Distinct Mechanisms of Surround Modulation for Dynamic Natural Scenes in Motion- and Scene-Selective Cortex

Merve Kiniklioglu[1*], David Pitcher[2], and Daniel Kaiser[1,3,4,5]

[1]Neural Computation Group, Department of Mathematics and Computer Science, Physics, Geography, Justus Liebig University Gießen, 35392 Gießen, Germany.

[2]Department of Psychology, University of York, Heslington, York, YO10 5DD, UK

[3]Cluster of Excellence "The Adaptive Mind", Universities of Giessen, Marburg, and Darmstadt, 35392 Gießen, Germany.

[4]Center for Mind, Brain and Behavior (CMBB), Universities of Giessen, Marburg, and Darmstadt, 35032 Marburg, Germany

[5]Center for Applied Computer Science and Data Science (ZAD), Justus Liebig University Gießen, 35392 Gießen, Germany.

[*]Corresponding author, mervekiniklioglu@gmail.com.

November 6, 2025

Abstract

Surround modulation refers to changes in neural responses to a central stimulus induced by its surrounding context, which can manifest as either suppression or facilitation. Although this phenomenon is well established for simple stimuli, its underlying neural mechanism during natural scene perception remains unclear. Using fMRI, we examined how motion congruence and categorical similarity between central and surrounding scenes shape surround modulation across the visual hierarchy. Central and surrounding scenes systematically varied in categorical similarity (identical exemplar, different exemplar of the same basic-level category, different basic-level category, or different superordinate category) and in motion direction (same or opposite direction). Neural responses were measured in primary visual cortex (V1), motion-selective cortex (hMT+), and scene-selective occipital and parahippocampal place areas (OPA and PPA). hMT+ showed robust motion-dependent surround suppression, which was stronger for same-direction motion. In contrast, V1 showed surround facilitation across all conditions, which was reduced when center and surround were identical and moved in the same direction, consistent with sensitivity to physical similarity. OPA and PPA primarily exhibited facilitation. Multivariate decoding between center-only and center-surround conditions complemented these findings, revealing motion-dependent surround modulation in hMT+, and category-dependent surround modulation in OPA and PPA. Across the visual hierarchy, surrounding scene context thus systematically attenuates redundant input or enhances informative differences: from low-level facilitation in V1 to motion-dependent suppression in hMT+ and category-dependent modulation in scene-selective regions.

Keywords: scene perception, motion perception, surround suppression, sur-

round modulation

Distinct Mechanisms of Surround Modulation for Dynamic Natural Scenes in Motion- and Scene-Selective Cortex

Introduction

Sensitivity to visual stimuli is strongly influenced by the surrounding context. A well-known example is surround modulation, in which the response of a neuron to a stimulus presented in the receptive field (RF) center is modulated by stimuli presented in the surround. This modulation can manifest as suppression (*surround suppression*) or facilitation (*surround facilitation*) (Cavanaugh et al., 2002a, 2002b; DeAngelis et al., 1994; Ichida et al., 2007; Petrov et al., 2005; Pihlaja et al., 2008; Shushruth et al., 2012; Walker et al., 1999; Williams et al., 2003; Xing & Heeger, 2001; Zenger-Landolt & Heeger, 2003). Such modulatory effects are generally attributed to antagonistic center–surround interactions within the visual cortex, particularly in motion-selective regions such as the middle temporal area (MT) (e.g., MT hypothesis; see Er et al., 2020; Kinriklioglu & Boyaci, 2025; Pack et al., 2005; Schallmo et al., 2018; Tadin et al., 2003, 2011; Turkozer et al., 2016).

These center–surround mechanisms are observed throughout the visual hierarchy. Surround modulation has been classically demonstrated in retinotopic areas of the early visual cortex (V1-V4) and in motion-selective regions such as MT and MST, suggesting that it reflects a canonical computation of contextual modulation within the visual system (Allman et al., 1985; Angelucci et al., 2017; Hallum & Movshon, 2014; Nurminen & Angelucci, 2014; Sundberg et al., 2009). More recently, surround modulation has also been observed in higher-level regions of the ventral visual stream, such as the lateral occipital complex (LOC), highlighting its prevalence across the visual hierarchy (Montoya et al., 2025).

The strength and form of surround modulation depend on stimulus properties. Both neural (Kiniklioglu & Boyaci, 2025; Schallmo et al., 2018; Shushruth et al., 2013; Turkozer et al., 2016; Zenger-Landolt & Heeger, 2003) and behavioral studies (Er et al., 2020; Kiniklioglu & Boyaci, 2022; Schallmo et al., 2018; Tadin et al., 2011; Turkozer et al., 2016) show that low-level features such as size, contrast, and motion direction systematically influence the strength of surround suppression. Typically, feature similarity between center and surround increases suppression, whereas dissimilarity reduces it and can even reverse the effect into facilitation. For example, neurons responding to a central grating are strongly suppressed by iso-oriented surrounds but can be facilitated when the surround is orthogonal (Cavanaugh et al., 2002b; Schallmo et al., 2016; Serrano-Pedraza et al., 2012). A comparable pattern is observed for motion, where suppression is typically stronger when the center and surround drift in the same direction than when they move in opposite directions (Allman et al., 1985; Born & Tootell, 1992; Cavanaugh et al., 2002b; Kiniklioglu & Boyaci, 2025; Lamme, 1995; Paffen, van der Smagt, et al., 2005).

Although these effects are well established with simple stimuli such as static or drifting gratings, it remains unclear whether surround modulation also occurs in the context of complex, dynamic scenes. Our recent behavioral work demonstrated that surround modulation is indeed observed with dynamic natural scenes, with suppression increasing as the categorical similarity between the center and surround scenes increases (Kiniklioglu & Kaiser, 2025). These results align with scene categorization studies showing that categorically congruent surrounds facilitates recognition, whereas incongruent or physically mismatched surrounds impair it (Faurite et al., 2024; Peyrin et al., 2021). Consistent with low-level studies (Kiniklioglu & Boyaci, 2022; Paffen, Alais, & Verstraten, 2005; Paffen, van der Smagt, et al., 2005), we also found that suppression decreased when the center and surround moved in opposite directions. Together, this evidence suggests that surround modulation in natural vision can be shaped by both low-level visual features, such as motion direction, and higher-level contextual information, such as categorical similarity.

While our behavioral results suggest a suppressive effect that scales with the dissimilarity between center and surround content, neurophysiological work has disagreed on whether suppression or facilitation prevails. In the macaque primary visual cortex (V1), suppression was stronger when the center and surround were contained homogeneous natural images (Coen-Cagli et al., 2015), whereas in the cat visual cortex, surrounds that perceptually completed a central pattern elicited facilitation (Onat et al., 2013). However, no human studies have investigated the neural mechanisms underlying the suppressive effects observed in behavioral studies with dynamic natural scenes.

To address this gap, we used functional magnetic resonance imaging (fMRI) to investigate the neural correlates of surround modulation in dynamic natural scenes. We examined responses across regions of interest spanning multiple levels of the visual hierarchy, including primary visual cortex (V1), motion-selective cortex (hMT+), and the scene-selective occipital place area (OPA) and parahippocampal place area (PPA). Participants viewed central scenes presented together with surrounding scenes that varied in categorical similarity across four levels: identical exemplar, different exemplar from the same basic-level category, different basic-level categories within the same superordinate category, and different superordinate categories. This manipulation enabled us to test how categorical similarity shapes neural signatures of surround modulation across the visual hierarchy. We also varied motion congruence, comparing conditions in which the center and surround drifted in the same versus opposite directions, to assess whether motion-related effects observed with simple stimuli generalize to naturalistic stimuli. By independently manipulating categorical similarity and motion congruence, we could test how both high- and low-level contextual factors influence surround modulation in dynamic natural scenes.

Methods

Participants

Twenty healthy volunteers (11 female, mean age = 26.7 years) participated in the study. One participant was excluded due to excessive head motion during scanning (see Preprocessing for motion-quality criteria), leaving a final sample of nineteen participants. The sample size (N = 19) is comparable to that used in previous studies investigating low-level surround modulation (Er et al., 2020; Kiniklioglu & Boyaci, 2025; Schallmo et al., 2016). All reported normal or corrected-to-normal vision. Written informed consent was obtained before participation, and participants received monetary compensation. The study was approved by the Ethics Committee of Justus Liebig University Giessen. All experimental protocols were in accordance with the Declaration of Helsinki.

Stimuli

Stimuli were panoramic videos created by moving static scene images behind a circular occluder (908 × 699 pixels; 5 pixels per frame). The central image was presented through a 1.9°-diameter aperture, surrounded by an annulus ranging from 2.5° to 10.4°. To separate the center and surround, the area between them (1.9°-2.5°) remained unstimulated. A cosine envelope was applied at the aperture boundaries to minimize sharp transitions at the edges.

Scene images were selected from two superordinate categories (indoor: restaurants, museums; outdoor: parks, residential areas), each featuring two basic-level categories, with two exemplars per basic-level category. The center scenes were always presented with a surrounding scene, except in center-only trials where the central stimulus appeared alone (see below). The resulting center–surround stimuli depicted natural scenes that varied in their categorical relationship between center and surround, with four conditions: identical exemplar condition, where the center and surround images were identical; different exemplar condition, where the center and surround images belonged to the same basic-level category (e.g., museums) but were not identical; different basic-level category condition, where the center and surround images belonged to the same superordinate category (e.g., indoor) but were from different basic-level categories; and different superordinate category condition, where the center and surround images belonged to different superordinate categories (see Figure 1). Motion congruency was manipulated with two conditions: in half of the trials, the center and surround scenes moved in the same direction, and in the other half, they moved in opposite directions.

Experimental Paradigm

We used a mixed event-related design in which center-only and center+surround trials were presented within the same run (Figure 1). In total, nine trial types (four category conditions × two motion directions, plus the center-only condition) were pre-

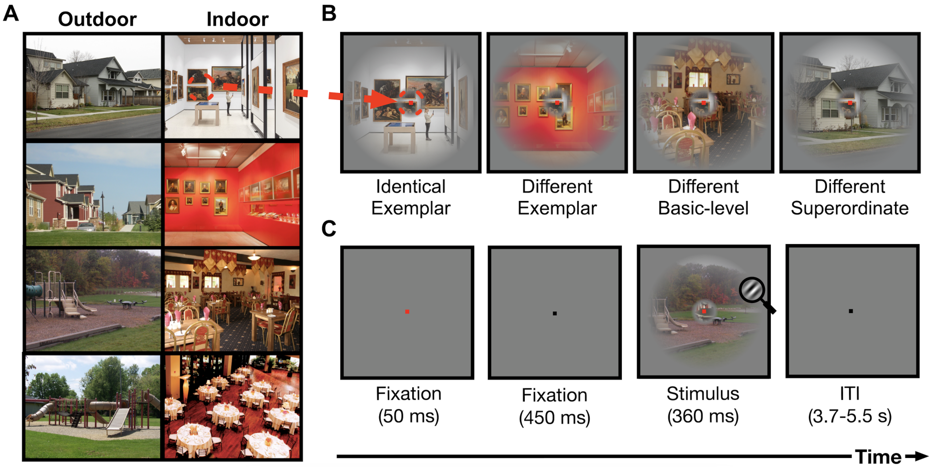

Figure 1: Stimuli and Paradigm. A) The natural scene images used to create the stimuli were drawn from two superordinate categories (indoor and outdoor) and two basic-level categories within each (restaurants, museums, parks, and residential areas), with two exemplars per basic-level category. B) Example stimuli for each category condition, shown from left to right: identical exemplar, different exemplar, different basic-level category, different superordinate category. C) Experimental design. Each trial began with a fixation point (500 ms), which briefly turned red to signal the upcoming stimulus, followed by presentation of a scene video (360 ms). Intertrial intervals varied randomly between 3.7 and 5.6 s. On 10% of trials, a high-contrast Gabor patch appeared at a random location within either the center or surround region, and participants reported its detection with a keypress.

sented in randomized order. Center-only trials served as a baseline for quantifying the suppressive effect of adding the surround.

Each trial began with a fixation point (500 ms), followed by a scene video (360 ms) on a mid-gray background. Participants were instructed to maintain fixation at the center of the display. Intertrial intervals varied randomly between 3.7 and 5.6 seconds, sampled from a uniform distribution. To ensure that participants attended to both center and surround, on 10% of trials a small Gabor patch (spatial frequency = 1 cycles/°, diameter = 0.7°, contrast = 98%) appeared at a random location within either the center or surround region. Participants reported its detection with a keypress, with a mean accuracy of 92.2% (SD = 7.6%). Each condition was presented eight times per run, and participants completed five runs. Each run lasted approximately 10 minutes, and short breaks were provided between runs.

Data Acquisition

MRI data were acquired on a 3T Siemens Magnetom Prisma scanner (Siemens Healthineers, Erlangen, Germany) equipped with a 64-channel head coil at the Bender Institute of Neuroimaging (BION), Justus Liebig University Giessen. High-resolution anatomical images were collected using a T1-weighted 3D sagittal MP-RAGE sequence (voxel size = 1 mm^3 isotropic, 176 slices). Functional images were acquired with a T2*-weighted EPI sequence (TR = 1850 ms, TE = 30 ms, voxel size = 2.2 mm^3, 58 slices). Visual stimuli were presented on a MR-compatible LED monitor (1920 × 1080 pixels, 120 Hz) positioned at the rear of the bore and viewed via a mirror mounted on the head coil at a distance of 140 cm. Stimuli were generated and presented using MATLAB (MathWorks, Natick, MA) and the Psychophysics Toolbox (Brainard, 1997). Participant responses were recorded using an MR-compatible fiber-optic response box. Each session began with an anatomical scan, followed by four localizer runs and five experimental runs, for a total duration of approximately 90 minutes.

Localizer Runs

hMT+: The hMT+ area of each participant was defined using established methods (Huk et al., 2002). Stimuli consisted of 200 randomly positioned white dots presented within a 12° aperture on a black background. In dynamic blocks, the dots moved along one of four trajectories: radial (expansion–contraction), angular (clockwise–counterclockwise rotation), horizontal (left–right), or vertical (up–down). Motion direction changed every 1.85 s to prevent adaptation. In static blocks, the dots were identical but remained fixed in position. Each block lasted 14.8 s, and dynamic and static blocks alternated eight times per run. Participants maintained central fixation and performed a color-change detection task in which they reported changes in the color of the fixation point.

V1: The V1 area of each participant was defined using a flickering checkerboard-patterned wedge paradigm, similar to established retinotopic mapping methods (Engel et al., 1997; Sereno et al., 1995). Because we were specifically interested in the V1/V2 boundary, we used alternating horizontal and vertical wedges instead of rotating or expanding stimuli (Greenberg et al., 2012; Slotnick & Yantis, 2003). The run consisted of 14.8-s blocks of horizontal and vertical wedges presented in alternation, repeated ten times in the run. As in the hMT+ localizer, participants maintained fixation and performed a color-change detection task.

OPA and PPA: Scene-selective areas were defined using 3-s movie clips from four categories: faces, scenes, objects, and scrambled objects (Küçük et al., 2024; Pitcher et al., 2011). Each category included 60 clips. Face clips featured seven children filmed against a black background in close-up, showing only their faces as they danced or interacted with toys or with adults who remained out of frame. Scene clips were recorded in 15 different locations, primarily pastoral settings, and filmed from a slowly moving car. Object clips depicted 15 distinct inanimate items (e.g., mobiles, wind-up toys, toy planes, tractors, and rolling balls). Scrambled-object clips were generated by dividing each object movie into a 15 × 15 grid and spatially shuffling the resulting segments within each frame. Participants completed eight blocks per category, each consisting of six videos randomly sampled from the full set of 60 exemplars in that category. Although the same actor, scene, or object could appear more than once, the large stimulus pool made such repetitions unlikely. As in the previous localizer runs, participants maintained fixation and reported changes in the color of the fixation point.

hMT+ and V1 Sub-ROI Localizer Run

Using an independent localizer run, we identified the voxels corresponding to the spatial location and size of the center stimuli as sub-ROIs within hMT+ and V1. This approach is commonly used to localize responses to low-level grating stimuli (Er et al., 2020; Kinniklioglu & Boyaci, 2025; Schallmo et al., 2018), but here we adapted it to ensure that the same spatially defined sub-ROIs could be applied in the analysis of naturalistic stimuli. In the localizer, participants viewed drifting high-contrast (98%) center and surround (i.e., annulus) gratings, matched in size and location to those used in the functional runs. The run consist of 14.8-s center blocks alternating with 14.8-s surround blocks, each separated by a 14.8-s blank period, repeated six times. As in the other localizers, participants maintained fixation and reported changes in the color of the fixation point.

Data Analysis

Preprocessing

Localizer Runs: Localizer data were preprocessed and analyzed using the FMRIB Software Library (FSL) (www.fmrib.ox.ac.uk/fsl) and Freesurfer (Dale et al., 1999; Fischl et al., 1999; Woolrich et al., 2001). High-resolution anatomical images were skull-stripped with BET. Preprocessing steps for functional images included motion correction with MCFLIRT, high-pass temporal filtering (100s), and BET brain extraction. Each participant's functional images were aligned to their own high-resolution anatomical image and registered to the standard Montreal Neurological Institute (MNI) 2-mm brain using FLIRT. The 3D cortical surface was constructed from anatomical images for each participant using FreeSurfer's *recon-all* command for visualizing statistical maps, anatomical delineation, and identifying ROIs.

Experimental Runs: Experimental data were preprocessed and analyzed in SPM12 (www.fil.ion.ucl.ac.uk/spm/). Preprocessing steps included geometric distortion correction with the SPM FieldMap toolbox, motion correction and coregistration of functional volumes to each participant's T1-weighted structural image. Structural images were segmented and normalized to MNI 2-mm standard space. After also normalizing the functional images, they were spatially smoothed with a 6 mm FWHM Gaussian kernel.

To assess data quality, head motion was quantified for each volume using framewise displacement (FD). Volumes with FD greater than 0.5 mm were flagged as motion outliers (Power et al., 2012). Runs with more than 30% flagged volumes were discarded (Parkes et al., 2018). Participants with two or more runs discarded were excluded (Ciric et al., 2017). One participant met this exclusion criterion and was removed from all analyses; all results reported below are based on the remaining $N = 19$ participants.

ROI Construction

For all ROI constructions, a general linear model (GLM) was applied using FSL's FMRI Expert Analysis Tool (FEAT). The predicted fMRI response in each trial was computed assuming a double-gamma hemodynamic response function (HRF). Nuisance regressors for linear motion (derived from MCFLIRT) were also included in the model. For removing temporal autocorrelations, FILM prewhitening was applied (Woolrich et al., 2001).

hMT+: For the hMT+ ROI, the statistical parametric maps (SPMs) of the dynamic versus static contrast were registered to Freesurfer and overlaid on the surface in the native space using the *tksurfer* command. Utilizing the MT label from FreeSurfer's anatomical delineation for guidance, voxels at the ascending tip of the inferior temporal sulcus and responding more to dynamic compared to static dots were identified as hMT+ and used as a mask for the hMT+ sub-ROI localization.

V1: For the V1 ROI, SPMs of horizontal versus vertical contrast were registered to Freesurfer and overlaid on the surface in the native space using the *tksurfer* program. Utilizing the V1 label from FreeSurfer's anatomical delineation for guidance and voxels that respond stronger to vertical than horizontal wedges, we drew the V1-V2 boundaries. The voxels that fell in or around the calcarine sulcus were identified as V1 and used as a mask for the V1 sub-ROI localization.

hMT+ and V1 sub-ROIs: For V1 and hMT+ sub-ROIs, we analyzed the independent localizer run data and identified the voxels that respond more strongly to the center compared to the surround grating within V1 and hMT+. The activated regions were identified as V1 and hMT+ sub-ROIs, respectively, and subsequently used as masks for the analysis of the experimental runs.

OPA and PPA: For the scene-selective ROIs, we identified the voxels with the greatest scene versus face and object contrast, using a voxelwise corrected threshold of $p < 0.05$, and a cluster extent threshold of 10 voxels. The occipital place area (OPA) was localized as the scene-selective cluster on the lateral surface near the transverse occipital sulcus, and the parahippocampal place area (PPA) as the scene-selective cluster on the ventral visual cortex. The identified regions were used as OPA and PPA masks. Because the center vs. surround contrast did not yield reliable voxel-level responses in OPA and PPA, we used the full ROIs rather than sub-ROIs for the analysis of the experimental runs.

Analysis of Experimental Runs

Univariate Analysis

The experimental runs were analyzed in SPM12 using a general linear model (GLM). For each run, separate regressors were specified for each of the nine conditions, and six motion parameters from realignment were included as nuisance regressors. Predicted BOLD responses were modeled by convolving event onsets with a double-gamma hemodynamic response function (HRF), and statistical parametric maps (SPMs) were generated to assess the main effect of each condition across runs. To quantify fMRI responses, parameter estimates (beta weights) for each condition were extracted from the predefined ROIs (hMT+, V1, OPA, and PPA) using custom code with SPM12 functions. For these ROI-based univariate analyses we used the unsmoothed preprocessed data to preserve fine-grained voxel-level signals. The beta weights were treated as fMRI responses and used in further statistical analyses.

To quantify changes in fMRI response due to the presence of the surround , we calculated a Suppression Index (SI), defined as:

$$SI = B_C - B_{C+S}; \quad C = center, \quad C + S = center + surround, \tag{1}$$

where $B_.$ denotes the fMRI response for the given condition. Negative SI values indicate surround facilitation, positive values indicate surround suppression, and an SI of 0 reflects no suppression or facilitation.

Surround modulation was assessed by testing SI values against zero using one-sample, two-tailed Student's t-tests with correction for multiple comparisons. We then performed a two-way Analysis of Variance (ANOVA) on SI values with categorical similarity (four levels: identical exemplar, different exemplar, different basic-level category, different superordinate category) and motion-direction congruence (two levels: same-direction, opposite-direction) as factors for each sub-ROI using SPSS Version 25 (IBM Corp., Armonk, NY). Finally, post hoc paired-sample t-tests were used to assess how categorical similarity affects surround modulation under each motion condition.

Multivariate Analysis

To investigate distributed patterns of activity, we performed ROI-based decoding using CoSMoMVPA (Oosterhof et al., 2016) on condition-wise GLM parameter estimates. For each participant and ROI (hMT+, V1, OPA, and PPA), beta images from each run and condition were assembled into voxel pattern vectors. We implemented two-way classification, contrasting *center-only* with *center+surround* trials across different category and motion-direction conditions, using linear discriminant analysis (LDA). Datasets were z-scored within the training folds and class-balanced within chunks (runs). Cross-validation used a leave-one-run-out partitioning scheme, and the mean accuracy across folds was taken as the decoding score.

This two-way decoding quantifies the separability of multivoxel response patterns between center-only and center+surround conditions. Accuracy above chance indicates that the presence of a surround reliably alters the distributed activity pattern across voxels relative to center-only. Higher decoding accuracy reflects a greater surround-induced change in these patterns, either due to shifts in overall response magnitude and/or reconfiguration of voxelwise patterns, and thereby indexes a stronger influence of the surround on neural representations. We thus interpret decoding accuracy as an indicator of the magnitude of surround modulation, while it cannot indicate the direction of the modulation (i.e., suppression vs facilitation).

Group-level significance was assessed by comparing subject-level accuracies against chance level (0.5) using one-sample t-tests (one-tailed), with false discovery rate (FDR) correction applied across conditions. For all ROIs (hMT+, V1, OPA, and PPA), we performed separate decoding analyses for all combinations of scene categories and for both motion directions. For assessing the effects of categorical similarity, data were merged across motion directions to increase statistical power, and decoding accuracies were compared across the four category conditions using a one-way repeated-measures ANOVA. For assessing the effects of motion congruence, data were merged across categories, and directional sensitivity was assessed by directly comparing decoding accuracy between same- and opposite-direction conditions using paired-samples t-test.

Results

Univariate analyses reveal distinct patterns of surround suppression and facilitation across visual regions

In hMT+, we observed surround suppression for conditions in which the center and surround moved in the same direction. Figure 2A shows the Suppression Index (SI; see Methods) for the hMT+ sub-ROI across four categorical similarity and two motion-direction conditions. SI values were significantly greater than zero only in the identical-exemplar condition for same-direction trials ($t(18)= 2.77$, $p_{\mathrm{FDR}} = 0.04$). A two-way ANOVA revealed a significant main effect of motion direction ($F(1,18)= 13.10$, $p = 0.002$, $\eta^2 = 0.42$), but no significant main effect of category ($F(3,54)= 1.17$, $p = 0.33$, $\eta^2 = 0.06$), or interaction between category and direction ($F(3,54)= 0.29$, $p = 0.83$, $\eta^2 = 0.02$). These results suggest that surround suppression in hMT+ depends

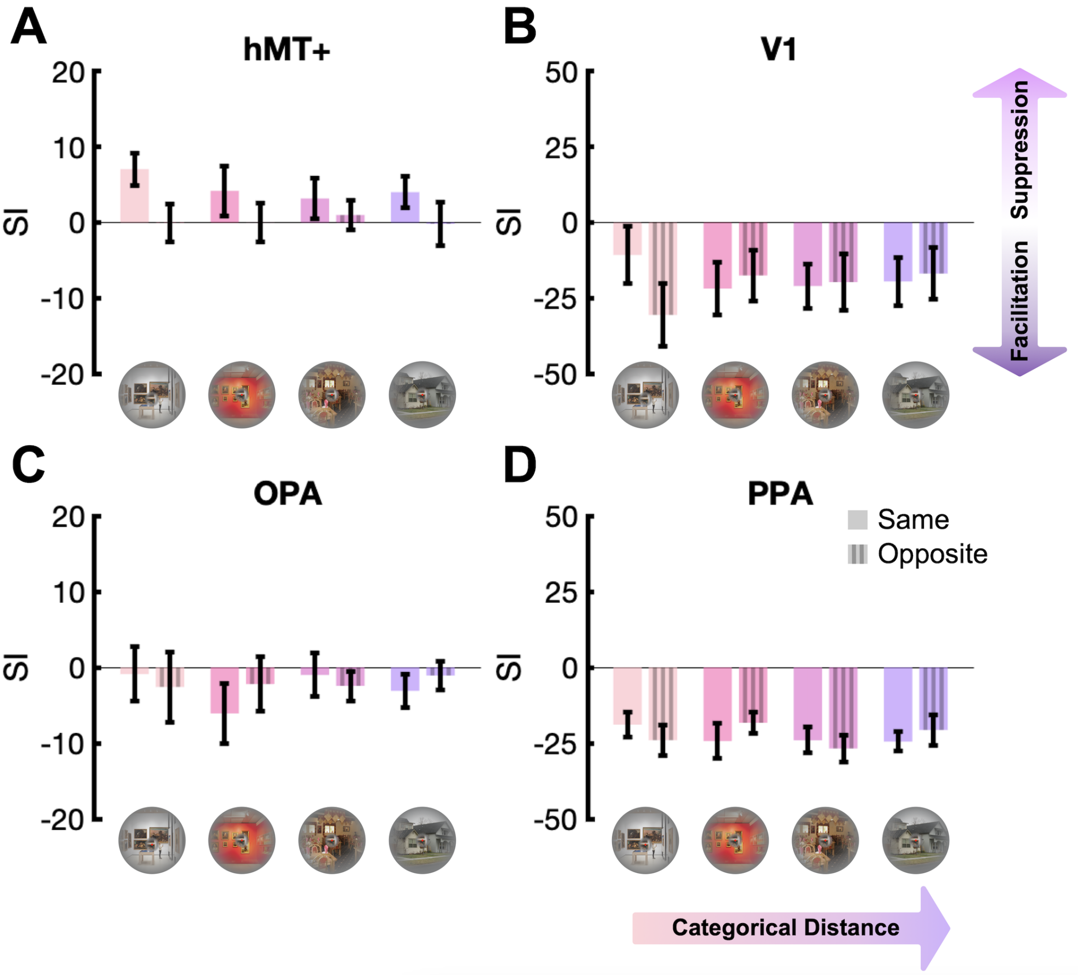

Figure 2: Univariate surround modulation effects. Suppression index (SI) values are shown across four category conditions: identical exemplar, different exemplar, different basic-level category, and different superordinate category, and two motion-direction conditions (same and opposite). Example stimuli for each category condition are displayed along the x-axis to illustrate the categorical hierarchy, with categorical distance increasing from left to right (from the identical exemplar to the different superordinate category conditions). Results are displayed for **A)** hMT+, **B)** V1, **C)** OPA, and **D)** PPA. Positive SI values indicate surround suppression; negative values reflect facilitation. Error bars represent the standard error of the mean (SEM).

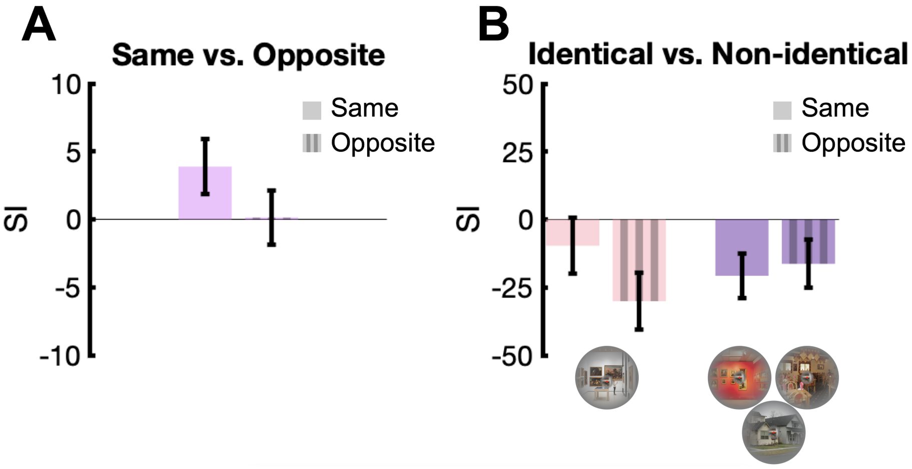

Figure 3: Averaged univariate surround modulation effects in hMT+ and V1. **A)** hMT+ SI values averaged across the four category conditions for same- and opposite-direction trials. **B)** V1 SI values for the identical exemplar condition versus the mean of the three other (non-identical) conditions (i.e., different exemplar, different basic-level category, and different superordinate category conditions), shown separately for same- and opposite-direction trials. Positive SI values indicate surround suppression; negative values reflect facilitation. Error bars represent the standard error of the mean (SEM).

primarily on the relative motion direction between center and surround.

As expected, the categorical similarity of the center and surround had no effect on the suppression strength, consistent with hMT+ being primarily motion selective rather than sensitive to scene content. Because the categorical similarity neither modulated SI nor interacted with motion direction, SI values were averaged across the four category conditions (Figure 3A). A paired-sample t-test on the averaged data confirmed that suppression was significantly stronger when the center and surround moved in the same direction than when they moved in opposite directions ($t(18) = 3.62$, $p = 0.002$). Together, these results demonstrate that hMT+ exhibits motion-dependent surround suppression under dynamic, natural scene conditions, consistent with perceptual suppression effects in our recent behavioral study using the same stimuli (Kiniklioglu & Kaiser, 2025), and with classical findings using simple gratings (Kiniklioglu & Boyaci, 2022; Paffen, Alais, & Verstraten, 2005; Paffen, van der Smagt, et al., 2005).

In V1, we observed surround facilitation across all conditions, in stark contrast to the suppression effects found in hMT+. Figure 2B shows the SI values for the V1 sub-ROI across four scene-category conditions and two motion-direction conditions. SI values were significantly lower than zero in all conditions except the identical-exemplar condition for same-direction trials ($t(18) = 0.95$, $p_{\mathrm{FDR}} = 0.18$). A two-way ANOVA revealed a significant interaction between motion direction and categorical similarity ($F(3{,}54) = 9.34$, $p < 0.001$, $\eta^2 = 0.34$), but no significant main effects of motion direction ($F(1{,}18) = 0.44$, $p = 0.51$, $\eta^2 = 0.02$) or category ($F(3{,}54) = 0.30$, $p = 0.83$, $\eta^2 = 0.02$).

Because the three non-identical category conditions (different exemplar, different basic-level category, and different superordinate category) did not differ from each other and showed no effect of motion direction (all ps$>$0.20), they were collapsed into a single “non-identical” condition, analyzed separately for same- and opposite-direction trials (Figure 3B). Post-hoc tests confirmed that the category × motion direction interaction observed in the ANOVA was driven by the identical-exemplar condition. Motion direction significantly affected responses only in the identical-exemplar condition, with reduced facilitation when the center and surround moved in the same direction compared to opposite directions ($t(18) = 3.47$, $p_{\mathrm{FDR}} = 0.003$). Moreover, for same-direction trials, facilitation was significantly weaker for identical exemplars than for the collapsed non-identical condition ($t(18) = 2.78$, $p_{\mathrm{FDR}} = 0.01$). This pattern suggests that similarity between center and surround modulates facilitation in V1, such that identical center–surround combinations yield weaker facilitation than non-identical ones, even within the same category. This result aligns with classical low-level findings showing the strongest suppression for iso-oriented uniform stimuli (Cavanaugh et al., 2002a; DeAngelis et al., 1994; Flevaris & Murray, 2015; Schallmo et al., 2016; Serrano-Pedraza et al., 2012; Walker et al., 1999) and is consistent with motion-dependent surround modulation observed in hMT+, though in V1 the effect manifested as reduced facilitation rather than suppression.

Scene-selective areas OPA and PPA both showed surround facilitation. Figures 2C and D show the SI values for these ROIs across four scene-category conditions and two motion-direction conditions. In OPA, SI values trended toward facilitation but did not reach significance (all $p_{\mathrm{FDR}} > 0.05$), except the different-exemplar ($t(18) = 2.44$, $p_{\mathrm{FDR}} = 0.03$) and different superordinate category condition ($t(18) = 2.52$, $p_{\mathrm{FDR}} = 0.03$) for same-direction trials. In contrast, PPA showed robust facilitation, with SI values significantly below zero across all conditions (all $t(18) > 3.76$, $p_{\mathrm{FDR}} < 0.001$). A two-way ANOVA in both OPA and PPA revealed no significant effects of motion direction, category, or their interaction (all ps > 0.05). These results contrast with the suppression observed in hMT+ and the reduced facilitation in V1 under identical-exemplar condition, suggesting that scene-selective regions rely on distinct surround modulation mechanisms under dynamic naturalistic stimulation. These higher-level scene areas may integrate information from both center and surround to support scene understanding rather than exhibiting classical surround suppression.

Multivariate analyses reveal motion-dependent surround modulation in hMT+

To assess how motion direction influences surround modulation, data were collapsed across the four category conditions prior to decoding to isolate motion-related effects. We then decoded the center-only versus center+surround conditions in V1, hMT+, OPA, and PPA under two motion-direction conditions: same-direction and opposite-direction.

As shown in Figure 4A, decoding accuracy in hMT+ was above chance for both same-direction ($p_{\mathrm{FDR}} < 0.001$) and opposite-direction trials ($p_{\mathrm{FDR}} = 0.003$). A paired-samples t-test revealed higher accuracy for same-direction than opposite-direction trials ($t(18) = 3.07$, $p = 0.007$), indicating that motion congruence enhances surround modulation in hMT+ multivoxel responses. This multivariate pattern closely mirrors the univariate results, which showed stronger surround suppression when center and surround moved in the same direction. Together, these analyses demonstrate that hMT+ activity reflects motion-dependent surround modulation both in overall response amplitude and in distributed activation patterns. These converging findings suggest that surround modulation in hMT+ is driven primarily by the relative motion direction between center and surround, consistent with classical low-level results (Kinikliogl u & Boyaci, 2022; Paffen, Alais, & Verstraten, 2005; Paffen, van der Smagt, et al., 2005).

As shown in Figure 4B–D, decoding accuracy was above chance in V1, OPA, and PPA for both same-direction (all $p_{\mathrm{FDR}} < 0.008$) and opposite-direction trials (all $p_{\mathrm{FDR}} < 0.001$). However, paired-samples t-tests revealed no significant differences in decoding accuracy between same- and opposite-direction trials (all $p_{\mathrm{FDR}} > 0.20$). Although V1, OPA, and PPA showed reliable above-chance decoding, the absence of direction-related differences in these regions suggests that their surround modulation reflects general contextual integration rather than motion congruence.

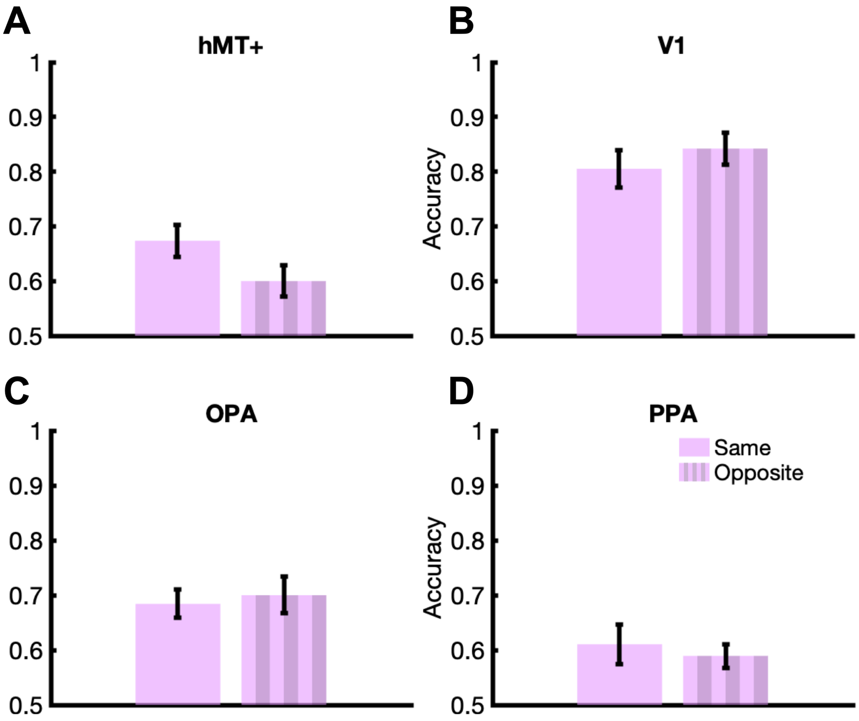

Figure 4: Multivoxel decoding accuracy for center-only and center+surround classification for the two motion-direction conditions: same-direction and opposite-direction. Mean decoding accuracy is shown for **A)** hMT+, **B)** V1, **C)** OPA, and **D)** PPA. Error bars indicate the standard error of the mean (SEM).

Multivariate analyses reveal category-dependent surround modulation in OPA and PPA

To assess how category similarity influences surround modulation, data were collapsed across the two motion-direction conditions prior to decoding. We then decoded the center-only versus center+surround conditions in V1, hMT+, OPA, and PPA across four category conditions: identical exemplar, different exemplar, different basic-level category, and different superordinate category.

As shown in Figures 5A and B, classification accuracy was significantly above chance for all category conditions in both hMT+ (all $p_{\mathrm{FDR}} < 0.01$) and V1 (all $p_{\mathrm{FDR}} < 0.001$). Despite robust decoding performance, one-way repeated-measures ANOVAs revealed no effect of category in either region (V1: $F(3, 54) = 1.93$, $p = 0.13$, $\eta^2 = 0.10$;

hMT+: $F(3, 54) = 0.56$, $p = 0.64$, $\eta^2 = 0.03$). These results indicate strong surround modulation in both areas, but this modulation did not depend on categorical similarity between center and surround, consistent with the roles of V1 and hMT+ in processing low-level visual features such as edges, orientation, contrast, and motion rather than higher-level, category-selective information.

As shown in Figure 5C, the classification accuracy was significantly above chance in OPA for all category conditions (all $p_{\mathrm{FDR}} < 0.002$). A one-way repeated-measures ANOVA revealed a main effect of category, $F(3, 54) = 4.12$, $p = 0.01$, $\eta^2 = 0.19$. Post-hoc comparisons showed higher decoding for the different superordinate-category condition than for both the identical- and different-exemplar conditions (all $t(18) > 3.01$, $p_{\mathrm{FDR}} < 0.02$), and marginally higher decoding for the different basic-level category condition than for the identical-exemplar condition ($t(18) = 2.04$, $p_{\mathrm{FDR}} = 0.06$). Decoding did not differ between the identical- and different-exemplar conditions, $t(18) = 0.70$, $p_{\mathrm{FDR}} = 0.51$.

These results suggest that surround modulation in OPA increases with categorical distance. Multivoxel responses were more strongly influenced by the surround when the center and surround scenes belonged to different categories, indicating that OPA is sensitive not only to the presence of contextual surround input but also to its semantic relationship with the center. This categorical sensitivity aligns with our behavioral findings (Kiniklioglu & Kaiser, 2025), in which surround suppression increased as categorical similarity decreased.

PPA showed a similar pattern, with decoding accuracy increasing as the categorical dissimilarity between center and surround increased. Decoding was above chance for all category conditions except the identical-exemplar condition ($p_{\mathrm{FDR}} = 0.17$; Figure 4D).

The main effect of category was significant, $F(3, 54) = 2.93$, $p = 0.04$, $\eta^2 = 0.14$. Post-hoc comparisons revealed higher decoding accuracy for the different superordinate-category condition than for the identical-exemplar condition ($t(18) = 2.59$, $p = 0.02$), and higher decoding accuracy for the different basic-level category condition than for the identical-exemplar condition ($t(18) = 2.18$, $p = 0.04$). Compared with OPA, surround modulation in PPA exhibited a more gradual and weaker increase in decoding accuracy with categorical distance. Taken together, these results extend surround modulation effects to the scene-selective cortex, demonstrating that surround modulation in these regions is shaped by higher-level categorical relationships.

Discussion

The present study investigated how motion congruence and categorical similarity between center and surround scenes shape neural surround modulation. Using fMRI, we measured neural responses while systematically varying the relationship of center and surround across four levels of categorical similarity (identical exemplar, different exemplar, different basic-level category, and different superordinate category) and two levels of motion congruence (same vs. opposite direction). Univariate analyses revealed distinct patterns of surround modulation across the visual hierarchy. hMT+ showed robust motion-dependent surround suppression, with stronger suppression when the center and surround moved in the same direction. In contrast, V1 exhibited surround facilitation across all conditions, which was reduced when the center and surround were identical and moved in the same direction. Scene-selective regions OPA and PPA also showed predominantly facilitation rather than suppression. Multivariate analyses complemented these findings: in hMT+, decoding between the center-only

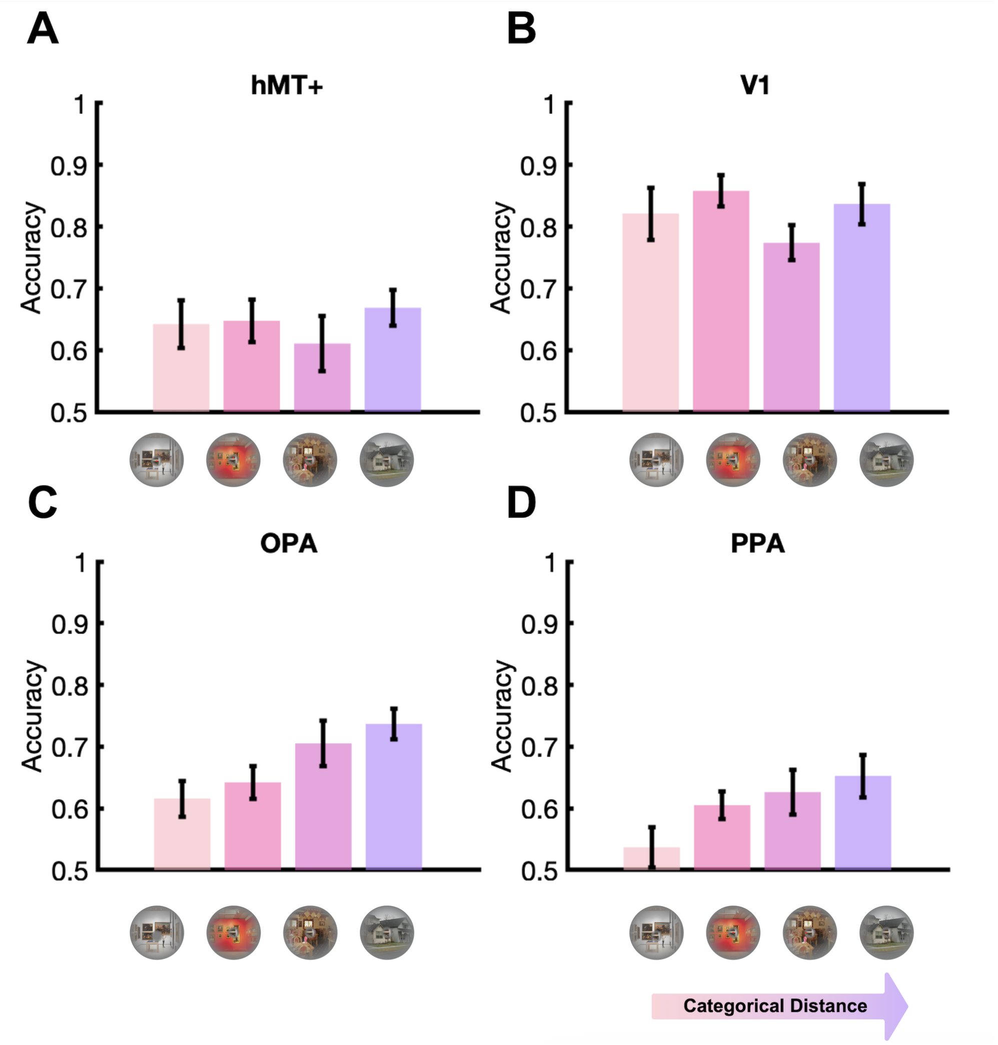

Figure 5: Multivoxel decoding accuracy for the center-only versus center+surround classification across the four category conditions: identical exemplar, different exemplar, different basic-level category, and different superordinate category. Example stimuli for each category condition are displayed along the x-axis to illustrate the categorical hierarchy, with categorical distance increasing from left to right (from the identical exemplar to the different superordinate category conditions). Mean decoding accuracy is shown for **A)** hMT+, **B)** V1, **C)** OPA, and **D)** PPA. Error bars represent the standard error of the mean (SEM).

versus center+surround conditions was stronger for same- than opposite-direction motion, whereas in OPA and PPA, surround modulation was category-dependent, with decoding accuracy increasing as categorical similarity between center and surround decreased.

These results provide the first neural evidence for surround modulation in humans with dynamic natural scenes. While surround suppression has been extensively characterized with simple stimuli such as gratings (Cavanaugh et al., 2002b; Schallmo et al., 2018; Tadin et al., 2003), its neural basis under naturalistic conditions has remained largely unexplored. The present findings demonstrate that surrounding scene context systematically modulates neural responses to central scenes across multiple stages of the visual hierarchy: from low-level facilitation in V1 to motion-dependent suppression in hMT+ and category-dependent modulation in scene-selective regions. This hierarchical organization suggests that surround modulation, a well-established mechanism in early visual processing, is also engaged during naturalistic scene perception, with both motion congruence and categorical similarity shaping how contextual information is integrated across the visual system.

Motion-dependent surround modulation in hMT+

hMT+ exhibited stronger suppression when the center and surround moved in the same direction than when they moved in opposite directions. Decoding analyses further confirmed that this surround modulation was driven specifically by the motion direction, independent of categorical similarity. This pattern aligns with classical neurophysiological findings showing that MT neurons show stronger suppression for same-direction motion and weaker or no suppression when the center and surround move in opposite directions (Allman et al., 1985; Born & Tootell, 1992; Cavanaugh et al., 2002b; Kastner et al., 1995; Lamme, 1995). Such motion-dependent modulation likely reflects local inhibitory interactions among direction-selective neurons in MT, consistent with the *MT hypothesis* of surround suppression (Tadin et al., 2003). By extending these low-level findings to naturalistic conditions, our results provide neural evidence that motion-based center–surround interactions are recruited during complex scene processing in hMT+. Furthermore, the close correspondence between our neural and behavioral findings using identical stimuli (Kiniklioglu & Kaiser, 2025) suggests that perceptual suppression during natural vision may arise from these motion-sensitive inhibitory mechanisms in hMT+.

Facilitation in V1 Reflects Sensitivity to Physical Similarity

Although facilitation in V1 was robust across conditions, it was significantly weaker for identical exemplars than for other category conditions in same-direction trials, indicating that greater physical similarity between the center and surround reduces facilitation. Moreover, motion direction affected responses only in the identical-exemplar condition, with facilitation decreasing when the center and surround moved in the same direction compared to opposite directions. This pattern aligns with classical findings showing the strongest suppression for iso-oriented, uniform stimuli (Cavanaugh et al., 2002a; DeAngelis et al., 1994; Flevaris & Murray, 2015; Schallmo et al., 2016; Serrano-Pedraza et al., 2012; Walker et al., 1999). The motion-dependent modulation also resembles suppression effects reported in hMT+; however, in V1, the influence manifests as reduced facilitation rather than suppression. These effects may arise from horizontal interactions within V1 or feedback from motion-selective areas such as hMT+, consistent with evidence that motion information can be transmitted across early visual areas (Angelucci & Bressloff, 2006; Angelucci et al., 2017; Hupé et al., 2001).

Supporting this interpretation, animal studies using naturalistic stimuli have shown that surround modulation in the primary visual cortex depends strongly on low-level image similarity. For example, Coen-Cagli et al. (2015) reported that macaque V1 exhibits stronger suppression when the center and surround consist of homogeneous natural image regions. This aligns broadly with our findings, which also reveal surround modulation driven by physical similarity, although it manifested as reduced facilitation rather than increased suppression. Consistent with this, multivariate analyses further showed robust surround modulation in V1, but this effect did not depend on categorical similarity or motion congruence. Together, these results indicate that surround modulation in V1 primarily reflects sensitivity to low-level physical similarity between center and surround, rather than higher-level, category-selective information.

Nevertheless, the nature of surround modulation in V1 remains debated. Some studies have reported suppression, suggesting that early visual mechanisms may contribute to the suppressive effects often observed in higher-level visual areas (Angelucci et al., 2017; Nurminen et al., 2009, 2013; Zenger-Landolt & Heeger, 2003), whereas others have reported surround facilitation (Er et al., 2020; Press et al., 2001). Importantly, several studies have demonstrated both suppression and facilitation in V1, depending on stimulus configuration, task demands, and attentional context (Flevaris & Murray, 2015; Ichida et al., 2007; Williams et al., 2003). It is therefore possible that the facilitation observed in our study reflects context-dependent modulation. Because the task did not explicitly require attention to the center–surround relationship, suppressive interactions may have been reduced, resulting in an overall facilitative response.

Alternatively, the facilitation observed in V1 may be influenced by feedback from scene-selective cortex. Surround modulation in V1 is thought to arise from a complex interplay of feedforward, horizontal, and feedback mechanisms (Nurminen & Angelucci, 2014; Nurminen et al., 2018). As part of this highly interconnected network, V1 receives extensive feedback from other visual areas (Shao & Burkhalter, 1996), including those associated with scene processing (Morgan et al., 2019; Rockland & Van Hoesen, 1994). Top-down inputs from scene-selective regions may bias early visual responses toward facilitation, dynamically regulating the balance between facilitation and suppression and thereby contributing to the overall pattern observed in V1.

Category-dependent modulation in scene-selective cortex

OPA and PPA primarily exhibited facilitation, likely reflecting their relatively large receptive fields that integrate information across broad regions of the visual field (e.g., Levy et al., 2001; Nasr et al., 2011; Silson et al., 2015). This integration may allow them to combine inputs from both center and surround to support scene understanding rather than exhibiting classical surround suppression. While univariate analyses revealed no category or motion effects, the more sensitive multivariate analyses showed clear category-dependent surround modulation, particularly in OPA. Multivoxel responses were more strongly influenced by the surround when the center and surround scenes belonged to different categories, indicating that OPA is sensitive not only to the presence of contextual information but also to its categorical relationship with the central scene. PPA showed a similar but weaker gradual increase in surround modulation with categorical distance, possibly reflecting the lower signal-to-noise ratio typically observed in ventral visual regions (e.g., Rua et al., 2018; Winawer et al., 2010).

Importantly, decoding revealed no motion-related effects in either region, indicating that surround modulation in scene-selective cortex is driven specifically by categorical similarity. These results are consistent with our behavioral findings (Kiniklioglu & Kaiser, 2025), suggesting a shared organizing principle across neural and perceptual levels whereby contextual influences strengthen as scenes become more categorically dissimilar. The correspondence between neural and behavioral results obtained with identical stimuli supports the idea that perceptual contextual effects during natural scene perception may reflect category-sensitive integrative processes in scene-selective regions.

Similar sensitivity to categorical context has been demonstrated in prior neuroimaging work on scene and object perception. Contextual incongruence between scene elements has been shown to alter neural responses in parahippocampal and occipitotemporal regions (Faivre et al., 2019; Peyrin et al., 2021; Rémy et al., 2014). For example, Peyrin et al. (2021) reported that categorically congruent (e.g., both man-made) but physically dissimilar (e.g., buildings vs. streets) peripheral scenes disrupted central scene categorization and increased activation in inferior frontal and occipitotemporal cortices. Consistent with this, we found that surround modulation in scene-selective cortex was stronger even when the center and surround belonged to the same superordinate category (e.g., two indoor scenes) but differed at the basic level (e.g., restaurant vs. museum).

Hierarchical surround modulation may support efficient scene representation

Real-world scenes show systematic spatial and semantic regularities (e.g., a bathroom typically contains a sink, mirror, and shower arranged within a characteristic layout). The visual system exploits these regularities to integrate information efficiently across multiple representational levels, supporting rapid understanding of complex environments (Bar, 2004; Kaiser et al., 2019; Võ, 2021). Computational and neurophysiological models suggest that such contextual integration relies on recurrent and inhibitory interactions that predictively modulate sensory input according to its relevance (Angelucci & Bressloff, 2006; Gilbert & Li, 2013; Rao & Ballard, 1999). Consistent with this view, behavioral and neuroimaging work demonstrates that coherent scene structure facilitates both scene and object perception (Chen et al., 2022; Davenport & Potter, 2004; Kaiser & Peelen, 2018; Kaiser et al., 2019, 2020, 2021; Võ & Wolfe, 2013). Together, these findings suggest that contextual mechanisms may enhance informative signals while attenuating redundant input, thereby promoting stable and efficient visual processing in natural environments.

Our results build on this framework by showing that categorically incongruent surrounds elicit stronger contextual modulation in scene-selective cortex than congruent surrounds. Whereas studies using simple stimuli have attributed surround modulation to the suppression of statistical redundancies in low-level features (Angelucci et al., 2017; Coen-Cagli et al., 2012; Nurminen & Angelucci, 2014; Schwartz & Simoncelli, 2001; Vinje & Gallant, 2000), our findings indicate that, at higher representational levels, contextual influences are governed primarily by categorical relationships. In natural scenes, categorically incongruent information may enhance neural responses because it violates contextual expectations and conveys higher informational value, whereas with simple stimuli, identical information may be suppressed because it is predictable and redundant. In contrast, V1 showed a pattern more consistent with low-level surround modulation, where facilitation decreased when the center and surround were physically identical, indicating sensitivity primarily to visual redundancy rather than categorical context. Both forms of modulation may therefore reflect a shared computational principle: optimizing neural representations by emphasizing informative differences while attenuating predictable input.

Together, these results reveal a hierarchical organization of surround modulation in natural vision. Early visual areas (V1) are primarily governed by physical similarity, mid-level motion areas (hMT+) are sensitive to motion congruence, and higher-level scene-selective regions (OPA and PPA) integrate categorical structure. This progression supports a predictive and efficient coding framework (Angelucci et al., 2017; Gilbert & Li, 2013; Rao & Ballard, 1999), whereby each stage of the visual hierarchy selectively suppresses predictable or redundant input and enhances informative discrepancies.

Limitations and future directions

Although the present study provides neural evidence for both categorical and motion-based surround modulation in natural vision, several questions remain open for future investigation. First, we did not explicitly manipulate low-level visual features such as spatial frequency or amplitude spectrum, which may covary with categorical similarity and contribute to surround modulation (Peyrin et al., 2021). Future work could examine these factors more directly by varying physical properties and categorical coherence independently to disentangle their respective influences.

Second, while we used sub-ROI localizers in V1 and hMT+ to identify voxels corresponding to the center stimulus, this approach could not be applied to OPA and PPA due to their larger receptive fields. Consequently, the ROIs in these regions may have included voxels responsive to both center and surround, so that the effects cannot be directly interpreted as modulatory effects on the representation of the central stimulus. Moreover, scene-selective areas are thought to contain fine-grained functional subfields with heterogeneous tuning to spatial scale, visual field position, and category information (e.g., Baldassano et al., 2013; Silson et al., 2015). The absence of univariate category effects in OPA and PPA may therefore reflect this spatial and functional heterogeneity, where the averaging across voxels with different category or positional preferences could mask category-specific responses. Category-specific modulatory effects may instead depend on multivariate response patterns in scene-selective cortex, necessitating pattern-based analysis approaches for characterizing this fine-scale organization.

Conclusions

Using dynamic natural scenes, we demonstrate that surround modulation is a canonical property of human visual processing, shaped by both motion-direction congruence and categorical similarity. Univariate and multivariate analyses revealed distinct mechanisms across the visual hierarchy: in hMT+, surround suppression depended on motion congruence (stronger for same-direction motion), consistent with classic center–surround interactions among direction-selective neurons; in V1, responses were predominantly facilitatory but weakened when center and surround were physically identical and moved in the same direction, reflecting sensitivity to low-level similarity; in scene-selective cortex (OPA, PPA), surround influences were largely facilitatory and scaled with categorical distance in multivoxel patterns, indicating sensitivity to higher-level categorical structure. Together, these findings extend surround modulation from simple stimuli to naturalistic vision, with contextual influences shifting from sensitivity to physical similarity in the early visual cortex to category-sensitive integration in higher-level regions. This hierarchical progression aligns with predictive and efficient coding accounts, whereby each stage attenuates predictable input and emphasizes informative discrepancies to support perceptual stability and efficient scene understanding.

Data Availability

The data set and analysis scripts will be available on the Open Science Framework.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG), grants SFB/TRR 135 (project no. 222641018); KA4683/5-1 (project no. 518483074); and under Germany's Excellence Strategy (EXC 3066/1, “The Adaptive Mind”, project no. 533717223). It was further supported by an European Research Council (ERC) Starting Grant (PEP, ERC-2022-STG 101076057). Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

MR imaging for this study was performed at the Bender Institute of Neuroimaging (BION) at Justus Liebig University Giessen, Germany.

We thank Tugce Dalmis for assistance with data collection.

Competing interests

The authors declare no competing interests.

References

Allman, J., Miezin, F., & McGuinness, E. (1985). Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (mt). *Perception*, *14*, 105–126. https://doi.org/10.1068/p140105

Angelucci, A., Bijanzadeh, M., Nurminen, L., Federer, F., Merlin, S., & Bressloff, P. C. (2017). Circuits and mechanisms for surround modulation in visual cortex. *Annual Review of Neuroscience*, *40*, 425–451. https://doi.org/10.1146/annurev-neuro-072116-031418

Angelucci, A., & Bressloff, P. C. (2006). The contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra- classical receptive field surround of primate v1 neurons. *Progress in Brain Research*, *154*, 93–120. https://doi.org/https://doi.org/10.1016/S0079-6123(06)54005-1

Baldassano, C., Beck, D. M., & Fei-Fei, L. (2013). Differential connectivity within the parahippocampal place area. *NeuroImage*, *75*, 228–237. https://doi.org/10.1016/j.neuroimage.2013.02.073

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629. https://doi.org/10.1038/nrn1476

Born, R. T., & Tootell, R. B. H. (1992). Segregation of global and local motion processing in primate middle temporal visual area. *Nature*, *357*, 497–499. https://doi.org/10.1038/357497a0

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Cavanaugh, J. R., Bair, W., & Movshon, J. A. (2002a). Nature and interaction of signals from the receptive field center and surround in macaque v1 neurons. *Journal of Neurophysiology*, *88*(4), 2530–2546. https://doi.org/10.1152/jn.00692.2001

Cavanaugh, J. R., Bair, W., & Movshon, J. A. (2002b). Selectivity and spatial distribution of signals from the receptive field surround in macaque v1 neurons. *Journal of Neurophysiology*, *88*, 2547–2556. https://doi.org/10.1152/jn.00693.2001

Chen, L., Cichy, R. M., & Kaiser, D. (2022). Semantic scene-object consistency modulates n300/400 eeg components, but does not automatically facilitate object representations. *Cerebral Cortex*, *32*(16), 3553–3567. https://doi.org/10.1093/cercor/bhab433

Ciric, R., Wolf, D. H., Power, J. D., Roalf, D. R., Baum, G. L., Ruparel, K., Shinohara, R. T., Elliott, M. A., Eickhoff, S. B., Davatzikos, C., Gur, R. C., Gur, R. E., Basset, D. S., & Satterthwaite, T. D. (2017). Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *NeuroImage*, *154*, 174–187. https://doi.org/10.1016/j.neuroimage.2017.03.020

Coen-Cagli, R., Dayan, P., & Schwartz, O. (2012). Cortical surround interactions and perceptual salience via natural scene statistics. *PLOS Computational Biology*, *8*(3), e1002405. https://doi.org/10.1371/journal.pcbi.1002405

Coen-Cagli, R., Kohn, A., & Schwartz, O. (2015). Flexible gating of contextual influences in natural vision. *Nature Neuroscience*, *18*(11), 1648–1655. https://doi.org/10.1038/nn.4128

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. segmentation and surface reconstruction. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559–564. https://doi.org/10.1111/j.0956-7976.2004.00719.x

DeAngelis, G. C., Freeman, R. D., & Ohzawa, I. (1994). Length and width tuning of neurons in the cat's primary visual cortex. *Journal of Neurophysiology*, *71*(1), 347–374. https://doi.org/10.1152/jn.1994.71.1.347

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*(2), 181–192.

Er, G., Pamir, Z., & Boyaci, H. (2020). Distinct patterns of surround modulation in v1 and hmt+. *NeuroImage*, *220*, 117084. https://doi.org/10.1016/j.neuroimage.2020.117084

Faivre, N., Dubois, J., Schwartz, N., & Mudrik, L. (2019). Imaging object-scene relations processing in visible and invisible natural scenes. *Scientific Reports*, *9*(1), 4567. https://doi.org/10.1038/s41598-019-38654-z

Faurite, C., Aprile, E., Kauffmann, L., Mermillod, M., Gallice, M., Chiquet, C., Cottereau, B. R., & Peyrin, C. (2024). Interaction between central and peripheral vision: Influence of distance and spatial frequencies. *Journal of Vision*, *24*(1), 3. https://doi.org/10.1167/jov.24.1.3

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis: Ii: Inflation, flattening, and a surface-based coordinate system. *NeuroImage*, *9*(2), 195–207. https://doi.org/10.1006/nimg.1998.0396

Flevaris, A. V., & Murray, S. O. (2015). Attention determines contextual enhancement versus suppression in human primary visual cortex. *The Journal of Neuroscience*, *35*(35), 12273–12280. https://doi.org/10.1523/JNEUROSCI.1409-15.2015

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, *14*(5), 350–363. https://doi.org/10.1038/nrn3476

Greenberg, A. S., Verstynen, T., Chiu, Y.-C., Yantis, S., Schneider, W., & Behrmann, M. (2012). Visuotopic cortical connectivity underlying attention revealed with white-matter tractography. *Journal of Neuroscience*, *32*(8), 2773–2782. https://doi.org/10.1523/JNEUROSCI.5419-11.2012

Hallum, L. E., & Movshon, J. A. (2014). Surround suppression supports second-order feature encoding by macaque v1 and v2 neurons. *Vision Research*, *104*, 24–35. https://doi.org/10.1016/j.visres.2014.10.004

Huk, A. C., Dougherty, R. F., & Heeger, D. J. (2002). Retinotopy and functional subdivision of human areas mt and mst. *Journal of Neuroscience*, *22*(5), 7195–7205.

Hupé, J.-M., James, A. C., Girard, P., Lomber, S. G., Payne, B. R., & Bullier, J. (2001). Feedback connections act on the early part of the responses in monkey visual cortex. *Journal of Neurophysiology*, *85*(1), 134–145. https://doi.org/10.1152/jn.2001.85.1.134

Ichida, J. M., Schwabe, L., Bressloff, P. C., & Angelucci, A. (2007). Response facilitation from the ”suppressive” receptive field surround of macaque v1 neurons. *Journal of Neurophysiology*, *98*(4), 2168–2181. https://doi.org/10.1152/jn.00298.2007

Kaiser, D., Häberle, G., & Cichy, R. M. (2020). Cortical sensitivity to natural scene structure. *Human Brain Mapping*, *41*(5), 1286–1295. https://doi.org/10.1002/hbm.24875

Kaiser, D., Häberle, G., & Cichy, R. M. (2021). Coherent natural scene structure facilitates the extraction of task-relevant object information in visual cortex. *NeuroImage*, *240*, 118365. https://doi.org/10.1016/j.neuroimage.2021.118365

Kaiser, D., & Peelen, M. V. (2018). Transformation from independent to integrative coding of multi-object arrangements in human visual cortex. *NeuroImage*, *169*, 334–341. https://doi.org/10.1016/j.neuroimage.2017.12.065

Kaiser, D., Quek, G. L., Cichy, R. M., & Peelen, M. V. (2019). Object vision in a structured world. *Trends in Cognitive Sciences*, *23*(8), 672–685. https://doi.org/10.1016/j.tics.2019.04.013

Kastner, S., Nothdurft, H. C., & Pigarev, I. N. (1995). Neuronal correlates of pop-out in cat striate cortex. *Vision Research*, *37*, 371–376. https://doi.org/10.1016/S0042-6989(96)00184-8

Kiniklioglu, M., & Boyaci, H. (2025). Hmt+ activity predicts the effect of spatial attention on surround suppression. *Journal of Vision*, *25*(4), 12. https://doi.org/10.1167/jov.25.4.12

Kiniklioglu, M., & Boyaci, H. (2022). Increasing the spatial extent of attention strengthens surround suppression. *Vision Research*, *199*, 108074. https://doi.org/10.1016/j.visres.2022.108074

Kiniklioglu, M., & Kaiser, D. (2025). Characterizing surround suppression with dynamic natural scenes [preprint]. *bioRxiv*. https://doi.org/10.1101/2025.08.11.669612

Küçük, E., Foxwell, M., Kaiser, D., & Pitcher, D. (2024). Moving and static faces, bodies, objects, and scenes are differentially represented across the three visual pathways. *Journal of Cognitive Neuroscience*, *36*(12), 2639–2651. https://doi.org/10.1162/jocn_a_02139

Lamme, V. A. F. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of Neuroscience*, *15*, 1605–1616. https://doi.org/10.1523/JNEUROSCI.15-02-01605.1995

Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center-periphery organization of human object areas. *Nature Neuroscience*, *4*(5), 533–539. https://doi.org/10.1038/87490

Montoya, S. A., Moser, H. R., Kamath, R. S., Chong, L. S., Grant, A. N., Marjańska, M., Sponheim, S. R., Engel, S. A., Olman, C. A., & Schallmo, M.-P. (2025). Center-surround processing in psychosis [preprint]. *medRxiv*. https://doi.org/10.1101/2025.08.15.25333221

Morgan, A. T., Petro, L. S., & Muckli, L. (2019). Scene representations conveyed by cortical feedback to early visual cortex can be described by line drawings [The official journal of the Society for Neuroscience]. *The Journal of Neuroscience*, *39*(47), 9410–9423. https://doi.org/10.1523/JNEUROSCI.0852-19.2019

Nasr, S., Liu, N., Devaney, K. J., Yue, X., Rajimehr, R., Ungerleider, L. G., & Tootell, R. B. H. (2011). Scene-selective cortical regions in human and nonhuman primates. *The Journal of Neuroscience*, *31*(39), 13771–13785. https://doi.org/10.1523/JNEUROSCI.2792-11.2011

Nurminen, L., & Angelucci, A. (2014). Multiple components of surround modulation in primary visual cortex: Multiple neural circuits with multiple functions? *Vision Research*, *104*, 47–56. https://doi.org/10.1016/j.visres.2014.08.018

Nurminen, L., Kilpeläinen, M., Laurinen, P., & Vanni, S. (2009). Area summation in human visual system: Psychophysics, fmri, and modeling. *Journal of Neurophysiology*, *102*(5), 2900–2909. https://doi.org/https://doi.org/10.1152/jn.00201.2009

Nurminen, L., Kilpeläinen, M., & Vanni, S. (2013). Fovea-periphery axis symmetry of surround modulation in the human visual system. *PLOS ONE*, *8*(2), 1–11. https://doi.org/10.1371/journal.pone.0057906

Nurminen, L., Merlin, S., Bijanzadeh, M., Federer, F., & Angelucci, A. (2018). Top-down feedback controls spatial summation and response amplitude in primate visual cortex. *Nature Communications*, *9*, 2281. https://doi.org/10.1038/s41467-018-04500-5

Onat, S., Jancke, D., & König, P. (2013). Cortical long-range interactions embed statistical knowledge of natural sensory input: A voltage-sensitive dye imaging study. *F1000Research*, *2*, 51. https://doi.org/10.12688/f1000research.2-51.v2

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). Cosmomvpa: Multi-modal multivariate pattern analysis of neuroimaging data in matlab/gnu octave. *Frontiers in Neuroinformatics*, *10*, 27. https://doi.org/10.3389/fninf.2016.00027

Pack, C. C., Hunter, J. N., & Born, R. T. (2005). Contrast dependence of suppressive influences in cortical area mt of alert macaque. *Journal of Neurophysiology*, *93*(3), 1809–1815. https://doi.org/10.1152/jn.00629.2004

Paffen, C. L. E., Alais, D., & Verstraten, F. A. (2005). Center–surround inhibition deepens binocular rivalry suppression. *Vision Research*, *45*, 2642–2649. https://doi.org/10.1016/j.visres.2005.04.018

Paffen, C. L. E., van der Smagt, J., Maarten, te Pas, S., & F.Verstraten. (2005). Center-surround inhibition and facilitation as a function of size and contrast at multiple levels of visual motion processing. *Journal of Vision*, *5*, 571–578. https://doi.org/10.1167/5.6.8

Parkes, L., Fulcher, B. D., Yücel, M., & Fornito, A. (2018). An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional mri. *NeuroImage*, *171*, 415–436. https://doi.org/10.1016/j.neuroimage.2017.12.073

Petrov, Y., Carandini, M., & McKee, S. (2005). Two distinct mechanisms of suppression in human vision. *Journal of Neuroscience*, *25*(38), 8704–8707. https://doi.org/https://doi.org/10.1523/JNEUROSCI.2871-05.2005

Peyrin, C., Roux-Sibilon, A., Trouilloud, A., Khazaz, S., Joly, M., Pichat, C., & Kauffmann, L. (2021). Semantic and physical properties of peripheral vision are used for scene categorization in central vision. *Journal of Cognitive Neuroscience*, *33*(5), 799–813. https://doi.org/10.1162/jocn_a_01689

Pihlaja, M., Henriksson, L., James, A., & Vanni, S. (2008). Quantitative multifocal fmri shows active suppression in human v1. *Human Brain Mapping*, *29*(9), 1001–1014. https://doi.org/10.1002/hbm.20442

Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C., & Kanwisher, N. (2011). Differential selectivity for dynamic versus static information in face-selective cortical regions. *NeuroImage*, *56*(4), 2356–2363. https://doi.org/10.1016/j.neuroimage.2011.03.067

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity mri networks arise from subject motion. *NeuroImage*, *59*(3), 2142–2154. https://doi.org/10.1016/j.neuroimage.2011.10.018

Press, W. A., Brewer, A. A., Dougherty, R. F., Wade, A. R., & Wandell, B. A. (2001). Visual areas and spatial summation in human visual cortex. *Vision Research*, *41*(10–11), 1321–1332. https://doi.org/10.1016/S0042-6989(01)00074-8

Rao, R., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*, 79–87. https://doi.org/10.1038/4580

Rémy, F., Vayssière, N., Pins, D., Boucart, M., & Fabre-Thorpe, M. (2014). Incongruent object/context relationships in visual scenes: Where are they processed in the brain? *Brain and Cognition*, *84*(1), 34–43. https://doi.org/10.1016/j.bandc.2013.10.008

Rockland, K. S., & Van Hoesen, G. W. (1994). Direct temporal-occipital feedback connections to striate cortex (v1) in the macaque monkey. *Cerebral Cortex*, *4*(3), 300–313. https://doi.org/10.1093/cercor/4.3.300

Rua, C., Wastling, S. J., Costagli, M., Symms, M. R., Biagi, L., Cosottini, M., Del Guerra, A., Tosetti, M., & Barker, G. J. (2018). Improving fmri in signal drop-out regions at 7 t by using tailored radio-frequency pulses: Application to the ventral occipito-temporal cortex. *MAGMA (New York, N.Y.)*, *31*(2), 257–267. https://doi.org/10.1007/s10334-017-0652-x

Schallmo, M.-P., Grant, A. N., Burton, P. C., & Olman, C. A. (2016). The effects of orientation and attention during surround suppression of small image features: A 7 tesla fmri study. *Journal of Vision*, *16*(10), 19. https://doi.org/10.1167/16.10.19

Schallmo, M.-P., Kale, A. M., Flevaris, A. V., Brkanac, Z., Edden, R. A., Bernier, R. A., & Murray, S. O. (2018). Suppression and facilitation of human neural responses. *eLife*, *7*, 1–23. https://doi.org/https://doi.org/10.7554/eLife.30334.001

Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature Neuroscience*, *4*(8), 819–825. https://doi.org/10.1038/90526

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K., Belliveau, J. W., Brady, T. J., Rosen, B. R., & Tootell, R. B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*(5212), 889–893.

Serrano-Pedraza, I., Grady, J. P., & Read, J. C. A. (2012). Spatial frequency bandwidth of surround suppression tuning curves. *Journal of Vision*, *12*(6), 24. https://doi.org/10.1167/12.6.24

Shao, Z., & Burkhalter, A. (1996). Different balance of excitation and inhibition in forward and feedback circuits of rat visual cortex [The official journal of the Society for Neuroscience]. *The Journal of Neuroscience*, *16*(22), 7353–7365. https://doi.org/10.1523/JNEUROSCI.16-22-07353.1996

Shushruth, S., Mangapathy, P., Ichida, J. M., Bressloff, P. C., Schwabe, L., & Angelucci, A. (2012). Strong recurrent networks compute the orientation tuning of surround modulation in the primate primary visual cortex. *The Journal of Neuroscience*, *32*, 308–321. https://doi.org/10.1523/JNEUROSCI.3789-11.2012

Shushruth, S., Nurminen, L., Bijanzadeh, M., Ichida, J. M., Vanni, S., & Angelucci, A. (2013). Different orientation tuning of near- and far-surround suppression in macaque primary visual cortex mirrors their tuning in human perception. *The Journal of Neuroscience*, *33*(1), 106–119. https://doi.org/10.1523/JNEUROSCI.2518-12.2013

Silson, E. H., Chan, A. W.-Y., Reynolds, R. C., Kravitz, D. J., & Baker, C. I. (2015). A retinotopic basis for the division of high-level scene processing between lateral and ventral human occipitotemporal cortex. *The Journal of Neuroscience*, *35*(34), 11921–11935. https://doi.org/10.1523/JNEUROSCI.0137-15.2015

Slotnick, S. D., & Yantis, S. (2003). Efficient acquisition of human retinotopic maps. *Human Brain Mapping*, *18*(1), 22–29. https://doi.org/10.1002/hbm.10077

Sundberg, K. A., Mitchell, J. F., & Reynolds, J. H. (2009). Spatial attention modulates center-surround interactions in macaque visual area v4. *Neuron*, *61*(6), 952–963. https://doi.org/10.1016/j.neuron.2009.02.023

Tadin, D., Lappin, J. S., Gilroy, L. A., & Blake, R. (2003). Perceptual consequences of centre- surround antagonism in visual motion processing. *Nature*, *424*, 312–315. https://doi.org/10.1038/nature01800

Tadin, D., Silvanto, J., Pascual-Leone, A., & Battelli, L. (2011). Improved motion perception and impaired spatial suppression following disruption of cortical area mt/v5. *Journal of Neuroscience*, *31*(4), 1279–1283. https://doi.org/https://doi.org/10.1523/JNEUROSCI.4121-10.2011

Turkozer, H. B., Pamir, Z., & Boyaci, H. (2016). Contrast affects fmri activity in middle temporal cortex related to center–surround interaction in motion perception. *Frontiers in Psychology*, *7*, 1–8. https://doi.org/10.3389/fpsyg.2016.00454

Vinje, W. E., & Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, *287*(5456), 1273–1276. https://doi.org/10.1126/science.287.5456.1273

Võ, M. L.-H. (2021). The meaning and structure of scenes. *Vision Research*, *181*, 10–20. https://doi.org/10.1016/j.visres.2020.11.003

Võ, M. L.-H., & Wolfe, J. M. (2013). Differential electrophysiological signatures of semantic and syntactic scene processing. *Psychological Science*, *24*(9), 1816–1823. https://doi.org/10.1177/0956797613476955

Walker, G. A., Ohzawa, I., & Freeman, R. D. (1999). Asymmetric suppression outside the classical receptive field of the visual cortex. *Journal of Neuroscience*, *19*(23), 10536–10553. https://doi.org/10.1523/JNEUROSCI.19-23-10536.1999

Williams, A. L., Singh, K. D., & Smith, A. T. (2003). Surround modulation measured with functional mri in the human visual cortex. *Journal of Neurophysiology*, *89*(1), 525–533. https://doi.org/10.1152/jn.00048.2002

Winawer, J., Horiguchi, H., Sayres, R. A., Amano, K., & Wandell, B. A. (2010). Mapping hv4 and ventral occipital cortex: The venous eclipse. *Journal of Vision*, *10*(5), 1–1. https://doi.org/10.1167/10.5.1

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of fmri data. *NeuroImage*, *14*(6), 1370–1386. https://doi.org/10.1006/nimg.2001.0931

Xing, J., & Heeger, D. J. (2001). Measurement and modeling of center-surround suppression and enhancement. *Vision Research*, *41*, 571–583. https://doi.org/10.1016/S0042-6989(00)00270-4

Zenger-Landolt, B., & Heeger, D. J. (2003). Response suppression in v1 agrees with psychophysics of surround masking. *Journal of Neuroscience*, *23*(17), 6884–6893. https://doi.org/10.1523/JNEUROSCI.23-17-06884.2003