Comparing Object Selectivity in Visual Cortex and Topographic Deep Artificial Neural Networks

Davide Cortinovis (davide.cortinovis@unitn.it)
Center for Mind/Brain Sciences (CIMeC), University of Trento
Rovereto, 38068, Italy

Martin Hebart (hebart@cbs.mpg.de)
Department of Medicine, Justus Liebig University Giessen
Giessen, 35390, Germany
Max Planck Institute for Human Cognitive and Brain Sciences
Leipzig, 04103, Germany

Stefania Bracci (stefania.bracci@unitn.it)
Center for Mind/Brain Sciences (CIMeC), University of Trento
Rovereto, 38068, Italy

Abstract

The occipitotemporal cortex (OTC) exhibits category selectivity, with specialized regions responding to specific object categories. Topographic Deep Artificial Neural Networks (TDANNs) have been proposed as mechanistic models of this spatial and functional organization. However, a direct comparison of the visual and semantic features driving functional selectivity in the two systems is lacking. We analyzed fMRI data from three participants viewing 200 images of distinct body parts and inanimate objects, and compared OTC selectivity with TDANN activations. Body-, hand-, and tool-selective regions all showed strong category preferences. TDANNs displayed similar, though weaker, selectivity with blurrier category boundaries, especially for tools. Texture scrambling revealed that TDANN selectivity partly relies on local features: body and hand selectivity persisted despite global shape disruption, while tool selectivity disappeared, possibly due to their higher similarity with the other inanimate categories. These results represent a first step toward better characterizing and comparing functional selectivity in visual cortex and topographic models.

Keywords: fMRI; category selectivity; topographic models; visual cortex

Introduction

The occipitotemporal cortex (OTC) exhibits functionally-selective responses to specific object categories, such as faces, bodies, and scenes (Kanwisher, 2010). Finer-grained distinctions have also been observed, such as separable responses to hands and whole-bodies (Bracci et al., 2010).

Recent computational modeling work have developed Topographic Deep Artificial Neural Networks (TDANNs) that propose a mechanistic explanation of the emergence of the spatial organization in OTC (Lee et al., 2020). TDANNs incorporate a spatial loss function that constraints neighbouring units to have correlated firing patterns, leading to the emergence of category-selective clusters (Margalit et al., 2024). TDANNs allow direct test on what computational advantages topographic organization confers onto a system (Deb et al., 2025; Qian et al., 2024). However, the categories tested so far, like faces and scenes, vary on many visual and semantic properties. Therefore, it remains unclear whether topographic models can replicate the finer-grained functional selectivity profile at the image level for categories sharing higher semantic similarity. Moreover, the extent to which these similarities reflect shared computational mechanisms and sensitivity to the same features is still unknown.

Here, we compare the functional selectivity for bodies, hands, and tools in visual cortex and TDANNs, and test the influence of mid-level features on the emergence of category-selectivity in computational models.

Methods

fMRI dataset. We collected fMRI data from three participants in six scanning sessions. The stimulus set consisted of 200 images depicting whole bodies (without visible hands), hands, tools (e.g., hammer, pliers), manipulable (glass, plate), and non-manipulable (air baloon, standing lamp) objects.

ROI selection. To maximize spatial precision, ROIs were defined on the native unsmoothed surface of each participant, based on data from a separate localizer session, with a contrast of category vs. all. Here, we report the results for three ROIs in lateral OTC: a right-lateralized body-selective region in the Lateral Occipital Sulcus (LOS-body), a left-lateralized hand-selective region in the posterior Inferior Temporal Gyrus (ITG-hand), and a left-lateralized tool-selective area more anteriorly in ITG (ITG-tool).

Network. The analyses were performed on the TDANN model ($n = 5$ initializations), developed by Margalit et al. (2024) and consisting of a pretrained self-supervised ResNet-18 (trained with ImageNet). Analyses targeted the last (VTC-like) topographic layer. Category-selective clusters were identified by selecting contiguous selective units with a contrasts of category vs. all ($t > 3.5$), and the top-25 most selective units within the clusters were selected for further analyses among those passing this threshold. This was done to select only units that show the highest selectivity and to select a comparable number of units as category-selective voxels in the brain.

Functional selectivity Analyses We quantified category preferences in OTC and TDANNs using:

- A *Selectivity Index (SI)*, defined as:

$$SI = \frac{\mu_{\text{category}} - \mu_{\text{others}}}{\mu_{\text{category}} + \mu_{\text{others}}}$$

  with μ corresponding to the mean activation per category. The significance of each category SI (vs. each of the other) was assessed using 10000 permutation tests, and all reported results are significant at $p < 0.004$ (Bonferroni corrected with $N = 12$ comparisons), unless we report no effect.

- *Top-N rank analysis*: Proportion of stimuli from each category among top 25 responses (e.g., how many hands there are in the 25 stimuli eliciting the highest activation).

Texture Analyses To investigate the influence of mid-level visual features in eliciting functional selectivity in TDANNs, we conducted two texture scrambling analyses:

- *Gatys-style texture synthesis*: Texturized versions of images generated by matching the feature correlations of mid-level layers of a VGG16 (Gatys, Ecker, & Bethge, 2015), preserving local feature statistics and object identity while disrupting global shape.
- *Texforms*: Synthetic texture representations that retain local features but disrupt both global shape and object identity (Freeman & Simoncelli, 2011; Long, Yu, & Konkle, 2018).

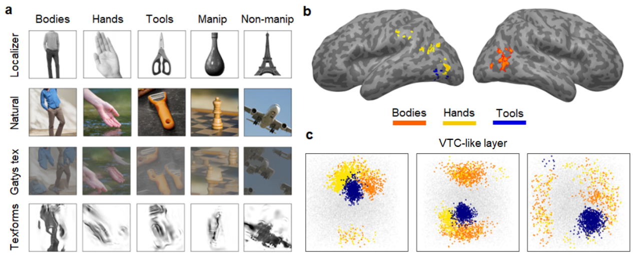

Figure 1: Stimulus set and ROIs. a) Example of stimuli used: localizer, natural images, and texture-scrambled versions of the natural images. b) Body, hand, and tool activations in an example participant. c) Body, hand, and tool clusters in the VTC-like layer of three initializations of the TDANN model.

Results

Testing category selectivity in visual cortex and TDANNs

Category-selective regions in the visual cortex showed distinct functional specialization: LOS-body exhibited strong selectivity for bodies ($SI = 0.56$), with 84% of its top-25 activating images depicting bodies, with a secondary presence of hands (8%). ITG-hand displayed sharp hand selectivity ($SI = 0.57$), primarily activating to hands (76%) with secondary responses to tools (16%) and bodies (8%). ITG-tool selectively responded to tools ($SI = 0.42$), which dominated its top-25 activations (72%), followed by a weaker sensitivity to manipulable objects (20%).

TDANNs developed similar albeit weaker category selectivity: body units showed selective responses to bodies ($SI = 0.44$) with 68% of bodies in the top-25 images, but retained some hand responses (28% in top-25); hand units similarly match cortical selectivity ($SI = 0.41$) with 52% of hand images in the top-25, but showed some sensitivity to non-hand categories (especially non-manipulable in the top-25: 20%), and tool units demonstrated positive tool responses ($SI = 0.25$; 48% top-25), but with no significant difference compared to the SI of the other inanimate objects.

Texture scrambling weakens but preserves functional selectivity in TDANNs

To test the influence of mid-level visual features in eliciting category selectivity in TDANNs, we conducted the same analyses with texturized images.

Gatys-texture scrambling reduced but preserved category specificity, with body units showing retained body preference ($SI = 0.28$) and 56% images of bodies in the top-25, but increased hand responses ($SI = 0.1$, 32% in top-25), hand units maintaining selectivity ($SI = 0.35$, 40% in top-25) but with increased sensitivity to tools (36% in top-25), and tool units showing little distinction between tools and other inanimate objects, especially manipulable (tool $SI = 0.23$, manipulable $SI = 0.15$, no difference between tools and other inanimate objects, tool 35% in top-25, and manipulable 40%).

The analysis with texforms revealed that body units maintained category preference ($SI = 0.23$, 60% in top-25), but with increased sensitivity to non-preferred stimuli (e.g., manipulable appeared 25% in top-25), hand units weakening in selectivity ($SI = 0.11$, 40% in top-25) while increasing responses to inanimate objects (60% combined in top-25), and tool units further losing specificity ($SI = 0.042$) with top-25 dominated by the other inanimate objects (68% combined).

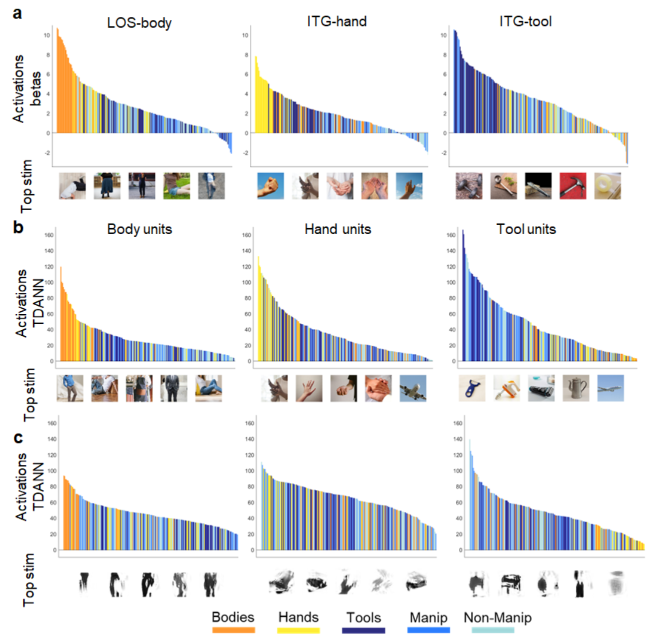

Figure 2: Functional selectivity. Sorted activations and top-5 images for a) body, hand, and tool selective areas in OTC, b) body, hand, and tool units for original images in TDANNs, c) body, hand, and tool units for texform images in TDANNs.

Discussion

In OTC, we observed a categorical and distinct selectivity profile for body, hand, and tool areas (Peelen & Downing, 2017). TDANNs, while showing similarity in terms of category preferences (especially for bodies and hands), had weaker selectivity and blurrier category boundaries, with more graded tuning for non-preferred categories. Moreover, the fact that selectivity is partially retained for texturized images suggest that TDANNs may rely on mid-level texture statistics for their functional organization (Jagadeesh & Gardner, 2022); at the same time, they struggle in capturing distinctions within inanimate objects, possibly because these stimuli present higher visual and semantic similarity (Cortinovis, Peelen, & Bracci, 2025). In future analyses we will test the influence of texture scrambling on the functional selectivity of category-selective areas, and we will examine how these topographic constraints directly influence, and potentially clash with, functional selectivity, to highlight the potential conflict or benefit of spatial organization onto functional correspondence with visual cortex.

References

Bracci, S., Ietswaart, M., Peelen, M. V., & Cavina-Pratesi, C. (2010). Dissociable neural responses to hands and non-hand body parts in human left extrastriate visual cortex. *Journal of neurophysiology*, *103*(6), 3389–3397.

Cortinovis, D., Peelen, M. V., & Bracci, S. (2025). Tool representations in human visual cortex. *Journal of Cognitive Neuroscience*, *37*(3), 515–531.

Deb, M., Deb, M., & Murty, N. (2025). Toponets: High performing vision and language models with brain-like topography. *arXiv preprint arXiv:2501.16396*.

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature neuroscience*, *14*(9), 1195–1201.

Gatys, L., Ecker, A. S., & Bethge, M. (2015). Texture synthesis using convolutional neural networks. *Advances in neural information processing systems*, *28*.

Jagadeesh, A. V., & Gardner, J. L. (2022). Texture-like representation of objects in human visual cortex. *Proceedings of the National Academy of Sciences*, *119*(17), e2115302119.

Kanwisher, N. (2010). Functional specificity in the human brain: a window into the functional architecture of the mind. *Proceedings of the national academy of sciences*, *107*(25), 11163–11170.

Lee, H., Margalit, E., Jozwik, K. M., Cohen, M. A., Kanwisher, N., Yamins, D. L., & DiCarlo, J. J. (2020). Topographic deep artificial neural networks reproduce the hallmarks of the primate inferior temporal cortex face processing network. *BioRxiv*, 2020–07.

Long, B., Yu, C.-P., & Konkle, T. (2018). Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proceedings of the National Academy of Sciences*, *115*(38), E9015–E9024.

Margalit, E., Lee, H., Finzi, D., DiCarlo, J. J., Grill-Spector, K., & Yamins, D. L. (2024). A unifying framework for functional organization in early and higher ventral visual cortex. *Neuron*, *112*(14), 2435–2451.

Peelen, M. V., & Downing, P. E. (2017). Category selectivity in human visual cortex: Beyond visual object recognition. *Neuropsychologia*, *105*, 177–183.

Qian, X., Dehghani, A. O., Farahani, A. B., & Bashivan, P. (2024). Local lateral connectivity is sufficient for replicating cortex-like topographical organization in deep neural networks. *bioRxiv*, 2024–08.