Representational Content of Oscillatory Brain Activity during Object Recognition: Contrasting Cortical and Deep Neural Network Hierarchies

Leila Reddy, Radoslaw Martin Cichy, Rufin Vanrullen

eNeuro

Research Article: New Research

Cognition and Behavior

Representational Content of Oscillatory Brain Activity during Object Recognition: Contrasting Cortical and Deep Neural Network Hierarchies

Leila Reddy,[1,2] Radoslaw Martin Cichy,[3] and Rufin VanRullen[1,2]

https://doi.org/10.1523/ENEURO.0362-20.2021

[1]Artificial and Natural Intelligence Toulouse Institute, Université de Toulouse 3, Toulouse 31052, France, [2]Centre National de la Recherche Scientifique, Centre de Recherche Cerveau et Cognition (CerCo), Toulouse 31052, France, and [3]Department of Education and Psychology, Freie Universität Berlin, Berlin 14195, Germany

Abstract

Numerous theories propose a key role for brain oscillations in visual perception. Most of these theories postulate that sensory information is encoded in specific oscillatory components (e.g., power or phase) of specific frequency bands. These theories are often tested with whole-brain recording methods of low spatial resolution (EEG or MEG), or depth recordings that provide a local, incomplete view of the brain. Opportunities to bridge the gap between local neural populations and whole-brain signals are rare. Here, using representational similarity analysis (RSA) in human participants we explore which MEG oscillatory components (power and phase, across various frequency bands) correspond to low or high-level visual object representations, using brain representations from fMRI, or layer-wise representations in seven recent deep neural networks (DNNs), as a template for low/high-level object representations. The results showed that around stimulus onset and offset, most transient oscillatory signals correlated with low-level brain patterns (V1). During stimulus presentation, sustained β ($\sim$20 Hz) and γ ($>$60 Hz) power best correlated with V1, while oscillatory phase components correlated with IT representations. Surprisingly, this pattern of results did not always correspond to low-level or high-level DNN layer activity. In particular, sustained β band oscillatory power reflected high-level DNN layers, suggestive of a feed-back component. These results begin to bridge the gap between whole-brain oscillatory signals and object representations supported by local neuronal activations.

Key words: brain oscillations; deep neural networks; fMRI; MEG; object recognition; representational similarity analysis

Significance Statement

Brain oscillations are thought to play a key role in visual perception. We asked how oscillatory signals relate to visual object representations in localized brain regions, and how these representations evolve over time in terms of their complexity. We used representational similarity analysis (RSA) between MEG oscillations (considering both phase and amplitude) and (1) fMRI signals (to assess local activations along the cortical hierarchy), or (2) feedforward deep neural network (DNN) layers (to probe the complexity of visual representations). Our results reveal a complex picture, with the successive involvement of different oscillatory components (phase, amplitude) in different frequency bands and in different brain regions during visual object recognition.

Received August 19, 2020; accepted April 15, 2021; First published April 26, 2021.
The authors declare no competing financial interests.

Author contributions: L.R., R.M.C., and R.V. designed research; R.M.C. performed research; R.M.C. and R.V. contributed unpublished reagents/analytic tools; L.R. analyzed data; L.R., R.M.C., and R.V. wrote the paper.

Introduction

Oscillatory neuronal activity is thought to underlie a variety of perceptual functions. Different frequency bands can carry information about different stimulus properties (e.g., whether the stimulus consists of coarse or fine object features; Smith et al., 2006; Romei et al., 2011), feedforward or feedback signals (van Kerkoerle et al., 2014; Bastos et al., 2015), or may reflect neuronal communication between different neuronal populations (Fries, 2005; Jensen and Mazaheri, 2010). Other studies have shown that different components of an oscillation (e.g., its power or phase) encode different types of sensory information (Smith et al., 2006).

Although neuronal oscillations are observed in different brain regions, and key theories hold that they reflect processing within, and communication between, brain regions (Fries, 2005; Jensen and Mazaheri, 2010), it has been difficult to pin down how large-scale brain oscillations are related to local patterns of neural activity, and how this relationship unfolds over time. This is because oscillatory activity is often studied with methods such as EEG or MEG, which have low spatial resolution. Although oscillatory signals with high spatial specificity can be recorded via local field potential recordings in humans or animals, these methods usually only target specific brain regions, and thus can only provide a partial view of oscillatory activity and its role in large-scale brain function. A direct link between large-scale oscillations and local neural activity is missing.

Here, we combine large-scale oscillatory signals recorded by MEG with local patterns of neural activity recorded with fMRI to bridge the gap between oscillatory components and different dimensions of object representation in the brain. Using representational similarity analysis (RSA; Kriegeskorte et al., 2008), we investigate the information carried by whole-brain oscillations obtained from MEG, and examine how this information evolves over time during an object recognition task.

We define three distinct dimensions of interest along which neural representations may unfold, and which are often conflated in the literature. First, we use the terms "early" and "late" to denote the temporal evolution of representations. Second, we differentiate between "low-level" and "high-level" stages of a processing hierarchy. Third, we consider the complexity of representations by distinguishing between "low-complexity" and "high-complexity" information (for example, higher-complexity might be characterized by additional nonlinear transformations of the input). In many information processing systems and in many typical experimental situations, these three dimensions are directly related to one another, as input information propagates over time through a succession of hierarchical stages, becoming more and more complex along the way. In such situations, the three dimensions of interest are in fact redundant and need not be further distinguished. But in systems with recurrence and feedback loops (like the brain), time, space, and information complexity are not always linearly related. For example, a lower hierarchical level (e.g., V1) can carry higher-complexity representations, later in time, as a result of feedback loops or lateral connections (Lamme and Roelfsema, 2000). In our terminology, such a representation would be classified as late in time, low-level in the hierarchy, yet high-complexity.

In this work, we consider two main hierarchical systems. We are interested in understanding information processing in the human brain, so we use V1 and IT fMRI brain representations, as done in a number of recent studies (Cichy et al., 2014; Khaligh-Razavi and Kriegeskorte, 2014). Representational similarity between MEG oscillations and this fMRI-based hierarchy can be interpreted in terms of early and late representations (based on the timing of the MEG oscillations), and in terms of low-level (V1) versus high-level (IT) hierarchical stages. To assess the complexity of representations independent of temporal evolution and cortical hierarchy of processing, we related our data to a second class of hierarchical systems: artificial feed-forward deep neural networks (DNNs), as done also in numerous recent studies (Cichy et al., 2016; Bankson et al., 2018; Hebart et al., 2018; Khaligh-Razavi et al., 2018; Kuzovkin et al., 2018). In these artificial networks, the hierarchical level (low-level vs high-level) is directly related to feature complexity (low-complexity vs high-complexity representations), because of the absence of feed-back or recurrent loops: the layer number directly reflects the number of nonlinear input transformations. For any MEG oscillatory signal, representational similarity with DNN activation patterns can thus inform us about representational complexity. In turn, any difference between DNN-based and brain-based RSA may be suggestive of feed-back or recurrent influences in the MEG oscillatory signals.

With this dual approach, we find an intricate picture of transient and sustained oscillatory signals that can be related to V1 and IT representations. Transient oscillatory components around stimulus onset and offset, as well as sustained β ($\sim$20 Hz) and γ ($>$60 Hz) power components resemble V1 representations, while phase-dependent sustained activity correlates best with IT representations. However, when compared with DNNs, some low-level V1-like components actually correlate more with higher DNN layers, suggesting that stimulus representations in primary brain regions may already include high-complexity information, presumably as a result of feedback or top-down influences (Kar et al., 2019; Kietzmann et al., 2019).

This work was supported by the European Research Council Consolidator Grant P-CYCLES 614244, the OSCI-DEEP Agence Nationale de la Recherche Grant ANR-19-NEUC-0004, and the Artificial and Natural Intelligence Toulouse Institute (ANITI) Grant ANR-19-PI3A-0004 (to R.V.); a NVIDIA Graphics Processing Unit (GPU) grant, Agence Nationale de la Recherche Grant Vis-Ex ANR-12-JSH2-0004, and Agence Nationale de la Recherche Grant AI-REPS ANR-18-CE37-0007-01 (to L.R.), and the Emmy Noether Grant CI241/1-1, Deutsche Forschungsgemeinschaft (DFG) grants (CI241/1-1, CI241/3-1) and a European Research Council Starting Grant (ERC-2018-StG) (to R.M.C.).

Correspondence should be addressed to Leila Reddy at leila.reddy@cnrs.fr.

https://doi.org/10.1523/ENEURO.0362-20.2021

In effect, our results narrow the gap between the description of neural dynamics in terms of whole-brain oscillatory signals and local neural activation patterns. Disentangling temporal evolution, hierarchical stage of processing and complexity of representations from each other, our approach allows for a more nuanced view on cortical information flow in human object processing.

Materials and Methods

Experimental paradigm and data acquisition

The data analyzed in this study was obtained from (Cichy et al., 2014), and detailed methods can be obtained from that paper.

Fifteen human subjects of either sex performed separate MEG and fMRI sessions while they viewed a set of 92 images. The image set consisted of human and non-human faces and bodies, and artificial and natural everyday objects. The 92-image stimulus set was taken from the Kiani image set (Kiani et al., 2007), which consists of cutout objects on a gray background (Fig. 1*A*).

In the MEG sessions, each image was presented for 0.5 s followed by an interstimulus interval (ISI) of 1.2 or 1.5 s. Every three to five trials, a target paperclip object was presented, and subjects' task was to press a button and blink whenever they detected this target image. Subjects performed two MEG sessions, of 2 h each. In each session they performed between 10 and 15 runs. Each image was presented twice in each run, in random order.

In each of two fMRI sessions, each image was presented for 0.5 s followed by an ISI of 2.5 or 5.5 s. Subjects' task in the fMRI sessions was to press a button when they detected a color change in the fixation cross on 30 null trials, when no image was presented. Each image was presented once in each fMRI run, and subjects performed 10–14 runs in each session.

The MEG data were acquired from 306 channels (204 planar gradiometers, 102 magnetometers, Elekta Neuromag TRIUX, Elekta) at the Massachusetts Institute of Technology. The MRI experiment was conducted on a 3T Trio scanner (Siemens), with a 32-channel head coil. The structural images were acquired using a T1-weighted sequence (192 sagittal slices, FOV = 256 mm^2, TR = 1900 ms, TE = 2.52 ms, flip angle = 9°). For the fMRI runs, 192 images were acquired for each participant (gradient-echo EPI sequence: TR = 2000 ms, TE = 32 ms, flip angle = 80°, FOV read = 192 mm, FOV phase = 100%, ascending acquisition gap = 10%, resolution = 2 mm, slices = 25).

MEG analysis, preprocessing

MEG trials were extracted with a 600-ms baseline before stimulus onset until 1200 ms after stimulus onset. A total of 20–30 trials were obtained for each stimulus condition, session, and participant. Each image was considered as a different stimulus condition.

Data were analyzed using custom scripts in MATLAB (MathWorks) and FieldTrip (Oostenveld et al., 2011). Data were downsampled offline to 500 Hz. For each trial and sensor, we computed the complex time-frequency (TF) decomposition using multitapers. Parameters used were: 50 distinct frequencies increasing logarithmically from 3 to 100 Hz, over a time interval of –600 to 700 ms with respect to stimulus onset, in steps of 20 ms. The length of the sliding time window was chosen such that there were two full cycles per time window. The amount of smoothing increased with frequency (0.4 * frequency).

From the complex number at each TF coordinate, we extracted two measures for each sensor and each stimulus condition: the power and the phase of the oscillation. For each channel and stimulus condition, on each trial, the power was first expressed in decibels, and then averaged across trials to obtain one power value per stimulus condition. The phase of the oscillation was obtained by first normalizing each trial to make each trial's vector in the complex domain of unit length, and then averaging across trials for each stimulus condition. The resultant average vector was then normalized to unit length, and the sine (real) and cosine (imaginary) components were extracted for each stimulus condition and each sensor.

MEG analysis, multivariate analysis (Fig. 1*B*)

At each TF coordinate and for each stimulus condition, we next arranged the 306 power values from the 306 MEG sensors into a 306-dimensional vector representing the power pattern vector for that stimulus condition. Similarly, at each TF coordinate and for each stimulus condition we concatenated the 306 sine and 306 cosine values into a 612-dimensional phase pattern vector for that stimulus condition.

We next computed two representational dissimilarity matrices (RDMs): one for power and one for phase, at each TF point. For each pair of stimulus conditions, the power (phase) pattern vectors were correlated using the Pearson correlation measure, and the resulting 1-correlation value was assigned to a 92×92 power (phase) RDM, in which the rows and columns corresponded to the images being compared. This matrix is symmetric across the diagonal. This procedure results in one power (phase) RDM at each TF point.

fMRI analysis (Fig. 1*C*)

The preprocessing steps for the fMRI data are described in detail in Cichy et al. (2014). For the multivariate analysis, two regions of interest (ROIs) were defined: V1 and IT. In each subject, for each ROI, voxel activation values were extracted for each stimulus condition, and the resulting values were arranged in a pattern vector for each stimulus condition. Then, in each ROI, for each pair of stimulus conditions, the corresponding pattern vectors were correlated using the Pearson correlation measure, and the resulting 1-correlation value was assigned to the 92×92 fMRI RDM. For further analysis, the fMRI RDMs were averaged across the 15 subjects, resulting in one RDM per ROI. The fMRI RDMs were provided by R. Cichy, D. Pantazis, and A. Oliva (Cichy et al., 2014).

MEG-fMRI RSA (Fig. 1*D*)

RSA consists in comparing two (or more) RDMs. RSA between the MEG and fMRI RDMs was performed by computing the partial Pearson's correlation between each

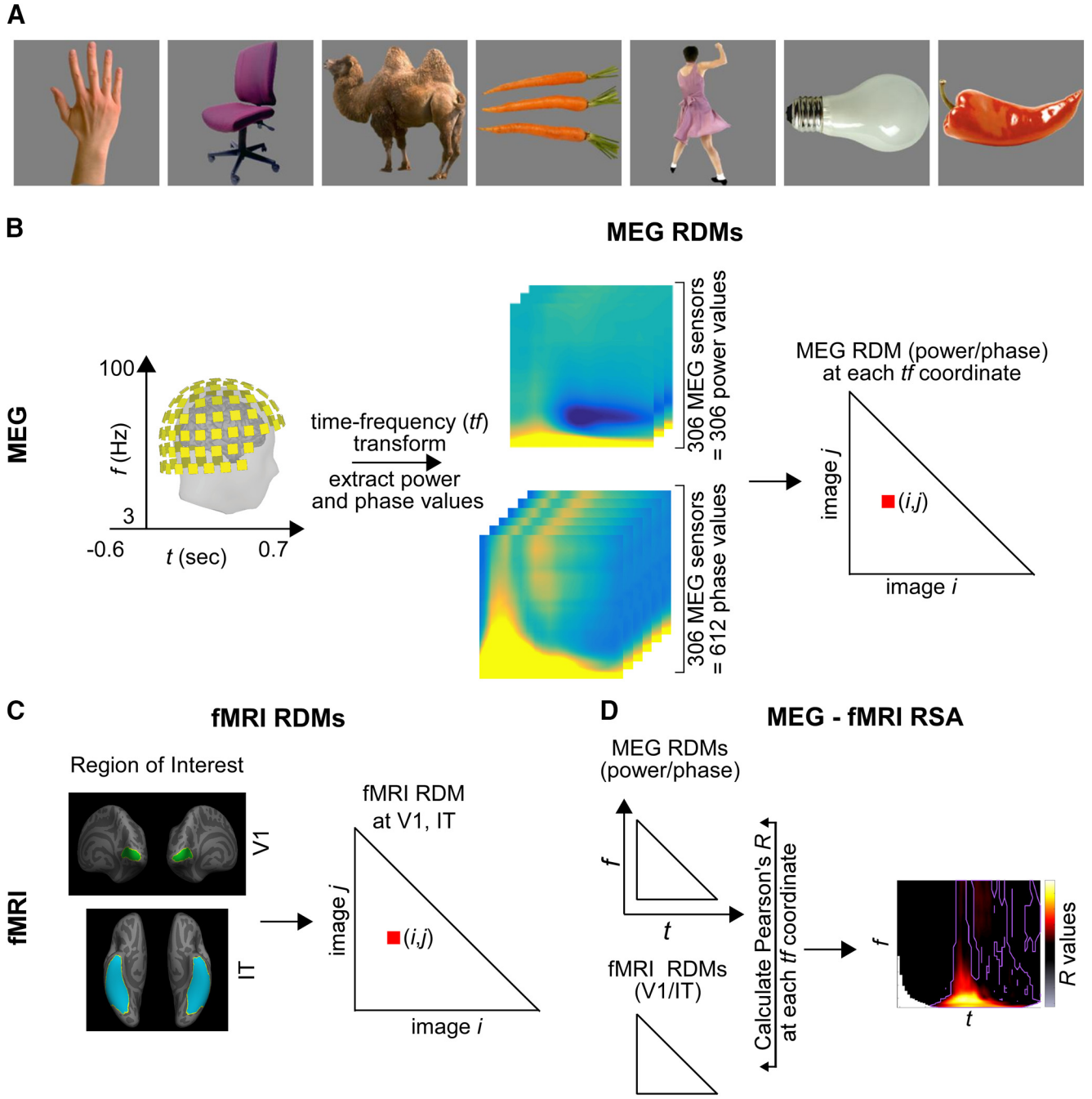

Figure 1. MEG-fMRI RSA analysis. ***A***, Examples from our 92-image set (***B***) MEG analysis and MEG RDMs. From the MEG signals, the complex TF transform was computed for each of the 306 MEG sensors. The amplitude and phase (separated into cosine and sine) values were extracted from the complex number at each TF coordinate, and a MEG RDM was constructed, reflecting the distance between oscillatory activation patterns for every pair of images (*i,j*; for details, see Materials and Methods). As a result, we obtained a power and phase MEG RDM at each TF coordinate for each participant. ***C***, fMRI RDMs were obtained from (Cichy et al., 2014). Two ROIs were defined: V1 and IT and one fMRI RDM was obtained for each ROI, and each participant, reflecting the distance between BOLD activation patterns for every pair of images (*i,j*). ***D***, RSA consists in comparing two (or more) RDMs. The MEG power or phase RDMs were compared with the fMRI RDMs (V1 or IT) by computing the partial Pearson's *R*. This step was performed at each TF coordinate, resulting in an RSA map of R values at each TF coordinate, for each subject and ROI.

MEG (phase or power) RDM with each fMRI RDM (V1 or IT), while partialling out any contribution from the other fMRI RDM (IT or V1). We chose to perform a partial correlation because the V1 and IT RDMs were positively correlated with each other ($r \sim 0.3$); compared with a standard correlation, the partial correlation allowed us to isolate the unique correlation of each fMRI RDM with the MEG RDM, while discarding their joint contribution.

This procedure resulted in four RSA maps per subject (power/phase MEG RDMs $\times$ V1/IT fMRI RDMs). Each

RSA map shows the R value between the MEG signals and the V1/IT activation patterns at each TF point (Figure 4). Significance of the RSA result was evaluated with a paired t test against 0, FDR corrected, $\alpha = 0.05$. The time-course of these effects is shown in Figure 5.

Clustering analysis

The MEG TF RDMs are heavily correlated with each other. To facilitate the interpretation of the information content of oscillatory signals, and to determine which features co-vary and which are independent, K-means clustering was performed on the MEG power and phase RDMs (Figure 2). Clustering was performed on the 4186-dimensional [(92 × 92–92)/2] RDMs across all (66) time points and (46) frequency points, combining the power and phase signals (resulting in 46 × 66 × 2 = 6072 data points to cluster in a 4186-dimensional space). K-means was implemented with the MATLAB function kmeans, with the correlational distance measure, five replicates, and the number of clusters going from 1 to 20. The optimal number of clusters was then determined with the elbow criterion defined as the point just before the local maximum of the second derivative of the residual sum of squares (corresponding to the point at which adding another cluster would only provide a marginal gain in variance explained). With this method, the first elbow occurred at $k = 7$ clusters.

The chosen clusters could be visualized by plotting the correlation distance (in the 4186-dimensional RDM space) between the cluster's centroid and every TF point, for both power and phase signals, resulting in two TF maps of "distance to cluster centroid" for each cluster (Figs. 6, 7).

RSA (using partial Pearson's correlation) was performed between each cluster's centroid and each of the fMRI RDMs (see below). For RSA with fMRI, this procedure resulted in two RSA values R_{V1} and R_{IT} (one each for V1 and IT). Since each cluster centroid could correspond to both V1 and IT to different degrees, the "cortical level" L of the cluster was positioned somewhere between V1/low-level and IT/high-level using the following equation:

$$L = \sigma\left({}^{(R_{IT}-R_{V1})}\!/\!_{(R_{IT}+R_{V1})}\right), \quad (1)$$

where σ denotes the sigmoid function. The measure L could vary between 0 (when the cluster's representational content was perfectly similar to V1) and 1 (when it was perfectly similar to IT).

Significance of RSA between the cluster centroids and the fMRI RDMs was computed with a surrogate test. On each iteration, the cluster centroid RDM was randomly shuffled and the partial correlation was computed between this shuffled RDM and the true RDM. This procedure was repeated for 10^5 iterations, and the proportion of iterations on which the shuffled RSA values were higher than the true RSA values was counted.

DNN RDMs

The MEG phase/power representations were also compared with representations in seven DNNs (so as to ensure that conclusions were not dependent on one specific network architecture): AlexNet (Krizhevksy et al., 2012), VGG16 (Simonyan and Zisserman, 2015), GoogleNet (Szegedy et al., 2015), InceptionV3 (Szegedy et al., 2017), ResNet50 (He et al., 2016), DenseNet121 (Huang et al., 2017), and EfficientNetB3 (Tan and Le, 2019), processing the same 92 images as in our MEG and fMRI data. However, in contrast to our 92-image stimulus set, which consisted of cutout objects on a gray background, the DNNs had been trained on images from ImageNet (millions of photographs with one or more objects in natural backgrounds). The networks had thus learned optimal representations for their training set, but in this representation space our 92 images tended to cluster into a remote "corner" (Fig. 3), with low dissimilarity (1-Pearson's R) values between images, and a resulting RDM of poor quality. To retrieve meaningful distances between the representations of the 92 images, we first performed a centering procedure: we centered the activation of each layer of each DNN by subtracting the mean activation of an independent set of 368 images from the Kiani image set. This independent image set consisted of four images from each of the categories in our 92-image set. Importantly, because the image set used for centering did not include any of the 92 images from our study, there was no circularity in the centering operation, nor any leakage of information between the representations of our 92 images. This centering procedure contributed to minimize the potential problems arising from differences between our dataset and the standard ImageNet dataset, but it did not completely alleviate these differences, as can be seen in Figure 3*C*.

RDMs were constructed for several convolutional layers of each network based on the layer activation values. There were five layers for AlexNet, 13 for VGG16, 12 for GoogleNet, 16 for InceptionV3, 17 for ResNet50 (hereafter referred to as ResNet), 14 for DenseNet121 (hereafter DenseNet), and eight for EfficientNetB3 (hereafter EfficientNet). These layers were chosen so as to span the entire network hierarchy, without making the analysis computationally intractable (as some networks can contain >200 layers to choose from). RSA was then performed (with the Spearman correlation) between these RDMs and the centroid of each cluster (see above for details of the clustering analysis). The layer with maximum RSA, normalized by the number of layers for this DNN, was taken to reflect the information content of this cluster between 0 (in the terminology defined in the Introduction, low-complexity corresponding to the DNN's first layer) and 1 (high-complexity, corresponding to the DNN's last layer), and finally averaged across the seven DNNs.

Code accessibility

Custom code can be made available on request.

Results

Fifteen participants viewed the same set of 92 images while fMRI and MEG data were recorded (in separate sessions). The image set consisted of human and non-human bodies and faces, and artificial and natural stimuli. Each

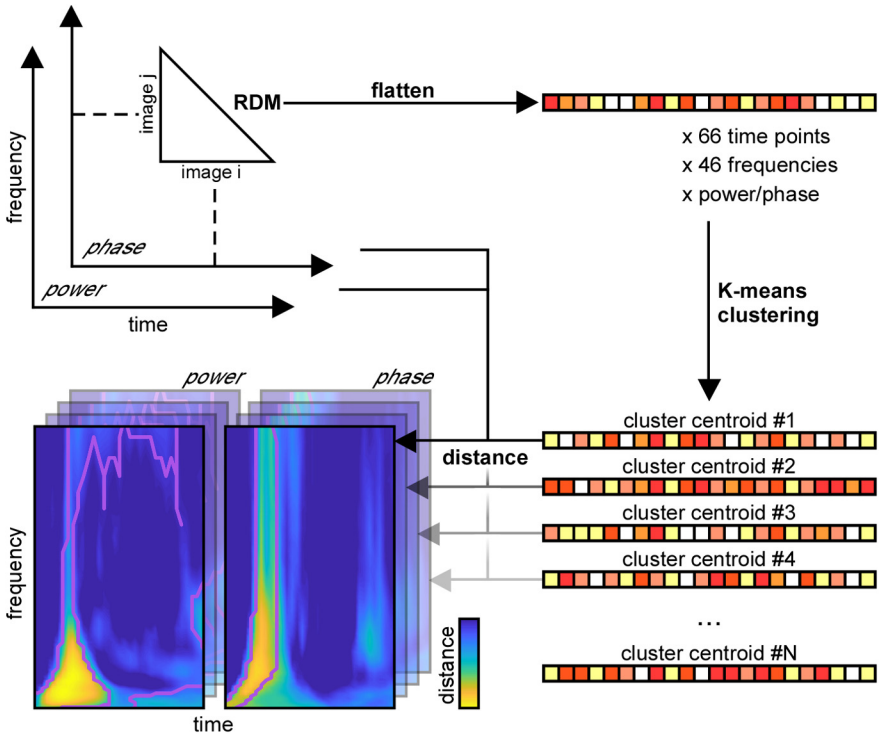

Figure 2. K-means clustering analysis procedure. Starting from the MEG RDM representations (top left, as described in Fig. 1*A*), we flatten each RDM data point into a vector. The entire set of vectors (across all TF coordinates and power/phase conditions) is entered into a K-means clustering algorithm (right), resulting in N clusters and their centroids. By measuring the distance of these centroids to all initial RDM data points, we obtain TF maps of "distance to centroid" (bottom left) that capture the main TF components (across both power and phase) of each cluster.

stimulus was presented for 0.5 s, followed by a 1.2- or 1.5-s baseline period.

To assess oscillatory components, we extracted stimulus-related activity from –600 to 1200 ms relative to stimulus onset from the MEG data. For each trial, and each sensor a TF decomposition was performed, and a power and phase value extracted at each time and frequency point. These values were used to compute RDMs at each TF point, separately for power and phase (Materials and Methods; Fig. 1*A*). Each element in the MEG RDMs indicates how distinct the corresponding images are in the MEG power or phase spaces, and the entire MEG RDM is a summary of how the 92-image stimulus set is represented in the MEG oscillatory power or phase at each TF point.

To assess local patterns of neural activity, we generated fMRI RDMs by performing comparisons between the local BOLD activation patterns of pairs of images in V1 and IT (Cichy et al., 2014). Two fMRI RDMs were obtained (Fig. 1*B*), one for V1 and one for IT. The fMRI RDMs are a measure of the representation of the image set in the voxel space of V1 and IT local neural activity.

Bridging the space, time, and frequency gap in object recognition

How similar is the oscillatory representation of the images to their representation in each brain region? The MEG RDMs (power and/or phase) at each TF point represent the stimulus set in a large-scale brain oscillatory activity space, while the fMRI RDMs represent the same image set via BOLD activity in a local population of neurons in two brain regions (V1 or IT). We evaluated the similarity of representations in the TF domain with those in the fMRI activation patterns by computing the partial Pearson's correlation between the MEG RDMs (phase or power) with the fMRI RDMs (V1/IT), at each TF point (Fig. 1*C*). This analysis resulted in four TF maps of R values (or RSA maps), which provide the unique correspondence between whole-brain oscillations and local patterns of neural activity in V1 and IT, at each TF point (Fig. 4). With these maps we can ask whether and when stimulus information contained in oscillatory phase or power at each TF point resembles BOLD activations in a given brain region (V1/IT), and potentially, which region it resembles more. The advantage of measuring partial correlation (instead of a standard correlation) is to discard the (potentially large) portion of the variance in oscillatory representations that is explained equally well by V1 or IT BOLD representations, owing to the fact that V1 and IT signals already share similarities. This way, we concentrate on the part of oscillatory representations that is uniquely explained by each brain ROI.

Our results show that different oscillatory components map to different brain regions at different moments in time. Overall, the absolute maximum of representational similarity with brain area V1 occurred in the α band $\sim$120 ms after stimulus for oscillatory power, whereas the

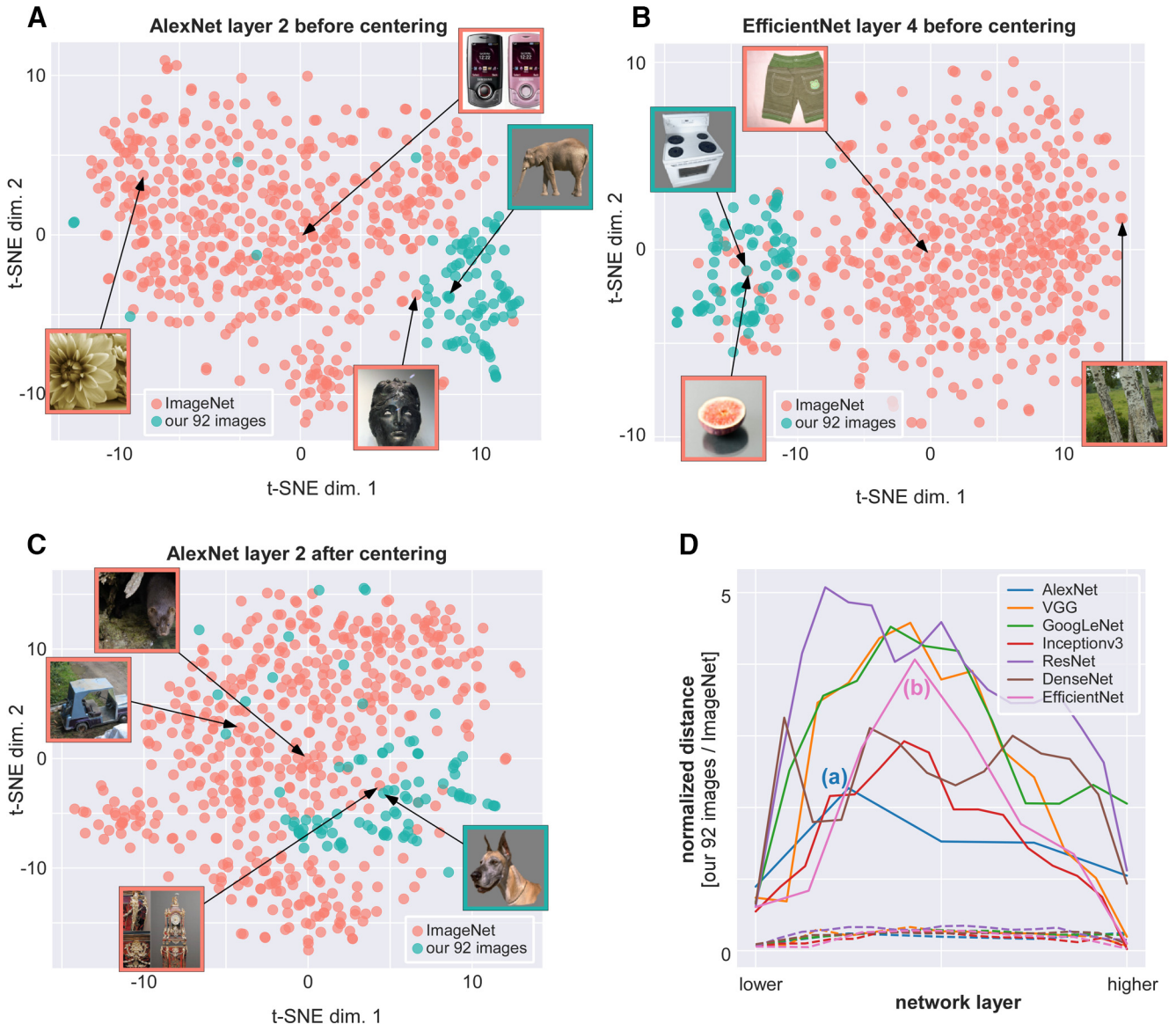

Figure 3. ***A***, ***B***, t-distributed stochastic neighbor embedding (t-SNE) visualizations of 500 ImageNet samples and the 92-image stimulus-set used in this study, across representative layers of two networks (***A***, AlexNet; ***B***, EfficientNet). To obtain these visualizations, the feature values of all images were subjected to a principal component analysis (PCA) of which only the first 100 dimensions were retained (so as to limit computational demands); then, t-SNE was applied, as implemented in the scikit-learn Python library, with parameters: [perplexity = 30, n_iter = 1000, learning_rate = 1.0, min_grad_norm = 0]. The DNNs used in this study had been trained on images from ImageNet, which consists of millions of photographs of one or more objects in natural backgrounds. In contrast, our 92-image stimulus set consists of cut-out images on a gray background. The DNNs learn optimal representations for the training images from ImageNet, i.e., different images from different categories are mapped to different regions of the representation space, and the whole space tends to be equally occupied by the training samples. However, as the t-SNE visualizations show, our 92 images are all projected into a remote corner of this space, meaning that the RDM distances between the 92 images are confounded by the mean vector (the pairwise Pearson distance depends more on the alignment with the mean vector, and less on the true physical distance between points). Inset images show the most stereotypical image of our 92-stimulus set (highlighted in green), the closest image from the ImageNet set (characterized, as expected, by an empty gray background), as well as one ImageNet sample near the space origin, and one on the opposite side of the feature space. To circumvent this problem, we used a re-centering approach as described in Materials and Methods. ***C***, The same layer of AlexNet as shown in ***A***, after re-centering. The 92-stimulus set is now closer to the center of the feature space. ***D***, Systematic measurement of the distance between the centroid of our 92-stimulus set and the space origin (normalized by the SD across our 92 images), for each layer of each DNN. The two DNN layers depicted in ***A***, ***B*** are labeled (a) and (b) on the corresponding curves. As a baseline, the dashed lines reflect the same distance measure, applied to the 500 ImageNet samples.

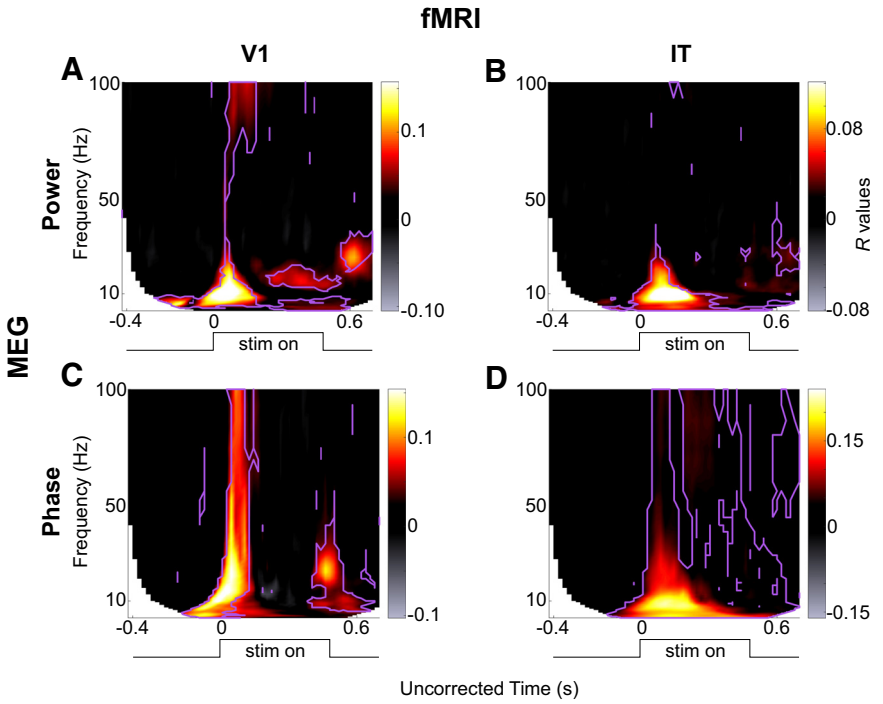

Figure 4. Results of the 2 × 2 RSA comparisons (MEG power/phase × fMRI V1/IT), averaged over all subjects. (***A***) MEG-Power x fMRI V1, (***B***) MEG-Power x fMRI-IT, (***C***) MEG-Phase x fMRI-V1, (***D***) MEG-Phase x fMRI-IT. The purple contours mark those regions in the maps that are significantly different from zero (paired *t* test against 0 across $N = 15$ subjects, FDR correction, $\alpha = 0.05$). Note that the absolute latencies are not directly comparable across frequencies, because of different smoothing windows applied at the different frequencies when performing the TF transform (hence, the *x*-axis is labeled as uncorrected time).

absolute maximum related to area IT occurred for θ and α phase ~200–300 ms. More generally, a strong increase in representational similarity was observed shortly after stimulus onset in all four maps. The frequency, latency and duration of these similarity effects depended however on the exact oscillatory signal (power, phase) and brain region (V1, IT). In terms of MEG power (Fig. 4*A*,*B*), the latencies (see also Fig. 5) respected the hierarchical order of visual processing (Nowak and Bullier, 1998) with an increase in representational similarity in the lower (<20 Hz) frequency bands occurring around the evoked response first for RSA with V1, and ~20–30 ms later for RSA with IT (paired *t* test against 0, FDR corrected, $\alpha = 0.05$). This latency difference is similar to that reported in Cichy et al. (2014), where the peak correspondence between the average MEG signal and V1 activity occurs ~30 ms before the peak with IT activity. The onset response in the V1 RSA map also consisted of high γ frequencies (>70 Hz), whereas this high-γ activity was not observed in the IT RSA map. This is compatible with recent findings showing that γ oscillations in early visual cortex are particularly prominent for certain stimuli, yet can be entirely absent for others (Hermes et al., 2015). A sustained low-β (20 Hz, 200–500 ms) and an offset high-β (30 Hz, ~600 ms) response also corresponded to V1 representations, although neither of these effects were observed in the IT RSA map (see also Fig. 5). In terms of stimulus representations in the MEG oscillatory phase (Fig. 4*C*,*D*), after an initial broadband (3–100 Hz) transient peak at stimulus onset corresponding to V1 representations, stimulus information carried by sustained oscillatory phase resembled IT representations in the low (<20 Hz) and high frequency (60 Hz) bands, and this resemblance persisted until the end of the trial. Phase representations corresponding to V1 patterns were observed again around stimulus offset, at α (~10 Hz) and β (20–30 Hz) frequencies (see also Fig. 5), in line with the known involvement of V1 neurons in OFF responses (Jansen et al., 2019).

These results thus suggest that different oscillatory components correspond to different brain regions at different TF points. However, since the RDMs in the TF space are heavily correlated with each other, it is difficult to ascertain from this analysis which power/phase features co-vary, and which effects occur independently. To better interpret the results shown in Figure 4, we turned to a clustering analysis. The clustering analysis allowed us to reduce the dimensionality of the dataspace and to determine which oscillatory signals occurred jointly, and which are independent. We performed k-means clustering jointly on the power and phase RDMs. That is, each RDM (one for each time point, frequency, and phase/power signal), was considered as an input data point for the clustering analysis, which returned the corresponding cluster index assigned to each point (see Fig. 2). The results of the clustering analysis for $k = 7$ clusters (the optimal number of clusters for our dataset) are shown in Figure 6 (see also Fig. 7). The first cluster (ranked by smallest average distance from cluster centroid) corresponded to early

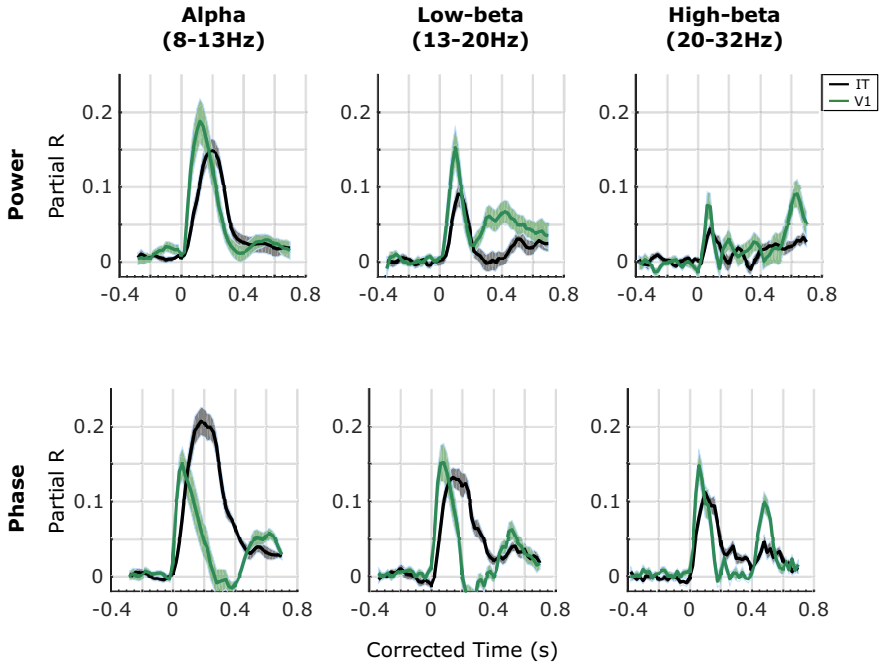

Figure 5. Profile of the results of the RSA with V1 (green lines) and IT (black lines) in oscillatory power (top row) and oscillatory phase (bottom row) in different frequency bands. To examine the RSA maps in more detail, we extracted their time courses in different traditional frequency bands: α (8–13 Hz), low-β (13–20 Hz), and high-β (20–32 Hz). In each of these frequency bands, we computed the average *R* values. Since the TF decomposition induces temporal smearing, and the amount of smearing differs for different frequencies, to interpret the latencies of the representational similarities, we corrected for this smearing effect. Specifically, to avoid underestimating the onset latencies, we corrected time by adding half the wavelet window duration at each frequency. Note that the same correction was applied to the two curves compared in each plot. Solid lines are the means across subjects, and the shaded areas correspond to the SEM across subjects.

broadband (0–100 Hz) phase and power RDMs, followed by sustained γ power (>60 Hz), and β power (20–30 Hz) at stimulus offset. The second cluster corresponded to broadband (0–100 Hz) and sustained (0.1–0. 4s) phase effects after stimulus onset, without any noticeable power effects. The third cluster consisted primarily of sustained (0.1–0.6 s) β (10–30 Hz) and low-γ (<60 Hz) power, without any noticeable phase effects. The fourth cluster reflected broadband phase effects (0–100 Hz) at stimulus offset (without associated power effects). The last three clusters (5–7) all displayed prestimulus effects in α-β power, or α or γ phase, characteristic of spontaneous, stimulus-unrelated activity that we did not investigate further (Fig. 7). The clustering analysis performed on the MEG RDMs thus identified four main clusters of power and phase oscillatory components that occurred at different time points and in different frequency bands after stimulus onset.

How do the oscillatory representations in each cluster, and their different time and frequency profiles relate to local processing in V1 and IT as measured by fMRI representations? To address this question, we performed RSA between the cluster centroids and the V1 and IT RDMs. The cluster centroids correlated to different degrees with both V1 and IT (all partial *R* values between 0.12 and 0.49; all significant at $p < 1e\text{-}5$ with a surrogate test; see Materials and Methods). To directly contrast the representational similarity of each brain area (V1, IT) to the cluster centroid, we combined the two RSA partial *R* values into a single scale (see Materials and Methods, Eq. 1). According to this scaling (Fig. 6, insets), the transient broadband phase and power effect with sustained γ power in cluster 1 corresponded best with V1 representations (i.e., low-level). Conversely, the broadband sustained phase effects of cluster 2 corresponded best to IT representations (high-level). The other two clusters (sustained β-γ power in cluster 3, broadband offset-transient phase in cluster 4) had more balanced similarity to both V1 and IT, with a slight inclination toward V1. Thus, transient oscillatory components occurring around stimulus onset correspond more closely to V1 representations, whereas the more sustained components could be either more IT-like, or less localized depending on the frequency of the oscillations. These results thus suggest a complex link between oscillatory representations and local processing in V1 or IT. To try to clarify these relationships we next turned to using DNNs as a template for object representations.

Assessing representational complexity with DNNs

The fMRI RDMs are a representation of the image set in the multi-voxel space of V1 and IT. However, these fMRI representations are static because the fMRI BOLD signal used to construct the RDMs was measured over a period of several seconds. Neuronal activity in these regions, on the other hand, is known to evolve over fairly rapid

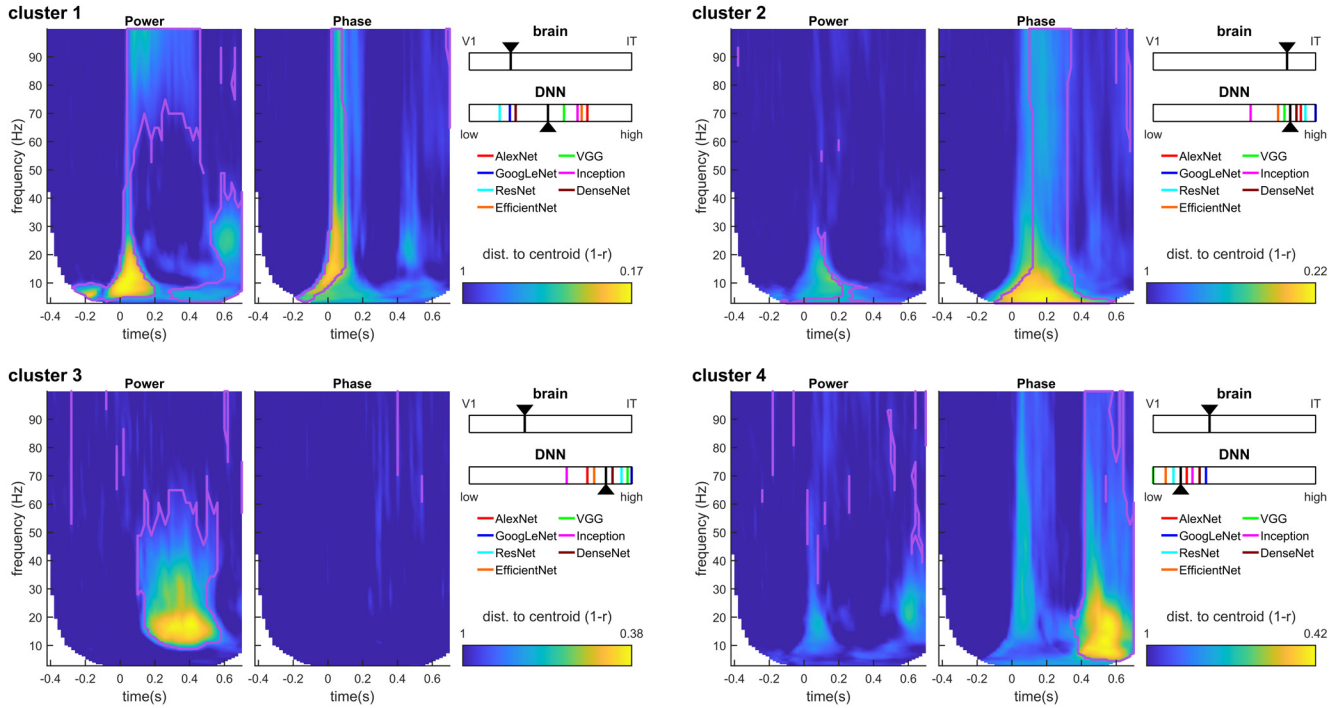

Figure 6. Clustering analysis. K-means clustering was performed on the MEG power and phase RDMs. Each TF plot shows the distance of each RDM from the centroid of the corresponding cluster. The purple lines correspond to the cluster boundaries as returned by the k-means algorithm, indicating that all points within the purple lines are assigned to this specific cluster based on their distance to the different cluster centroids. The distance to centroid (color scale) reflects how "stereotypical" each RDM is for the corresponding cluster (i.e., how close to the cluster centroid), a continuous scale that complements the discrete cluster assignment. For example, although cluster 3 simultaneously encompasses oscillatory power across many frequencies from 10 to 65 Hz, we can see that low-β frequencies (13–20 Hz) are the most stereotypical for this cluster. The insets show the relative degree of RSA between the cluster centroid and V1/IT (top), or the cluster centroid and the DNN layer hierarchy (bottom). For the DNNs, the layer with maximum RSA, normalized by the number of layers in the DNN hierarchy, and averaged across the seven DNN types (colored ticks), was taken as the layer that corresponded to each cluster centroid (black arrowhead).

timescales, on the order of hundreds of milliseconds as a result of feedback and top-down signals (Roelfsema et al., 1998; Lamme and Roelfsema, 2000). The fMRI RDMs are thus limited representations of the image set, potentially mixing low-complexity and high-complexity brain activity from different moments in each trial. Therefore, while it is tempting to interpret the oscillatory signals composing cluster 1 as low-complexity, because they are more V1-like, and those forming cluster 2 as high-complexity (more IT-like), such a conclusion would be premature as it ignores the dynamics of neural responses within and across brain regions, and how these neural responses evolve over different timescales. To obtain a complementary picture of low and high-level object representations, we considered the representations of our image set in different layers of feed-forward DNNs pretrained on a large dataset of natural images. To ensure the generality of our results we assessed seven different DNNs: AlexNet (Krizhevksy et al., 2012), VGG16 (Simonyan and Zisserman, 2015), GoogleNet (Szegedy et al., 2015), InceptionV3 (Szegedy et al., 2017), ResNet (He et al., 2016), DenseNet (Huang et al., 2017), and EfficientNet (Tan and Le, 2019). Activity in each layer of these DNNs is not influenced by top-down or recurrent connections, and consequently represents a truly hierarchical evolution in the complexity of image representations, from low to high complexity. Indeed, several studies have suggested that DNN representations approximate the feed-forward cascade of the visual processing hierarchy in the brain (Khaligh-Razavi and Kriegeskorte, 2014; Cichy et al., 2016). Performing RSA between MEG oscillatory RDMs and DNN layer RDMs should thus reveal which features of the MEG oscillatory representations correspond to low-complexity vs high-complexity object representations.

An RDM was obtained for several representative convolutional layers of the seven DNNs. RSA was then performed between the cluster centroids of the MEG RDMs and the DNN RDMs. For each cluster and DNN, the layer with maximum RSA was determined, and scaled between 0 (lowest layer, low-complexity information) and 1 (highest layer, high-complexity information) based on the number of layers in the DNN hierarchy. Despite notable differences between the seven DNNs, the analysis revealed that clusters 2 and 3 mapped best to higher DNN layers, cluster 1 to intermediate layers, and only cluster four had similarity to lower layers. This is in stark contrast with the results of fMRI RSA, which had ranked clusters 2, 4, 3, and 1 in order of decreasing complexity. The most striking

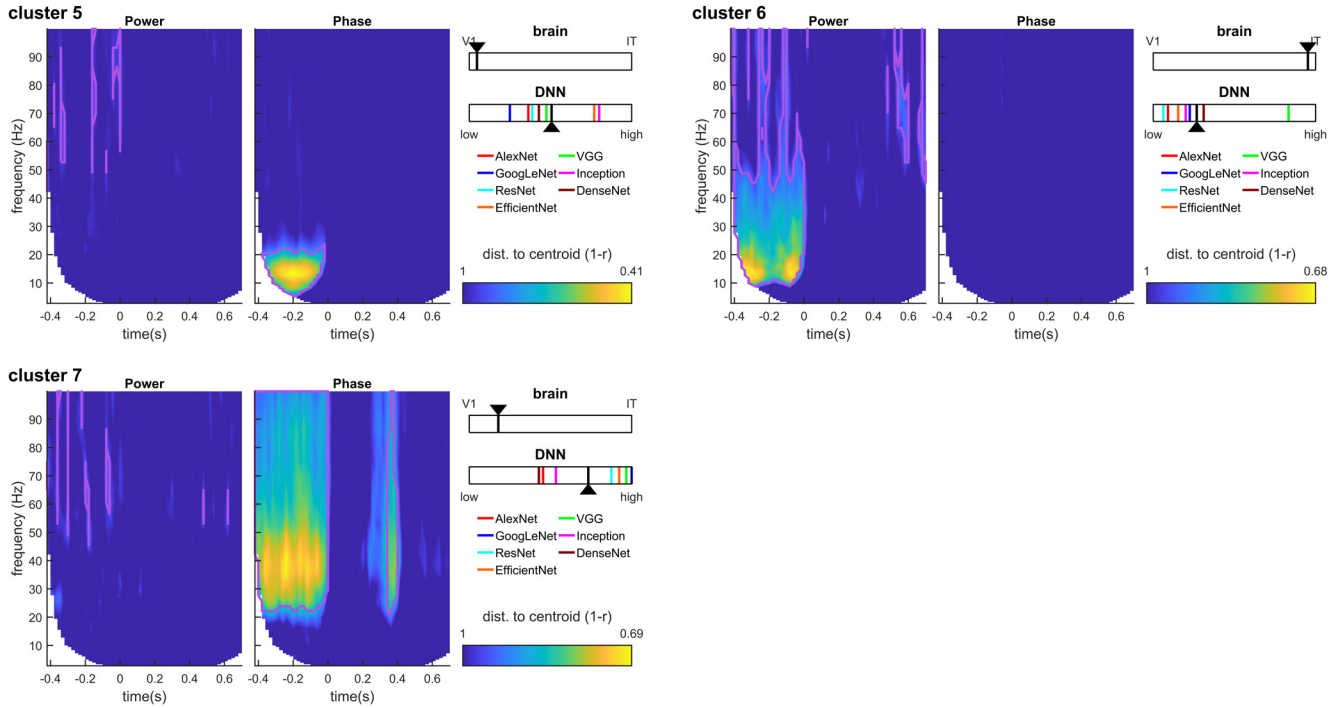

Figure 7. Clustering results for clusters 5–7. We identify these clusters as noise components because (1) their distance to the cluster centroid is typically higher than for other clusters, and (2) they mainly map onto prestimulus oscillatory activity. Prestimulus oscillations, while accounting for a sizeable portion of the (notoriously noisy) MEG signal variance, cannot possibly encode the identity of a stimulus that has not been presented yet. Prestimulus α is a well-studied oscillatory component reflecting the attention state of the observer, and whose phase is known to modulate the subsequent ERP amplitudes and latencies; as such, it is not surprising that the phase of this oscillatory component would induce a separate cluster of RDM patterns (cluster 5). Similarly, prestimulus α-β power (cluster 6) and γ phase (cluster 7) could reflect preparatory attention or motor signals (including muscular artifacts) not related to stimulus identity. Notations as in Figure 6. Note that for consistency with the previous figure, we continue to report the V1-IT RSA scaling value in the insets; however, the corresponding correlation values were systematically lower for these clusters, and should thus be interpreted with caution (V1 partial correlations for clusters 5–7: [0.04, 0.00, 0.08], IT partial correlations: [0.00, 0.04, 0.03]). In comparison, V1 partial correlations for clusters 1–4 ranged from 0.14 to 0.43, and IT partial correlations from 0.12 to 0.44. Similarly, we also report the DNN layer for which the correlation to the cluster centroid was maximal; however, this maximal correlation was consistently lower than in the previous figure, as expected for a prestimulus component (peak correlations averaged across DNNs for clusters 5–7: [0.02, 0.03, 0.16], compared with values ranging from 0.24 to 0.45 for clusters 1–4).

difference is obtained for cluster 3 (sustained β-γ power): a high-complexity representation according to DNNs, but closer to V1 than to IT according to fMRI. Based on the logic above, this cluster is likely to reflect feed-back signals that carry high-complexity object information (visible in high DNN layers) down to lower brain regions (visible in V1 BOLD signals).

Discussion

Our results (summarized in Fig. 8) show that MEG oscillatory components at different frequencies carry stimulus-related information at specific times, which can be linked, via RSA, to stimulus representations in different brain regions (V1, IT), and with different representational complexity (as measured by DNNs). Importantly, the representational dynamics of brain oscillations can be very differently expressed by power versus phase signals. At stimulus onset and offset, broadband phase transients (possibly related to fluctuations in evoked potential latencies) carry mainly low-complexity or intermediate-complexity information (Fig. 6, clusters 1 and 4). However, during stimulus presentation, sustained phase information is visible across all frequencies, and consistently maps to high-level and high-complexity representations (IT and high DNN layers, cluster 2). Oscillatory power components (clusters 1 and 3) tend to correlate with both V1 and IT fMRI representations (with an inclination toward V1); however, onset-transient low-frequency ($<$20 Hz) power together with sustained high-frequency ($>$60 Hz) power (i.e., cluster 1) correspond best to intermediate DNN layers, whereas sustained β-γ power (20–60 Hz) clearly maps to the highest DNN layers (cluster 3).

It is important to note that some of the TF components revealed here could be specific to the conditions of our experiment. For example, brain oscillations are often modulated by the subjective state, the participant's attention or the task instructions. As such, it is likely that different oscillatory patterns would be obtained for tasks involving active behavior rather than passive viewing of the images. Similarly,

spurious muscular activity (including ocular saccades or microsaccades) could be an important contribution to the observed TF components (Yuval-Greenberg et al., 2008), as long as this activity would systematically differ between the different stimulus classes.

Our findings complement those reported in an earlier study that contrasted oscillatory power measured from intracranial electrodes with object representations across various layers of AlexNet (Kuzovkin et al., 2018). These authors linked γ power in lower visual areas to object content in lower DNN layers, while γ power in higher visual areas (as well as θ-band activity) was related to higher DNN layers. Possible differences with our own observations could be explained by the spatial scale of the electro-magnetic signals recorded (more localized in intracranial electrodes, more widespread in MEG), our consideration of phase in addition to power components, our use of 6 other DNNs in addition to AlexNet, or by more specific aspects of our analysis pipeline such as the DNN layer centering procedure (Fig. 3) or the K-means clustering (Fig. 2).

In our study, we found no simple mapping between low/high-level (or low/high-complexity) representations and oscillatory components (power/phase) or frequency. Both low-frequency (θ, α) and high-frequency (β, γ) oscillatory signals can carry either low- or high-level/complexity representations at different times (e.g., clusters 2 vs 4). Similarly, both phase and power signals can carry either low or high-level representations (e.g., clusters 1 vs 3). The picture that emerges is a rather intricate one, in which successive interactions between different oscillatory components in different brain regions and at different frequencies reflect the different stages of neural processing involved in object recognition.

How can the representational content of a given oscillatory component (phase or power) be interpreted in functional terms, at the level of neural populations and their interactions? For example, a sustained phase component (such as the broadband component in cluster 2) means that for some extended period of time (here, roughly between 150 and 350 ms), the exact phase of oscillatory signals (here, across multiple frequency bands, from δ to high γ) will systematically vary with the stimulus identity. Note that this is not about increased phase locking, but about the phase values themselves, and their differences between images. Such systematic phase differences between stimulus classes could arise if the underlying oscillatory processes come into play with different delays, e.g., as a result of information routed through slightly distinct circuits. As for oscillatory power, the differences (for example, the sustained β power differences summarized in Cluster 3) would imply that some image classes tend to result in higher amplitudes and others in weaker amplitudes. This could arise, for example, in a scenario where the oscillation is selectively triggered by certain images (those that the neural population is selective to, e.g., animate vs inanimate, natural vs man-made, etc.).

Our results highlight the importance of complementing MEG-fMRI RSA with another measure of representational

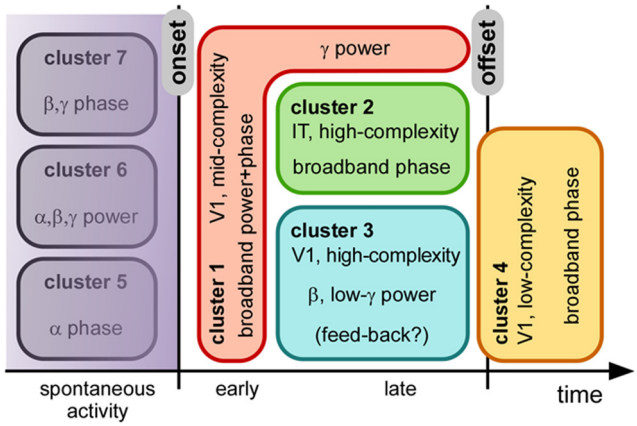

Figure 8. Illustrative summary. The different clusters identified are plotted schematically as a function of time, and their main oscillatory characteristics (frequency band, power/phase) are indicated, together with the corresponding brain region (V1/IT) and the corresponding DNN layer (low/mid/high-complexity).

content such as feed-forward DNNs (Cichy et al., 2016; Bankson et al., 2018; Hebart et al., 2018; Khaligh-Razavi et al., 2018). fMRI BOLD signals are often analyzed such that they reflect a single static representation. Thus, they cannot distinguish dynamics in local patterns as, for example, early feedforward and later feedback activity. By design, feedforward DNN layers cannot be dynamically influenced by feedback signals, and could be considered to provide a template for low-complexity versus high-complexity representations during the different stages of image processing. Perhaps the best illustration of this notion stems from the discrepancy between fMRI and DNN RSA for MEG cluster 3, which suggests that sustained β-γ power during stimulus presentation could reflect feedback signals: best corresponding to V1 fMRI activity (low-level), but higher DNN layers (high-complexity). Without this additional information (e.g., looking at Fig. 4*A* alone), one might have interpreted sustained β power as a strictly low-level signal. The observed distinction between sustained power effects at lower frequencies (β and low-γ, cluster 3) versus higher frequencies (high-γ, cluster 1) is consistent with a large number of recent studies that reported a functional distinction between γ-band and β-band signals, respectively supporting feed-forward and feedback communication (Fontolan et al., 2014; van Kerkoerle et al., 2014; Bastos et al., 2015; Michalareas et al., 2016). Future work could attempt to separate feedforward from feedback signals (e.g., with backward masking and/or layer-specific fMRI), to confirm the differential contribution of γ-band and β-band oscillatory frequencies to feedforward versus feedback object representations, as determined with RSA.

In addition to their involvement in the transmission of feedforward and feedback signals, several studies have shown that different oscillatory signals can carry distinct information about stimulus properties (Smith et al., 2006; Romei et al., 2011; Schyns et al., 2011; Mauro et al., 2015). Here, we considered whether oscillatory

components in different frequencies correspond to lower-complexity or higher-complexity stimulus processing stages. Our results suggest that most oscillatory brain activity, at least at the broad spatial scale that is measured with MEG, reflects already advanced stimulus processing in object detection tasks. This result can be seen in Figure 6, where most oscillatory components are more related to higher-level DNN layer representations, with the exception of the offset-transient (cluster 4). Indeed, one might have expected that stimulus representations at both stimulus onset and offset are more reflective of transient low-level and low-complexity processing. However, while both onset and offset signals (clusters 1 and 4) are better matched to V1 than IT (low-level; see also Fig. 4*C*,*D*), in terms of DNN activations the offset-transient (cluster 4) appears to be of much lower-complexity and the onset-transient of higher-complexity (cluster 1). A tentative explanation could be that the continued presence of the stimulus after the onset-transient supports a rapid refinement of object representations, which would not be the case for the offset-transient (because the stimulus is absent from the retina). Indeed, it is remarkable that, aside from this offset-transient broadband phase activity (cluster 4), no other oscillatory signal was found to reflect low-level DNN layers (i.e., low-complexity information).

One possible explanation for the relative dearth of oscillatory components reflecting low-level DNN layers could be that neural oscillations are a circuit-level property, rather than a single-neuron property; this could provide a better match for high-level DNN layers that pool across large numbers of inputs. In any case, such a bias, if it exists, would only be relative, as we did find at least one oscillatory component (related to cluster 4) that better matched low-level DNNs.

In conclusion, our results help characterize the representational content of oscillatory signals during visual object perception. By separately considering hierarchical level (V1/IT) and representational complexity (based on DNNs), we narrow the gap between whole-brain oscillations and visual object representations supported by local neural activation patterns.

References

Bankson BB, Hebart MN, Groen IIA, Baker CI (2018) The temporal evolution of conceptual object representations revealed through models of behavior, semantics and deep neural networks. Neuroimage 178:172–182.

Bastos AM, Vezoli J, Bosman CA, Schoffelen JM, Oostenveld R, Dowdall JR, De Weerd P, Kennedy H, Fries P (2015) Visual areas exert feedforward and feedback influences through distinct frequency channels. Neuron 85:390–401.

Cichy RM, Pantazis D, Oliva A (2014) Resolving human object recognition in space and time. Nat Neurosci 17:455–462.

Cichy RM, Khosla A, Pantazis D, Torralba A, Oliva A (2016) Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. Sci Rep 6:27755.

Fontolan L, Morillon B, Liegeois-Chauvel C, Giraud AL (2014) The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. Nat Commun 5:4694.

Fries P (2005) A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn Sci 9:474–480.

He K, Zhang X, Ren S, Sun J (2016) Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 770–778..

Hebart MN, Bankson BB, Harel A, Baker CI, Cichy RM (2018) The representational dynamics of task and object processing in humans. Elife 7:e32816.

Hermes D, Miller KJ, Wandell BA, Winawer J (2015) Stimulus dependence of gamma oscillations in human visual cortex. Cereb Cortex 25:2951–2959.

Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 2261–2269.

Jensen O, Mazaheri A (2010) Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Front Hum Neurosci 4:186.

Jansen M, Jin J, Li X, Lashgari R, Kremkow J, Bereshpolova Y, Swadlow HA, Zaidi Q, Alonso JM (2019) Cortical balance between ON and OFF visual responses is modulated by the spatial properties of the visual stimulus. Cereb Cortex 29:336–355.

Kar K, Kubilius J, Schmidt K, Issa EB, DiCarlo JJ (2019) Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. Nat Neurosci 22:974–983.

Khaligh-Razavi SM, Kriegeskorte N (2014) Deep supervised, but not unsupervised, models may explain IT cortical representation. PLoS Comput Biol 10:e1003915.

Khaligh-Razavi SM, Cichy RM, Pantazis D, Oliva A (2018) Tracking the spatiotemporal neural dynamics of real-world object size and animacy in the human brain. J Cogn Neurosci 30:1559–1576.

Kiani R, Esteky H, Mirpour K, Tanaka K (2007) Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. J Neurophysiol 97:4296–4309.

Kietzmann TC, Spoerer CJ, Sörensen LKA, Cichy RM, Hauk O, Kriegeskorte N (2019) Recurrence is required to capture the representational dynamics of the human visual system. Proc Natl Acad Sci USA 116:21854–21863.

Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis - connecting the branches of systems neuroscience. Front Syst Neurosci 2:4.

Krizhevksy A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25:1097–1105.

Kuzovkin I, Vicente R, Petton M, Lachaux JP, Baciu M, Kahane P, Rheims S, Vidal JR, Aru J (2018) Activations of deep convolutional neural networks are aligned with gamma band activity of human visual cortex. Commun Biol 1:107.

Lamme VA, Roelfsema PR (2000) The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci 23:571–579.

Mauro F, Raffone A, VanRullen R (2015) A bidirectional link between brain oscillations and geometric patterns. J Neurosci 35:7921–7926.

Michalareas G, Vezoli J, van Pelt S, Schoffelen JM, Kennedy H, Fries P (2016) Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. Neuron 89:384–397.

Nowak L, Bullier J (1998) The timing of information transfer in the visual system. In: Cerebral cortex (Kaas JH, Rockland K, and Peters A, eds), pp 205–241. New York: Plenum.

Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011:156869.

Roelfsema PR, Lamme VA, Spekreijse H (1998) Object-based attention in the primary visual cortex of the macaque monkey. Nature 395:376–381.

Romei V, Driver J, Schyns PG, Thut G (2011) Rhythmic TMS over parietal cortex links distinct brain frequencies to global versus local visual processing. Curr Biol 21:334–337.

Schyns PG, Thut G, Gross J (2011) Cracking the code of oscillatory activity. PLoS Biol 9:e1001064.

Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale visual recognition. Proceedings of the ICLR 2015 conference (International Conference on Learning Representations), San Diego, CA, May 7–9.

Smith ML, Gosselin F, Schyns PG (2006) Perceptual moments of conscious visual experience inferred from oscillatory brain activity. Proc Natl Acad Sci USA 103:5626–5631.

Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, pp 1–9.

Szegedy C, Ioffe S, Vanhoucke V (2017) Inception-v4, Inception-ResNet and the impact of residual connections on learning. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, California, pp 4278–4284.

Tan M, Le Q (2019) Efficientnet: rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning, Vol 97, pp 6105–6114.

van Kerkoerle T, Self MW, Dagnino B, Gariel-Mathis MA, Poort J, van der Togt C, Roelfsema PR (2014) Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. Proc Natl Acad Sci USA 111:14332–14341.

Yuval-Greenberg S, Tomer O, Keren AS, Nelken I, Deouell LY (2008) Transient induced gamma-band response in EEG as a manifestation of miniature saccades. Neuron 58:429–441.