

Using population receptive field models to elucidate spatial integration in high-level visual cortex

Sonia Poltoratski (sonia09@stanford.edu)

Stanford University Dept. of Psychology, 450 Serra Mall,
Stanford, CA 94305, USA

Kendrick Kay (kay@umn.edu)

University of Minnesota Center for Magnetic Resonance Research, 2021 6th St SE,
Minneapolis, MN 55455, USA

Kalanit Grill-Spector (kalanit@stanford.edu)

Stanford University Dept. of Psychology, 450 Serra Mall,
Stanford, CA 94305, USA

Abstract:

While spatial information and biases have been consistently reported in high-level face regions, the functional contribution of this information toward face recognition behavior is unclear. Here, we propose that spatial integration of information plays a critical role in a hallmark phenomenon of face perception: holistic processing, or the tendency to process all features of a face concurrently rather than independently. We sought to gain insight into the neural basis of face recognition behavior by using a voxelwise encoding model of spatial selectivity to characterize the human face network using both typical face stimuli, and stimuli thought to disrupt normal face perception. We mapped population receptive fields (pRFs) using 3T fMRI in 6 participants using upright as well as inverted faces, which are thought to disrupt holistic processing. Compared to upright faces, inverted faces yielded substantial differences in measured pRF size, position, and amplitude. Further, these differences increased in magnitude along the face network hierarchy, from IOG- to pFus- and mFus-faces. These data suggest that pRFs in high-level regions reflect complex stimulus-dependent neural computations that underlie variations in recognition performance.

Keywords: fMRI; high-level vision; face perception; pRF modeling; holistic processing; visual perception

Introduction

High-level visual processing in the ventral 'what' pathway is thought to be abstracted from and invariant to spatial properties of objects, like their location or size in the visual field; this allows us to efficiently recognize objects across the near-infinite number of 2D images they might project onto the retina (DiCarlo

& Cox, 2007). However, this classic account is challenged by consistent findings of spatial biases and information in the ventral temporal cortex (VTC), the end-stage of the ventral stream (Hasson, Levy, Behrmann, Hendler, & Malach, 2002; Kobatake & Tanaka, 1994; Levy, Hasson, Avidan, Hendler, & Malach, 2001; Sayres & Grill-Spector, 2008). Recently, our group has used population receptive field (pRF) methods to quantify voxel-wise spatial selectivity in face-selective regions in the human ventral temporal cortex (VTC). However, it is unknown if and how pRF properties in VTC contribute to visual perception. That is, do these representations of space exert only a general influence on where in the visual field high-level recognition is optimal, or do they play a specific role in determining how we recognize objects?

In the current work, we used the compressive spatial summation (CSS) pRF model (Kay, Weiner, & Grill-Spector, 2015; Kay, Winawer, Mezer, & Wandell, 2013) to relate spatial representations in face-selective areas to a hallmark of face recognition behavior: holistic processing. A prominent theory of face perception describes that efficient and accurate recognition requires the holistic processing of an entire face at once, rather than processing individual facial features. While human behavior supports this theory (Richler, Cheung, & Gauthier, 2011; Richler, Palmeri, & Gauthier, 2012; Yin, 1969; Young, Hellawell, & Hay, 1987), the neural mechanisms underlying holistic processing have proven elusive and computationally underspecified. We posit that, at a minimum, holistic processing requires spatial integration of information across features, and that pRF measurements may



reflect such integration. To evaluate the relationship between spatial representations and face recognition behavior, we mapped pRFs in early visual cortex (V1-hV4) and three constituent regions of the human face network: one region on the inferior occipital gyrus, IOG-faces (sometimes termed OFA), and two regions on the fusiform gyrus, pFus-faces (posterior fusiform; FFA-1) and mFus-faces (mid fusiform; FFA-2).

Methods & Materials

fMRI Approach

Scanning protocol Six experienced fMRI participants (3 women) ages 24-30 participated in the experiment. All fMRI experiments were conducted on a 3T GE fMRI scanner equipped with an eye tracker, which monitored subjects' fixation. We used a 32-channel head coil and single-shot EPI with 2.2 mm isotropic voxels and 2s TR. Standard preprocessing (motion correction, linear trend removal) was performed using mrVista tools. Retinotopic and face-selective regions were defined in a separate mapping session using standard phase-encoded retinotopic mapping (Engel, Glover, & Wandell, 1997) and a functional localizer (Stigliani, Weiner, & Grill-Spector, 2015), respectively.

Population receptive field (pRF) mapping

Background The pRF model (Dumoulin & Wandell, 2008; Wandell & Winawer, 2015) is a method for estimating the location and area in the visual field to which a particular voxel responds. Our model assumes a circular 2D Gaussian pRF with a compressive nonlinearity (Kay et al., 2013). PRF properties for a voxel reflect the combined receptive fields of the neural population in the voxel and are well-aligned with single-neuron RF properties (Wandell & Winawer, 2015).

Mapping procedure Each pRF mapping run presented participants with randomly intermixed upright and inverted faces at 25 spatial locations, as shown in Figure 1b. Timing for each run and participant was randomized, and faces appeared in 4s blocks at 2Hz at each spatial location. Participants completed 8-12 mapping runs, performing a difficult 1-back task on letters at fixation throughout.

Model implementation Prior to comparing pRF properties mapped with upright vs. inverted faces, we sought to account for low-level visual differences between the two stimulus conditions. To do so, we coded the stimulus location as an absolute contrast image reflecting visual features of the faces, rather

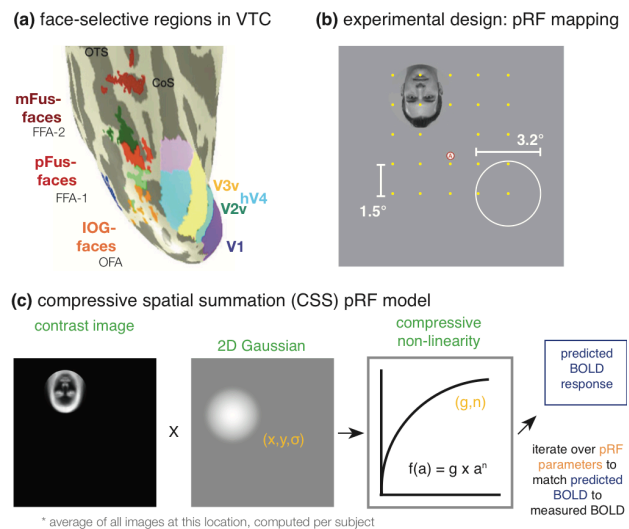


Figure 1: (a) Target face-selective and early visual ROIs. (b) pRF mapping stimulus layout. (c) Implementation of the CSS pRF model.

than as a simple binary mask indicating face location (see Figure 1c). The implementation of the CSS model is depicted in Figure 1c and described further in (Kay et al., 2013). To fit the CSS model, we first estimated Betas for the inverted and upright conditions in 25 spatial locations via standard general linear modeling. We then optimized pRF parameters to best fit the Betas for each voxel, separately for the upright and inverted conditions. This yielded, for each condition type, estimates of the voxel's pRF position (X,Y), size (σ), gain, and exponent n.

Results & Discussion

In early visual areas V1-hV4, we found no differences in pRF properties mapped with upright vs. inverted faces (Table 1). However, at late stages of the face-processing hierarchy, we observed for inverted faces: (a) substantial shifts of pRF centers downward, (b) reductions in pRF size, (c) reductions in gain, and (d) reductions in model goodness-of-fit (r^2). Moreover, the magnitude of these differences increased hierarchically from IOG-faces, where differences between upright and inverted mapping were not significant, to pFus- and mFus-faces (Figure 3; Table 1). A simulation analysis revealed that changes in signal-to-noise ratio could not account for the observed shift and size changes; that is, reducing the gain of responses in the upright condition and adding noise until the r^2 matched that of the inverted condition does not produce systematic changes in position or size.

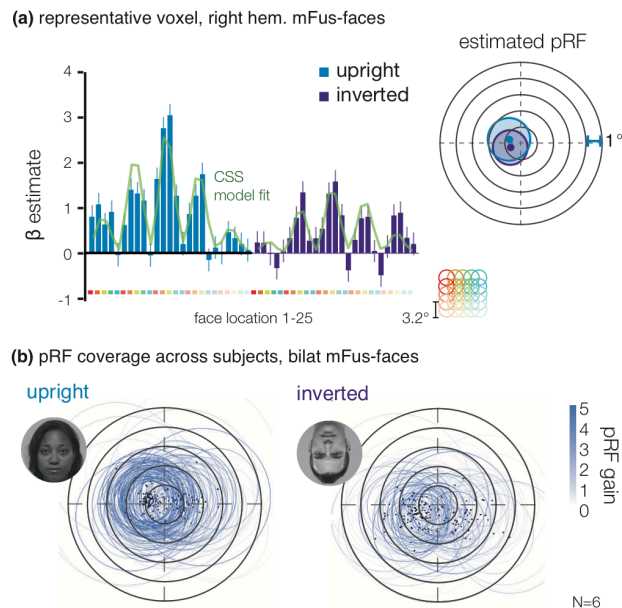


Figure 2: **(a)** Estimated Betas and model fits for a single representative voxel in the right mFus-faces ROI. Parameters of the fit CSS model (green line) can be translated to a pRF in retinotopic space (inset). **(b)** Visual field coverage across pRFs in mFus-faces. scaled by estimated gain.

The observed changes in size, position, and gain are not present in any retinotopic visual areas, nor are they seen at the earliest node of the face network, IOG-faces. This implies that the current implemented model, which utilizes the spatial location of visual features in the mapping stimulus (e.g. Figure 4, top left), is sufficient to account for spatial responses to both upright and inverted faces in early regions. However, we see that a modification to the model is needed to characterize spatial representations in fusiform regions: ideally, a good candidate model should produce convergent pRF estimates across both the upright and inverted conditions. Prior work

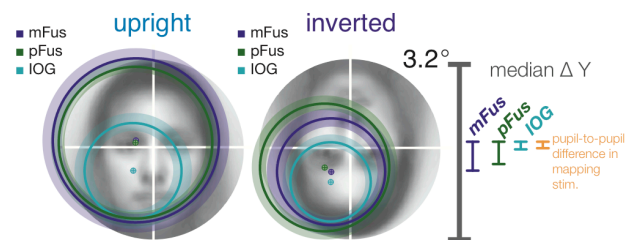


Figure 3: Summary of median pRF location and size in face-selective areas, plotted over the average of all mapping face stimuli at presented size (3.2°). Shaded regions illustrate \pm SE across subjects' estimated pRF size.

indicates the importance of internal face features, and in particular the eyes, in both driving responses in face-selective cortex (Issa & Dicarlo, 2012) and in determining behavioral performance (Royer et al., 2018). To evaluate whether the location of the eyes or of external features better predicts spatial responsiveness in mFus- and pFus-faces, we re-fit the CSS pRF model using these alternative codings of the mapping stimuli (Figure 4).

Notably, the internal features coding (right column) produces equivalent estimates of pRF position across the mapping stimuli. This implies that spatial responses in fusiform face-selective regions are selectively driven by the location of internal face features, rather than all contrast in the image (left column) or simply the location of the eyes (middle column). However, this model does not provide a fully parsimonious account of pRFs across upright and inverted faces, as differences in size and gain persist. Additional work is needed to refine this model and to evaluate whether differential weighting of facial features may yield more equivalent estimates across upright and inverted stimuli; human recognition has been shown to differentially weight certain facial features, e.g. the eyes and mouth are more informative than the nose for recognition (Schyns, Bonnar, & Gosselin, 2002).

The current work points to a neural mechanism for a prominent, but computationally underspecified, aspect of high-level face recognition: holistic processing across features, which is thought to underlie normal face recognition. We mapped pRFs in the face network with stimuli thought to disrupt holistic processing, e.g. inverted faces, and observed substantial differences in their position, size, and gain relative to those mapped with upright faces. As this occurred even when participants did not actively engage in face perception, performing a challenging

Table 1: pRF parameter changes by ROI (*bold = t-test significance across subs, $p < .05$*)

	ΔY	ΔX	Δ size	Δ gain
V1	.08° (SE=.04°)	.02° (.04°)	.11° (.07°)	.28 (.39)
hV4	.11° (.08°)	.03° (.07°)	.26° (.13°)	1.2 (.59)
IOG- faces	.23° (.15°)	.05° (.08°)	.30° (.17°)	.97 (.51)
pFus- faces	.45° (.20°)	.03° (.16°)	.43° (.15°)	2.0 (.67)
mFus- faces	.62° (.17°)	.14° (.20°)	1.04° (.26°)	3.2 (.94)

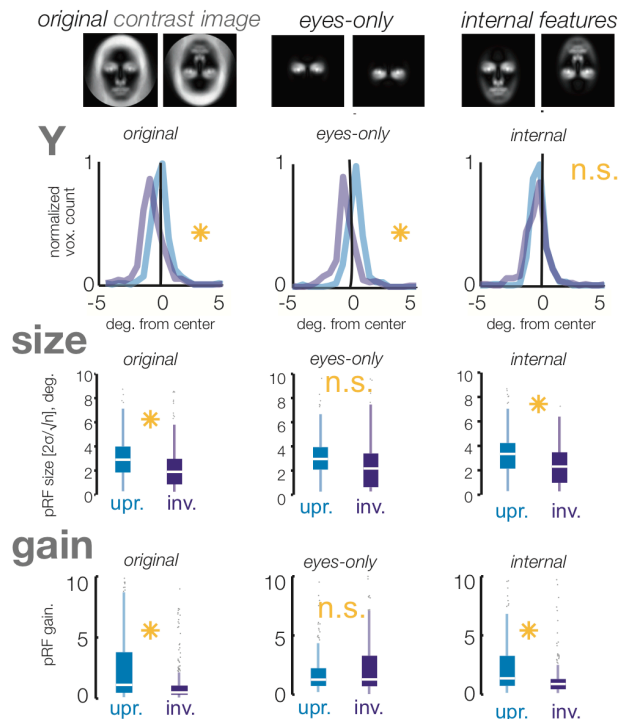


Figure 4: Manipulations of stimulus coding and their effects on estimated pRF parameters in mFus-faces. Significance (*) is marked for t -test $p < .05$ across subjects.

task away from the mapping stimuli, spatial representations in the VTC may functionally constrain recognition behavior, rather than simply reflect it.

Acknowledgments

This work was supported by NSF Grant BSC-1756035.

References

- Dicarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, *11*, 333–341.
- Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *Neuroimage*, *39*, 647–660.
- Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex*, *7*, 181–192.
- Hasson, U., Levy, I., Behrmann, M., Hendler, T., & Malach, R. (2002). Eccentricity Bias as an Organizing Principle for Human High-Order Object Areas. *Neuron*, *34*, 479–490.
- Issa, E. B., & Dicarlo, J. J. (2012). Precedence of the Eye Region in Neural Processing of Faces. *The Journal of Neuroscience*, *32*, 16666–16682.
- Kay, K. N., Weiner, K. S., & Grill-Spector, K. (2015). Attention reduces spatial uncertainty in human ventral temporal cortex. *Current Biology*, *25*, 595–600.
- Kay, K., Winawer, J., Mezer, A., & Wandell, B. (2013). Compressive spatial summation in human visual cortex. *Journal of Neurophysiology*, *110*, 481–494.
- Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology*, *71*, 856–867.
- Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center-periphery organization of human object areas. *Nature Neuroscience*, *4*, 533–539.
- Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic Processing Predicts Face Recognition. *Psychological Science*. doi:10.1177/0956797611401753
- Richler, J. J., Palmeri, T. J., & Gauthier, I. (2012). Meanings, mechanisms, and measures of holistic processing. *Frontiers in Psychology*, *3*, 553.
- Royer, J., Blais, C., Charbonneau, I., Déry, K., Tardif, J., Duchaine, B., et al. (2018). Greater reliance on the eye region predicts better face recognition ability. *Cognition*, *181*, 12–20.
- Sayres, R., & Grill-Spector, K. (2008). Relating retinotopic and object-selective responses in human lateral occipital cortex. *Journal of Neurophysiology*, *100*, 249–267.
- Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science : a Journal of the American Psychological Society / APS*, *13*, 402–409.
- Stigliani, A., Weiner, K. S., & Grill-Spector, K. (2015). Temporal Processing Capacity in High-Level Visual Cortex Is Domain Specific. *Journal of Neuroscience*, *35*, 12412–12424.
- Wandell, B. A., & Winawer, J. (2015). Computational neuroimaging and population receptive fields. *Trends in Cognitive Sciences*, *19*, 1–9.
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*, 141–145.
- Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*, 747–759.