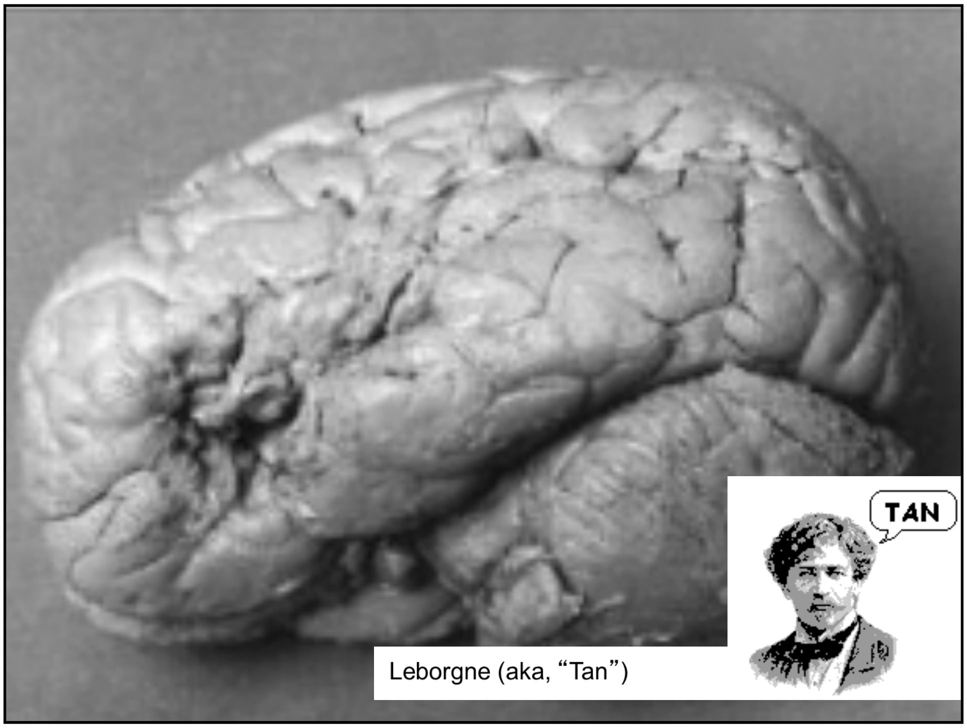
Language in the Brain

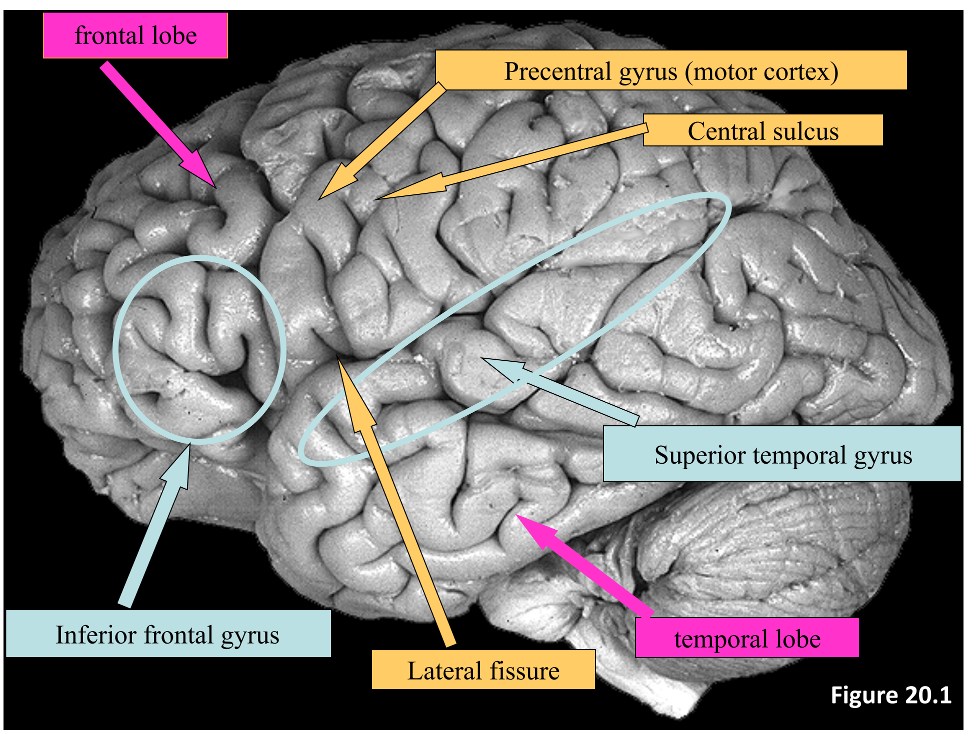
I. Lesions: Broca's and Wernicke's areas
II. Neural representation of meaning
III. Bilinguals and sign language

Paul Broca

(1824-1880)

Aphasia:
"The collective deficits in language comprehension and production that accompany neurological damage"

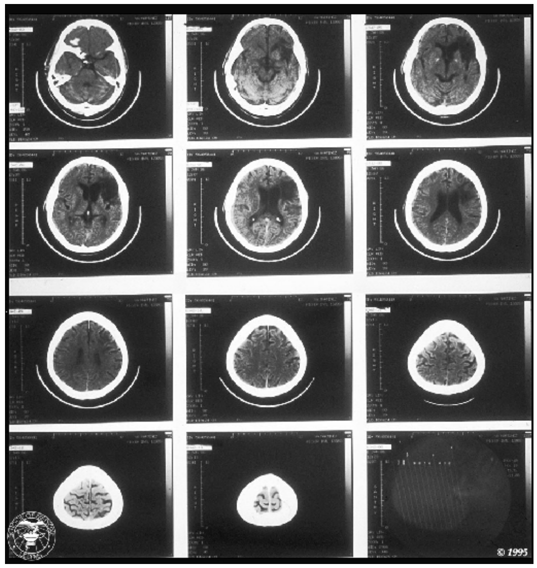

TAN
Leborgne (aka, "Tan")


frontal lobe
Precentral gyrus (motor cortex)
Central sulcus
Superior temporal gyrus
Inferior frontal gyrus
Lateral fissure
temporal lobe
Figure 20.1

(CT scan: computer
assisted tomography)

rostral (anterior)

right



left

Caudal (posterior)

Only damage in the left hemisphere
results in aphasia:

Paul Broca (1864)

**"Nous parlons avec
l' hemisphere gauche!"**

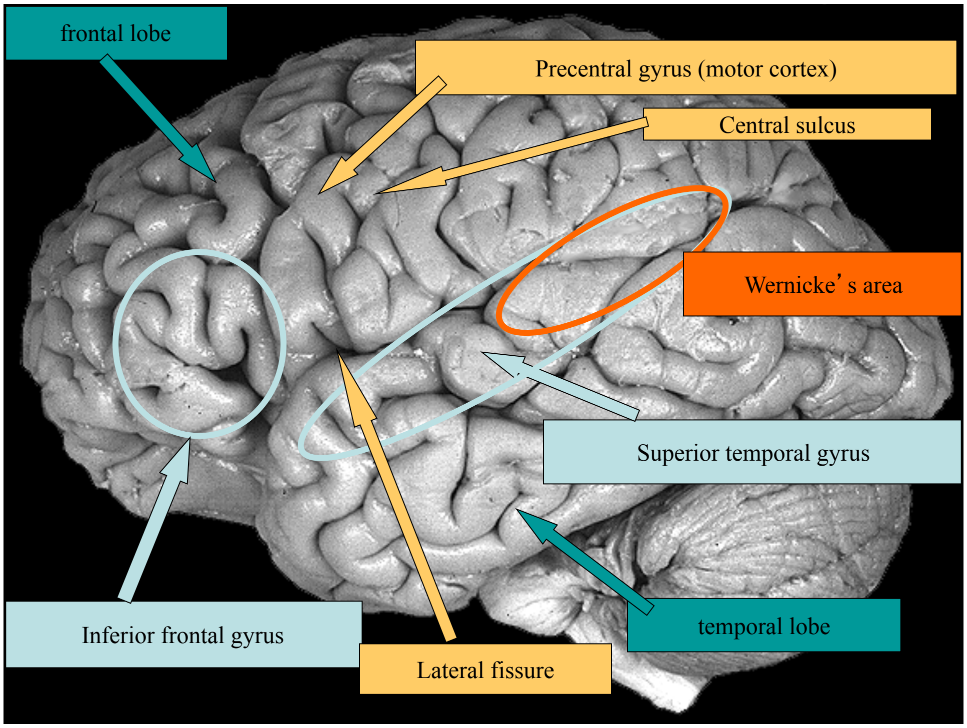
Carl Wernicke
(1848-1904)

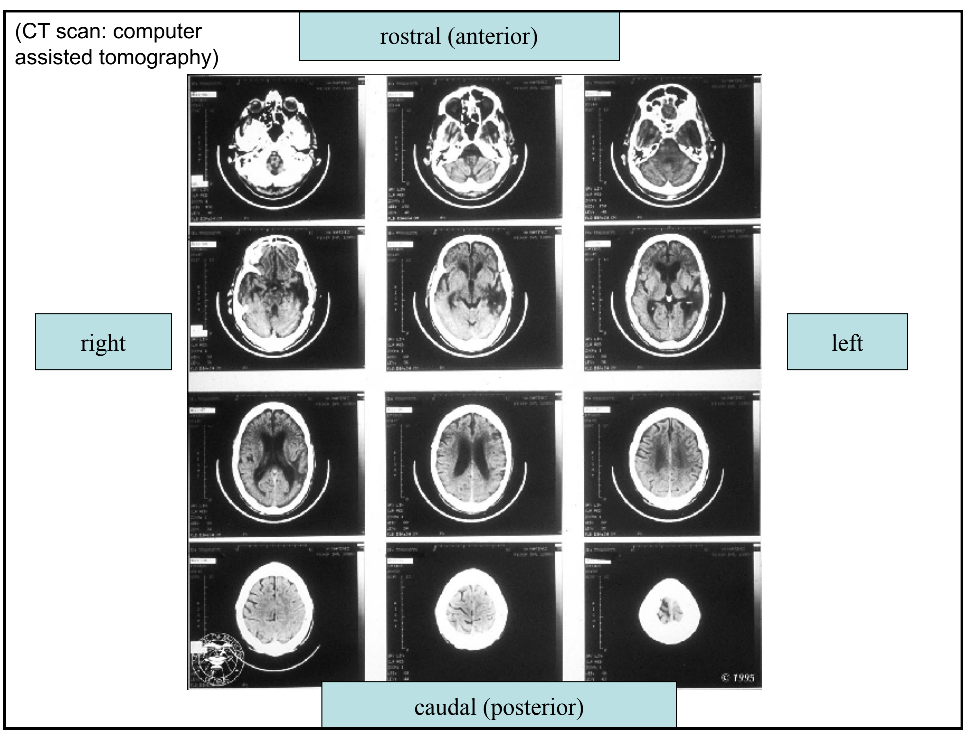
Carl Wernicke
(1848–1905)

“I called my mother on the television and did not understand the door. It was not too breakfast, but they came from far to near. My mother is not too old for me to be young.”

Wernicke’s aphasia

Fluent speech, but nonsensical; loss of ability to understand language


frontal lobe
Precentral gyrus (motor cortex)
Central sulcus
Wernicke's area
Superior temporal gyrus
Inferior frontal gyrus
temporal lobe
Lateral fissure


(CT scan: computer assisted tomography)
rostral (anterior)
right
left
caudal (posterior)
© 1995

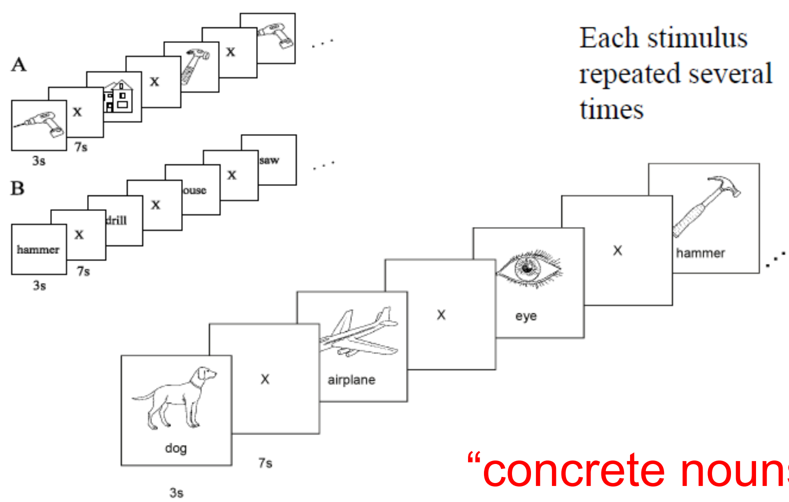
Language in the Brain

I. Lesions: Broca's and Wernicke's areas
II. Neural representation of meaning
III. Bilinguals and sign language

fMRI experiment

Tom Mitchell et al

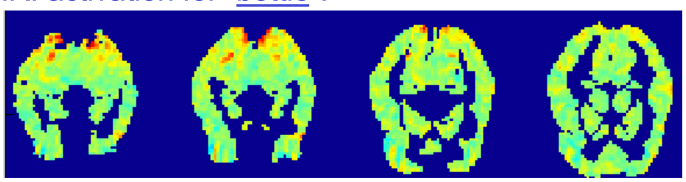
Typical stimuli



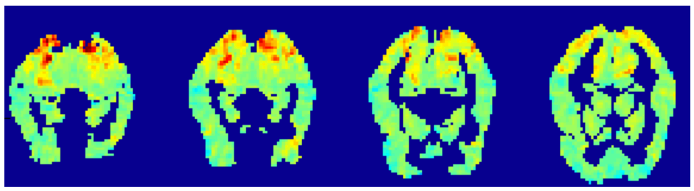
Each stimulus repeated several times

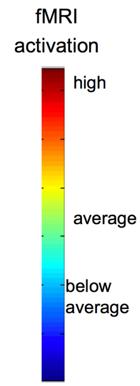
"concrete nouns"

fMRI activation for "bottle":

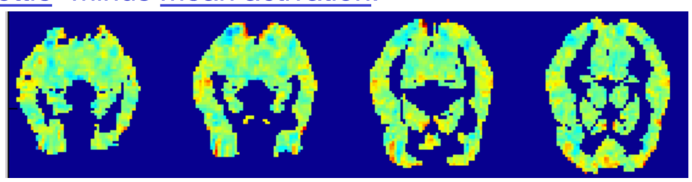


Mean activation averaged over 60 different stimuli:





"bottle" minus mean activation:

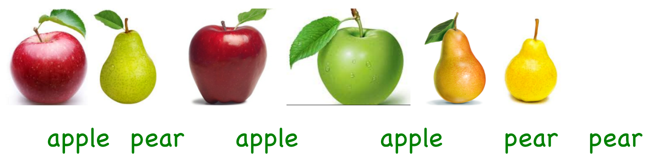


Neuroscience Research Questions

- Can we observe differences in neural activity as people think about <u>different concepts</u>?
- Is the neural activity that represents concepts <u>localized or distributed</u>?
- Are neural representations <u>similar across people</u>?
- Can we discover <u>underlying principles</u> of neural representations? (e.g., are representations built up from more primitive components?)

Classifiers

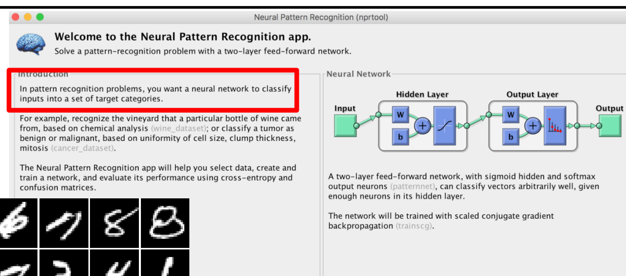
- A classifier accepts a description of an object and predicts what class it belongs to
- Supervised learning works from labeled training examples:

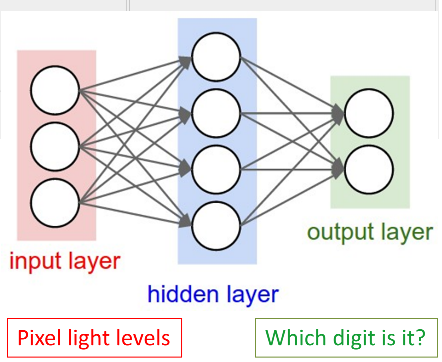


and then tests on new examples:

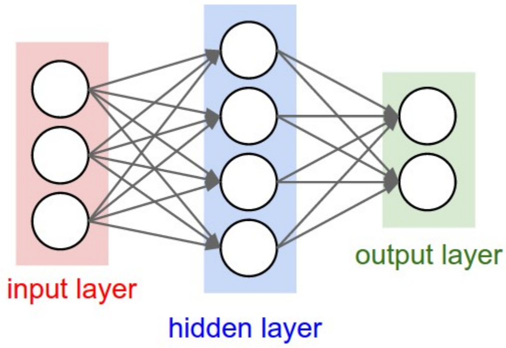
- Approach:
  - Describe each example by values for a set of features
    <color, size, shape, has-stem, has-leaf, texture,…>
    <red, small, round, yes, yes, smooth,…>
  - Training examples must be different from testing examples

Lab 4: ANNs as classifiers





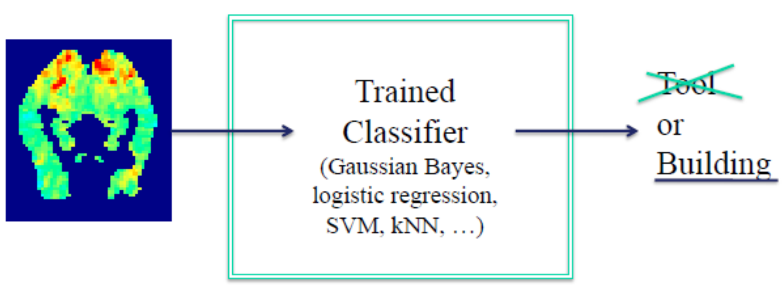
Neural Network Classifier as



- Model of how the brain perceives
- Tool for applications like face recog, navigation
- **Tool for seeing what information is in an experimentally measured neural signal**

Q1: Can one classify mental state from fMRI images?

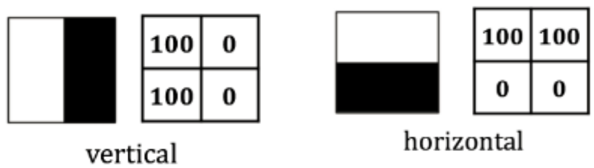
Given 84 nouns, present word, and capture fMRI data
Training example is fMRI output and presented word
Train on 83 and then test on 1 (repeat 84 times)



(classifier as virtual sensor of mental state)

Remember Lab 5: template matching for face recognition

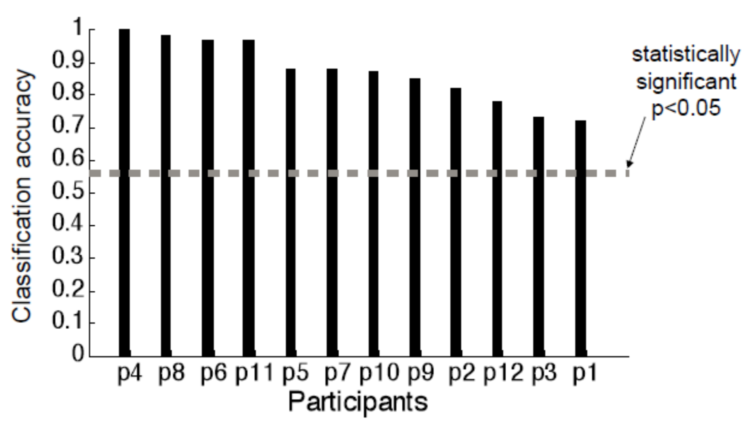
Consider a very simple example where we have two known image patterns corresponding to a vertical or horizontal edge, as shown below:



Suppose we are given a "mystery" image and want to determine whether it has a vertical or horizontal edge pattern:

Classification task: is person viewing a "tool" or "building"?

I.e. Compare current fMRI activation pattern to average "tool" pattern and average "building" pattern—choose whichever "template" it is closer to.

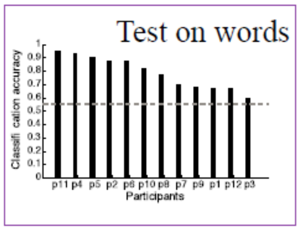


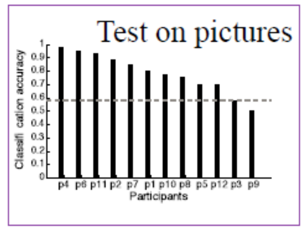
But is it learning just the *appearance* of the stimulus (the letter sequence) or its *meaning*?

Question 2: Is our classifier capturing neural activity encoding stimulus <u>meaning</u> or <u>appearance</u>?

Can we train on word stimuli, then decode picture stimuli?

<u>YES</u>: We can train classifiers when presenting English words, then decode category of picture stimuli, or Portuguese words



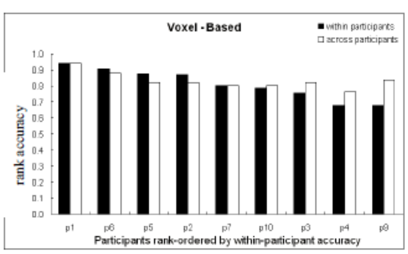


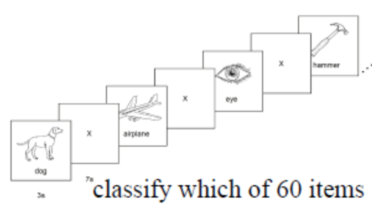
Therefore, the learned neural activation patterns must capture how the brain represents the <u>meaning</u> of input stimulus

Question 3: Are representations similar across people?

Can we train classifier on data from a collection of people, then decode stimuli for a new person?

<u>YES</u>: We can train on one group of people, and classify fMRI images of new person





Therefore, seek a theory of neural representations common to all of us (and of how we vary)

Question 4: Can we discover underlying principles of neural representations?

I.e. can activation patterns for novel words be predicted as combinations of known feature-related activity patterns?

Is the neural code for language "compositional"?

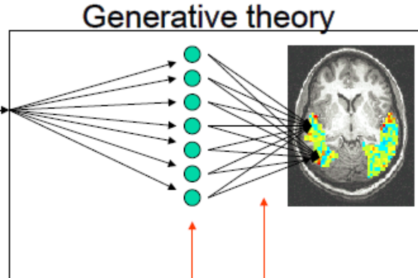
arbitrary word → **Generative theory** of word representation → predicted brain activity

Idea: Predict neural activity from corpus statistics of stimulus word

[Mitchell et al., *Science*, 2008]

Generative theory

"telephone" →



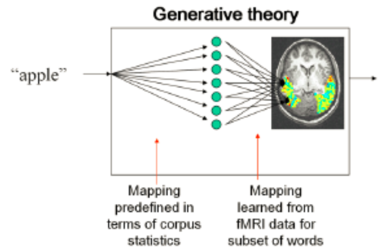
→ predicted activity for "telephone"

Statistical features from a trillion-word text corpus

Mapping learned from fMRI data

I.e. define a limited number of "**semantic features**" to characterize each word by its **set of feature weights**—coordinates in **semantic feature space**!

Which corpus statistics?

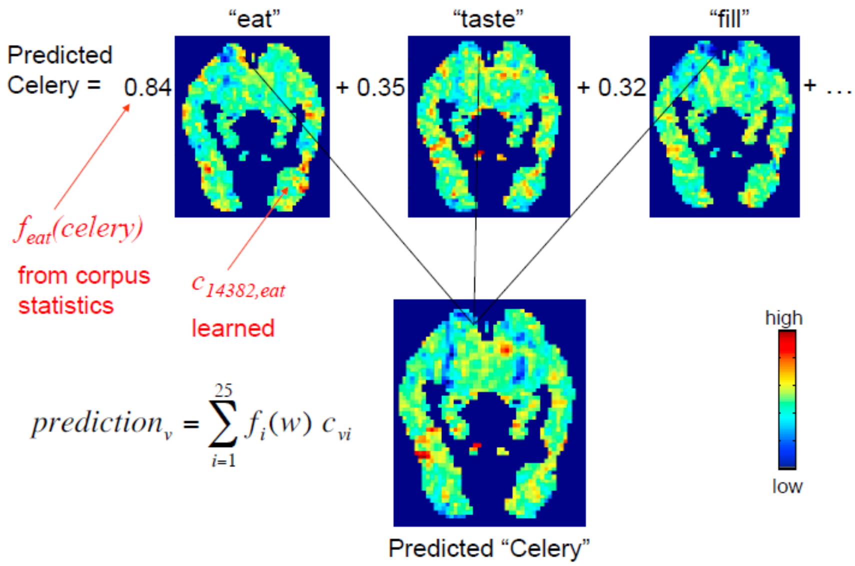


- Feature *i* = co-occurrence frequency of stimulus noun with verb *i*
- The model uses 25 verbs:
  - Sensory: *see, hear, listen, taste, touch, smell, fear,*
  - Motor: *rub, lift, manipulate, run, push, move, say, eat,*
  - Abstract: *fill, open, ride, approach, near, enter, drive, wear, break, clean*

Semantic feature values: "**celery**"
0.8368, eat
0.3461, taste
0.3153, fill
0.2430, see
0.1145, clean
0.0600, open
0.0586, smell
0.0286, touch
…
…
0.0000, drive
0.0000, wear
0.0000, lift
0.0000, break
0.0000, ride

Semantic feature values: "**airplane**"
0.8673, ride
0.2891, see
0.2851, say
0.1689, near
0.1228, open
0.0883, hear
0.0771, run
0.0749, lift
…
…
0.0049, smell
0.0010, wear
0.0000, taste
0.0000, rub
0.0000, manipulate

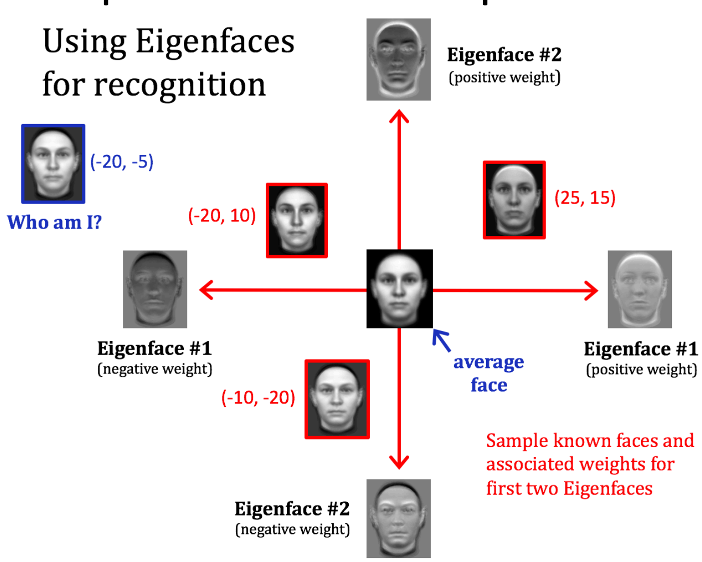
Predicted Activation is Sum of Feature Contributions



$$prediction_v = \sum_{i=1}^{25} f_i(w)\, c_{vi}$$

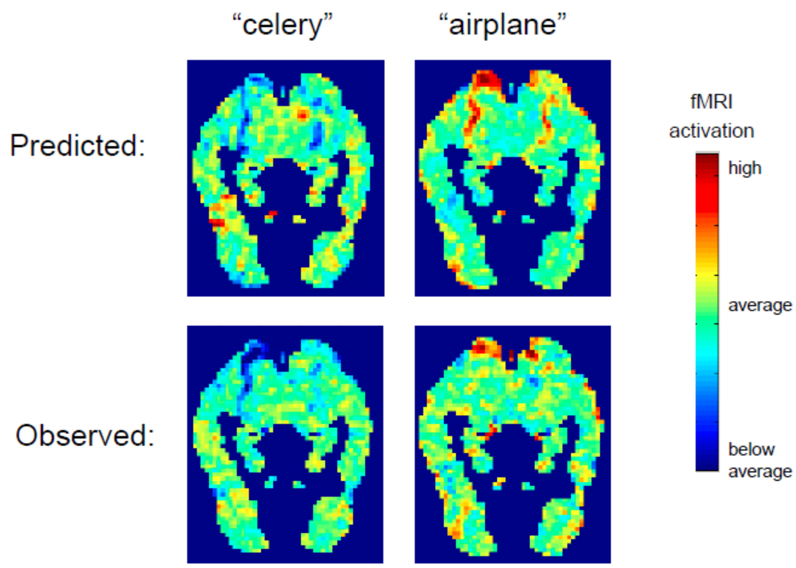
Remember LAB 3: Eigenfaces are Principal Components of Face Space

Using Eigenfaces for recognition



Represent each face Image by a set of Eigenface **weights**.

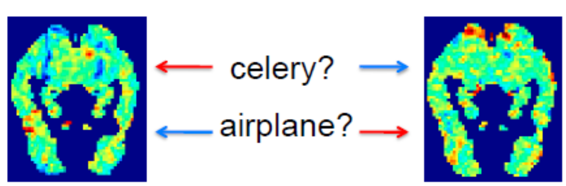
→"Dimensionality Reduction:" Many fewer weights Than pixels!



Predicted and observed fMRI images for "celery" and "airplane" after training on 58 other words.

Evaluating the Computational Model

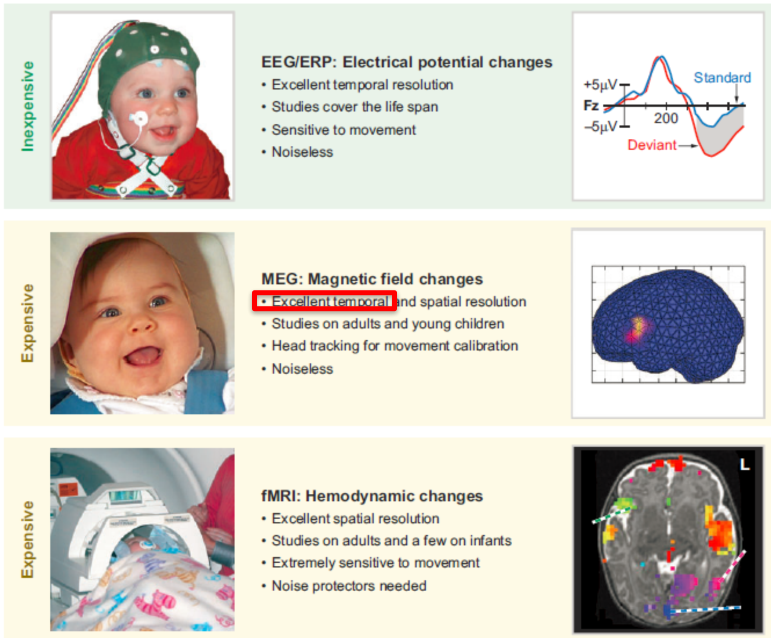
- Train it using 58 of the 60 word stimuli
- Apply it to predict fMRI images for other 2 words
- Test: show it the observed images for the 2 held-out, and make it predict which is which



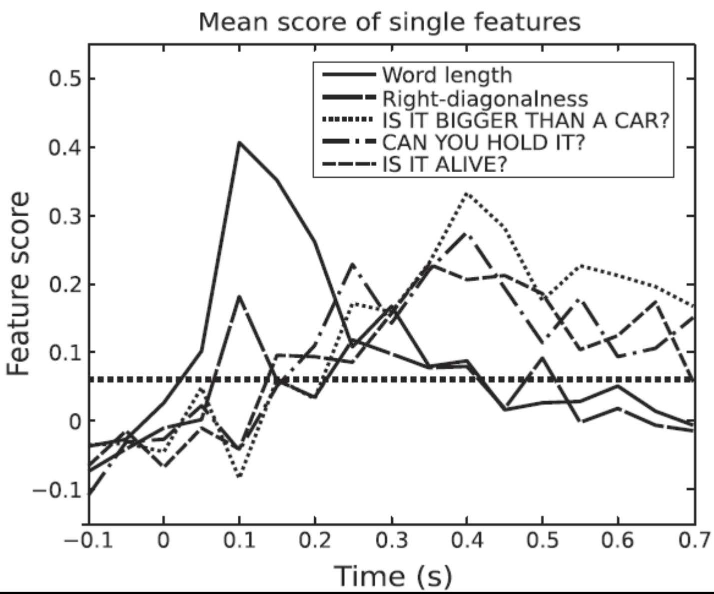
1770 test pairs in leave-2-out:
- Random guessing → 0.50 accuracy
- Accuracy above 0.61 is significant ($p<0.05$)

Mean accuracy over 9 subjects: 0.79

Neuroscience techniques used with infants



MEG experiment to determine WHEN info about specific features appears in the brain: sliding window classification analysis



Perceptual features first, semantic features later

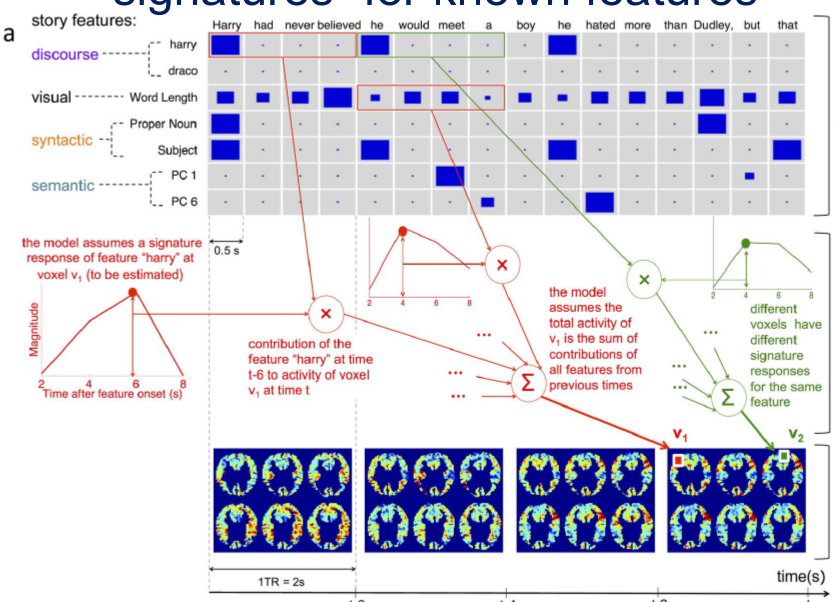
Sudre 2012

Simultaneously Uncovering the Patterns of Brain Regions Involved in Different Story Reading Subprocesses

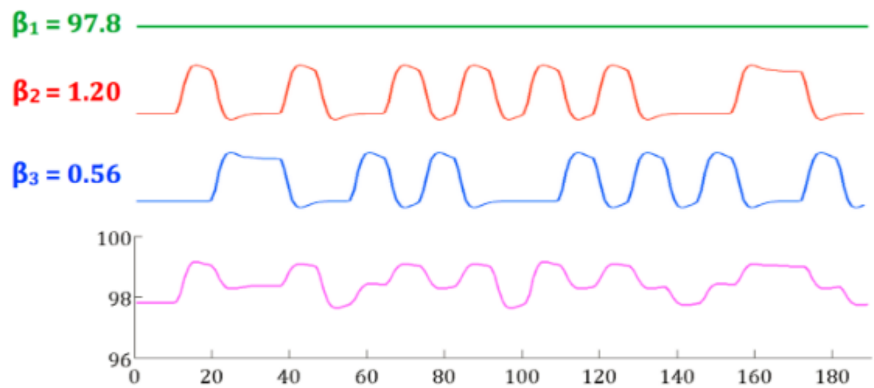
Abstract

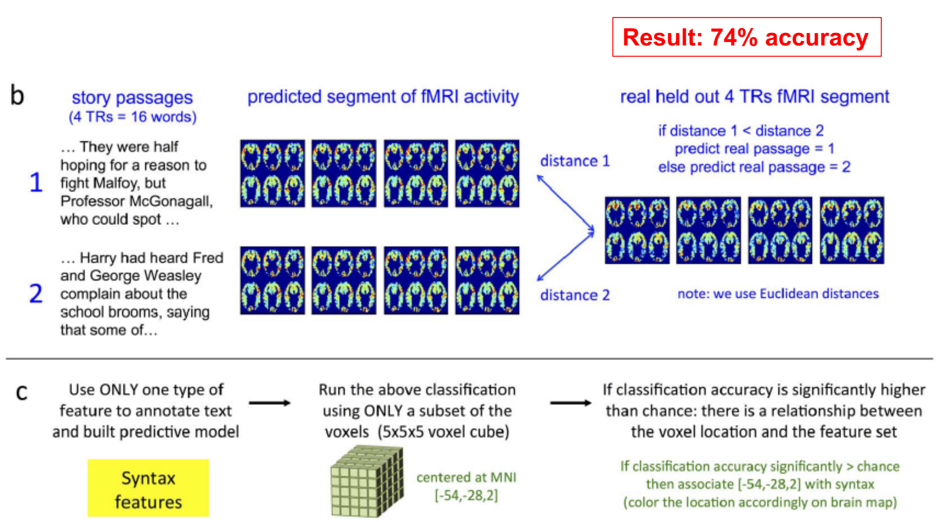
Story understanding involves many perceptual and cognitive subprocesses, from perceiving individual words, to parsing sentences, to understanding the relationships among the story characters. We present an integrated computational model of reading that incorporates these and additional subprocesses, simultaneously discovering their fMRI signatures. Our model predicts the fMRI activity associated with reading arbitrary text passages, well enough to distinguish which of two story segments is being read with 74% accuracy. This approach is the first to simultaneously track diverse reading subprocesses during complex story processing and predict the detailed neural representation of diverse story features, ranging from visual word properties to the mention of different story characters and different actions they perform. We construct brain representation maps that replicate many results from a wide range of classical studies that focus each on one aspect of language processing and offer new insights on which type of information is processed by different areas involved in language processing. Additionally, this approach is promising for studying individual differences: it can be used to create single subject maps that may potentially be used to measure reading comprehension and diagnose reading disorders.

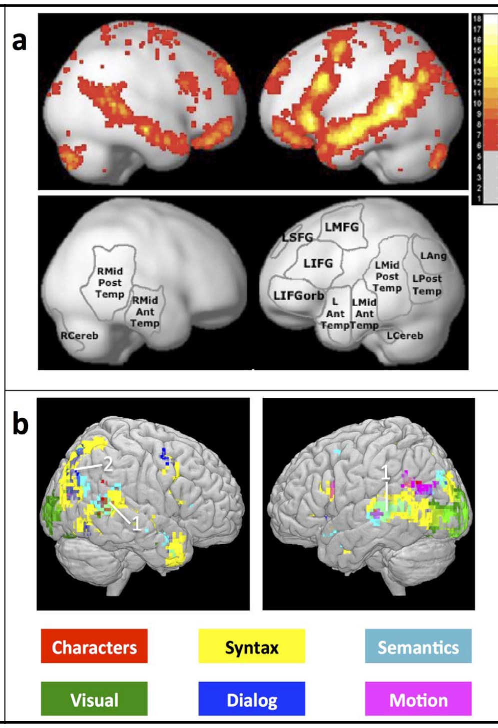
Predict fMRI activation by adding up "signatures" for known features



Lab 8: General Linear Model!







Sentences > nonwords

RESULTS: Different info in different brain regions