A neural encoding model of area PL, the earliest face selective region in monkey IT

Charles F. Cadieu*, Elias B. Issa*, and James J. DiCarlo

Progress has been made in understanding early sensory areas (e.g. olfactory bulb, retinae, V1) by constructing encoding models at the neural level. However, our understanding of human face processing has been hindered by a lack of such models. Building models of face selective neurons would lead to a mechanistic explanation of cognitive phenomena observed in the psychophysics and fMRI communities. We provide a first step towards this goal by modeling the earliest face selective region in monkey IT, area PL, which has been proposed as a gateway to face processing and may serve as a face detection module. We tested a wide array of image-based encoding models and found that hierarchical models that pool over local features captured PL responses across 3043 images at 87% cross-validated explained variance (65% mean explained variance for sites, n=150). Models with the highest explanatory power incorporated localized sub-features, feature rectification, and a tolerance operation over space and scale. Those models also demonstrated similar properties to PL according to a phenomenological 'scorecard' (e.g. similar rankings across face parts and non-face images). We compared these models with the 'word model' in the field -- that 'face neurons' signal face presence -- by measuring human judgements of 'faceness' (n=210 subjects). These judgements did not match the phenomenological scorecard and correlated poorly with PL responses (22% explained variance). Together these results provide new perspective on the early stages of face processing in IT: PL is better viewed as a non-linear image operation than as a cognitive indicator of face presence. In summary, this work is the first to create image-computable encoding models of face selective neurons. These models may bridge the gap between the cognitive understanding of face processing and a mechanistic understanding of the neural basis of face processing.

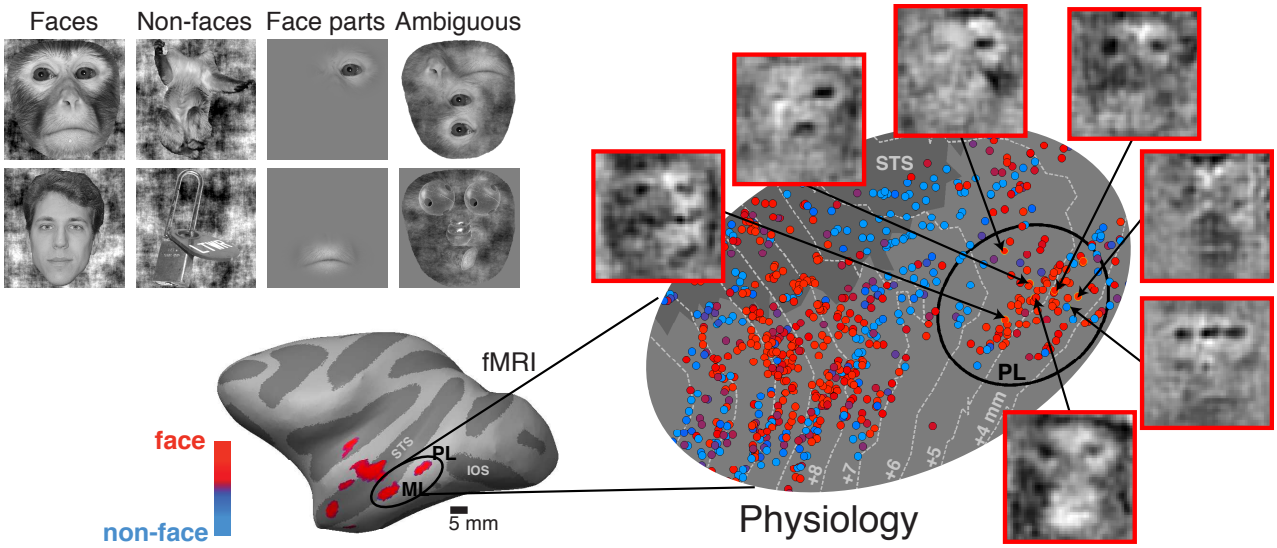

Figure 1 - Stimuli and neural recordings from area PL. Upper-left) We used stimuli that spanned a large range of image variation. **Lower-left)** fMRI maps of face versus object selectivity were used to target recordings of area PL. **Right)** Local, high-resolution neural maps were constructed by using an x-ray based system for electrode localization, and sites in physiologically defined area PL (black circle) were tested across a battery of images spanning faces to non-faces (Issa & DiCarlo, 2012). Models were fit to the population average response and to individual sites. Visualizations of models fit to individual sites are shown as an iconic image with an arrow pointing to the modeled site. Many of the estimated models contained eyelike features or even coarsely resembled faces.

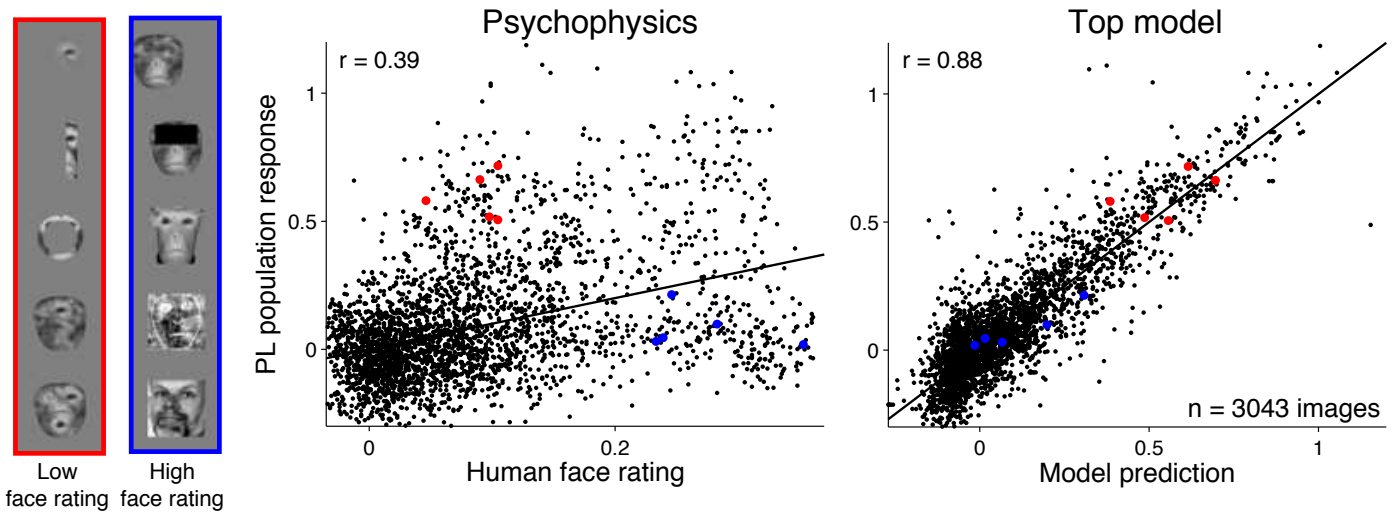

Figure 2 - Comparison of human psychophysics and a top performing image-based encoding model in predicting neural responses. Human psychophysical judgements of 'faceness' were poor predictors of the population response in PL -- many images that humans judged as highly face-like (examples in blue) elicited weak neural responses, while PL responded strongly to many images judged by humans as weakly face-like (examples in red). Our top performing models, however, correctly predicted neural responses to both types of images, demonstrating the dissociation between a semantic notion of a face and the true image operation of PL.

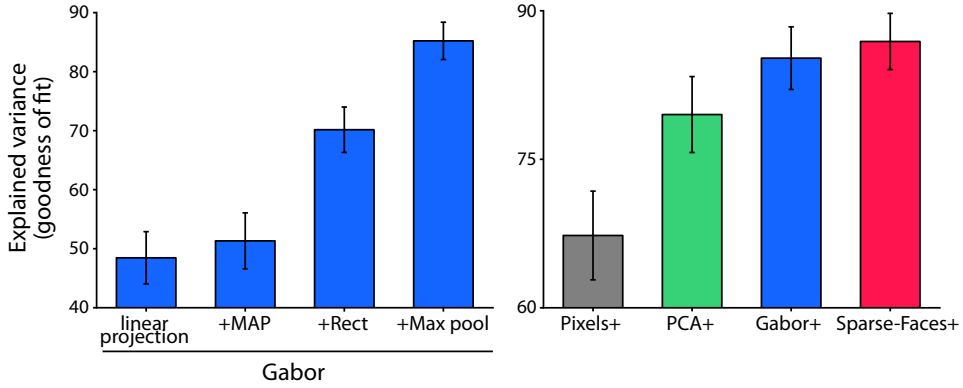

Figure 3. Effect of model properties on goodness of fit. Left) We found that specific modeling choices improved the goodness of fit for PL. Specifically, we found that a baseline simple linear Gabor model (Gabor, linear projection) could be improved upon by using the Gabor functions in a sparse encoding model and using the MAP estimates of the coefficients as regressors (+MAP). We also found that adding half-wave rectification (+Rect) and a non-linear pooling operation (max) to provide tolerance to translation and scale changes (+Max) increased goodness of fit. **Right)** We investigated various choices of basis functions or features: Pixels, PCA, Gabor functions, and a dictionary of sparse components learned on a face database (Sparse-Faces). We found that using functions localized in space and frequency (both Gabor and Sparse-Faces) improved goodness of fit.

References

Issa EB & DiCarlo JJ (2012). Precedence of the eye region in neural processing of faces. J Neurosci (in press).