Object recognition

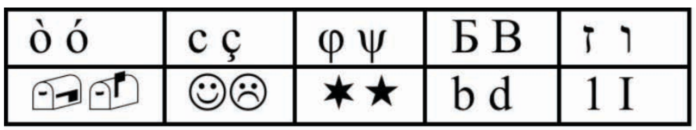
Gabriel Kreiman

http://kreiman.hms.harvard.edu

gabriel.kreiman@tch.harvard.edu

Four key features of visual object recognition

1. Selectivity



2. Tolerance (scale, rotation, etc.)

3. Speed (Potter 1969, Thorpe 1996)

4. Capacity (Standing 1973, Brady 2008)

Why visual shape recognition?

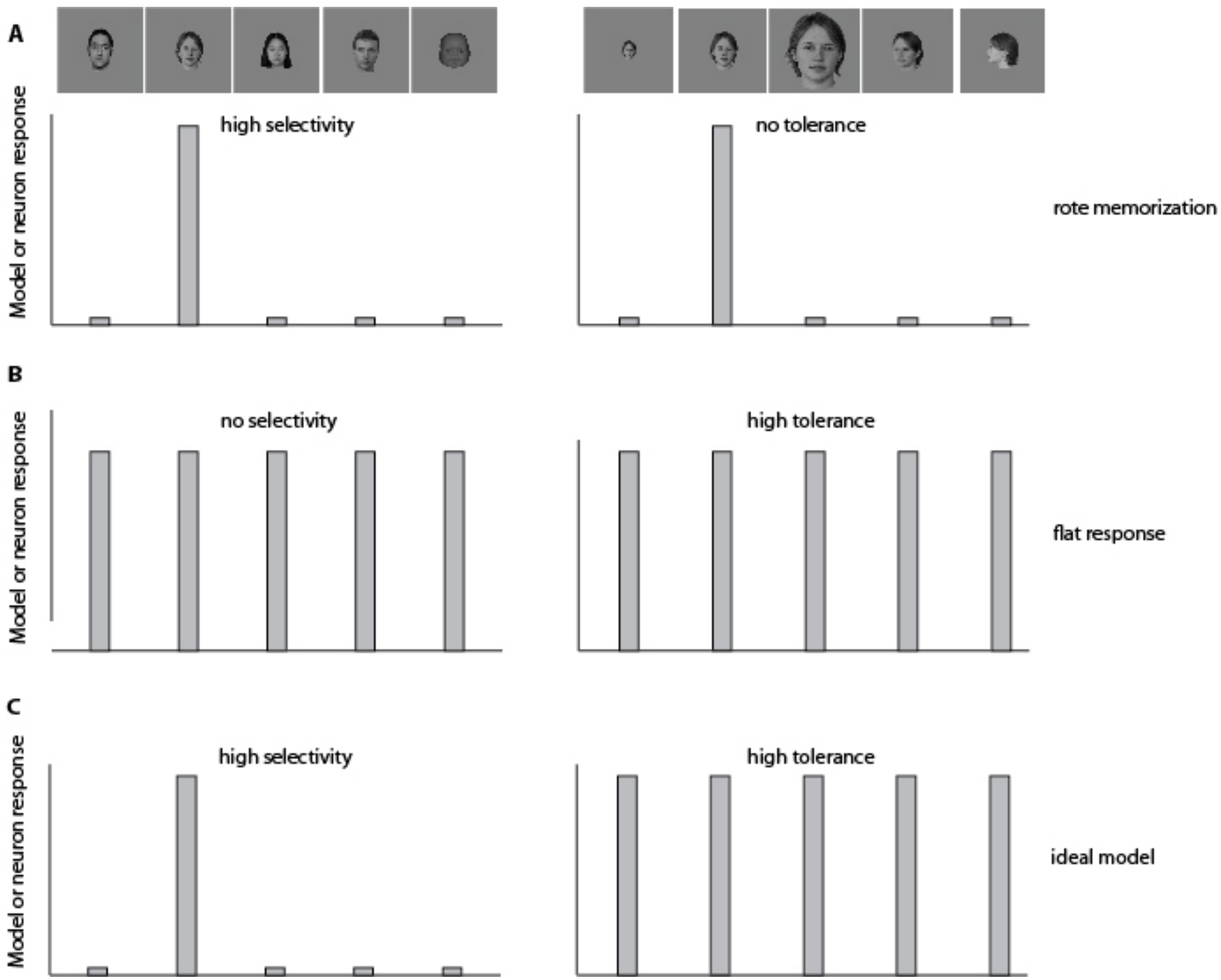
- Navigation
- Recognizing danger
- Recognizing food
- Social interactions
- Recognizing far away signals
- High speeds
- (Reading/Symbols)

Applications

- Pattern recognition
  - ATM machines without passwords
  - Automatic analyses of clinical images (e.g. tumor present?)
  - Security
  - Pointing cell phone to a person and knowing who he/she is
  - Automatic behavioral analysis (e.g. biological experiments)
  - Automatic navigation
  - Cars: detecting pedestrians and other vehicles

- Clinical
  - Visual prosthetics: Helping visually impaired people by “reading-out” and “writing-in” information directly into visual cortex
  - Cognitive disorders

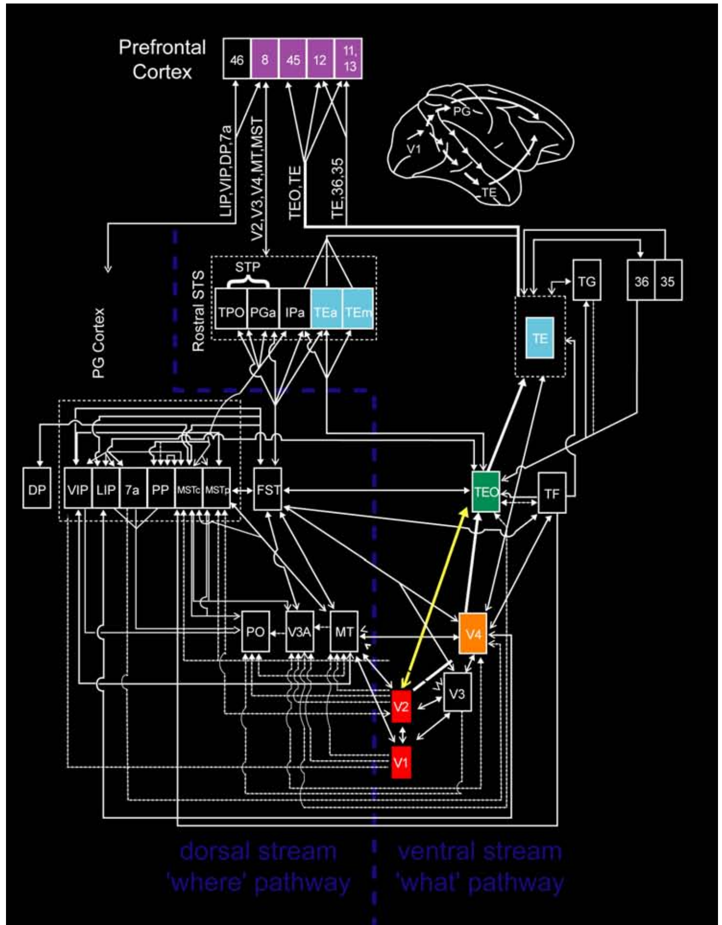
Why is vision difficult?



Coarse circuitry of the primate visual system

Notes:

1. This diagram is an oversimplification
2. A large number of areas in the primate brain are involved in vision
3. Connections are bi-directional
4. Stereotypical “cannonical” circuitry
5. We do not understand the function of most of these connections



Felleman and Van Essen. *Cerebral Cortex*

Lesions provide important insights into function

Table 1 Identity recognition and familiarity ratings for target and nontarget faces (patient E.H.)

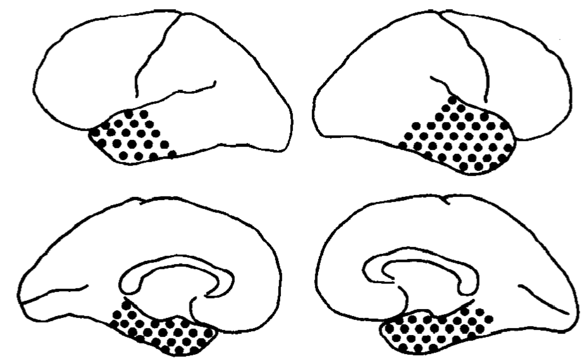
	N	Identity recognition (% correct)	Average familiarity rating (s.d. in parentheses)
Retrograde-family experiment			
Target	8	0	6.0 (0.0)
Nontarget	42	—	6.0 (0.0)
Retrograde-famous experiment			
Target	8	0	6.0 (0.0)
Nontarget	42	—	6.0 (0.0)

-Unable to visually recognize friends, famous people, relatives, even self

-Could not learn to recognize new faces (but could learn to recognize new people from voice and other cues)

-Normal language, memory, learning, non-face object recognition

-Many normal visual functions

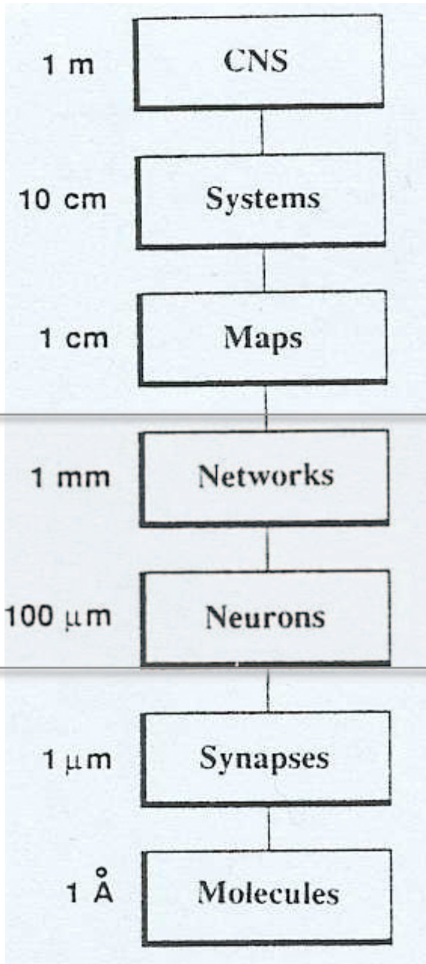


Distribution of lesion sites in cases of face agnosia

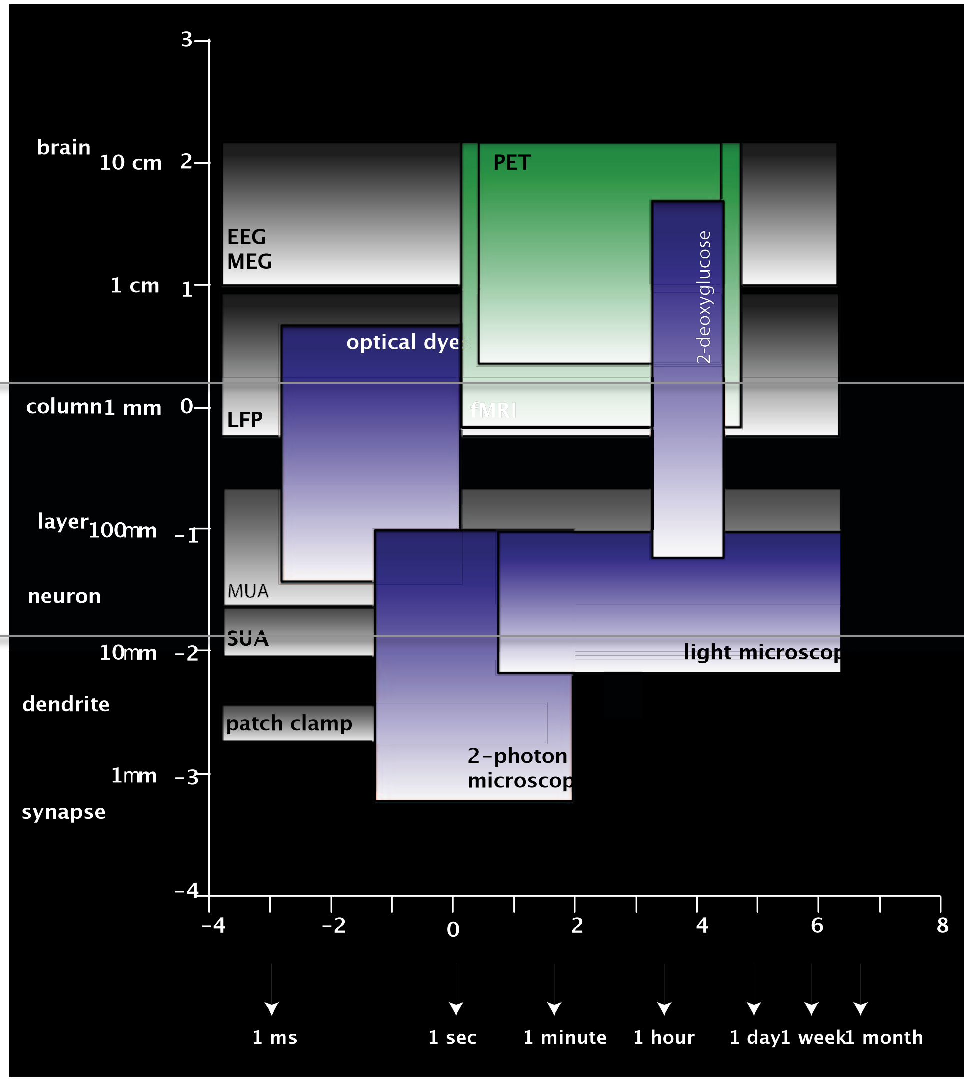
Damasio et al. *Face agnosia and the neural substrates of memory.* Annual Review of Neuroscience (1990). **13**:89-109

Lesions in macaque inferior temporal cortex lead to object recognition deficits (Dean 1976)

Every problem has a “natural scale”

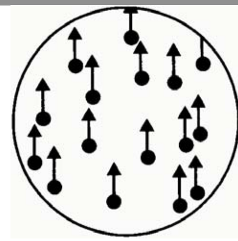


Churchland and Sejnowski
Science 1988

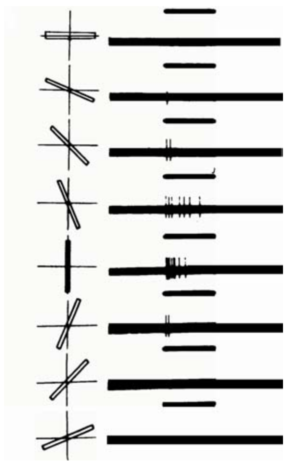


Kreiman. *Physics of Life Reviews* 2004

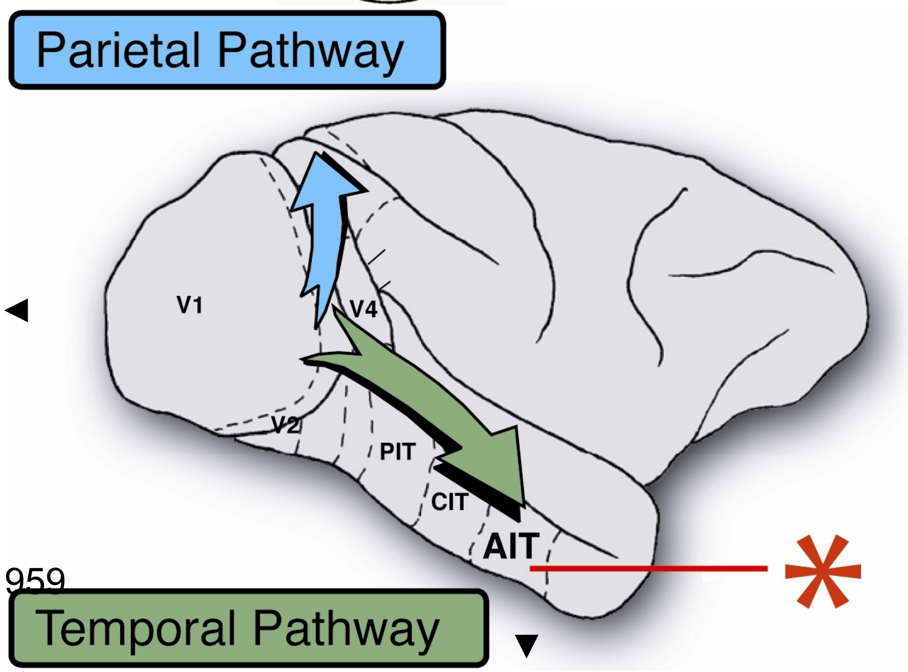
Functional anatomy of the primate visual system

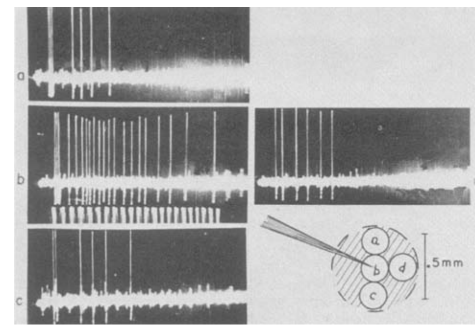


Newsome *et al*. *Nature* 1989

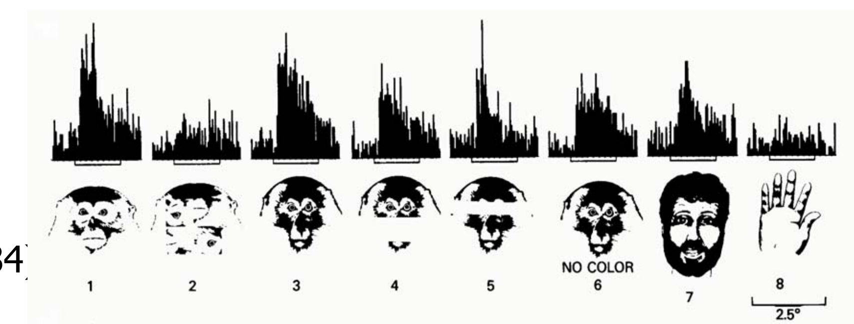


Hubel & Wiesel. *J. Physiol.* 1959



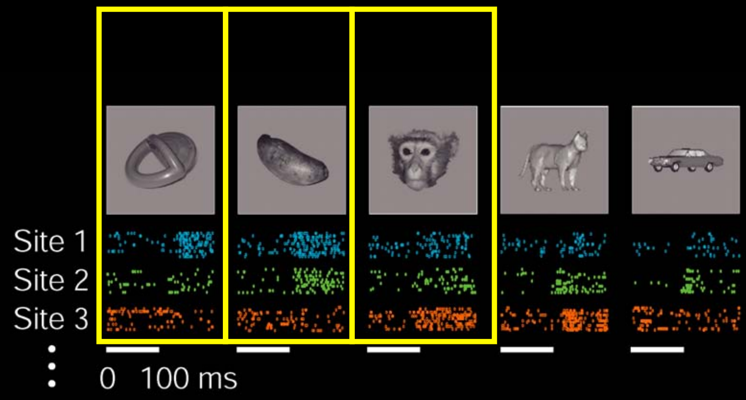


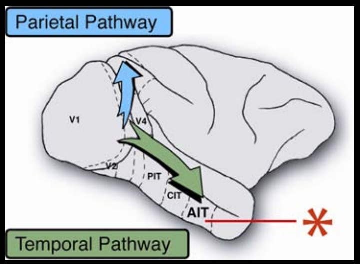
Kuffler, *J. Neurophys* 1953



Desimone *et al. J. Neurosci.* 1984

Quantifying selectivity and tolerance in macaque inferior temporal cortex

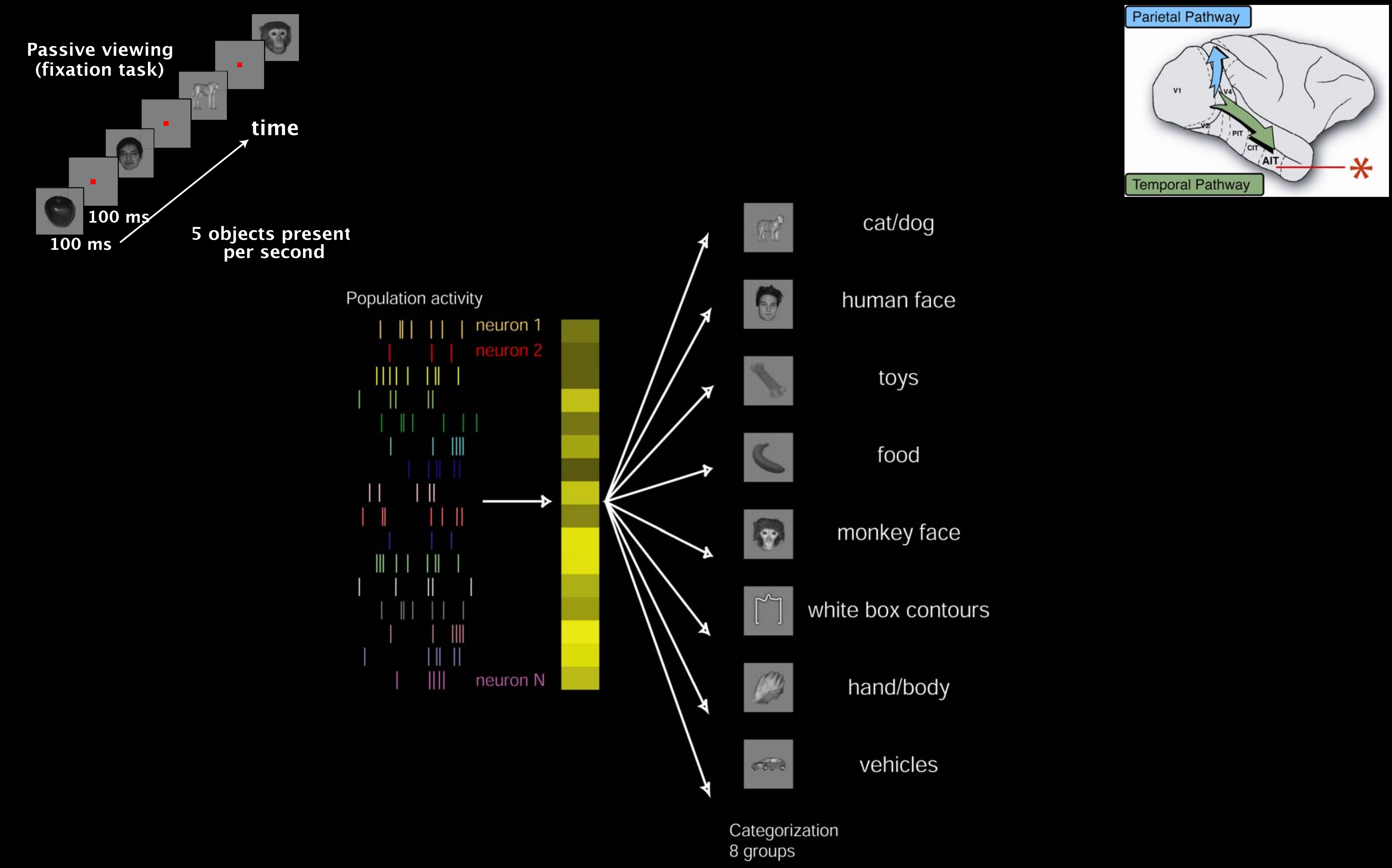




Neuron 1	Neuron 2	Neuron 3	Object
Yes	No	No	1
Yes	Yes	No	2
Yes	Yes	Yes	3

Chou Hung, Jim DiCarlo, Tomaso Poggio

Using machine learning to decode object information from neuronal populations in monkeys



Neuronal population activity

Classifier prediction

Vehicle

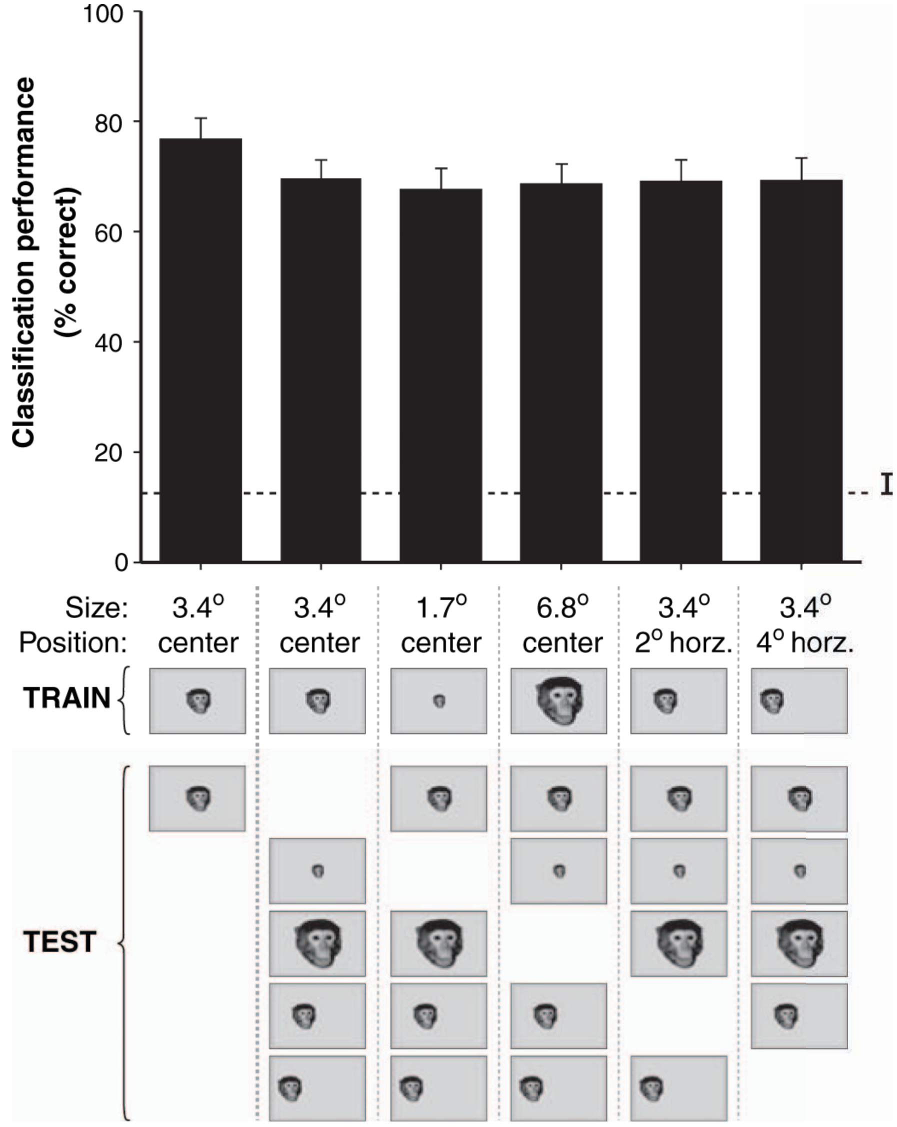
<u>Categorization</u>

- Toy
- Body
- Human Face
- Monkey Face
- Vehicle
- Food
- Box
- Cat/Dog

256 units
Categorization: ~90% (chance = 12.5%)
Identification: ~70% (chance = 1.3%)

Video speed: 1 frame/sec
Actual presentation rate: 5 objects/sec

Scale and position tolerance in inferior temporal cortex



Shape tuning in V4: example

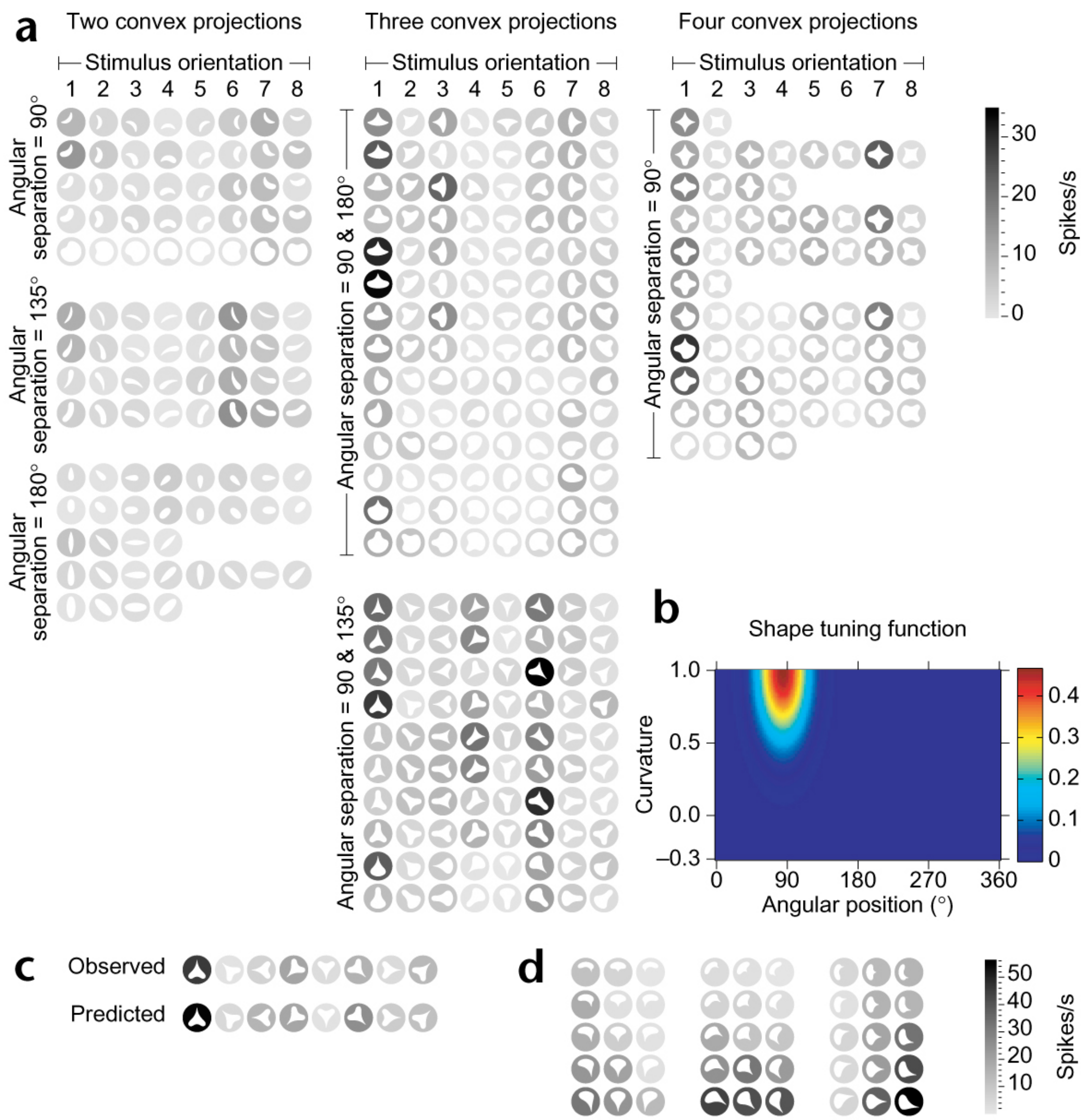
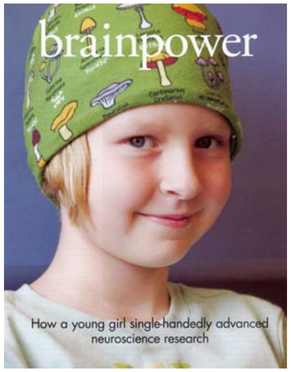
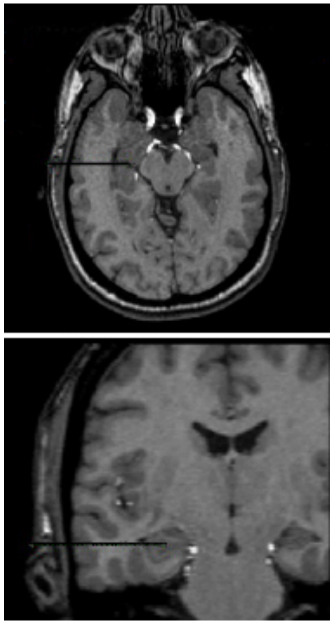


Fig. 1. Single neuron shape-tuning example. (**a**) Responses of an individual V4 neuron are represented by shades of gray surrounding each stimulus icon. The response to each stimulus was averaged across five presentations. The scale bar (right) shows that mean response rates ranged from 0 (light gray) to 34 (dark gray) spikes/s. The stimulus set comprised most of the geometrically feasible combinations of five standard boundary fragments: sharp convex, medium convex, broad convex, broad concave and medium concave curves. Each combination was presented at eight orientations (rows), or fewer if rotational symmetry made some orientations redundant. The stimuli are arranged here into three large blocks (left, middle, right) according to how many convex projections they contained (two, three or four, respectively). They are also blocked in the vertical direction according to the angular separations between convex projections. The stimuli were presented in red (the optimal color for this cell) at the cell's receptive field center (0.32° left of and 1.32° below fixation). (**b**) Gaussian shape-tuning function describing the response pattern in (**a**). The vertical axis represents boundary curvature, and the horizontal axis represents angular position of boundary fragments with respect to the shape's center of mass. The color scale (right) indicates normalized predicted response. The tuning peak corresponds to sharp convex curvature (1.0) near the top of the shape (84.6°). (**c**) Comparison of observed responses to responses predicted by the Gaussian tuning function, for the heart-shaped stimulus at eight orientations. Gray-level scale is the same as in (**a**). (**d**) Auxiliary test of object-centered position tuning for a different neuron.

Neurophysiological recordings in the human brain





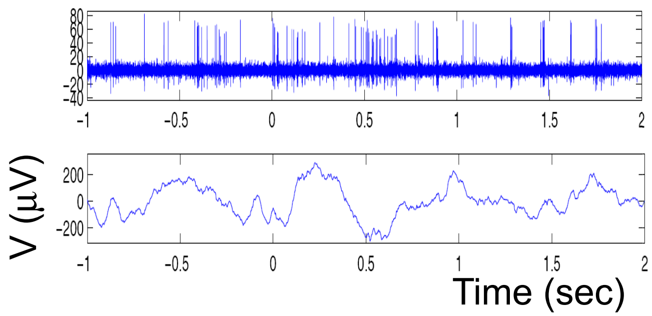
V (µV)

Time (sec)

- Patients with pharmacologically intractable epilepsy
- Multiple electrodes implanted to localize seizure focus
- Targets typically include the temporal lobe (inferior temporal cortex, fusiform gyrus), medial temporal lobe (hippocampus, entorhinal cortex, amygdala and parahippocampal gyrus)
- Patients stay in the hospital for about 7-10 days

Itzhak Fried (UCLA)
Joseph Madsen (CHB)
Alex Golby (BWH)
Stanley Anderson (BWH)

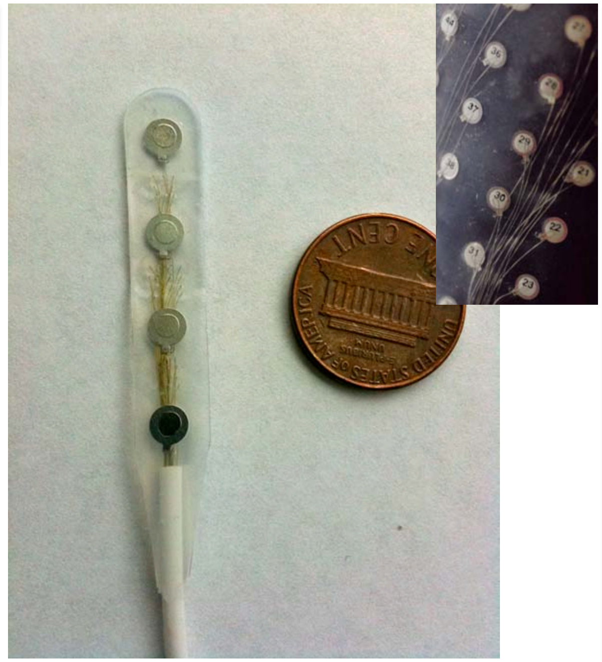
A panoply of different types of electrodes



•Targets typically include the medial temporal (hippocampus, entorhinal cortex, amygdala and parahippocampal gyrus)

•40 micron diameter, impedance ~ 1 MOhm

•Action potentials, LFPs

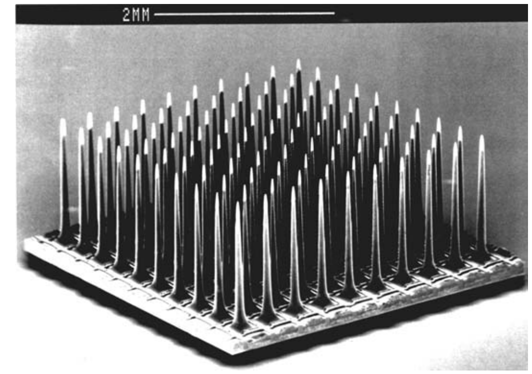


•Subdural (temporal cortex, frontal cortex)

•Low impedance (<1 kOhm)

•High impedance microwires (~ 1MOhm)

•Large coverage



- Utah array electrodes

•Impedance ~ 1 MOhm, 96 microwires, 40 micron diameter

•Local measurements

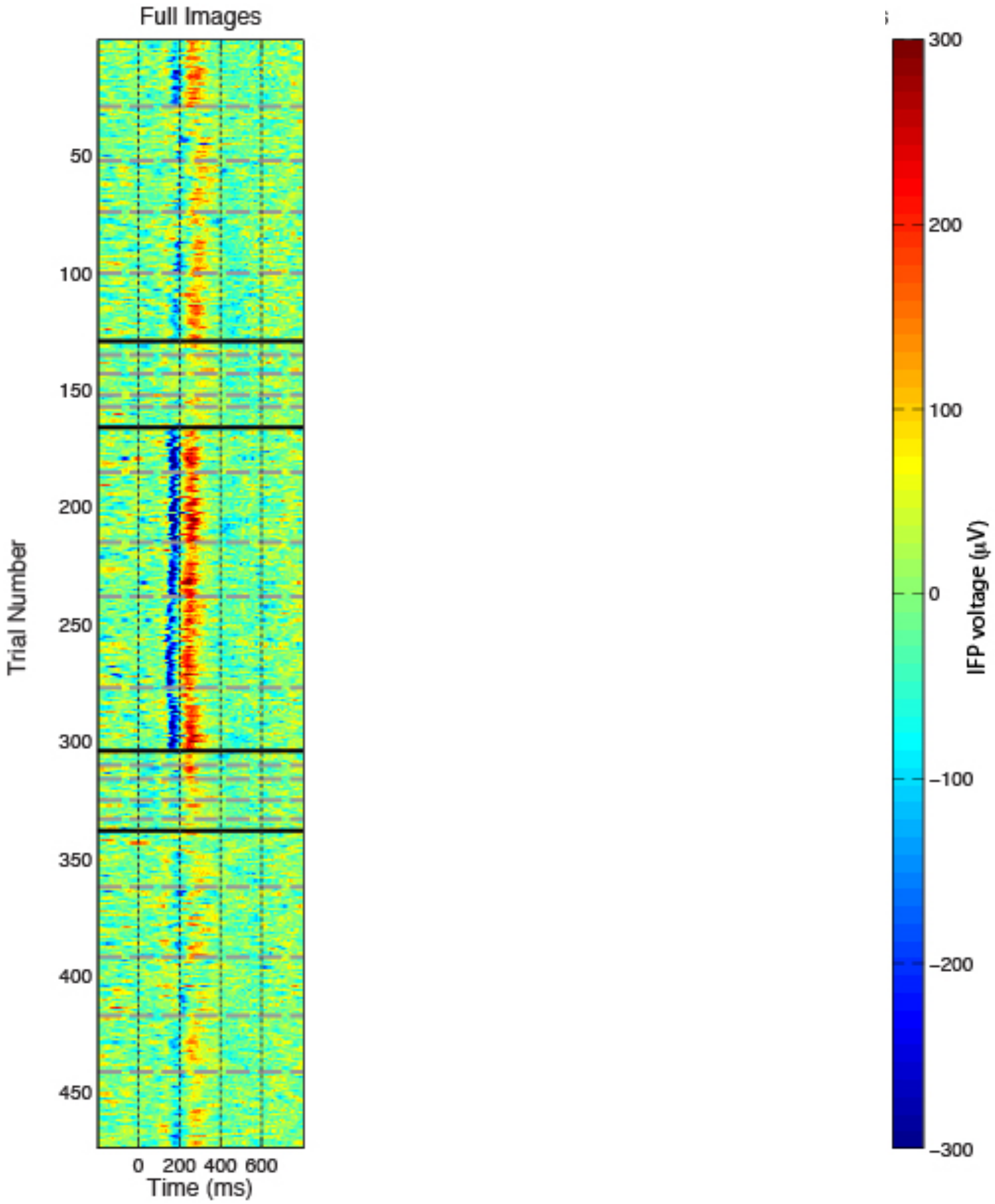
•Action potentials, LFPs

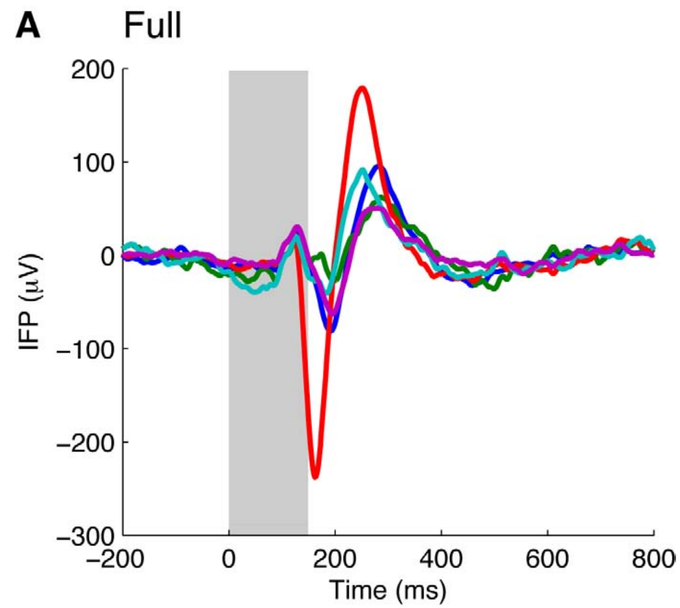
Itzhak Fried (UCLA), Joseph Madsen (CHB), Alex Golby (BWH), Stanley Anderson (Hopkins)
Jed Singer, Radhika Madhavan, Arjun Bansal, Hanlin Tang, Daniel Millman

Example of selectivity and tolerance in the human medial temporal lobe

Itzhak Fried, Quian Quiroga, Christof Koch

Example selective responses in field potential recordings



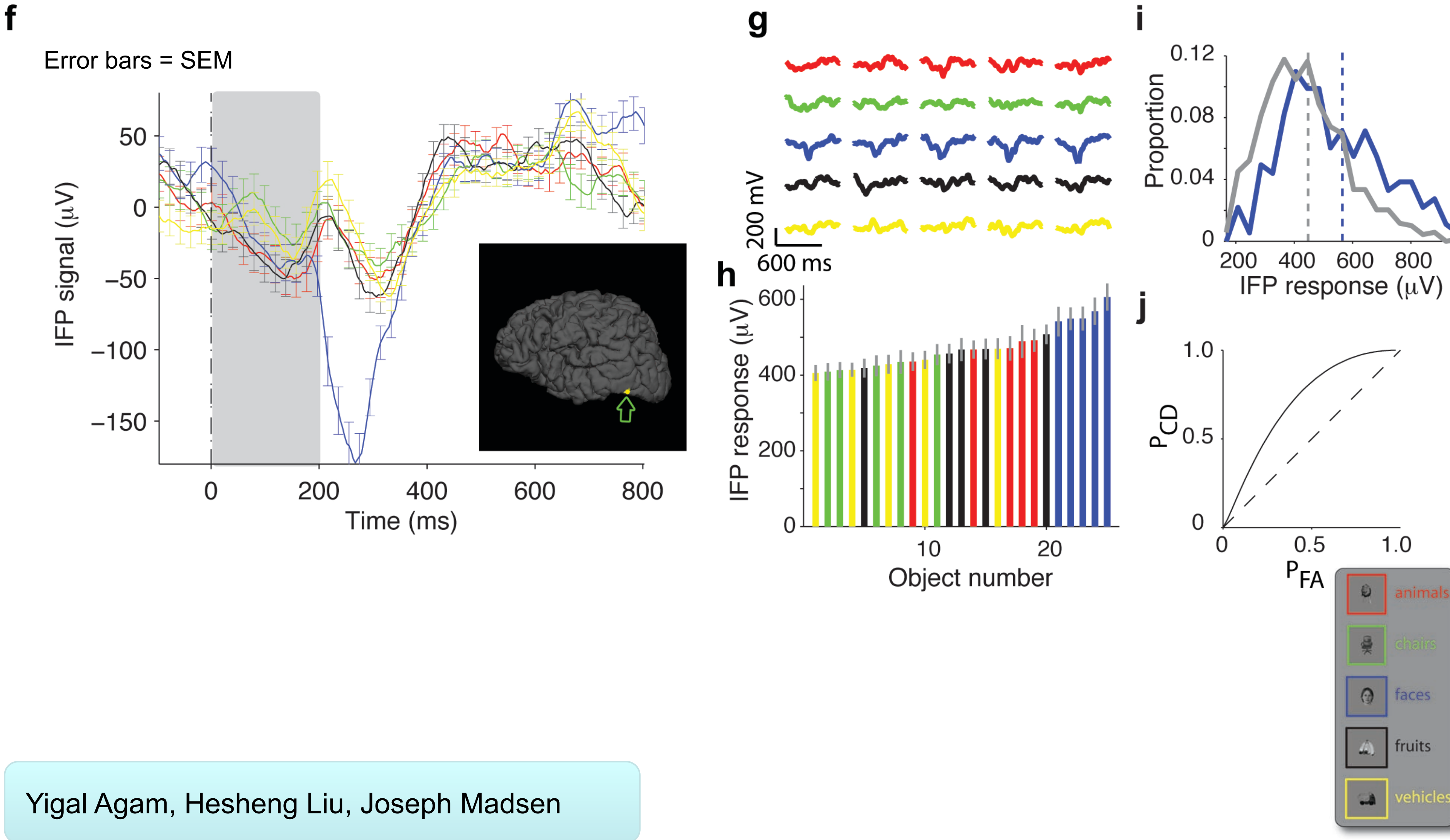


Inferior temporal gyrus

Selectivity in human visual cortex - Example

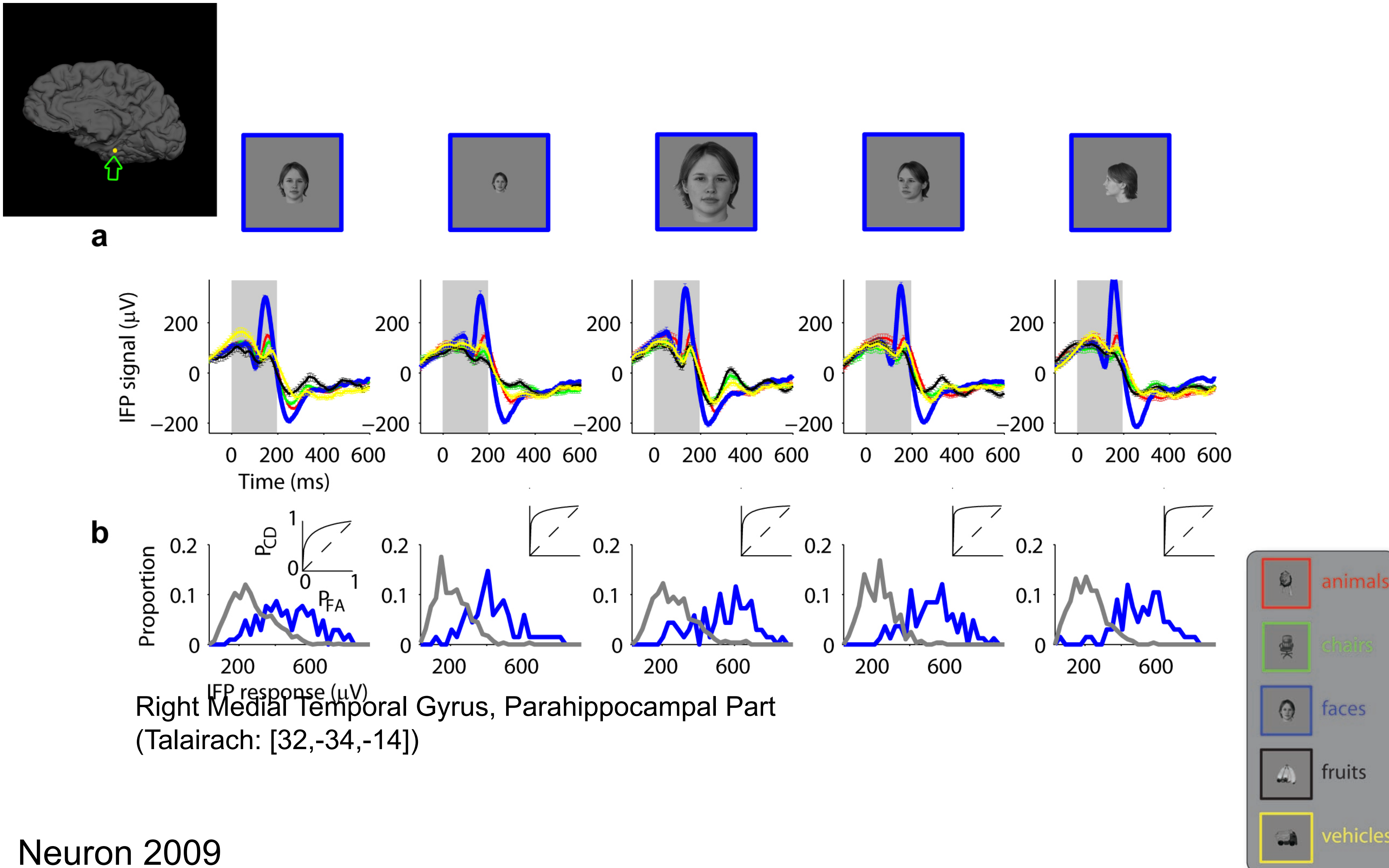
Left Inferior Occipital Gyrus and Sulcus
Talairach: [-48.8,-69.1,-11.8]

Classification performance = 65±5% (change=50%)



Yigal Agam, Hesheng Liu, Joseph Madsen

Tolerance to scale and rotation changes - Example

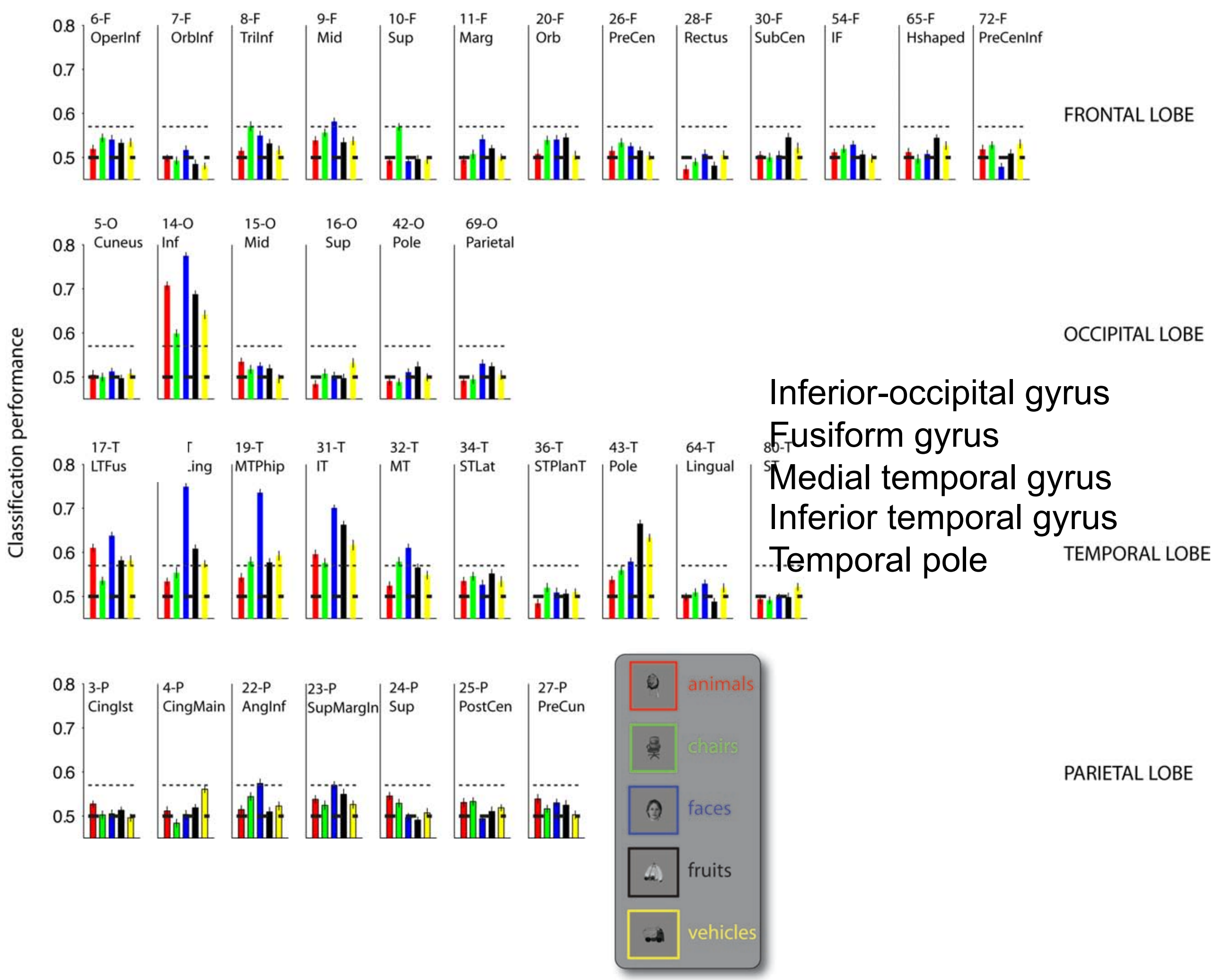


Right Medial Temporal Gyrus, Parahippocampal Part
(Talairach: [32,-34,-14])

Neuron 2009

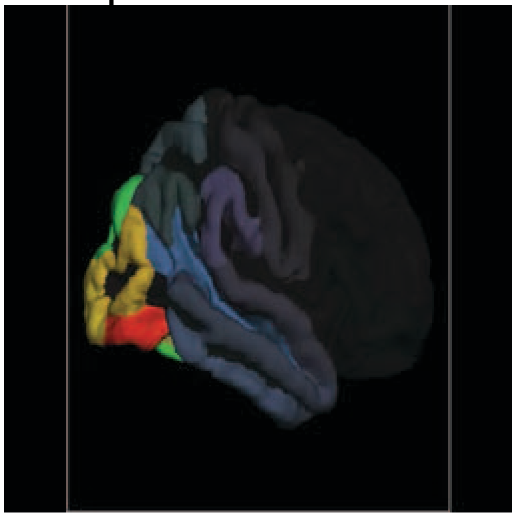
Location, location, location: Stronger selectivity in the temporal lobe

2205 electrodes
27 subjects

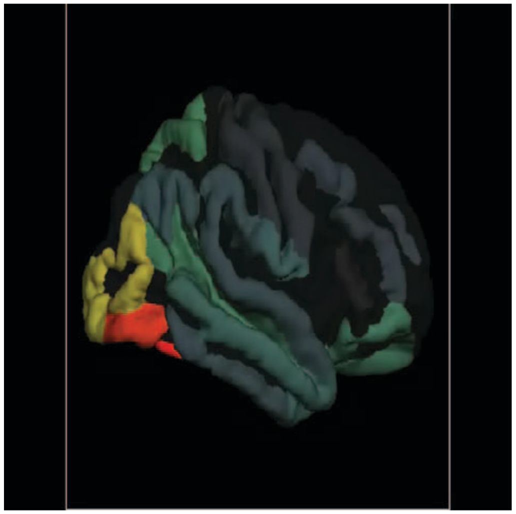


Inferior-occipital gyrus
Fusiform gyrus
Medial temporal gyrus
Inferior temporal gyrus
Temporal pole

Responsive



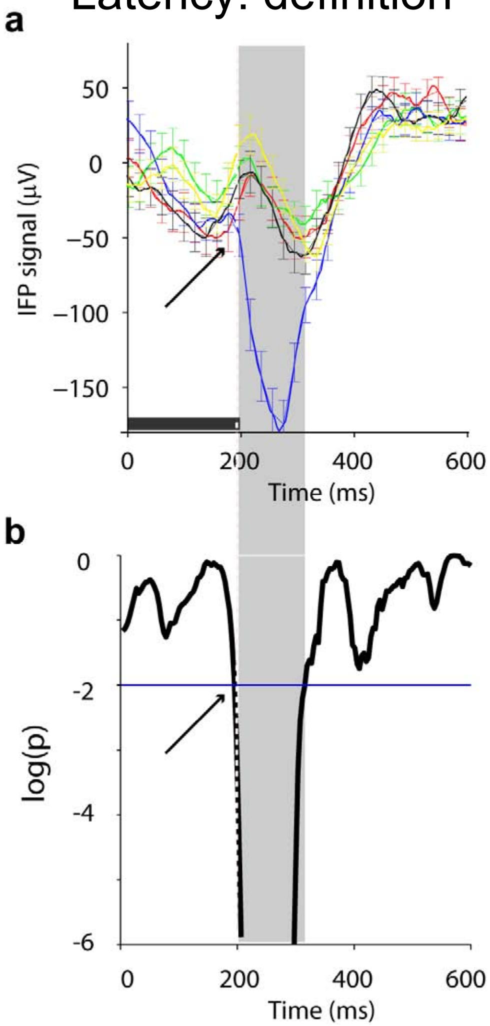
Selective



Jed Singer

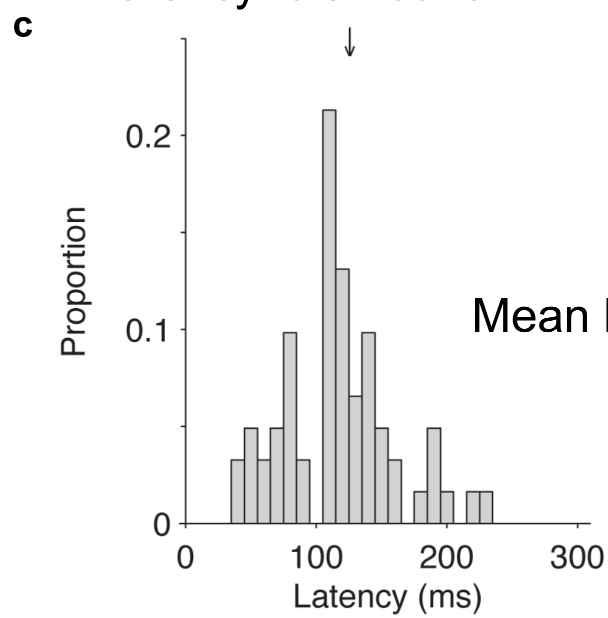
Timing, timing, timing: Selective responses within ~ 150 ms

Latency: definition



Point-by-point ANOVA
25 consecutive points with $p<0.01$ → selective
10 consecutive points with $p<0.01$ → latency
~ Thorpe et al 1996

Latency: distribution

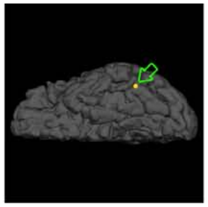


Mean latency = 115 ms

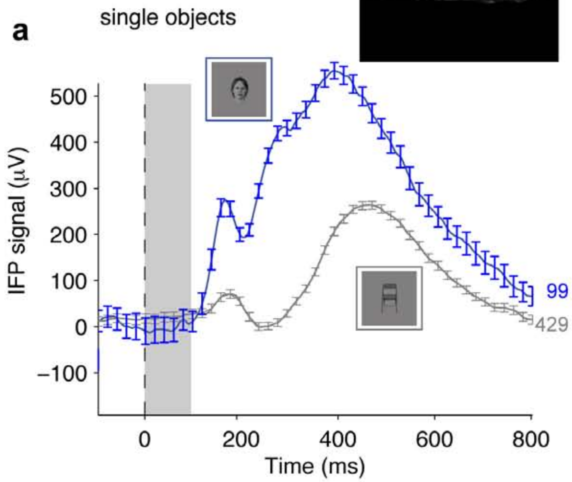
These latencies are consistent with:

- Latencies in macaque monkeys (e.g. Hung et al Science 2005)
- Human scalp recordings (e.g. Thorpe et al Nature 1996)
- Human psychophysics (e.g. Potter et al 1969)

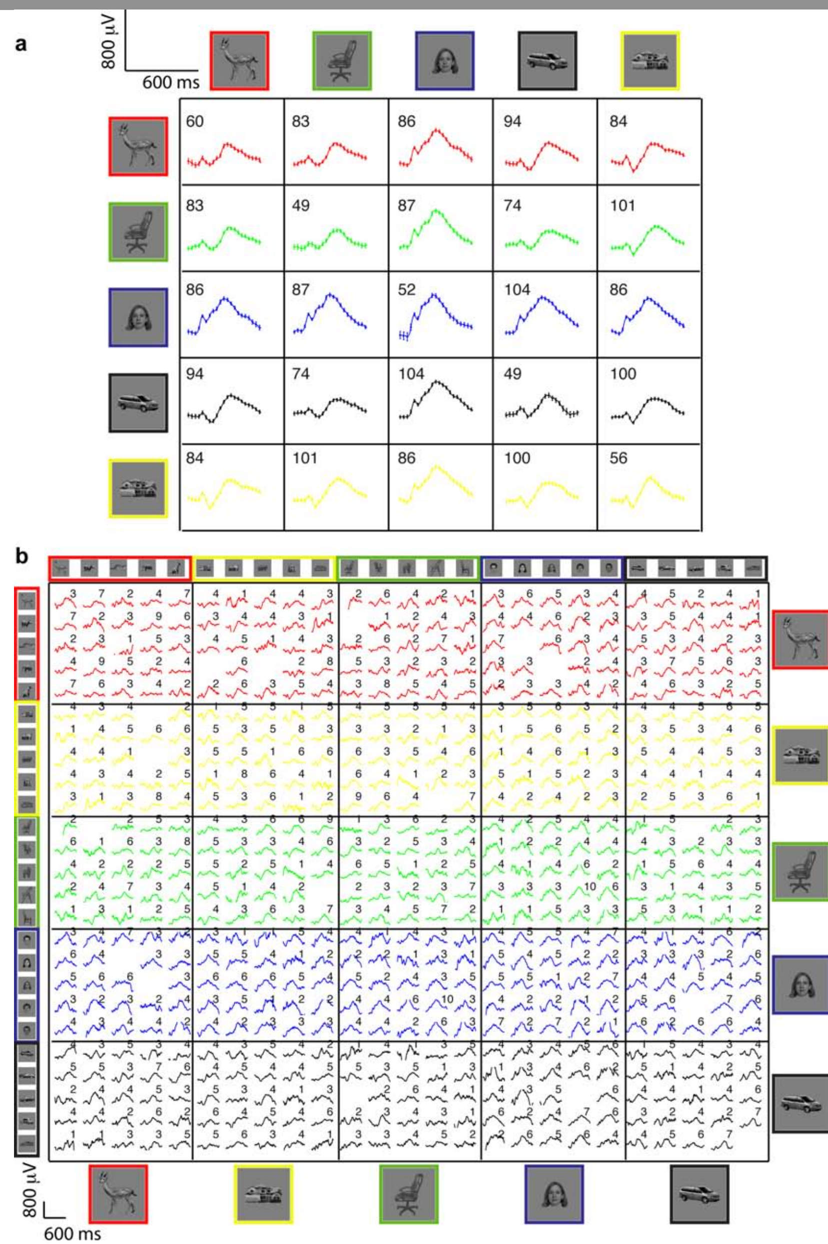
Clutter tolerance in field potential recordings (Example)



Left Occipito-Temporal Fusiform Gyrus [-42,-44,-24]



Same electrode, all object pairs



All category pairs
(5 x 5)

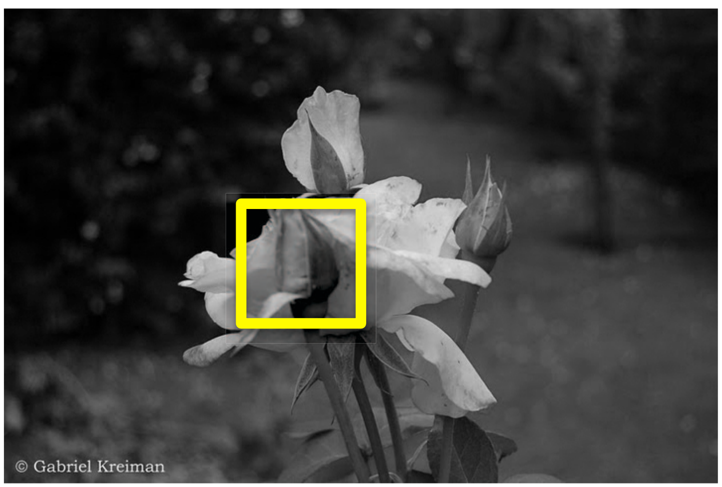
All exemplar pairs
(25 x 25)

Theory and computer models are critical to understand vision

Computational models can

- Integrate existing data
- Explain apparently disparate observations
- Quantify and formalize knowledge
- Suggest experimentally-testable predictions
- Provide a useful engineering tool

A flower, as seen by a computer



23	16	13	12	13	13	12	12	12	14	16	19	21	22	25	24	20	90	127	101
31	22	13	13	12	12	11	11	13	16	18	18	23	22	21	19	39	83	96	78
34	24	16	14	13	12	21	14	13	17	15	22	15	29	42	82	147	118	63	36
30	20	15	13	14	12	26	34	10	11	79	139	88	91	119	174	172	137	96	78
20	14	12	12	14	14	21	77	35	16	136	148	110	109	127	137	168	157	144	175
13	10	10	12	15	16	14	81	86	52	155	123	91	114	149	120	154	139	138	186
9	9	9	11	14	17	18	54	110	111	143	99	105	104	148	128	103	148	162	172
9	8	9	11	14	18	20	26	97	99	99	91	116	116	141	139	77	88	117	156
9	9	12	12	15	18	15	29	107	99	88	86	121	124	115	123	79	78	98	92
9	10	11	13	15	16	30	97	121	112	98	68	102	125	115	101	100	60	105	109
9	9	11	14	17	13	96	127	145	115	95	60	90	114	118	98	107	72	60	111
9	10	12	13	16	17	117	128	122	114	89	65	94	108	118	116	117	93	59	67
10	10	10	7	9	78	152	127	118	114	77	72	95	109	116	120	128	96	68	50
7	1	10	54	114	166	145	121	125	113	65	70	97	107	110	107	103	93	67	54
33	92	129	151	157	158	146	130	125	104	66	77	100	105	111	108	94	85	62	58
145	144	135	142	151	152	149	137	131	98	69	82	102	111	102	93	89	84	59	54
125	125	140	156	144	150	145	133	128	98	74	87	110	110	106	93	86	80	56	48
147	147	161	143	143	144	138	129	121	94	69	86	107	106	102	91	82	77	50	43
182	181	164	140	143	140	132	128	121	97	71	82	100	109	97	91	93	80	44	40
188	174	143	147	146	144	137	127	119	97	78	83	100	105	104	92	86	81	46	38

A brute force approach to object recognition

Task: Recognize the handwritten "A"

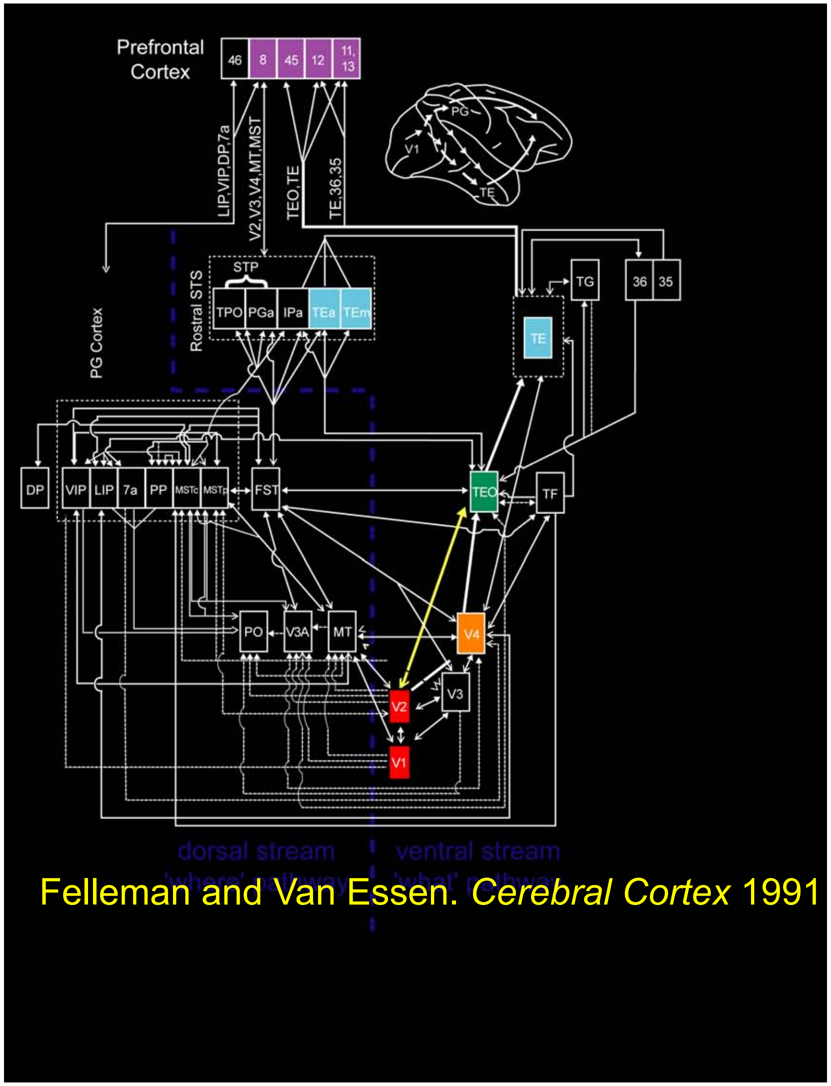
A "brute force" solution:

- Use templates for each letter
- Use multiple scales for each template
- Use multiple positions for each template
- Use multiple rotations for each template
- Etc.

Problems with this approach:

- Large amount of storage for each object
- No extrapolation, no intelligent learning
- Need to learn about each object under each condition

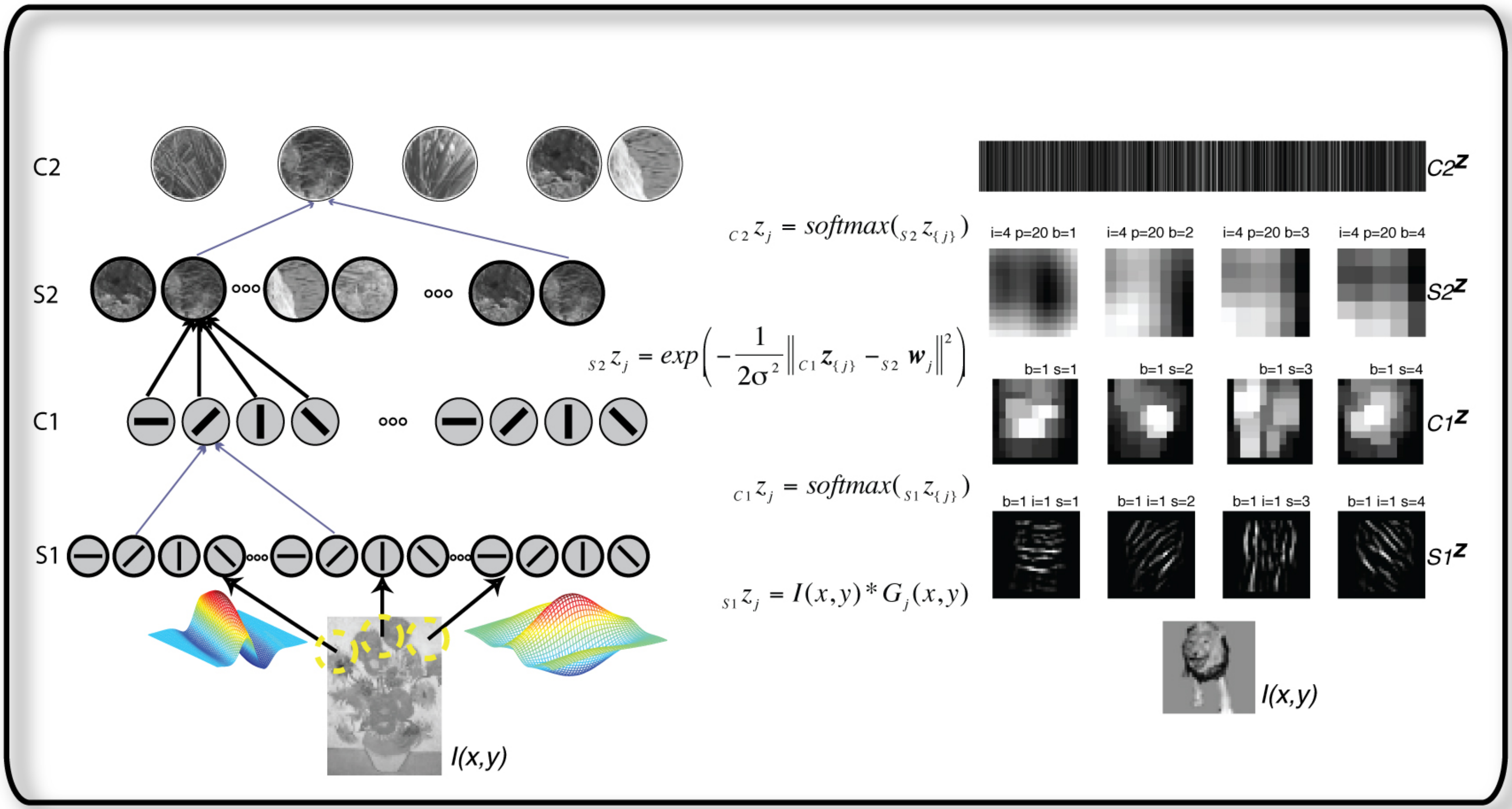
Towards a computational model of ventral visual cortex



Felleman and Van Essen. *Cerebral Cortex* 1991

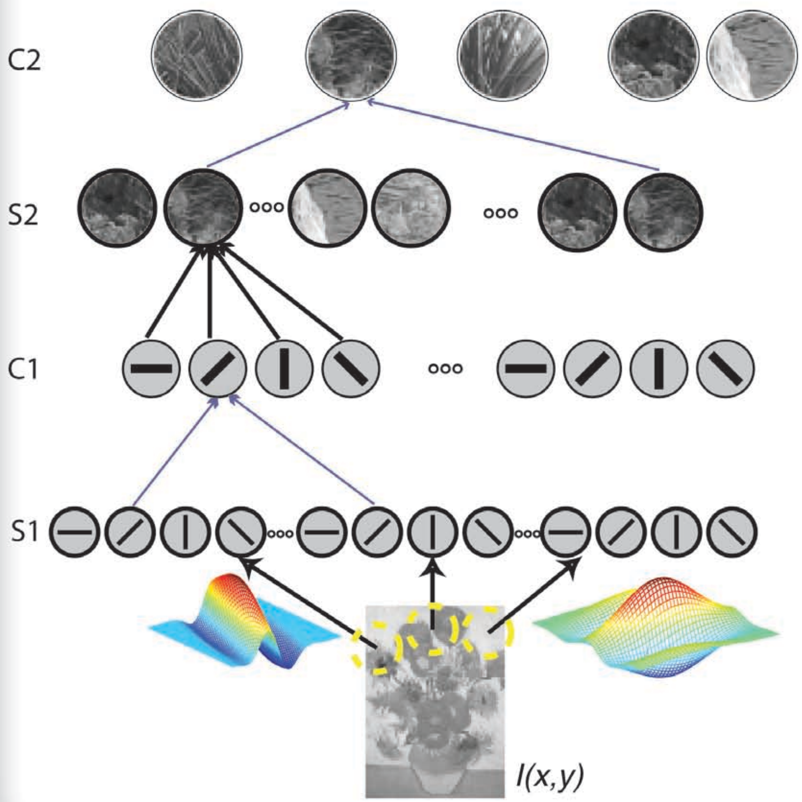
Fukushima 1980, Hubel&Wiesel 1959; Mel 1997; Olshaussen et al 1993; LeCun et al 1998; VanRullen&Thorpe 2002; Amit&Mascaro 2003; Wersing and Korner 2003; Deco and Rolls 2001;

A biologically-inspired, bottom-up, hierarchical model of object recognition



Cadieu, Knoblich, Kouh, Riesenhuber, Serre, Poggio

A biologically-inspired, bottom-up, hierarchical model of object recognition



Selectivity and invariance in the model

1) Apply a battery of Gabor filters to the input image (4 orientations o, 16 scales s) $_{o}S1_{s}$

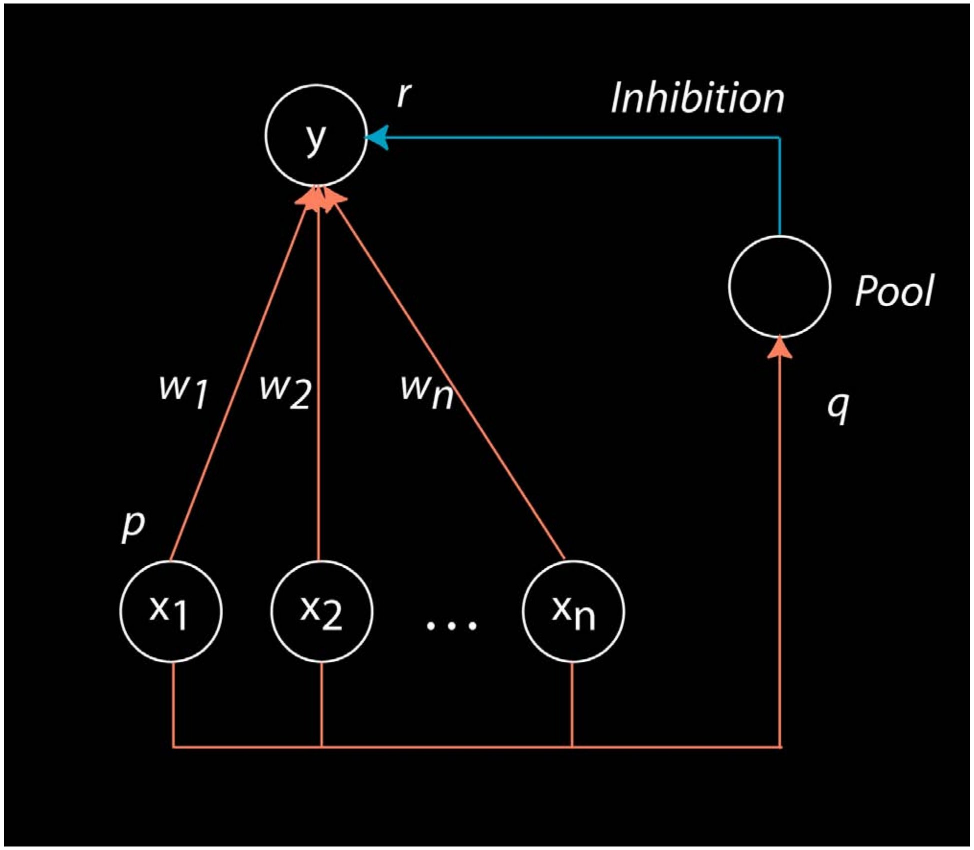
2) Take soft-max over scales and positions (local neighborhood) $_{o}C1_{s}$

Unsupervised Training:
Extract K patches P_i $(i=1,\ldots,K)$ n_i x n_i and all 4 orientations from *C1* maps from natural images.

3) For each *C1* image and at each position compute $\exp(-\gamma||X - P_i||^2)$ for all image patches X and each patch P learned during training $_{i}S2_{\Sigma}$

4) Take soft-max over (all) positions & scales for each S2: obtain position and scale invariant responses $_{i}C2$

Biophysical implementation of cortical nonlinear operations



$$y = \frac{\sum_{j=1}^{n} w_j x_j^p}{k + \left(\sum_{j=1}^{n} x_j^q\right)^r}$$ Canonical

$$y = \sum_{j=1}^{n} x_j^2$$ Energy model

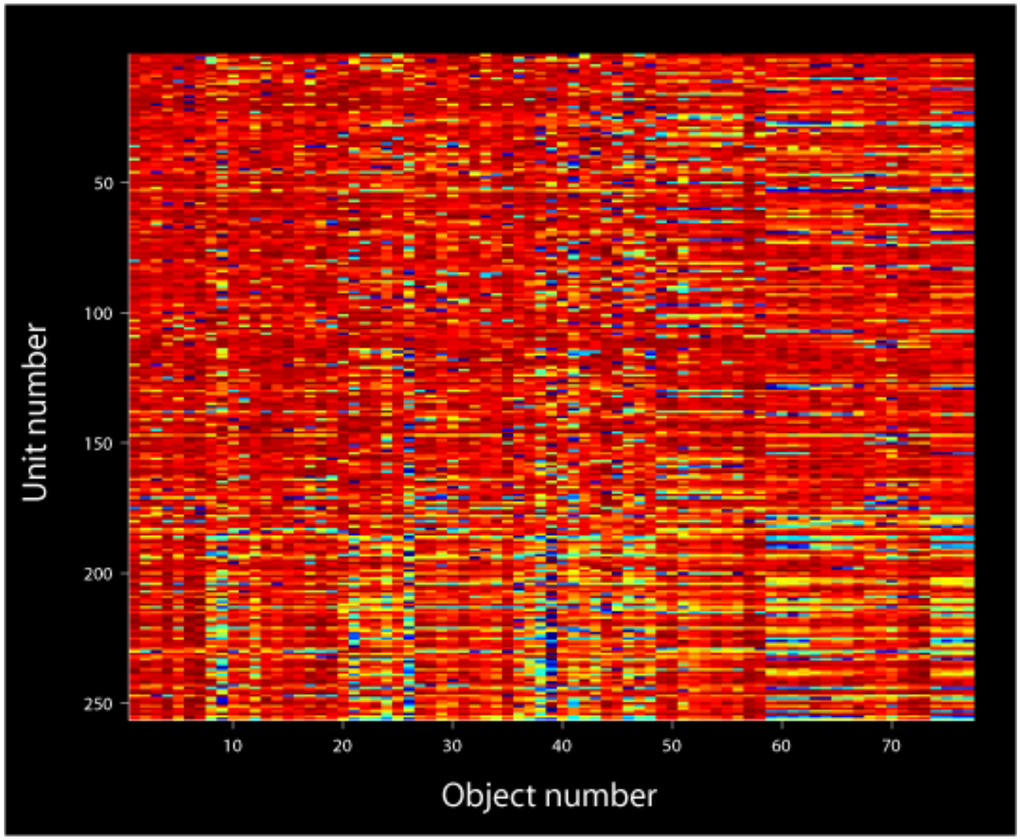
$$y = \frac{\sum_{j=1}^{n} x_j^2}{k + \sum_{j=1}^{n} x_j^2}$$ Sigmoid-like

$$y = \frac{\sum_{j=1}^{n} w_j x_j}{k + \sum_{j=1}^{n} x_j^2}$$ Gaussian-like

$$y = \frac{\sum_{j=1}^{n} x_j^3}{k + \sum_{j=1}^{n} x_j^2}$$ Max-like

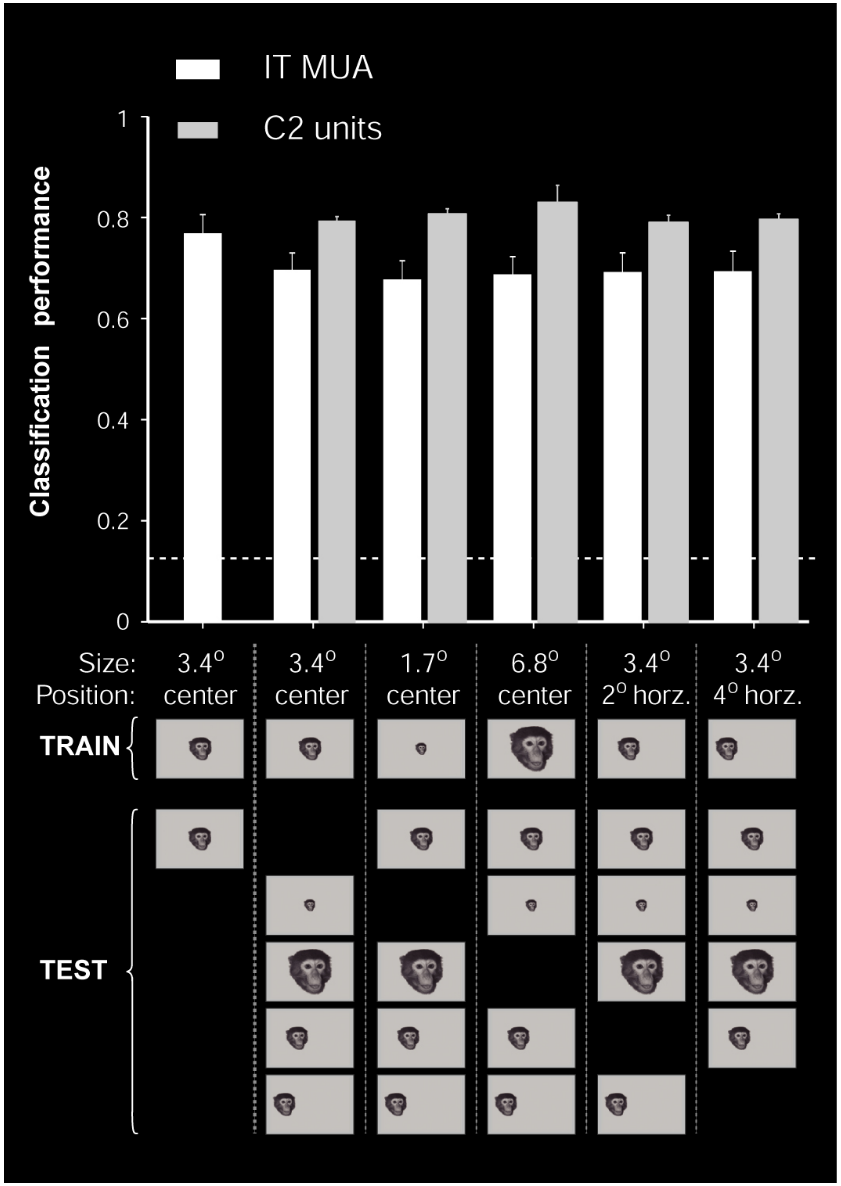
Neural Computation 2008

Example: responses of the top-level units



Images from Hung et al Science 2005

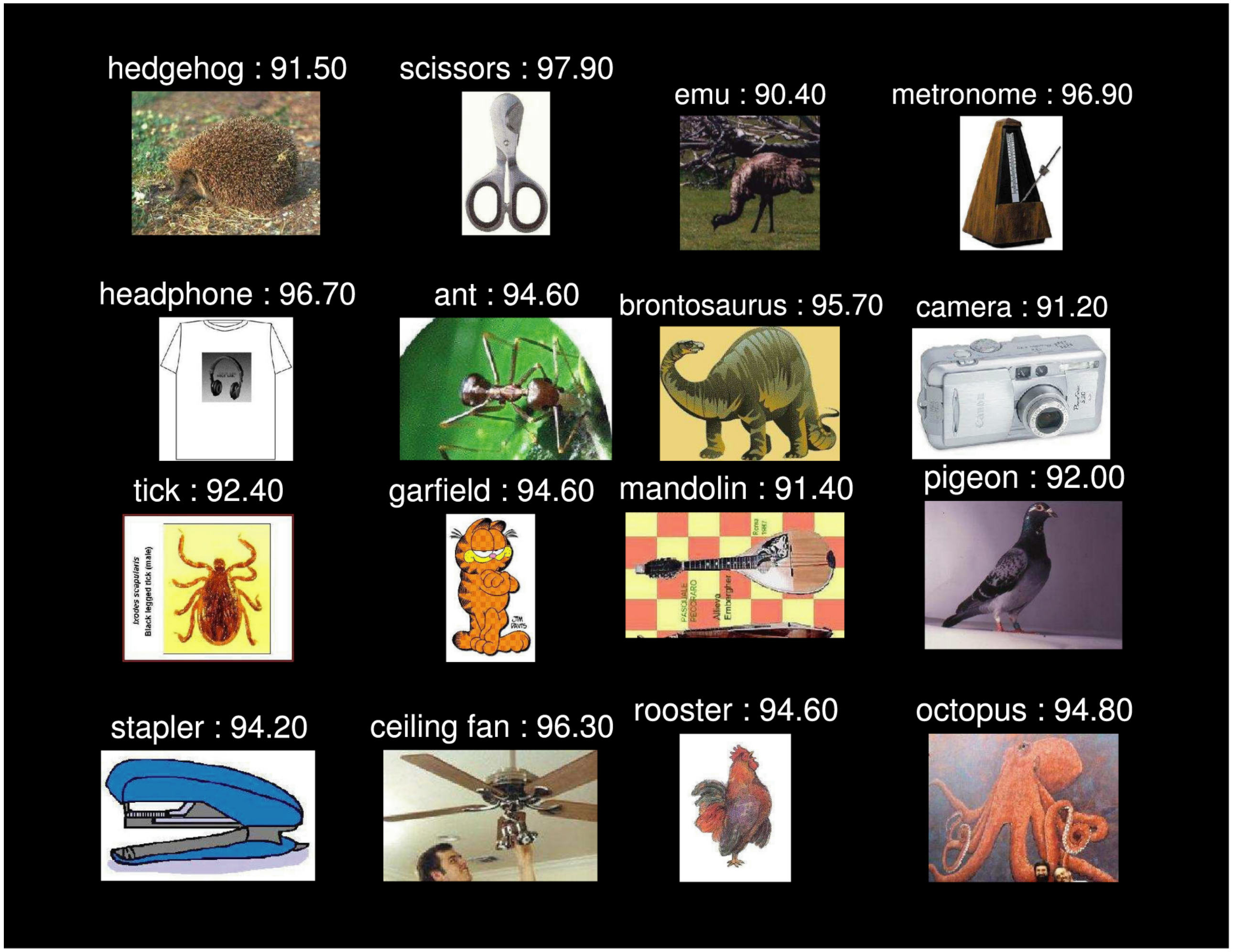
Scale and position tolerance in inferior temporal cortex and model units



The model performs quite well in comparison with state-of-the-art AI systems

Performance on Caltech101 dataset*

ROC areas for category vs. background



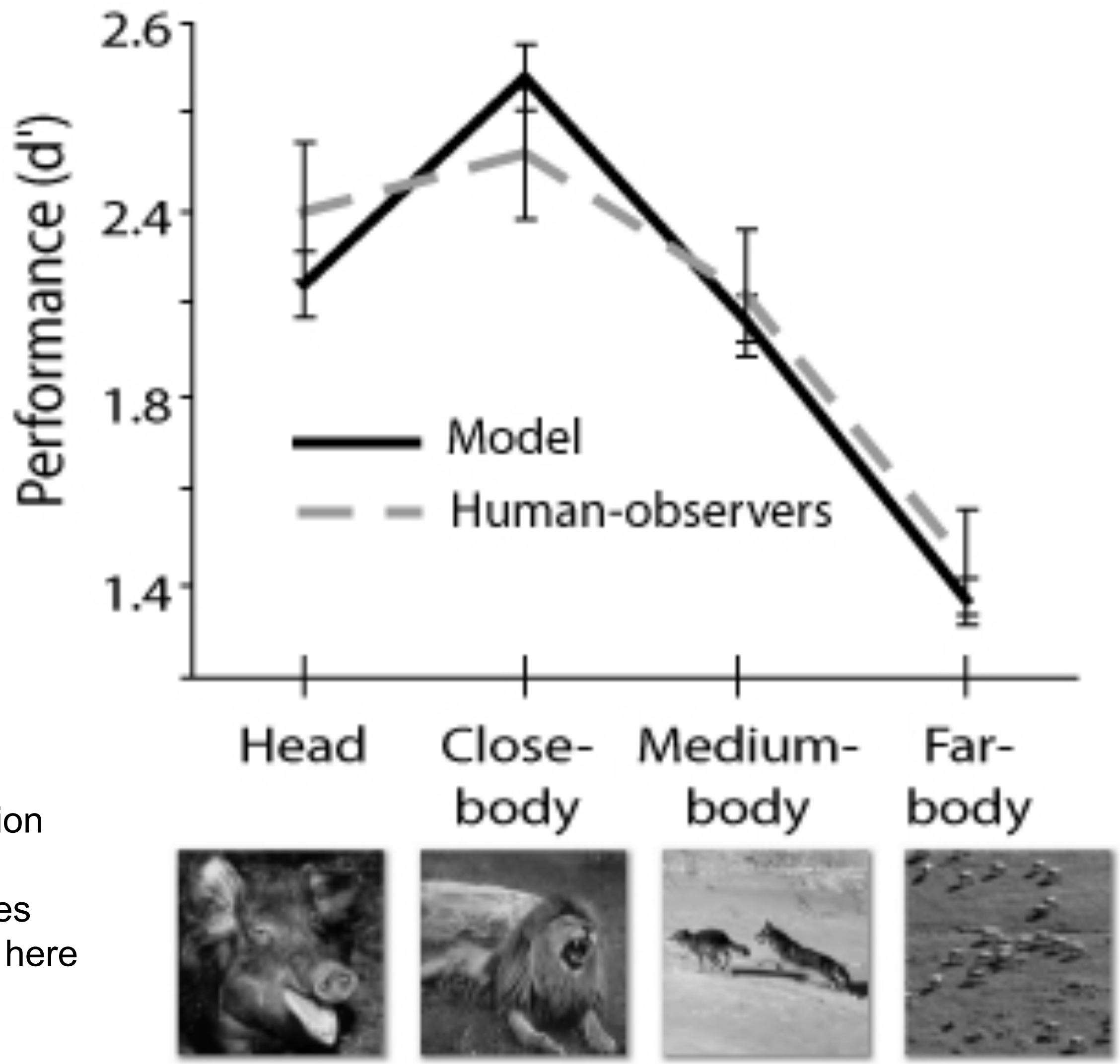
*

1.Performance influenced by low-level image properties

2.Several transformations not examined here

Thomas Serre; CVPR 2005

The model achieves human level performance in a rapid categorization task*



Thomas Serre; PNAS 2007

*

1. Human performance is lowered by rapid presentation followed by mask
2. Low-level image properties are likely to play a key role here

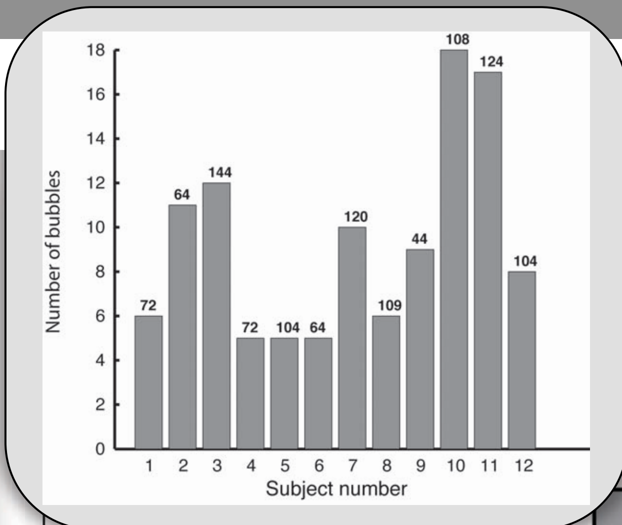
Examining the neurophysiology of object completion

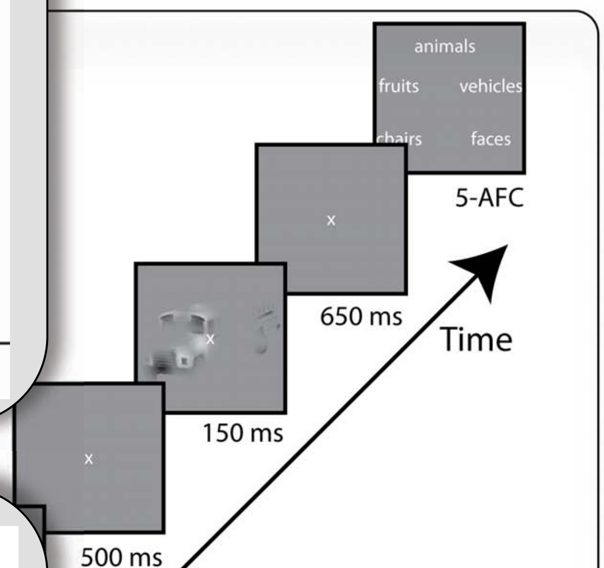
20 bubbles

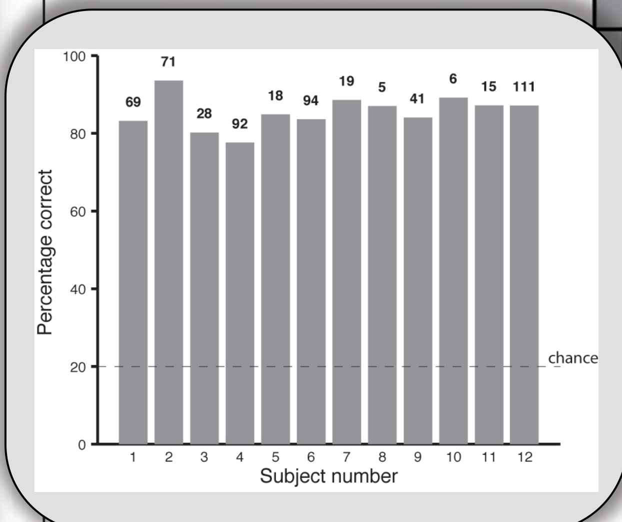
10 bubbles

6 bubbles

4 bubbles



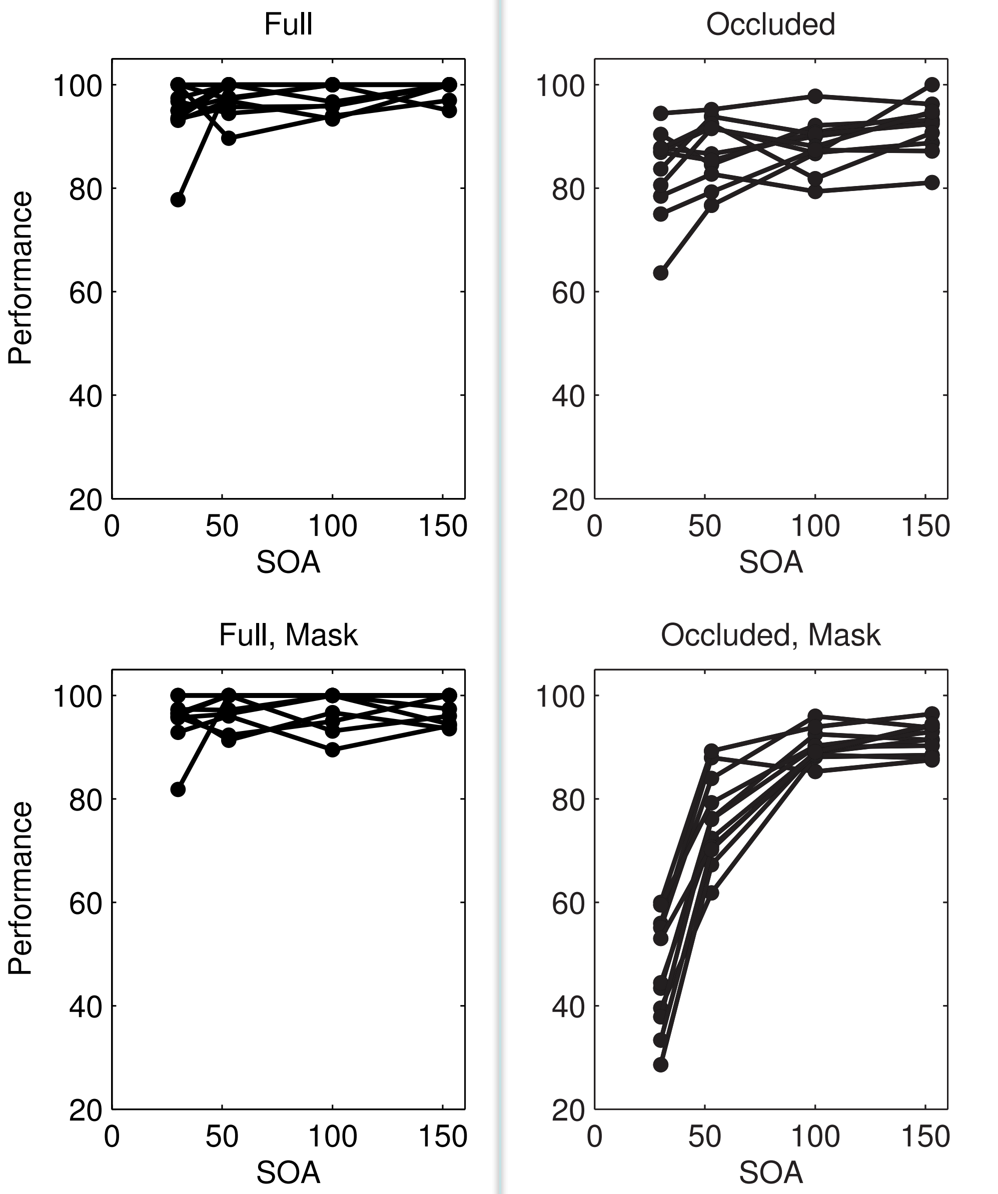




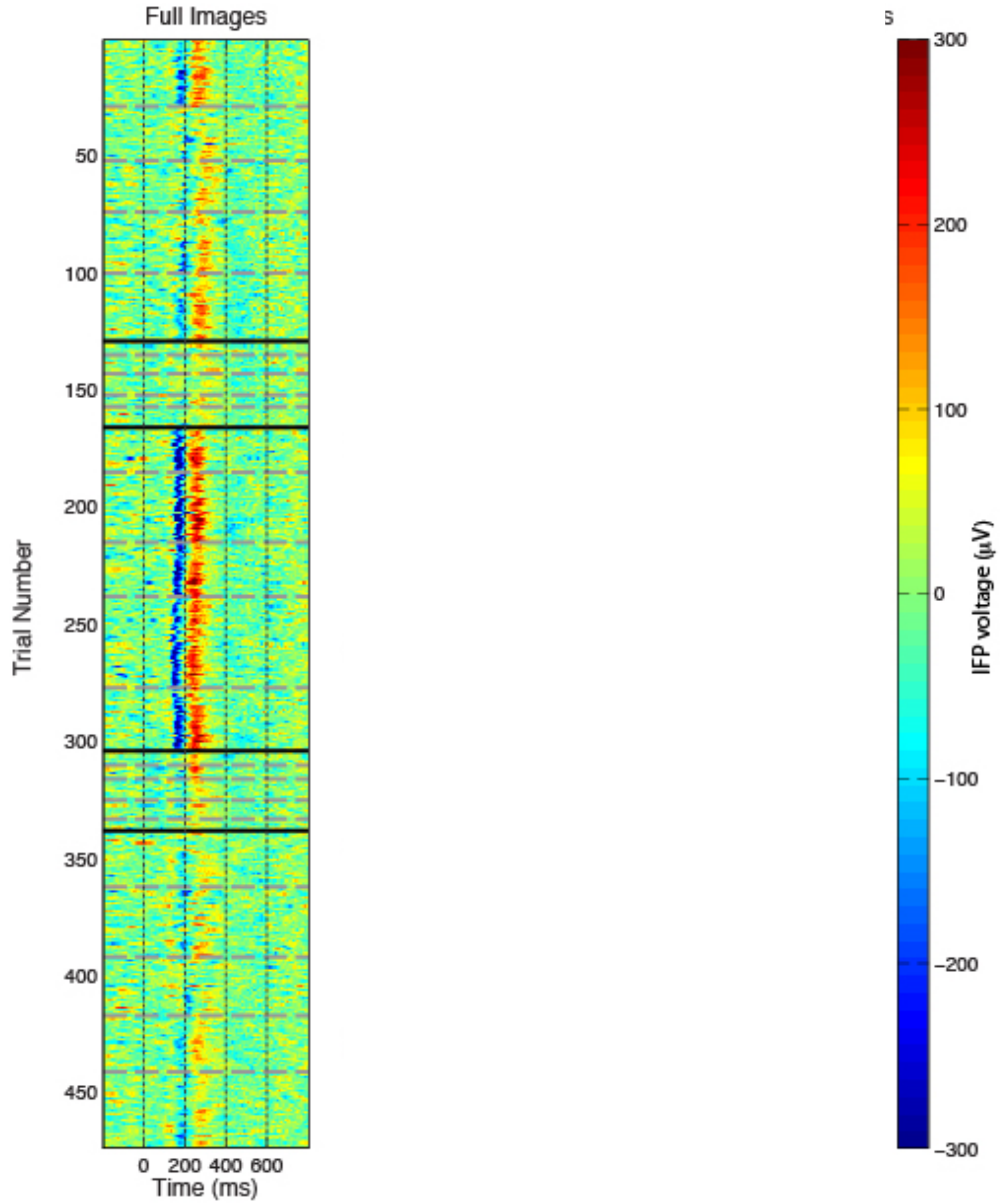
12 subjects (12-40 yrs old)
1129 electrodes
grayscale, contrast normalized stimuli
~4.5 degrees visual angle
Alternative forced choice categorization
Eye tracking in 3 subjects
5 categories
5 exemplars per category
randomized order
bubbles adjusted to ~80% performance

Calin Buia, Hanlin Tang, Joseph Madsen

Performance in object completion task

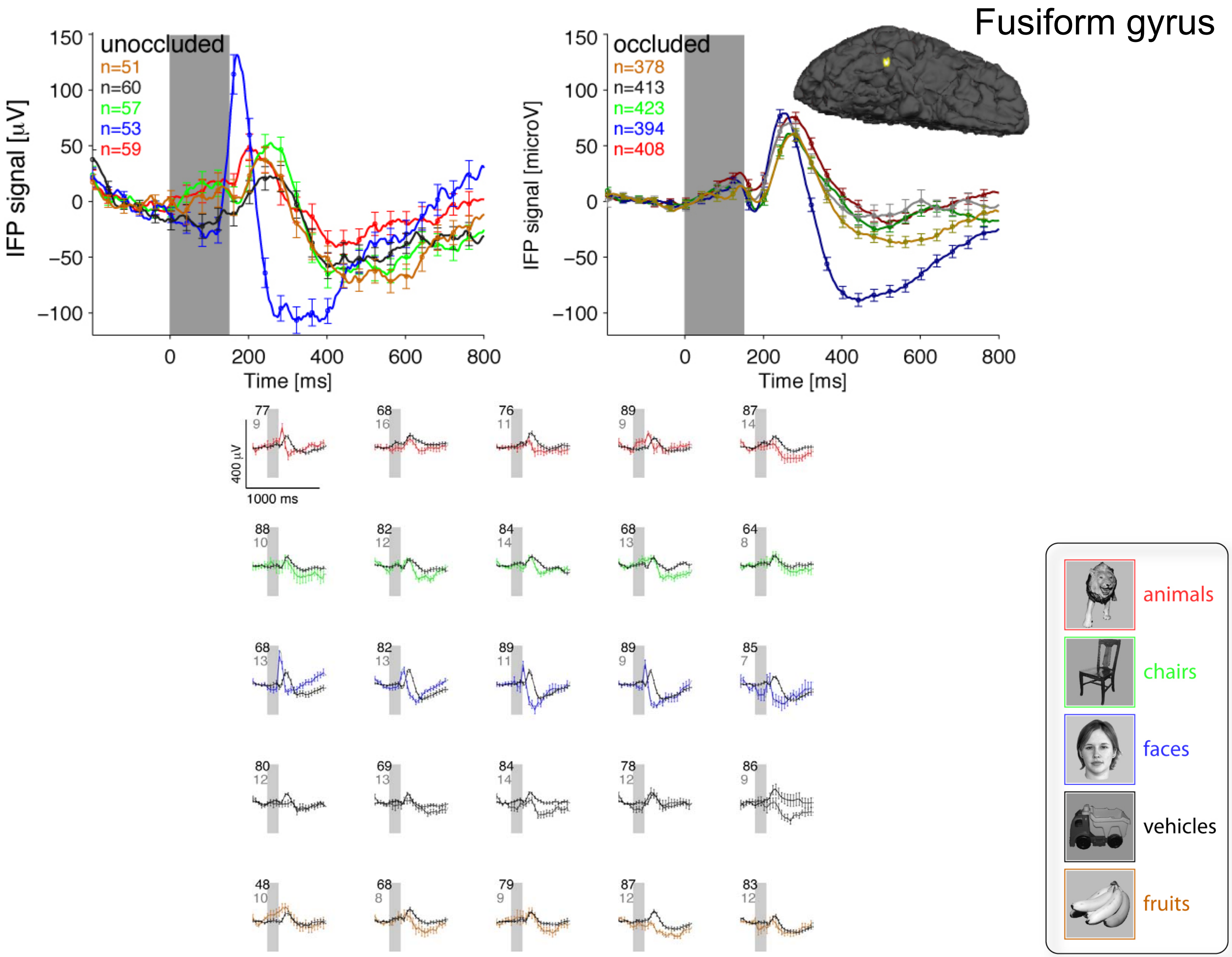


Example responses during object completion (single trials)



Responses during object completion task (Example 1)

Average across repetitions and exemplars

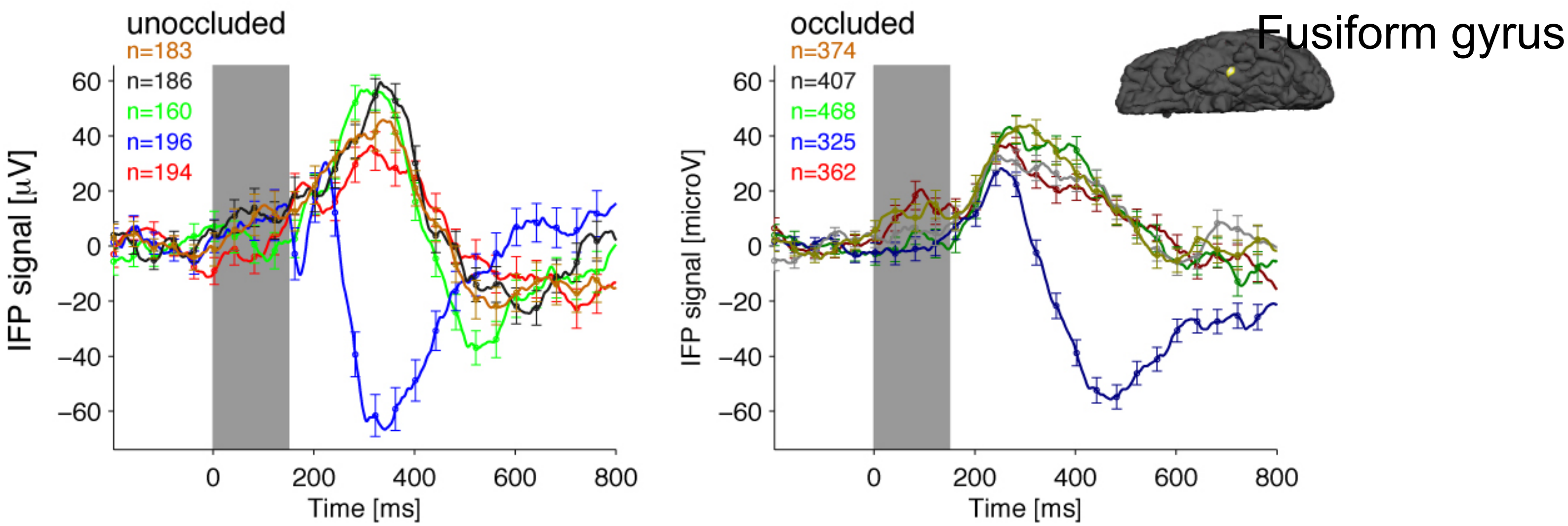


Average across repetitions

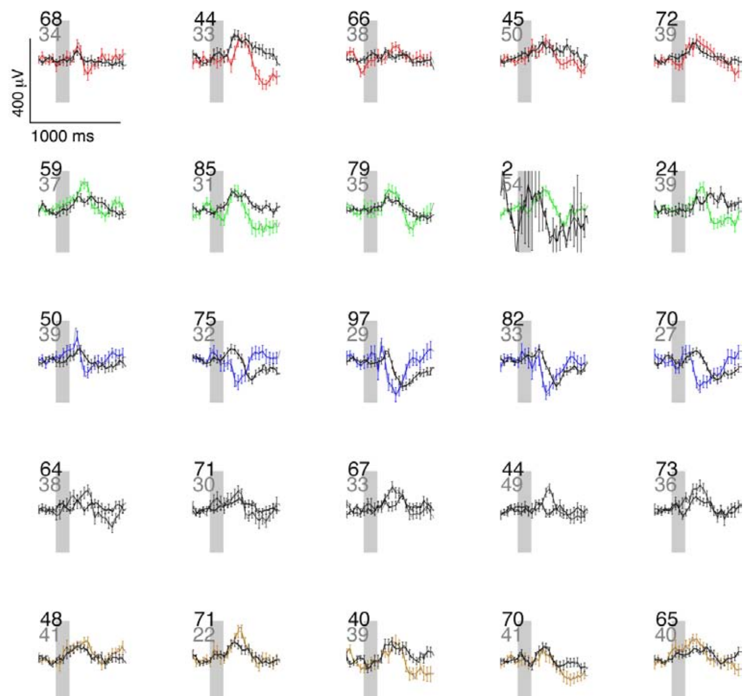
m00026
channel=49

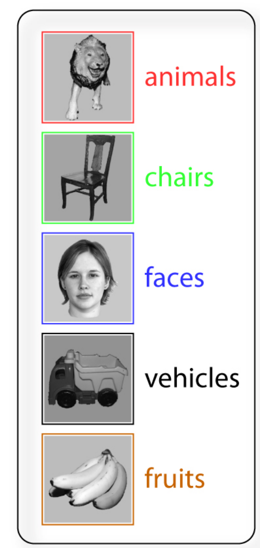
Responses during object completion task (Example 2)

Average across repetitions and exemplars



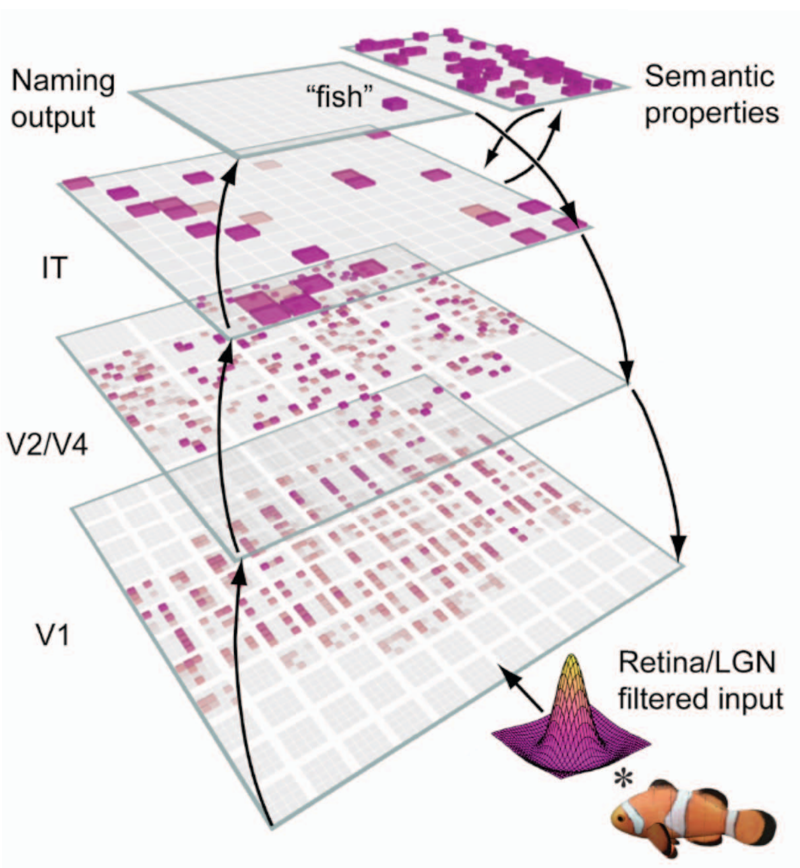
Average across repetitions

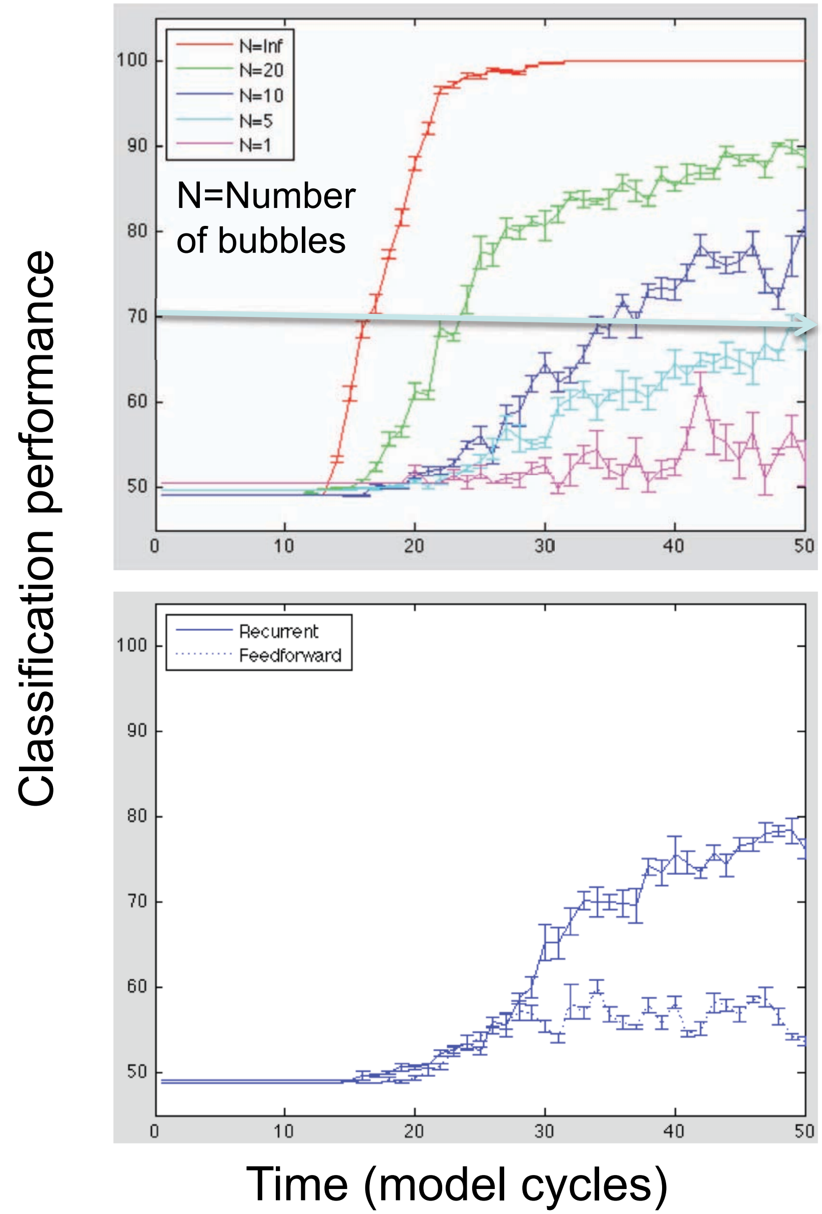




m00032
channel=21

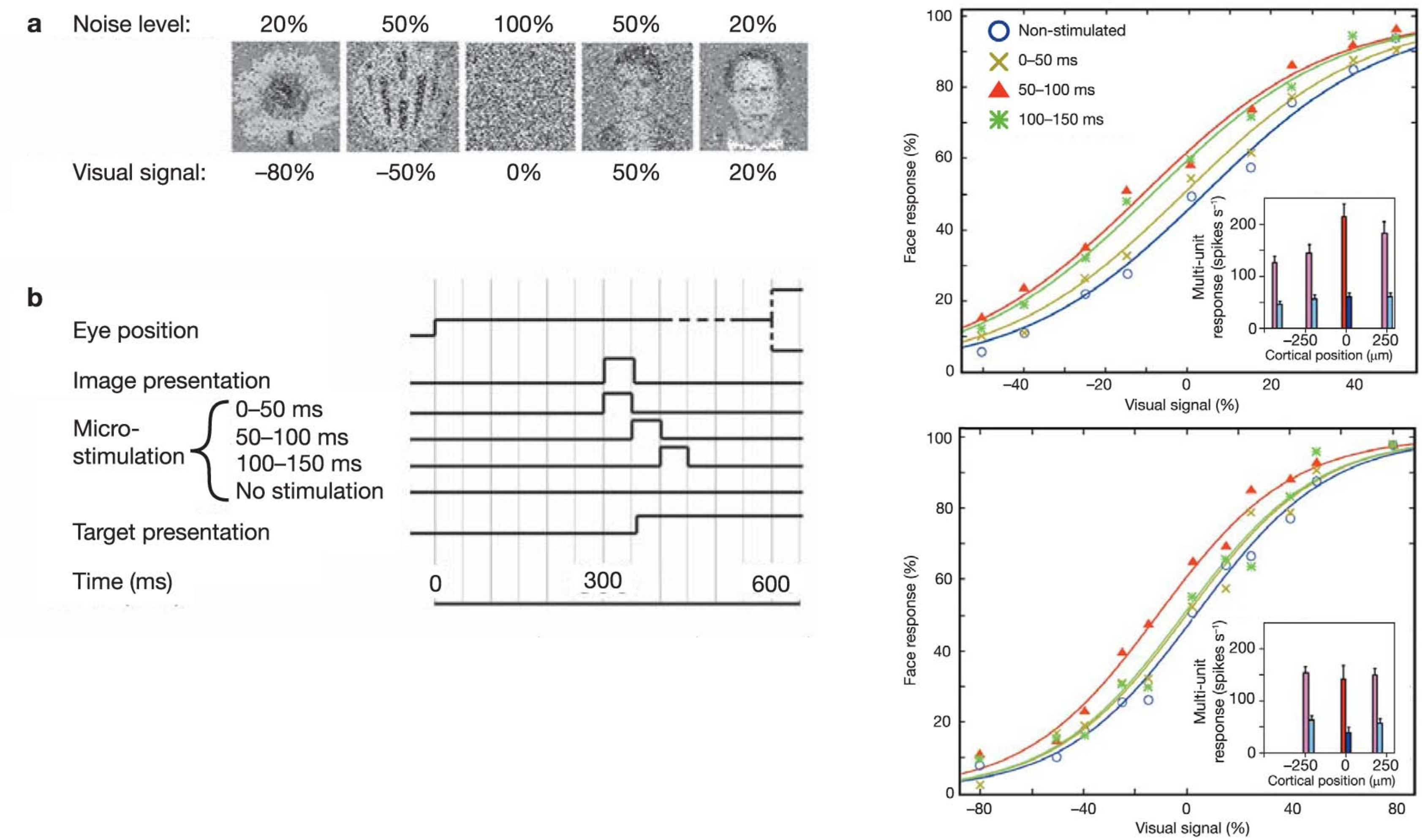
Top-down connections help perform object completion





Dean Wyatte, Randall O'Reilly, Hanlin Tang

Electrical stimulation can bias object recognition decisions



Afraz et al. *Microstimulation of inferotemporal cortex influences face categorization.* Nature (2006) **442**: 692-695.

Electrical stimulation in the human brain

Penfield & Perot. *The brain's record of auditory and visual experience. A final summary and discussion. Brain* (1963) **86**:595-696

Object recognition

Gabriel Kreiman

http://kreiman.hms.harvard.edu

gabriel.kreiman@tch.harvard.edu