Neuron, Volume *110*

Supplemental information

Predictive coding of natural images by V1 firing rates and rhythmic synchronization

Cem Uran, Alina Peter, Andreea Lazar, William Barnes, Johanna Klon-Lipok, Katharine A. Shapcott, Rasmus Roese, Pascal Fries, Wolf Singer, and Martin Vinck

Supplementary Information for: Predictive coding of natural images by V1 firing rates and rhythmic synchronization.

Cem Uran[a,e,g,*], Alina Peter[a,*], Andreea Lazar[a], William Barnes[a,b], Johanna Klon-Lipok[a,b], Katharine A Shapcott[a,c], Rasmus Roese[a], Pascal Fries[a,d], Wolf Singer[a,b,c], Martin Vinck[a,e,g,f]

[a] *Ernst Strüngmann Institute (ESI) for Neuroscience in Cooperation with Max Planck Society, 60528 Frankfurt, Germany*
[b] *Max Planck Institute for Brain Research, 60438 Frankfurt, Germany*
[c] *Frankfurt Institute for Advanced Studies, 60438 Frankfurt, Germany*
[d] *Donders Institute for Brain, Cognition and Behaviour, Department of Biophysics, Radboud University Nijmegen, 6525 AJ Nijmegen, Netherlands*
[e] *Donders Centre for Neuroscience, Department of Neuroinformatics, Radboud University Nijmegen, 6525 AJ Nijmegen, Netherlands*
[f] *Lead contact*
[g] *Correspondence to cem.uran@esi-frankfurt.de and martin.vinck@esi-frankfurt.de*

---

[*] Equally contributing

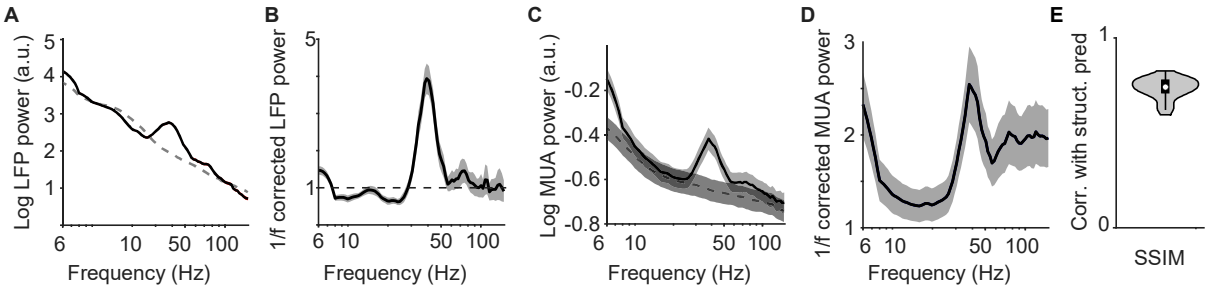

Figure S1: **Example recordings, and relationship of structural predictability with perceptual similarity. Relates to Figure 1 of Main text.** **(A)** For the example image shown in Figure 1A, LFP spectra showed a characteristic $1/f$ trend during the baseline period (dashed line), and an increase in gamma-band activity during the stimulus period (solid line). **(B)** After $1/f$ correction, a clear gamma-peak in the LFP was visible. **(C-D)** MUA spiking activity showed a gamma-peak during visual stimulation both in the raw spectra and in the $1/f$-corrected spectrum (solid line). **(E)** Spearman correlation of the perceptual similarity measure SSIM to the structural predictability measure defined in Figure 1 across images. SSIM is based on a combination of structural similarity (computed as pixel-by-pixel correlations) but additionally factors in contrast and luminance correlations between images. In our measure of structural predictability, we omitted the dependence of SSIM on contrast and spatial frequency to avoid an influence of these low-level image factors. Data in panels A-D are represented as mean ± SEM.

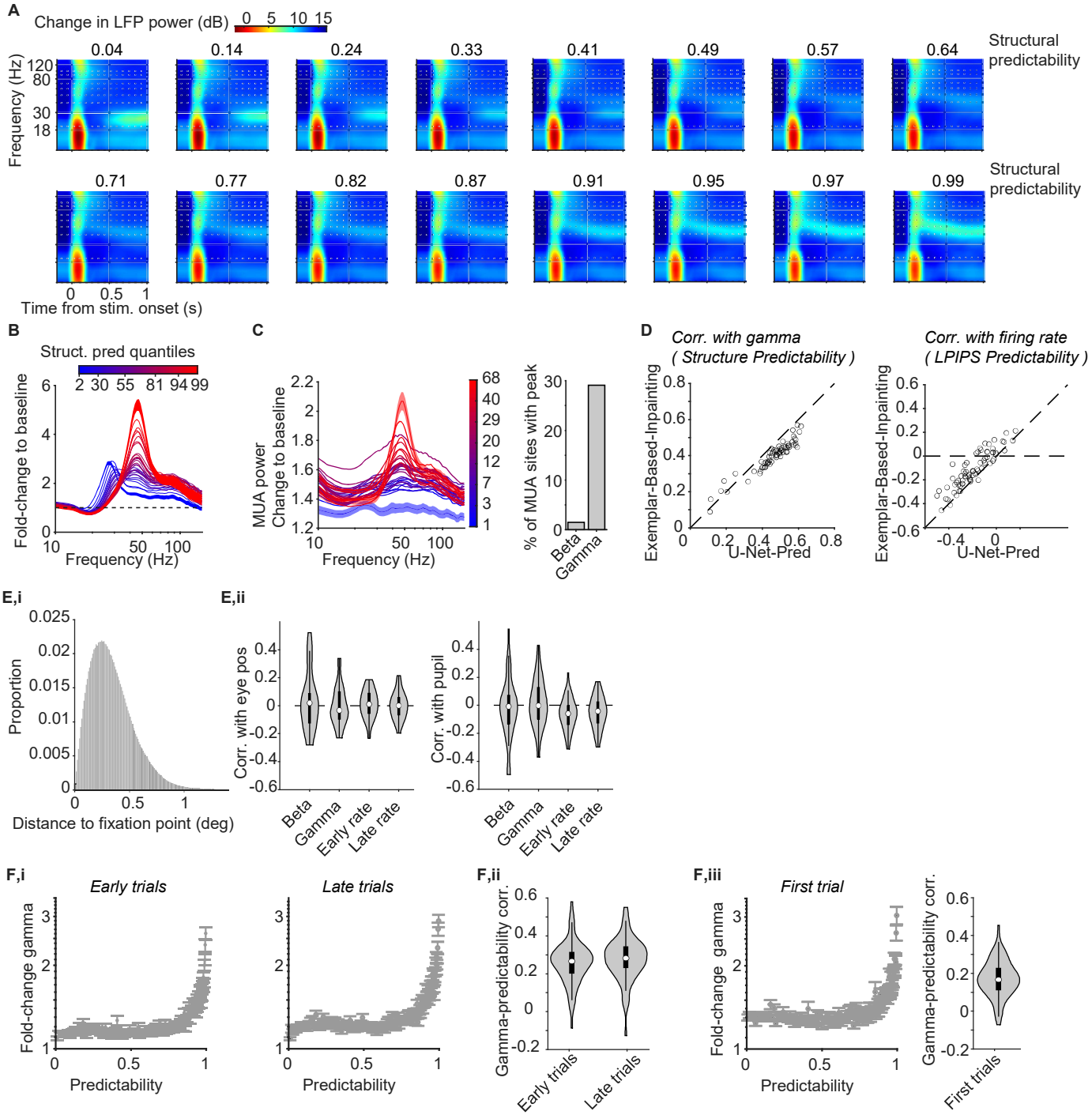

Figure S2: **Relationship of neural activity with structural predictability. Relates to Figure 2 of Main text.** **(A)** Time-frequency representations for 16 different levels of predictability (similar to Figure 2A). Time-frequency representations were computed using discrete prolate spheroidal sequences (multi-tapers for +/-5 Hz smoothing) with a 300ms time window sliding in steps of 50ms. **(B)** Similar to Figure 2B-left, but now shown average baseline-corrected LFP power spectra. SEMs are shown for the lowest and highest quantile of predictability. **(C)** MUA power as a function of the product of structural predictability and luminance-contrast, as shown in Figure 4C, for monkey H. Blue-to-red colors indicate quantiles of contrast × predictability. MUA power shows a gradual increase in γ amplitude. Right: Percentage of sites (across channels and recording sessions) with a detectable MUA peak in the beta or the gamma-frequency range, across all three monkeys. Note that many sites show detectable γ peaks, but only very few sites show detectable β peaks in the MUA. **(D)** Comparison between predictability derived via deep learning methods and exemplar-based inpainting. Instead of making image predictions with a deep neural network with U-net architecture, we generated predictions using exemplar-based inpainting (Criminisi et al., 2004). Similar to the main analyses, we then computed predictability either as structural predictability or LPIPS-predictability. Predictability derived via exemplar-based inpainting correlated more poorly with γ for the structural predictability, but also shows a clear positive correlation with γ across channels. For firing rates, LPIPS correlations are stronger with deep learning methods (U-Net) than with exemplar-based inpainting. **(E,i)** Histogram of absolute eye-movement deviations across all time-points, trials and monkeys, showing that eye movements were restricted in a relatively narrow range. **(E,ii)** Left: Correlations of average eye-movement deviation with β, γ, early and late firing rates. Average eye-movement deviation was computed for each image separately. Correlations were non-significant for all conditions. Right: Correlations of average pupil diameter with β, γ, early and late firing rates. Average pupil diameter was computed for each image separately. Correlations were non-significant for β and γ but significantly negative for early and late firing rates ($P < 0.05$, T-test). **(F,i)** γ as a function of structural predictability for early and late trials (first half and last half of trials in a session). Similar to Figure 2C. **(F,ii)** Correlations of γ with predictability for early and late trials. Similar to Figure 2D. **(F,iii)** Left: γ as a function of predictability for the first trial in the session. Right: Correlation of γ with predictability. Data in panels B,C,F are represented as mean ± SEM.

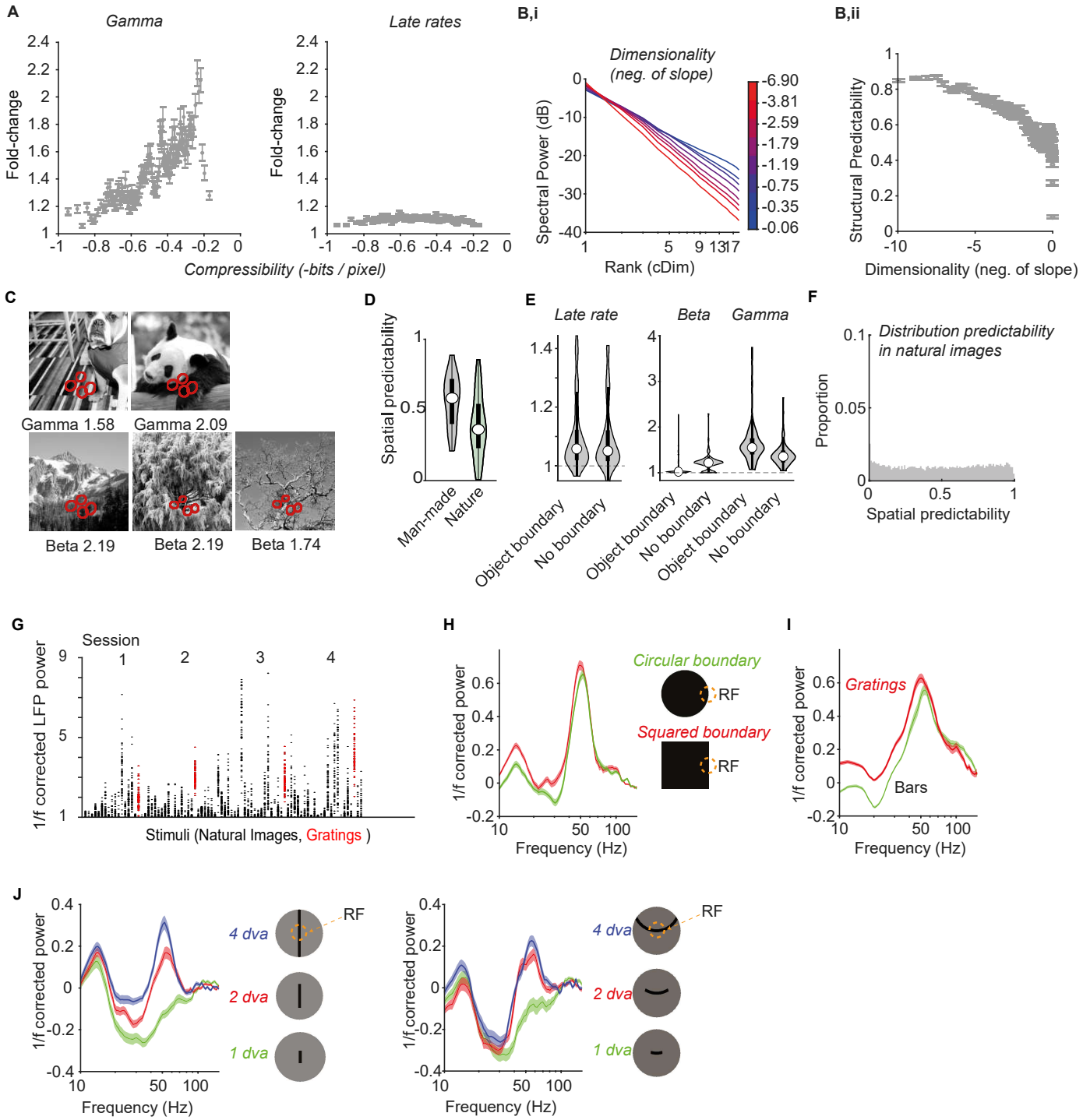

Figure S3: **Relationship of neural activity with compressibility, dimensionality and image categories. Relates to Figure 3 of Main text.** **(A)** γ-synchronization and firing rates across bins of compressibility. **(B,i)** Dimensionality was determined by first taking the two-dimensional fast Fourier transform of the RF image-patch, taking the rotational average, and ranking the spectral components by magnitude. Dimensionality was then defined as the slope of the resulting spectrum. An image with low dimensionality can be represented only by a few spatial frequency components. **(B,ii)** Images with low dimensionality had high spatial predictability (R across bins = -0.91, $P < 0.001$). **(C)** Example images with β and γ values (both as fold-change) for a set of RFs. **(D)** Average structural predictability for the random images category, separate for nature and man-made contents. Shown are the median, 25/75% percentiles and data distribution across images. Man-made images have higher spatial predictability (T-Test, $P < 0.001$). **(E)** Fold-changes in neural activity for image patches that contain object boundaries versus image patches that do not. Comparisons were significant for β ($P < 0.001$) and γ ($P < 0.001$), but not for early and late rates ($P = 0.94$ and $P = 0.73$, T-test). **(F)** Histograms of spatial (structural) predictability for natural images that that fulfilled the criteria on global luminance contrast (> 0.2), luminance (between 40 and 200) and spatial frequency (average centroid greater than 0.5 degrees of visual angle). **(G)** Gamma-band responses for 4 sessions that included a stationary grating stimulus at 2c/d spatial frequency, presented in vertical or horizontal orientation. Each dot represents a trial. Stationary gratings yield strong gamma, however several natural stimuli in those sessions induced stronger gamma-band activity. **(H-J)** contain separate experiments with artificial stimuli ran in separate sessions: **(H)** Boundary stimuli generate strong γ responses, both for curves and square filled surface stimuli (average across 14 channels, 1 session). **(I)** Bar stimuli generate comparable γ response to gratings. Square-wave grating stimuli presented were 6 degrees in diameter with a spatial frequency of 0.5,1,2,4 c/d. Bar stimuli were matched in width to the respective spatial frequency. Average across 28 channels from monkey H and monkey A. **(J)** Curved bar (1 session, average across 14 channels) and straight bar stimuli (4 sessions, average across 53 channels) generate γ responses with characteristic size tuning beyond 1dva. Data in panels A,B,H-J are represented as mean ± SEM.

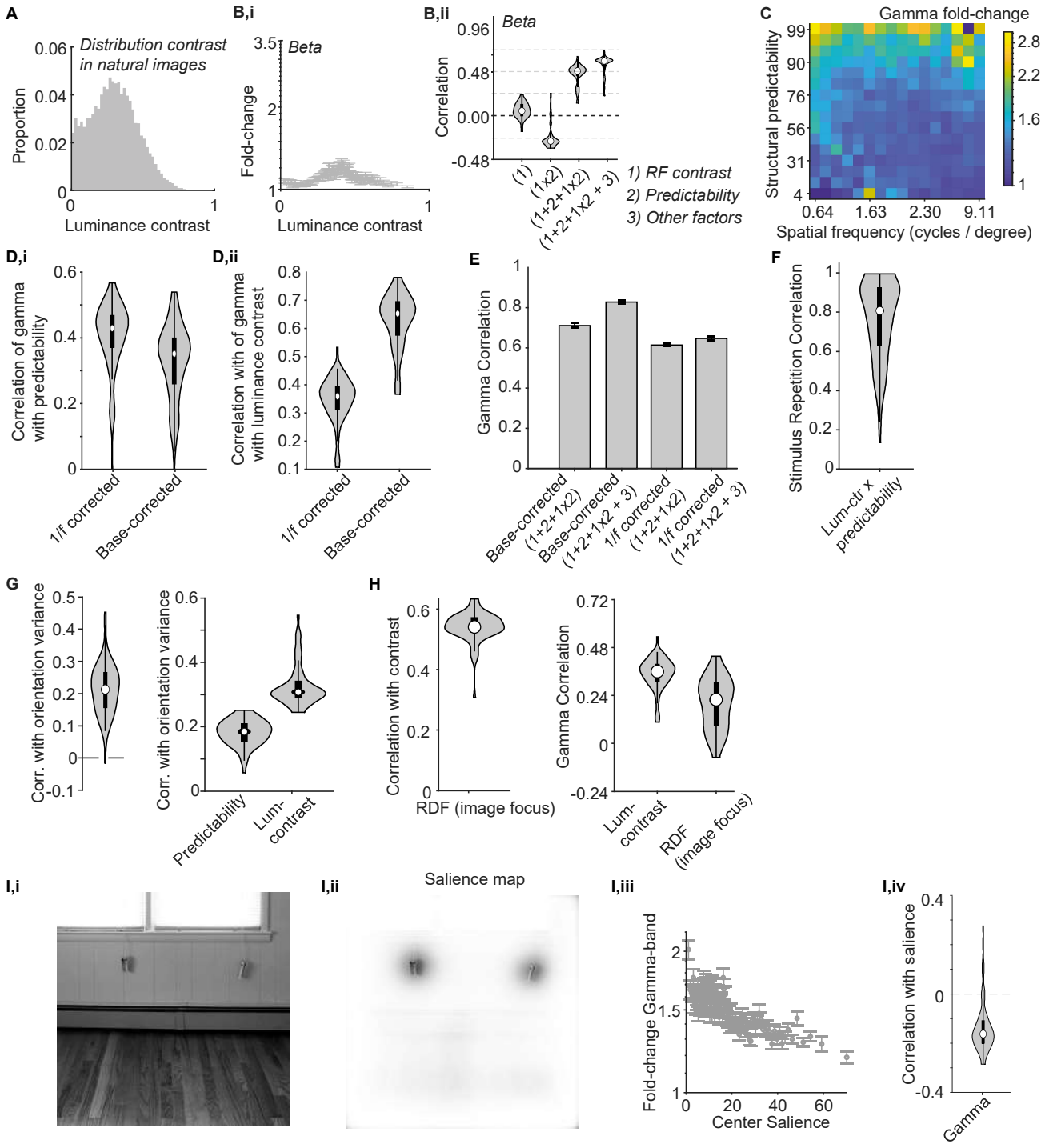

Figure S4: **Relationship of neural activity with luminance contrast and salience. Relates to Figure 4 of Main text.** **(A)** Histograms of luminance contrast (RMS, root-mean-square, see Methods) for images that fulfilled the criteria on global luminance contrast (> 0.2), luminance (between 40 and 200) and spatial frequency (average centroid greater than 0.5 degrees of visual angle). **(B,i)** Average 1/f corrected β-peak amplitude for different levels of luminance contrast, similar to Figure 4B. β power peaks at intermediate luminance contrasts. **(B,ii)** Similar to Figure 4C. Beta was very weakly correlated with luminance contrast and strongest for intermediate contrasts. Including other image factors as predictors led to a small increase in explained variance for β synchronization. **(C)** Gamma-band activity (fold-change) as a function of different values of spatial frequency in the RF and structural predictability. Predictability increases gamma across a wide range of spatial frequency values **(D)** Correlation of γ derived either with $1/f$ correction or as the baseline-corrected power. Predictability correlates more strongly with the $1/f$ derived power (D,i), whereas contrast correlates much more strongly with the baseline-corrected power (D,ii) ($P < 0.001$ for both) which most likely reflects the influence of firing rates on baseline-corrected power. **(E)** Prediction of gamma power, using either baseline-corrected power and 1/f corrected power. Shown are prediction models with the same definition as in panel B-ii and Figure 4C. For this figure panel, only sites with an SNR value above the median are used, proposed as a quality metric of neural recordings by (Pospisil and Bair, 2021). Baseline-corrected power generally shows higher correlations, however to avoid an influence of firing rates we focused on 1/f corrected power in the main text. **(F)** Correlation of luminance-contrast x predictability interaction term across repeats of the same stimulus set in a different session (avg+/-sem = 0.76+/0.01, n=5 repeat sessions). This was computed by (i) computing the (contrast x pred) interaction term for each trial based on the eye position and the estimated RF position in the image, (ii) averaging the interaction terms across trials, and (iii) correlating the interaction term across the images in between the two sessions. This was done for all channels separately. Note that we did not factor in variability due to RF estimation, or training of the predictive neural network. **(G)** Correlation of the orientation variance measure of (Hermes et al., 2019) with γ, predictability and luminance contrast. **(H)** Left: Spearman correlation between image focus (related to gamma according to Brunet and Fries (2019)) and luminance contrast. Right: Correlation between γ and luminance contrast or focus. Correlation of gamma with luminance contrast was significantly higher than for image focus (paired T-test, $p < 0.001$). **(I)** Salience analyses. Salience was extracted using state-of-the-art deep neural networks for salience that have been trained on human eye movement data and validated on macaque eye movements (I,i and I,ii for an example image). We observed a negative correlation between the salience at the RF location and γ at the RF location across images (I,iii and I,iv). Data in panels B,E,I are represented as mean ± SEM.

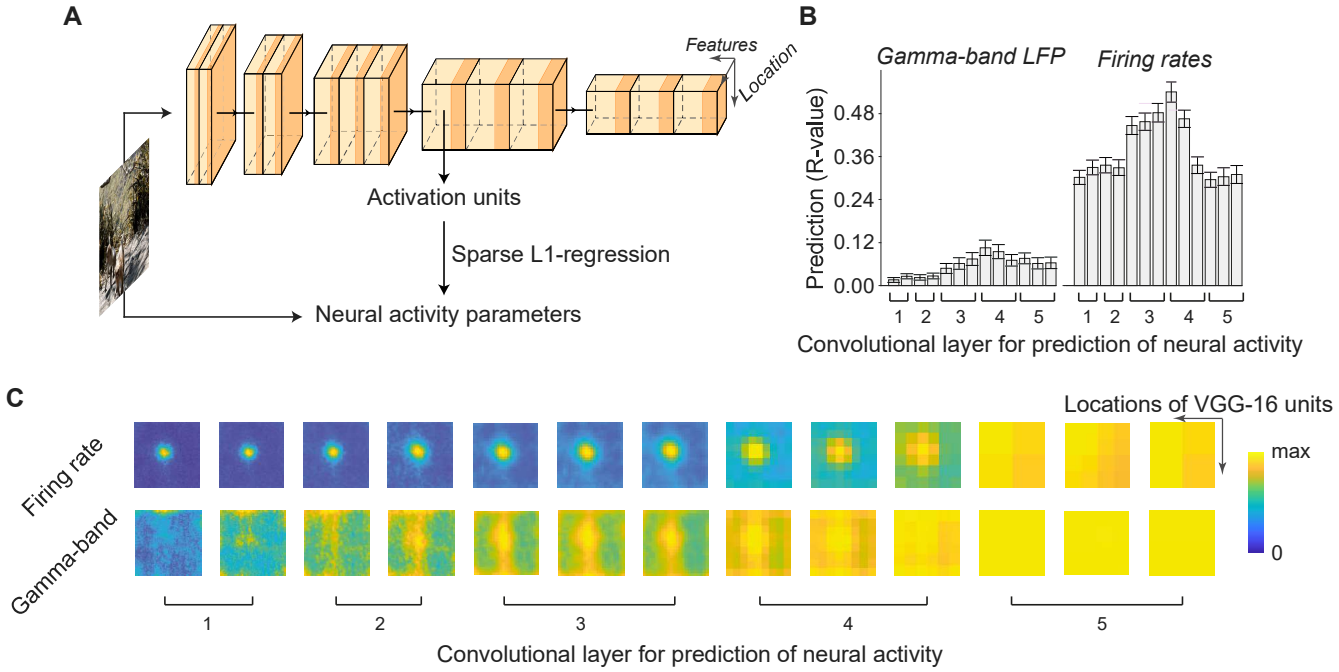

Figure S5: **A feedforward neural network for object recognition explains firing rates relatively well, but poorly accounts for γ-synchronization. Relates to Figure 5 of Main text. (A)** For each recording site, we determined different neural activity parameters. The image patch centered on the RF of the recording site was then passed into the CNN for object recognition (OR-CNN; in this case the VGG-16), and we computed the activation of every OR-CNN artificial neuron (AN) whose RF overlapped with the recording site. Sparse L1-regression with cross-validation was used to predict neural activity from OR-CNN ANs with RFs at the center of the image. **(B)** Regression prediction accuracy of different neural activity parameters depending on OR-CNN layer. Regression prediction accuracy for late (200-600 ms) firing rates was significantly higher for middle (5-9) than early (1-4) and deep (10-13) convolutional layers ($P < 0.001$, paired T-test). For γ, regression prediction accuracy was significantly higher for middle ($P < 0.001$) and deep ($P < 0.05$) than early layers. Data are represented as mean ± SEM. For early rates and β see Figure 5A-B. **(C)** Prediction accuracy depending on the RF location of OR-CNN ANs in the image. In this case, we predicted neural activity from all units in a 3x3 image using sparse L1-regression. Shown are the prediction weights, which reveal circular RFs for firing rates already in the earliest layer.

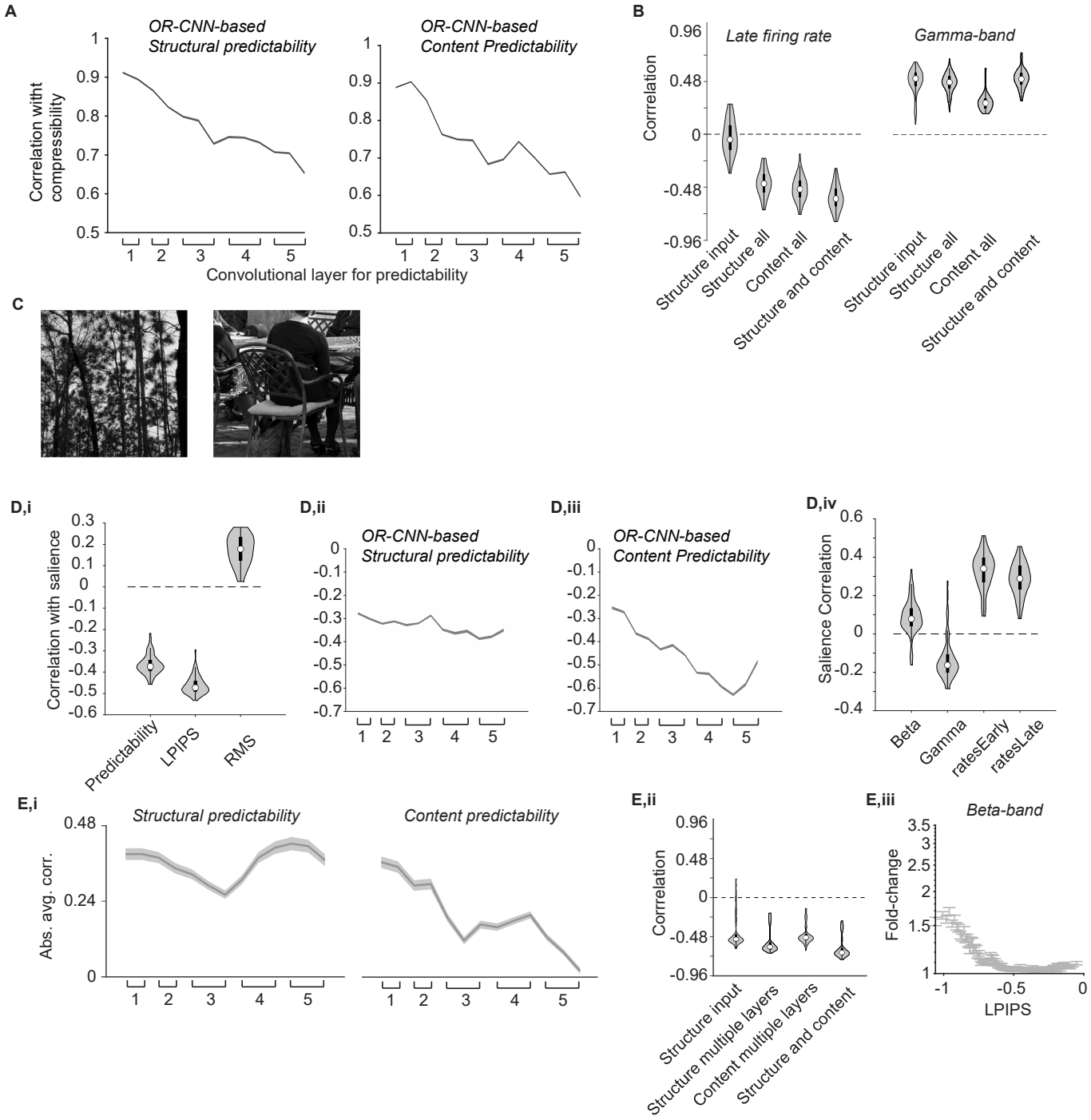

Figure S6: **Relationship of neural activity, perceptual similarity and salience with low- and high-level predictability. Relates to Figure 6 of Main text.** **(A)** Structural predictability and content predictability defined across different layers (Figure 6) vs. compressibility. Both correlations were negative (R=-0.97, R = -0.92, $P < 0.001$). **(B)** Left to right: (1) Correlation with structural predictability at the input image level; (2) regression model based on structural predictability across layers 1conv2, 2conv2, 3conv3, 4conv3, 5conv3; (3) regression based on content predictability in those layers; (4) regression based on both content and structure predictability in those 5 layers. For γ, the full regression model did not explain more variance than structural predictability at the input level ($P = 0.40$), and structure explained more variance than content ($P < 0.001$, paired T-test). For rates, the model with content explained more variance than structure ($P < 0.001$), and the model with both content and structure explained more variance than either content or structure ($P < 0.001$ for both). **(C)** Example of an image with relatively weak (i.e. lower than median) low-level predictability but relatively strong (i.e. higher than median) high-level predictability on the left (content loss conv1,1 = 0.63, content loss conv5,1 = 0.05). On the right example with weak high-level predictability but relatively strong low-level predictability (content loss conv1,1 = 0.05, content loss conv5,1 = 0.8). **(D)** Salience analyses: **(D,i)** Pearson-r correlation of image salience with structural predictability, LPIPS-predictability and luminance contrast across recording sites (with a minimum RMS contrast of 0.1). **(D,ii-D,iii)** Correlation of OR-CNN-based structural predictability and OR-CNN-based Content predictability across OR-CNN layers. **(D,iv)** Correlation of salience with different neural activity measures. **(E)** Analyses on β related to Figure 6: **(E,i)** Correlation of β-power with structural (left) and content (right) predictability across VGG-16 layers. β showed a significant decrease in absolute correlation with predictability across layers for content ($R = 0.9$, $P < 0.001$). Similar to Figure 6D. **(E,ii)** Left to right: (1) Correlation with structural predictability at the input image level; (2) regression model based on structural predictability across layers 1conv2, 2conv2, 3conv3, 4conv3, 5conv3; (3) regression based on content predictability in those layers; (4) regression based on both content and structure predictability in those 5 layers (similar to panel B). **(E,iii)** Correlation of LPIPS with β synchronization across bins of LPIPS predictability (correlation was significant at $P < 0.001$) (similar to Figure 6C). Data in panels A,D,E are represented as mean ± SEM.

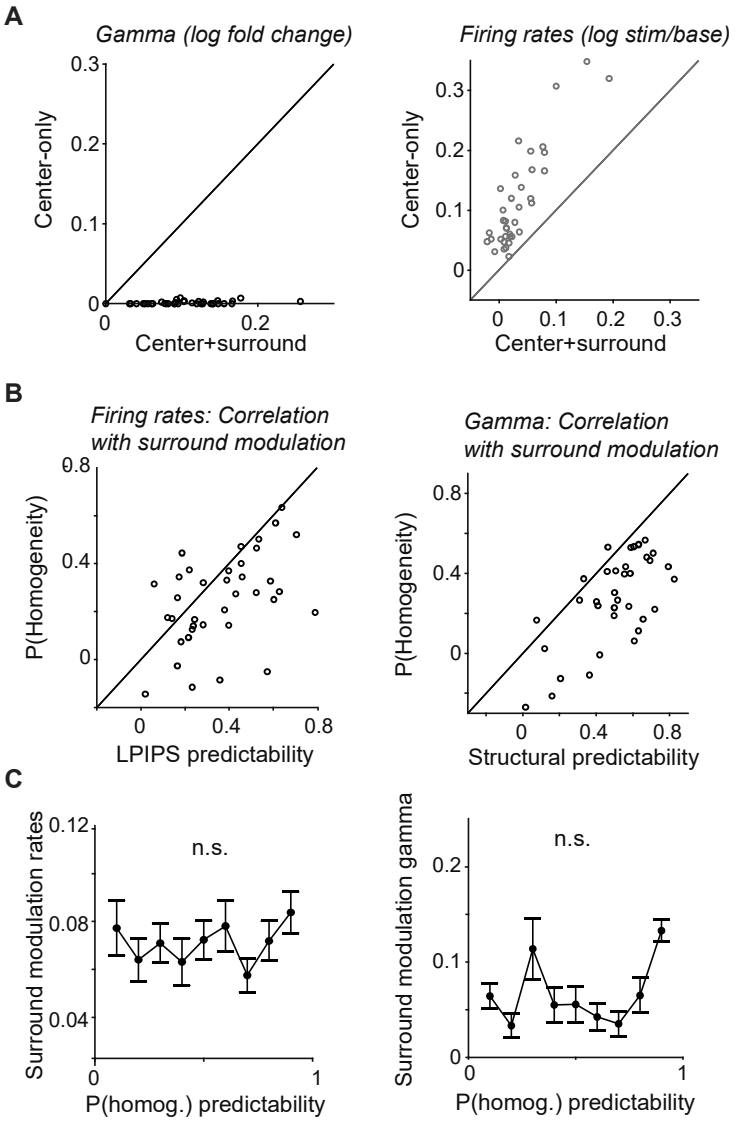

Figure S7: **Prediction of surround modulation via different measures of center-surround predictability and homogeneity. Relates to Figure 7 of Main text. (A)** Left: Comparison of γ-synchronization ($\log_{10}$ fold-change) between the center+surround (full image) vs. center-only (0.5 and 1dva pooled). Right: Surround suppression for the late firing-intensity, and showing $\log_{10}(stim/base)$. Note that suppression for early firing-intensity was significantly weaker than for late firing-intensity (paired T-test, $P < 0.001$). Each circle represents a channel. **(B)** For each image, we computed structural and LPIPS predictability based on the 4×4 dva patch around the center. In addition, we computed P(homogeneity), the center-surround homogeneity measure of (Coen-Cagli et al., 2015), a bimodally distributed, probabilistic measure of center-surround homogeneity based on energy correlations between Gabor-like simple-cell responses (Coen-Cagli et al., 2015). These measures were then correlated with the surround modulation of firing rates and γ-synchronization. Left: Spearman correlation between surround modulation of firing rates and LPIPS-predictability vs. Spearman correlation between surround modulation and P(homogeneity). Each circle corresponds to a channel for which the stimulus was centered on the channel's RF. LPIPS-predictability correlated more strongly with surround suppression than P(homogeneity) ($P < 0.01$, pairwise T-test). Right: Spearman correlation between surround modulation of γ-synchronization and structural predictability vs. Spearman correlation between surround modulation and P(homogeneity). Correlation was stronger for Structural Predictability than P(homogeneity) ($P < 0.001$, pairwise T-test). **(C)** Left: Average surround modulation (center-only minus center+surround) as a function of different levels of P(homogeneity) for firing rates (left) and γ-synchronization (right). Correlation was not significant ($P > 0.59$ and $P = 0.49$). Data are represented as mean $\pm$ SEM.

layer	1c1	1c2	2c1	2c2	3c1	3c2	3c3	4c1	4c2	4c3	5c1	5c2	5c3
pixels	3	5	10	14	24	32	40	60	76	92	132	164	196
dva (A & I)	0.07	0.12	0.25	0.35	0.60	0.80	1.00	1.50	1.90	2.30	3.30	4.10	4.90
dva (H)	0.06	0.10	0.19	0.27	0.46	0.62	0.77	1.15	1.46	1.77	2.54	3.15	3.77

Table S1: **Receptive Field sizes in the VGG-16 across layers. Relates to Figure 5 of Main text.** The abbreviation dva stands for degrees in visual angle. This was shown separately for monkey H and monkey A and I because of the differences in the number of pixels per degree.