

Category-selectivity together with a Normalization Model Predicts the Response to Multi-category Stimuli along the Category-selective Visual Cortex

Libi Kliger (libikl@mail.tau.ac.il)

The School of Psychological Sciences, Tel Aviv University
Tel Aviv, Israel

Galit Yovel (gality@tauex.tau.ac.il)

The School of Psychological Sciences and Sagol School of Neuroscience, Tel Aviv University
Tel Aviv, Israel

Abstract:

According to the normalization framework the neural response of a single neuron to multiple stimuli is normalized by the response of its surrounding neurons. High-level visual cortex is composed of clusters of neurons that are selective to the same category. In an fMRI study, we show that the normalization model, together with the profile of category-selectivity of a given cortical area, can predict its response to multi-category stimuli. We measured the response to a face and a body (or a face and an object) presented alone or simultaneously and estimated the contribution of each category to the multi-category representation by fitting a linear model. Results show that the response to multi-category stimuli is a weighted mean of the response to each of its components. The coefficients were correlated with the selectivity profile of the cortical region. These findings suggest that the functional organization of category-selective cortex, i.e., neighboring patches of neurons, each selective to a single category, bias the response to certain categories, for which such clusters of neurons exist, and give them priority in the representation of cluttered visual scenes.

Keywords: fMRI; Normalization model; Category-selective visual cortex; High level vision.

Introduction

It is well established that high-level visual cortex is composed of areas that are selective to different categories such as faces bodies or objects - that reside in neighboring locations. Most fMRI studies have primarily studied the response of these areas to a single stimulus drawn from a single category such as a face, a headless body or an object. However, in real life we are surrounded by complex visual stimuli, typically composed of multiple objects from multiple categories. Even the simple stimulus of a whole person is composed of two categories – a face and a body – that are typically studied separately. In the current study we therefore asked how category-selective visual cortex represents such multi-category stimuli.

The neural response to multiple stimuli has been studied previously in single unit recording studies. These studies showed that the firing rate to a preferred stimulus decreases

when the stimulus is presented together with a non-preferred stimulus (Reynolds, Chelazzi, & Desimone, 1999; Zoccolan, Cox, & DiCarlo, 2005, for a review see: Reynolds & Heeger, 2009). A normalization model was proposed to account for these results. According to the normalization model, the response of a neuron to a stimulus is normalized by the response of the surrounding neurons to this stimulus (Reynolds & Heeger, 2009). When a preferred stimulus is presented together with a non-preferred stimulus, neighboring neurons that are selective to the non-preferred stimulus normalize the response of the neuron, resulting in a lower response to the pair of stimuli relative to the response to the preferred stimulus when presented alone. This operation is described by the following equation (Eq. 1) in which the measured response of a specific neuron (i.e. neuron j), $R_j(A + B)$ equals to the response of the neuron to two stimuli, $A_j + B_j$, divided by the sum of the responses of the surrounding neurons to the two stimuli, $\sum_k A_k + \sum_k B_k$.

$$\text{Eq. 1: } R_j(A + B) = \gamma \frac{A_j + B_j}{\sigma + \sum_k A_k + \sum_k B_k}$$

In neuroimaging studies, we cannot estimate the specific parameters of the normalization equation, thus, previous studies have examined the relationship between the response to multiple stimuli and the response to each of its components (Baecck, Wagemans, & de Beeck, 2013; Baldassano, Beck, & Fei-Fei, 2016; Kaiser, Strnad, Seidl, Kastner, & Peelen, 2014; Macevoy & Epstein, 2009; MacEvoy & Epstein, 2011; Reddy, Kanwisher, & Vanrullen, 2009; Song, Luo, Li, Xu, & Liu, 2013). These studies reported that the response to multiple categories is either the mean or a deviation from the mean of the response to their component stimuli. However, as specified below, the normalization model enables us to further make specific predictions about the relative contribution of each of the components to the representation of the multi-category stimulus based on the profile of category selectivity of a given cortical region.

The goal of the current study was to examine the representation of multi-category stimuli in category-selective



cortex. In particular, we estimated the contribution of each of the stimulus components to the representation of the multi-category stimulus. We then tested whether results were consistent with the predictions of the normalization model as specified below.

Predictions

The representation of multi-category stimuli was tested for a face and a body or a face and an object in cortical areas selective to faces, bodies or objects. Here we specify the predictions for the response to a face and a body based on the normalization equation in areas that are either selective to a face or a body or both (Figure 1A-E). Similar predictions apply also for the representation of a face and an object: A face-selective area is formed by a cluster of face-selective neurons. Thus, a face-selective neuron in this area is surrounded by neurons that are also face-selective. Based on the normalization equation we can predict that the response to the face and a body will be dominated by the response to the face- (Figure 1C). Similarly, the response of a body-selective neuron within a body-selective area to a face and a body would be dominated by the response to the body (Figure 1D). Recently, Bao & Tsao (2018) found support for these predictions in a single unit recording study in the macaque face- and body-selective neurons. In addition, Face- and body-selective areas usually reside in neighboring locations and the border between them contains two populations of neighboring neurons that are selective to either a face or a body. In an area that has similar proportion of neurons that are selective to faces and bodies, the face and the body will contribute equally to the representation of the face and body. Thus, the response to the face and body when presented together will be the mean of the responses to each of them when presented alone (Figure 1E).

More generally, we predict that the response to a face and body presented together will be a weighted mean of the response to each of its categories, a face and a body, when presented alone (Eq. 2). The coefficients are determined by the proportions of the surrounding neurons that are selective to each of its categories (Eq. 3 & Eq. 4) and therefore vary along the category-selective cortex. Based on these equations, we can predict that the difference between the coefficients will be determined by the relative selectivity to the face and to the body in a given cortical area. Additionally, we can predict that the sum of the beta coefficients will be approximately 1 ($\beta_F + \beta_B \approx 1$). Alternatively, if no normalization occurs, we expect that the response to the face and the body will be the sum of the responses to each of them when presented alone ($\beta_F \approx 1$; $\beta_B \approx 1$).

$$\text{Eq. 2: } R_j(F\&B) = \beta_F \cdot R_j(F) + \beta_B \cdot R_j(B)$$

$$\text{Eq. 3: } \beta_F = \frac{\sigma + \sum_k F_k}{\sigma + \sum_k F_k + \sum_k B_k}$$

$$\text{Eq. 4: } \beta_B = \frac{\sigma + \sum_k B_k}{\sigma + \sum_k F_k + \sum_k B_k}$$

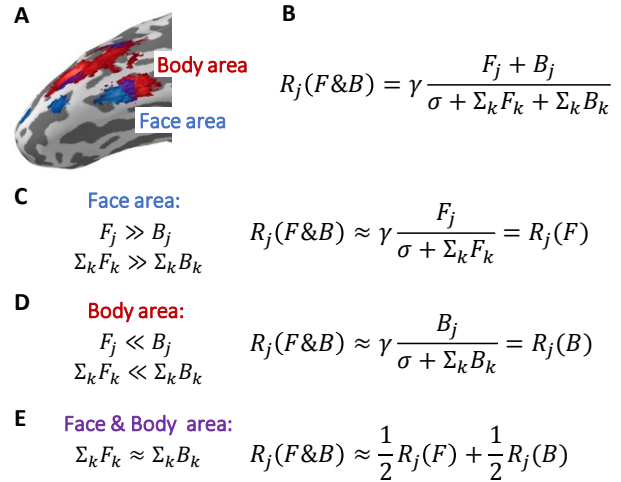


Figure 1: (A) Face- and body-selective areas in the occipito-ventral and lateral areas (blue: Face area; red: Body area; purple: Areas that are selective to both Face and Body, typically found in the border between face and body areas). (B) The response to Face+Body based on the normalization equation. $R_j(F + B)$: measured neuronal response to a Face & Body; F_j : neuron response to a face; B_j : neuron response to the body; $\sum_k F_k$: response of the surrounding neurons to the face; $\sum_k B_k$: response of the surrounding neurons to the body; γ, σ : neuron-specific model parameters. (C) The response to a Face +Body in a face-selective area. (D) The response to a Face+Body in a body-selective area. (E) The response to a Face+Body in an area that is selective to both faces and bodies

To summarize, based on the normalization model we predicted that the response to multiple-category stimuli will be approximately a weighted mean of the response to each of its categories. Importantly, we predict that the contribution of each of the categories to the response of the multi-category stimuli will be determined by the selectivity to these categories in any given cortical area.

Methods

16 healthy subjects participated in an fMRI experiment. The experiment included 3 different runs: Face-Body runs, Face-Object runs and a functional localizer (Figure 2A-C).

A searchlight analysis was performed within the category-selective cortices of each subject: A moving mask of a sphere of 27 voxels was applied. For each sphere, two separate linear models were fitted: a linear model that predicts the response to Face+Body based on the responses to the face and the body when presented alone (Eq. 5); and a linear model that predicts the response to Face+Object based on the response to the face and the object when presented alone (Eq. 6).

$$\text{Eq. 5: } \text{Face} + \text{Body} = \beta_F^{(FB)} \cdot \text{Face} + \beta_B^{(FB)} \cdot \text{Body} + \epsilon^{(FB)}$$

$$\text{Eq. 6: } \text{Face} + \text{Object} = \beta_F^{(FO)} \cdot \text{Face} + \beta_O^{(FO)} \cdot \text{Body} + \epsilon^{(FO)}$$

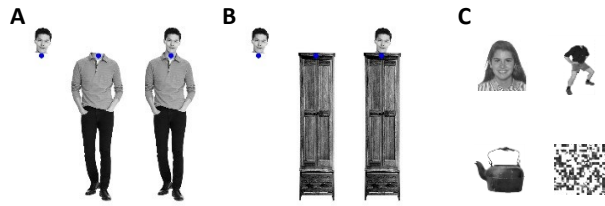


Figure 2: (A) Face-Body stimuli set: faces, bodies, and face+body stimuli, taken from the same images. Subjects were instructed to fixate on the blue dot and indicate whenever two identical images were presented in a row (1-back task). These data were used to estimate the contribution of the face and the body to the face+body representation. (B) Face-Object stimuli set: faces, objects, and faces+objects, all taken from the same images. We used wardrobes as the objects, which were matched to the body stimuli in terms of low-level visual properties. Subjects were instructed to fixate on the blue dot and performed a 1-back task. These data were used to estimate the contribution of the face and the object to the face+object representation. (C) Functional localizer stimuli set: faces, bodies, objects and scrambled objects. Subjects performed a 1-back task. Functional localizer data were used to define category-selective regions of interest and to evaluate the selectivity to specific categories, independently from the data that were used to estimate the contribution of each part to the multi-category representation.

To define the relative contribution of each of the categories to the multi-category stimuli, for each model and each sphere we calculated the difference between the beta-coefficients. In addition, for each voxel of each subject we evaluated the relative selectivity to faces compared to bodies and to faces compared to objects based on the functional localizer data. We then performed a Pearson correlation between the difference in beta-coefficients and the relative selectivity across voxels of the category-selective cortices.

Results

Figure 3A depicts the beta coefficients for the face and the body, i.e. the contribution of the face and the body to the face+body response of all spheres within the category-selective cortices of all subjects. The coefficients are scattered along the weighted mean line, indicating a sum of coefficients that is not significantly different from 1 [mean sum=1.015, $t(14)=0.978$, $p=0.438$]. The color of each dot indicates the selectivity to the face relative to the body, as measured by the independent functional localizer. The difference in the contribution of the face and the body to the face+body representation, (i.e. the difference between the beta coefficients) is correlated with the selectivity to the face relative to the body as predicted [mean $r=0.405$, $t(14)=8.249$, $p<0.001$].

In order to generalize our results to different categories, we performed similar analysis for a pair of a face and an object. Figure 3B depict the contribution of the face and the object

to the face+object model of all spheres within the category selective cortices of all subjects, with the color of each dot indicating the relative selectivity to the face compared to the object. Similarly to the face+body results, the beta-coefficients are scattered along the weighted mean line with a sum of coefficients that is not significantly different from 1 [mean sum=1.016, $t(14)=1.564$, $p=0.140$]. Moreover, the difference in the contribution of the face and the object to the face+object representation (i.e., the difference between the coefficients) is correlated with the selectivity to the face relative to the object as expected [mean $r=0.415$, $t(14)=10.848$, $p<0.001$].

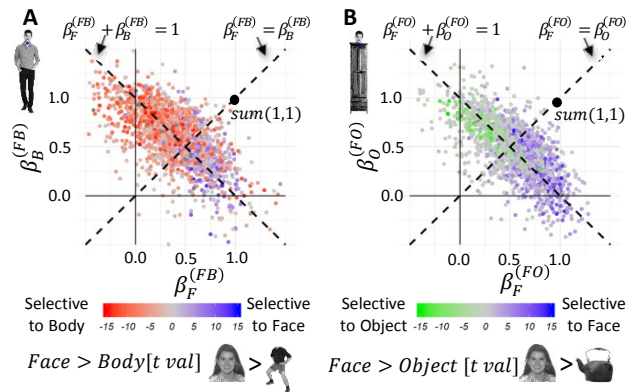


Figure 3: (A) The beta coefficients of all spheres in the category selective cortices of all subjects for the face-body model (Eq. 5). The color of each dot indicates the selectivity for the face relative to the body. The difference between the contribution of the face and the body to the face+body representation is correlated with the relative selectivity to the face and the body. (B) Same as A for the face+object model (Eq. 6). The color of each dot indicates the selectivity for the face relative to the object. The difference between the contribution of the face and the object to the face+object representation is correlated with the selectivity to the face relative to the object.

In order to compare the spatial distribution of the beta-coefficients and category selectivity, we plotted the difference between the coefficients and the difference between the selectivity to each pair of categories on brain surface maps (Figure 4A-D). Figure 4A shows the difference between the face and body coefficients (i.e., difference between the contribution of the face and the contribution of the body to the face+body representation) of one representative subject along his category selective cortex. Figure 4B shows the selectivity to the face relative to the selectivity to the body for the same subject. It can be seen that cortical areas that show higher contribution of the face to the face+body representation correspond to face-selective clusters (red in both figures), and that areas that show higher contribution of the body to the face+body representation correspond to body-selective clusters (blue in both figures). Figure 4C shows the difference between the contribution of

the face and the object to the face+object representation for the same subject. Figure 4D shows the selectivity to the face relative to the object. Similarly to the face+body results, areas that show higher contribution of the face to the face+object representation correspond to face-selective clusters (red in both figures), and areas that show higher contribution of the object to the face+object representation correspond to object-selective clusters (blue in both figures).

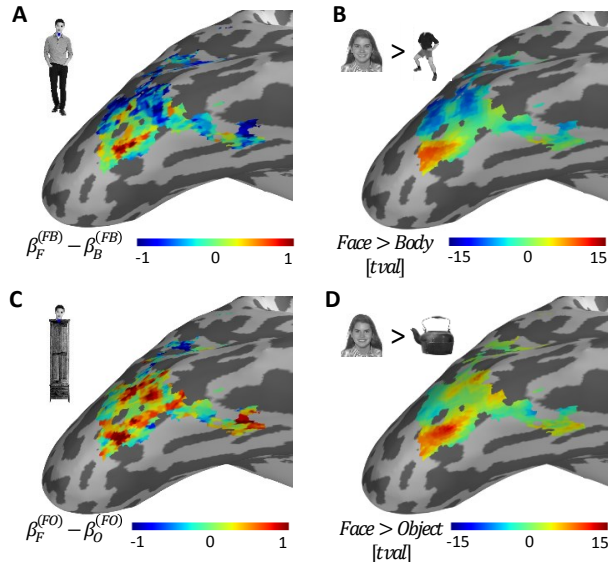


Figure 4: Results of one representative subject plotted on the cortical surface. (A) The difference between the contribution of the face and the body to the face+body representation as indicated by the difference between the regression coefficients. (B) Selectivity to faces relative to bodies (t map of Face>Body). Selectivity was determined based on independent functional localizer data. (C) The difference between the contribution of the face and the object to the face+object representation as indicated by the difference between the regression coefficients. (D) Selectivity to faces relative to objects (t map of Face<Body) based on independent functional localizer data.

Conclusions

To summarize, we found that, in line with the normalization model, the response to multi-category stimuli in category-selective cortex is a weighted mean of the response to each category. Moreover, the contribution of each category to the response of the complex stimulus in a given cortical area is determined by its profile of category selectivity.

We conclude that the functional organization of category-selective cortex, i.e., neighboring patches of neurons that each of them is selective to a single category, determines the representation of complex visual scenes. Clusters of category-selective neurons bias the response towards their preferred categories. This organization gives priority to these categories in cluttered visual scenes.

Acknowledgments

This work is supported by a grant from the Israeli Science Foundation (ISF 446/16).

References

- Baek, A., Wagemans, J., & de Beeck, H. P. (2013). The distributed representation of random and meaningful object pairs in human occipitotemporal cortex: The weighted average as a general rule. *NEUROIMAGE*, *70*, 37–47.
- Baldassano, C., Beck, D. M., & Fei-Fei, L. (2016). Human-object interactions are more than the sum of their parts. *Cerebral Cortex*, 1–13.
- Bao, P., & Tsao, D. Y. (2018). Representation of multiple objects in macaque category-selective areas. *Nature Communications*, *9*(1), 1–16.
- Kaiser, D., Strnad, L., Seidl, K. N., Kastner, S., & Peelen, M. V. (2014). Whole person-evoked fMRI activity patterns in human fusiform gyrus are accurately modeled by a linear combination of face- and body-evoked activity patterns. *Journal of Neurophysiology*.
- Macevoy, S. P., & Epstein, R. a. (2009). Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Current Biology: CB*, *19*(11), 943–7.
- MacEvoy, S. P., & Epstein, R. a. (2011). Constructing scenes from objects in human occipitotemporal cortex. *Nature Neuroscience*, *14*(10), 1323–9.
- Reddy, L., Kanwisher, N. G., & Vanrullen, R. (2009). Attention and biased competition in multi-voxel object representations. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(50), 21447–21452.
- Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive Mechanisms Subserve Attention in Macaque Areas V2 and V4. *The Journal of Neuroscience*, *19*(5), 1736–1753.
- Reynolds, J. H., & Heeger, D. J. (2009). The Normalization Model of Attention. *Neuron*, *61*(2), 168–185.
- Song, Y., Luo, Y. L. L., Li, X., Xu, M., & Liu, J. (2013). Representation of Contextually Related Multiple Objects in the Human Ventral Visual Pathway. *Journal of Cognitive Neuroscience*, *25*(8), 1261–1269.
- Zoccolan, D., Cox, D. D., & DiCarlo, J. J. (2005). Multiple object response normalization in monkey inferotemporal cortex. *The Journal of Neuroscience*, *25*(36), 8150–64.