Representational tuning models: Uniting representational similarity analysis and neural encoding models.

Stan Bergey (s.r.bergey@tilburguniversity.edu)
Department of Cognitive Science and Artificial Intelligence, Tilburg University, Warandelaan 2
5037 AB Tilburg, Netherlands

Luis Martín-Roldán Cervantes (l.martin.roldan.cervantes@gmail.com)
EngD Program Software Technology, Technical University Eindhoven, De Zaale
5600 MB Eindhoven, Netherlands

Jorge Almeida (jorgealmeida@fpce.uc.pt)
PROACTION lab, Faculty of Psychology and Educational Sciences, University of Coimbra, Rua do Colégio Novo
3001-802, Coimbra, Portugal

Ben Harvey (b.m.harvey@uu.nl)
Department of Experimental Pschology, Utrecht University, Heidelberglaan 1
3584 CS Utrecht, Netherlands

Abstract

In fMRI analysis, neural encoding models reveal how individual voxels respond as a function of continuous stimulus parameters, and how response function parameters change within and between brain areas. Conversely, representational similarity analysis reveals structure in the responses of a region of interest (ROI) to any stimuli. Here we develop representational tuning models to unite these approaches. These response models first rescale the representational dissimilarity matrix (for an ROI) to a 2-dimensional representational space then find (for each voxel) the Gaussian function within this 2D space that best predicts the voxel's response to all stimuli. By deriving continuous response function parameters from the ROI's responses, this approach requires no a priori hypothesis of stimulus parameters underlying the response function. It thereby allows application of neural encoding models to arbitrary stimulus sets. We test this approach for responses from the Natural Scenes Dataset within 12 visual field maps. We show that representational tuning models significantly predict voxels' responses to natural images in higher-level (but not early) visual field maps, especially when sampling from other visual field maps, and we demonstrate that the principal components of representational spaces reflect the spatial structure of responses across the cortical surface within an ROI.

Keywords: Representational Similarity Analysis, fMRI, neural tuning, vision, object recognition

Introduction

Parametric neural population encoding models (Dumoulin & Wandell, 2008b; Kay et al., 2008) are widely used to characterize the response functions of individual fMRI voxels and how these change within and between brain areas. These describe a response function of a continuous stimulus parameter like visual field position (Harvey & Dumoulin, 2011), auditory frequency (Thomas et al., 2015), numerosity (Harvey et al., 2013) or event timing (Harvey et al., 2020; Hendrikx et al., 2022). But many stimuli, like objects, differ in complex high-dimensional ways. This makes it difficult to determine which stimulus parameter(s) the neural response function follows. Here, representational similarity analysis (RSA) (Kriegeskorte et al., 2008) is widely used to analyze the relationships between patterns of responses to different stimuli. As RSA uses response patterns of larger brain areas, it does not characterize responses of individual voxels, or how these change within and between brain areas.

Multidimensional scaling of representational dissimilarity matrices (RDMs) (Figure 1A) makes response pattern differences more interpretable by projecting them to distances in a two-dimensional (2D) space (Figure 1B). Here we hypothesized that a single voxel's response to arbitrary stimuli could be described by a parametric function in the 2D representational space (Figure 1C) of the brain area the voxel lies in. We further hypothesized that, for fMRI data, the two dimensions recovered by multidimensional scaling (i.e. principal coordinates) may, in part, follow the structure of the cortical surface within the brain area from which the RDM is derived.

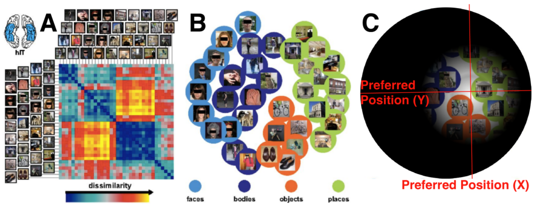

Figure 1: Schematic of the representational tuning model. (A) An RDM describes correlations between responses to arbitrary image pairs in a brain area. (B) This undergoes multidimensional scaling to give a 2D representational space. (C) For each voxel in the brain area, we find the Gaussian in this 2D representational space that best describes the voxel's response amplitude to all images. The Gaussian center is the voxel's preferred position in the 2D space

Methods

We used of the Natural Scenes Dataset (Allen et al., 2022), 7T fMRI responses to between 7,000 and 10,000 images per participant, each presented at most 3 times. This includes the response amplitude (beta) of every voxel's response to every image presentation from 8 participants, plus visual field mapping data.

We used the visual field mapping data to identify visual field maps in the early visual cortex (V1-3), lateral visual stream (LO-1&2, TO-1&2) and ventral visual stream (hV4, VO-1&2, PHC-1&2), our regions of interest (ROIs). These small ROIs each correspond to one visual image representation.

Subsequent analysis used responses to natural scene images in the same participants. We first made a training set of these responses, averaging response amplitudes to the same image across all but the last presentation of the image. We used this training set to determine an RDM: the difference (1-correlation) between the response patterns to every image pair. We derived a 2D representational space using classical multi-dimensional scaling, giving each image a position in the ROI's 2D representational space.

For each voxel in the ROI, we then fit the isotropic (circular) Gaussian that best describes the voxel's response amplitudes to each image, considering the image's positions in the 2D space. We evaluated the model's fit on an independent test set: responses to the final single presentation of the same images. We then tested whether model fits in each visual field map had a median cross-validated variance explained ($CVVE$) significantly above zero.

In each visual field map in each hemisphere, we quantified the spatial structure of voxels' fitted preferred representational tuning position using (Spearman) correlation between the cortical surface distance between each pair of voxels and the distance between their preferred representational tuning position.

We then extended this within-ROI approach by fitting the response function within the representational space of an ROI other than that the voxel lay in, a between-ROI model. For the voxels in each target ROI (whose responses we try to explain), we selected the source ROI that best fit the training set responses, then evaluated the resulting model's predictions on the test set.

Results

Figure 2 shows how the CVVE of both within-ROI and between-ROI models across Subject 1's cortical surface (Fig 2a) and in the average in each visual field maps across all subjects. The within-ROI model can explain some of the variance, particularly in anterior, higher-order visual field maps in the lateral and ventral occipito-temporal cortex (Fig 2b, left). It predicted response variance significantly above 0 in V1, PHC-2, LO-2 and TO-1. While CVVE is consistently low (under 0.08 R^2) in these areas, it is often around half of the correlation between response amplitudes in the training and test sets, the noise ceiling. The response functions' extents also decreased from posterior to anterior visual field maps, so voxels responded to progressively more specific parts of the representational space (Fig 2c). Lastly, the distance between the preferred representation space position of any pair of voxels was significantly correlated with the cortical surface distance between these voxels (Fig 3), so the representation tuning properties changed systematically across the cortical surface within each ROI.

For the between-ROI models, all ROIs' average CVVE was significantly above 0 except in V3, hV4, LO-1 and LO-2, meaning our model can capture variance over most of the visual system, and can explain more variance in PHC-1, PHC-2, TO-1 and TO-2.

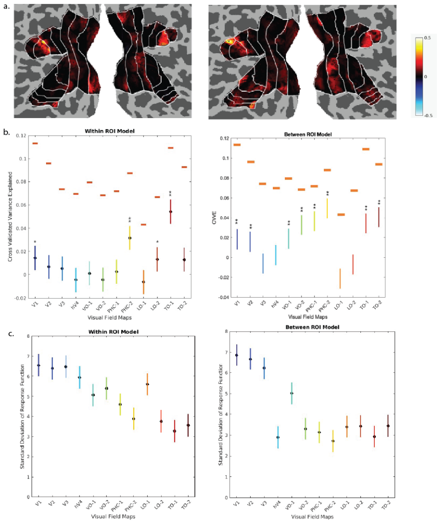

Figure 2: **a** The cross-validated variance explained of each voxel on the cortical surface (for Subject 1). **b** Comparisons of average $CVVE$ across both hemispheres and all subjects. The star symbols indicate where the distribution of $CVVE$ for each ROI is significantly above zero ($*p < 0.05, **p < 0.01$). The orange bars are the average noise ceiling of each ROI across all subjects.

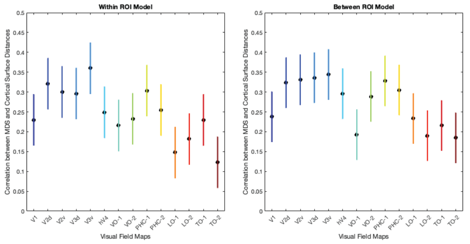

Figure 3: Average Spearman Correlation coefficients (ρ) between distances between voxels' preferred Representational Space positions and distances on the cortical surface across all subjects.

Discussion

Our results show that individual fMRI voxel's responses can be meaningfully described as a sample of the representational space within the voxel's brain area or another area, particularly for higher-order visual areas. This shows that parametric encoding models can be used to describe neural responses for arbitrary sets of stimuli, without needing hypothetical parameters describing the stimulus states. Our between-ROI model performed better than our within-ROI model, likely through its more flexible source ROI selection.

Our results also demonstrate that each voxel's fitted preferred position varies systematically across the cortical surface within each ROIs. This implies that the two dimensions obtained by multi-dimensional scaling of RDMs from fMRI data (i.e. the first two principal coordinates in principal coordinate analysis) reflect the structure of the cortical surface within the ROI from which the RDM is derived. However, the strength of correlations between distances on the cortical surface and representational space do not closely follow the CVVE across ROIs. Therefore, well-fitting models are not simply capturing visual field map structure. Likewise, our models' poor performance in early visual field maps suggests they are not simply picking up the spatial structure of image contrast (Dumoulin & Wandell, 2008a; Kay et al., 2008), but instead higher-level object or image properties that drive later responses.

Finally, our results show that this method can also be applied using the representational space of a different ROI than the target voxel lies in. This allows us to ask how neural responses in a voxel can be explained as a sampling of a representation space in a distant brain area, much like connective field modelling (Haak et al., 2013) describes how higher-level visual spatial responses can be explained as a sample from an earlier visual field map. This has the potential to reveal relationships between stimulus representations in different brain areas.

Acknowledgments

This work was supported by: Netherlands Organization for Scientific Research (452.17.012 to B.M.H.); Portuguese Foundation for Science and Technology (IF/01405/2014 to B.M.H.); European Research Council (ERC) (Starting Grant 802553 “ContentMAP” and Widening Programme Grant 101087584 ”CogBooster” to J.A.)

Bibliography

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., Nau, M., Caron, B., Pestilli, F., Charest, I., et al. (2022). A massive 7t fmri dataset to bridge cognitive neuroscience and artificial intelligence. *Nature neuroscience*, *25*(1), 116–126.

Dumoulin, S. O., & Wandell, B. A. (2008a). Population receptive field estimates in human visual cortex. *Neuroimage*, *39*(2), 647–660.

Dumoulin, S. O., & Wandell, B. A. (2008b). Population receptive field estimates in human visual cortex. *NeuroImage*, *39*(2), 647–660. `https://doi.org/10.1016/j.neuroimage.2007.09.034`

Haak, K. V., Winawer, J., Harvey, B. M., Renken, R., Dumoulin, S. O., Wandell, B. A., & Cornelissen, F. W. (2013). Connective field modeling. *NeuroImage*, *66*, 376–384. `https://doi.org/10.1016/j.neuroimage.2012.10.037`

Harvey, B. M., & Dumoulin, S. O. (2011). The Relationship between Cortical Magnification Factor and Population Receptive Field Size in Human Visual Cortex: Constancies in Cortical Architecture. *Journal of Neuroscience*, *31*(38), 13604–13612. `https://doi.org/10.1523/JNEUROSCI.2572-11.2011`

Harvey, B. M., Dumoulin, S. O., Fracasso, A., & Paul, J. M. (2020). A network of topographic maps in human association cortex hierarchically transforms visual timing-selective responses. *Current Biology*, *30*(8), 1424–1434.

Harvey, B. M., Klein, B. P., Petridou, N., & Dumoulin, S. O. (2013). Topographic representation of numerosity in the human parietal cortex. *Science*, *341*(6150), 1123–1126.

Hendrikx, E., Paul, J. M., van Ackooij, M., van der Stoep, N., & Harvey, B. M. (2022). Visual timing-tuned responses in human association cortices and response dynamics in early visual cortex. *Nature Communications*, *13*(1), 3952. `https://doi.org/10.1038/s41467-022-31675-9`

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, *452*(7185), 352–355.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 4.

Thomas, J. M., Huber, E., Stecker, G. C., Boynton, G. M., Saenz, M., & Fine, I. (2015). Population receptive field estimates of human auditory cortex. *Neuroimage*, *105*, 428–439.