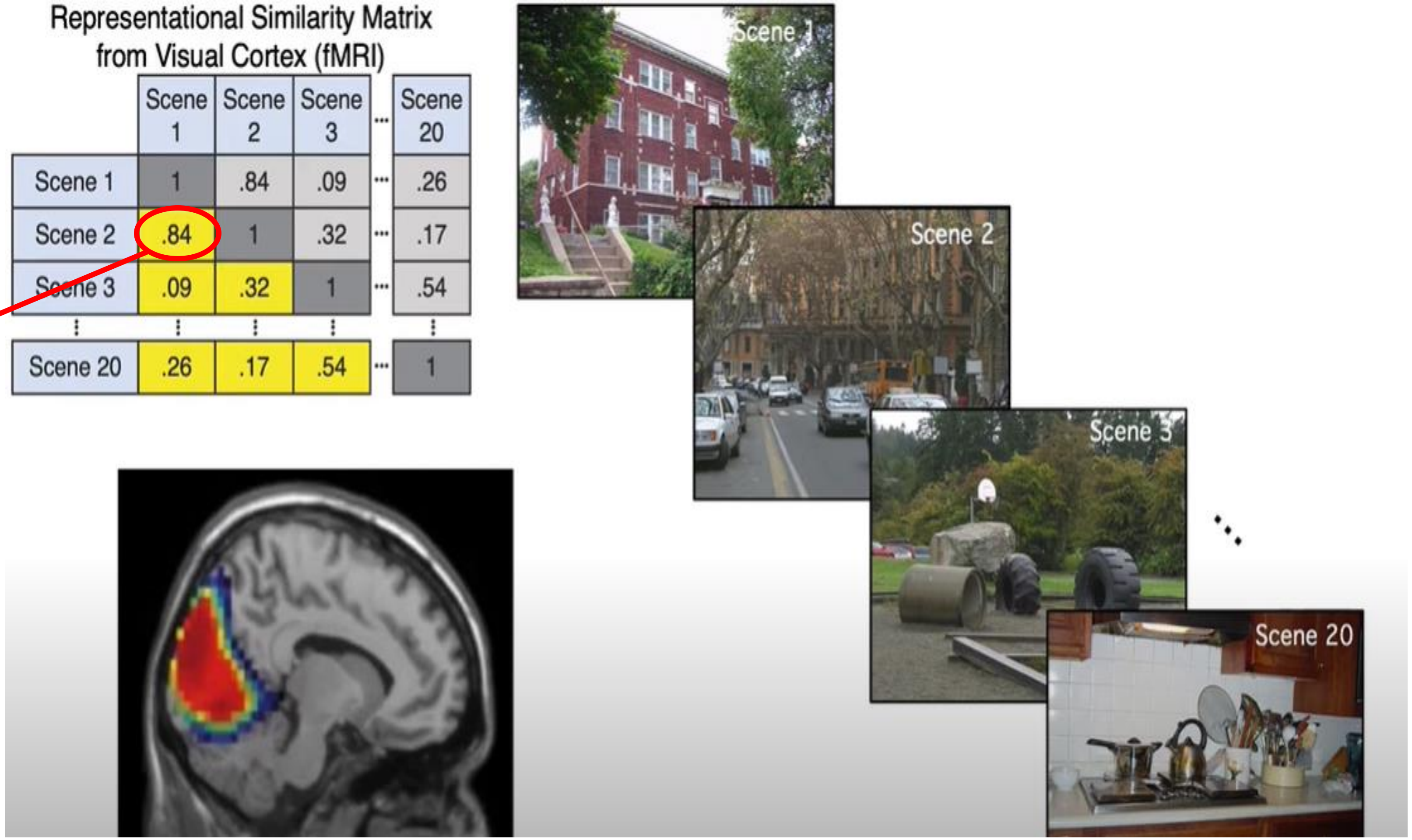
Periera et al. 2018

Evaluating Decoding Models: Rank Accuracy



Periera et al. 2018

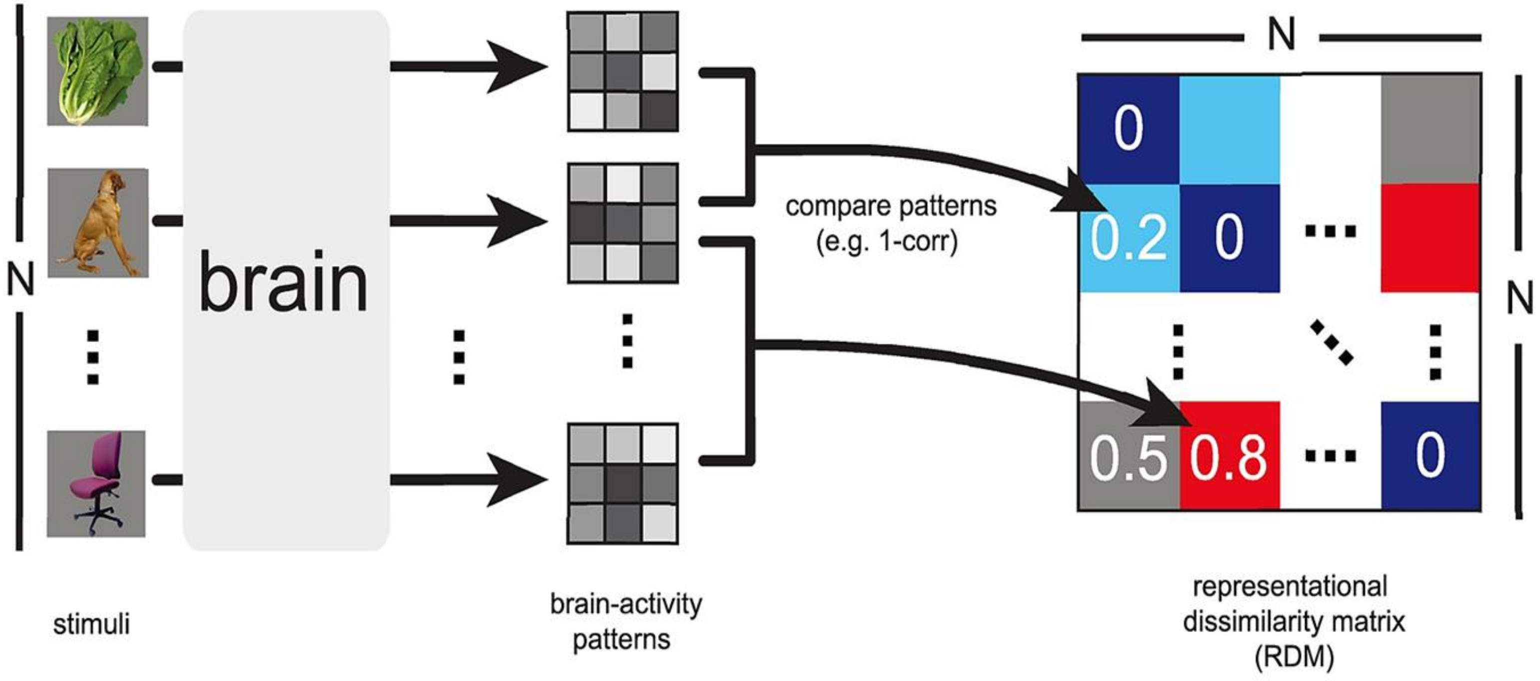
Representational Similarity Matrix (RSM)



corr(Scene1, Scen2)

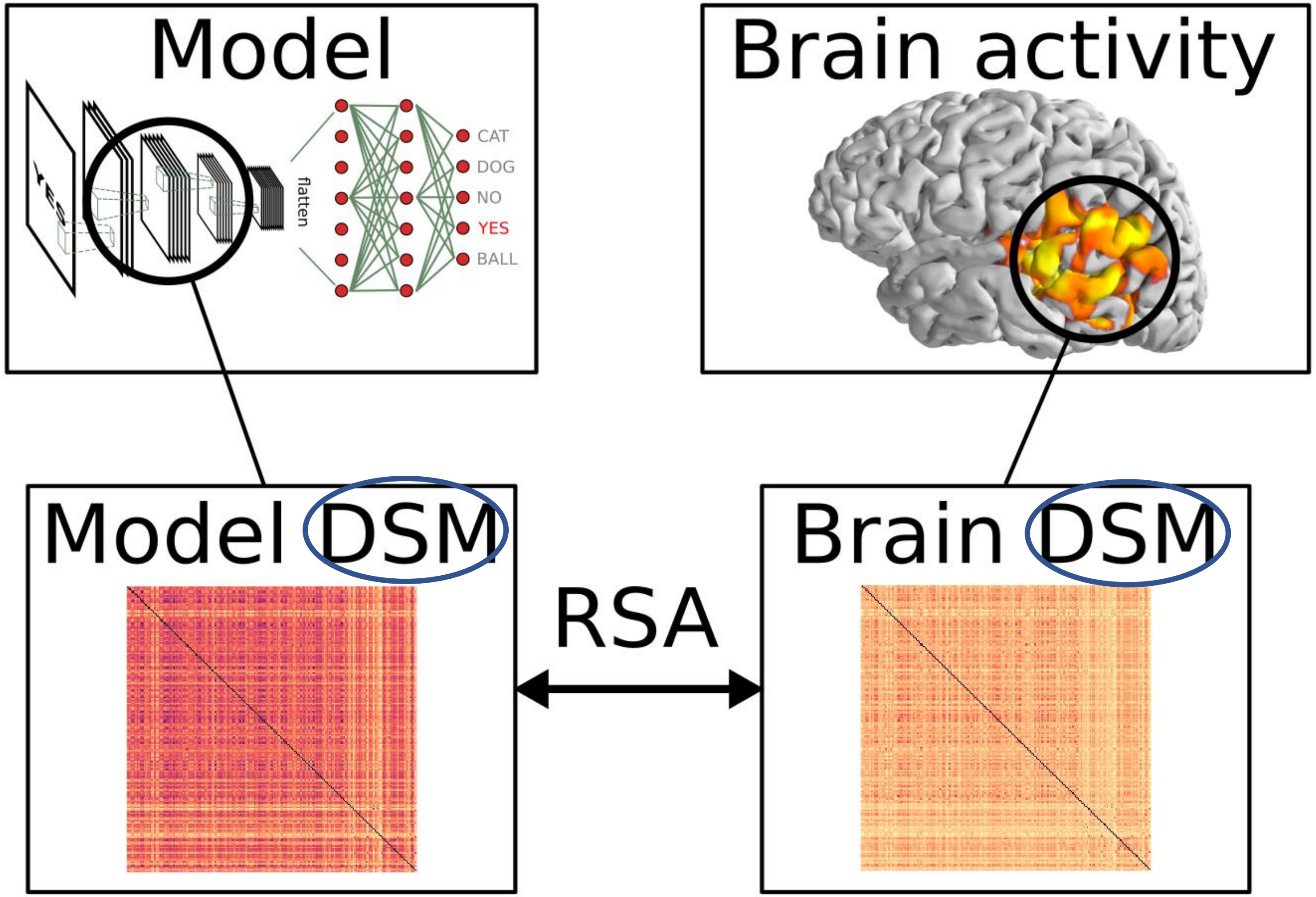
Moussa et al. 2012

Representational Dissimilarity Matrix (RDM)



Hamed et al. 2014

Representation Similarity Analysis



DSM = RDM

Kriegeskorte et al. 2018

Outline

- Introduction to Brain Decoding
- Decoding models
  - Linear Models
  - Non-Linear Models (including DNNs)
- Language
  - Periera et al. 2018, Gauthier et al. 2019, Huth et al. 2023, Oota et al. 2022

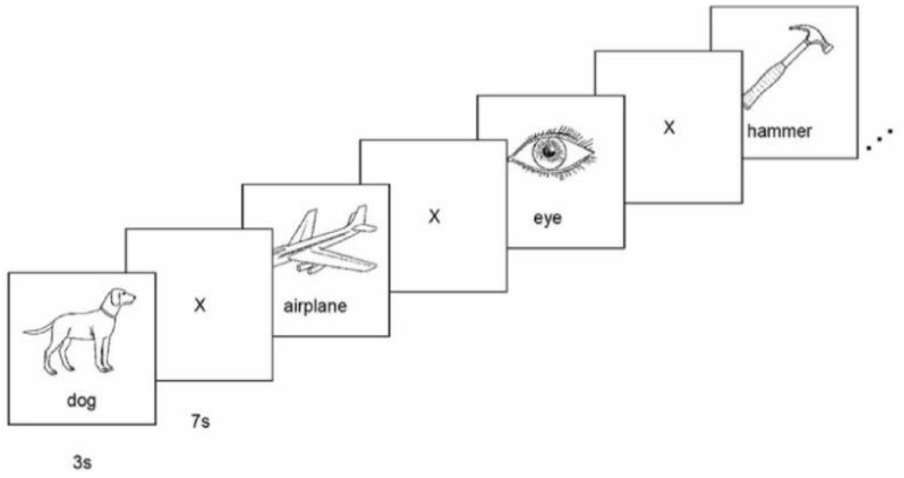
Linguistic Brain Decoding

- Toward Word-level Universal Brain Decoder
- Does injecting linguistic structure into language models lead to better alignment with brain recordings?
- Multi-view and Cross-view Decoding

Periera et al. 2018, Gauthier et al. 2019, Huth et al. 2023, Oota et al. 2022

Classical Decoders

- Classical decoding solutions extracting linguistic meaning from imaging data have been largely limited to
  - concrete nouns,
  - using similar stimuli for training and testing,
  - small number of semantic categories.

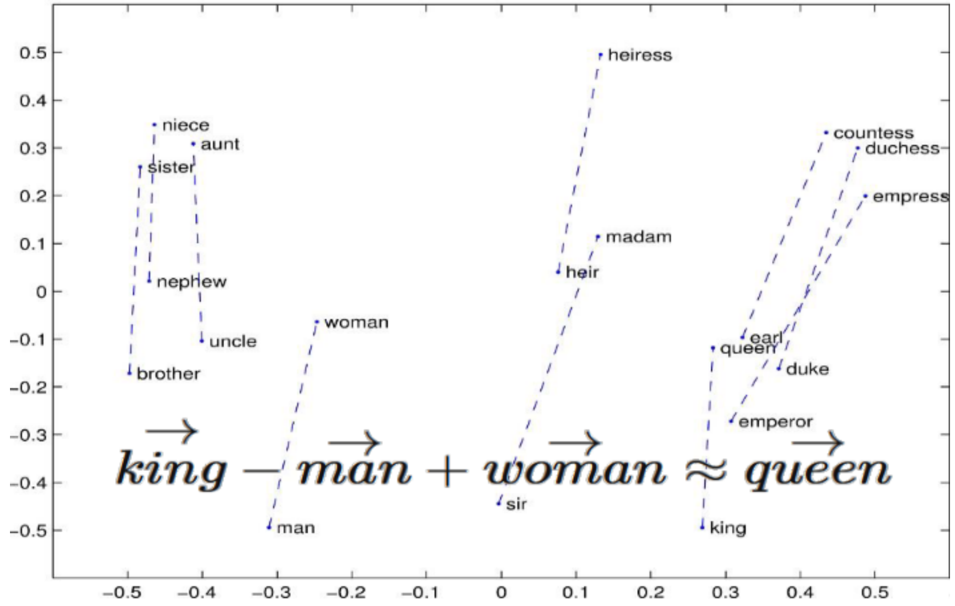


Category	Exemplar 1	Exemplar 2
animals	bear	cat
body parts	arm	eye
buildings	apartment	barn
building parts	arch	chimney
clothing	coat	dress
furniture	bed	chair
insects	ant	bee
kitchen utensils	bottle	cup
man made objects	bell	key
tools	chisel	hammer
vegetables	carrot	celery
vehicles	airplane	bicycle

Mitchell et al. 2008

Toward a universal decoder

- Presented a new approach for building a brain decoding system:
  - words and sentences are represented as vectors in a semantic space constructed from massive text corpora.
  - wide variety of both concrete and abstract topics from two separate datasets.
  - subject reads naturalistic linguistic stimuli on potentially any topic, including abstract ideas (ex., pleasure, justice, love, etc).



GloVE

Pennington et al. 2014

Pereira et al. 2018

Dataset Details (Experiment-1)

Concept
Word

Bird

1. The bird flew around the cage.
2. The nest was just big enough for the bird.
3. The only bird she can see is the parrot.
4. The bird poked its head out of the hatch.
5. The bird holds the worm in its beak.
6. The bird preened itself for mating.

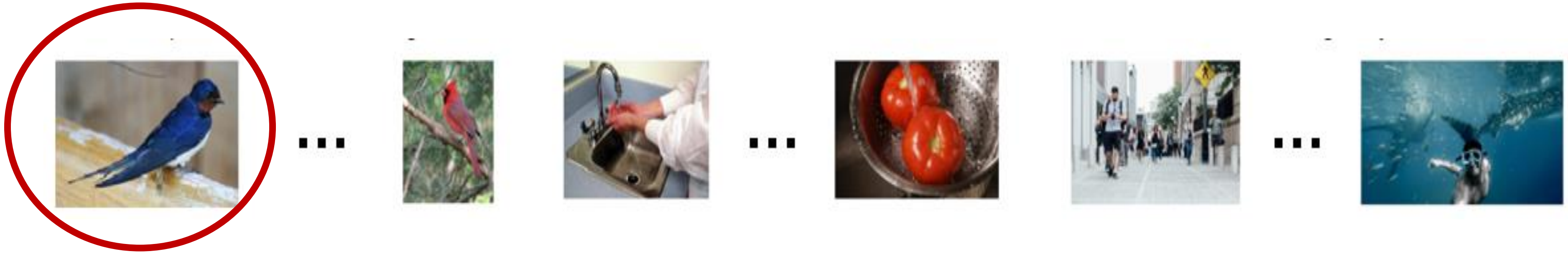
Wash

1. To make the counter sterile, wash it.
2. The dishwasher can wash all the dishes.
3. He likes to wash himself with bar soap.
4. She felt clean after she could wash herself.
5. You have to wash your laundry beforehand.
6. The maid was asked to wash the floor.

Unaware

1. She was unaware of how oblivious he really was.
2. She was unaware of her status.
3. Unprejudiced and unaware, she went full throttle.
4. Unaware of current issues, he is a terrible candidate.
5. He was unaware of how uninterested she was.
6. He was unaware of the gravity of the situation.

⇨ Concept + Sentence View



⇨ Concept + Picture View

Bird	Wash	Unaware
Nest, Flock, Bird, Mating, Beak, Winged	Clean, Shower, Sink, Wash, Soap, Laundry	Unprepared, Unprotected, Unaware, Inexperienced, Unwilling, Unconcerned

⇨ Concept + Wordcloud View

Periera et al. 2018

Dataset Details (Experiment-1)

- 180 Concepts
  - 128 nouns
  - 22 verbs
  - 29 adjectives
  - 1 function word
- 16 subjects
- AAL atlas (180 regions)
- Gordon atlas (333 regions)

Periera et al. 2018

Dataset Details (Experiments 2 and 3)

Topic Concept Topic

Experiment 2:

Musical instruments (clarinet)

A clarinet is a woodwind musical instrument.
It is a long black tube with a flare at the bottom.
The player chooses notes by pressing keys and holes.
The clarinet is used both in jazz and classical music.

Musical instruments (accordion)

An accordion is a portable musical instrument with two keyboards. One keyboard is used for individual notes, the other for chords. Accordions produce sound with bellow that blow air through reeds. An accordionist plays both keyboards while opening and closing the bellows.

Musical instruments (piano)

The piano is a popular musical instrument played by means of a keyboard. Pressing a piano key causes a felt-tipped hammer to hit a vibrating steel string. The piano has an enormous note range, and pedals to change the sound quality. The piano repertoire is large, and famous pianists can give solo concerts.

Experiment 3:

Skiing (passage 1)

I hesitantly skied down the steep trail that my buddies convinced me to try. I made a bad turn, and I found myself tumbling down. I finally came to a stop at a flat part of the slope. My skis were nowhere to be found, and my poles were lodged in a snow drift up the hill.

Skiing (passage 2)

A major strength of professional skiers is how they use ski poles. Proper use of ski poles improves their balance and adds flair to their skiing. It minimizes the need for upper body movements to regain lost balance while skiing.

Skiing (passage 3)

New ski designs and stiffer boots let skiers turn more quickly. But faster and tighter turns increase the twisting force on the legs. This has led to more injuries, particularly to ligaments in the skier's knee.

Gambling (passage 1)

When I decided to start playing cards, things went from bad to worse. Gambling was something I had to do, and I had already spent close to $10,000 doing it. My friends were sick of watching me gamble my savings away. The hardest part was the horror of leaving a casino after losing money I did not have.

Gambling (passage 2)

Good data on the social and economic effects of legalized gambling are hard to come by. Some studies indicate that having a casino nearby makes gambling problems more likely. Gambling may also be associated with personal bankruptcies and marriage problems.

Gambling (passage 3)

Over the past generation, there has been a dramatic expansion of legalized gambling. Most states have instituted lotteries, and many have casinos as well. Gambling has become a very big but controversial business.

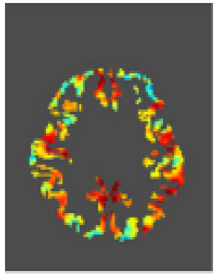
Periera et al. 2018

Informative Voxel Selection

Stimulus: Apartment

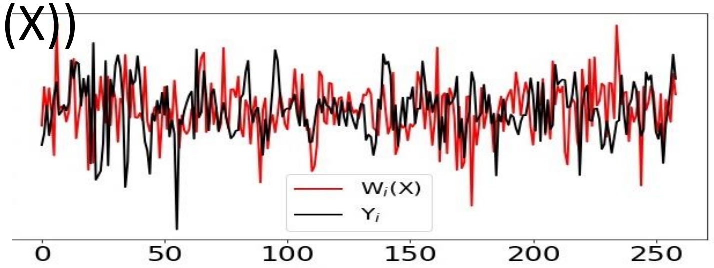
Present

3D Image



Correlation across feature dimensions

Pearson Correlation (R) = Corr(Y, W(X))



X Input

W Ridge Regression

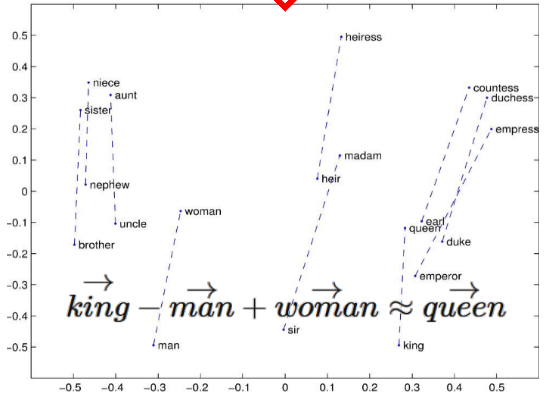
Y Output

Voxel + 26 neighbors in 3D

Text semantic vector for "apartment"

Stimulus: Apartment

Present

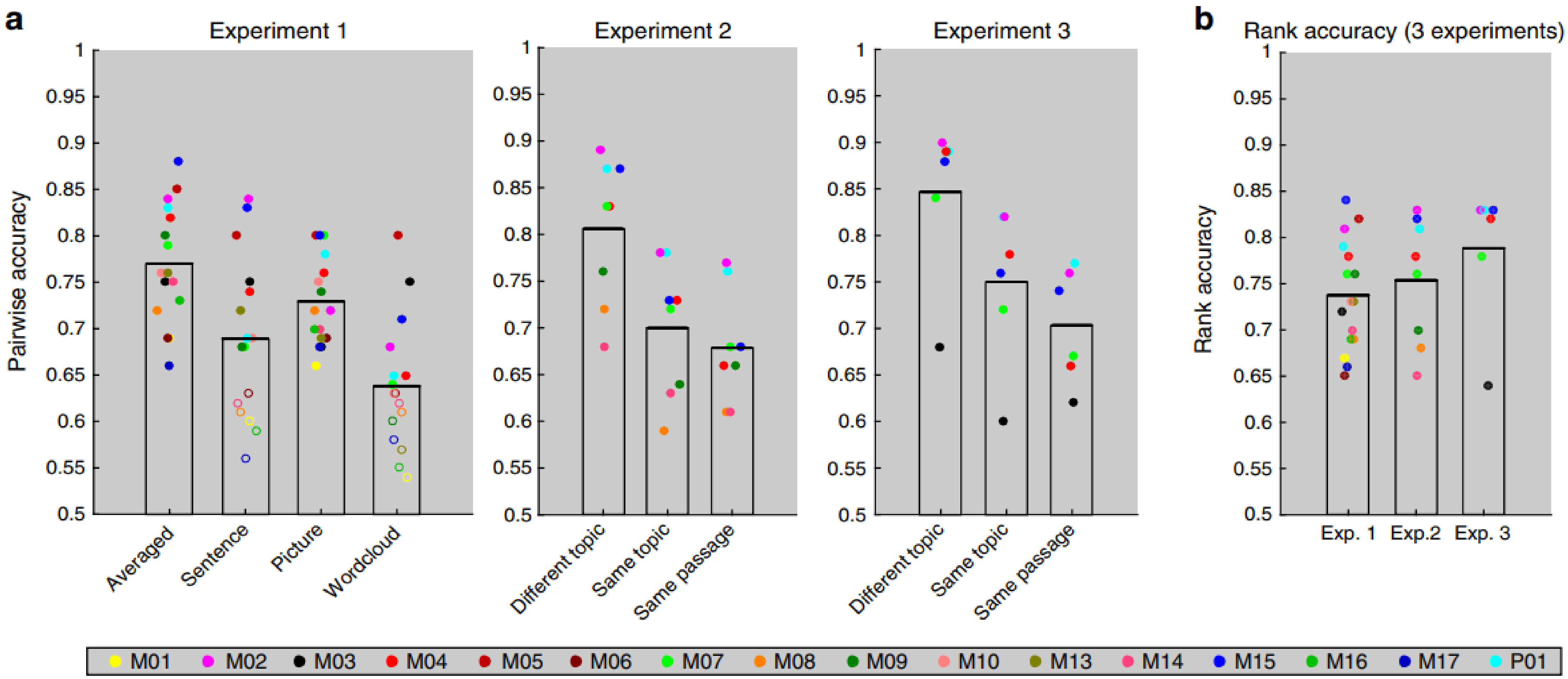


GloVE

$V_1 - R_1$
$V_2 - R_2$
....
$V_n - R_3$

Select 5000 voxels based on top-5000 correlation scores

Pairwise and Rankwise Results

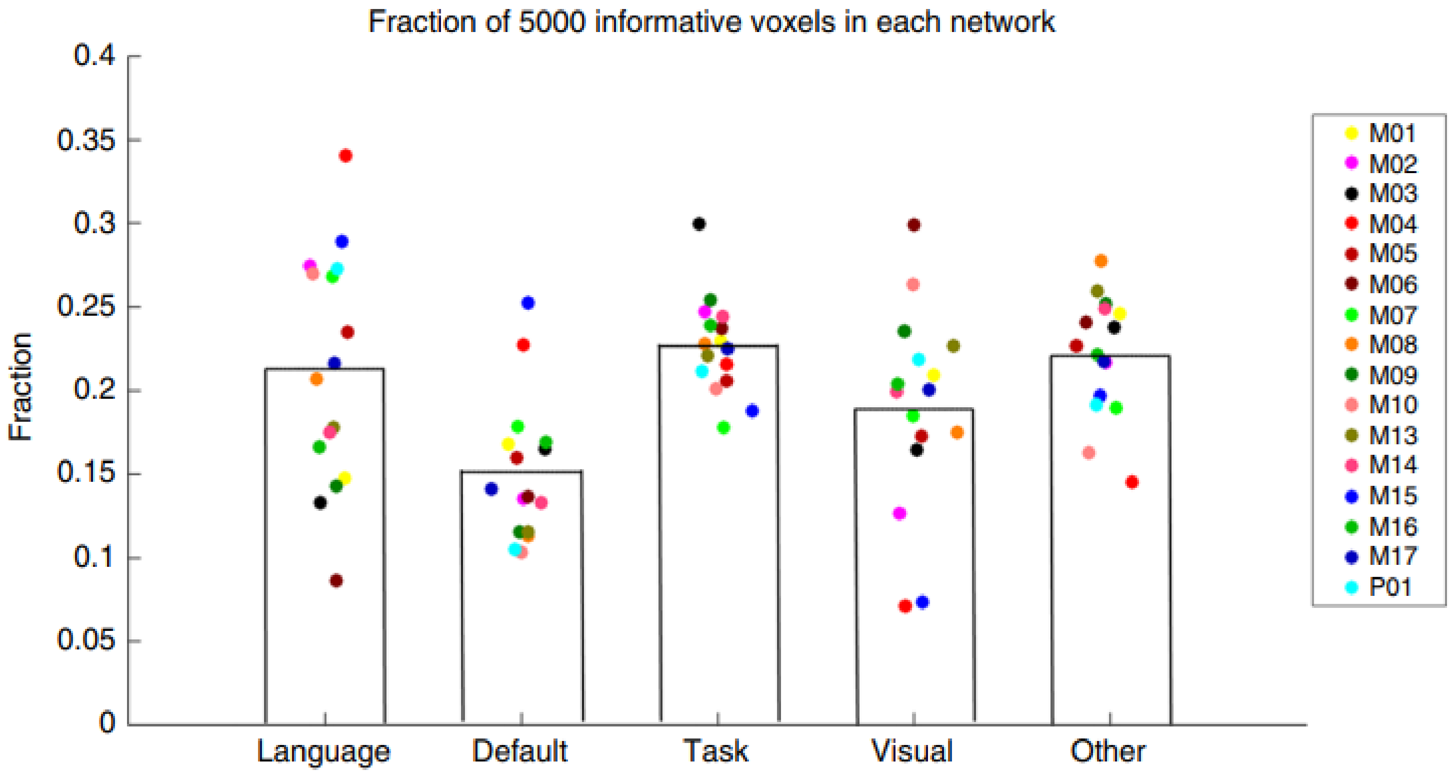


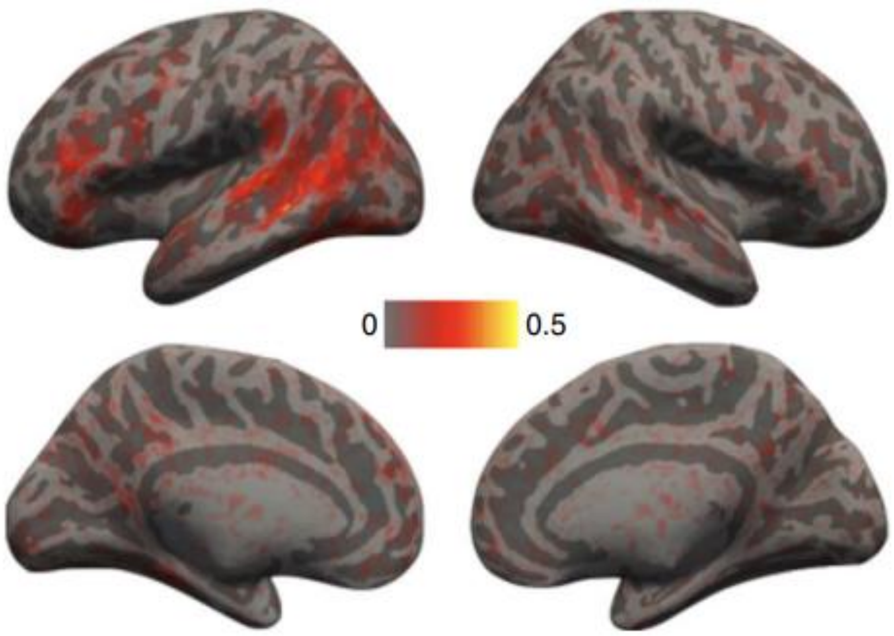
Decoder built from Expt 1 could distinguish sentences at all levels of granularity

Universal Decoder!

Periera et al. 2018

Distribution of Informative Voxels





5000 informative voxels are roughly evenly distributed among the four networks
Overall, LN contains a relatively higher proportion of informative voxels, compared to its size!

Brain activation patterns consistent across 16 Ss

Periera et al. 2018

Insights

- Presented a viable approach for building a universal decoder, capable of extracting a representation of mental content from linguistic materials.
- The semantic resolution of brain-based decoding of mental content will continue to improve rapidly
  - given the progress in the development of distributed semantic representations

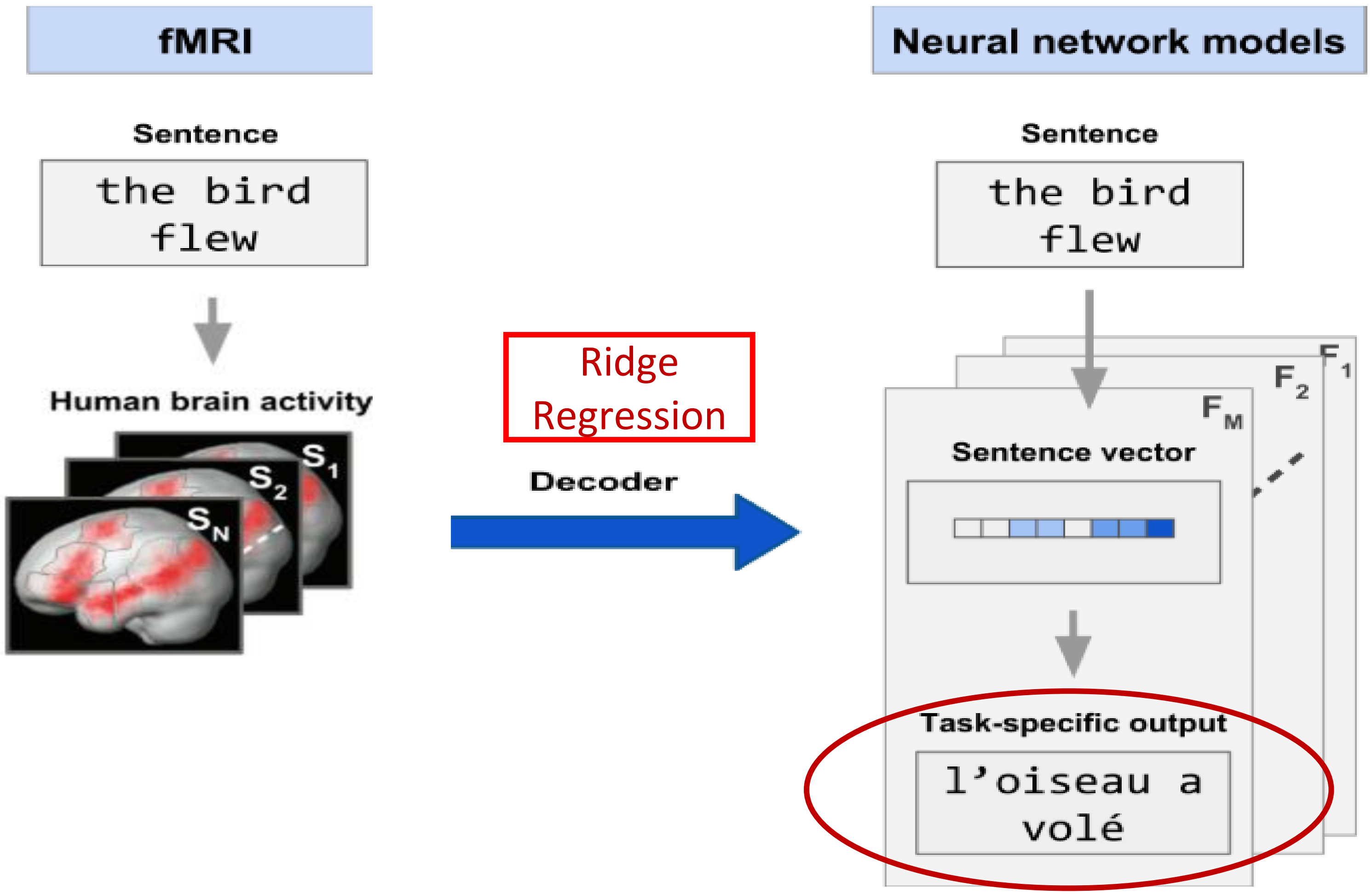
Periera et al. 2018

Linguistic Brain Decoding

- Toward Word-level Universal Brain Decoder
- Linking artificial and human neural representations of language
- Multi-view and Cross-view Decoding

Periera et al. 2018, Gauthier et al. 2019, Huth et al. 2023, Oota et al. 2022

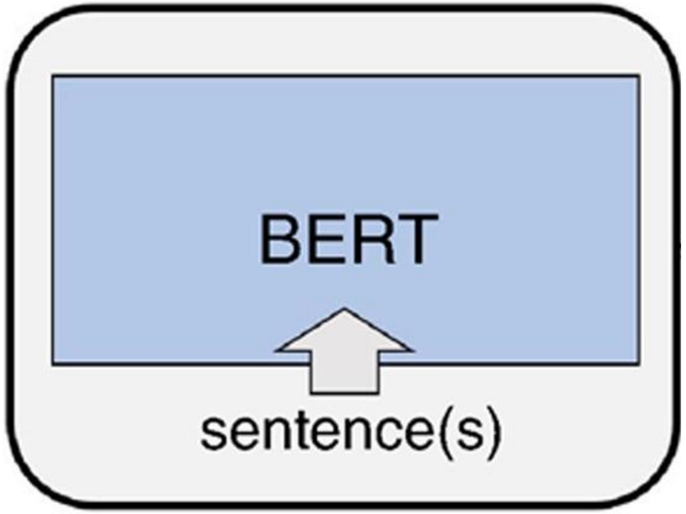
Linking artificial and human neural representations of language



- Evaluate the link between human brain activity and neural network models as the models are optimized for different tasks.
- To investigate why these mappings are successful?
- Uncovering the parallel representational contents shared between human brains and neural networks

Gauthier et al. 2019

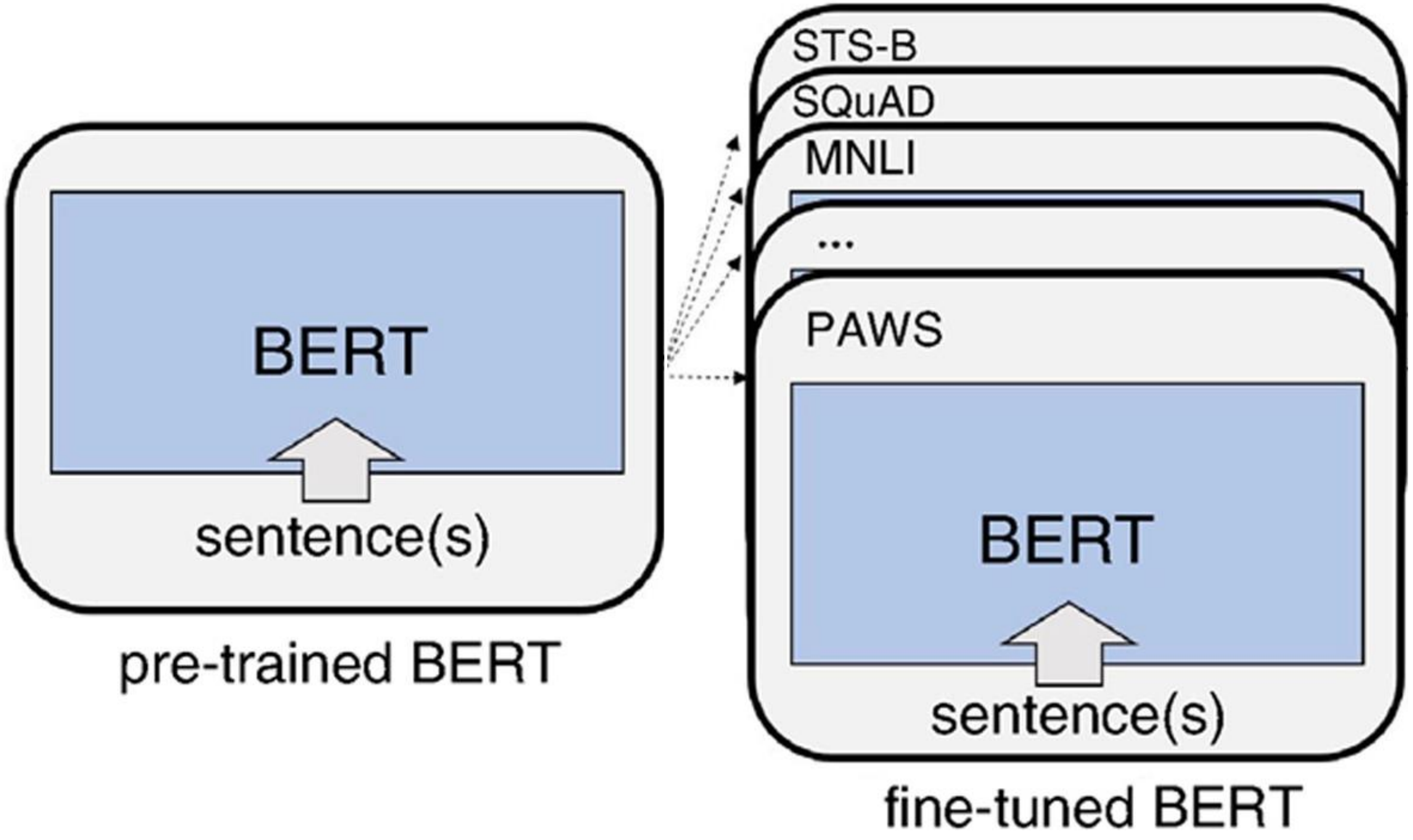
Pretrained vs. Task-specific language models



pre-trained BERT

Devlin et al. 2019

Pretrained vs. Task-specific language models



Natural Language Understaning Tasks

- Paraphrase
- Question Answering
- Sentiment Analysis
- Natural Language Inference

Article: Endangered Species Act
Paragraph: " *. . . Other legislation followed, including the Migratory Bird Conservation Act of 1929, a 1937 treaty prohibiting the hunting of right and gray whales, and the Bald Eagle Protection Act of 1940. These later laws had a low cost to society—the species were relatively rare—and little opposition was raised.*"

Question 1: "*Which laws faced significant opposition?*"
Plausible Answer: *later laws*

Question 2: "*What was the name of the 1937 treaty?*"
Plausible Answer: *Bald Eagle Protection Act*

Squad-2.0: Question Answering

Devlin et al. 2019, Bowon et al. 2020

Custom Tasks

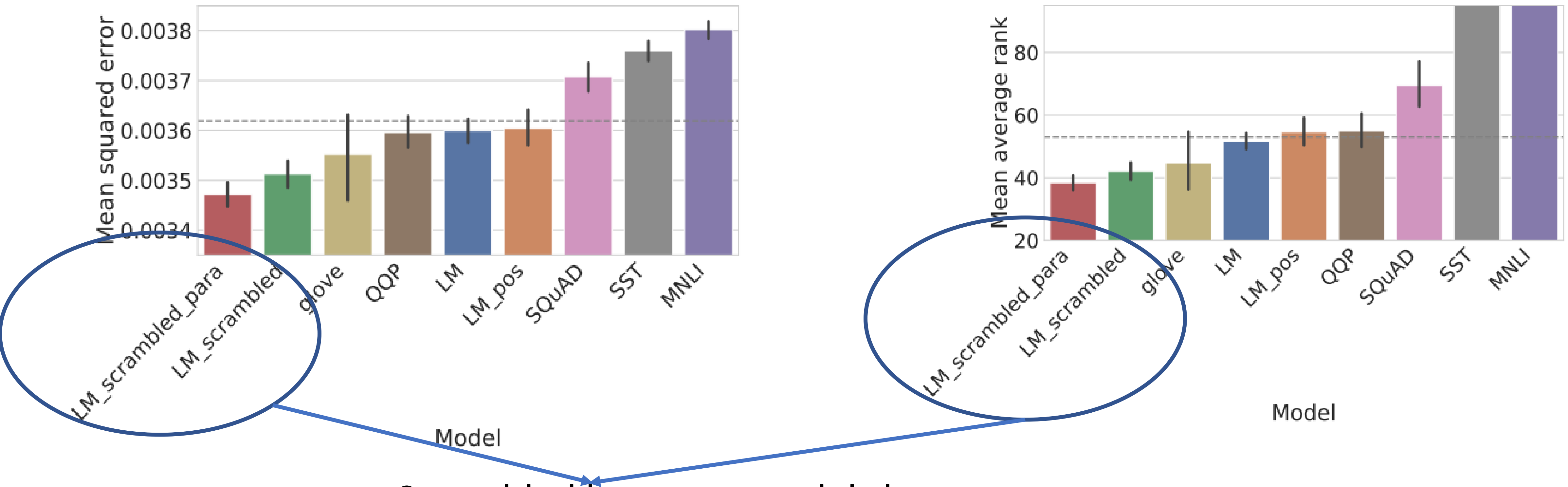
- Scrambled language modeling:
  - LM-scrambled: deals with sentence inputs where words are shuffled within sentences
  - LM-scrambled-para, uses inputs where words are shuffled within their containing paragraphs in the corpus.

Fingers are used for grasping, writing, grooming and other activities.
grasping are used for Fingers, grooming, writing and other activities.

This is Los Angeles. And it's the height of summer. In a small bungalow off of La Cienega, Clara serves homemade chili and chips in red plastic bowls -- wine in blue plastic.
This is Los Angeles. And the height it's of summer. In a bungalow off small of La Cienega, Clara serves homemade chili and chips in red plastic bowls -- wine in blue plastic.

Gauthier et al. 2019

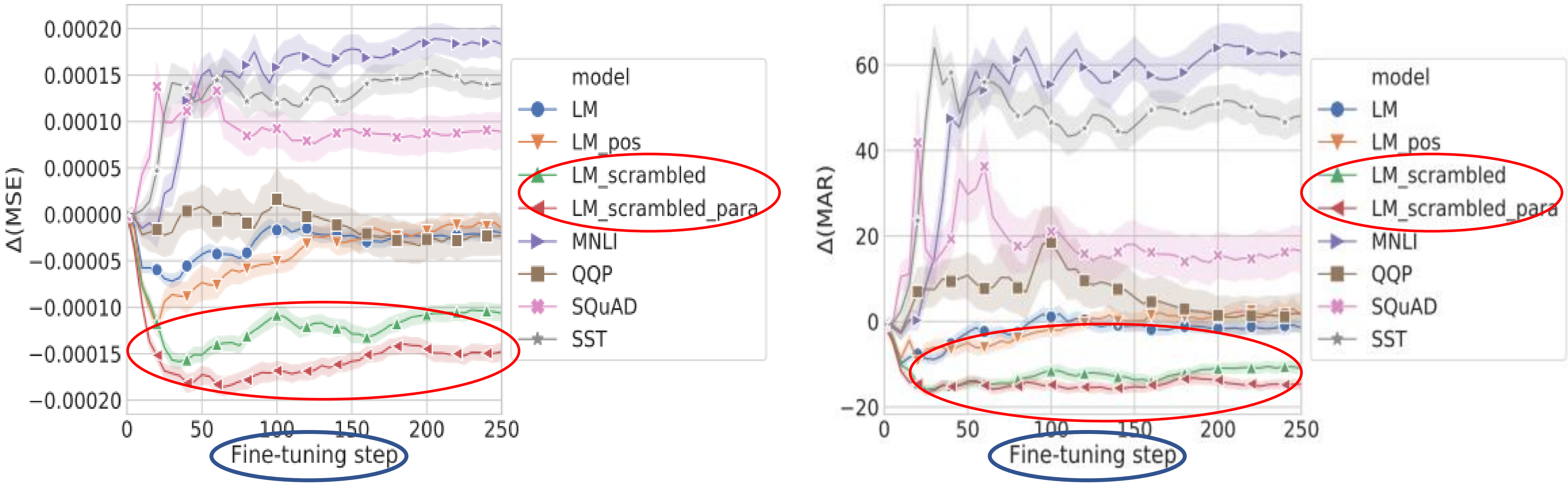
Brain decoding performance



Scrambled language models have shown better performance!!

Gauthier et al. 2019

Brain decoding performance trajectories over fine-tuning time



Gauthier et al. 2019

Summary

- Set of scrambled language modeling tasks which best match the structure of brain activations among the models tested.
  - models optimized for LM- scrambled and LM-scrambled-para — the models which improve in brain decoding performance

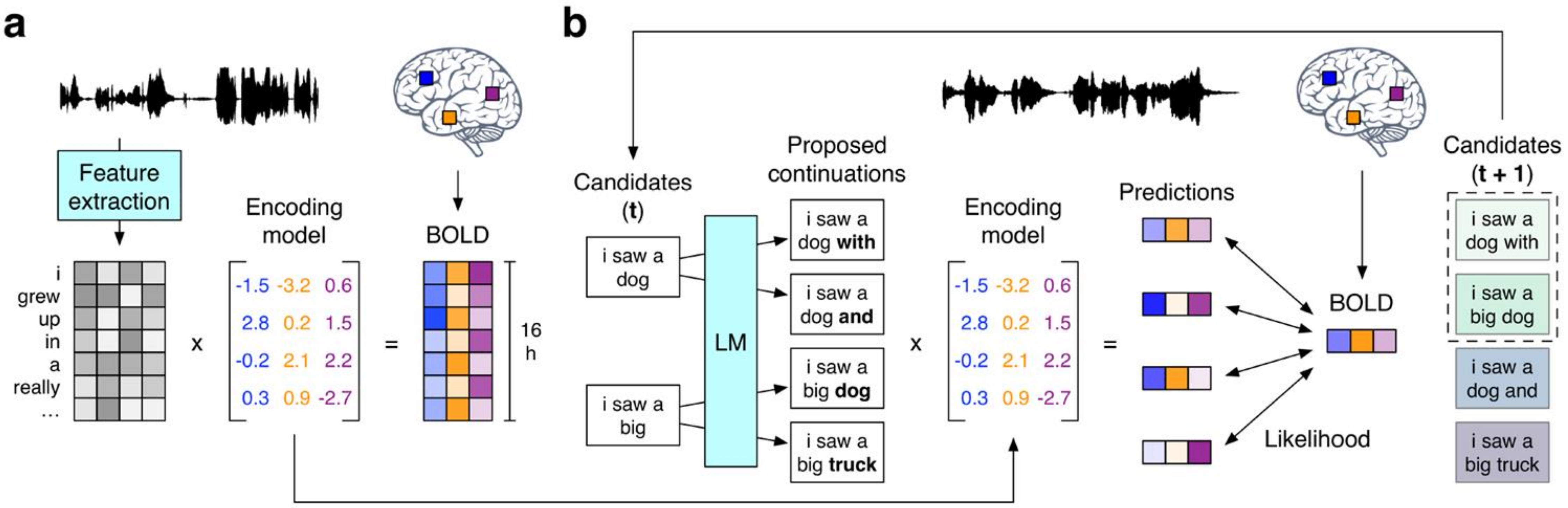
Gauthier et al. 2019

Linguistic Brain Decoding

- Toward Word-level Universal Brain Decoder
- Linking artificial and human neural representations of language (contd)
- Multi-view and Cross-view Decoding

Periera et al. 2018, Gauthier et al. 2019, Huth et al. 2023, Oota et al. 2022

Continuous Language Decoder



Tang, LaBel, Jain & Huth (2023)

Continuous Language Decoder

C

Actual stimulus	Decoded stimulus
i got up from the air mattress and pressed my face against the glass of the bedroom window expecting to see eyes staring back at me but instead finding only darkness	i just continued to walk up to the window and open the glass i stood on my toes and peered out i didn't see anything and looked up again i saw nothing
i didn't know whether to scream cry or run away instead i said leave me alone i don't need your help adam disappeared and i cleaned up alone crying	started to scream and cry and then she just said i told you to leave me alone you can't hurt me i'm sorry and then he stormed off i thought he had left i started to cry
that night i went upstairs to what had been our bedroom and not knowing what else to do i turned out the lights and lay down on the floor	we got back to my dorm room i had no idea where my bed was i just assumed i would sleep on it but instead i lay down on the floor
i don't have my driver's license yet and i just jumped out right when i needed to and she says well why don't you come back to my house and i'll give you a ride i say ok	she is not ready she has not even started to learn to drive yet i had to push her out of the car i said we will take her home now and she agreed

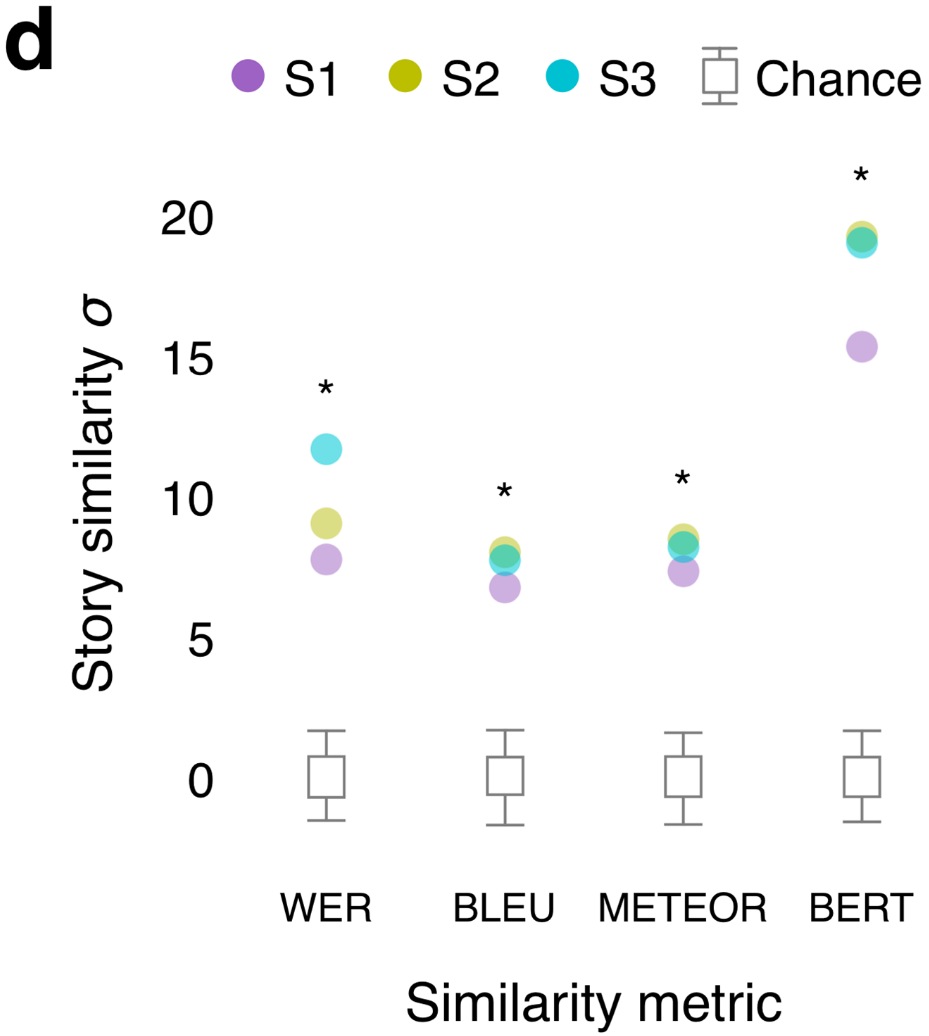
Exact

Gist

Error

Tang, LaBel, Jain & Huth (2023)

Continuous Language Decoder



Tang, LaBel, Jain & Huth (2023)

Summary

- Continuous language representations of semantic meaning can be decoded (reconstructed) from non-invasive brain recordings (fMRI),
- Given novel brain recordings, decoder generates intelligible word sequences that recover the meaning of perceived speech, imagined speech, and even silent videos, demonstrating that a single language decoder can be applied to a range of semantic tasks.
- Exciting possibility enabling future multipurpose brain-computer interfaces!

Tang, LaBel, Jain & Huth (2023)

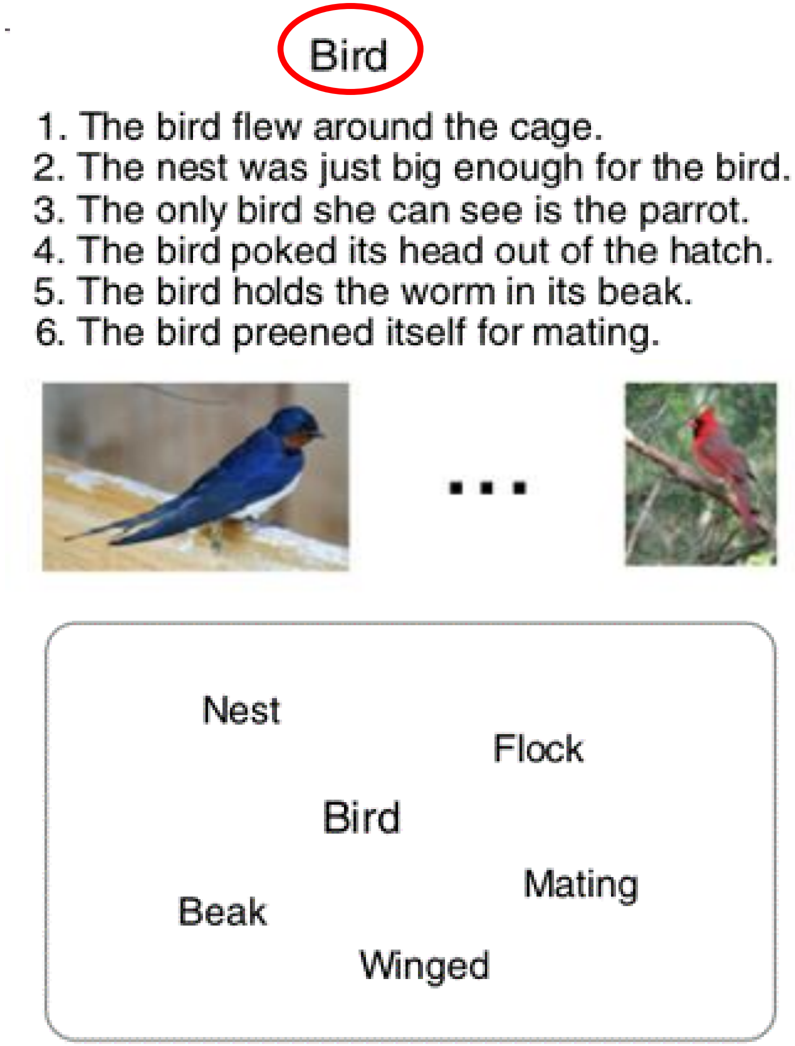
Linguistic Brain Decoding

- Toward Word-level Universal Brain Decoder
- Linking artificial and human neural representations of language
- Multi-view and Cross-view Decoding

Periera et al. 2018, Gauthier et al. 2019, Huth et al. 2023, Oota et al. 2022

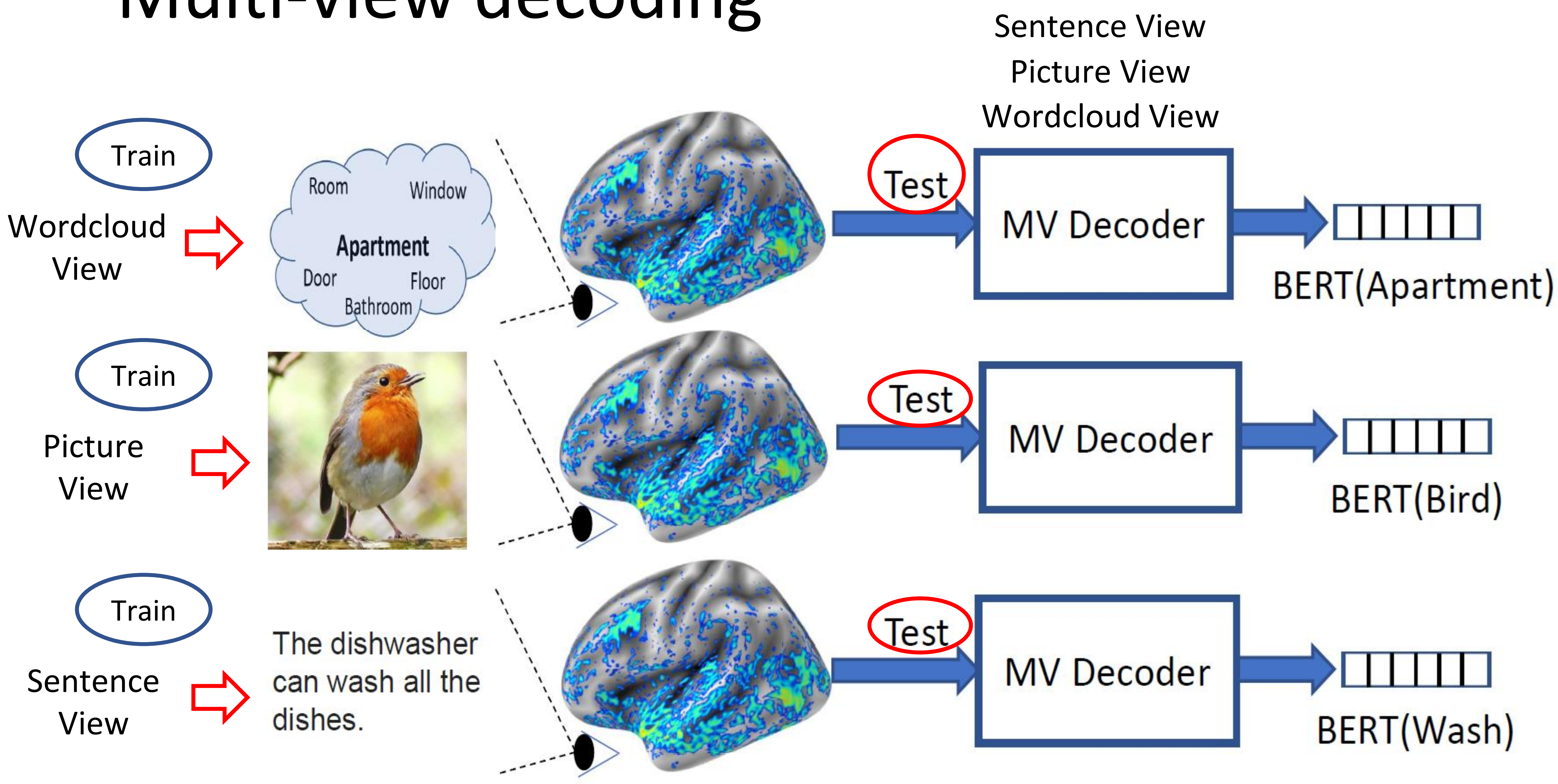
Multi-view and Cross-ViewBrain Decoding

- Human brains have the unique capability of language acquisition:
  - the process of learning the language
  - understand the meaning of concepts from multiple modalities such as images, text, speech, and videos.
- Prior works focus on single-view brain decoding using traditional feature engineering.
- However, how the brain captures the meaning of linguistic stimuli across multiple views is still a critical open question in neuroscience.
- Consider three different views of the concept bird:
  - (1) sentence using the target word,
  - (2) picture presented with the target word label, and
  - (3) word cloud containing the target word along with other semantically related words.
- Earlier works have explored which of these three different views provides richer information to understand the concept.



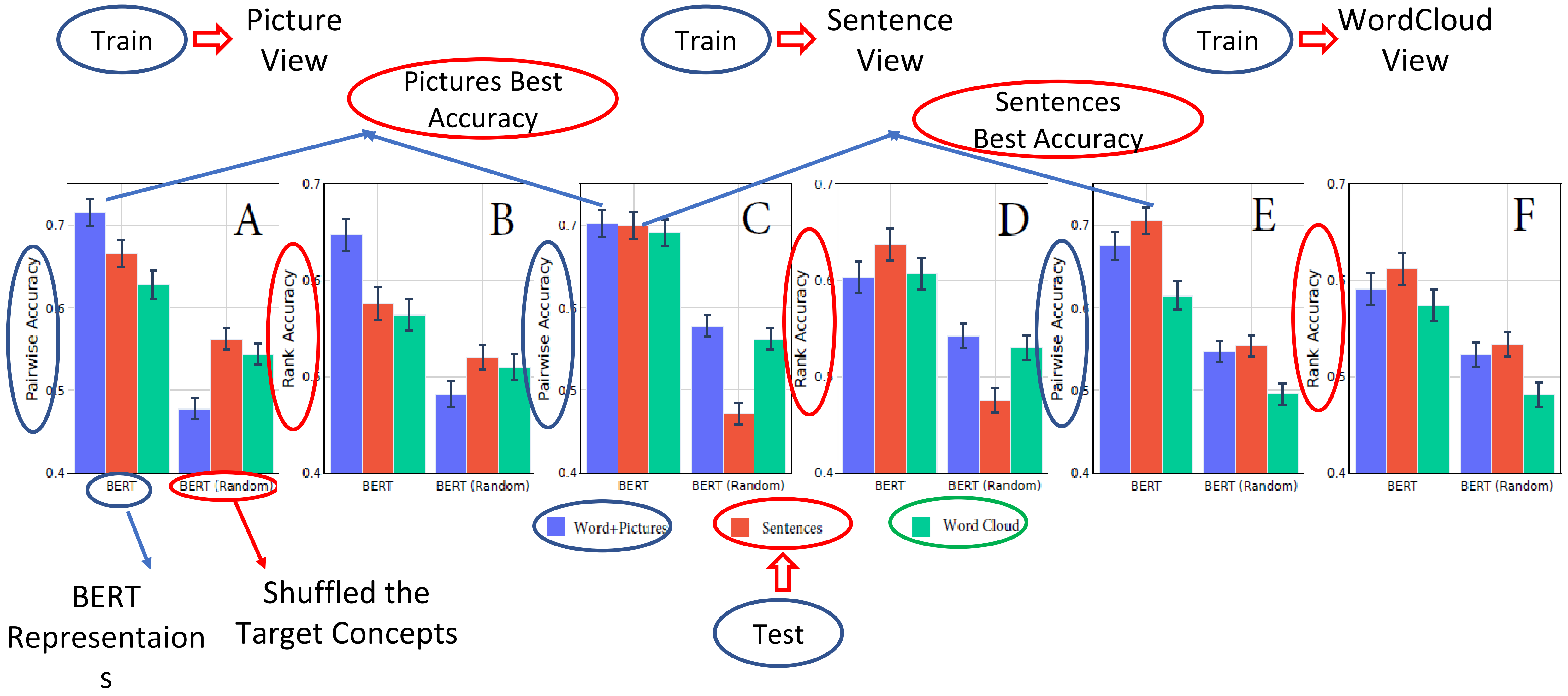
Oota et al. 2022

Multi-view decoding



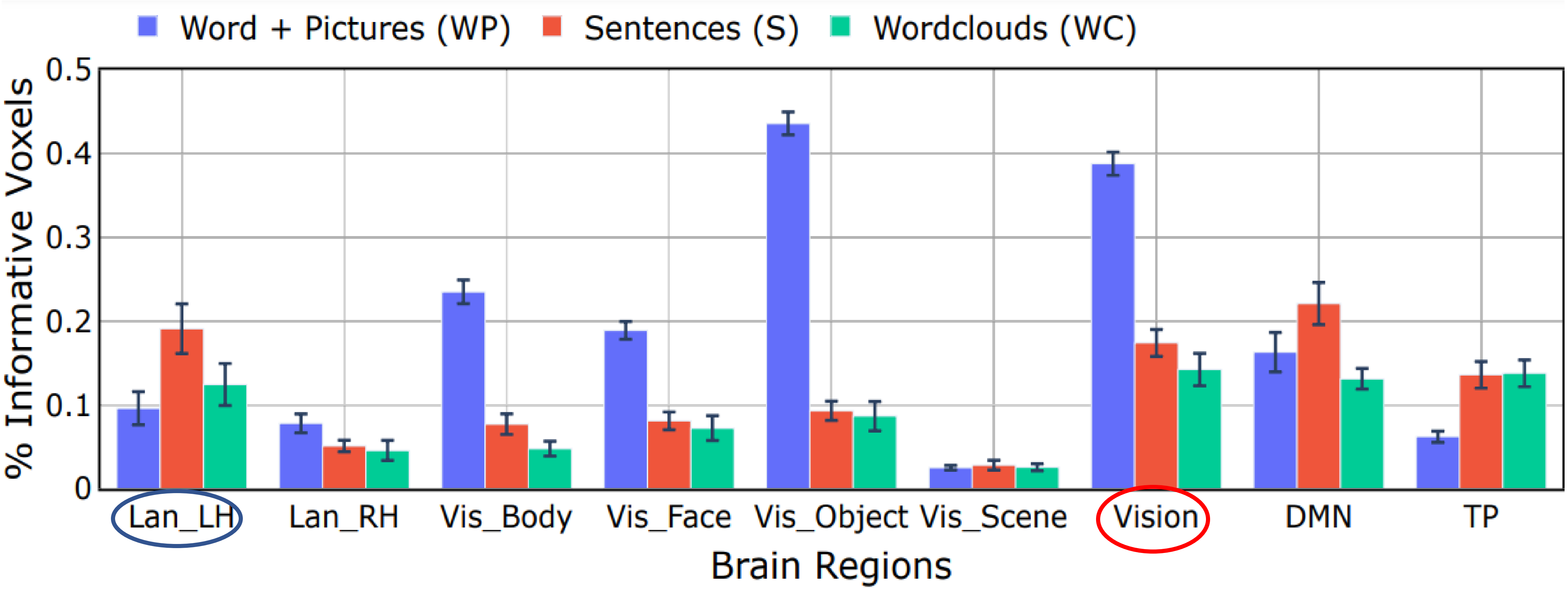
Oota et al. 2022

Multi-view decoding results



Oota et al. 2022

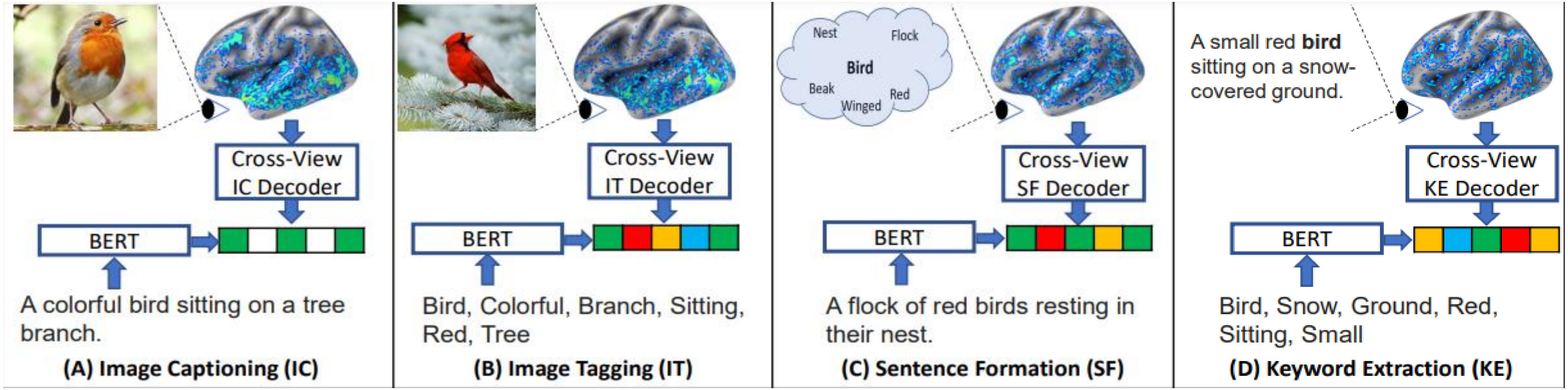
Distribution of Informative Voxels



Oota et al. 2022

Cross-view Decoding

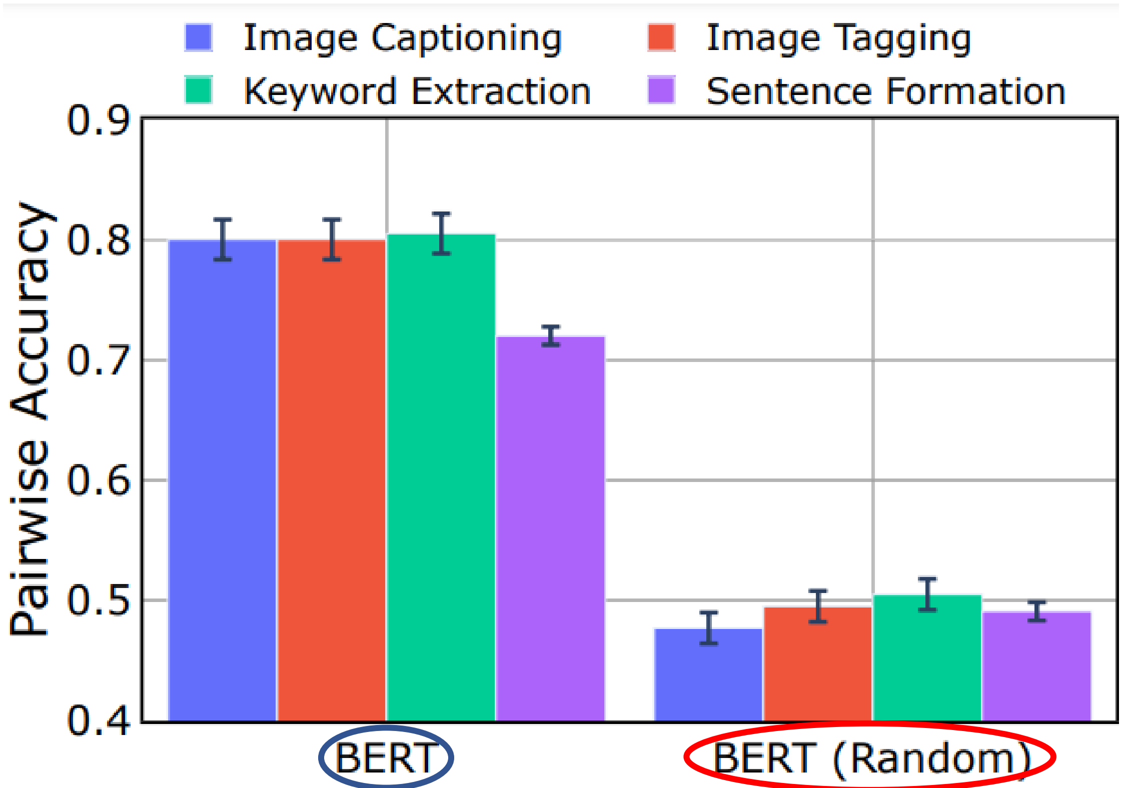
Train ⇨ Picture View | Train ⇨ Picture View | Train ⇨ Wordcloud View | Train ⇨ Sentence View

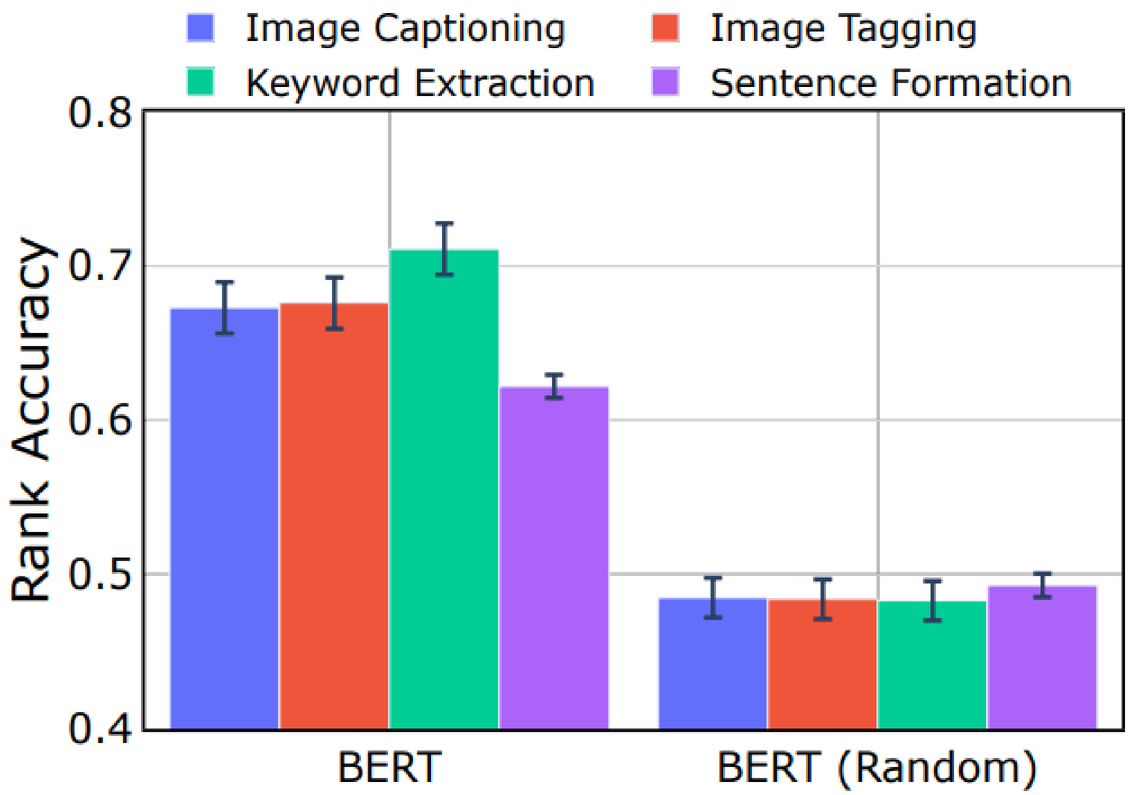


Test ⇨ Caption | Test ⇨ Visual words | Test ⇨ Sentence | Test ⇨ Keywords

Oota et al. 2022

Cross-view Decoding results





BERT Representaions

Shuffled the Target Concepts

Oota et al. 2022

Summary

- Cross-view and Multi-view decoding tasks establish that the information contained in the brain response is rich and capable of driving multiple downstream tasks.

Oota et al. 2022

Linguistic Brain Decoding

- Toward Word-level Universal Brain Decoder
- Linking artificial and human neural representations of language
- Multi-view and Cross-view Decoding

Periera et al. 2018, Gauthier et al. 2019, Huth et al. 2023, Oota et al. 2022

Agenda

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

References

- Pereira, Francisco, et al. "Toward a universal decoder of linguistic meaning from brain activation." *Nature communications* 9.1 (2018).
- Sun, Jingyuan, et al. "Towards sentence-level brain decoding with distributed representations." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. No. 01. 2019.
- Affolter, Nicolas, et al. "Brain2word: decoding brain activity for language generation." arXiv preprint arXiv:2009.04765 (2020).
- Abdou, Mostafa, et al. "Does injecting linguistic structure into language models lead to better alignment with brain recordings?." arXiv preprint arXiv:2101.12608 (2021).
- Sun, Jingyuan, et al. "Neural encoding and decoding with distributed sentence representations." IEEE Transactions on Neural Networks and Learning Systems 32.2 (2020): 589-603.
- Oota, Subba Reddy, et al. "Cross-view Brain Decoding." arXiv preprint arXiv:2204.09564 (2022).
- Gauthier, Jon, and Roger Levy. "Linking artificial and human neural representations of language." EMNLP/IJCNLP (1). 2019.
- Shen, Guohua, et al. "Deep image reconstruction from human brain activity." PLoS computational biology 15.1 (2019): e1006633.
- Beliy, Roman, et al. "From voxels to pixels and back: Self-supervision in natural-image reconstruction from fMRI." Advances in Neural Information Processing Systems 32 (2019).
- Shen, Guohua, et al. "End-to-end deep image reconstruction from human brain activity." Frontiers in Computational Neuroscience (2019): 21.

References

- Nishimoto, Shinji, et al. "Reconstructing visual experiences from brain activity evoked by natural movies." *Current biology* 21.19 (2011): 1641-1646.
- Anumanchipalli, Gopala K., Josh Chartier, and Edward F. Chang. "Speech synthesis from neural decoding of spoken sentences." Nature 568.7753 (2019): 493-498.
- Schrimpf, Martin, et al. "The neural architecture of language: Integrative modeling converges on predictive processing." Proceedings of the National Academy of Sciences 118.45 (2021): e2105646118.
- Wehbe, Leila, et al. "Simultaneously uncovering the patterns of brain regions involved in different story reading subprocesses." PloS one 9.11 (2014): e112575.

Deep Learning for Brain Encoding and Decoding

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

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

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




Agenda

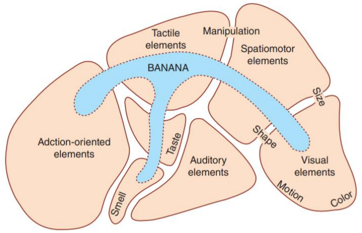
- Introduction to Brain encoding and decoding [30 min]
- Stimulus Representations [1 hour 30 min]
- Coffee break [30 min]
- Deep Learning for Brain Decoding [1 hour 30 min]
- Lunch break [1 hour 15 min]
- **Deep Learning for Brain Encoding [1 hour 30 min]**
- Coffee break [30 min]
- Advanced Methods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

Agenda

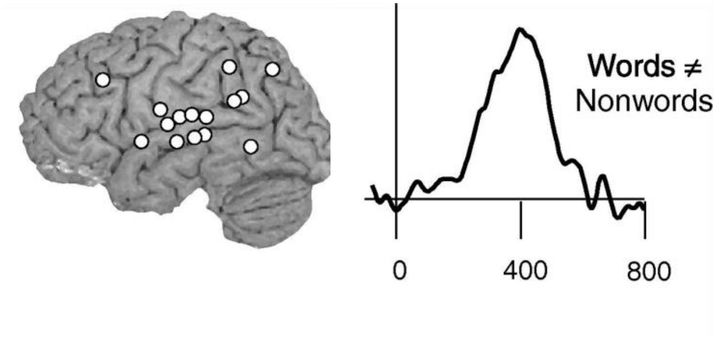
- Introduction to Brain encoding and decoding [30 min]
- Stimulus Representations [1 hour 30 min]
- Coffee break [30 min]
- Deep Learning for Brain Decoding [1 hour 30 min]
- Lunch break [1 hour 15 min]
- **Deep Learning for Brain Encoding [1 hour 30 min]**
  - **Classic findings & common approaches**
  - More recent findings utilizing deep learning
- Coffee break [30 min]
- Advanced Methods [1 hour 15 min]
- Summary and Future Trends [15 min]

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

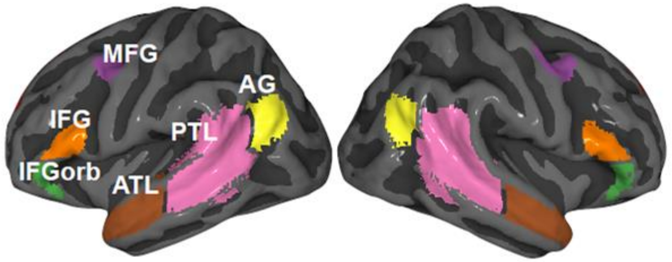
What



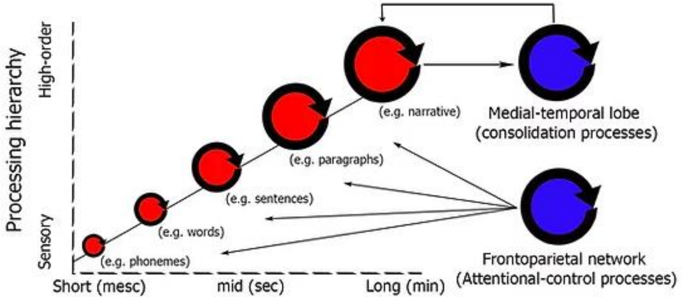
When



Where



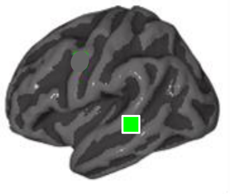
How



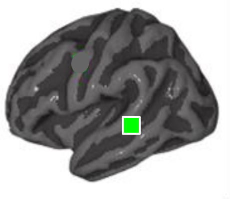
Encoding models have a causal interpretation

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

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



y_{train}



Train: $f(<0,1,\ldots 0>) \approx$

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

Part of Speech: Noun $<0, 1, \ldots 0>$

stimulus properties

sim.

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)

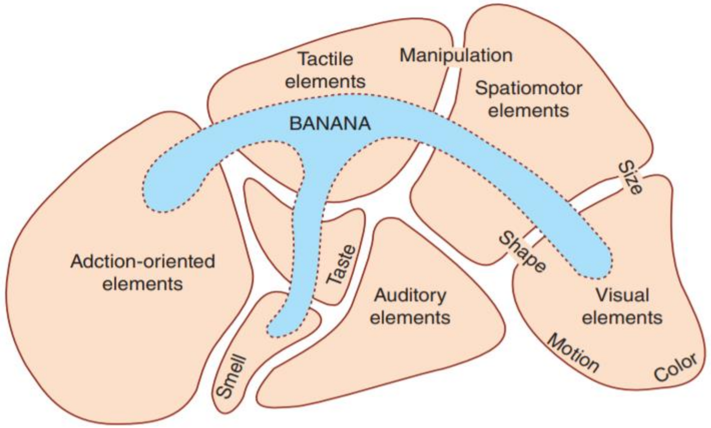


figure from Kemmerer, 2014; adapted from Thompson-Schill et al. 2006

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

Empirical evidence for distributed organization for attributes related to:

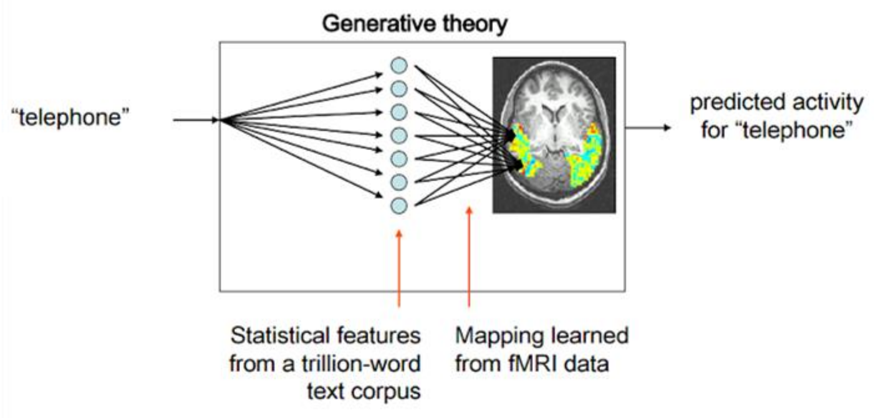
- audition [Kiefer et al., 2008]
- color [Simmons et al., 2007]
- shape [Chao et al., 1999]
- motion [Damasio et al., 1996]
- olfaction and taste [Goldberg, Perfetti, et al., 2006a; Goldberg, Perfetti, et al., 2006b]

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.

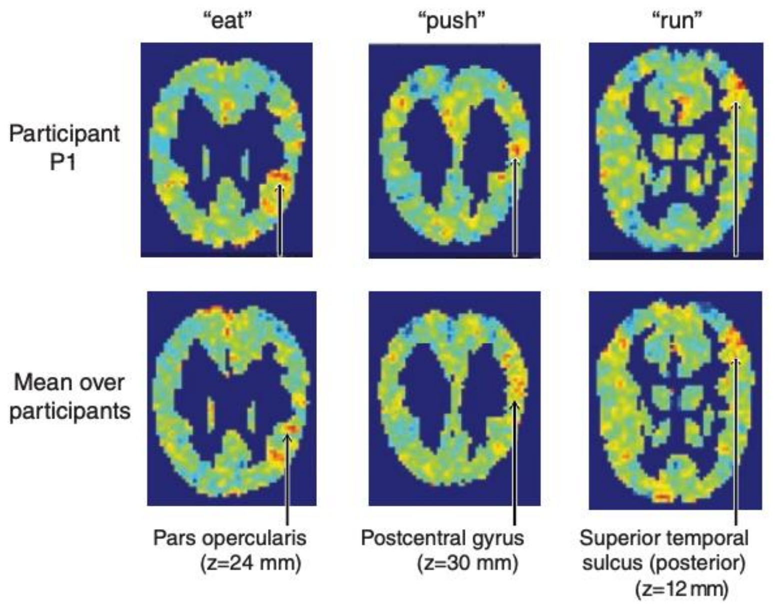
Classic encoding model finding: Language

- 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

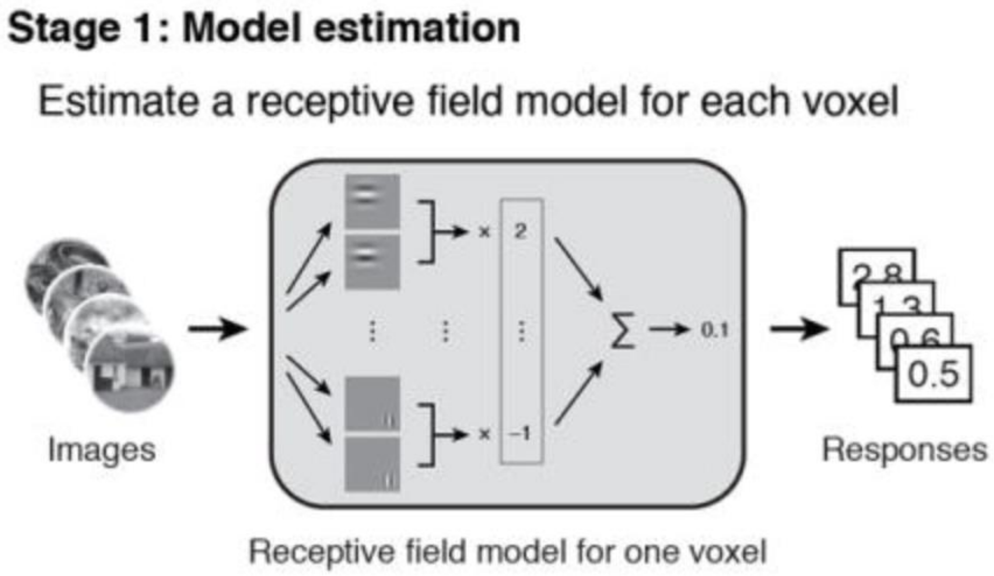
Mitchell, Tom M., Svetlana V. Shinkareva, Andrew Carlson, Kai-Min Chang, Vicente L. Malave, Robert A. Mason, and Marcel Adam Just. "Predicting human brain activity associated with the meanings of nouns." science 320, no. 5880 (2008): 1191-1195.

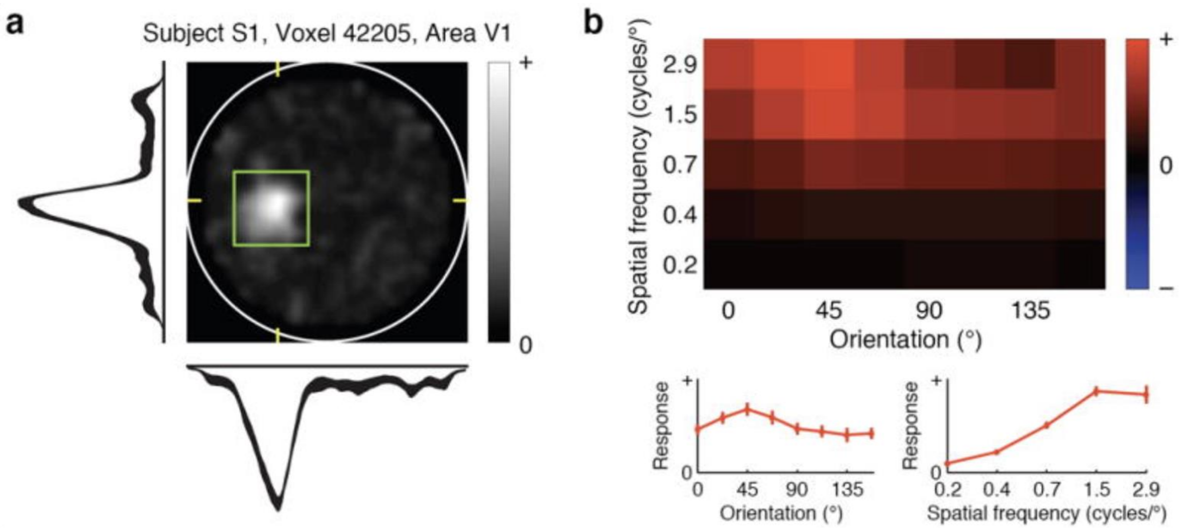
Classic encoding model finding: Vision

- 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





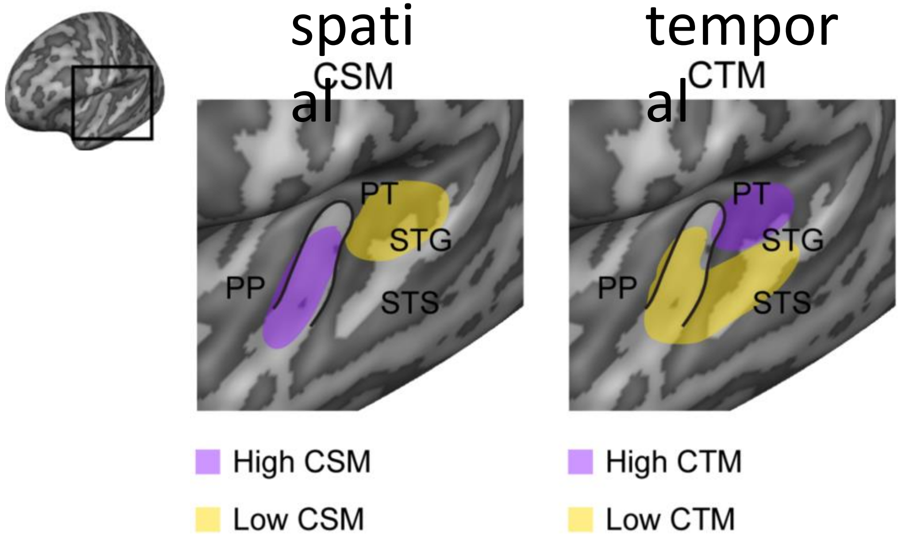
Kay, Kendrick N., Thomas Naselaris, Ryan J. Prenger, and Jack L. Gallant. "Identifying natural images from human brain activity." *Nature* 452, no. 7185 (2008): 352-355.

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

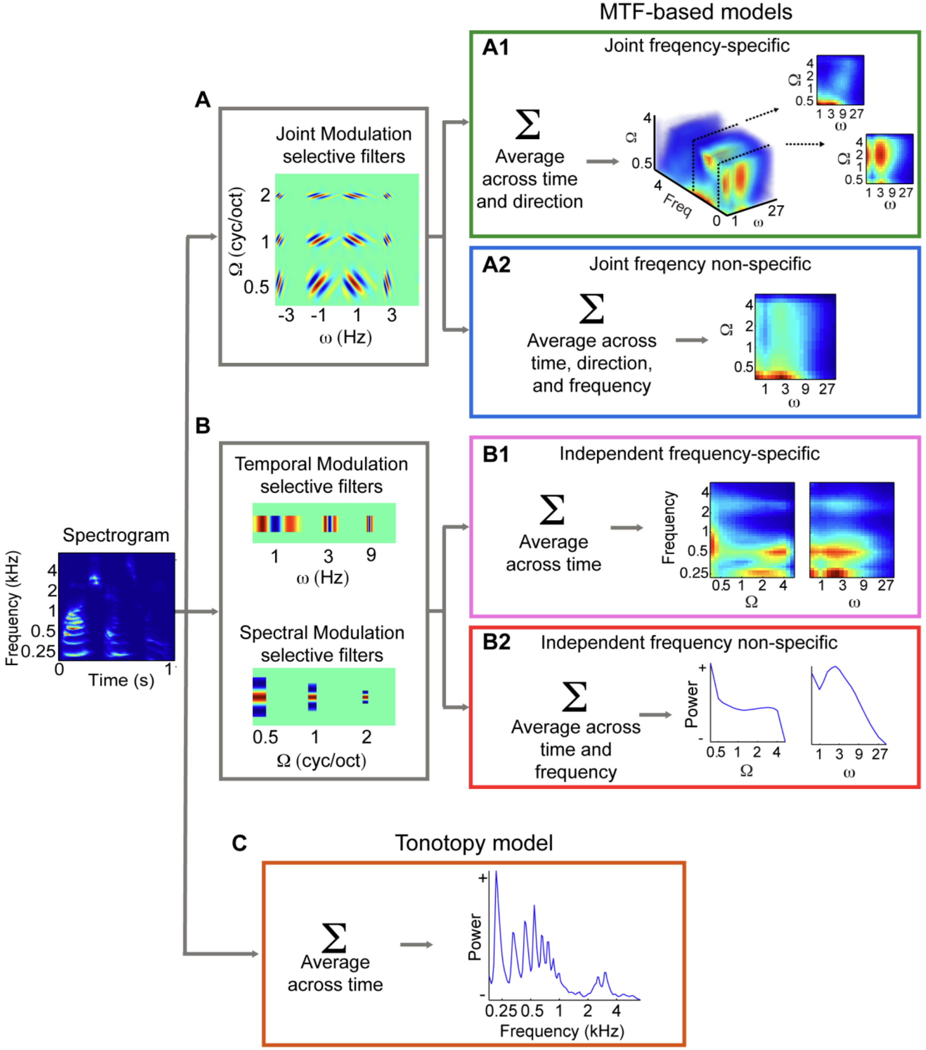
spatial

temporal



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

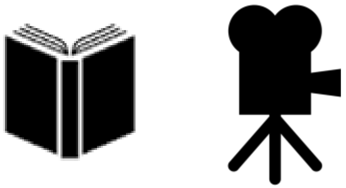
anterior/ventral auditory: fine-grained spectral & low temporal precision



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

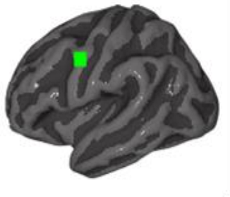
more naturalistic stimuli



more stimulus properties that affect brain activity

simple stim. representations explain less variance in brain activity

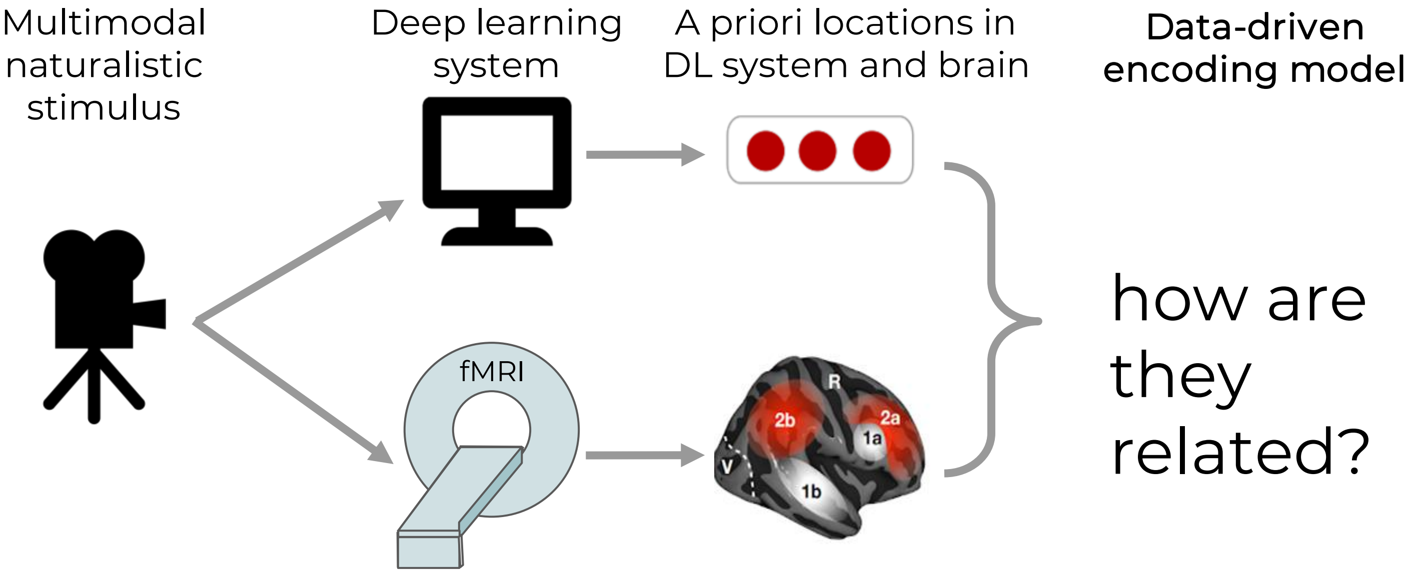
$f(<0,1,\ldots\ 0>) \approx$



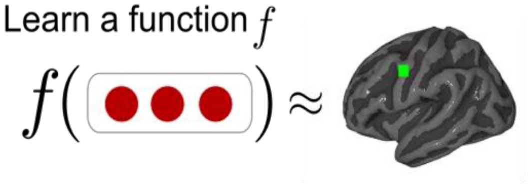
DeepMind’s New AI Taught Itself to Be the World’s Greatest Go Player
Singularity Hub

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**

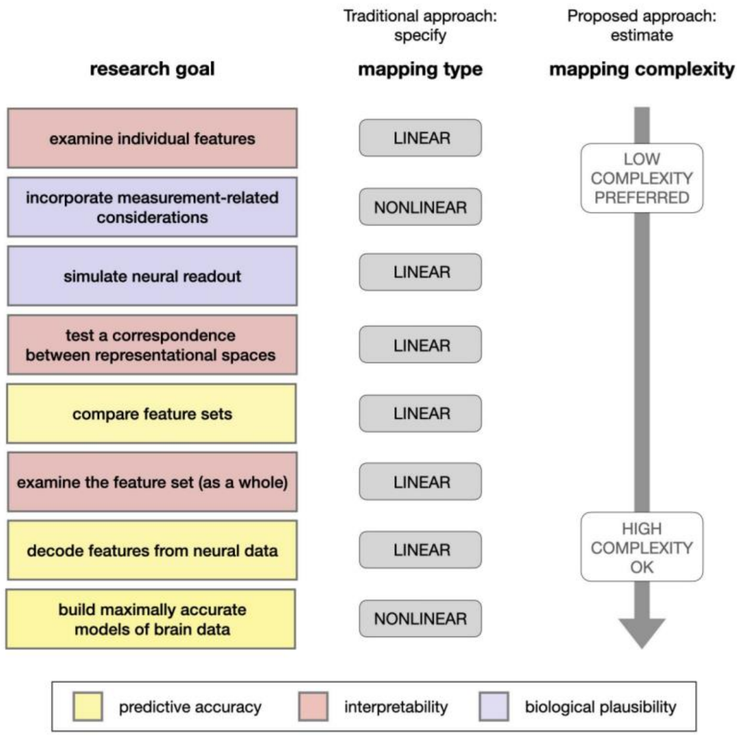


function f often modeled as linear

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

Considerations for f

Linear vs non-linear

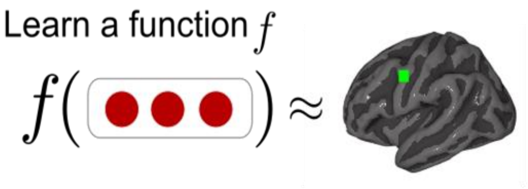


Ivanova, Anna A., Martin Schrimpf, Stefano Anzellotti, Noga Zaslavsky, Evelina Fedorenko, and Leyla Isik. "Is it that simple? Linear mapping models in cognitive neuroscience." bioRxiv (2021).

Encoding: **training and evaluation**

function f often modeled as linear

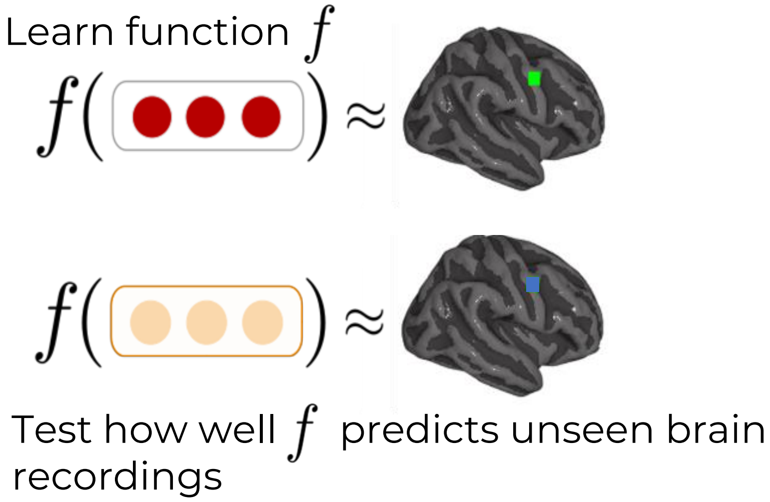
[Mitchell et al. 2008, Nishimoto et al., 2011; Sudre et al., 2012; Wehbe et al., 2014]



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**



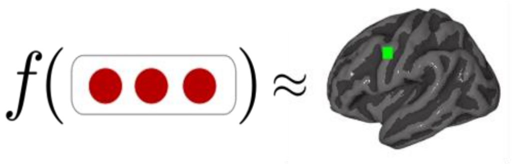
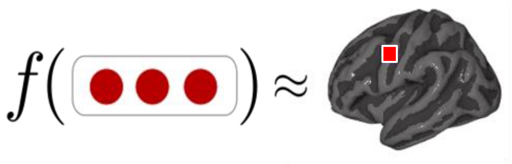
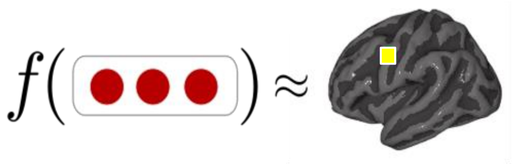
- Goal: find a mapping from stimulus representation to brain data that **generalizes** to new brain data

- 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

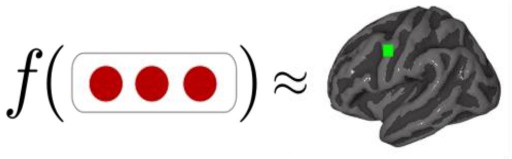
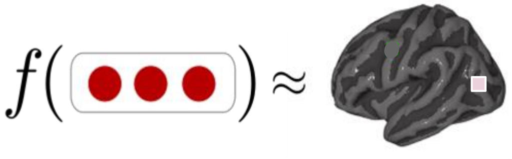
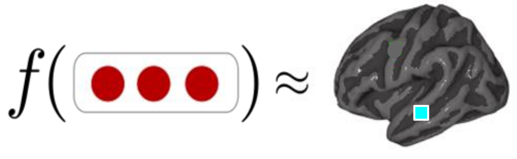
- Independent model per participant

P1 **P2** … **PN**

- Independent model per voxel / sensor-timepoint

P1, v1 **P1, v2** … **P1, vm**

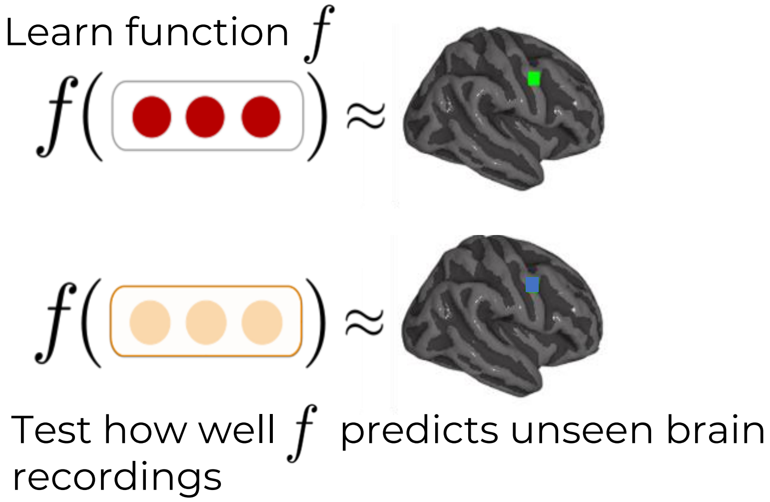
Encoding: fMRI specifics

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**

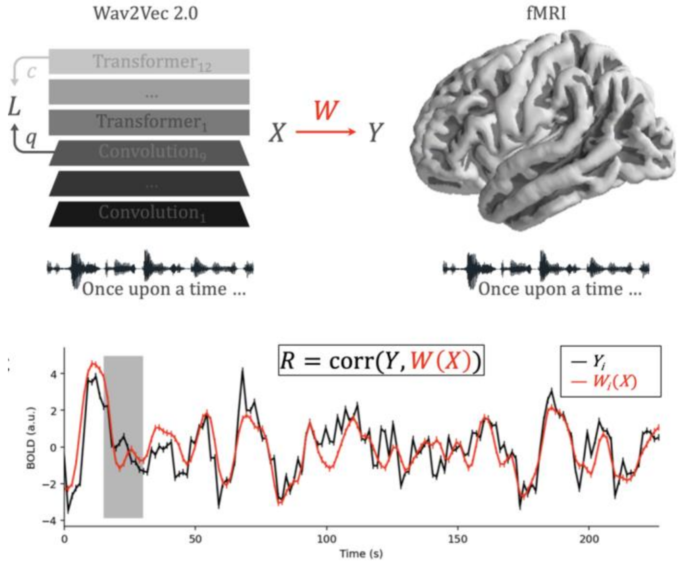
- Predict data heldout from training by applying learned function to corresponding stimulus representations

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



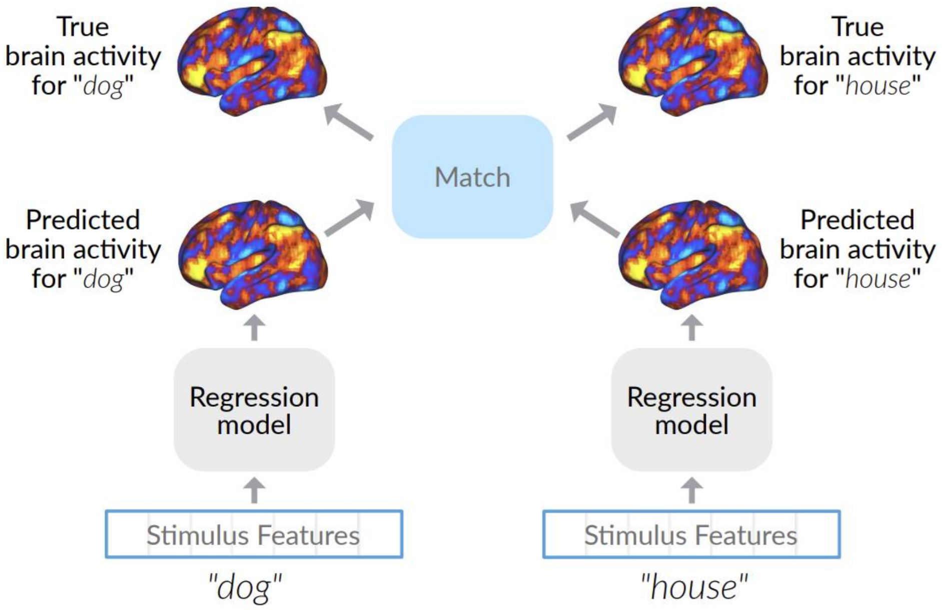
Encoding: evaluation **metrics**

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



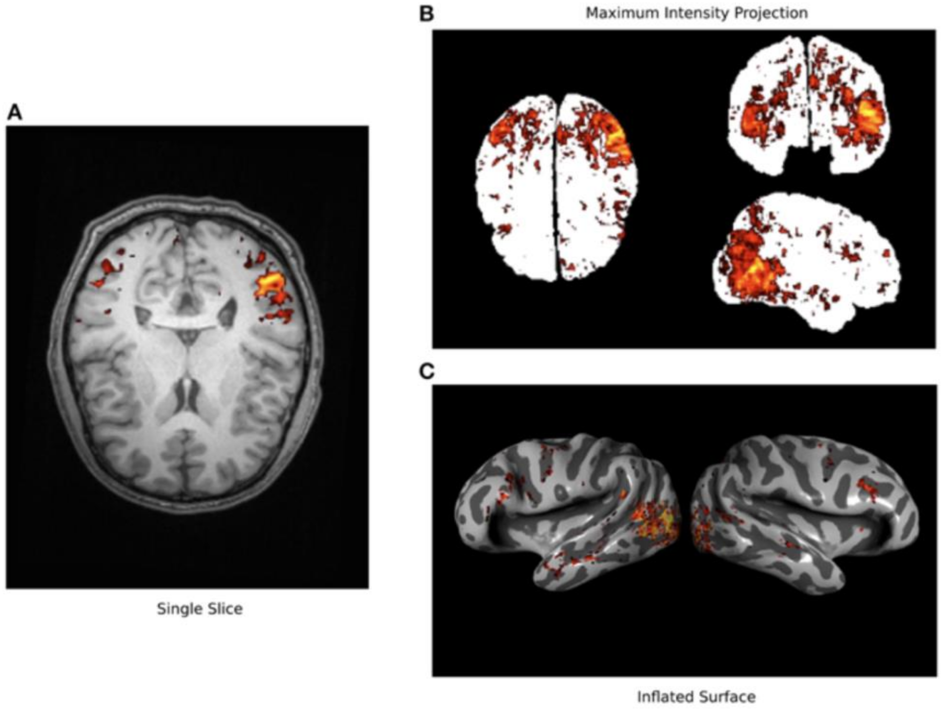
[Toneva, Mariya, Otilia Stretcu, Barnabás Póczos, Leila Wehbe, and Tom M. Mitchell. "Modeling task effects on meaning representation in the brain via zero-shot meg prediction." Advances in Neural Information Processing Systems 33 (2020): 5284-5295.]

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.

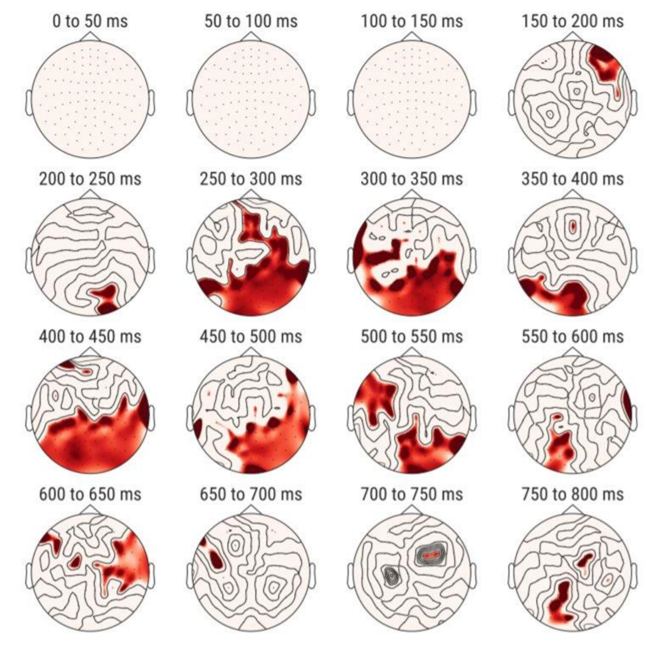
Encoding: performance **visualization**

fMRI



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 [1 hour 30 min]
- Coffee break [30 min]
- Deep Learning for Brain Decoding [1 hour 30 min]
- Lunch break [1 hour 15 min]
- **Deep Learning for Brain Encoding [1 hour 30 min]**
    - Classic findings & common approaches
    - **More recent findings utilizing deep learning**
- Coffee break [30 min]
- Advanced Methods [1 hour 15 min]
- Summary and Future Trends [15 min]

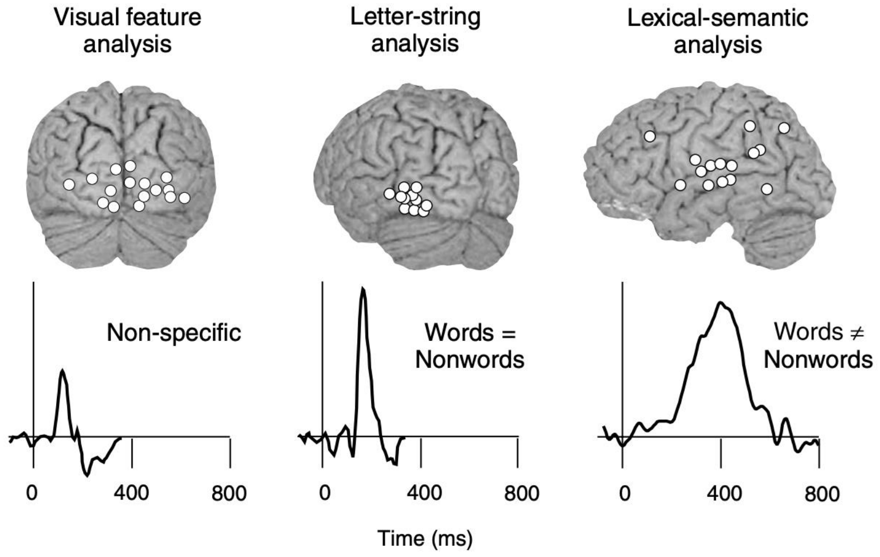
More recent work utilizing progress in DL for encoding

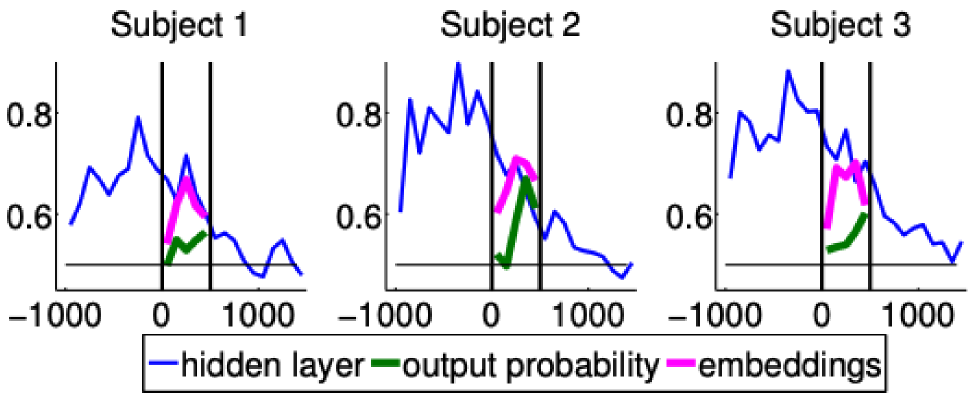
- Using representations of stimuli from deep learning systems
- **Language:**
  - Wehbe et al. 2014; Jain and Huth, 2018; Toneva and Wehbe, 2019; Caucheteux and King, 2020/2022; Schrimpf et al. 2020/2021; Goldstein et al. 2021/2022
- Vision:
  - Yamins et al. 2014; Cichy et al. 2016; Konkle and Alvarez,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

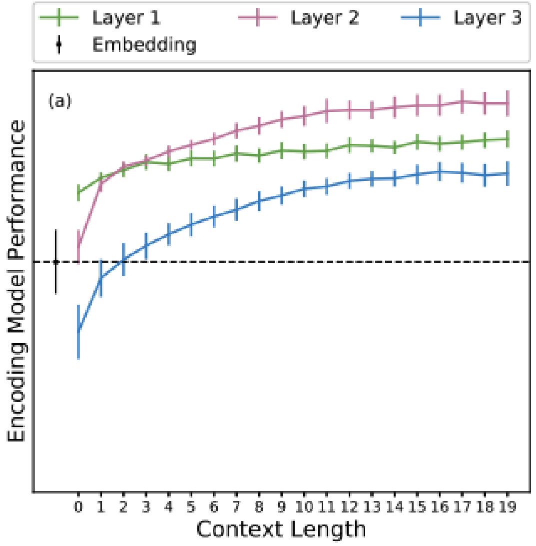


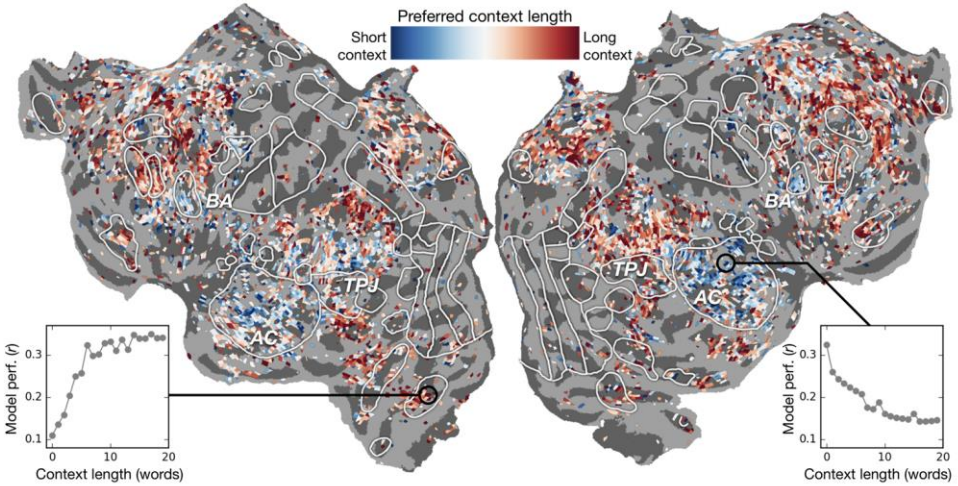


Wehbe, Leila, Ashish Vaswani, Kevin Knight, and Tom Mitchell. "Aligning context-based statistical models of language with brain activity during reading." In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 233-243. 2014.

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



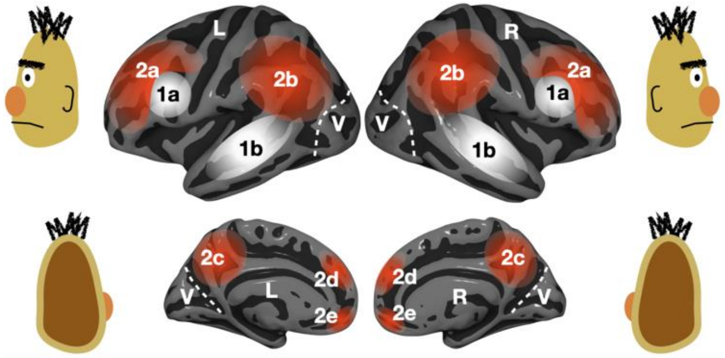


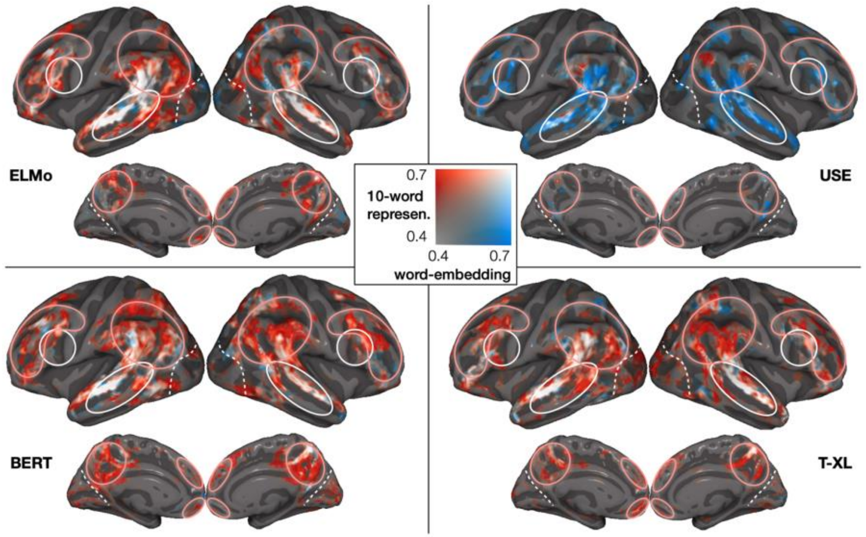
alignment between fMRI & recurrent NLP representations w/ varying context;
best alignment with middle layer

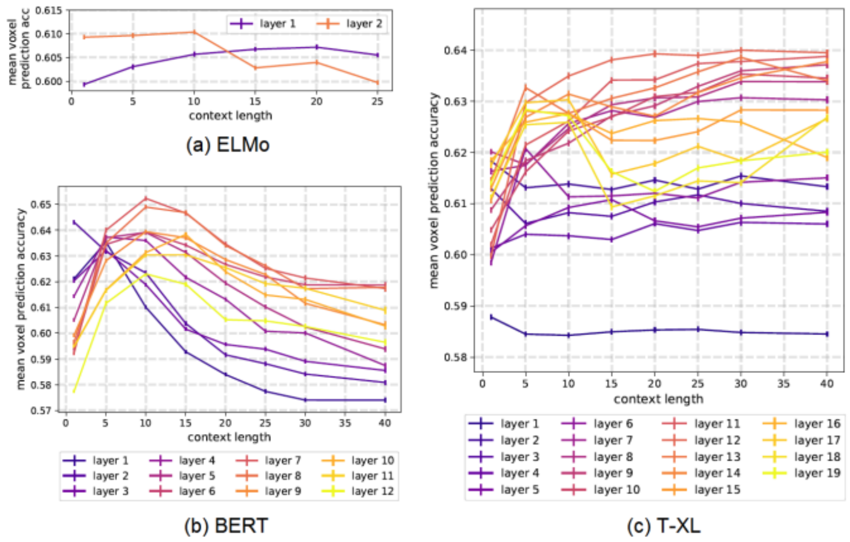
Jain, Shailee, and Alexander Huth. "Incorporating context into language encoding models for fMRI." Advances in neural information processing systems 31 (2018).

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

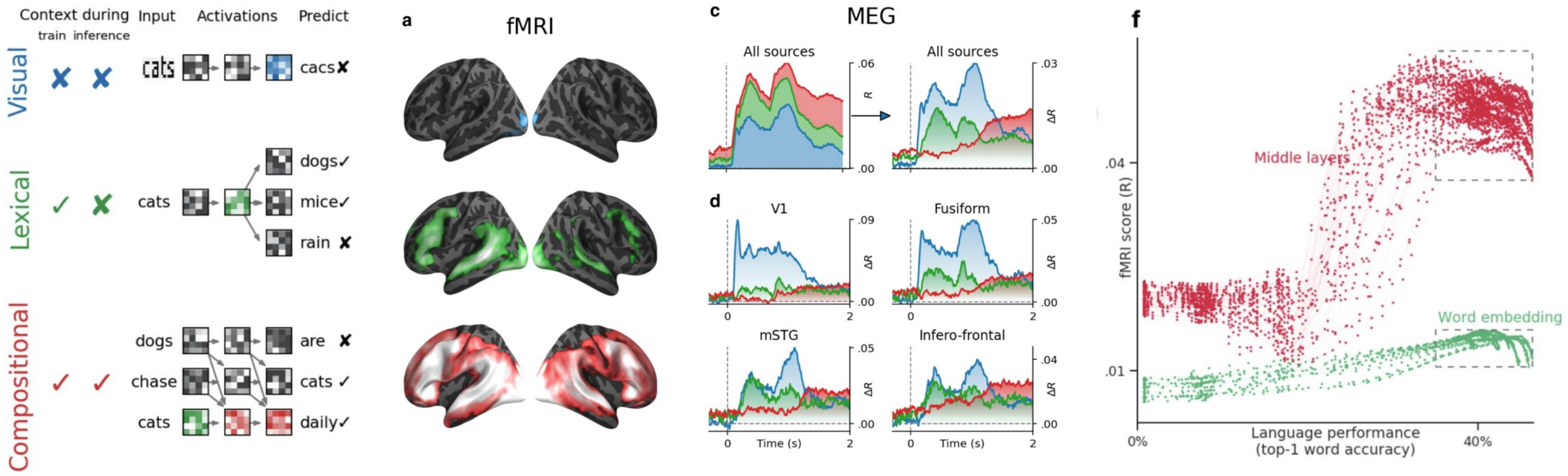
Toneva, M., & Wehbe, L. (2019). Interpreting and improving natural-language processing (in machines) with natural language-processing (in the brain). Advances in Neural Information Processing Systems, 32.

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

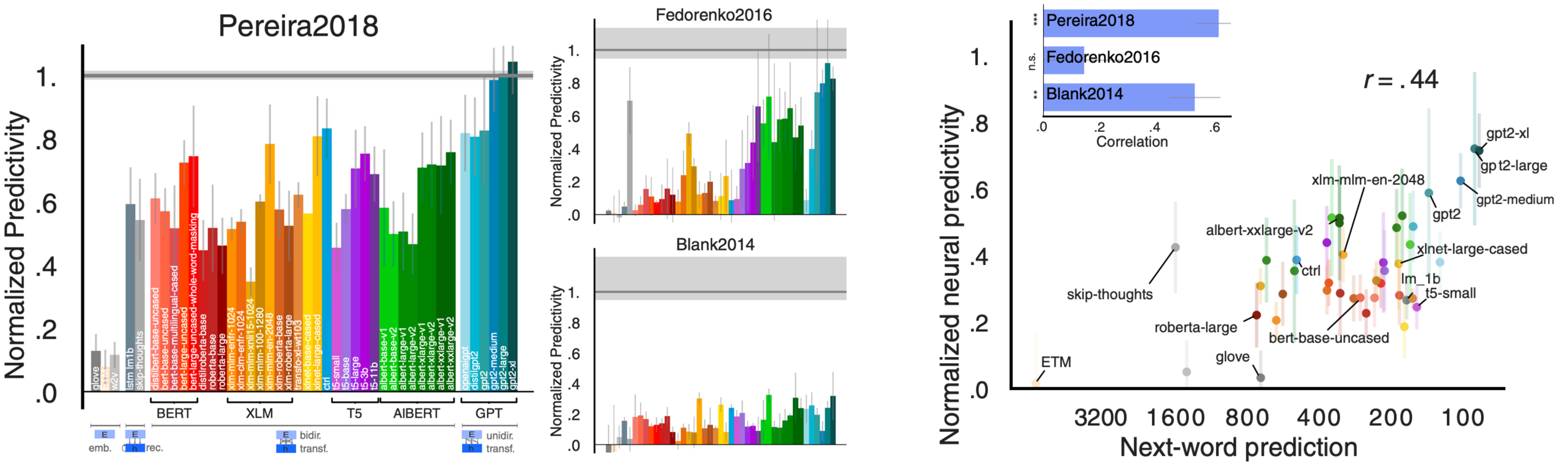


Caucheteux, Charlotte, and Jean-Rémi King. "Brains and algorithms partially converge in natural language processing." Communications biology 5, no. 1 (2022): 1-10.

Language: work utilizing DL progress

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

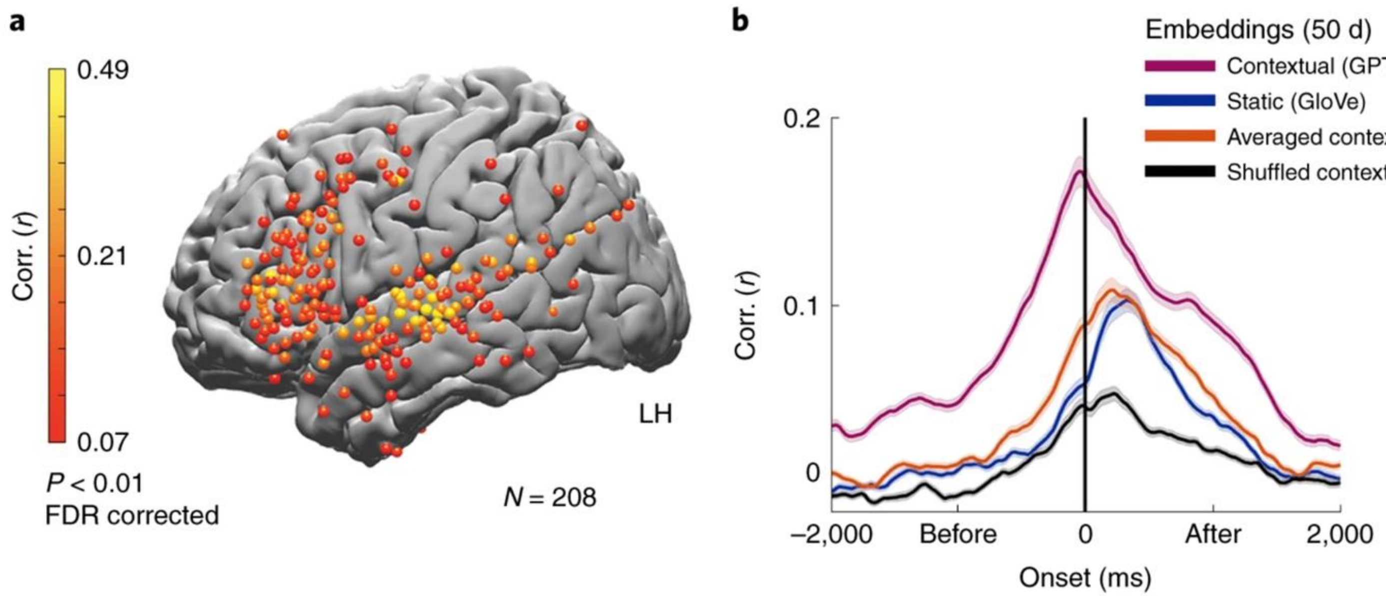
some NLP systems can predict fMRI and ECoG up to 100% of estimated noise ceiling



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

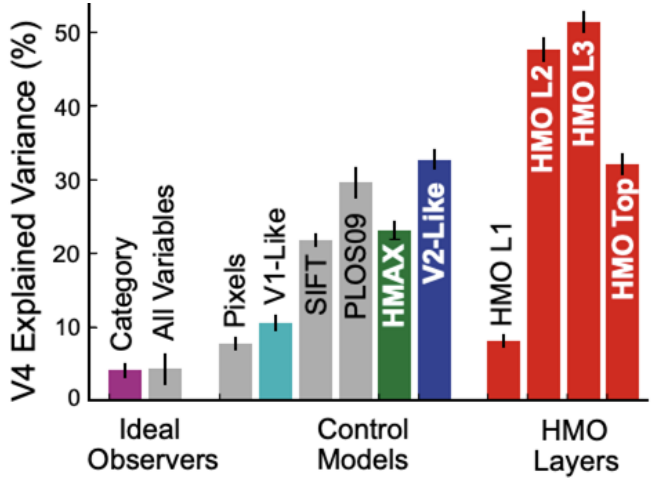
Goldstein, Ariel, Zaid Zada, Eliav Buchnik, Mariano Schain, Amy Price, Bobbi Aubrey, Samuel A. Nastase et al. "Shared computational principles for language processing in humans and deep language models." Nature neuroscience 25, no. 3 (2022): 369-380.

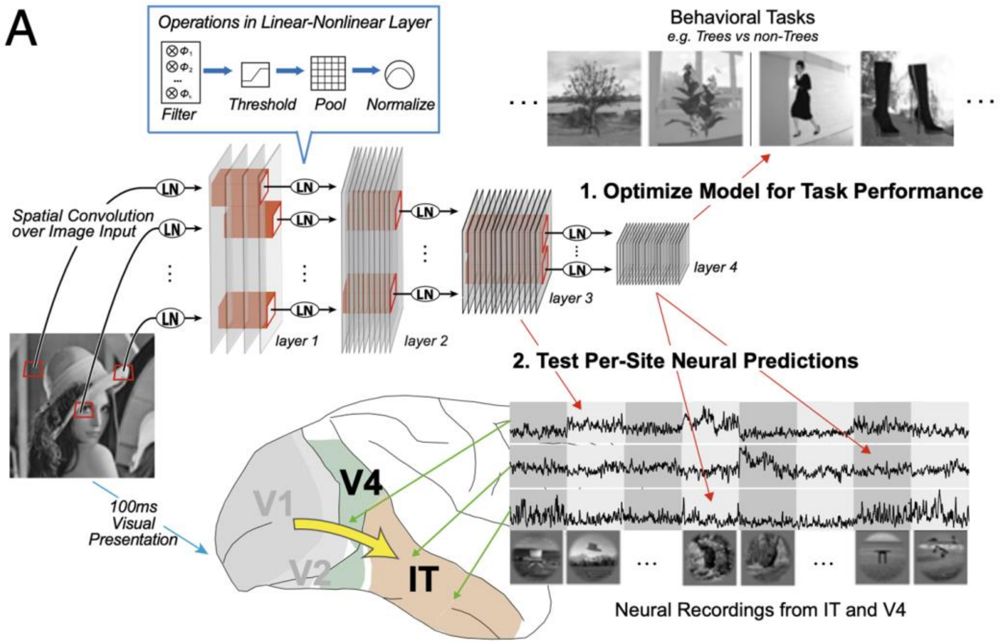
Recent work utilizing progress in DL for encoding

- 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

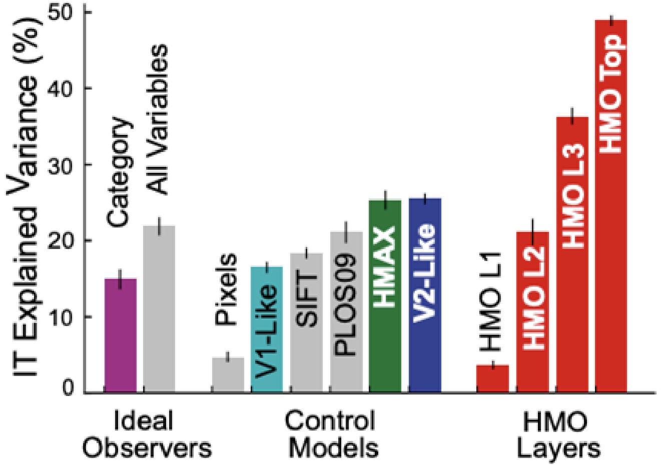
Vision: work utilizing DL progress

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





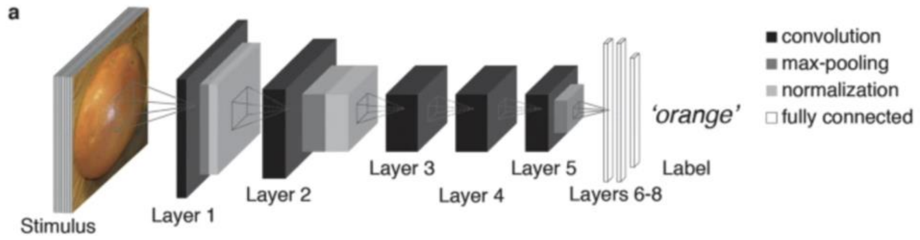
Highest layer in CNN model most predictive of IT; intermediate layers most predictive of V4

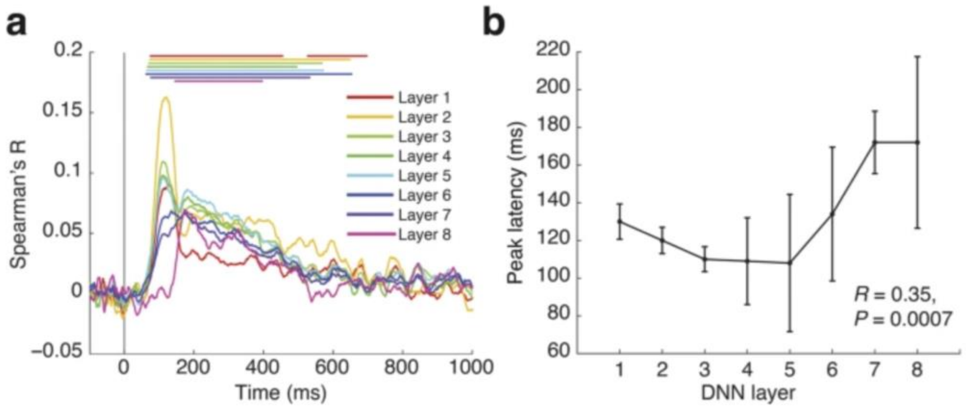


Yamins, Daniel LK, Ha Hong, Charles F. Cadieu, Ethan A. Solomon, Darren Seibert, and James J. DiCarlo. "Performance-optimized hierarchical models predict neural responses in higher visual cortex." *Proceedings of the national academy of sciences* 111, no. 23 (2014): 8619-8624.

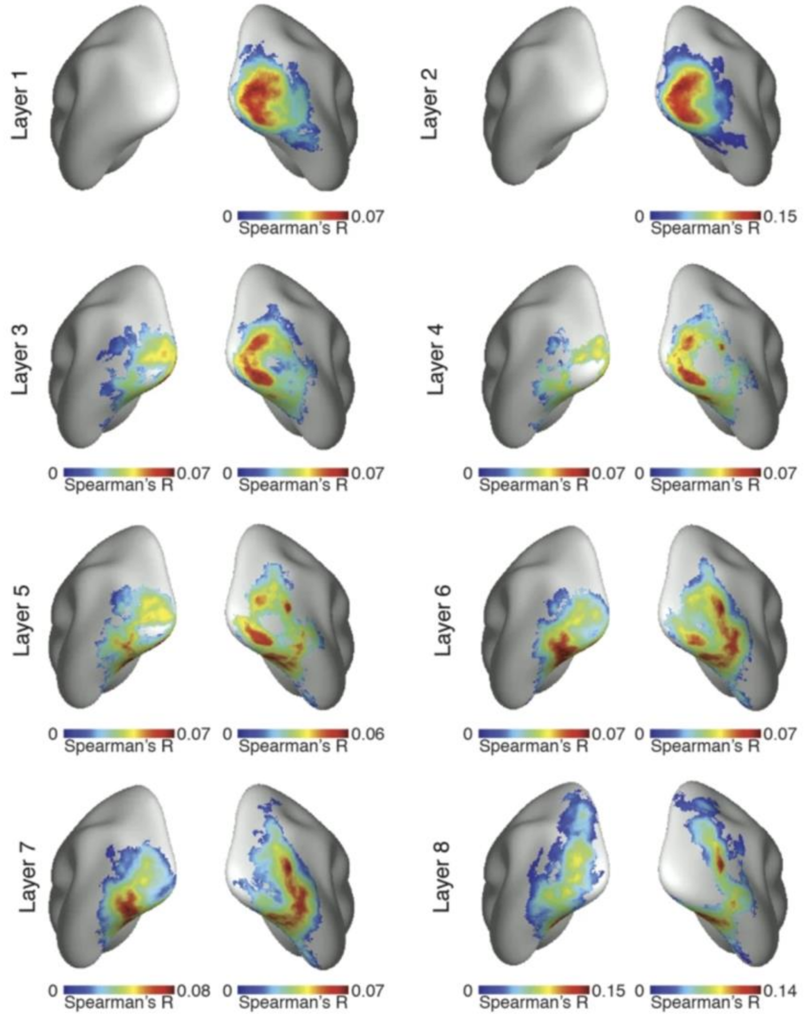
Vision: work utilizing DL progress

- 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

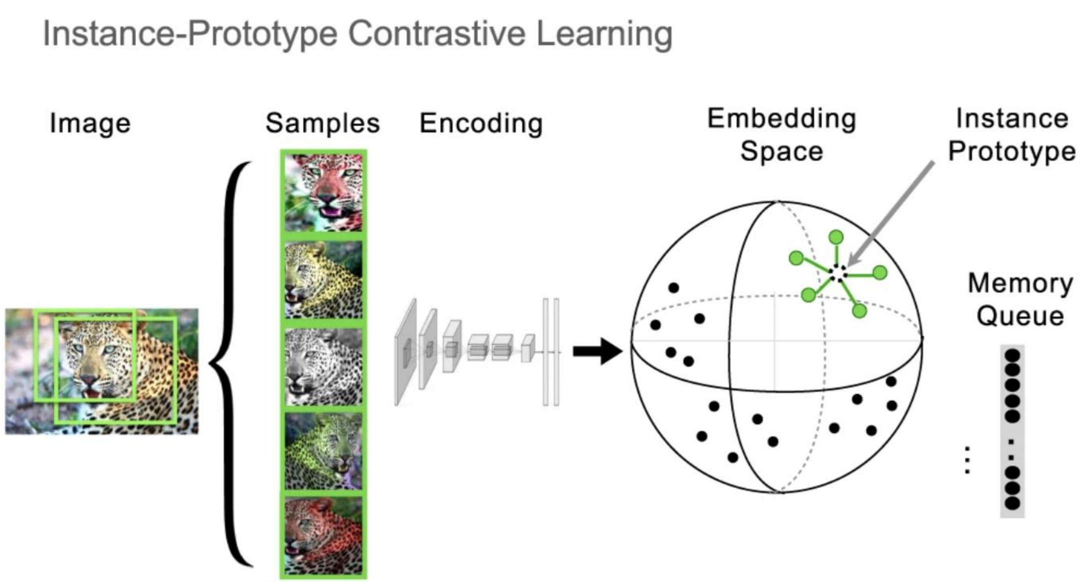


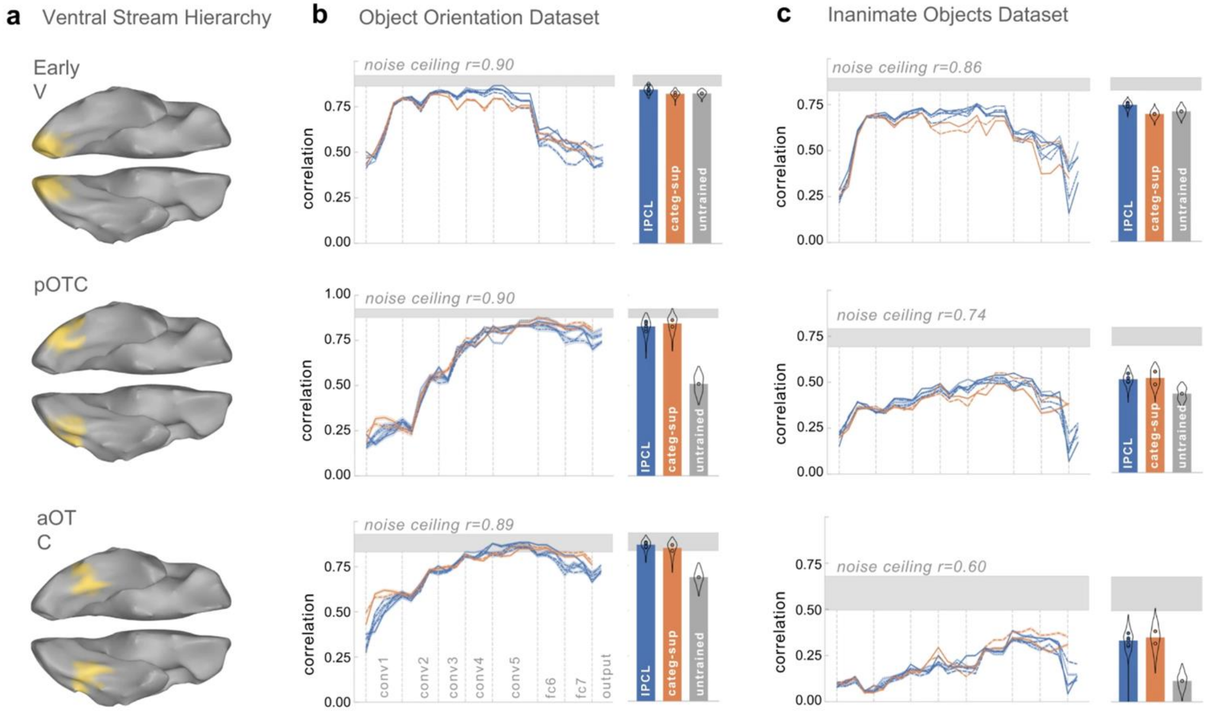
Cichy, Radoslaw Martin, Aditya Khosla, Dimitrios Pantazis, Antonio Torralba, and Aude Oliva. "Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence." Scientific reports 6, no. 1 (2016): 1-13.

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



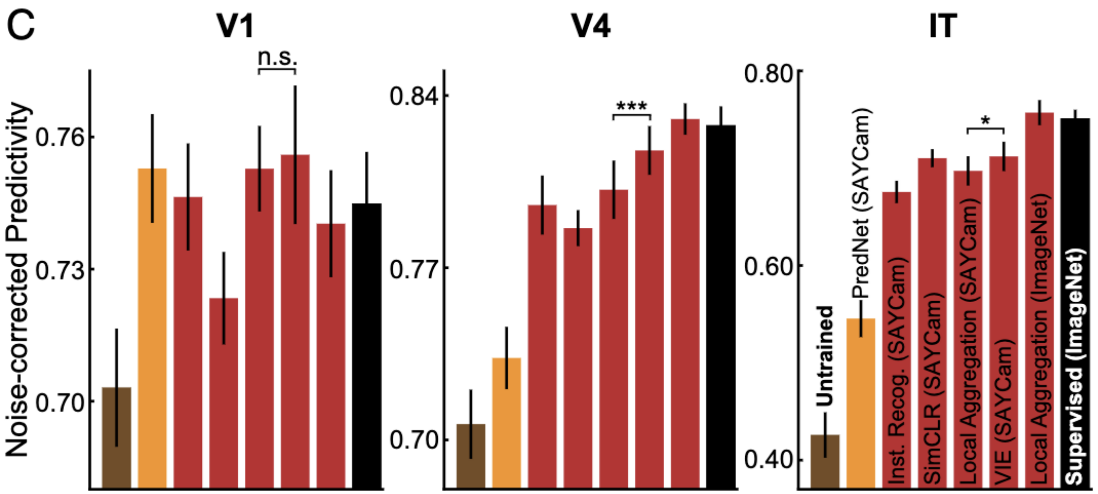


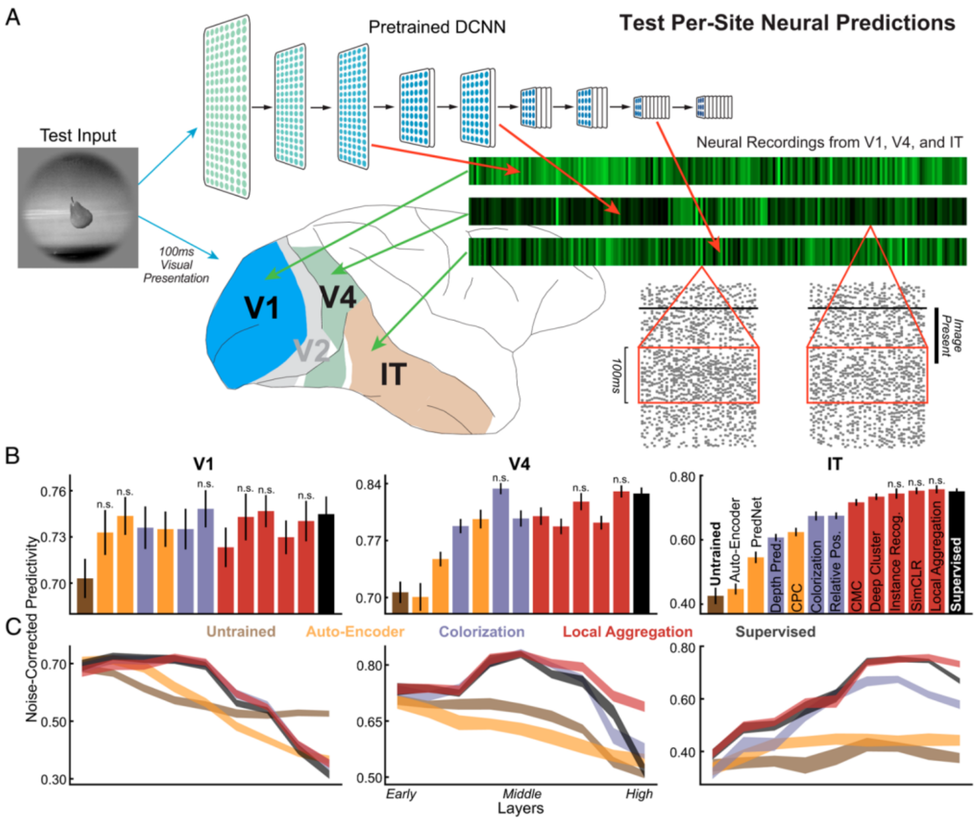
Konkle, Talia, and George A. Alvarez. "A self-supervised domain-general learning framework for human ventral stream representation." Nature communications 13, no. 1 (2022): 1-12.

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





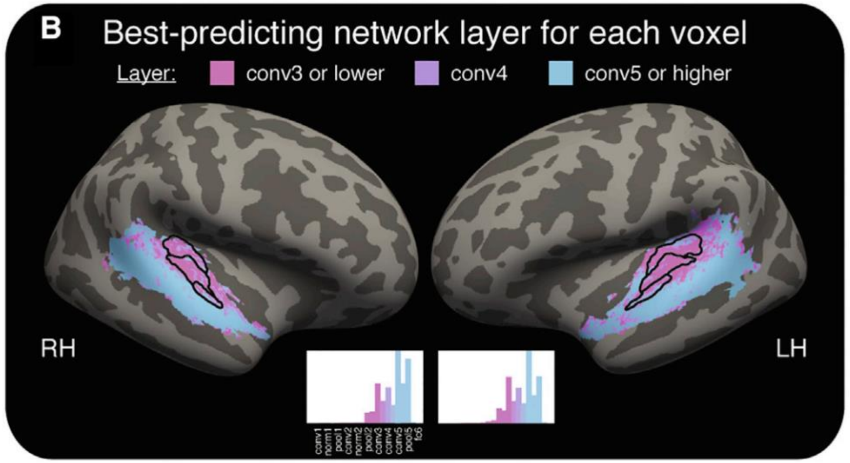
[Zhuang, Chengxu, Siming Yan, Aran Nayebi, Martin Schrimpf, Michael C. Frank, James J. DiCarlo, and Daniel LK Yamins. "Unsupervised neural network models of the ventral visual stream." Proceedings of the National Academy of Sciences 118, no. 3 (2021): e2014196118.]

Recent work utilizing progress in DL for encoding

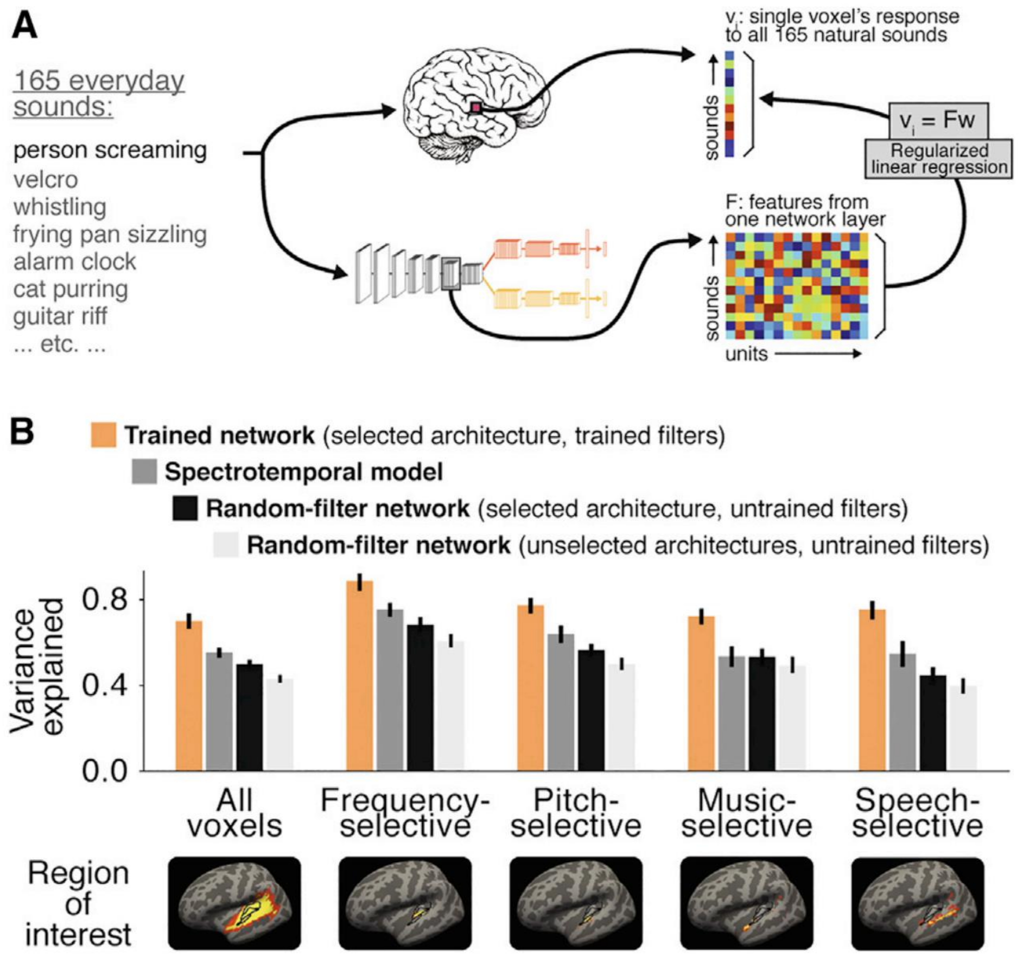
- Using representations of stimuli from deep learning systems
  - Data-driven
- Language:
  - Wehbe et al. 2014; Jain and Huth, 2018; Toneva and Wehbe, 2019; Caucheteux and King, 2020/2022; Schrimpf et al. 2020/2021; Goldstein et al. 2021/2022
- Vision:
  - Yamins et al. 2014; Cichy et al. 2016; Konkle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

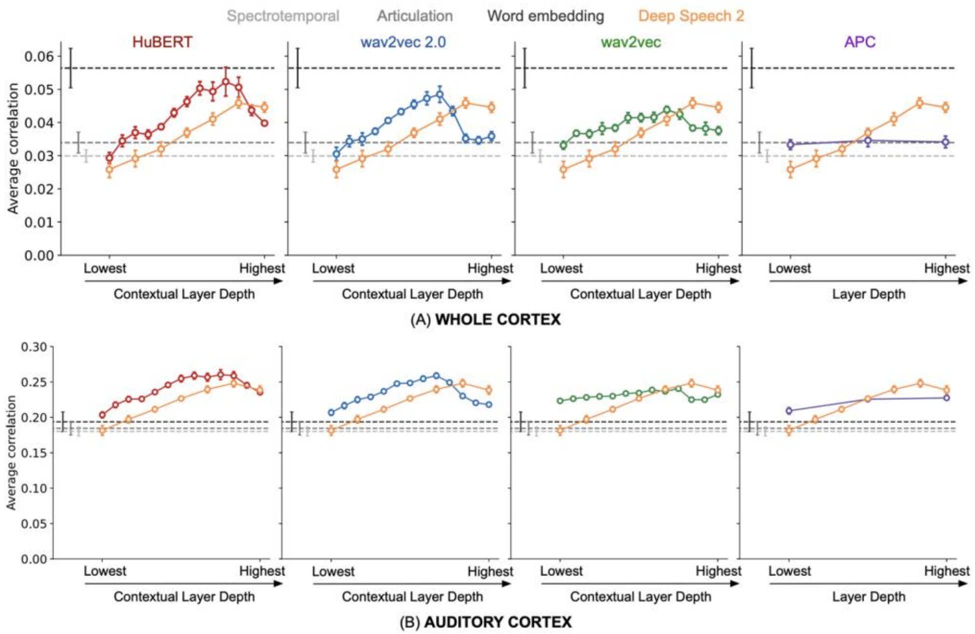


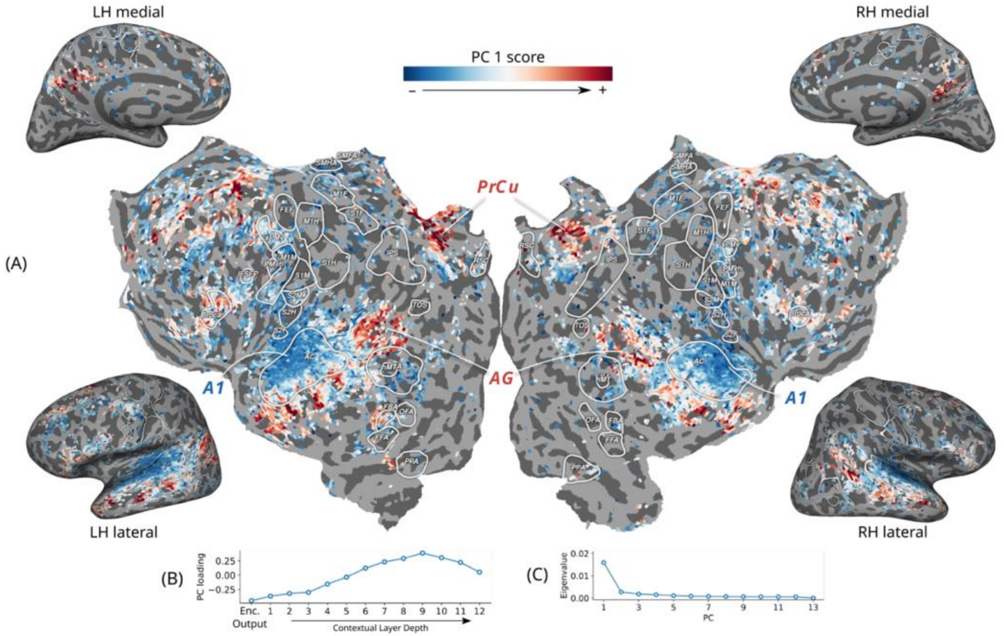
Kell, Alexander JE, Daniel LK Yamins, Erica N. Shook, Sam V. Norman-Haignere, and Josh H. McDermott. "A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy." Neuron 98, no. 3 (2018): 630-644.

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

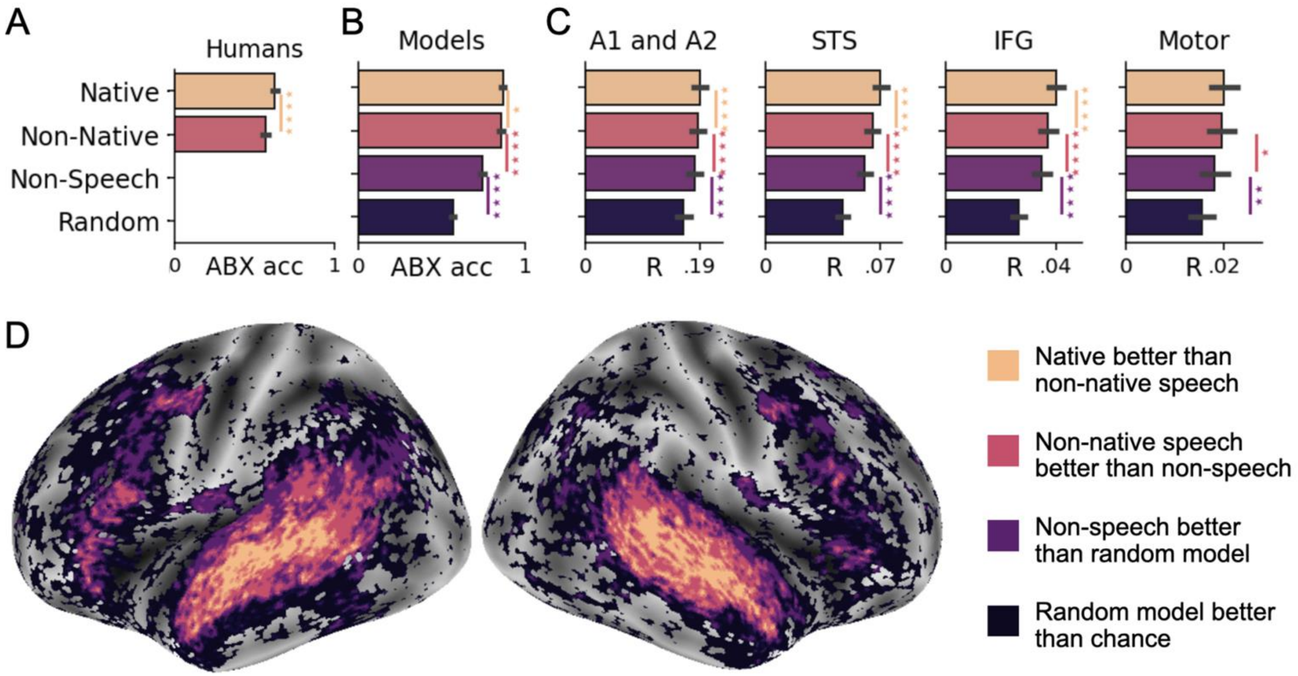




[Vaidya, Aditya R., Shailee Jain, and Alexander G. Huth. "Self-supervised models of audio effectively explain human cortical responses to speech." ICML (2022).](#)

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 reveal 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 [1 hour 30 min]
- Coffee break [30 min]
- Deep Learning for Brain Decoding [1 hour 30 min]
- Lunch break [1 hour 15 min]
- **Deep Learning for Brain Encoding [1 hour 30 min]**
  - Classic findings & common approaches
  - More recent findings utilizing deep learning
- Coffee break [30 min]
- Advanced Methods [1 hour 15 min]
- Summary and Future Trends [15 min]

Deep Learning for Brain Encoding and Decoding

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

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

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




8

Agenda

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

Challenges in using DL for cognitive modeling

- Not designed to specifically model brain processing

NLP systems: Designed to predict upcoming words

Harry never thought ***???***

Harry never thought he ***???***

Harry never thought he would ***???***

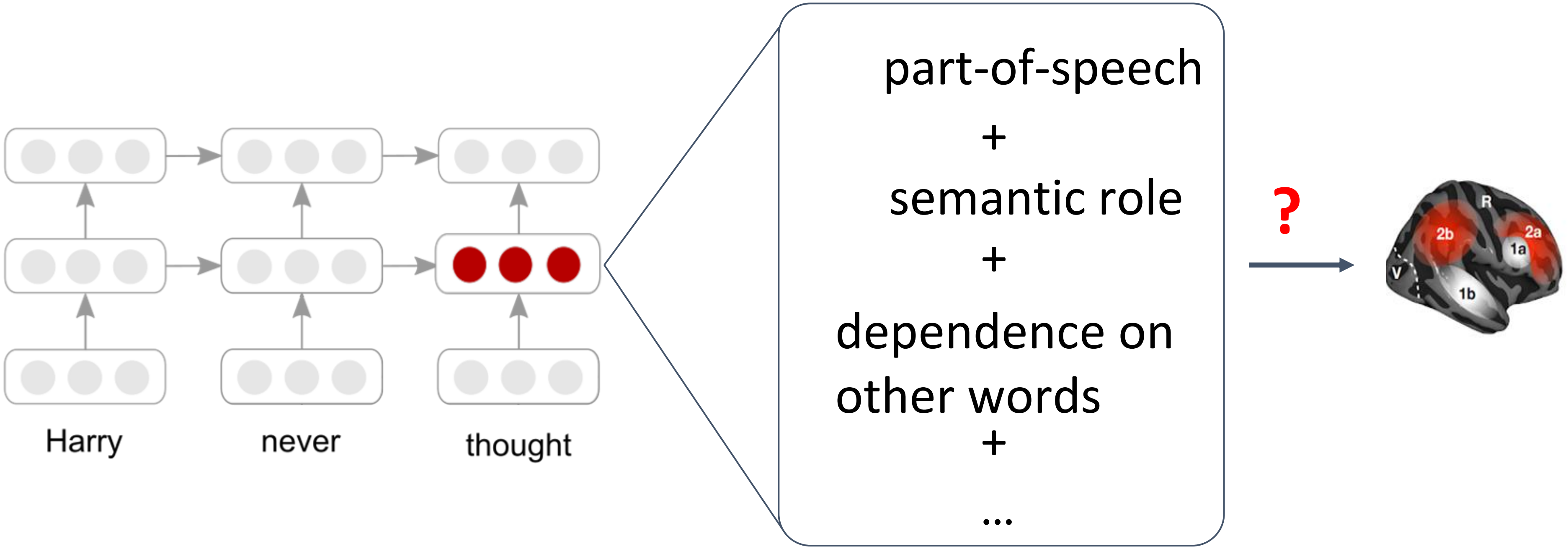
...

Challenges in using DL for cognitive modeling

- Not designed to specifically model brain processing
  - Training DL models using brain recordings
  - Task-based modeling

Challenges in using DL for cognitive science

- Not designed to specifically model brain processing
  - Training DL models using brain recordings
  - Task-based modeling
- Can be difficult to interpret due to multiple sources of information



Challenges in using DL for cognitive science

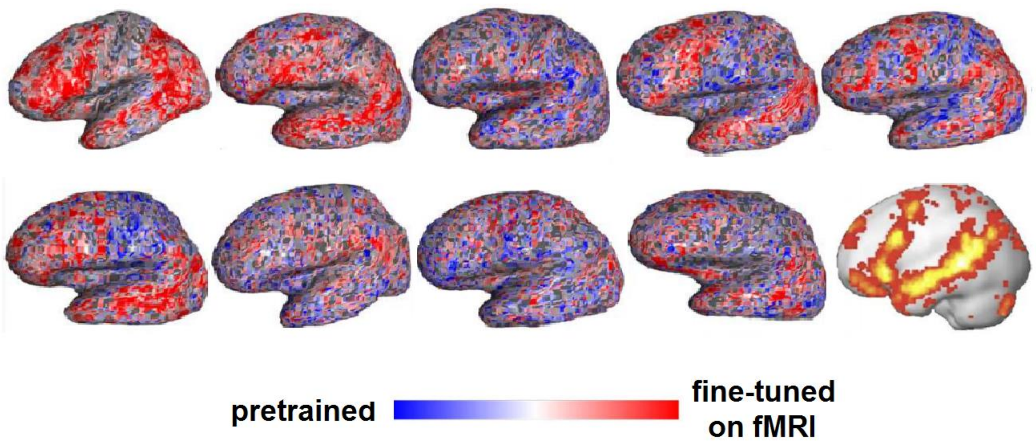
- Not designed to specifically model brain processing
  - Training DL models using brain recordings
  - Task-based modeling
- Can be difficult to interpret due to multiple sources of information
  - Disentangling contributions of different info sources to brain predictions

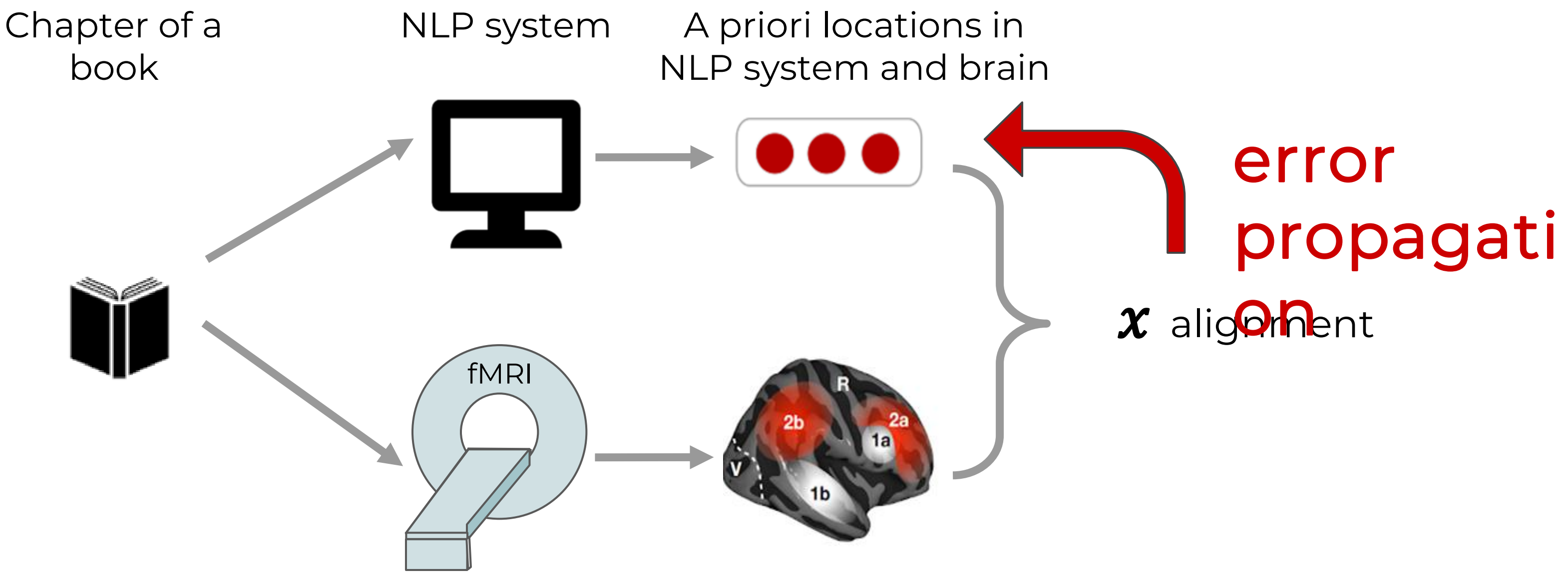
Challenges in using DL for cognitive science

- Not designed to specifically model brain processing
  - **Training DL models using brain recordings**
  - Task-based modeling
- Can be difficult to interpret due to multiple sources of information
  - Disentangling contributions of different info sources to brain predictions

Training DL models using brain recordings

- Stimuli: one chapter of Harry Potter
- Stimulus representation: brain-optimized NLP model
- Brain recording & modality: fMRI & MEG, reading



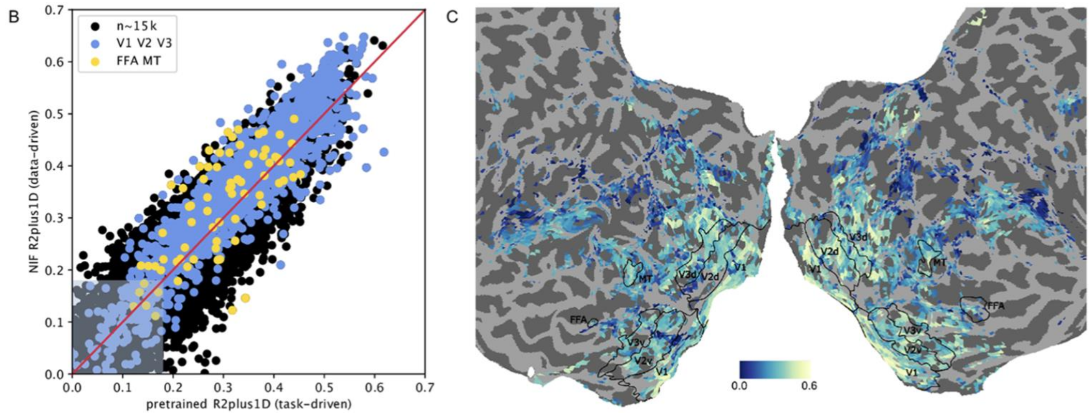


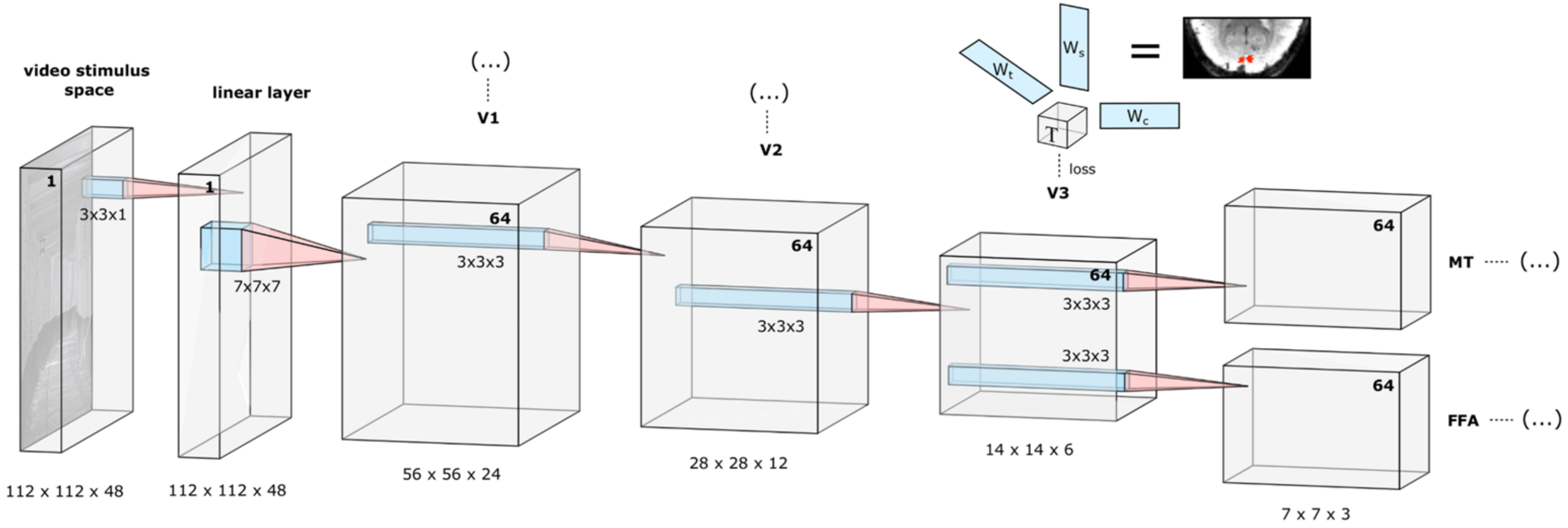
Brain-optimized NLP model predicts unseen fMRI recordings better, especially in canonical language regions

Schwartz, Dan, Mariya Toneva, and Leila Wehbe. "Inducing brain-relevant bias in natural language processing models." Advances in neural information processing systems 32 (2019).

Training DL models using brain recordings

- Stimuli: movie and TV show clips
- Stimulus representation: brain-optimized CNN
- Brain recording & modality: fMRI, vision



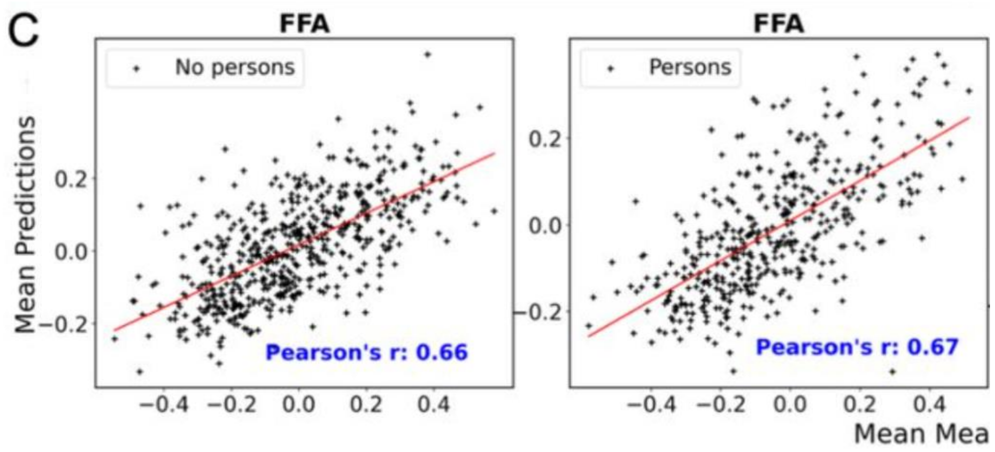


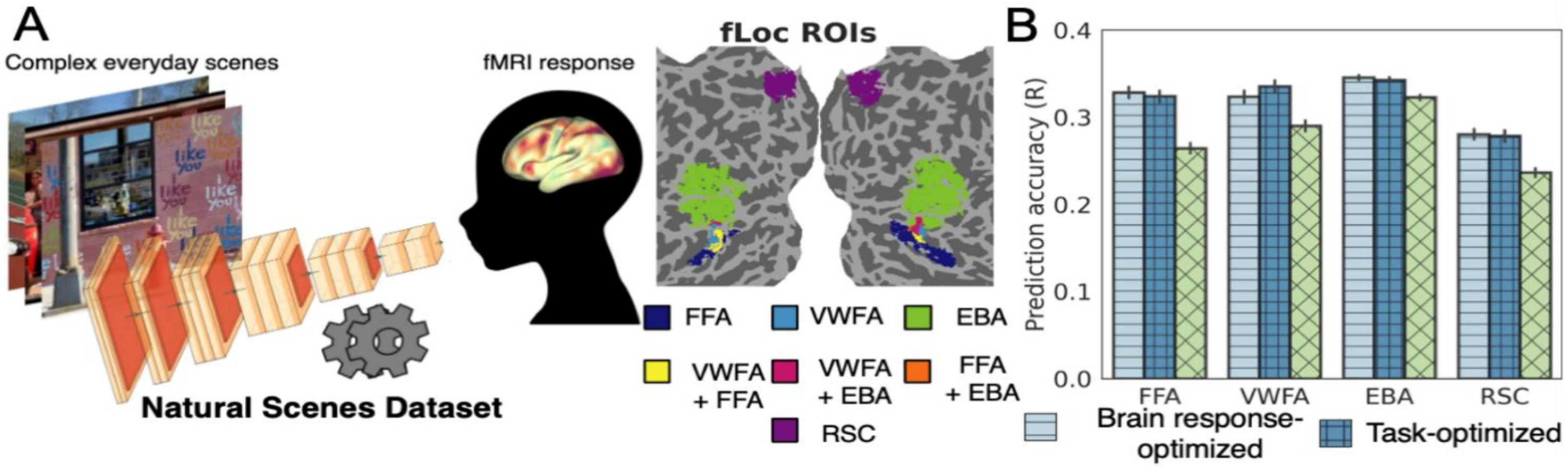
Brain-optimized vision model trained entirely on fMRI recordings ~= task-optimized networks for predicting brain recordings in early and high-level ROI

Seeliger, Katja, Luca Ambrogioni, Yağmur Güçlütürk, Leonieke M. van den Bulk, Umut Güçlü, and Marcel AJ van Gerven. "End-to-end neural system identification with neural information flow." PLOS Co

Training DL models using brain recordings

- Stimuli: images natural scenes
- Stimulus representation: brain-optimized CNN
- Brain recording & modality: fMRI, vision



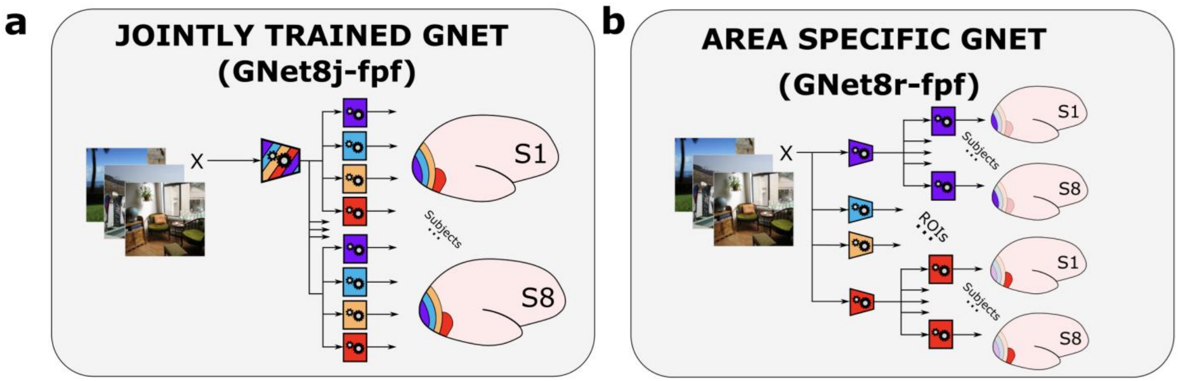


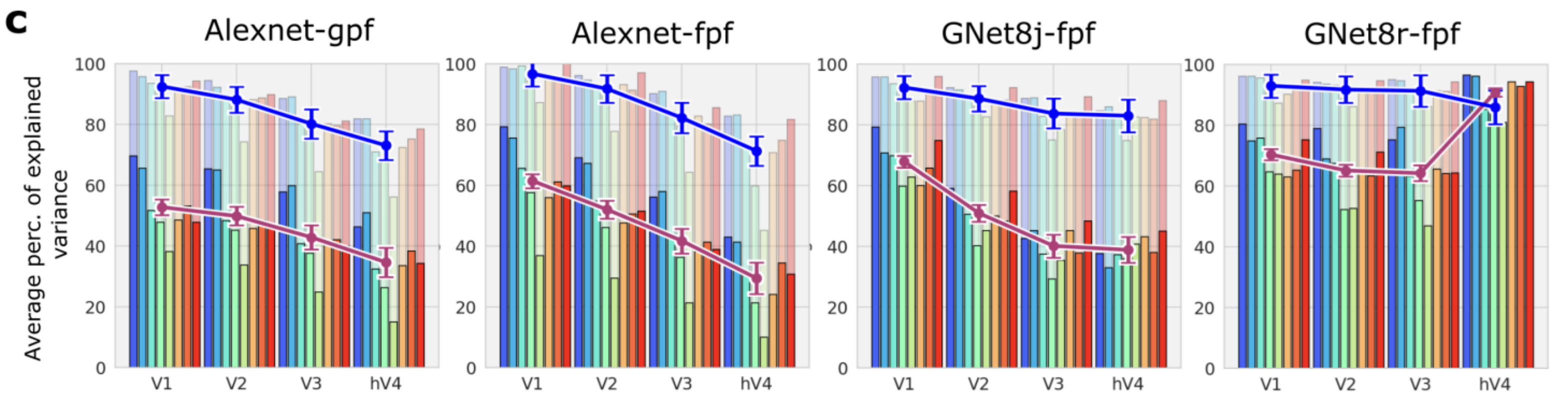
Brain-optimized vision model can predict brain signals corresponding to a category of stimuli that it was never trained on

Khosla, Meenakshi, and Leila Wehbe. "High-level visual areas act like domain-general filters with strong selectivity and functional specialization." bioRxiv (2022).

Training DL models using brain recordings

- Stimuli: images natural scenes
- Stimulus representation: brain-optimized CNN
- Brain recording & modality: fMRI, vision





Brain-optimized vision model can learn representations that do not follow a strict hierarchy

St-Yves, Ghislain, Emily J. Allen, Yihan Wu, Kendrick Kay, and Thomas Naselaris. "Brain-optimized neural networks learn non-hierarchical models of representation in human visual cortex." Nature Communications (2023).

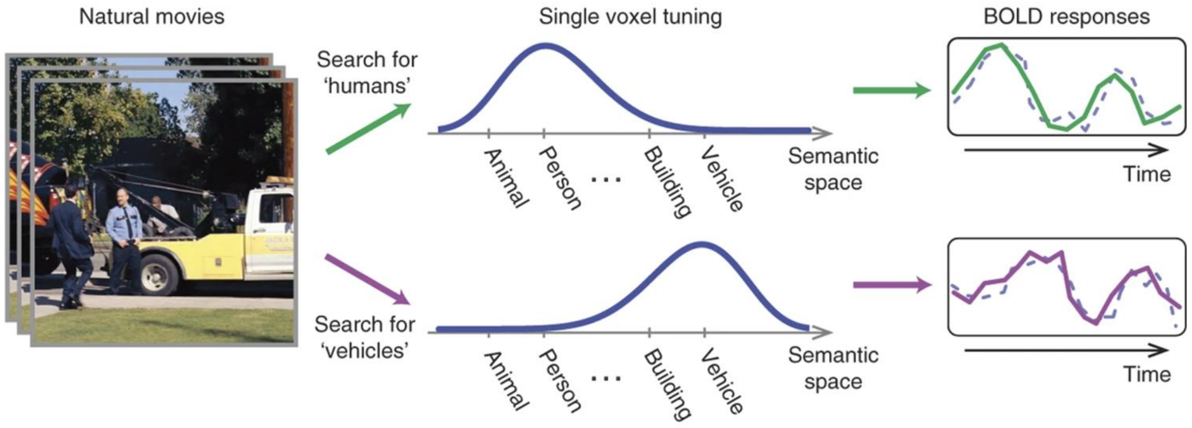
Challenges in using DL for cognitive modeling

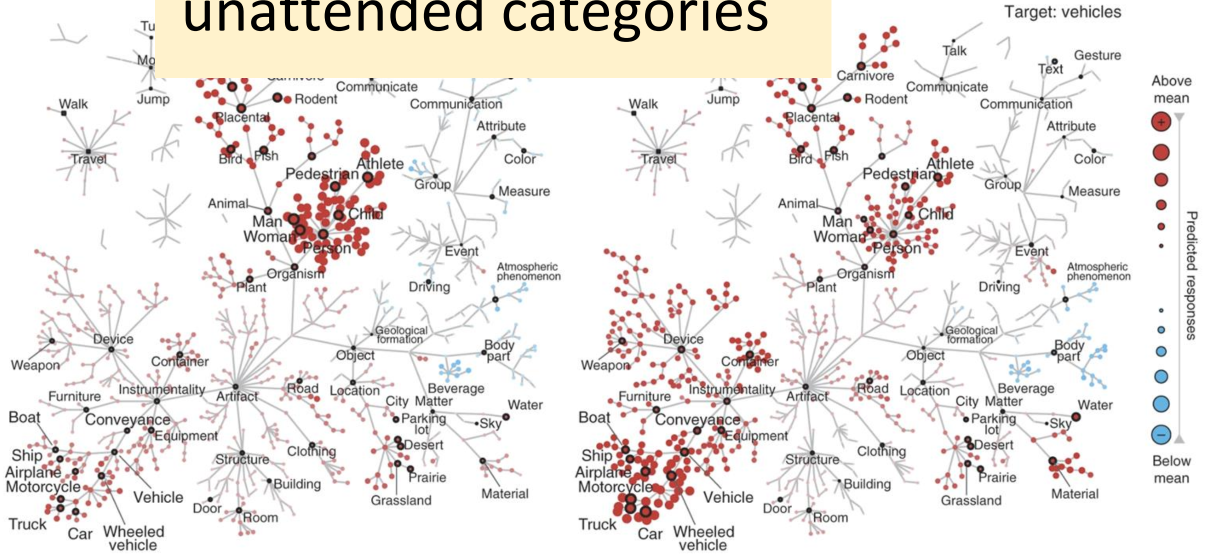
- Not designed to specifically model brain processing
  - Training DL models using brain recordings
  - **Task-based modeling**
- Can be difficult to interpret due to multiple sources of information
  - Disentangling contributions of different info sources to brain predictions

Tasks affect processing

- Stimuli: natural movies
- Task: visual search for vehicles or humans
- Stimulus representation: object and action labels from WordNet
- Brain recording & modality: fMRI, vision

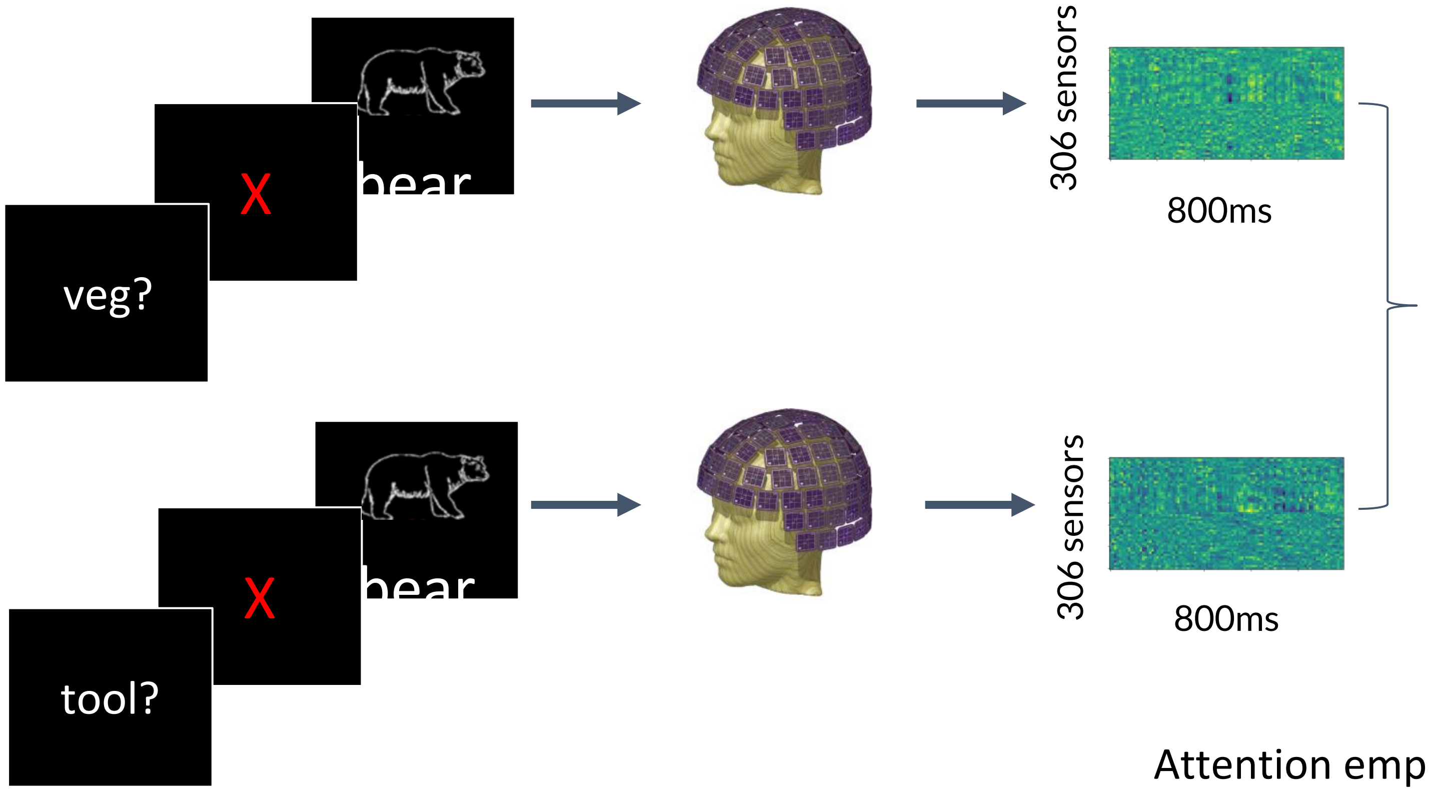
Category-based attention during natural vision alters representation of both attended and unattended categories





Çukur, Tolga, Shinji Nishimoto, Alexander G. Huth, and Jack L. Gallant. "Attention during natural vision warps semantic representation across the human brain." Nature neuroscience 16, no. 6 (2013): 763-770.

Tasks affect processing



Systematic difference due to different question tasks

Can we model as a function of the task **AND** stimulus?

Attention emphasizes task-relevant information

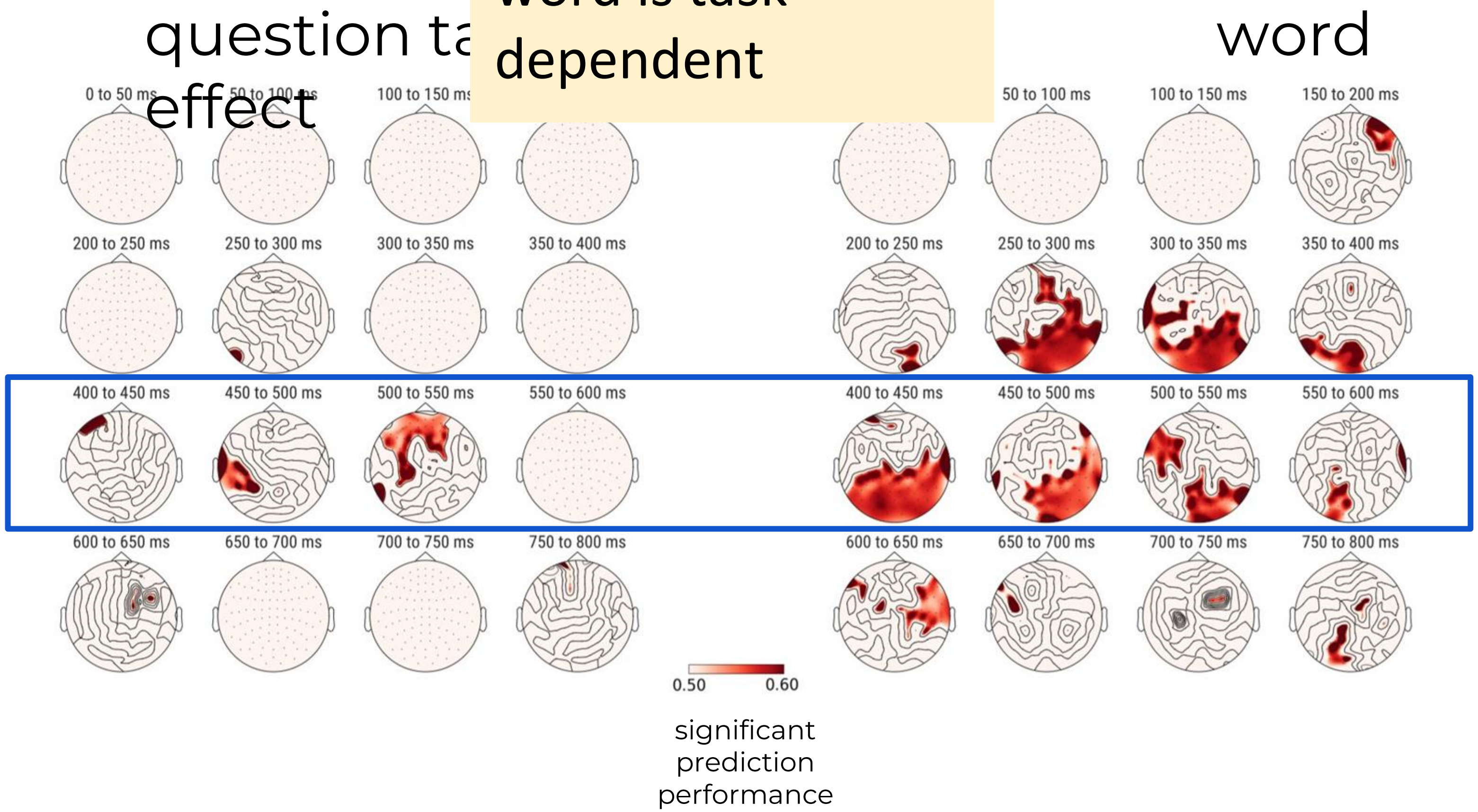
Mechanism?

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.

Tasks affect processing

- Stimuli: concrete nouns + line drawings
- Task: answer Yes/No questions about noun
- Stimulus representation: human judgments
- Brain recording & modality: MEG, reading

The end of semantic processing of a word is task-dependent

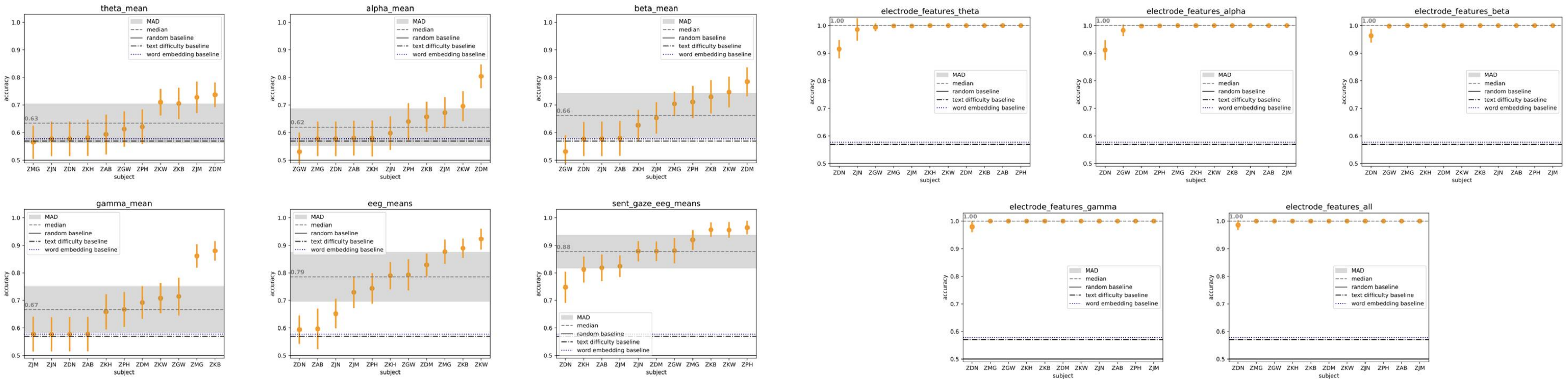


Toneva, Mariya, Otilia Stretcu, Barnabás Póczos, Leila Wehbe, and Tom M. Mitchell. "Modeling task effects on meaning representation in the brain via zero-shot meg prediction." Advances in Neural Information Processing Systems 33 (2020): 5284-5295.

Tasks affect processing

- Stimuli: sentences
- Task: searching for specific relations
- Stimulus representation: word embeddings
- Brain recording & modality: EEG, reading

Possible to predict whether a person is passively reading or performing a task with the text based on EEG recordings



Hollenstein, Nora, Marius Tröndle, Martyna Plomecka, Samuel Kiegeland, Yilmazcan Özyurt, Lena A. Jäger, and Nicolas Langer. "Reading task classification using EEG and eye-tracking data." arXiv preprint arXiv:2112.06310 (2021).

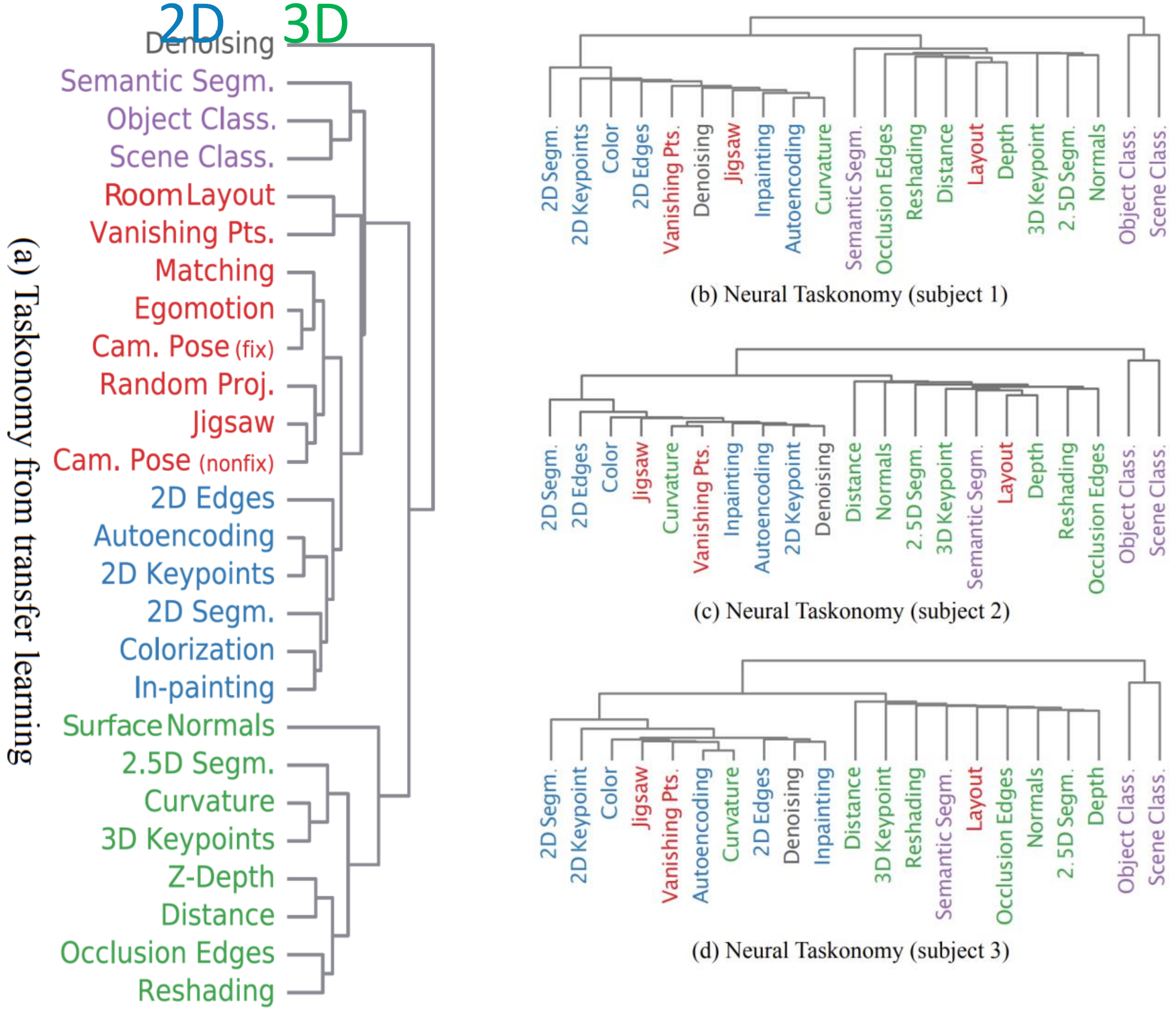
Tasks affect processing

- Stimuli: images of natural scenes
- Stimulus representation: task-optimized CNNs for a range of tasks
- Brain recording & modality: fMRI, vision

Vision tasks with higher transferability make similar predictions for brain responses from different regions

Semantic Low-dim. Geometric

2D 3D

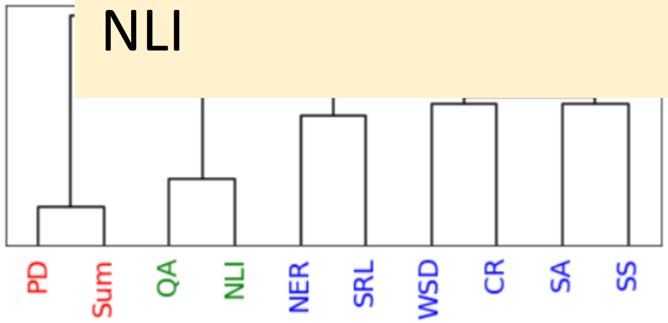


Wang, Aria, Michael Tarr, and Leila Wehbe. "Neural taskonomy: Inferring the similarity of task-derived representations from brain activity." Advances in Neural Information Processing Systems 32 (2019).

Tasks affect processing

- Stimuli: passages and narratives
- Stimulus representation: task-optimized NLP models for a range of tasks
- Brain recording & modality: fMRI, reading & listening of different stimuli

Reading fMRI best explained by coref. resolution, NER, shallow syntax parsing
Listening fMRI best explained by paraphrasing, summarization, NLI

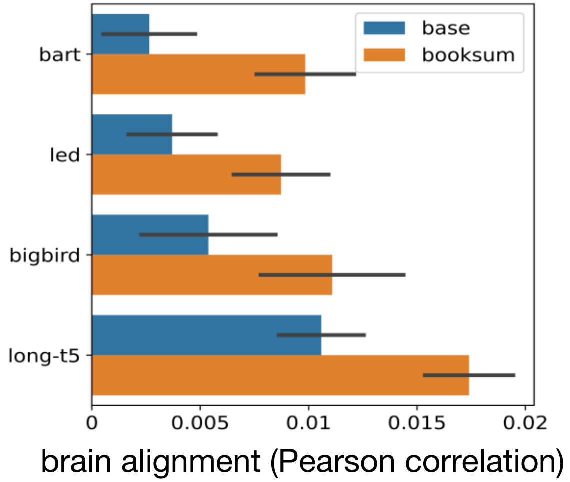


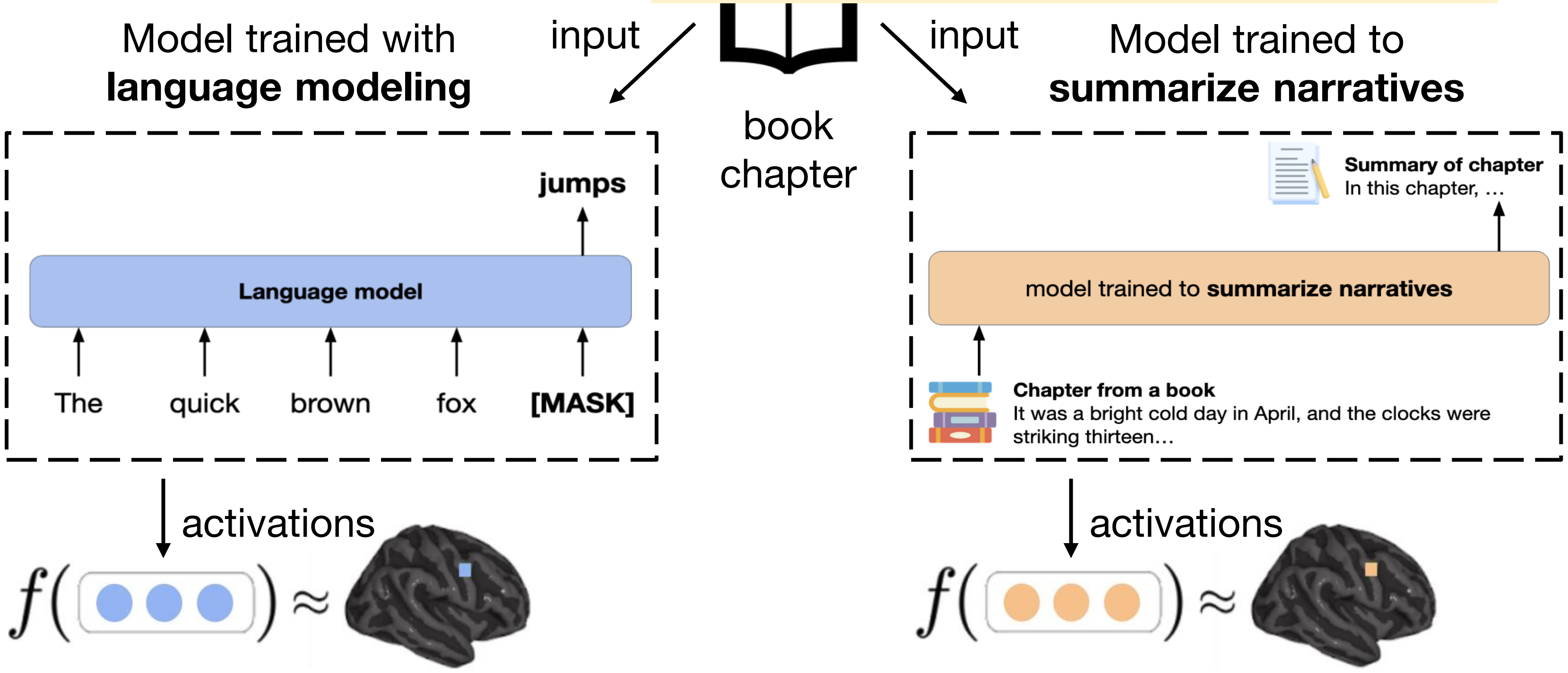
Oota, Subba Reddy, Jashn Arora, Veeral Agarwal, Mounika Marreddy, Manish Gupta, and Bapi Raju Surampudi. "Neural Language Taskonomy: Which NLP Tasks are the most Predictive of fMRI Brain Activity?." *arXiv preprint arXiv:2205.01404* (2022).

Tasks affect processing

- Stimuli: one chapter of Harry Potter
- Stimulus representation: summarization-optimized language models
- Brain recording & modality: fMRI, reading

Training language models to summarize narratives improves brain alignment, especially during important narrative elements (Characters, emotions, etc.)





[Aw, K.L., and Mariya Toneva. Training language models to summarize narratives improves brain alignment" ICLR 2023]

Challenges in using DL for cognitive modeling

- Not designed to specifically model brain processing
  - Training DL models using brain recordings
  - Task-based modeling
- Can be difficult to interpret due to multiple sources of information
  - **Disentangling contributions of different info sources to brain predictions**

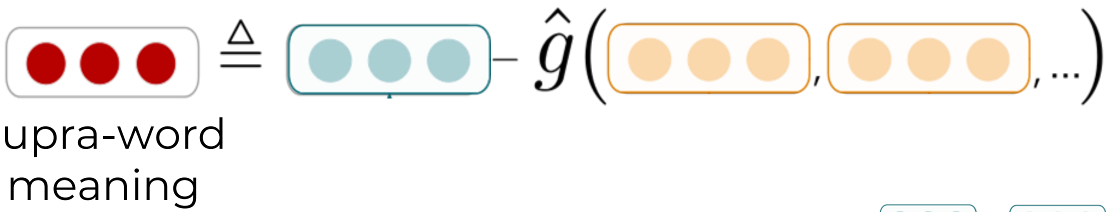
Disentangling contributions of different info sources to brain predictions

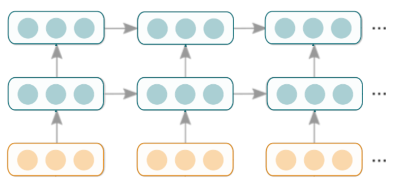
“Mary finished the apple”

supra-word meaning may contain concept of:

- eating
- apple core
- …

Isolating supra-word meaning is a type of intervention

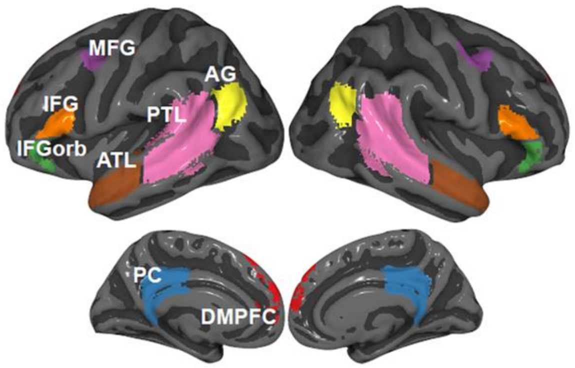


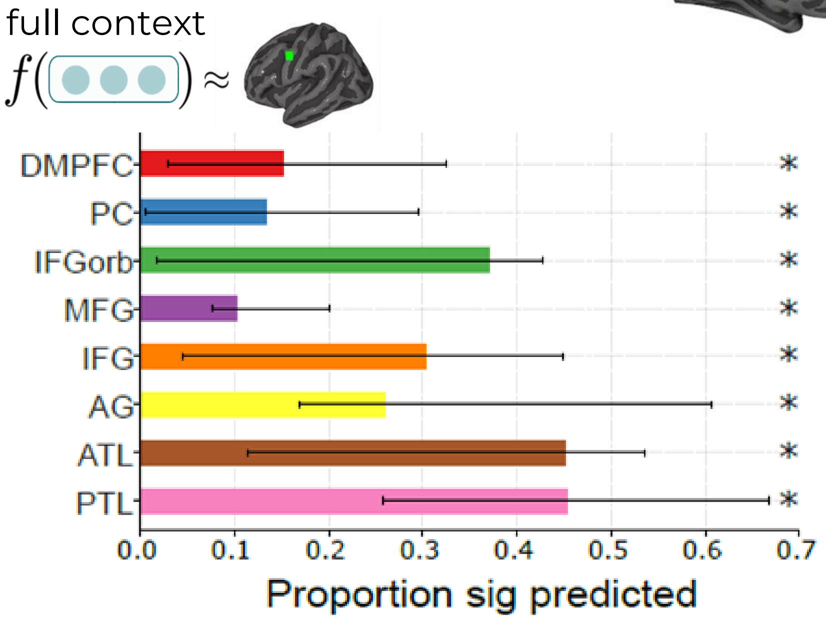


Toneva, Mariya, Tom M. Mitchell, and Leila Wehbe. "Combining computational controls with natural text reveals aspects of meaning composition." Nature Computational Science (2022)..

Disentangling contributions of different info sources to brain predictions

- Stimuli: one chapter of Harry Potter
- Stimulus representation: disentangled embeddings from pretrained NLP models
- Brain recording & modality: fMRI & MEG, reading





Bilateral PTL and ATL process supra-word meaning

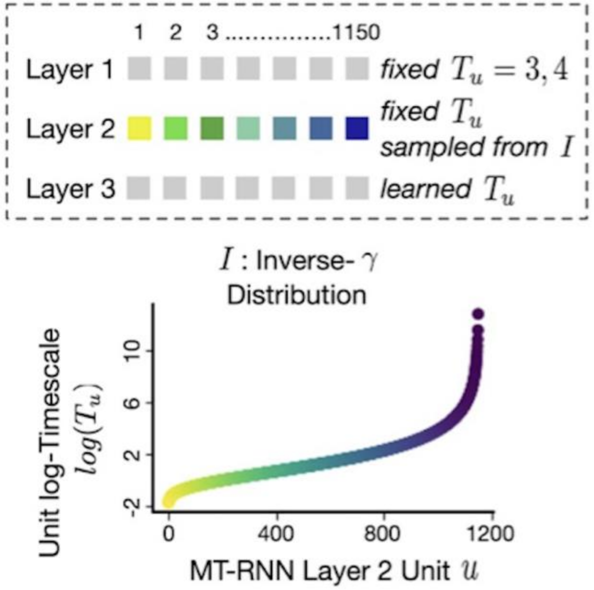
Word-level information important for prediction of most language regions

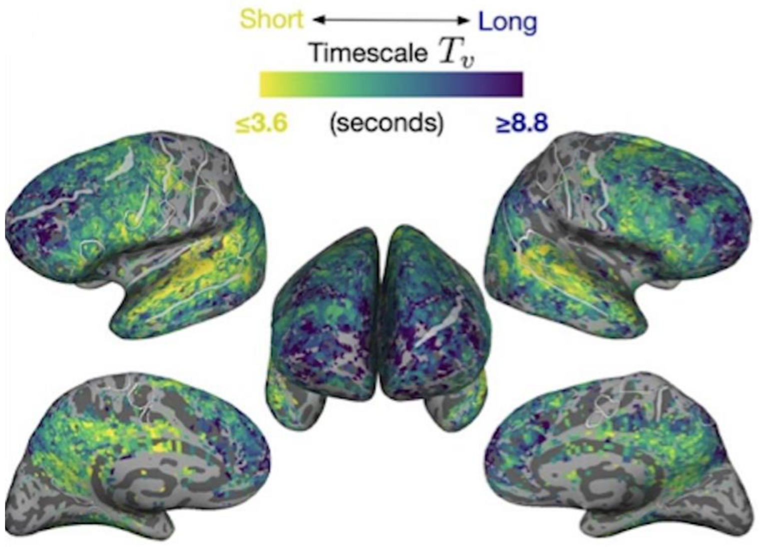
Toneva, Mariya, Tom M. Mitchell, and Leila Wehbe. "Combining computational controls with natural text reveals aspects of meaning composition." Nature Computational Science (2022)..

Disentangling contributions of different info sources to brain predictions

- Stimuli: story
- Stimulus representation: multi-timescale NLP model
- Brain recording & modality: fMRI, listening

Utilizing an NLP model that explicitly represents different timescale of information allows the voxel-wise estimation of the preferred timescales



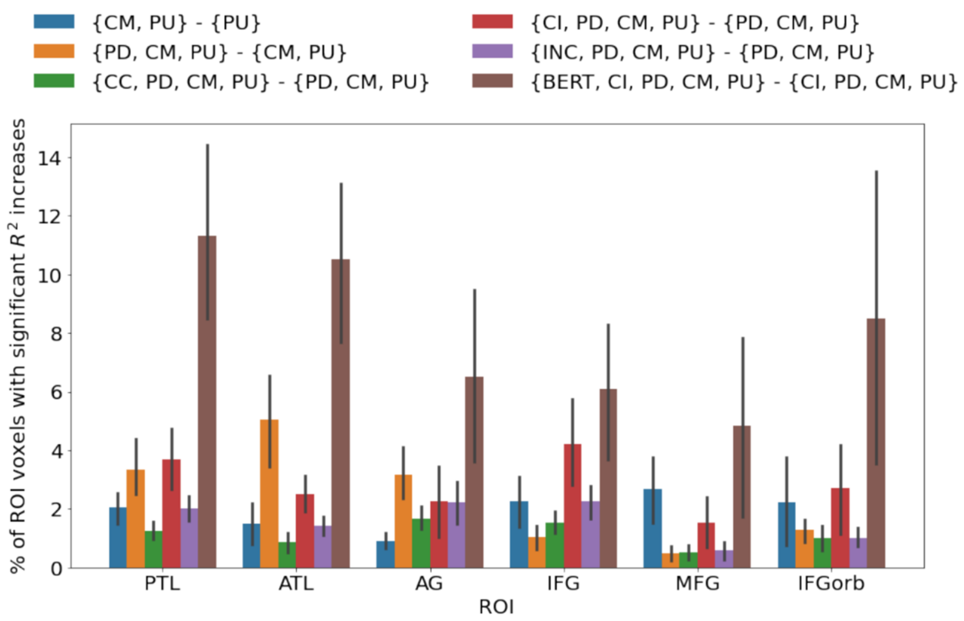


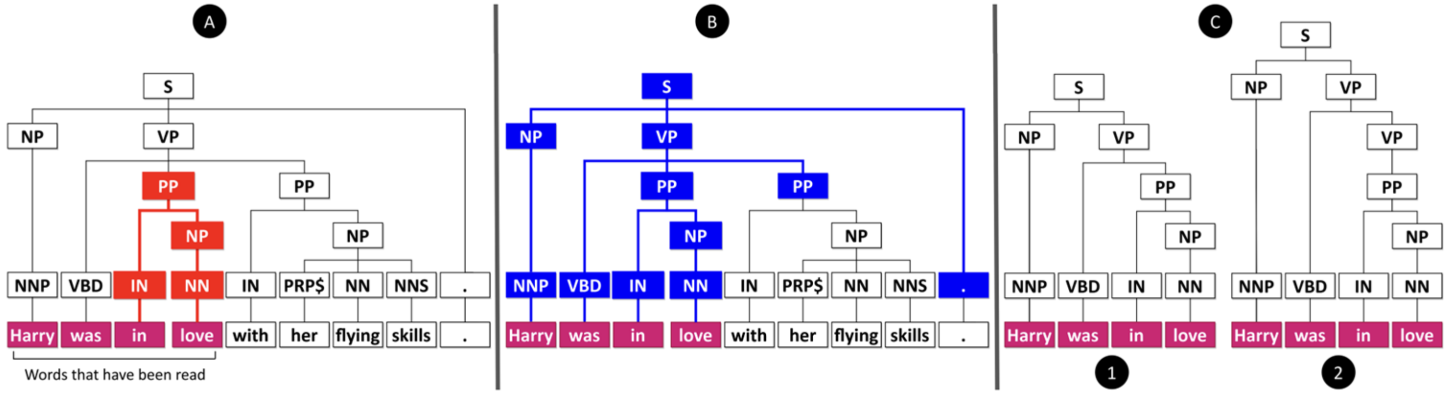
Figures provided by Shailee Jain

able multi-timescale models for predicting fMRI responses to continuous natural speech." Advances in Neural Information Processing Systems 33 (2020): 13738-

Disentangling contributions of different info sources to brain predictions

- Stimuli: one chapter of Harry Potter
- Stimulus representation: syntactic tree representations & pretrained NLP model
- Brain recording & modality: fMRI, reading





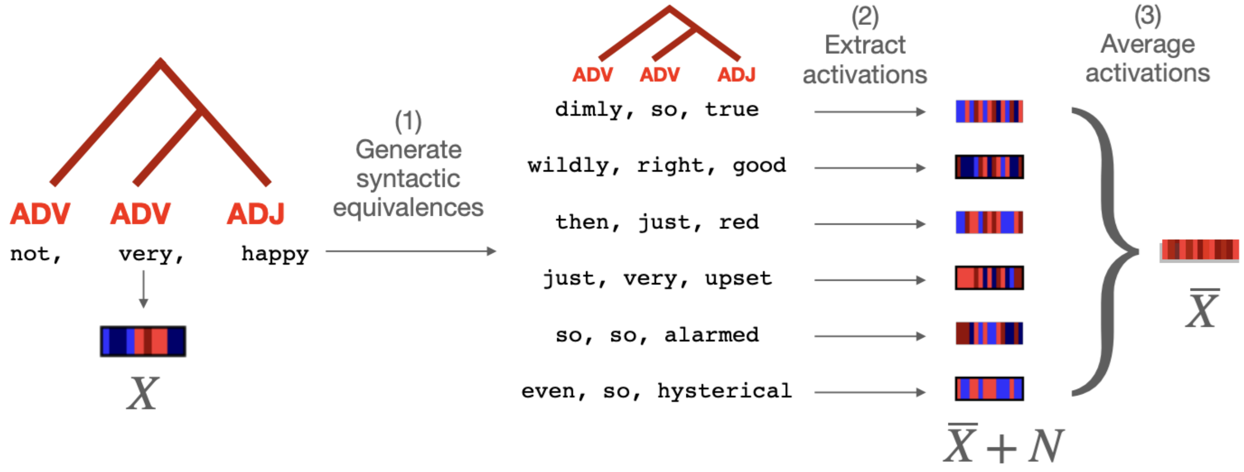
Syntactic structure-based features explain additional variance in language regions over complexity metrics

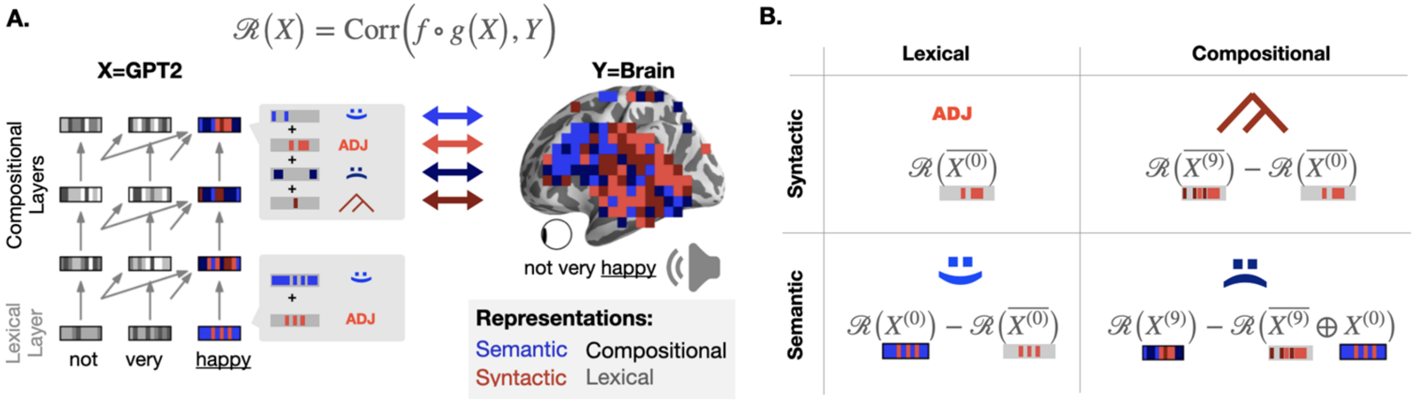
Regions predicted by syntactic and semantic are difficult to distinguish

Reddy, Aniketh Janardhan, and Leila Wehbe. "Can fMRI reveal the representation of syntactic structure in the brain?." Advances in Neural Information Processing Systems 34 (20

Disentangling contributions of different info sources to brain predictions

- Stimuli: story
- Stimulus representation: pretrained NLP models
- Brain recording & modality: fMRI, listening





Caucheteux, Charlotte, Alexandre Gramfort, and Jean-Remi King. "Disentangling syntax and semantics in the brain with deep networks." In *International Conference on Machine Lear*

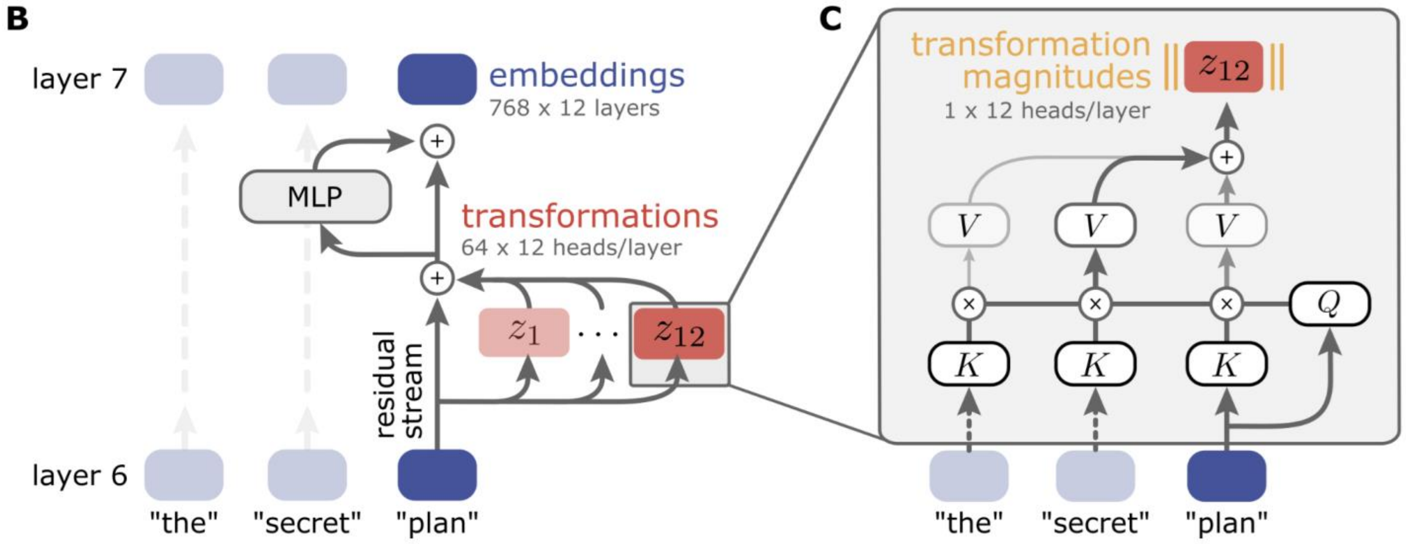
Compositional representations recruit a wider cortical network than word-level representations

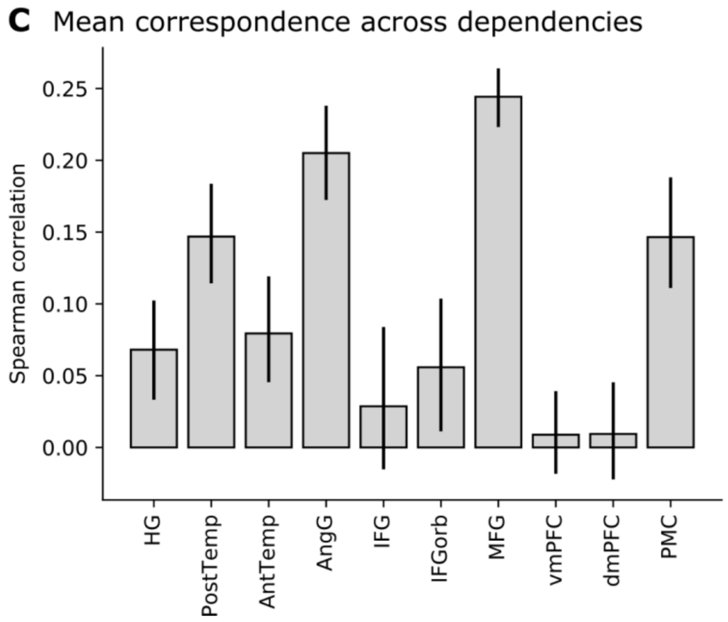
Syntax and semantics not associated with separate modules

Disentangling contributions of different info sources to brain predictions

- Stimuli: story
- Stimulus representation: pretrained NLP model
- Brain recording & modality: fMRI, listening

Decomposing NLP embeddings into attention heads reveals correlations between syntactic
m
fM

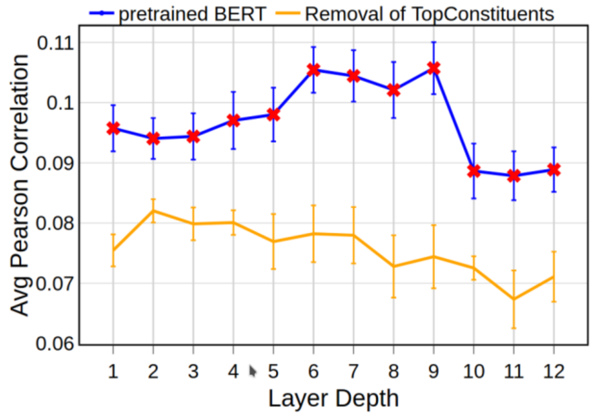


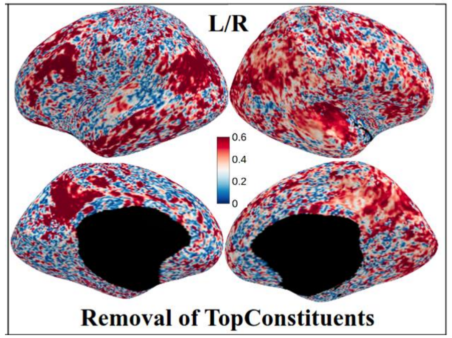


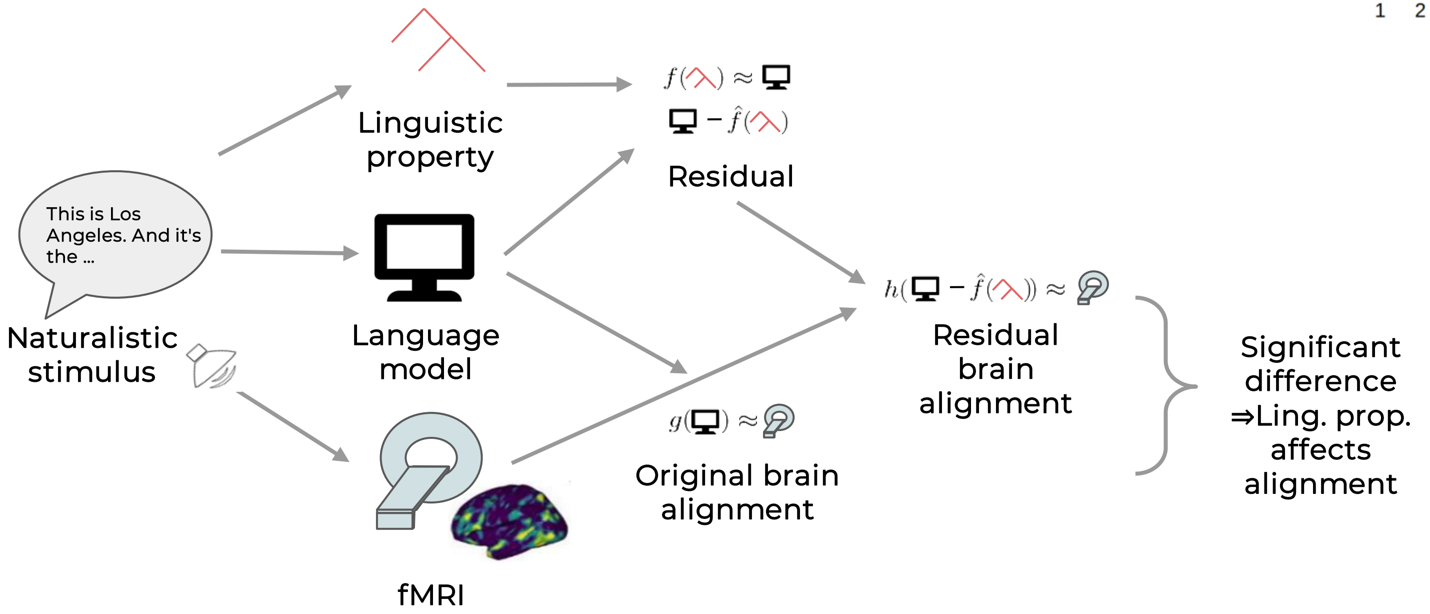
Kumar, Sreejan, Theodore R. Sumers, Takateru Yamakoshi, Ariel Goldstein, Uri Hasson, Kenneth A. Norman, Thomas L. Griffiths, Robert D. Hawkins, and Samuel A. Nastase. "Reconstructing the cascade of language processing in the brain using the internal computations of a transformer-based language model." bioRxiv (2022).

Disentangling contributions of different info sources to brain predictions

- Stimuli: story
- Stimulus representation: pretrained NLP model
- Brain recording & modality: fMRI, listening



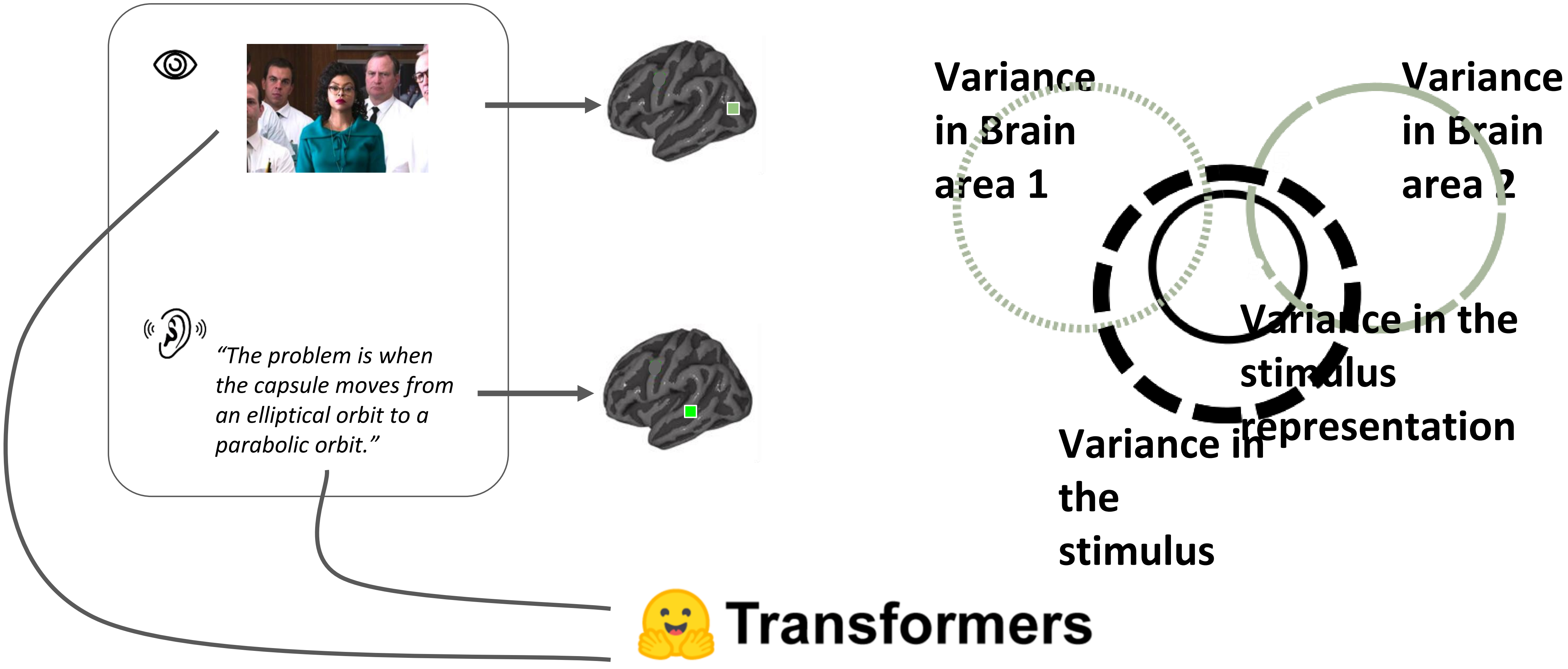




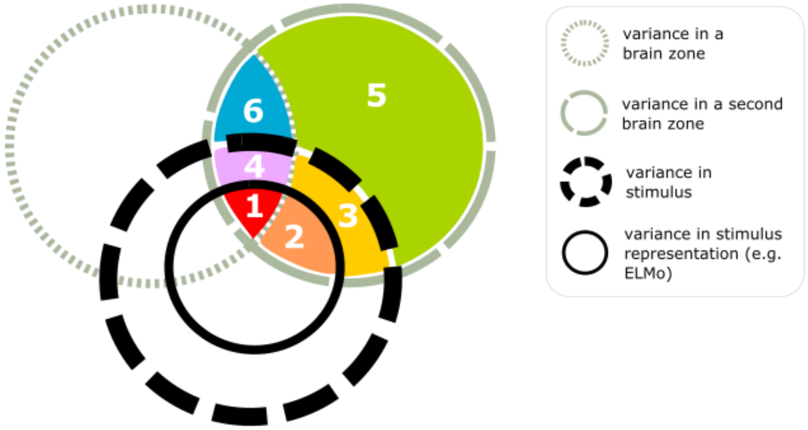
Syntactic properties contribute the most to the brain alignment trend across layers of language models

Oota, S., Manish Gupta, and Mariya Toneva. "Joint processing of linguistic properties in brains and language models" arXiv (2022).

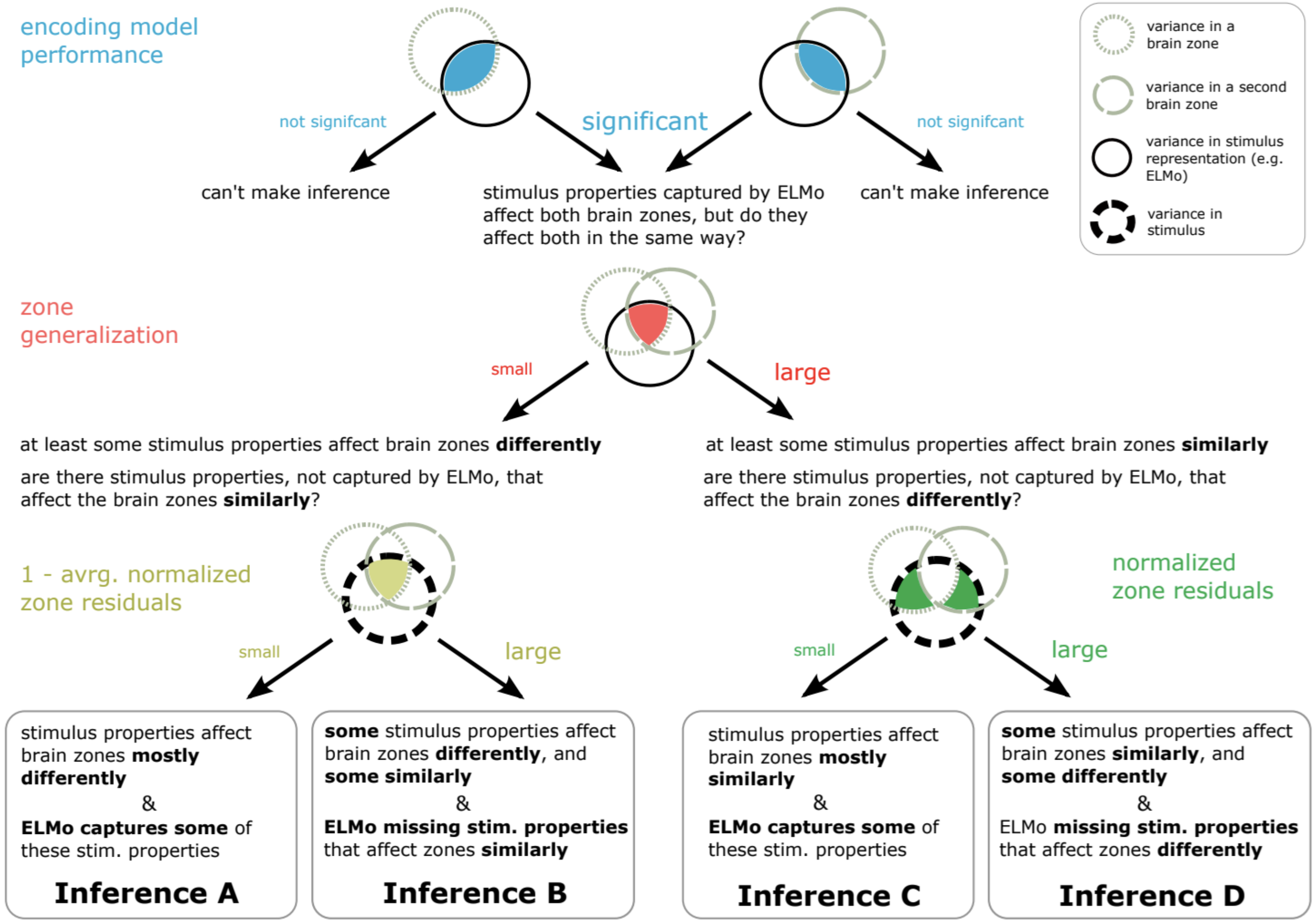
Complex stimulus representations make it difficult to infer the effect of a stimulus on multiple brain areas



Framework to determine whether a complex stimulus affects two brain areas in a similar way



1+4 similar effect
2+3 different effect
1 similar effect of stim. properties captured by ELMo
4 similar effect of stim. properties missing from ELMo
2 different effect of stim. properties captured by ELMo
3 different effect of stim. properties missing from ELMo
5 different noise
6 similar noise



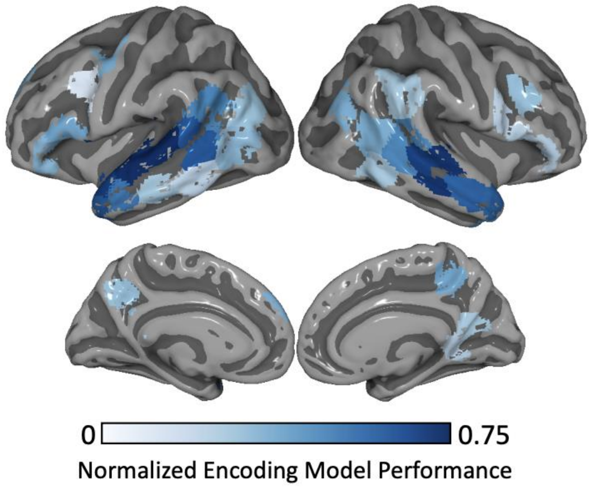
Toneva, Mariya, Jennifer Williams, Anand Bollu, Christoph Dann, and Leila Wehbe. "Same cause; different effects in the brain." Causal Learning and Reasoning (2022).

Framework reveals differences in processing across language network areas

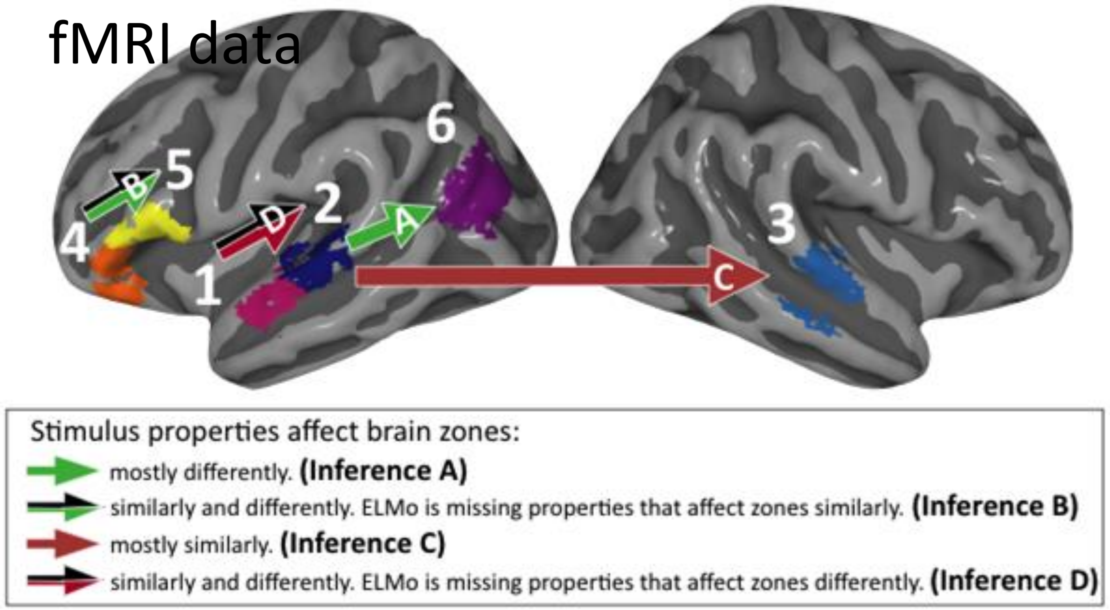
- Stimuli: movie
- Stimulus representation: pretrained NLP model
- Brain recording & modality: fMRI, view & listen

Framework reveals differences in processing across language network areas

Example of each type of effect in movie fMRI data



Encoding model perf. significant in all language areas



Toneva, Mariya, Jennifer Williams, Anand Bollu, Christoph Dann, and Leila Wehbe. "Same cause; different effects in the brain." Causal Learning and Reasoning (2022).

Challenges in using DL for cognitive modeling

- Not designed to specifically model brain processing
  - Training DL models using brain recordings
  - Task-based modeling
- Can be difficult to interpret due to multiple sources of information
  - Disentangling contributions of different info sources to brain predictions

Deep Neural Networks and Brain Alignment: Brain Encoding and Decoding

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

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

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




Agenda

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

Outline

1. Summary
2. Future trends

Summary

- Exciting times: publicly accessible neuroimaging data of various tasks starting to be avaliable now!
  - Opportunities:
    - Data ahead of theory, so it's an open field for theoretical and methodological innovation!
    - Encoding models can be interpreted as process models constraining brain-computational theories (Kriegeskorte and Douglas, 2019).
    - Decoding models serve as a test for the presence of information in neural responses (Karamolegkou et al., 2023)
    - Decoding is relevant for cognitive neuroscientists interested in how semantic information is represented in the brain.
    - Computational linguists are interested in the cognitive plausibility of distributional models. (Minnema & Herbelot, ACL 2019)
    - DL is helpful in uncovering patterns in brain responses and may lead to theories of information organization in the brain.

  - Challenges:
    - Hypothesis-driven data collection might be more helpful
    - Individual variability is the norm in neuroimaging data!
    - Neuroimaging data is more complex, noisy as compared to classical datasets used by DL researchers

Summary

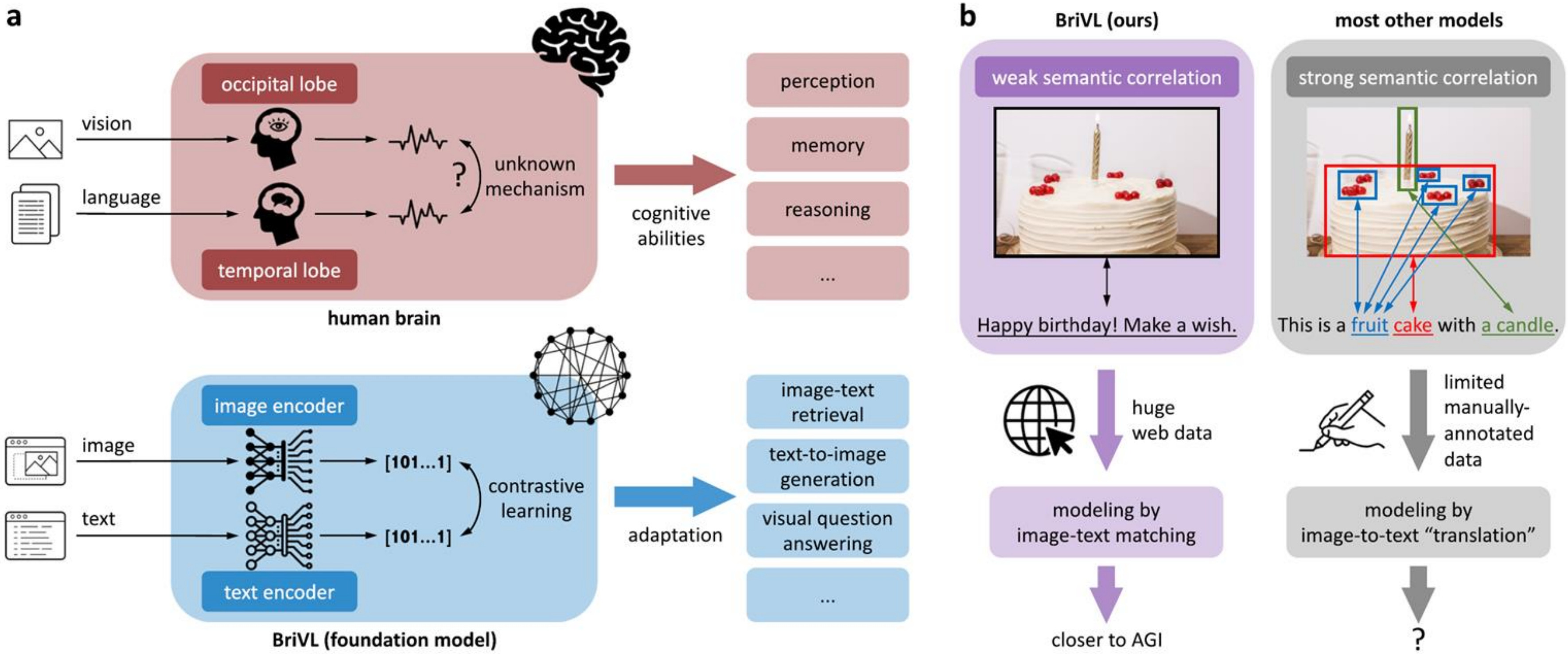
- This Tutorial:
- Stimulus representation schemes
  - Vision: CNN-based
  - Language: Transformer-based
- Datasets available (Reading/Listening/Viewing tasks in EEG, MEG, fMRI)
- Decoding
  - Word-level Universal Brain Decoder; Continuous Lang Decoding; Multi-view and Cross-view Decoding
- Encoding
  - Classical findings; More recent DL-based models
- Advance methods
  - Tuning/Training DL models using brain recordings
  - Task-based modeling

Outline

1. Summary
2. Future trends: DNNs & The Brain

DNNs & The Brain: Multi-modal, Multi-task

- Brain response to a stimulus is multi-modal, multi-task related
    - Cross-view and multi-view decoding (Oota et al 2022a)
    - Visio-linguistic encoding (fusion of vision and language information) (Oota et al 2022b)
    - Task-based representations give better brain alignment (Neural Taskonomy: Oota et al 2022c)
    - Multimodal foundation model (Fei et al 2022)

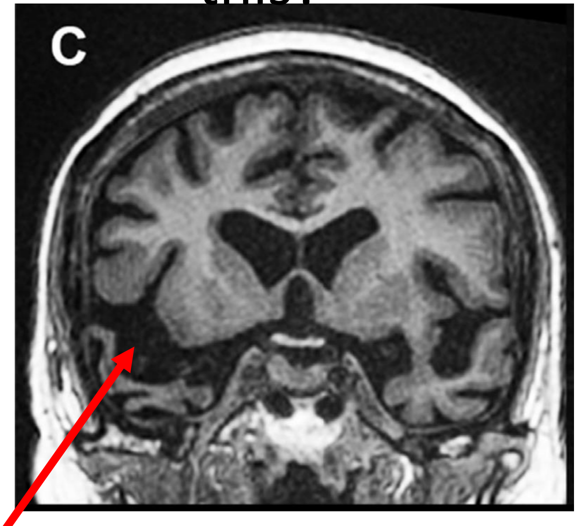


Fei, Lu, Gao et al (2022). Towards artificial general intelligence via a multimodal foundation model. *Nature Communications* 13:3094
doi.org/10.1038/s41467-022-30761-2

DNNs & Brain Damage

- DL models of encoding and decoding have not yet been put through the brain-damage experiments. Ex. Semantic Dementia

Animal habitat task.
The patient is asked:
Where would you find this?



Rt Ant Temporal Lobe Damage (Patient 8)

Stimulus	Response
FROG	In water
COW	On a farm
DUCK	On ponds. I see them on the river when I go walking.
SQUIRREL	In the woods, in the country. They are wild.
MONKEY	In trees, in Africa.

Do DL Models exhibit such degradation with damage to units?

Snowden, Harris, Thompson, Kobylecki, Jones, Richardson, Neary (2018). Semantic dementia and the left and right temporal lobes, *Cortex*, 107(188-203). https://doi.org/10.1016/j.cortex.2017.08.024.

Multilinguality

- How do multilingual participants represent information?
  - Different language families and typologies (verb-framed vs satellite
  - Multiple scripts
- How do brain activations align to modern LLMs that perform language translation among multiple languages apparently seamlessly?
- Bi/Multilingual Advantage and what does it mean for DL models?
  - studies have shown superior executive function (inhibitory control), memory in multilingual participants
  - Potential representational differences in simultaneous and sequential multilinguals
- Link between Language and Cognition
- What can DL models contribute to Bi/Multilingual Literature?

A big thank you!

Tutorial, Code and Material:

Material from IJCAI 2023 Tutorial would be uploaded soon!

(Past): Deep Learning for Brain Encoding and Decoding, Cogsci-2022

https://tinyurl.com/DL4Brain

(Past): Language and the Brain: Deep Learning for Brain Encoding and Decoding, IJCNN 2023

https://tinyurl.com/DLBrainIJCNN2023

Thanks!

- Questions
  - subba-reddy.oota@inria.fr
  - gmanish@microsoft.com
  - raju.bapi@iiit.ac.in
  - mtoneva@mpi-sws.org
- Connect with us:
  - https://www.linkedin.com/in/subba-reddy-oota-11a91254/
  - http://aka.ms/manishgupta, https://sites.google.com/view/manishg/
  - https://sites.google.com/view/bccl-iiith/home
  - http://www.mtoneva.com