Mixing hierarchical edge detection and medial axis models of object perception

Daniel D Leeds[1], Michael J Tarr[2,3]

[1]Computer and Information Science Department, Fordham University, Bronx, NY, [2]Center for the Neural Basis of Cognition, Carnegie Mellon University (CMU), Pittsburgh, PA, [3]Department of Psychology, CMU

COMPUTER AND INFORMATION SCIENCES DEPARTMENT

FORDHAM UNIVERSITY THE JESUIT UNIVERSITY OF NEW YORK

CNBC CENTER FOR THE NEURAL BASIS OF COGNITION

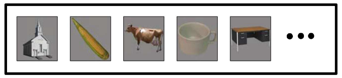
Carnegie Mellon University

Diverse models of cortical perception

- Multiple potential frameworks have been proposed to characterize cortical object perception
- Prominent models capturing holistic shapes and local-features show promise accounting for cortical responses to visual objects (Leeds 2013; Hung 2012; Yamins 2014)
- We show mixing holistic Shock graph and local-features SIFT models sometimes – but not always – provides a better account of cortical activity than does either model alone

Methods

- Participants shown photos of 60 real-world objects, 6 x each, passive viewing

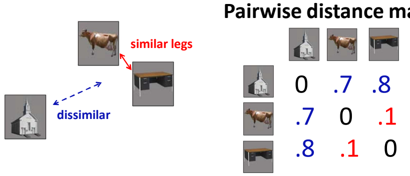


- BOLD signals recorded with slow event-related design (2 sec TR, partial coverage) for 5 subjects

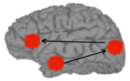
Representational dissimilarity analysis

- Representational dissimilarity: use pairwise distance matrix to show how stimuli are grouped by each neural and computational representation

Pairwise distance matrix



- **Neural:** 123-voxel sphere of responses at each location



- **Computational models:**
  - SIFT – bag of multi-scale Gabor-based visual words (similarities to Convolutional Neural Nets)
  - Shock graph – enhanced medial axis

- **Neural vs model comparison:** Spearman correlation between elements of neural and model distance matrices (Leeds 2013)

Joint SIFT/Shock graph model

Investigate contributions of each model to explain neural groupings (neural distance matrices) with **linear**, **bilinear**, and **quadratic** terms

Find $c_1, \ldots, c_6$ to best-fit

$$\text{neural}(i,j) \approx \boxed{\begin{aligned}\text{combine}(i,j) = c_1 + c_2\,\text{SIFT}(i,j) + c_3\,\text{Shock}(i,j)\\ + c_4\,\text{SIFT}(i,j)\,\text{Shock}(i,j)\\ + c_5\,\text{SIFT}(i,j)^2 + c_6\,\text{Shock}(i,j)^2\end{aligned}}$$

for each pair of stimuli *(stim$_i$, stim$_j$)*

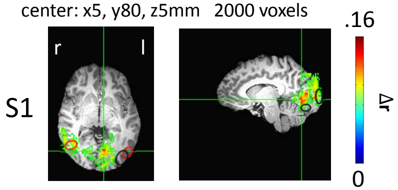
Find $r^{combine}$=corr(**neural**, **combine**)

Correlation increases from joint model

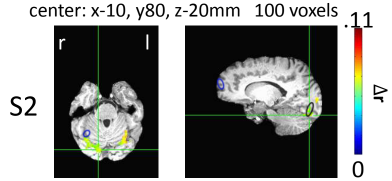
LOC localizer SIFT only Shock graph only Joint region center

q<1e-4 $\Delta r = r^{combine}-\max(r^{sift}, r^{shock})$ *via permutation test*

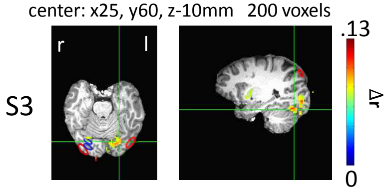
center: x5, y80, z5mm 2000 voxels



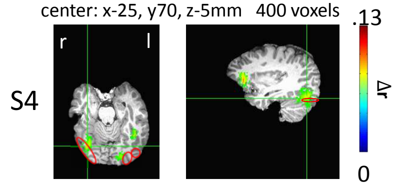
center: x-10, y80, z-20mm 100 voxels



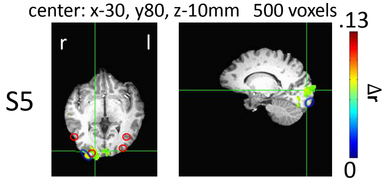
center: x25, y60, z-10mm 200 voxels



center: x-25, y70, z-5mm 400 voxels



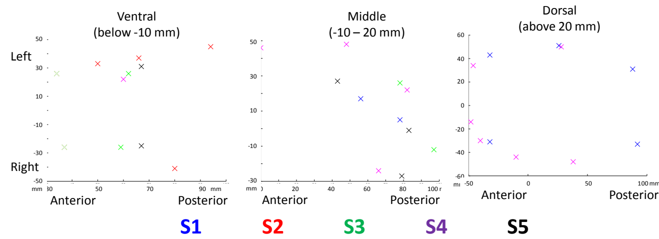
center: x-30, y80, z-10mm 500 voxels



Largest region of joint model correlation increase

- In medial occipital lobe
- Span 5-15 mm^3 volume
- Vary in overlap with LOC localizer
- Low overlap with regions of top SIFT and Shock correlations
- Provide 50-100% increase in match above single model (from r=.1 to r=.2)

Distribution of centers for regions of correlation increase

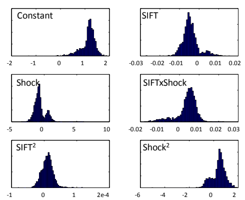


Centers are located

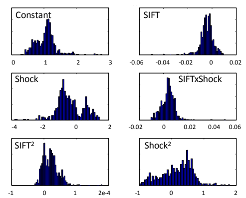
- Bilateral
- Posterior ventral (+ dorsal)
  - Inferior occipital
  - Fusiform
  - Lateral occipital
  - Lingual gyrus
  - Superior parietal
- Anterior dorsal
  - Prefrontal cortex
  - Smaller regions

Joint coefficients for increased correlation

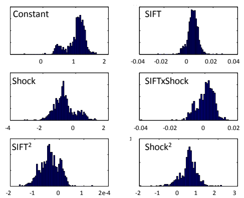
S1



S2



S4



- Typically 1-2 common coefficients per combination-model term
- Typically ~1 standard deviation away from 0
- Typical scale (-10 to 10 or -0.01 to 0.01) and values per combination-model term across subjects

Discussion

- **Non-linear combination of diverse models better accounts for voxel responses** in subset of mid- and high-level visual areas
- When combination provides true improvement, best fit **combination coefficients fairly consistent** across subjects and regions
- **Future directions:**
  - Expand to Convolutional Neural Networks
  - Expand to further non-linear model combinations

References

D. Leeds et al. Comparing visual representations across human fMRI and computational vision. *Jrnl of Vis,* 13(13):25, 1-27, 2013.

C. Hung et al. Medial axis shape coding in macaque inferotemporal cortex. *Neuron*, 74(6):1099–1113, 2012.

D. Yamins et al. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *PNAS*, 111(23): 8619-8624, 2014.

N. Kriegeskorte, et al. Representational similarity analysis – connecting the branches of systems neuroscience. *Front in Sys Neur*, 2(4): 2008.

Nowak, et al. Sampling strategies for bag-of-features image classification. In *Computer Vision - ECCV 2006*, 2006.

K. Siddiqi et al. Shock graphs and shape matching. *Intl Jrnl of Comp Vis*, 35:1, 1999.

Acknowledgments

Fordham University Faculty Research Grant; **Pennsylvania Department of Health**'s Commonwealth Universal Research Enhancement Program, **NIH** EUREKA Award #1R01MH084195A01, and the Temporal Dynamic of Learning Center at UCSD (NSF Science of Learning Center #SMA-1041755)